



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

CENTRO UNIVERSITARIO UAEM VALLE DE MÉXICO

El Comercio electrónico: análisis del consumidor en México. Utilizando datos de corte transversal aplicando la técnica econométrica modelos logit.

TESIS

Que para obtener el Título de

LICENCIATURA EN ACTUARÍA

Presenta

C. Carolina Suárez Cruz

Asesor: Dra. en E. Jéssica Gámez Arroyo

Co-Asesor: Dr. en E. Eduardo Rosas Rojas



Atizapán de Zaragoza, Edo. de Méx., septiembre de 2024

RESUMEN

La presente investigación tiene por objetivo analizar el comportamiento y hábitos de compra de los consumidores que residen en México en el comercio, por medio de variables sociodemográficas como lo es la edad, sexo, escolaridad, estado civil etc., variables relacionadas con gustos y preferencias, además de contemplar variables económicas. Todo bajo una perspectiva del enfoque de la teoría del consumidor de la microeconomía y con la metodología de los modelos econométricos logit.

Con la llegada de la pandemia por COVID-19 a principios del año 2020, el comercio electrónico tomó un papel importante para satisfacer necesidades y convertirse en una forma de reactivar la economía, durante este periodo de pandemia, en donde, consiguió un crecimiento de 300% en Latinoamérica que se estimaba llegaría entre 5 y 10 años de acuerdo con la revista Forbes (2021) y un crecimiento del 81% en México (AMVO, 2021).

Es por esto, que apoyados en la información de corte transversal que brinda la Encuesta Nacional de Disponibilidad y Uso de las Tecnologías de Información en los Hogares (ENDUTIH) de 2021 se plantea la metodología ya mencionada de modelos logit con la que se determina la probabilidad de compra de un mexicano en el comercio electrónico, dado que la encuesta proporciona datos que permiten conocer la disponibilidad y en qué se utilizan las tecnologías dentro del hogar y a nivel usuario. Para así brindar resultados a la industria del Ecommerce, logrando anticipar las decisiones de consumo por parte de la población.

ABSTRACT

The objective of this research is to analyze the behavior and purchasing habits of consumers residing in Mexico in commerce, by means of sociodemographic variables such as age, sex, schooling, marital status, etc., variables related to tastes and preferences, as well as economic variables. All this under the perspective of the consumer theory approach of microeconomics and with the methodology of the econometric logit models.

With the arrival of the COVID-19 pandemic in early 2020, e-commerce took on an important role to satisfy needs and become a way to reactivate the economy during this period of pandemic, where it achieved a 300% growth in Latin America that was estimated to reach between 5 and 10 years according to Forbes magazine (2021) and a growth of 81% in Mexico (AMVO, 2021).

Translated with DeepL.com (free version) This is why, based on the cross-sectional information provided by the National Survey of Availability and Use of Information Technologies in Households (ENDUTIH) of 2021, the aforementioned methodology of logit models is used to determine the probability of purchase of a Mexican in e-commerce, given that the survey provides data that allows knowing the availability and use of technologies within the household and at the user level. In order to provide results to the Ecommerce industry, anticipating the consumption decisions of the population.

CONTENIDO

Introducción	1
Capítulo 1. Análisis Del Consumidor. Un Enfoque Microeconómico	4
1.1. Preferencias Del Consumidor	5
1.2. Utilidad Del Consumidor	9
1.3. Restricción Presupuestal	10
1.4. Demanda Individual	13
1.5. El Consumidor en el Ecommerce	17
Capítulo 2. El Comercio Electrónico	19
2.1. Antecedentes Mundiales	19
2.2. Situación Actual Del Comercio Electrónico Y El Impacto De La Pandemia Por Covid-19	24
2.3. Comercio Electrónico En México	26
Capítulo 3. Análisis Estadístico De La Información	31
3.1. Descripción De La Base De Datos ENDUTIH	31
3.2. Análisis Estadístico Y Gráfico De Las Variables	32
Capítulo 4. Aplicación Del Modelo Logit Al Comercio Electrónico	54
4.1. Descripción De Metodología Econométrica: Modelos Logit	54
4.2. Aplicación Del Modelo Logit	60
4.3. Resultados y Discusión Del Modelo Logit	72
Conclusiones	79
Bibliografía	84
Anexos	91
Anexo 1. Lista de variables del modelo con el coeficiente resultante y su conversión en Odd ratios y probabilidades	91

Índice De Gráficas

Gráfica 1. Proporción de personas que hacen uso de internet en México 2021.....	33
Gráfica 2. Tiempo promedio de uso de internet en horas en México 2021.....	35
Gráfica 3. Tiempo en años usando internet por los mexicanos 2021.....	35
Gráfica 4. Nivel socioeconómico en los usuarios de internet en México 2021.....	36
Gráfica 5. Distribución por género en los usuarios de internet en México 2021.	37
Gráfica 6. Usuarios de internet por rangos de edades en México 2021.	38
Gráfica 7. Usuarios de internet por rangos de edades en México 2021.	39
Gráfica 8. Usuarios de internet que utilizan computadora, laptop o tablet México 2021.	40
Gráfica 9. Horas promedio de uso de computadoras, laptop o tablet por usuarios de internet.	40
Gráfica 10. Distribución de usuarios de internet por zonas urbanas o rurales en México. ..	41
Gráfica 11. Distribución de usuarios de internet por tamaño de localidad en México 2021.	42
Gráfica 12. Distribución de usuarios de internet por Nivel de estudios en México 2021....	43
Gráfica 13. Proporción de usuarios de internet que asisten a la escuela en México 2021. ..	44
Gráfica 14. Rol de los usuarios de internet dentro de un hogar en México 2021.....	45
Gráfica 15. Proporción de usuarios de internet económicamente activos en México 2021. 45	
Gráfica 16. Proporción de usuarios de internet económicamente activos en México 2021. 46	
Gráfica 17. Tipo de conexión a internet en los hogares en México 2021	47
Gráfica 18. Usuarios de internet que se conectan por medio de red celular en México 2021.	47
Gráfica 19. Usuarios de internet que se utilizan redes sociales en México 2021.....	48
Gráfica 20. Usuarios de internet que se utilizan Facebook en México 2021.	49
Gráfica 21. Usuarios de internet que se utilizan Whatsapp en México 2021.....	49
Gráfica 22. Usuarios de internet que se utilizan Instagram en México 2021.....	50
Gráfica 23. Usuarios de internet que se utilizan Messenger en México 2021.....	50
Gráfica 25. Usuarios de internet que se utilizan Twitter en México 2021.	51
Gráfica 24. Usuarios de internet que se utilizan YouTube en México 2021.....	52

Gráfica 26. Usuarios de internet que realizaron una compra por medio de internet en México 2021.....	52
Gráfica 27. Modelo logístico.....	523

Introducción

La presente investigación tiene por objeto de estudio analizar el comportamiento y hábitos de compra de los consumidores que residen en México en el comercio electrónico (e-commerce), a través de variables sociodemográficas (edad, sexo, escolaridad, estado conyugal), gustos y preferencias y económicas (ingresos, clasificación de la población en PEA¹ y PNEA², mercado laboral), bajo una perspectiva del enfoque de la teoría del consumidor de la microeconomía y con la metodología de los modelos econométricos logit.

Con el paso del tiempo las Tecnologías de la Información y Comunicación (TIC) han tomado un papel importante en la vida cotidiana del ser humano. Este avance trajo consigo la evolución del comercio, abriendo paso al ahora conocido comercio electrónico o e-commerce que se define como cualquier forma de transacción comercial de bienes y servicios, en la cual las partes interactúan de forma electrónica, en lugar de hacerlo a través de intercambios físicos. Según la Oficina de Promoción de la Sociedad de la Información (ISPO) de la Comisión Europea: el comercio electrónico consiste en realizar electrónicamente transacciones comerciales. Está basado en el tratamiento y transmisión electrónica de datos, incluidos texto, imágenes y video (Kabba, 2008; citado en Basantes, 2016).

Las TIC son la columna vertebral del comercio electrónico, su importancia es vital para muchas actividades de la vida y del acontecer económico. De aquí se deriva el internet, que es el alma del comercio electrónico, en cualquiera de sus modalidades (Oropeza, 2018; citado en Medel, 2020).

El uso de cualquier forma de comercio electrónico ha ido cambiando en la última década, la tecnología en hardware y software se va transformando al igual que la forma de hacer negocios, buscando una manera de optimizar los procesos, brindando un mejor servicio y mejor calidad en todo aspecto por tal motivo las formas de comprar o vender en línea es tan diverso actualmente (Hernández y Mendoza, 2018). Con esto en mente, tras la pandemia por COVID-19 a principios del año 2020, la población tuvo que mantenerse en contingencia

¹ PEA (Población Económicamente Activa).

² PNEA (Población No Económicamente Activa).

sanitaria y parte de la población entró en cuarentena, aunado a esto, las actividades económicas que no eran de primera necesidad se vieron forzadas a detener sus rutinas. Ante este suceso el comercio electrónico por la forma de compra rápida y entregas a domicilio tomó un papel importante, logrando satisfacer necesidades y se volvió una forma de reactivar la economía (Forbes, 2021) logrando un crecimiento de un 300% en Latinoamérica durante la pandemia que se estimaba llegaría entre 5 y 10 años de acuerdo con la revista Forbes (2021).

De acuerdo con Statista (Statista, 2020 citado en BlackSip, 2020) los ingresos de comercio electrónico en países líderes de Latinoamérica son los siguientes:

Tabla 1. Ingresos por ventas totales de comercio electrónico en países de Latinoamérica en 2018 y 2019. Cifras en millones de dólares

País	2018	2019
Colombia	\$6.000	\$8.000
Argentina	\$11.000	\$19.000
México	\$18.000	\$28.000
Brasil	\$28.000	\$39.000

Fuente: (Statista, 2020 citado en BlackSip, 2020)

Las grandes categorías y actividades del comercio electrónico en el mundo están dominadas sustancialmente por cinco grandes categorías: ‘comida y cuidado personal’, ‘muebles y electrodomésticos’, ‘juguetes y hobbies’, ‘electrónica de consumo (tecnología)’ y moda (ropa y calzado)’. Statista separa estas categorías por subcategorías puntuales, especificando la tendencia de compra mundial para cada uno de ellos, balanceando cifras de finales de 2018 y principios de 2019. (Statista, 2020; BlackSip, 2020). En México de acuerdo con un estudio de la Asociación Mexicana de Venta Online (AMVO,2021), las tendencias de consumo se centran en las categorías de comida, medicamentos, tecnología, seguidos de medios y entretenimiento.

Tras ver el crecimiento que ha tenido el comercio electrónico en los últimos años, se ha vuelto necesario analizar a los consumidores para poder identificar las tendencias de consumo, así como patrones del comportamiento de los compradores apoyados con la información de corte

transversal que brinda la Encuesta Nacional de Disponibilidad y Uso de las Tecnologías en los Hogares (ENDUTIH) de 2021, dado que nos proporciona información sobre dónde y qué compra la población, así como características socioculturales.

Se propone plantear un modelo logit para determinar la probabilidad de compra de un mexicano por medio del comercio electrónico, y así brindar resultados para la industria del e-commerce, logrando prever las decisiones de consumo por parte de la población.

Capítulo 1. Análisis Del Consumidor. Un Enfoque Microeconómico

En este primer capítulo se abordan los factores que intervienen en el consumidor que lo llevan a tomar una decisión en el momento de elegir entre dos o más bienes y servicios. Por lo que se revisan elementos relevantes como la teoría del consumidor, demanda, utilidad, preferencias, restricción presupuestal, precios, entre otras, con el fin de comprender la forma en la que un consumidor racional realiza la elección de compra de entre un conjunto de planes de consumo, de acuerdo con sus necesidades y respetando la cantidad de ingreso que tenga disponible.

La economía de un país se integra por miles de consumidores, los cuales toman decisiones individuales en la búsqueda de su máximo provecho (Cue, 2014). Pindyck & Rubinfeld (2009) mencionan que la conducta del consumidor se puede entender mediante los siguientes 3 pasos:

1. Las preferencias de los consumidores: en donde se describen las razones por las cuales los individuos prefieren un bien a otro.
2. Restricciones presupuestarias: Normalmente los consumidores consideran los precios. Por lo que se debe considerar como segundo paso es que los agentes económicos tienen una riqueza limitada que les restringe las cantidades de bienes que pueden adquirir.
3. Elecciones de los consumidores: Una vez que los individuos tienen identificadas su preferencia y su riqueza limitada, comienzan a realizar las elecciones de sus planes de consumo que maximicen su utilidad. En donde dichas combinaciones dependerán de los precios que tienen los bienes, lo que lleva a una mejor comprensión de la demanda, es decir, cómo afecta el precio del bien en las decisiones de las cantidades a consumir.

Dicho lo anterior Cue (2014) menciona que, en principio, la teoría económica establece que los consumidores son libres de elegir lo que quieren consumir en el mercado de acuerdo con sus preferencias, sin dejar de lado la restricción del ingreso monetario con el que cuentan. Bajo esta idea, el consumidor es soberano en el mercado al decidir cuáles son los bienes y servicios que desea adquirir.

1.1. Preferencias Del Consumidor

Cada consumidor tiene preferencias heterogéneas lo que significa que son diferentes entre ellos. A pesar de esto se identifican algunos patrones de consumo entre los individuos, los cuales en muchas ocasiones dependen de los tipos de bienes (de lujo, normales, inferiores, etc.) y de la elasticidad de los mismos. Se establece que dichas preferencias van cambiando con lentitud, por lo que se acepta que son relativamente estables en el tiempo; sin embargo, el uso de internet ha incidido significativamente en los hábitos de consumo, lo que ha generado que las preferencias cambien con mayor velocidad (Martínez, 2008).

Los agentes económicos son expuestos a diferentes planes de consumo, estos son evaluados por el consumidor, quien tendrá un comportamiento para cada plan y podrá decidir cuál de estos planes de consumo es estrictamente preferible en el sentido de cual le ha resultado de mayor agrado (Martínez, 2008). Si el consumidor tiene una preferencia por una cesta antes que otra significa que va a elegir la que más le satisfaga siempre y cuando tenga la posibilidad de adquirirla. Por otro lado, si el individuo al tener dos cestas diferentes se siente satisfecho al mismo nivel por ambas entonces estará mostrando una conducta indiferente (Varian, 2010).

Cuando un consumidor tiene preferencia por uno de los planes de consumo o se muestra indiferente ante estos se puede decir que prefiere débilmente al plan A que al B . Dichas relaciones de preferencias se encuentran relacionadas entre sí, por ejemplo, si A es débilmente preferible que B y B es débilmente preferible que A se concluye que A y B son cestas indiferentes por lo que se dice que A es al menos tan buena como B y viceversa (Varian, 2010).

Las relaciones que se abordaron en párrafos anteriores son denominadas como preferencias del consumidor y satisfacen tres propiedades (Martínez, 2008):

Completas: En este supuesto se plantea que A y B son dos bienes o servicios cualesquiera, el consumidor podrá compararlos en sí y decidir si A es preferible antes que B , si B es preferible antes que A hablando de que existe una inclinación particular hacia una de las dos opciones

que se le presentan. Por otra parte, también puede mostrar indiferencia ante estos dos planes de consumo decidiendo que A y B son preferidos por igual (Nicholson, 2005).

Transitivas: Dentro de este supuesto se plantea que las decisiones del consumidor son consistentes, evitando que exista relaciones de preferencias circulares (Martínez, 2008). Es decir, si afirma que A es preferible antes que B y B es preferible antes que C entonces A es preferible antes C . Este supuesto será acertado si se asume que el individuo toma sus decisiones con información completa (Nicholson, 2005).

Continuas: Este supuesto explica que, si un consumidor prefiere A antes que B , todos aquellos planes de consumo similares al plan A serán también preferidos antes que B . Bajo este supuesto se permite descartar una discontinuidad en las preferencias, lo que ayuda a formular en términos matemáticos (Nicholson, 2005).

De acuerdo con Martínez (2008) se necesitan considerar algunos supuestos adicionales a los que se mencionan anteriormente para poder derivar una teoría de consumo, que son la continuidad, convexidad y monotonía de las preferencias.

- Continuidad: Se tienen dos planes de consumo A y A' , es decir, se considera una secuencia de planes de A . En donde bajo este supuesto de continuidad se puede decir que A es al menos tan bueno A' . Si lo anterior se cumple entonces aquellos planes que sean muy cercanos al plan A serán al menos tan buenos como A' .
- Convexidad: Este supuesto indica que el consumidor dado un plan de consumo compuesto por dos bienes este va a requerir de cantidades más grandes de uno de los dos bienes para sacrificar unidades de otro si se quiere conservar el nivel de satisfacción, ligándose directamente con la tasa marginal de sustitución. En otras palabras, el supuesto se refiere a la tendencia de diversificar de un consumidor.
- Monotonía: En este supuesto exige que el individuo mejore las cantidades adicionales de los bienes (cuanto más, mejor). Este supuesto se divide en monotonía débil que menciona que un plan de consumo A que contiene al menos la misma cantidad de todos los bienes que otro plan A' , entonces es por lo menos tan bueno como este; y la monotonía fuerte en donde considera que un plan de consumo A que contiene al

menos la misma cantidad que el plan A' y más de alguno de los bienes es estrictamente mejor que este (Colell, 1995).

La manera gráfica para poder representar las preferencias del consumidor que se han mencionado en párrafos anteriores es con la curva de indiferencia, la cual representa todas aquellas combinaciones de cesta de consumo disponibles que reportan el mismo nivel de satisfacción para un individuo, por lo que el individuo se mostrará indiferente ante las cestas que conforman dicha curva. La forma de esta curva va a describir en qué medida está dispuesto un consumidor a sustituir un bien por otro (Pindyck & Rubinfeld, 2009).

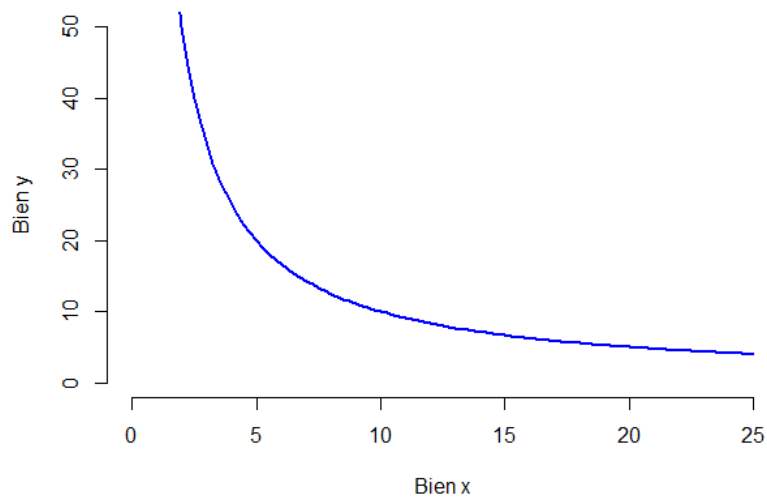


Figura: 1 Curva de indiferencia (Nicholson, 2005)

Derivado de la curva de indiferencia surge el mapa de preferencias el cual se va a conformar de una serie de curvas de indiferencia. Cada curva representa las combinaciones que le reportan la misma utilidad al consumidor, es importante destacar que una curva de indiferencia más alta indica una mayor utilidad (Pindyck & Rubinfeld, 2009).

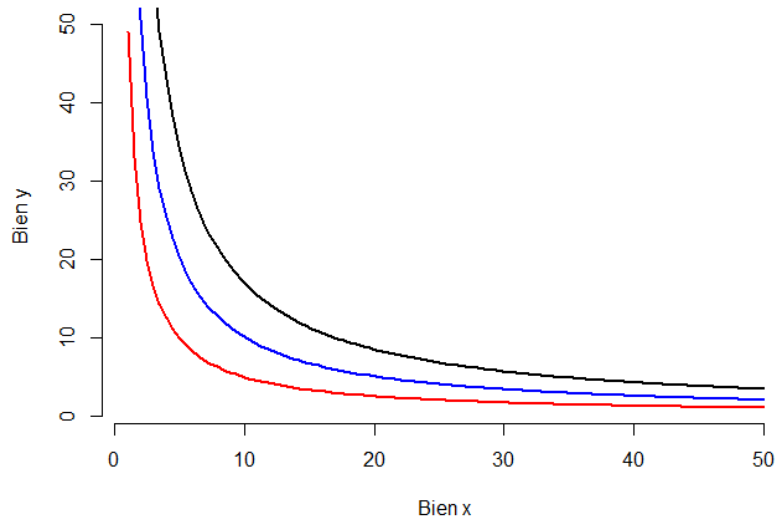


Figura 2. Mapa curvas de indiferencia (Nicholson, 2005)

Pindyck & Rubinfeld (2009) menciona que es importante considerar que las curvas de indiferencia tienen pendiente negativa, esto debido al principio de monotonía que hace referencia a cuanto más, mejor. Si una curva de indiferencia tiene pendiente positiva expresa que un consumidor se mostrará indiferente entre dos planes de consumo, aunque una de ellas tuviese “más” cantidades del bien *A* como del bien *B*.

La magnitud de la pendiente de una curva de indiferencia recibe el nombre de Tasa Marginal de Sustitución (TMS) y es la tasa a que una persona renunciará al bien *A* (que se mide en el eje *y*) para obtener una unidad adicional del bien *B* (medido en el eje *x*) conservando la indiferencia del consumidor, es decir que se conserva sobre la misma curva de indiferencia. Dependiendo la inclinación de esta pendiente se puede ver lo siguiente (Parkin, 2010):

- Si una curva de indiferencia tiene una pendiente *pronunciada* entonces la tasa marginal de sustitución será *alta*. Esto se debe a que el individuo estará dispuesto a renunciar a una gran cantidad del bien *A* para obtener una unidad adicional del bien *B* permaneciendo indiferente.
- Si una curva de indiferencia tiene una pendiente *plana* entonces la tasa marginal de sustitución será *baja*, debido a que el individuo está dispuesto a renunciar a una cantidad pequeña del bien *A* para obtener una unidad adicional del bien *B* permaneciendo indiferente.

La tasa marginal de sustitución decreciente es la clave para las preferencias, dado que es una tendencia general según la cual un consumidor está dispuesto a renunciar a menos del bien A para obtener una unidad adicional del bien B sin perder la indiferente conforme la cantidad de B aumenta (Parkin, 2010).

1.2. Utilidad Del Consumidor

En épocas antiguas tanto filósofos como economistas han hecho referencia a la utilidad como un indicador de bienestar. Era considerada como la forma numérica de medir la felicidad de una persona. Sin embargo, con el paso del tiempo estos economistas no pudieron encontrar una métrica para la utilidad. Tras la reformulación de la teoría de la conducta del consumidor esta conducta se queda en función de las preferencias, en donde la utilidad pasa a ser una forma de describir dichas preferencias (Varian, 2010), convirtiéndose así en un recurso que se utiliza para simplificar la ordenación de las cestas de mercado.

De acuerdo con Parkin (2010) la utilidad es el beneficio o satisfacción que una persona obtiene del consumo de un bien o servicio. Bajo este concepto, las elecciones que realice un consumidor se verán influenciadas por este factor de utilidad, dado que este buscará maximizar su utilidad.

Dentro de la utilidad cabe mencionar dos conceptos importantes: La utilidad total y la utilidad marginal; en donde la utilidad total es el beneficio total que tiene un agente económico de los diversos planes de consumo, dicha utilidad va a depender del nivel de consumo, por lo que el beneficio será favorecido tras existir una mayor cantidad de consumo y la utilidad marginal es el cambio de utilidad total que se deriva de un aumento de una unidad en la cantidad consumida del bien o servicio (Parkin,2010).

La función de utilidad es un instrumento con el cual se asigna un número a los planes o cestas de consumo posibles, en donde el número mayor es asignado al que tenga una mayor preferencia (Varian, 2010). Se expresa de la siguiente manera:

$$utilidad = U(x_1, x_2, \dots x_n; otros\ elementos)$$

En donde x hace referencia a las cantidades de los bienes o servicios que puede elegir un individuo y “otros elementos” es para mantener presente que se mantienen constantes diversos aspectos que se relacionan con el bienestar del consumidor (Nicholson, 2005). La única propiedad importante de una asignación de utilidad es la forma en que ordena las cestas de bienes (Varian, 2010).

Si la ordenación de preferencias es completa, transitiva, reflexiva y continua, entonces las preferencias se pueden representar a través de una función de utilidad continua (Cue, 2005).

Como se menciona en párrafos anteriores la utilidad no puede ser cuantificada, sin embargo, si puede ser ordenar de acuerdo con las preferencias de cada consumidor y es a lo que se denomina utilidad ordinal, debido a que el mismo individuo le asigna la importancia a cada cesta (cervantes, 2016).

Por otro lado, en algunas teorías sobre la utilidad consideran la magnitud como un punto importante, denominándose teoría de utilidad cardinal, en la cual se supone que la magnitud de la diferencia entre la utilidad de dos cestas de consumo tiene un significado. Bajo esta teoría se puede hacer dar una respuesta a preguntas sobre ¿cómo se sabe que a un individuo le gusta el doble una cesta que otra? Además, se pueden proponer diferentes definiciones como que al consumidor le gusta una cesta el doble si está dispuesto a pagar el doble por esta. Al final el propósito de la utilidad cardinal es asignarle niveles de utilidad a cada cesta de manera que la magnitud de los números asignados tuviera una importancia práctica (Varian, 2010).

1.3. Restricción Presupuestal

Uno de los aspectos importantes en la teoría del consumidor es la dualidad entre la cantidad y el precio. Esta dualidad expresa la relación entre dos bienes por un lado y los precios por el otro. Diciendo así que, el consumidor podrá elegir entre maximizar la función de utilidad sujeto a restricciones presupuestales o minimizar el gasto en una serie de bienes, siempre y cuando, dicha función de utilidad permanezca constante (Mora, 2013).

Dentro de la teoría de la demanda se plantea que un consumidor tiene un nivel de renta, el cual tiene planeado gastar en su totalidad en la adquisición de bienes y servicios en el

mercado. Con frecuencia un consumidor realiza una estimación de la cantidad de dinero que se destinará a la adquisición de los planes de consumo. Dimensiona los gastos a los que se va a enfrentar, por lo tanto, esta estimación de gastos se basa o depende de la disposición de recursos económicos, a lo que se le llama restricción presupuestal (Anaya, 2017).

El precio y el ingreso tienen un gran impacto en la magnitud del conjunto de oportunidades de un consumidor, lo que lleva a un término relevante que es el ingreso real en cual de acuerdo con Case (2012) se define como el conjunto de oportunidades a comprar de bienes y servicios reales para los individuos considerando los precios y el ingreso de dinero.

Por lo tanto, si un conjunto de precios disminuye y el ingreso se mantiene constante entonces el costo de oportunidad del consumidor crece mejorando la posición del consumidor, logrando así que el ingreso real aumente aun con la misma cantidad de ingresos. Por otra parte, cuando el ingreso tiene un incremento y a su vez aumentan los precios en mayor medida se dice que el ingreso real del consumidor se disminuye (Case, 2012).

De acuerdo con Nicholson (2005) en la mayoría de los problemas económicos es que no todas las cestas de consumo son factibles. La recta presupuestaria indica todas las combinaciones de A y B con las que la cantidad de dinero gastado es igual a la renta. Por lo tanto, esta recta la podemos expresar de la siguiente forma (Pindyck & Rubinfeld, 2009):

$$I = P_A A + P_B B$$

En donde I es la cantidad de renta fija; A y B son la cantidad adquirida del bien; P_A y P_B son el precio de los bienes A y B respectivamente.

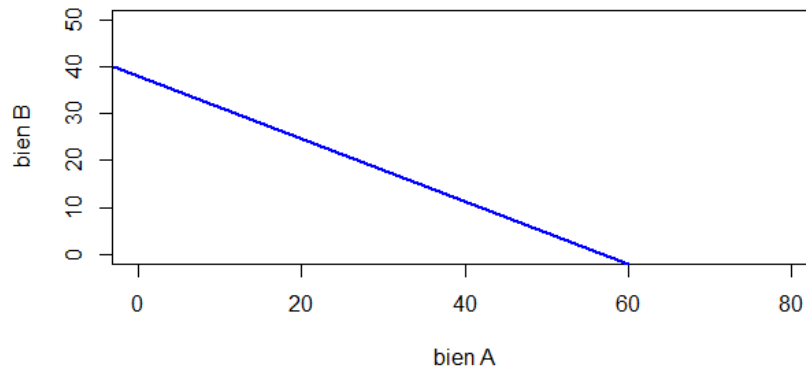


Figura 2. Restricción presupuestal. Elaboración propia

De acuerdo con Varian (2010) bajo la estructura de la ecuación de la recta presupuestaria cuestiona ¿cómo afectaría un cambio en alguno de los elementos que componen dicha recta? Cuando hay un cambio en el ingreso (I) la recta va a tener un desplazamiento en paralelo hacia dentro o fuera dependiendo de si es un incremento o un decremento, esto debido a que el espacio de oportunidades para el consumidor esta dado por el ingreso.

Por otro lado, si hay un cambio en los precios la recta tendrá cambio en cuanto a la inclinación, es decir, la pendiente de la recta tendrá una afectación. Dicha inclinación se deriva de que la cantidad accesible para el consumidor va a cambiar, es decir, si hay un aumento en uno o en ambos bienes el consumidor podrá adquirir menos cantidades con su mismo ingreso; por el contrario, si hay una disminución de precios en esta ocasión el consumido podrá acceder a mayores cantidades de los bienes.

La pendiente de la restricción presupuestaria mide la tasa a la que un consumidor puede cambiar un bien por otro y es negativa, caso similar a lo que sucede con la Tasa Marginal de Sustitución (TMS). Es importante observar que la pendiente de la restricción presupuestaria es igual al precio relativo de los dos bienes (Gregory, 2012).

Una vez que se abordaron los términos de la utilidad y la restricción presupuestal se puede buscar maximizar la utilidad de un consumidor dado su nivel de ingreso a lo que se conoce como equilibrio del consumidor, el cual es un punto en donde se optimizan los recursos con una canasta de bienes y servicios que permita al individuo maximizar su bienestar considerando el nivel de ingreso y los precios de los bienes (Cervantes, 2016).

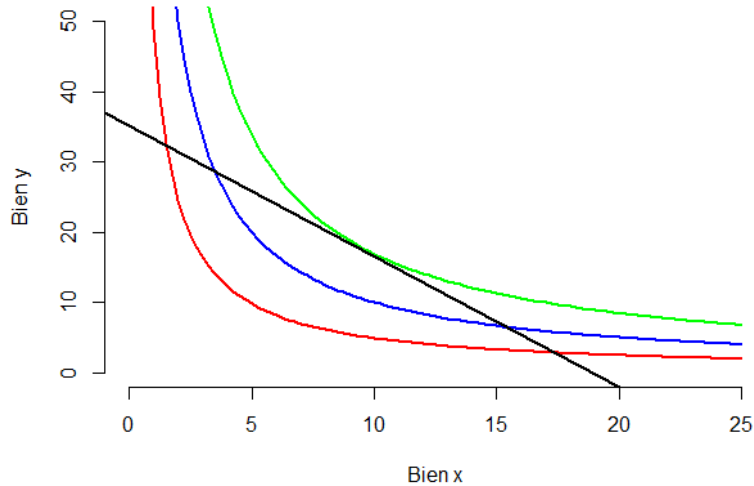


Figura 3. Equilibrio del consumidor. Elaboración propia

Siguiendo con Cervantes (2016) menciona que el equilibrio se consigue cuando la curva de indiferencia es tangente a la recta presupuestaria, es decir, cuando la tasa marginal de sustitución es igual a sus precios relativos, recordando que la tasa marginal de sustitución se puede expresar como el cociente entre la utilidad marginal del bien X y la tasa marginal del bien Y , el cual se puede igualar a la pendiente de la recta de la siguiente forma:

$$TMS_{x,y} = -\frac{UMg_x}{UMg_y} = -\frac{P_x}{P_y}$$

Como se observa en la figura 4 la curva color verde es aquella que logra maximizar la utilidad del consumidor dado su restricción presupuestaria. En las curvas de indiferencia color rojo y azul se encuentran por debajo de la recta presupuestal, por lo que indican que son planes accesibles por el consumidor sin embargo no le brindan el máximo beneficio.

1.4. Demanda Individual

De acuerdo con Mora (2013) la demanda representa la cantidad que un individuo desea comprar de un conjunto de bienes, ya sea que se exprese en términos de precio y el ingreso o en términos de utilidad y precios. Normalmente la demanda óptima o Marshallianas de los bienes depende de los precios y la renta.

Cuando una persona demanda algo significa que lo desea, puede pagarlo y ha hecho un plan definido para comprarlo. Existen varios factores que influyen en las elecciones de los planes de compra, y uno de ellos es el precio, en donde la ley de la demanda establece que: “Si los demás factores permanecen constantes (supuesto de *ceteris paribus*), cuanto más alto sea el precio de un bien menor será la cantidad demandada de dicho bien, y cuanto más bajo sea el precio de un bien mayor será la cantidad demandada del mismo” (Parking, 2010).

La demanda se deriva de a partir de la maximización de la utilidad que también son conocidas como demandas Marshallianas (Anaya, 2017).

Vial y Zurita (2011) mencionan la existencia de dos tipos de demanda:

- La demanda ordinaria o Marshalliana ya mencionada anteriormente la cual consiste en que por un bien A es una función que asigna para cada nivel de ingreso m y precios de los bienes p_1, p_2 , la cantidad consumida de x_A que permite alcanzar el mayor nivel de utilidad posible, dado el conjunto de posibilidades del individuo, la cual se expresa de la siguiente manera:

$$x_A^M = x_A(m, p_1, p_2)$$

En términos generales expresa que la función de demanda ordinaria surge de la maximización de la utilidad individual sujeta a la restricción presupuestaria.

- La demanda compensada o Hicksiana que consiste en que para el bien A es una función que se asigna, para cada nivel de utilidad u y precios de los bienes p_1, p_2 , la cantidad consumida de x_A que permite alcanzar el nivel de utilidad u al mínimo costo posible y se puede expresar cómo:

$$x_A^H = x_A(u, p_1, p_2)$$

La función de demanda compensada surge de la minimización de costos, sujeto a un determinado nivel de utilidad

La curva de demanda del individuo relaciona la cantidad que compra un consumidor de un bien con su precio cuando todos los demás factores que influyen en los planes de compra de

los consumidores permanecen constantes. Esta curva tiene dos importantes propiedades que son (Pindyck & Rubinfeld, 2009):

1. El nivel de utilidad que puede llegar a alcanzar va a variar a medida que nos desplazamos a lo largo de la curva. Es decir, cuanto más bajo es el precio del producto, más alto es el nivel de utilidad.
2. En todos los puntos que conformen la curva de demanda el consumidor maximiza la utilidad satisfaciendo la condición según la cual la relación marginal de sustitución (RMS) de la cesta A por B debe ser igual a la relación de precios de los A y B . Cuando baja el precio de los alimentos, también disminuyen la relación de precios y la RMS.

En términos generales la ecuación de la demanda se puede reducir si se contempla el principio de ceteris paribus y considerando el precio del bien x (P_x) y la cantidad demanda del mismo bien (Q_x^d), resultando la siguiente función (Cervantes, 2026):

$$Q_x^d = F(P_x)$$

Seguendo con Cervantes (2016) explica que dicha función de demanda cumple con algunas características:

1. Esta función se grafica bajo un eje x que representa la cantidad demandada del bien por el consumidor y el eje y que representa el precio de dicho bien.
2. Asociación negativa entre el precio y la cantidad, en otras palabras, ante un aumento de precio disminuye la cantidad y viceversa.
3. Tiene una pendiente negativa, dicha pendiente es a la que se le conoce como ley de la demanda y muestra una relación inversa entre el precio del bien y la cantidad demandada de dicho bien.

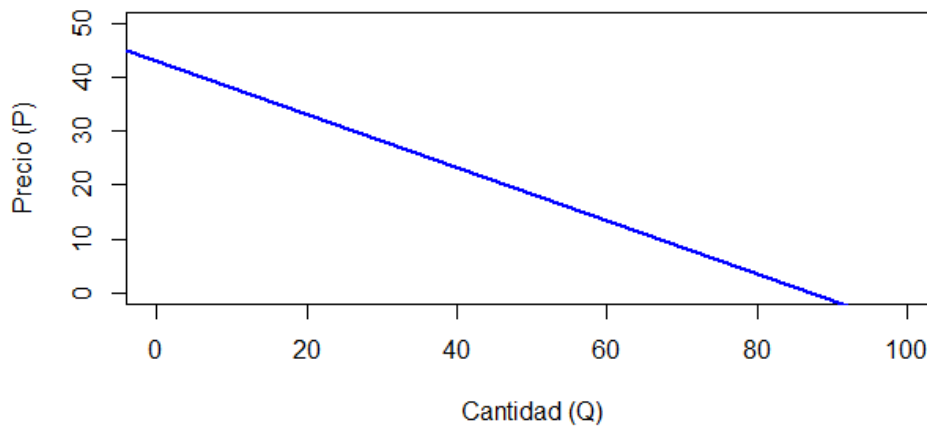


Figura 4. Curva de demanda. Elaboración propia

La curva de demanda va a sufrir un desplazamiento si existe alguna modificación en la cantidad demandada a un precio dado. La cantidad demandada cambia por ejemplo cuando se descubre que un bien cuenta con beneficios médicos, naturalmente los consumidores buscarán adquirirlo a cualquier precio provocando un aumento en la demanda, que gráficamente se observa con un desplazamiento en paralelo hacia el exterior (a la derecha). Por otro lado, si existe un cambio que provoque una reducción de la cantidad demandada el desplazamiento de la curva será hacia la parte interna (a la izquierda) (Gregory, 2012).

Existen diversas formas funcionales de la curva de demanda, no necesariamente es una línea recta, en otras palabras, la relación entre el precio y la cantidad demandada puede tomar otras formas por ejemplo logarítmica o cuadrática (Cuadros, 2012).

Otro de los cambios que afecta a la demanda es el ingreso, pero esto ¿de qué forma afectaría? La forma en la que la demanda se ve afectada se liga al tipo de bien. Cuando la demanda de un bien disminuye tras una disminución del ingreso entonces se dice que es un bien normal. Por el contrario, si la demanda de un bien aumenta ante una disminución del ingreso entonces estamos hablando de un bien inferior (Gregory, 2012).

Por otro parte existe bienes sustitutos en los cuales el aumento de la demanda de un bien provoca la caída de la demanda de otro. Contrario a esto existen bienes complementarios en

los cuales el aumento de la demanda de un bien ocasiona el aumento de demanda de otro bien (Gregory, 2012).

1.5. El Consumidor en el Ecommerce

El comercio electrónico en su modalidad B2C (Business to Consumer) es el que permite la interacción entre consumidor y empresa que oferta productos, utilizando el internet como medio de comunicación y negociación entre ambas partes para concretar la compra y venta. Este modelo de comercio es mayormente utilizado por las personas que desean adquirir sus bienes o servicio en tiendas virtuales (Vega, 2022).

Los clientes o consumidores son las personas físicas que compran bienes o servicios y son los que mantienen la relación con las empresas que se los brindan, constituyendo de esta manera el eje principal de cualquier empresa centrada en el cliente. Por lo tanto, para las empresas que tienen como prioridad al cliente, están en una posición en la que deben definir estrategias que brinden beneficios a largo plazo y priorizar la relación que mantienen con sus consumidores. Además de realizar segmentaciones de mercado para saber dónde y cómo se pueden aplicar las estrategias de fidelización dependiendo del tipo de cliente con el que están tratando (Guadarrama, 2015; citado en Vega, 2022).

Los individuos son exigentes al momento de elegir los productos de sus preferencias, como bien se menciona en la teoría del consumidor saben como catalogar su portafolio de consumo en aquellos que satisfacen ampliamente sus necesidades y cuales se quedan por debajo de las expectativas esperadas referentes a la alta calidad, precio justo y diseño innovador. Estas características se han vuelto obligatorias, elevando de esta forma el nivel de exigencia aun rango mayor (Peña, 2015; citado en Vega, 2022).

Una vez mencionados los puntos anteriores, de acuerdo con datos de la AMVO (2021) en su análisis sobre el ecommerce en México durante 2021, 7 de cada 10 compradores mexicanos están satisfechos con el canal digital y reconocen como beneficios de la compra en internet los siguientes puntos:

- Encontrar productos que no están disponibles en tiendas físicas.
- Recibir compras a domicilio.

- Ahorro de tiempo.
- Permite evitar las aglomeraciones en tiendas físicas.
- Se puede hacer comparación de precio y variedad antes de realizar la compra.

En los puntos mencionados anteriormente, se muestra que el comercio electrónico tiene beneficios adicionales para los consumidores, que les permite simplificar la adquisición de sus bienes y servicios al no tener la necesidad de salir a una tienda física, permitiendo ahorrar tiempos, costos y agilizar la comparación entre diferentes tiendas virtuales tanto en precios como en portafolios de productos, dando lugar a que el individuo pueda evaluar más opciones de compra que le maximicen su beneficio y este dentro de sus límites presupuestarios.

Por la parte de las categorías que prefieren comprar por internet son: comida a domicilio, moda, belleza y cuidado personal y como se muestra en la figura 6, las categorías que presentan crecimientos en términos de preferencia por parte de los consumidores mexicanos en internet comprando 2021 contra 2020 son: comida a domicilio, belleza y cuidado personal, farmacia y supermercado. Mientras que categorías de electrodomésticos, consolas y videojuegos, herramientas, deportes y automotriz están teniendo una reducción de preferencia.

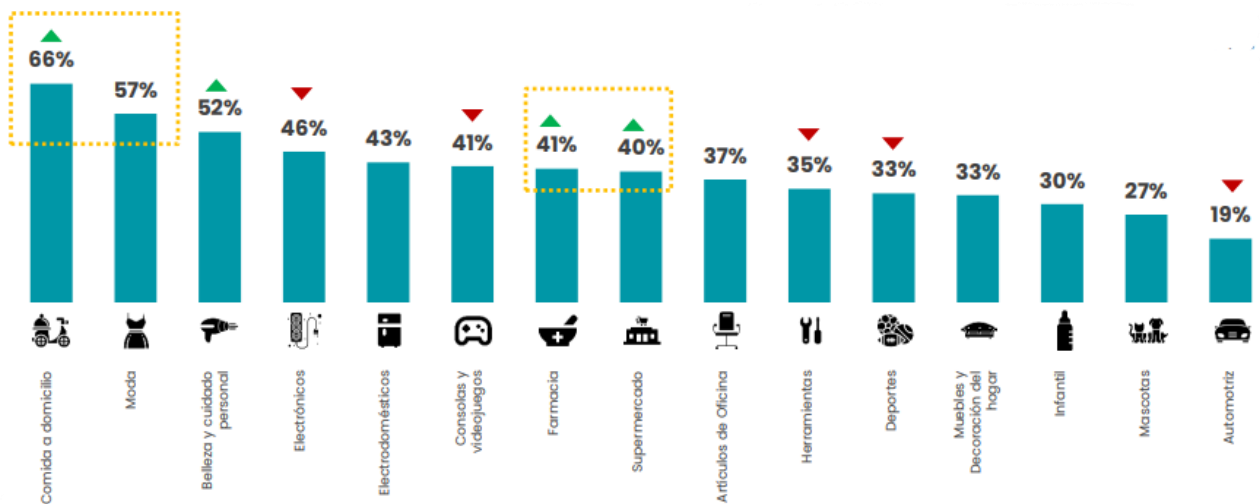


Figura 6. Preferencia de compra por categorías de en México 2020.

Fuente: AMVO (2021)

Capítulo 2. El Comercio Electrónico

En este capítulo se aborda el concepto e historia del comercio electrónico, considerando el contexto histórico a nivel mundial en donde se hace mención de las zonas con mayor desarrollo, las categorías de mayor consumo en el mundo, además de profundizar en la evolución de la adopción de internet por la población mundial y la brecha digital existente en diferentes países, debido a que el internet se considera la columna vertebral del comercio electrónico.

También se abarca el contexto de México en donde se identifican tendencias y estudios realizados por diferentes agencias de investigación, enfocado tanto en la evolución de la industria como tendencias de consumo en compradores.

De igual forma se contemplan un espacio sobre la pandemia por COVID-19 en 2020 que marcó un antes y un después en las ventas por internet debido al cierre de tiendas físicas y confinamientos de la población donde se identifica un crecimiento en el comercio electrónico del 300% de un año a otro, siendo este el crecimiento que se esperaba llegaría dentro de 5 o 10 años (Forbes, 2021)

2.1. Antecedentes Mundiales

Con el paso del tiempo las Tecnologías de la Información y Comunicación (TIC) han tomado un papel importante en la vida cotidiana del ser humano. Este avance trajo consigo la evolución del comercio, abriendo paso al ahora conocido comercio electrónico o Ecommerce que se define como cualquier forma de transacción comercial de bienes y servicios, en la cual las partes interactúan de forma electrónica, en lugar de hacerlo a través de intercambios físicos. Según la Oficina de Promoción de la Sociedad de la Información (ISPO) de la Comisión Europea: el comercio electrónico consiste en realizar electrónicamente transacciones comerciales. Está basado en el tratamiento y transmisión electrónica de datos, incluidos texto, imágenes y video (Kabba, 2008; citado en Basantes, 2016).

Laudon y Guercio (2013) analizan al comercio electrónico como transacciones comerciales digitales que ocurren entre organizaciones, entre individuos y entre organizaciones e

individuos. Consideran como transacciones digitales todas aquellas que se realizan por medio de tecnología digital tomando en cuenta que una gran parte de estas provienen principalmente de internet, la web o aplicaciones móviles.

Las TIC's son la columna vertebral del comercio electrónico, su importancia es vital para muchas actividades de la vida y del acontecer económico. De aquí se deriva el internet, que es el alma del comercio electrónico, en cualquiera de sus modalidades (Oropeza, 2018; citado en Medel, 2020).

De acuerdo con cifras del Banco Mundial (2020) en las últimas décadas se ha visto un incremento en el porcentaje de la población mundial que hace uso de internet. En la figura 6 se puede comprobar dicha tendencia creciente, llegando en 2020 con cobertura del 60% de la población.

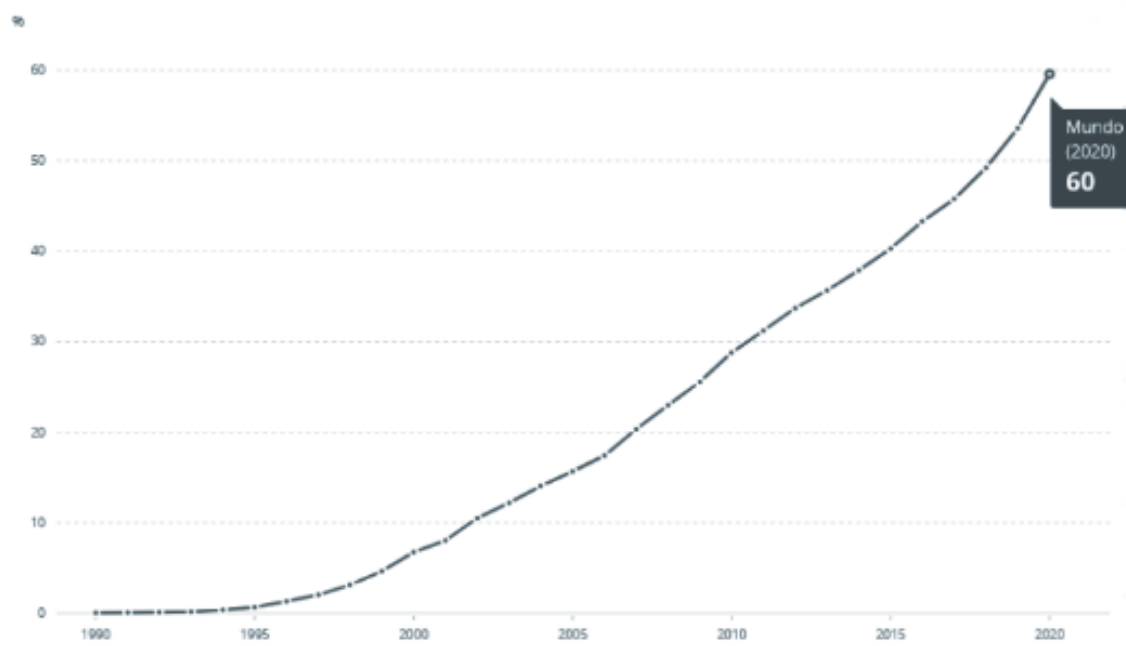


Figura 6. Evolución de cobertura de la población mundial que hace uso de internet periodo de 1990-2020.

Fuente: Banco Mundial (2020)

Existen importantes diferencias, entre países y también a escala nacional, en la capacidad de conexión a Internet y de cómo utilizarla. A esta diferencia la Conferencia de las Naciones Unidas Sobre el Comercio y Desarrollo (UNCTAD, 2021) la define como brecha digital, la

cual está presente en los accesos a banda ancha como la red 3G, 4G, 5G y en la adopción de teléfonos inteligentes.

De acuerdo con la empresa Meltwater citada en Marketing4ecommerce (2023) tras el monitoreo de medios en línea a nivel mundial realizado en enero de 2023 se registran las regiones de Europa Norte, Europa Occidental y América del Norte como las regiones con mayor adopción de internet en su población conservando una cobertura por arriba del 90% como se observa a continuación en la figura 7.

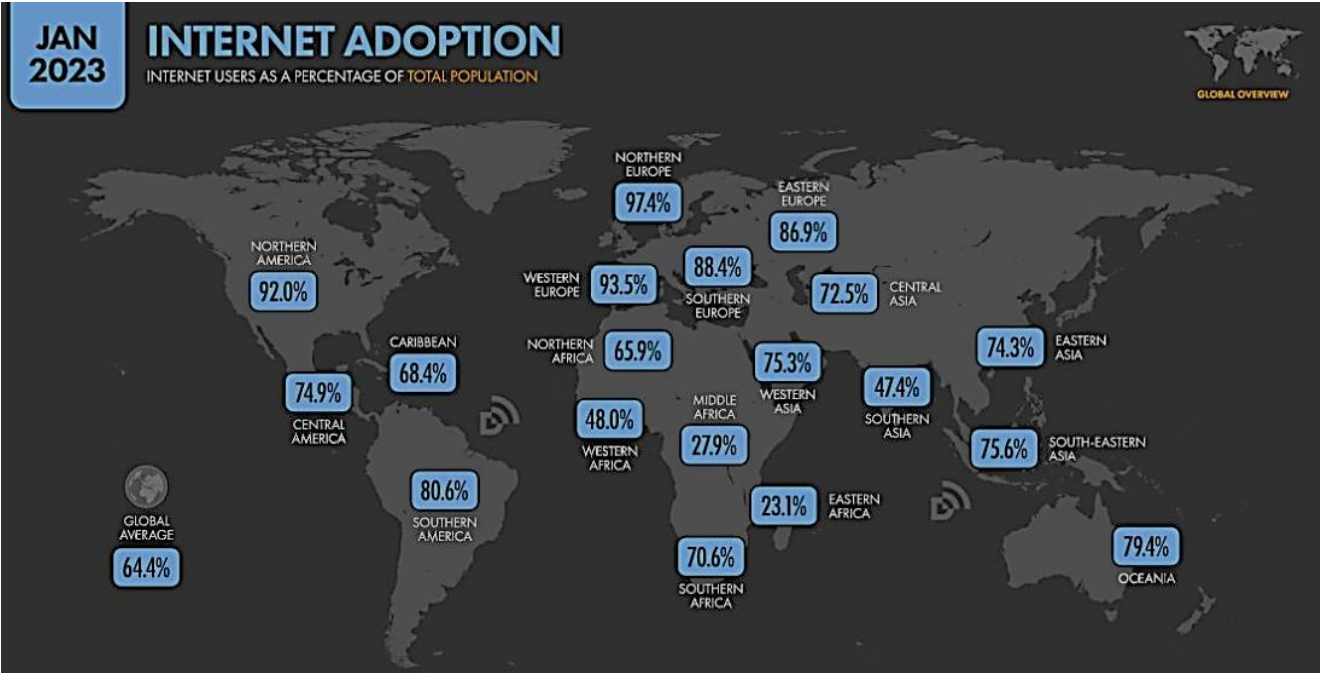


Figura 7 Cobertura de usuarios con acceso a internet por regiones del mundo.

Fuente Meltwater citada en Marketing4ecommerce (2023)

Desde la perspectiva de la cobertura de personas que realizaron una compra por el canal online a nivel mundial y por regiones se muestra en la ilustración 8 que las regiones de Sudamérica, Asia y Europa para 2020 fueron las que tuvieron una adopción mayor de compras en Ecommerce (Statista, 2021). En donde Europa coincide con un mayor número de usuarios que tienen acceso a internet que se mencionan en el párrafo anterior.

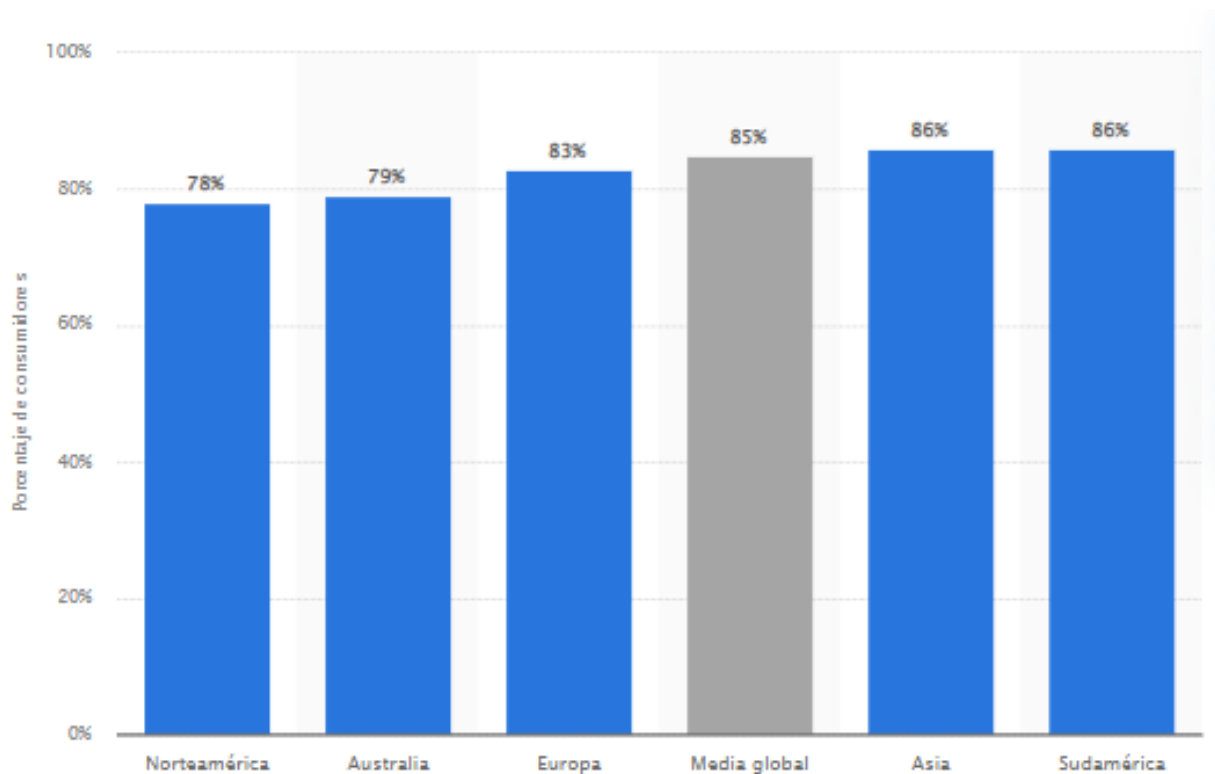


Figura 8 Porcentaje de personas que compraron online a nivel mundial en 2020, por región.

Fuente: Statista 2021

La evolución del comercio electrónico depende en gran parte de la capacidad o la preparación de un país para participar en la economía digital y logra beneficiarse de ella (UNCTAD, 2021).

El uso de cualquier forma de comercio electrónico ha ido evolucionando en la última década, así como la tecnología en hardware y software y la forma de hacer negocios, con el objetivo de buscar una manera de optimizar los procesos, brindando un mejor servicio y mejor calidad en todo aspecto, por tal motivo las formas de comprar o vender en línea es tan diverso actualmente (Hernández y Mendoza, 2018).

De acuerdo con Nielsen (2022), para el año 2021 Asia produjo más del 50% de todas las ventas globales de comercio electrónico con mercados desarrollados como China, Corea del Sur y Japón. Como mercado electrónico pionero, China representó casi una cuarta parte de las ventas minoristas totales de bienes de consumo con la participación de la generación más joven del país. Por otra parte, en el mercado europeo El Reino Unido es el mercado

electrónico más maduro, donde las categorías de alimentos frescos tienen un papel importante en las plataformas online.

Mientras tanto en el continente americano específicamente América Latina se encuentra en una curva de crecimiento pronunciada con las ventas de bienes de consumo online que aumentaron un 35.4% en 2021 (Nielsen, 2022).

Las grandes categorías y actividades del comercio electrónico en el mundo están dominadas sustancialmente por cinco grandes rubros: ‘comida y cuidado personal’, ‘muebles y electrodomésticos’, ‘juguetes y hobbies’, ‘electrónica de consumo (tecnología)’ y moda (ropa y calzado)’. Statista las separa por subcategorías puntuales, especificando la intención de compra mundial para cada uno de ellos, balanceando cifras de finales de 2018 y principios de 2019 (Statista, 2020; BlackSip, 2020).

Se proyecta que el crecimiento del comercio electrónico continúe con su consolidación durante los próximos años debido a que los consumidores ya adaptaron sus hábitos de consumo al comercio electrónico (eMarketer,2021; BlackSip, 2021). Como se puede ver en la figura 9 se proyecta que para 2023 \$5.920 billones de dólares representando un crecimiento del 8.9% respecto al año 2022 (eMarketer, 2023).

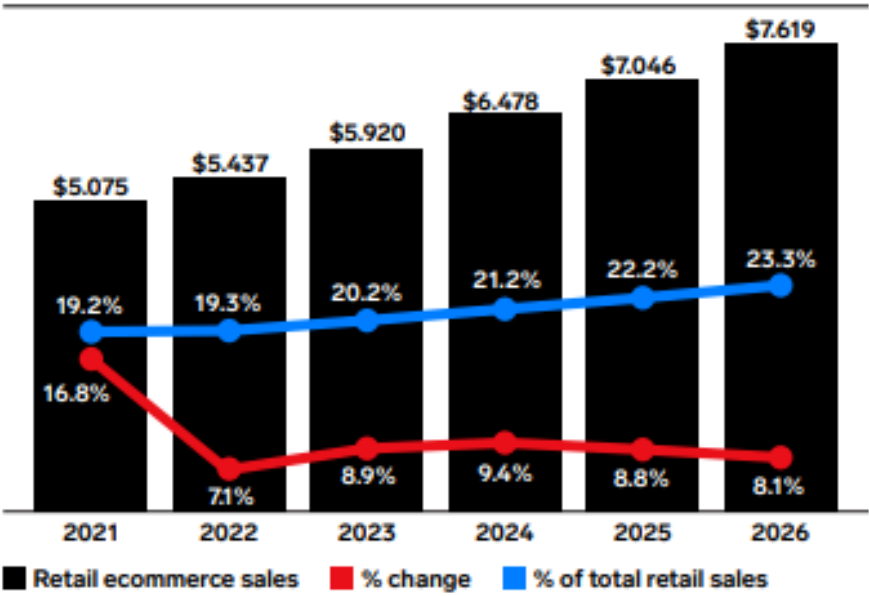


Figura 9 Ventas minoristas mundiales en billones de dólares comercio electrónico, porcentaje de cambio y porcentaje de ventas minoristas totales. Fuente: eMarketer, 2023.

2.2. Situación Actual Del Comercio Electrónico Y El Impacto De La Pandemia Por Covid-19

De acuerdo con la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD) (2021) el comercio electrónico mundial alcanzó un valor de 26.7 millones siendo un aumento del 4% de 2018 a 2019, esta cifra equivale a un 30% del Producto Interno Bruto (PIB) mundial

Continuando con el párrafo anterior, debido a las restricciones de circulación que se implementaron en cada país para hacer frente a la pandemia las compras en línea se vieron impulsadas en varios de ellos siendo la República de Corea la que registró un mayor crecimiento.

Observando la participación de las ventas minoristas realizadas por medio de internet en la figura 10, es notable el crecimiento del año 2019 a 2020, en donde crece 4 puntos porcentuales pasando de 14.3% a un 18.5%. Para posteriormente pasar a una etapa de estabilización en 2021 y 2022 con crecimientos no tan pronunciados (eMarketer, 2023).

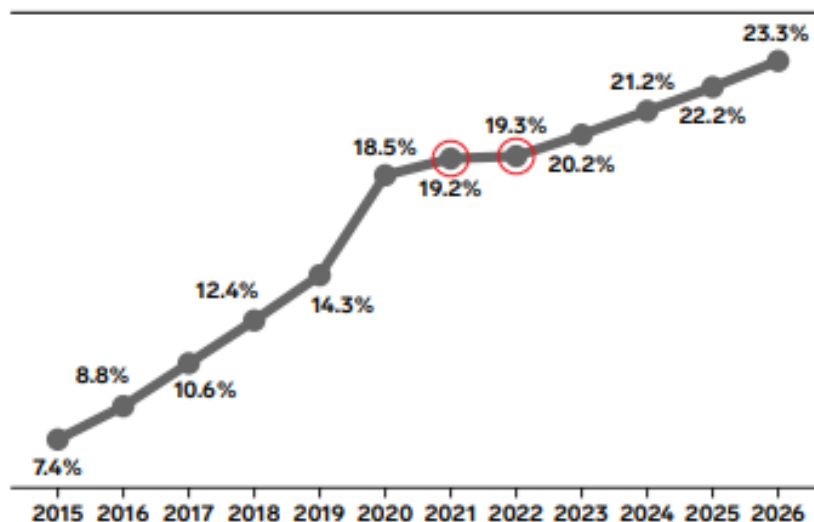


Figura 10 Porcentaje de participación de ventas minoristas en Ecommerce en las ventas minoristas mundiales totales. Fuente: eMarketer, 2023

Días (2020) en una conferencia de La Red Latinoamericana y del Caribe para la investigación en Servicios (REDLAS)³ realizó un estudio en el cual muestra el cambio en el valor de ventas brutas que tuvo el comercio electrónico interno con empresas como Mercado Libre que se muestra en la Figura 11, colocando los crecimientos en porcentaje que tuvo trimestralmente en los diferentes países de Latinoamérica, en donde el segundo trimestre (Q2) de 2020 fue donde se presentó el mayor crecimiento a triple dígito debido a que en ese periodo se dio inicio al confinamiento por pandemia.

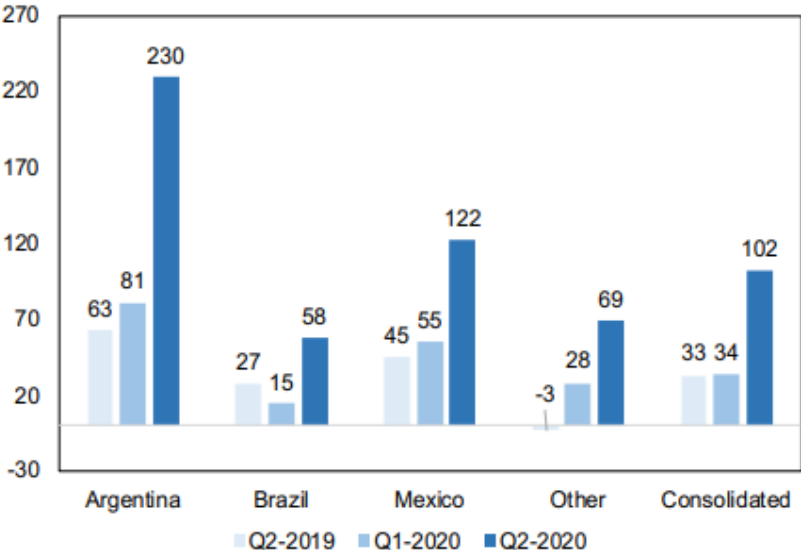


Figura 11. Cambio trimestral en porcentaje de volumen bruto de ventas en Mercado Libre.

Fuente: REDLAS, Días (2020)

De igual forma Días (2020) menciona el crecimiento en términos porcentuales de las ventas brutas de PayU mes a mes en donde también hubo un mayor crecimiento durante julio de 2020 en todos los países en donde tiene presencia, llegando a crecimientos de triple dígito en algunos países como Perú, Brasil, Argentina y Chile.

³ REDLAS es una comunidad de investigadores que tiene como objetivo fomentar la investigación a nivel internacional y difundir el estudio de los temas. REDLAS pertenece a la Comisión Económica para América Latina y el Caribe (CEPAL) organismo dependiente de las Naciones Unidas cuya responsabilidad es contribuir al desarrollo económico de la región y fortalecer las relaciones entre países con las demás naciones del mundo. Para más información consultar <https://www.cepal.org>

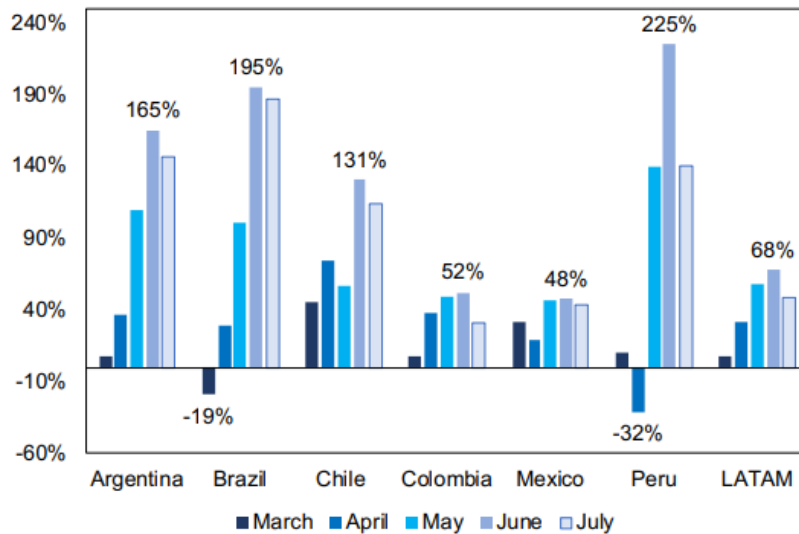


Figura 12. Cambio mensual en porcentaje de volumen bruto de ventas en PayU de marzo a julio de 2020.

Fuente: REDLAS, Días (2020)

Como se menciona durante esta sección es innegable el cambio en el Ecommerce durante la pandemia por COVID-19. Debido a la forma de compra rápida y entregas a domicilio tomó un papel importante, logrando satisfacer necesidades y se volvió una forma de reactivar la economía, logrando un crecimiento de un 300% durante este año en Latinoamérica durante la pandemia que se estimaba llegaría entre 5 y 10 años de acuerdo con la revista Forbes (2021).

2.3. Comercio Electrónico En México

Tras la llegada del internet en México para el año 1987 con la conexión a un servidor en Texas en 1990 se lanzaron las conexiones públicas y privadas fomentando la aparición de nombres de dominios registrado. Para 1992 solo existía “mx” años posteriores surgieron “com.mx” y “gob.mx”. Con esto en 1993 nace el comercio electrónico en México con la creación del “com.mx”, causando que las empresas comenzaran a hacer uso del internet para realizar sus transacciones comerciales (García, 2020).

Cordero (2019) en su trabajo menciona que, de acuerdo con diferentes autores, la relación que existe entre comprador y vendedor da lugar a diferentes tipos de comercio electrónico, por ejemplo:

- B2B o business to business: Comercio electrónico de bienes o servicios, así como las transacciones de información de procesos comerciales realizados entre empresas.
- B2C o business to consumer: Comercio electrónico entre empresas y los consumidores finales. Los bienes y servicios se pueden adquirir mediante un catálogo online y se trabaja en una plataforma para el cierre de la negociación.
- C2C o consumer to consumer: Comercio electrónico entre consumidores finales sin intermediarios, donde unos actúan como compradores y otros como vendedores.

Los tipos de comercio electrónico mencionados en el párrafo anterior, los que destacan en las ventas minoristas online son B2C y C2C. Debido a que la adopción de ventas por medio de internet en empresas de gran tamaño como Nike, Adidas etc. Las cuales son consideradas como ventas de tipo empresa a consumidor. Por otro lado, la existencia de los Marketplace permite que exista un intercambio entre consumidores rompiendo con barreras de comunicación y de intercambio de bienes entre los diferentes entes económicos.

Ante la evolución de la forma de realizar compras en México para el año 2014, casi dos décadas después de la llegada del Ecommerce a México, se constituye una organización civil sin fines de lucro bajo el nombre de la Asociación Mexicana de Ventas Online (AMVO)⁴ con el propósito de apoyar e impulsar el Ecommerce y la economía digital en México.

El Ecommerce en México al paso de los años se ha visto un crecimiento en ventas llegando a 2020 con un crecimiento anual del 81% derivado de la pandemia por COVID-19 como se observa en la figura 13. Aunque la economía mexicana para este año mostraba debilidad el comercio electrónico se mantenía en tendencia creciente ya que la población usaba este tipo de comercio como una alternativa eficiente para satisfacer sus necesidades de consumo (AMVO, 2023).

⁴ Para conocer más sobre la Asociación Mexicana de Venta Online (AMVO) consultar <https://www.amvo.org.mx>

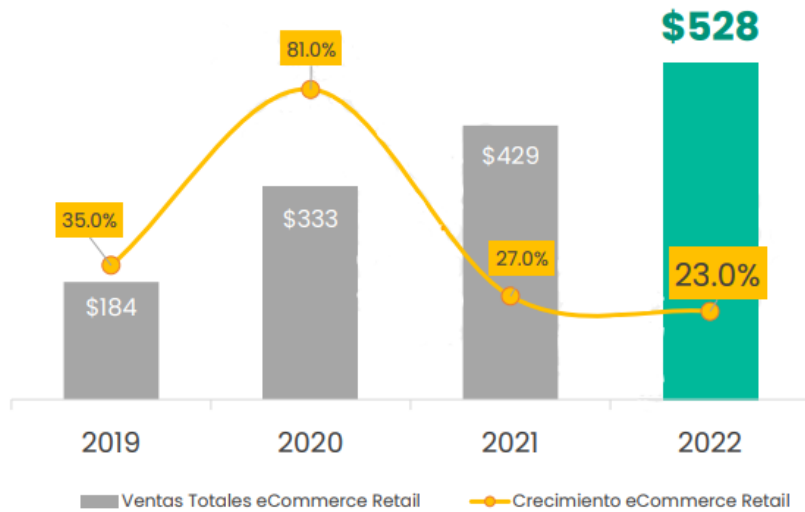


Figura 13. Evolución de Ventas minoristas anuales en México de Ecommerce en miles de millones de pesos y el crecimiento anual en porcentaje de las ventas.

Fuente: AMVO (2023)

Para el año 2022 México se posicionaba dentro de los países con mayor crecimiento en Ecommerce a nivel mundial, siendo de aquellos países que tiene un crecimiento por encima del crecimiento promedio mundial como se muestra en la Figura 14 (eMarketer, 2023; citado en AMVO, 2023).

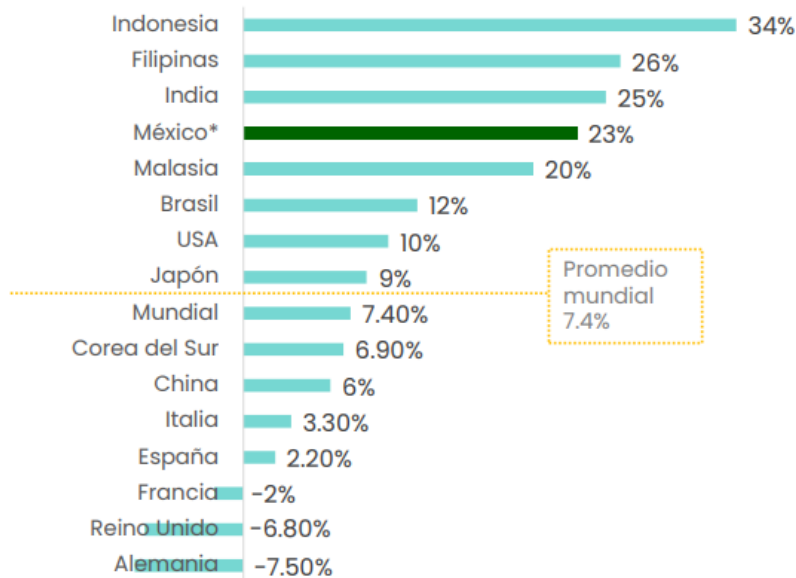


Figura 14. Crecimiento en ecommerce durante 2022 a nivel mundial.

Fuente: eMarketer (2023); citado en AMVO (2023).

Por otra parte, para mejorar la comprensión del consumidor en el uso del Ecommerce menciona indicadores para ellos como son: la audiencia en internet, patrones de tráfico de internet, intensidad y alcance del uso y finalmente la demografía y acceso (Cordero, 2019). La existencia de indicadores es fundamental dado que es la forma de expresar en términos numéricos las oportunidades a capitalizar para las empresas dedicadas a este tipo de comercio.

De acuerdo con un estudio de eMarketer (2022) pronostica que las ventas del comercio electrónico para México se mantengan en tendencia creciente como se muestra en la Figura 15, aunque en el cambio de año contra año se espera que tenga descenso, el pronóstico de 2024 a 2026 es que se mantendrá con crecimientos alrededor del 10%.

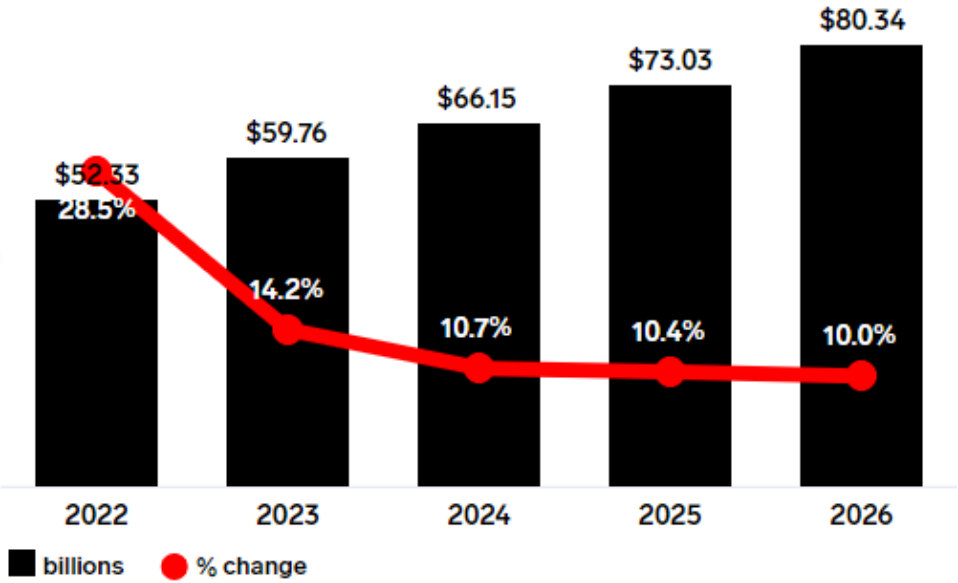


Figura 15. Proyección de Ventas minoristas anuales de Ecommerce en miles de millones de pesos y el cambio anual en porcentaje de las ventas en México.

Fuente: eMarketer (2022).

Como bien se ha mencionado a lo largo del capítulo, el comercio electrónico llega a México para quedarse y continuar creciendo año con año. Tras la pandemia se incrementa el valor de ventas y se posiciona como una forma de mantener activa la economía ante la contingencia sanitaria.

Bajo este contexto de previsión de crecimiento del Ecommerce a futuro la necesidad de mantenerse informado acerca de las tendencias, gustos y preferencias de los consumidores resulta fundamental para mantener dicha tendencia de crecimiento e incluso llegar a superar expectativas, basada en un manejo de estrategias adecuada sobre el consumidor.

Es por esta razón, es que la presente investigación se enfoca en los usuarios de internet que hacen uso del comercio electrónico y en aquellos que aún no participan en este tipo de comercio, lo que los convierte en clientes potenciales. En donde el objetivo es entender su comportamiento y de esta forma las empresas puedan aprovechar la oportunidad de incrementar su portafolio de productos y favorecer a los individuos facilitando el satisfacer sus necesidades de consumo.

Capítulo 3. Análisis Estadístico De La Información

En este capítulo se especifica la forma en que se construyó la base de datos que se utiliza para la presente investigación, así como el análisis estadístico de la misma. Una de las grandes ventajas en México es que cuenta con el Instituto Nacional de Estadística y Geografía (INEGI) que brinda información estadística, geográfica y económica a nivel nacional por medio de encuestas aplicadas a diferentes muestras representativas de la población dentro de la república mexicana a diferentes niveles de granularidad, es decir, ya sea encuestas que sean representativas a nivel entidad federativa, a nivel municipio.

3.1. Descripción De La Base De Datos ENDUTIH

En 1983 por decreto presidencial se crea en México el Instituto Nacional de Estadística y Geografía (INEGI) con el objetivo de recopilar y difundir información estadística, geográfica y económica a nivel nacional. Una de las formas para obtener dicha información es aplicando encuestas a una muestra representativa de la población que radica en el territorio nacional (INEGI, 2021)

INEGI cuenta con 10 encuestas aplicadas a hogares de la república mexicana, en donde se encuentra la Encuesta Nacional de Disponibilidad y Uso de las Tecnologías de la Información en los Hogares (ENDUTIH) que tiene por objetivo investigar en qué medida y a través de que medios la población tiene acceso y utiliza las tecnologías de la información y comunicación, aplicando el cuestionario a hogares e individuos de seis o más años de edad de manera anual para un panorama de comportamiento tanto en hogares como en individuos (INEGI, 2020).

Los datos que brinda la ENDUTIH permiten al Instituto Federal de Telecomunicaciones (IFT) realizar análisis con los cuales se basa para establecer políticas públicas que lleven a una sociedad de la información y el conocimiento (ENDUTIH, 2021).

Dentro de la investigación se aborda información de la ENDUTIH 2021, la cual recopila datos que permiten obtener variables sociodemográficas (edad, sexo, escolaridad, estado conyugal, etc.); gustos y preferencias (variables relevantes desde la perspectiva

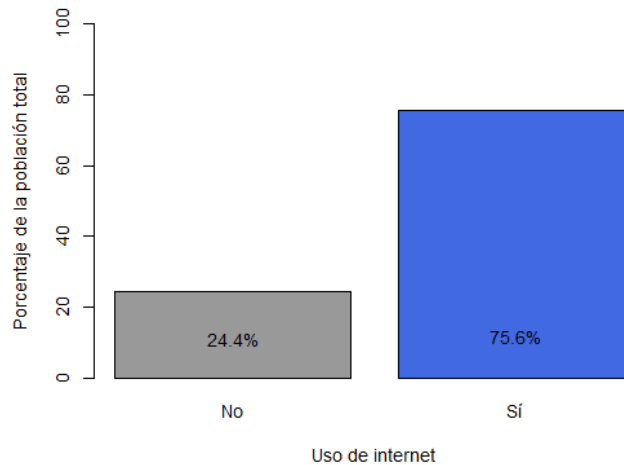
microeconómica; económicos (clasificación de la población en PEA y PNEA, mercado laboral); demográficas (lugar de residencia, inclusión financiera, acceso a las TIC, etc.). Dichos datos a nivel vivienda, hogar e individuo, cada uno con un factor de expansión que indica a cuántos hogares, viviendas o individuos está representando la unidad encuestada.

El periodo de levantamiento de esta encuesta se ubica entre el 2 de agosto al 24 de septiembre de 2021 con un total de 65, 179 viviendas a nivel nacional. La cual se encuentra diseñada para poder dar resultados a nivel nacional, nacional urbano y rural, y por entidad federativa (ENDUTIH, 2021).

3.2. Análisis Estadístico Y Gráfico De Las Variables

Como se menciona en la sección anterior, la ENDUTIH a grandes rasgos brinda información sobre el uso y disponibilidad de las tecnologías de información en los hogares y en la población dentro del territorio mexicano. Dicho esto, dentro del cuestionario aplicado para esta encuesta se encuentran dos preguntas fundamentales para el desarrollo de la presente investigación, las cuales son ¿tiene acceso a internet? y ¿ha realizado una compra por medio de internet?

Partiendo de las preguntas fundamentales para la investigación y bajo los resultados de la ENDUTIH 2021 se estima una población total de 117.1 millones de habitantes mayores de 6 años en el territorio mexicano, de los cuales el 75.6% (88.56 millones) afirma haber hecho uso de internet durante el 2021.



Gráfica 1. Proporción de personas que hacen uso de internet en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

Se considera que para realizar una compra en Ecommerce se debe tener acceso a internet dado que es la columna vertebral de este tipo de comercio, por lo que la población de interés para fines de la presente investigación se reduce a los 88.56 millones de personas que cuentan con acceso a este servicio.

Tras observar la desnivelada proporción de mexicanos que aún no participan en las compras por Ecommerce, se consideran las siguientes dieciséis variables con las que se evalúa los factores que influyen la compra por internet:

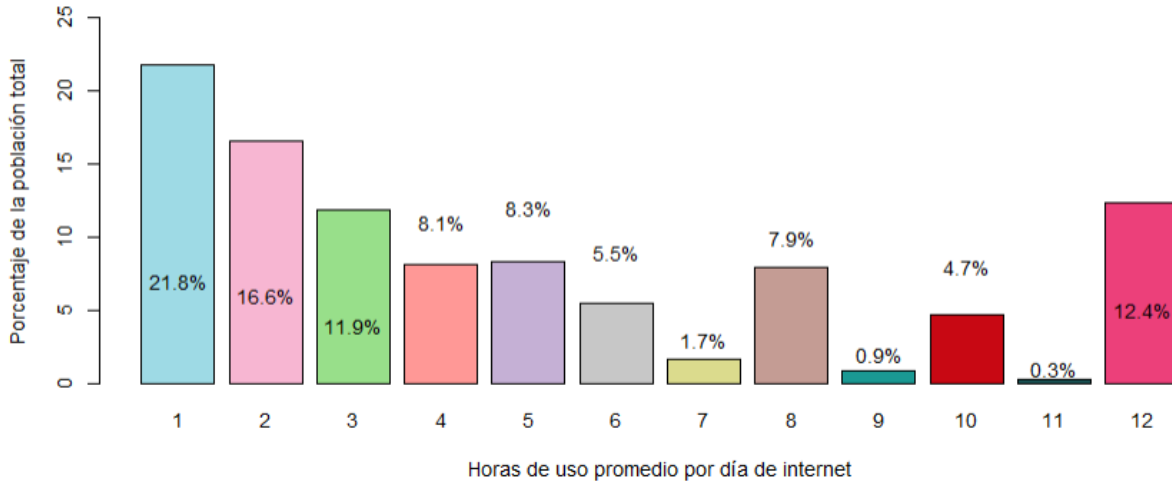
- Horas promedio por día de uso de internet.
- Tiempo utilizando internet en años.
- Estrato socioeconómico.
- Género.
- Edad.
- Dispone de celular.
- Disponibilidad de celular tipo smartphone.
- Acceso a computadoras, tablet o laptop.
- Horas promedio de uso de computadora, tablet o laptop.
- Tipo de población (Urbana o Rural).
- Tamaño de localidad.
- Nivel de estudios.

- Ser estudiante.
- Parentesco con el jefe de familia.
- Económicamente activo.
- Tipo de actividad que desempeña.
- Tipo de conexión a internet en el hogar.
- Se conecta mediante su smartphone a internet por red móvil.
- Uso de redes sociales.
- Utiliza Facebook.
- Utiliza WhatsApp.
- Utiliza Instagram.
- Utiliza Messenger.
- Utiliza Twitter.
- Utiliza YouTube.

A continuación, se describe cada una de las variables con el objetivo de identificar patrones y sesgos en los datos, obteniendo los siguientes resultados:

- **Horas promedio por día de uso de internet.**

Una de las preguntas consecuentes de saber que el individuo tiene acceso a internet es saber el tiempo que le dedican al día. Es decir, cuánto tiempo están expuestos los usuarios a diferentes mensajes publicitarios que los incentivan a adquirir servicios o productos por este medio. Con base en la gráfica, una gran proporción de usuarios solo hace uso de internet en promedio 1 hora al día, seguido de dos horas y 12 horas por lo que se detectan picos en ambos extremos de la línea temporal.

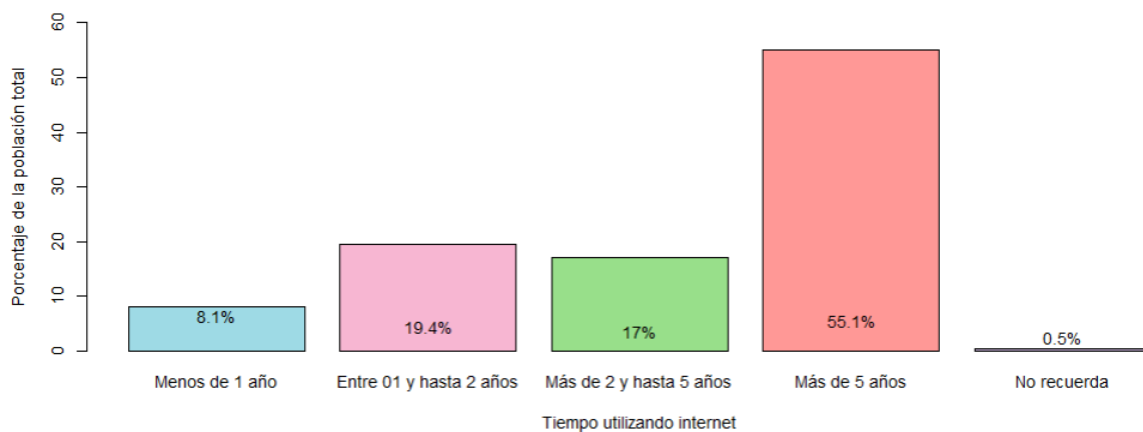


Gráfica 2. Tiempo promedio de uso de internet en horas en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Tiempo utilizando internet en años.**

Otra de las variables que se considera es el tiempo que llevan utilizando internet, el cual es un indicador de cuánta experiencia tienen y que tan ágiles pueden ser al momento de navegar en la red. Lo que se muestra en la gráfica 3 es que el 55% de las personas que han hecho uso de internet lo han venido haciendo desde hace más de 5 años mientras que la proporción de aquellos que recién comienza a hacer uso del servicio es de tan sólo un 8%, dejando en claro que es mayor la población con más experiencia en la red.

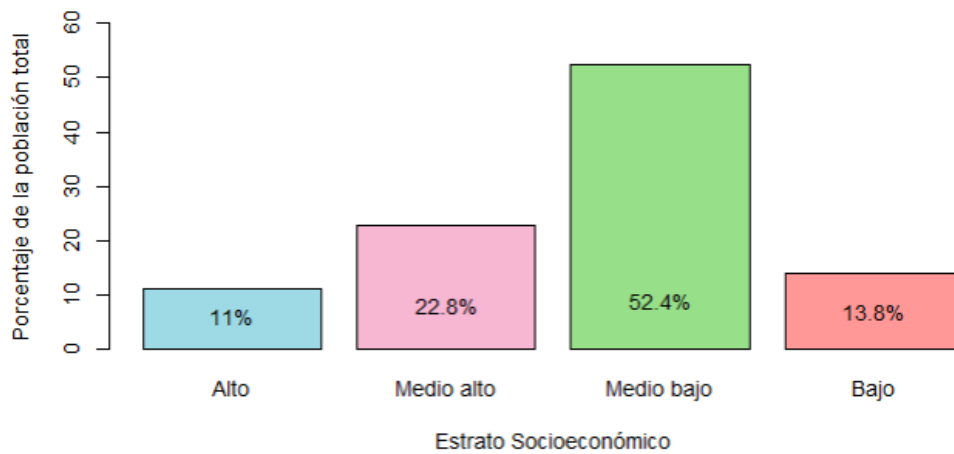


Gráfica 3. Tiempo en años usando internet por los mexicanos 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Estrato socioeconómico**

Como bien se sabe dentro de la población existe una distinción por niveles socioeconómicos, dependiendo del nivel de ingreso, grado de educación y calidad de vida entre otros factores que lo determinan. Por lo que resulta de interés indagar sobre la distribución de los usuarios de internet en estos grupos para poder identificar si la clasificación interfiere en la decisión de comprar por medio de internet. Lo que se observa en la gráfica 4, es que los dos grandes grupos en donde se alojan los usuarios son en los niveles socioeconómicos Medio bajo y Medio alto, siendo el estrato Medio bajo el de mayor peso con un 52.4% de usuarios.

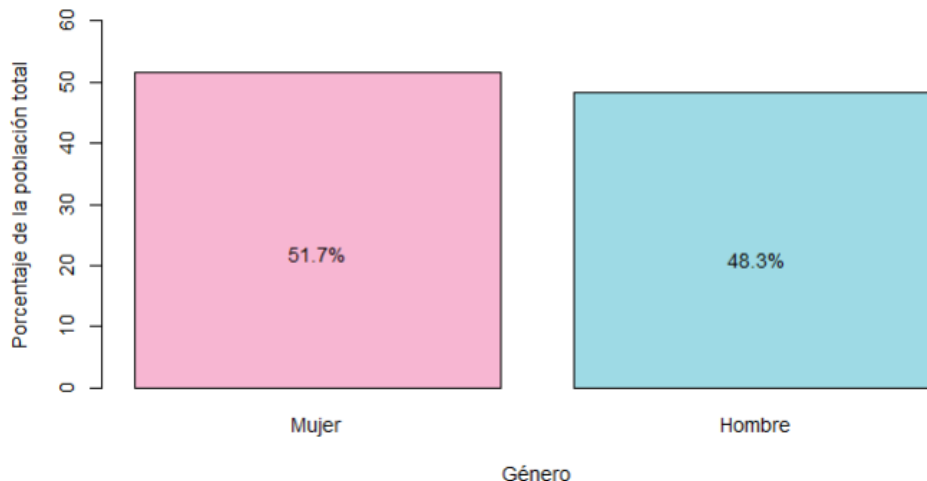


Gráfica 4. Nivel socioeconómico en los usuarios de internet en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Género**

Se considera la variable de género con la intención de comprobar la posibilidad de un sesgo que resulte en un factor que influya en la probabilidad de compra en un usuario. Sin embargo, se observa en la gráfica 5 que existe una distribución casi simétrica para la variable, sobresaliendo 3.4 puntos porcentuales el género Mujer, dada esta diferencia puede resultar con signo positivo favoreciendo a la probabilidad de compra cuando el usuario es mujer.

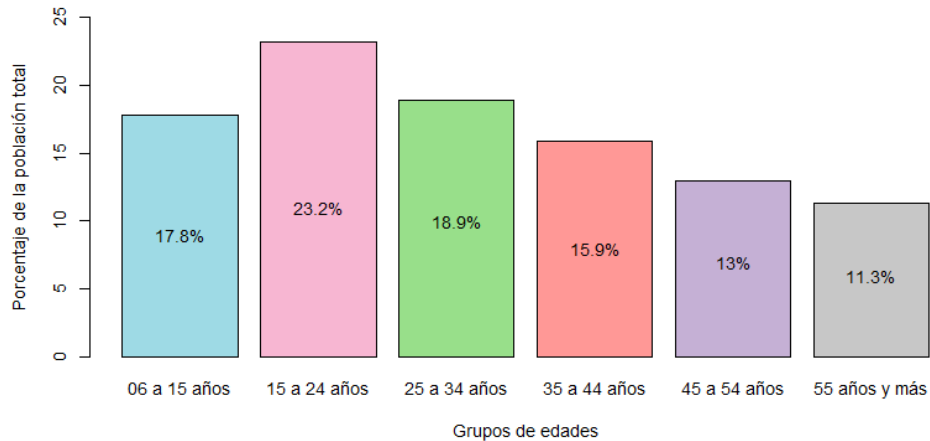


Gráfica 5. Distribución por género en los usuarios de internet en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Edad**

La variable se constituye de las edades en años cumplidos de los usuarios. Se cambia la estructura por grupos, los cuales permiten identificar sesgos en los rangos. Dentro de la hipótesis se plantea que la edad va a presentar un signo positivo expresando que a mayor edad mayor será la probabilidad de compra, sin embargo, al llegar a una edad avanzada la probabilidad el signo puede volverse negativo debido a que el usuario no tiene el mismo conocimiento en el tema como lo tienen las generaciones más jóvenes. Esta hipótesis se sustenta en observar la gráfica 6 donde los usuarios de internet se concentran principalmente en edades que comprenden de los 6 hasta los 34 años, resultando el rango de 15 a 24 años, el que tiene una mayor participación y para edades más avanzadas comienza a disminuir el número de usuarios.

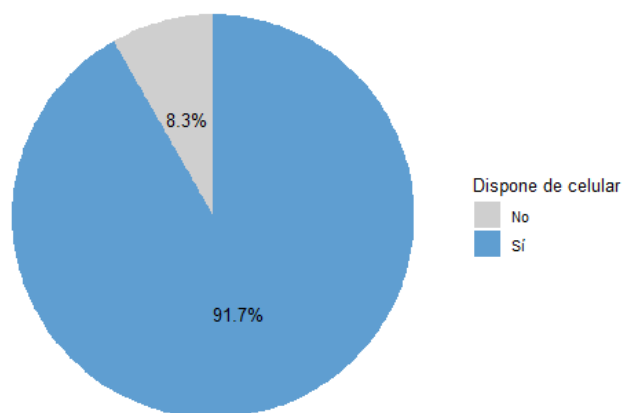


Gráfica 6. Usuarios de internet por rangos de edades en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Disponibilidad de celular**

El gobierno de navarra menciona que los celulares o teléfonos móviles, como su nombre lo indica, contienen las funciones de redes móviles que son las mismas de las redes de cable convencional de la telefonía fija, es decir, llamar y recibir llamadas. Contando con funciones auxiliares que dependen cada modelo como reproductor de música, radio, tomar fotografías, videos calculadora, agenda, entre otras. Sin embargo, estos a diferencia de los teléfonos tipo smartphones o celulares inteligentes, no cuentan con el sistema operativo que permite la instalación de aplicaciones adicionales para ampliar sus funciones.

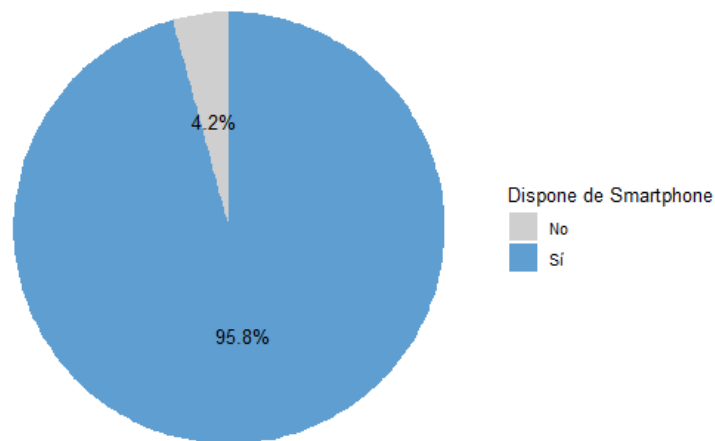


Gráfica 7. Usuarios de internet que disponen de un teléfono celular en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Disponibilidad de celular tipo smartphone.**

Quicios, Sevillano y Ortega (2013; citado en Saussure 2016) definen un smartphone como un teléfono móvil inteligente el cual cuenta con sistema de gestión de información y sus características técnicas son similares a las de una laptop. Por lo que una de las formas de realizar compras por internet es por medio de aplicaciones móviles disponibles para este tipo de teléfonos y aunado a esto, con el acceso a internet desde estos dispositivos permiten que el 95.8% de los usuarios estén más propensos a ceder ante una compra debido a la facilidad y disponibilidad instantánea.

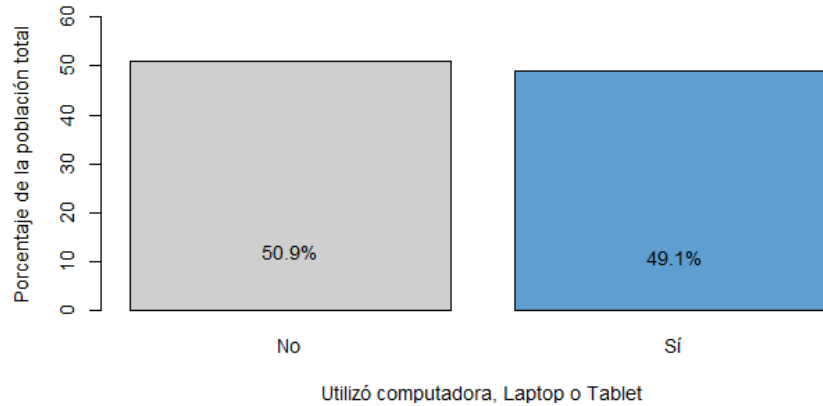


Gráfica 8. Usuarios de internet que disponen de un celular tipo smartphone en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Acceso a computadoras, Tablet o laptop**

De acuerdo con datos de Americas Market Intelligence (AMI) en 2022 el 28% de las compras en Ecommerce se realizaron por medio de una computadora de escritorio (desktop). Bajo este contexto se considera apropiado evaluar el haber utilizado una computadora, laptop o tablet como un factor que interviene en las compras online. De acuerdo con las proporciones de la población se mantiene casi simétrica la distribución de los usuarios como se muestra en la gráfica 8.

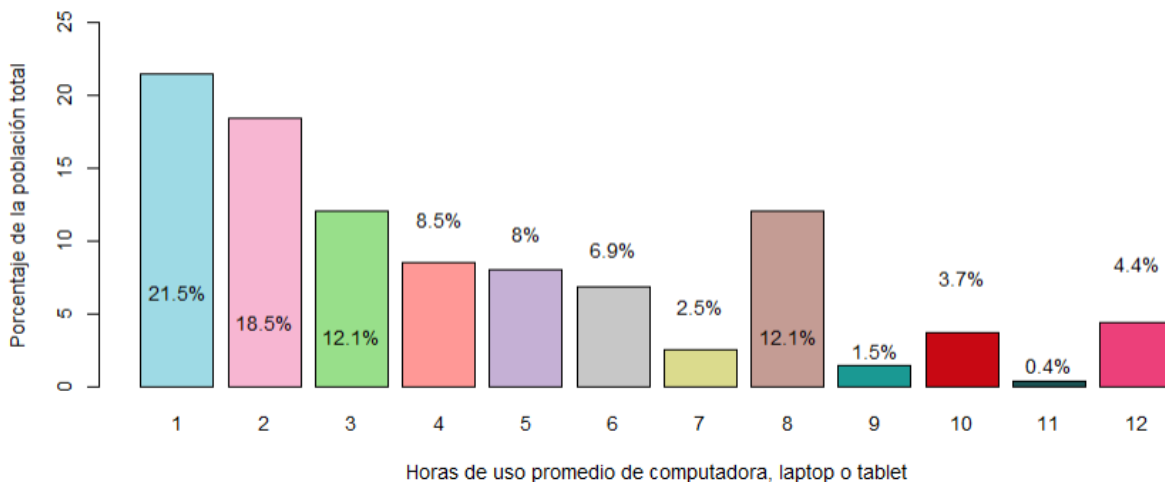


Gráfica 9. Usuarios de internet que utilizan computadora, laptop o tablet México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Horas promedio de uso de computadora, Tablet o laptop.**

Como se muestra en la variable anterior el uso de computadoras para realizar compras juega un papel importante, por lo que es importante también considerar el tiempo promedio que pasan un usuario utilizando estos dispositivos al día. En donde se observa que los tiempos más relevantes para los usuarios que declaran haber utilizado uno los dispositivos son de una, dos y ocho horas en promedio al día. Detectando un caso similar que en el tiempo de uso de internet en donde existe un pico en los promedios más bajos y posteriormente se ve un tercer pico en tiempos mayores.



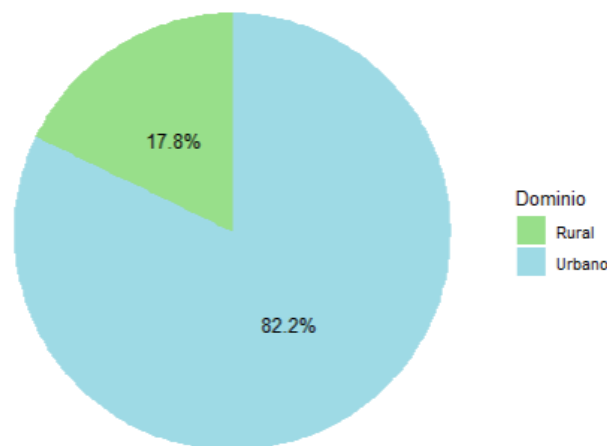
Gráfica 10. Horas promedio de uso de computadoras, laptop o tablet por usuarios de internet de individuo es México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Tipo de población (Urbana o Rural).**

De acuerdo con INEGI el número de habitantes en la localidad es lo que define la población rural y urbana. Considera rural cuando tiene menos de 2,500 habitantes y urbana cuando tiene más de 2,500 habitantes. Además de esta diferencia entre ambos tipos de localidades también se considera que las zonas urbanas son aquellas en donde el acceso a diferentes servicios como luz, agua e incluso existe una mayor cantidad de empleos mientras que las zonas rurales son directamente relacionadas con el sector primario dado el elevado número de áreas verdes que no se ven dentro de una urbana (CEPAL, 2022).

Dicho lo anterior el propósito de la variable dentro del análisis es identificar en qué tipo de zonas se encuentran los usuarios de internet y con esto inferir si el tipo de zona afecta la probabilidad de compra en Ecommerce o resulta ser una variable poco significativa dentro del modelo. Con base en datos de la ENDUTIH se identifica que el 82% de los usuarios de internet pertenecen a una zona urbana, por lo que se espera que la variable tenga un efecto positivo sobre la compra cuando el usuario pertenece a este ámbito.



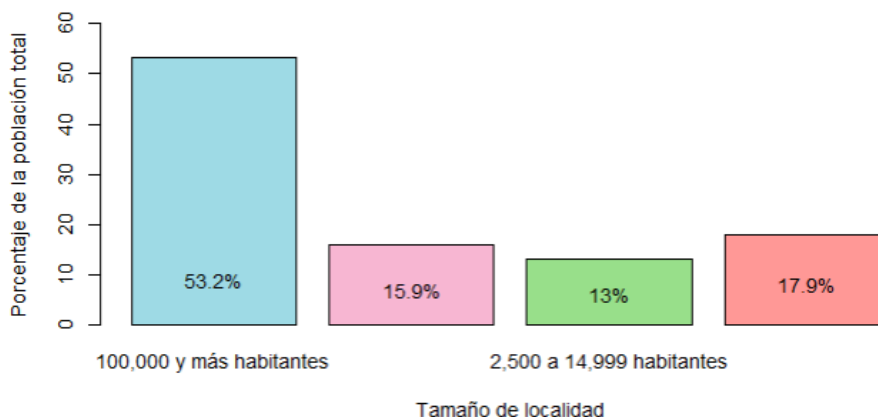
Gráfica 11. Distribución de usuarios de internet por zonas urbanas o rurales en México.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Tamaño de localidad.**

Como bien se menciona en la variable tipo de población las zonas habitadas se clasifican de acuerdo con el número de habitantes. Aunque predomine el ámbito urbano cuya población

se sabe que supera los 2,500 habitantes, es interesante cuestionarse sobre que tamaño de localidades se encuentra la mayor parte de los usuarios. Por lo que, en esta variable se clasifica a los usuarios de acuerdo con el tamaño de la localidad en la que radican, donde se observa que el 83% de ellos radican en localidades que tienen más de 100,000 habitantes. Teniendo como hipótesis que cuanto más grande sea la localidad afecta positivamente la probabilidad de compra.

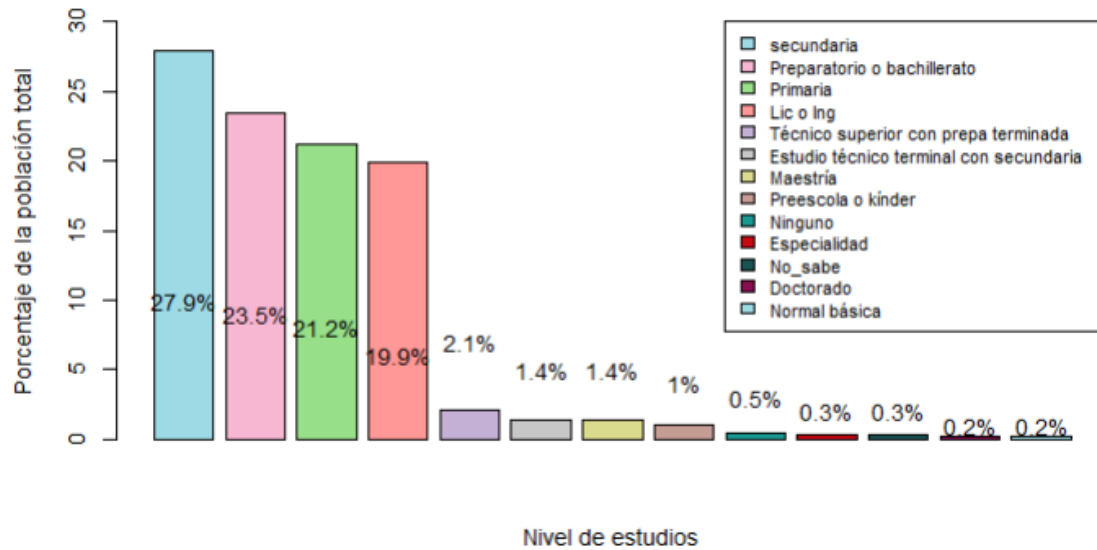


Gráfica 12. Distribución de usuarios de internet por tamaño de localidad en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Nivel de estudios.**

Parte del análisis que se realiza se enfoca en comprobar si el grado de estudios cursado por los usuarios de internet toma relevancia en las compras de Ecommerce. Teniendo como hipótesis que a mayor escolaridad mayor será la probabilidad de compra, dado que entre mayor nivel de estudio el ingreso al que aspira el usuario es mayor. Sin embargo, cómo se observa en la gráfica 12 dentro de los usuarios la mayor proporción tienen secundaria terminada, seguido de la preparatoria y nivel primaria; dejando los niveles más altos con menor proporción de usuarios.

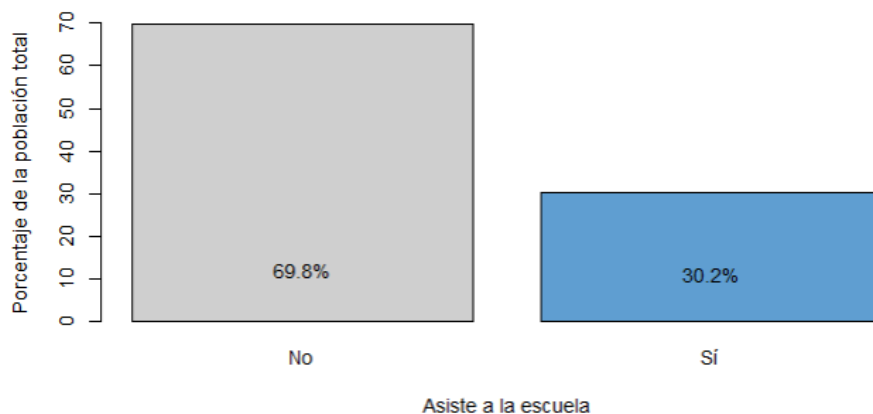


Gráfica 13. Distribución de usuarios de internet por Nivel de estudios en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Asiste a la escuela.**

Esta variable hace referencia a si el usuario es estudiante; fue seleccionada con el fin de probar la hipótesis en donde se plantea que un estudiante que no es económicamente independiente en su totalidad y que no cuenta aún con remuneraciones por jornadas laborales, lo que lleva a que su presupuesto disponible sea bajo, por lo tanto el mundo del Ecommerce no lo percibe como un medio factible para llevar a cabo sus compras. Observando la distribución de las respuestas de los usuarios se percibe que solo el 30% de estos declara asistir a la escuela. Por lo que da un primer indicio a que puede ser correcta la hipótesis planteada.



Gráfica 14. Proporción de usuarios de internet que asisten a la escuela en México 2021.

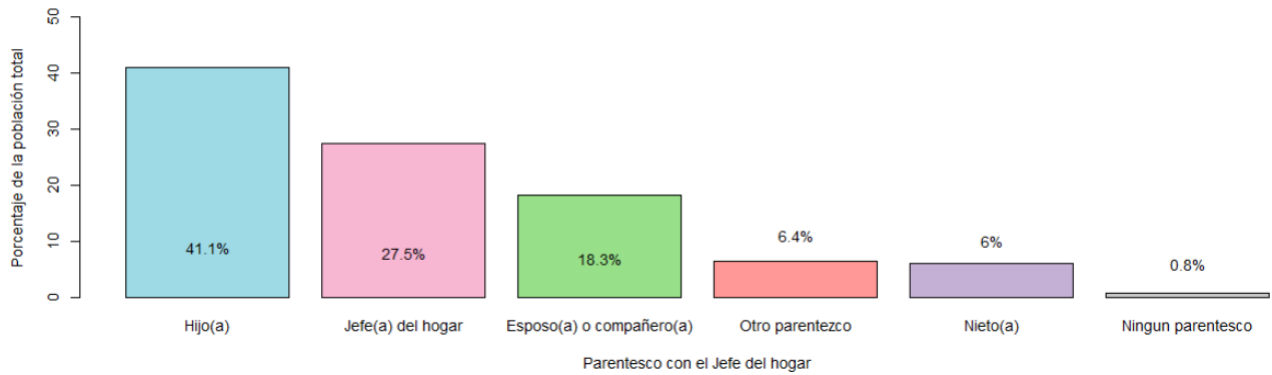
Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Parentesco con el Jefe de familia.**

Una de las preguntas dentro de la encuesta es ¿Cuál es el parentesco que tiene el individuo elegido con el jefe o jefa del hogar? De acuerdo con INEGI se define al Jefe(a) de familia como la persona reconocida como tal por los demás integrantes del hogar, dicho reconocimiento se le asigna al miembro de mayor jerarquía, ya sea por ser el proveedor de ingresos o por ser quien toma las decisiones.

Dicho lo anterior, se considera relevante contemplar el rol que tiene el usuario dentro del hogar para identificar como interactúa el rol del usuario ante las compras de internet, con lo que se pretende identificar si el jefe (a) de familia y el compañero (a) disminuye la probabilidad de que se realice una compra en Ecommerce dado que entre estos dos roles son los que proveen y administran el ingreso del hogar en la mayoría de los casos, restringiéndolo a sólo aquellos bienes que les resultan factibles. Con esta información brinda un panorama sobre el tipo personas se dirigen las plataformas con su contenido, permitiendo afinar lo que muestran al público haciéndolo de mayor interés.

Describiendo la variable en la gráfica 14 se puede observar que la mayor proporción de los usuarios de internet tienen un rol familiar de Hijos(as) con 41% de la población, seguido de los Jefes (as) del hogar con un 27.5% de la población.

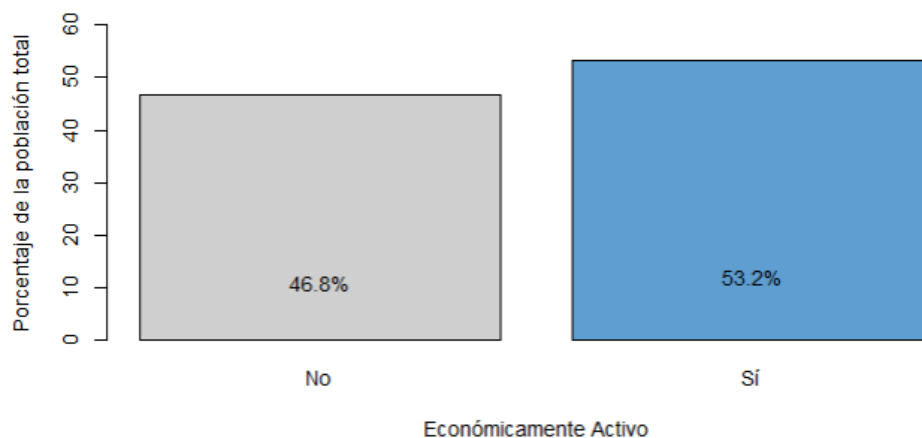


Gráfica 15. Rol de los usuarios de internet dentro de un hogar en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Económicamente activo.**

Para poder adquirir bienes y servicios en cualquier tipo de comercio los consumidores deben tener un ingreso, el cual destinan a dicha adquisición. Por lo tanto, una persona que pertenece a la Población Económicamente Activa (PEA) se considera económicamente activa por lo que se define como la persona con edad suficiente para ejercer actividades productivas de bienes y servicios (INEI, 2017). Bajo este concepto se considera que una persona con entradas de capital puede tener una mayor probabilidad de comprar por medio del comercio electrónico; al describir la variable se identifica que un 53% de usuarios de internet se encuentra activo, por lo que da lugar a que la hipótesis planteada sea verdadera.

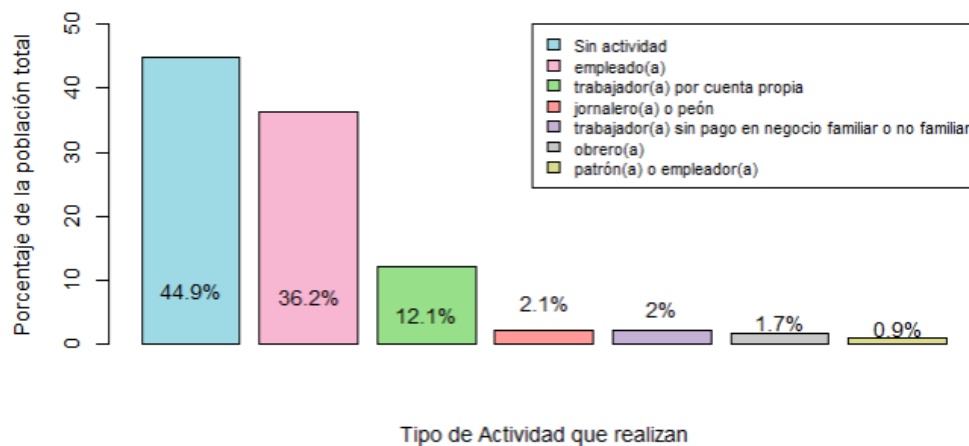


Gráfica 16. Proporción de usuarios de internet económicamente activos en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Tipo de actividad que desempeña.**

Una vez que se ha detectado a la población económicamente activa toma relevancia el tipo de actividad que realizan para generar sus ingresos. Se pretende comprobar que cuanto más alto es el puesto laboral del usuario mayor será la probabilidad de compra dado que un puesto de mayor jerarquía se relaciona directamente con un mayor ingreso. En la gráfica 16 se identifica que la mayor proporción de usuarios no realiza una actividad (sin actividad) dado que pertenecen a la PNEA, sin embargo, dejando de lado esto la actividad que tiene una mayor población es de tipo empleado(a) con 36% de los usuarios económicamente activos, seguido de trabajador por cuenta propia con un 12%.

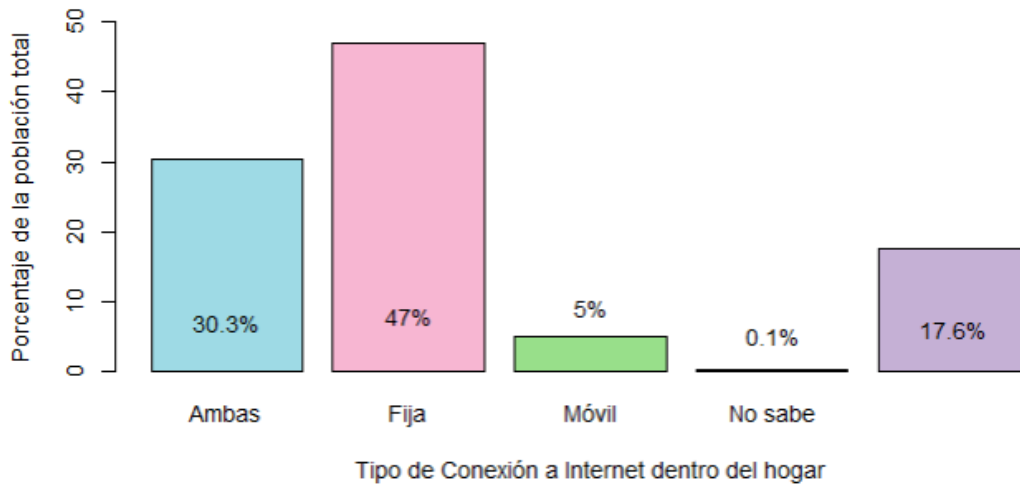


Gráfica 17. Proporción de usuarios de internet económicamente activos en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Tipo de conexión a internet en el hogar.**

Ahora que se conoce a la población que tiene acceso a internet y el tiempo que destinan a usar este servicio, abre paso a cuestionarse ¿cuál es el tipo de conexión con la que cuentan los usuarios dentro sus hogares? Esta es una de las preguntas que también considera la ENDUTIH, resultando que el tipo de conexión fija es el más común dentro de los hogares dado que 47% de ellos cuentan con este servicio y un 30% cuenta con este tipo de conexión más un tipo de conexión móvil como bien se muestra en la gráfica 17.

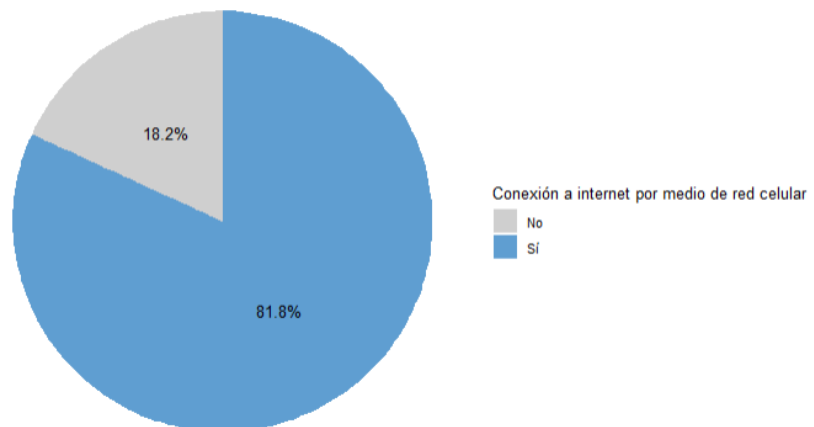


Gráfica 18. Tipo de conexión a internet en los hogares en México 2021

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Se conecta mediante su smartphone a internet por red móvil**

Además de considerar el tipo de conexión que tiene el usuario dentro del hogar se contempla como variable de interés si tiene acceso a internet desde su smartphone mediante red móvil o también conocida como datos móviles. En donde se identifica que casi el 82 % de los usuarios de internet si cuentan con este tipo de conexión que les permite acceder a la red desde diferentes lugares fuera o dentro del hogar, lo que deja como hipótesis que estos usuarios tengan una mayor probabilidad de compra dado que no están limitados a solo poder hacer compras en su hogar.

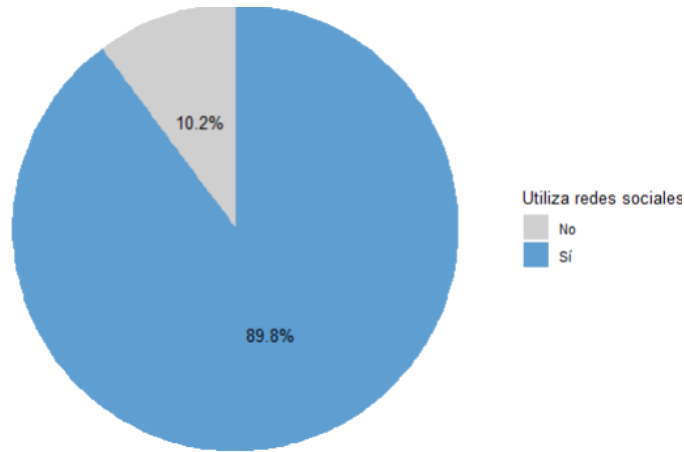


Gráfica 19. Usuarios de internet que se conectan por medio de red celular en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Utiliza redes sociales**

Parte de entender al usuario es verificar si tiene contacto con redes sociales dado que se busca probar si su uso beneficia en la compra de Ecommerce por las estrategias de marketing digital las cuales son todas aquellas que ocurren en la web y buscan algún tipo de compra por parte del usuario (Selman, 2017; citado en Velázquez 2021). Siguiendo con la gráfica 19 se idéntica que casi un 90% de los usuarios de internet hicieron uso de redes sociales.



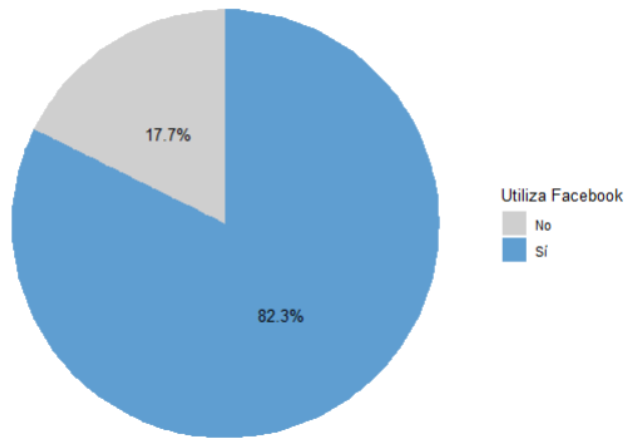
Gráfica 20. Usuarios de internet que se utilizan redes sociales en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

Cómo bien se observa en la variable anterior existe una gran cantidad de personas que utilizan redes sociales y aunado a esto resulta de interés saber que red social está ligada a favorecer las compras online considerado Facebook, WhatsApp, Messenger e Instagram las cuales cuentan con un mayor número de usuarios en México de acuerdo con Statista (2023) y dentro de la investigación se integra Twitter y YouTube para comprobar las hipótesis planteadas. Bajo los datos de la ENDUTIH se identifica el siguiente comportamiento para estas redes durante 2021:

- **Utiliza Facebook**

De acuerdo con los datos de la encuesta el 82% de los usuarios de internet hicieron uso de Facebook. Dejándola como la segunda red más utilizada en 2021. Dado el tipo de contenido grafico que tiene y la proporción de usuario con las que cuenta el sitio se espera que favorezca la compra en Ecommerce.

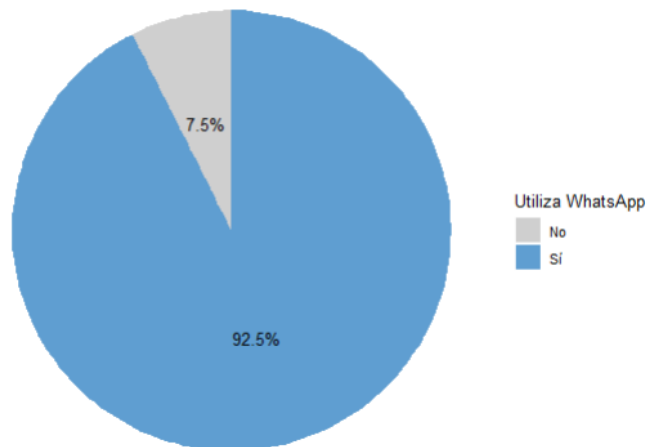


Gráfica 21. Usuarios de internet que se utilizan Facebook en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Utiliza WhatsApp**

Otra de las redes sociales a mencionar es WhatsApp la cual de acuerdo con la ENDUTIH es la red social más ocupada durante 2021 en donde el 92.5% de los usuarios de internet declaran hacer uso de esta aplicación. Por lo que se plantea como hipótesis que es una variable cuyo signo se espera sea positivo, diciendo así que favorece la probabilidad de compra online dado el alto uso que tiene.

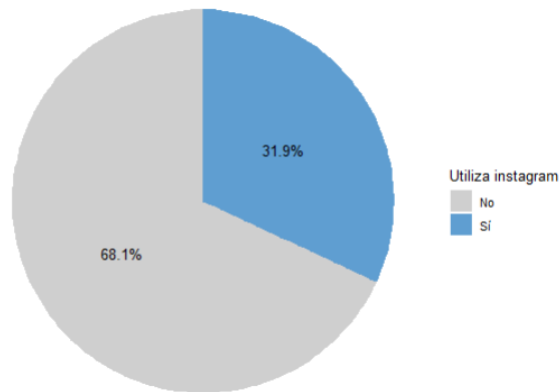


Gráfica 22. Usuarios de internet que se utilizan Whatsapp en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Utiliza Instagram**

Si bien esta red social es usada por casi el 32% de la población bajo el estudio de Statista se evalúa dentro del análisis. Donde se plantea que el signo que indique el modelo va a ser positivo, aunque la mayor cantidad de personas no hacen uso de esta red social por el tipo de contenido se espera que tenga relevancia sobre la compra en Ecommerce.

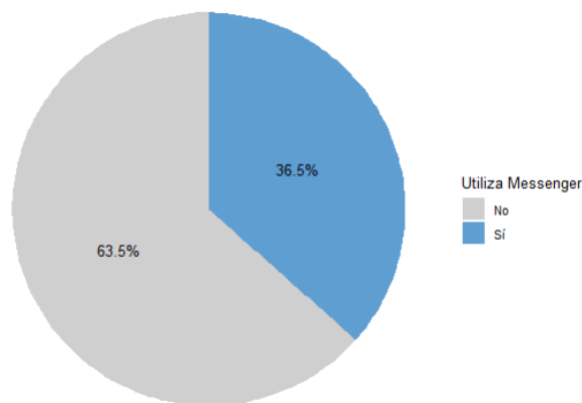


Gráfica 23. Usuarios de internet que se utilizan Instagram en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

- **Utiliza Messenger**

Se considera el uso de Messenger, que aun siendo la tercera red social más utilizada se espera un signo negativo dado que es un complemento de Facebook para mensajería, por lo que se considera que no favorece la probabilidad de compra al no ser un canal con alto uso de los marketplace para realizar ofertas que orillen al usuario a realizar compras por internet.

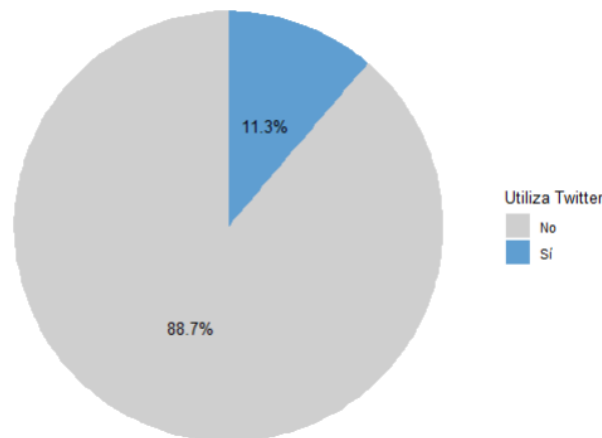


Gráfica 24. Usuarios de internet que se utilizan Messenger en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Utiliza Twitter**

Como parte de las redes sociales se coloca como adicional Twitter debido a que es una de las redes sociales populares colocándose en la sexta posición (Statista, 2023). La ENDUTIH (2021) identifica que solo un 11.3% de los usuarios de internet usaron esta red; comparado con las redes sociales anteriores se puede considerar que tiene una baja adopción, sin embargo, esto no garantiza que no pueda ser un factor significativo ante la compra por Ecommerce.



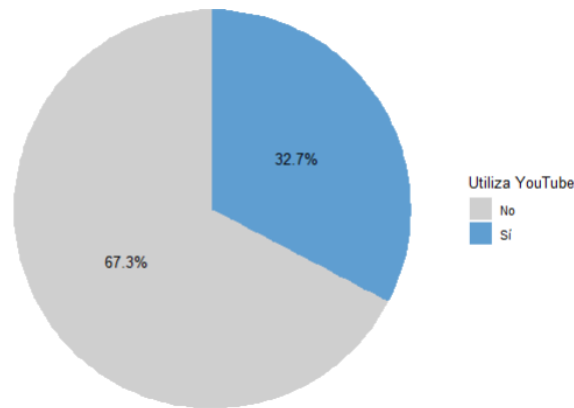
Gráfica 25. Usuarios de internet que se utilizan Twitter en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

- **Utiliza YouTube**

Por último, se considera el uso de YouTube, la plataforma de video; utilizando datos de la ENDUTIH 2021 se identifica que 32.7% de usuarios de internet declara haber usado dicha plataforma durante este ese año. De acuerdo con su página oficial⁵ parte de los servicios que tiene a disposición del público es el ingresar publicidad por medio de su iniciativa YouTube Advertising, en donde mencionan que sus usuarios afirman que es 2 veces más probable que compren un producto o servicio que vieron en YouTube. Por lo que, bajo este dato esta plataforma resulta de interés para verificar el impacto que tiene en la decisión de las compras de los usuarios de internet por medio del comercio electrónico.

⁵ https://www.youtube.com/intl/ALL_mx/ads/how-it-works/

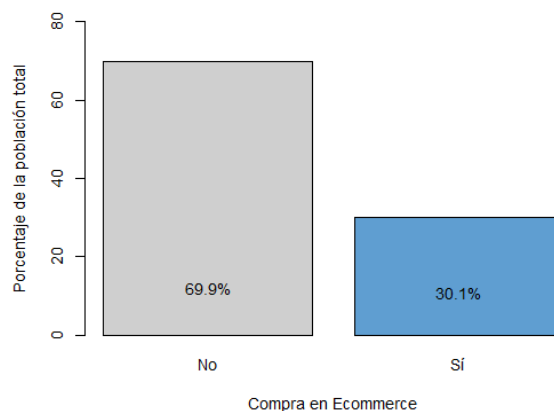


Gráfica 26. Usuarios de internet que se utilizan YouTube en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

Una vez descritas y analizadas cada una de las variables independientes que se consideran en la presente investigación procedemos a describir a la variable dependiente, la cual se obtiene de la pregunta ¿Ha realizado compras por medio de internet en los últimos 12 meses? Si bien ya se ha mencionado que 88.56 millones de mexicanos hacen uso de internet, esto no implica que todos ya realicen compras por internet.

Con las respuestas que dieron los encastados para la ENDUTIH (2021) se estima que solo el 30% realizó una compra por internet durante 2021. Dejando un área de oportunidad de casi un 70% equivalente a 61.9 millones de usuarios que tienen potencial para convertirse en compradores.



Gráfica 27. Usuarios de internet que realizaron una compra por medio de internet en México 2021.

Fuente: Elaboración propia con datos de la ENDUTIH 2021

Tras observar la proporción de personas que aún no compran por internet queda evidenciado el tamaño de oportunidad que existe en el mercado para el Ecommerce. Por lo que la presente investigación brinda esa perspectiva de cuales elementos se pueden considerar los Marketplace en México para llegar a estos usuarios que también son consumidores en el mundo físico y convertirlos en nuevos compradores dentro de su sitio.

Capítulo 4. Aplicación Del Modelo Logit Al Comercio Electrónico

En capítulos anteriores se ha expuesto la historia del Ecommerce tanto en el mundo como en México, destacando su crecimiento durante la pandemia y también postpandemia. En donde, tras el crecimiento acelerado surge la necesidad de analizar al consumidor que participa en este tipo de comercio, con el objetivo de entender su comportamiento identificando las particularidades que están influyendo para que un individuo realice o no compras por este medio.

Para dicho análisis se plantea un modelo logit dado su estructura en donde la variable dependiente es dicótoma y los resultados son interpretados como probabilidades, permite que se pueda responder a la pregunta de ¿qué probabilidad tiene un individuo con acceso a internet de realizar una compra por internet dadas sus características sociodemográficas, económicas, gustos y preferencias?

4.1. Descripción De Metodología Econométrica: Modelos Logit

En análisis de origen económico es natural encontrar variables categóricas o también conocidas como dummies, debido a que son aquellas que indican ausencia o presencia de alguna característica o atributo, principalmente en información sobre individuos y familias por mencionar algunas (Quintana y Mendoza, 2016), así como sucede para el caso de la ENDUTIH en donde se encuentran variables categóricas, por ejemplo, género, si la persona cuenta con acceso a internet, si el entrevistado es económicamente activo, entre otras.

De acuerdo con Quintana y Mendoza (2016) cuando la variable binaria es dependiente, el modelo de regresión se interpreta como probabilidades, lo que permite responder a preguntas como ¿Cuál es la probabilidad de que un consumidor realice una compra por medio del comercio electrónico?

Al interpretarse como probabilidades los modelos de regresión con variable binaria dependiente, se relaciona directamente con la Función de Distribución o Acumulativa (FDA) de una variable aleatoria. Gujarati (2003) menciona que por razones tanto históricas como prácticas, las FDA que suelen seleccionarse para representar los modelos de respuesta

dicotómica son la logística que da lugar a los modelos logit y la normal que da lugar a los modelos probit. Resultando así que la función de distribución logística (acumulativa) es la siguiente:

$$P_i = \frac{1}{1 + e^{-Z}} = \frac{e^Z}{1 + e^Z}$$

Donde $Z = \beta_1 + \beta_2 X_i$

Con esto se verifica que a medida que Z se encuentra dentro de un rango de $-\infty$ a $+\infty$, P_i se encuentra dentro de un rango de 0 a 1, y que P_i no está linealmente relacionado con Z , es decir con X_i . Al no ser lineal P_i no solo en X_i , sino también en la β , no se puede realizar la estimación de los parámetros con el método de Mínimos Cuadrados Ordinarios (MCO). Esto no representa un problema porque puede linealizarse de la siguiente forma:

Sabiendo que P_i es la probabilidad de que el evento suceda está dada por la función de distribución logística, entonces $(1 - P_i)$ es igual a la probabilidad de que no suceda el evento, representada como:

$$1 - P_i = \frac{1}{1 + e^Z}$$

Por consiguiente, podemos escribir:

$$\frac{P_i}{1 - P_i} = \frac{1 + e^Z}{1 + e^{-Z}} = e^Z$$

Con esto ahora se tiene que $\frac{P_i}{1 - P_i}$ es la razón de las probabilidades que están a favor de que suceda el evento.

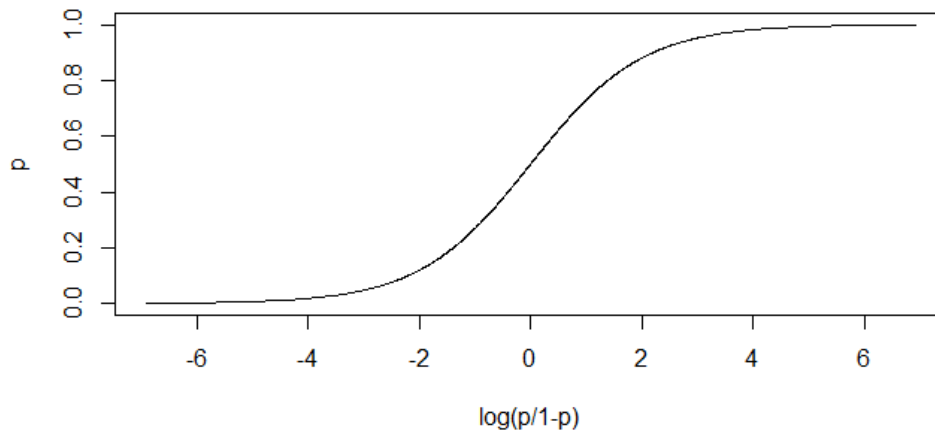
Ahora, si se toma el logaritmo natural de la ecuación anterior, se llega a:

$$L_i = \ln \left(\frac{P_i}{1 - P_i} \right) = Z = \beta_1 + \beta_2 X_i + u_i$$

Simplificando queda se la siguiente manera:

$$L_i = \ln \left(\frac{P_i}{1 - P_i} \right) = \beta_1 + \beta_2 X_i + u_i$$

L_i es el logaritmo de la razón de las probabilidades, que no sólo es lineal en X_i , sino también lo es para los parámetros. Siendo así que L_i se llama logit, y es de aquí el nombre de modelo logit para modelos como el anterior.



Gráfica 28. Modelo logístico

En la Gráfica 27 sobre el modelo logit se le conoce como sigmoide, en donde la curva es convexa al origen de forma creciente hasta el punto de inflexión convirtiéndose posterior a este en cóncava decreciente. De igual forma se observa que inicia en cero y termina en uno debido a que, como se menciona en el párrafo anterior se está realizando el cálculo de probabilidades, las cuales están delimitadas en este intervalo de 0 a 1.

Del modelo anterior los coeficientes se interpretan de la siguiente manera (Gujarati, 2003):

- β_2 es la pendiente, que mide el cambio en L_i que viene dado por un cambio unitario en X , en otras palabras, indica el cambio del logaritmo de las posibilidades a favor de que suceda el evento a medida que cambia en una unidad.
- β_1 es el intercepto, siendo el valor del logaritmo que indica las posibilidades de que el evento suceda cuando la variable independiente X_i en cero.

Para la estimación de los parámetros cuando se tienen datos a nivel individual, como lo es en el caso de los datos de la ENDUTIH que son a nivel usuario se recomienda usar el método de máxima verosimilitud por medio del software estadístico R.

Antes de la llegada de los modelos logit y probit, el modelo que se implementaba para la variable regresada dicotómica era el modelo lineal de probabilidad (MPL) que se estimaba mediante el método de Mínimos Cuadrados Ordinarios (MCO), lo que indica que debe cumplir con el supuesto de normalidad en los residuos, pero sí tendrán heteroscedasticidad. Esto no representa un problema como tal, sino que el problema de estos modelos radica en que, como bien sabemos las probabilidades van de 0 a 1 y tras hacer la estimación de los parámetros por MCO los valores en ocasiones serán inferiores a cero o superiores a uno, por lo que en estos casos se tendrá que asumir que si es negativo el estimador entonces el valor que debe tomar es cero, por el otro lado si es mayor a uno entonces se asumirá que el valor que tomará el estimador será uno (Gujarati, 2003).

Ante las limitaciones del modelo lineal de probabilidad surgen los modelos logit y probit que garantizan que las probabilidades estimadas estarán entre 0 y 1, además la regresión logística no supone linealidad como en modelos como de regresión clásica y puede existir heteroscedasticidad, lo que permite las estimaciones sea adapten mejor a la realidad, aunque requiere que las observaciones sean independientes y que las variables exploratorias estén relacionadas linealmente al logaritmo de la variable dependiente (Garson, 2014; citado en Quintana y Mendoza, 2016).

Continuando con el párrafo anterior y siguiendo con Garson (2014), quien menciona que debido a la no linealidad los modelos logit son estimados bajo el método de Máxima Verosimilitud. Este método indica que se elige como valor estimado del parámetro, aquél con mayor probabilidad de ocurrir según lo observado, en otras palabras, elige el valor más compatible con los datos observados, siempre suponiendo que es correcto el modelo propuesto (Molinero, 2003).

Los modelos logit forman parte de los conocidos Modelos Lineales Generalizados, siendo el logit un caso particular de estos, en los que la distribución es binomial y la función de enlace el algoritmo de las razones de probabilidad, es decir la función logit. Como bien se menciona ya no se busca un valor para y con una combinación lineal de los valores de $x_1 \dots x_i$ si no que se busca el logaritmo de las razones de probabilidad de $y = 1$ que se obtiene de dividir las probabilidades de acierto "1" por la probabilidad de fracaso "0" (Paladino, 2017).

Ahora bien, complementando el tema de la interpretación de los coeficientes resultantes de un modelo de este tipo, se identifica que no son tan intuitivos en su interpretación como lo es en una regresión lineal, dado que el modelo se compone de una función binomial y los resultados se refieren como probabilidades logarítmicas (UCLA, 2021).

Garson (2014) menciona los conceptos en los que se centra el modelo logit, en donde se encuentran los siguientes tres:

1. Odds: es ratio formado de la probabilidad de que ocurra el evento entre la probabilidad de que no ocurra el evento. De acuerdo con Paladino (2017) la transformación de probabilidad a odds ratio es monótona, es decir, a medida que incrementa la probabilidad incrementan los odds ratio y viceversa.
2. Log odds: Es el coeficiente resultante del modelo, denominado logit y es el logaritmo natural de las probabilidades de que la variable dependiente sea igual a algún valor, por lo que el logaritmo de las probabilidades es igual al logaritmo natural de la probabilidad de que se produzca el suceso dividido por la probabilidad de que no se produzca.
3. Odds Ratio: Es el ratio entre dos odds. Dentro del modelo logit es la medida de la magnitud de efecto que tiene el estimador sobre la variable independiente y se calcula con la exponencial de los coeficientes obtenidos.

En la siguiente tabla se coloca la relación entre probabilidad, odds y Log Odds en donde se idéntica la transformación monótona que existe entre sí.

Tabla 2: Relación entre Probabilidad, Odds y log Odds

Probabilidad (p)	Odds	Log Odds
0.001	0.001001	-6.906755
0.01	0.010101	-4.59512
0.15	0.176471	-1.734601
0.2	0.250000	-1.386294
0.3	0.428571	-0.8472978
0.35	0.538462	-0.6190392
0.4	0.666667	-0.4054651
0.45	0.818182	-0.2006707

0.5	1.000000	0
0.55	1.222222	0.2006707
0.6	1.500000	0.4054651
0.65	1.857143	0.6190392
0.7	2.333333	0.8472978
0.75	3.000000	1.098612
0.8	4.000000	1.386294
0.85	5.666667	1.734601
0.9	9.000000	2.197225
0.999	999.000000	6.906755
0.9999	9999.000000	9.21024

Fuente: Obtenido de UCLA (2021). Advanced Research Computing. Statistical Methods and data Analytics.

Una vez que se explicaron los términos anteriores, de acuerdo con Ferre (2014) los odds ratio son una buena medida del tamaño del efecto que tienen las variables exógenas sobre la variable dependiente. Cuya interpretación es que valores superiores a 1 indican que si el predictor aumenta los odds de la variable dependiente también aumenta. Inversamente si el valor es menor a 1 entonces el aumento de los odds de la variable dependiente provoca un decremento sobre la variable explicada. Para realizar esta transformación de a odds ratio se necesita obtener el exponente de las probabilidades logarítmicas o de los log odds.

Además de poder interpretar las probabilidades logarítmicas como odds ratios, también se puede llegar a una interpretación en términos de probabilidades bajo el siguiente procedimiento usando la ecuación del modelo logit:

$$L_i = \ln \left(\frac{P_i}{1 - P_i} \right)$$

Bajo la ecuación anterior se busca valor de la probabilidad (P)

$$\text{Exp}(L_i) = \frac{P_i}{1 - P_i}$$

$$(\text{Exp}(L_i))(1 - P_i) = P_i$$

Se desarrolla la multiplicación de ambos elementos

$$\text{Exp}(L_i) - \text{Exp}(L_i)P_i = P_i$$

Agrupación por componentes con variables en común

$$\text{Exp}(L_i) = P_i + \text{Exp}(L_i)P_i$$

Factorizando resulta de la siguiente manera

$$\text{Exp}(L_i) = P_i(1 + \text{Exp}(L_i))$$

Obtenido de esta forma el valor de la probabilidad

$$\frac{\text{Exp}(L_i)}{1 + \text{Exp}(L_i)} = P_i$$

Como bien se menciona anteriormente, con la ecuación final es posible interpretar los resultados de un modelo logit en términos probabilidades, en donde con la tabla 1 se identifica la relación de los log odds con las probabilidades diciendo así que cuanto más negativo sea el coeficiente obtenido por la regresión el resultado de éxito en variable explicada será menos probable, mientras que cuanto más positivo sea el coeficiente la probabilidad de éxito se incrementa.

Una vez que se explica y analiza la relación que existe entre los odds, odds ratios, log odds y probabilidades, para fines de esta investigación se implementa la interpretación de los coeficientes como probabilidades logarítmicas, debido a que el signo resultante para cada uno permite que a primera vista identificar si es una variable que impulsa la probabilidad de compra en Ecommerce por su signo positivos o si reduce la probabilidad de compra por un signo negativo. Sin embargo, en el anexo 1 se presentan las transformaciones de los coeficientes resultantes del modelo.

4.2. Aplicación Del Modelo Logit

Como se ha mencionado a lo largo de la investigación, se evalúa la probabilidad de realizar una compra en Ecommerce dadas las características sociodemográficas, económicas, gustos y preferencias del consumidor que es a su vez es usuario de internet utilizando datos de la

ENDUTIH 2021 y tras abordar cada una de las variables en el capítulo anterior se inicia con el proceso de modelado.

Un primer modelo que se desarrolla es aquel que considera todas las variables mencionadas y explicadas en el capítulo anterior. Debido a que se consideran diferentes variables cualitativas denotando las características del usuario se realiza el ajuste para poder separar cada variable categórica en distintas variables.

```
data_usuarios<-dummy_cols(data_usuarios,select_columns = c('Tiempo_utilizando_internet'  
, 'tipo_actividad', 'tipo_conexion_internet_hogar', 'ESTRATO', 'PAREN', 'nivel_estudio'  
, 'grupos_edad', 'TLOC', 'uso_redes_sociales', "utiliza_facebook", "utiliza_Twitter"  
, "utiliza_Instagram", "utiliza_LinkedIn", "utiliza_Snapchat", "utiliza_Whatsapp"  
, "utiliza_Youtube", "utiliza_Piterest", "utiliza_Messenger", "utiliza_Tumblr"  
, "utiliza_Otras_redes"),remove_selected_columns = FALSE)
```

Posterior a la división de las variables independientes se implementa un análisis de correlación. Este procedimiento es muy importante previo al desarrollo del modelo, ya que la correlación es una forma de medir la relación lineal entre dos variables cuantitativas continuas, es decir, si existe una covariación (varían conjuntamente) en el comportamiento de cada variable comparada con las demás variables incluidas en el modelo. Dando lugar a lo que se conoce como índice de correlación denotado como r (Vinuesa, 2016).

```
correlacion<-round(cor(d),1)
```

E índice de correlación es la medida del tamaño de efecto que existe de una variable sobre otra, cuyo valor puede variar entre -1 y $+1$, ambos extremos indicando correlación perfecta negativa y positiva respectivamente. Este indicador se puede interpretar de la siguiente forma (Vinuesa, 2016):

- Correlación despreciable: $r < |0.1|$
- Correlación baja: $|0.1| < r \leq |0.3|$
- Correlación mediana: $|0.3| < r \leq |0.5|$
- Correlación Fuerte o alta: $r > |0.5|$

Dicho lo anterior, con el análisis aplicado se detectaron variables con correlación perfecta por lo que se excluyó una de ellas. De esta manera el primer modelo, en donde la variable dependiente, en este caso la compra en Ecommerce, queda en función o explicada por todas

aquellas variables que no resultan tener un problema de multicolinealidad, la cual se define como la relación lineal “perfecta” o exacta, derivando en una indeterminación de los coeficientes de regresión (Gujarati, 2003).

Tras observar los resultados del primer modelo, se aplican pruebas de bondad de ajuste que permiten considerarlo un buen modelo como lo es el Pseudo-R cuadrado de McFadden, el cual consiste en comparar la mejora en la maximización del algoritmo de verosimilitud cuando es explicada la variable dependiente en función de las variables independientes, frente al modelo de la variable dependiente en función de una constante. Teóricamente, el Pseudo-R se mueve entre los valores 0 y 1; por lo tanto, cuando el resultado se encuentra entre 0.2 y 0.4 se considera que el modelo cuenta con un buen ajuste y muy bueno cuando es superior a estos valores (McFadden, 1976; citado en Tordable, 2016).

Al aplicar la prueba de McFadden el Pseudo-R cuadrado es de 0.298375 para el primer modelo el cual considera todas las variables, por lo que se puede decir que es un buen modelo de acuerdo con lo mencionado en el párrafo anterior.

```
> DescTools::PseudoR2(model1, which = "McFadden")
McFadden
0.298375
```

Aunque el primer modelo que contempla todas las predictoras presenta un buen ajuste se recurre a aplicar un algoritmo de selección de variables; implementando así la metodología de Step Wise, el cual es una combinación de los métodos Forward y Backward en donde comienza con la introducción progresiva de variables y en cada etapa se plantea si todas las variables agregadas deben permanecer en el modelo (González, 2015).

Posterior a implementar este algoritmo se tiene como resultado diferentes modelos, de los cuales se tienen que elegir el mejor, por lo que se considera una medida global de cada modelo, llamada Criterio de Akaike (AIC), en donde se busca el valor más pequeño, ya que en ese caso indica existencia una mayor verosimilitud y pocos parámetros (González, 2015).

```
> resumen_m1$aic
[1] 38286.9
> resumen_m2$aic #mejor ajuste
[1] 38269.8
```

Como bien se menciona en los párrafos anteriores, posterior a aplicar el método de Step Wise se compara el criterio de Akaike de cada modelo, siendo el modelo 2 el mejor, el cual se muestra a continuación:

compra_ecommerce

= *horas_uso_prom_x_dia_internet* + *dispone_celular*
+ *celular_smartphone* + *uso_compu_lap_tablet*
+ *uso_compu_lap_tablet* + *conexion_por_red_cel*
+ *Tiempo_utilizando_internet_Entre 01 y hasta 2 años*
+ *Tiempo_utilizando_internet_Más de 5 años*
+ *Tiempo_utilizando_internet_Menos de 1 año*
+ *tipo_actividad_jornalero(a) o peón*
+ *tipo_actividad_patrón(a) o empleador(a)*
+ *tipo_actividad_Sin actividad*
+ *tipo_actividad_trabajador(a) por cuenta propia*
+ *tipo_conexion_internet_hogar_Móvil*
+ *tipo_conexion_internet_hogar_Sin conexión en el hogar*
+ *ESTRATO_Alto* + *ESTRATO_Medio alto* + *ESTRATO_Medio bajo*
+ *PAREN_Esposo(a) o compañero(a)* + *PAREN_Jefe(a) del hogar*
+ *nivel_estudio_Doctorado* + *nivel_estudio_Lic o Ing*
+ *nivel_estudio_Maestría* + *nivel_estudio_Preescuela o kínder*
+ *nivel_estudio Primaria* + *nivel_estudio_secundaria*
+ *grupos_edad_15 a 24 años* + *grupos_edad_25 a 34 años*
+ *grupos_edad_35 a 44 años* + *grupos_edad_55 años y más*
+ *TLOC_100,000 y más habitantes*
+ *TLOC_15,000 a 99,999 habitantes*
+ *TLOC_2,500 a 14,999 habitantes* + *uso_redes_sociales_Sí*
+ *utiliza_facebook_Sí* + *utiliza_Twitter_Sí* + *utiliza_Instagram_Sí*
+ *utiliza_Youtube_Sí* + *utiliza_Messenger_Sí*

Dicho modelo también fue valorado con la Pseudo-R cuadrada de McFadden el cual resulta ser de 0.298211, cayendo dentro de los valores que permiten a este segundo modelo considerarlo como un buen modelo.

```
> DescTools::PseudoR2(modelo_step_wise, which = "McFadden")
McFadden
0.298211
```

Una vez validado la bondad de ajuste, se comienza con la revisión de los coeficientes obtenidos en donde cabe resaltar que todos son estadísticamente significativos y cuya interpretación se presenta a continuación:

Tabla 3. Coeficientes obtenidos del modelo logístico

Variable	Estimate	Pr(> z)	interpretación
(Intercept)	-4.47	0.00	El coeficiente del intercepto es el origen de la ordenada, el cual resulta ser significativo con signo negativo indicando que las probabilidades de que un usuario realice una compra por Ecommerce son menores que el no realizar la compra por este medio. Diciendo así que existen más personas que hoy no hacen uso del comercio electrónico que las que si recurren a él como una vía de satisfacción de necesidades.
horas_uso_prom_x_dia_internet	0.06	0.00	El coeficiente de la variable presenta un signo positivo por lo tanto ante un aumento de una unidad en las horas de uso promedio de internet al día aumenta la probabilidad de compra en Ecommerce en 0.06 probabilidades logarítmicas. Cuanto más tiempo pase navegando por la red, un usuario podría adoptar las compras online.

dispone_celular1	0.26	0.02	Para esta variable se obtiene un coeficiente positivo por lo tanto indica que si el usuario dispone de un teléfono celular aumenta 0.26 las probabilidades logarítmicas de que realice una compra en Ecommerce.
celular_smatphone1	0.37	0.01	En esta variable se verifica que si el teléfono celular que posee el usuario es de tipo smartphone entonces aumenta 0.37 las probabilidades logarítmicas de que realice una compra en Ecommerce.
uso_compu_lap_tablet1	0.67	0.00	El uso de computadora resulta ser una variable con mayor significancia estadística y cuyo coeficiente es positivo por lo que indica que si el individuo en efecto utiliza este dispositivo las probabilidades logarítmicas de que realice una compra online aumenta +0.67
asiste_escuela1	-0.23	0.00	Dentro del modelo se evalúa a los usuarios que son estudiantes y como resultado se obtiene un coeficiente negativo, represando de esta forma que el ser estudiante reduce las probabilidades logarítmicas de que el individuo realice una compra por Ecommerce.
conexion_por_red_cell1	0.44	0.00	Si el usuario cuenta con conexión a internet entonces las probabilidades logarítmicas de que realice una compra por medio del comercio electrónico aumentan 0.44.
`Tiempo_utilizando_internet_Entre 01 y hasta 2 años`1	-0.23	0.00	En esta variable se identifica si el usuario lleva entre 1 y 2 años utilizando internet entonces las probabilidades logarítmicas de realizar una compra por este medio se reducen en -0.23.

`Tiempo_utilizando_internet_Más de 5 años`1	0.41	0.00	Evaluando el tiempo de experiencia en cuanto al uso de internet cuando este es mayor de 5 años las probabilidades logarítmicas de realizar una compra por internet aumentan 0.41.
`Tiempo_utilizando_internet_Menos de 1 año`1	-0.43	0.00	En la siguiente variable se hace referencia a si el usuario tiene menos de un año utilizando internet, con lo cual afecta de forma negativa la probabilidad logarítmica en -0.43 de que un individuo realice una compra online.
`tipo_actividad_jornalero(a) o peón`1	-0.31	0.01	Considerando el tipo de actividad que realiza el individuo, por lo que indica que si el usuario es jornalero o peón las probabilidades logarítmicas de que realice una compra por internet disminuyen -0.31.
`tipo_actividad_patrón(a) o empleador(a)`1	0.63	0.00	Si el usuario tiene un papel de patrón o empleador aumenta 0.63 las probabilidades logarítmicas de comprar mediante el comercio electrónico.
`tipo_actividad_Sin actividad`1	-0.22	0.00	Cuando el usuario no desempeña una actividad las probabilidades logarítmicas de realizar compras medio de internet se reducen -0.22.
`tipo_actividad_trabajador(a) por cuenta propia`1	0.29	0.00	Si el usuario es trabajador por su cuenta propia entonces las probabilidades logarítmicas de que recurra al comercio eléctrico se incrementan en +0.29.
tipo_conexion_internet_hogar_Móvil1	-0.34	0.00	Las probabilidades logarítmicas de que un usuario de internet haga una compra por este medio se ven influenciadas negativamente en -0.34 si este tiene un tipo de conexión móvil dentro del hogar disminuyendo.

`tipo_conexion_internet_hogar_Sin conexión en el hogar`1	-0.52	0.00	Si el usuario no cuenta con algún tipo de conexión a internet dentro del hogar entonces las probabilidades logarítmicas de que realice una compra en Ecommerce se reduce en -0.52.
ESTRATO_Alto1	0.45	0.00	El estrato hace referencia al nivel socioeconómico al que pertenece el usuario por lo tanto de acuerdo con los resultados del modelo, si el individuo es de un nivel Alto entonces las probabilidades logarítmicas de que realice una compra online aumentan 0.45.
`ESTRATO_Medio alto`1	0.28	0.00	Si el usuario pertenece a un nivel socioeconómico Medio alto las probabilidades logarítmicas de recurrir al comercio electrónico se ven favorecidas en +0.28.
`ESTRATO_Medio bajo`1	0.17	0.00	Cuando el individuo es de un nivel medio Bajo aumenta las probabilidades logarítmicas de realizar una compra en Ecommerce +0.17.
`PAREN_Esposo(a) o compañero(a)`1	0.26	0.00	Una de las variables a contemplar es el parentesco que existe entre el individuo y el jefe del hogar por lo que sí es el esposo (a) o compañero(a) las probabilidades logarítmicas de recurrir al comercio electrónico aumentan +0.26.
`PAREN_Jefe(a) del hogar`1	0.26	0.00	Si el individuo es el jefe (a) del hogar aumentan 0.26 las probabilidades logarítmicas de realizar una compra online.

nivel_estudio_Doctorado1	1.24	0.00	Se evalúa el grado o nivel de estudios que ha alcanzado el usuario por lo que si tiene un doctorado aumentan las probabilidades logarítmicas de realizar una compra en Ecommerce 1.24 resultando ser aquellos que presentan una fuerte relación con la variable dependiente.
`nivel_estudio_Lic o Ing`1	0.31	0.00	Cuando el usuario tiene una licenciatura o una ingeniería las probabilidades logarítmicas de que realice una compra por medio del comercio electrónico en +0.31.
nivel_estudio_Maestría1	0.56	0.00	si el agente económico tiene una maestría las probabilidades logarítmicas de realizar una compra online aumenta +0.56.
`nivel_estudio_Preescolar o kínder`1	-2.57	0.01	Cuando el usuario solo ha concluido preescolar entonces las probabilidades logarítmicas de que realice una compra por Ecommerce se ven reducidas en -2.57.
nivel_estudio Primaria1	-0.56	0.00	Si el usuario solo llegó al nivel de estudios primaria se reducen las probabilidades logarítmicas -0.56 de que realice una compra en línea.
nivel_estudio_secundaria1	-0.17	0.00	Por último, el tener nivel de estudios hasta secundaria reduce las probabilidades de realizar una compra en Ecommerce en -0.17
`grupos_edad_15 a 24 años`1	0.62	0.00	La edad en las personas resulta de factor relevante para saber a qué tipo de público es más propenso a poder acercar al comercio electrónico, por lo que si el individuo tiene entre 15 y 24 años entonces las probabilidades de recurrir a este tipo de comercio aumentan las probabilidades logarítmicas +0.62.

`grupos_edad_25 a 34 años`1	0.65	0.00	Si el usuario tiene entre 25 y 34 años, las probabilidades de comprar en línea aumentan +0.65, siendo este el rango de edad con más posibilidades de recurrir a estas compras.
`grupos_edad_35 a 44 años`1	0.41	0.00	Cuando el usuario tiene una edad entre 35 y 44 años entonces las probabilidades de que realice compras por internet aumentan +0.41.
`grupos_edad_55 años y más`1	-0.42	0.00	Si el usuario se encuentra dentro del último rango de edades englobando a los mayores de 55 años las probabilidades logarítmicas de que recurra al comercio electrónico reducen -0.42.
`TLOC_100,000 y más habitantes`1	0.27	0.00	El tamaño de la localidad en la que radica el usuario resulta ser una variable significativa para el modelo, en donde se muestra que si este pertenece a una localidad de 100,000 y más habitantes entonces aumentan las probabilidades logarítmicas en +0.27 de que realice una compra online.
`TLOC_15,000 a 99,999 habitantes`1	0.19	0.00	Cuando el individuo pertenece a una localidad de entre 15,000 y 99,999 habitantes las probabilidades logarítmicas de que realice una compra en Ecommerce aumenta +0.19.
`TLOC_2,500 a 14,999 habitantes`1	0.17	0.00	Si el usuario pertenece a una localidad de 2,500 a 14,999 habitantes entonces las probabilidades logarítmicas de que realice una compra en línea aumentan +0.17.

uso_redes_sociales_Sí1	0.39	0.00	El uso de redes sociales se ha vuelto algo común en los internautas, por lo que se evalúa dentro del modelo la relación que tiene con respecto a la compra por medio de Ecommerce. Lo que resulta en que el usar redes sociales afecta de manera positiva las probabilidades logarítmicas de realizar dichas compras en +0.39
utiliza_facebook_Sí1	0.49	0.00	Si el individuo usa Facebook aumenta las probabilidades logarítmicas en +0.49 de comprar por comercio electrónico.
utiliza_Twitter_Sí1	0.51	0.00	Si el usuario utiliza Twitter aumentan las probabilidades logarítmicas de compra en +0.51
utiliza_Instagram_Sí1	0.72	0.00	Si el individuo utiliza Instagram beneficia las probabilidades incrementado las en +0.72 de realizar una compra online.
utiliza_Youtube_Sí1	0.12	0.00	Si el usuario declaro que utiliza YouTube se dice que las probabilidades logarítmicas de realizar una compra en línea aumentan +0.12
utiliza_Messenger_Sí1	0.11	0.00	Si el individuo utiliza Messenger las probabilidades logarítmicas de que recurra al Ecommerce para comprar un bien aumenta +0.11

Ahora bien, una vez que se analizan los coeficientes obtenidos de la regresión logística, se puede identificar el efecto y magnitud que tiene cada variable sobre la probabilidad de compra por internet. Por lo que, se procede a replicar el modelo a toda la base de datos, es decir, evaluar a cada individuo, con el objetivo de obtener para cada usuario de internet, la probabilidad que tiene de realizar una compra en línea una vez que se ha contemplado la totalidad de variables dentro del modelo, aplicando el comando predict() en R Project.

```
resultados_model2<-as.data.frame(predict(modelo2,type="response"))
```

Posterior a la aplicación de modelo se filtra la base para obtener solo a los usuarios de internet que declaran en la encuesta no haber realizado una compra por medio del comercio electrónico.

```
sin_compra_ecommerce<-filter(data_usuarios_predict,compra_ecommerce==0)
```

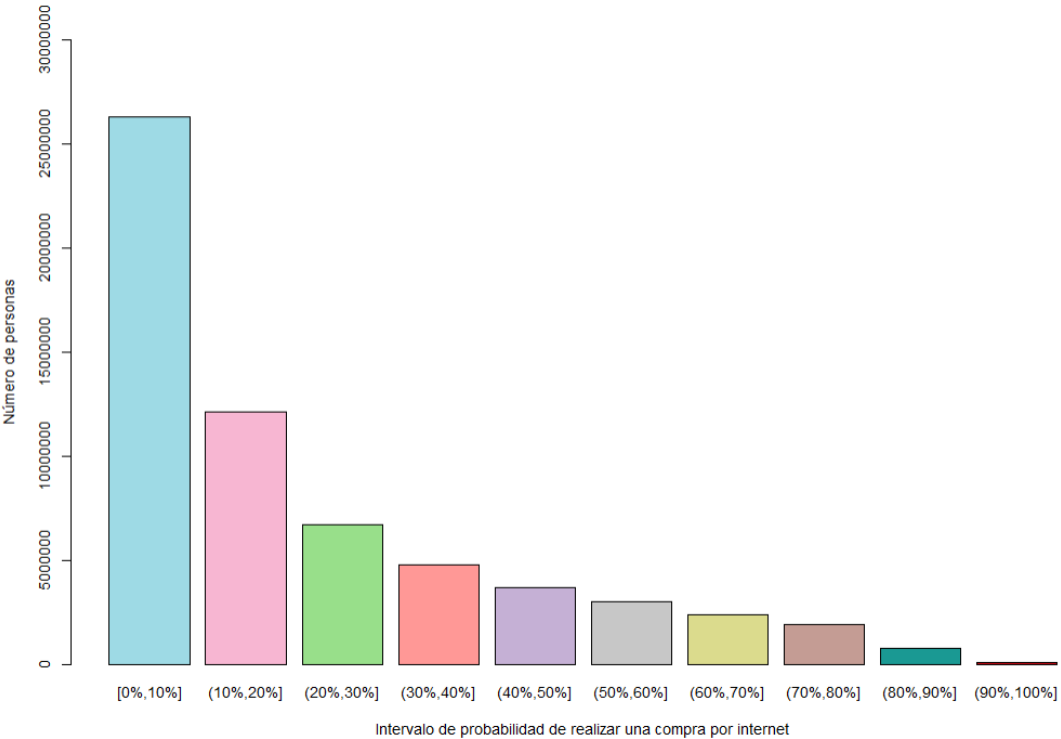
De esta manera, como se menciona en secciones anteriores existen un total de 61.9 millones de individuos con acceso a internet que no realizan compras por este medio, lo que indica que es el mercado potencial. En la tabla 4 se muestra la distribución de estos usuarios por intervalos de probabilidad de compra que se obtuvo del modelo logístico, diciendo así que el 42.5% de los usuarios (26.3 millones de individuos) tienen menos del 10% de probabilidad de comprar por medio del comercio electrónico, mientras que solo un 13.3% (8.25 millones de individuos) tienen una probabilidad mayor o igual a 50% de realizar una compra por internet.

Tabla 4. Población de usuarios de internet que aún no compran por Ecommerce distribuida por en intervalos de probabilidad de compra.

Intervalo	usuarios de Internet que aún no compran por Ecommerce	Frecuencia relativa
[0%,10%]	26,305,728	42.5%
(10%,20%]	12,118,573	19.6%
(20%,30%]	6,735,523	10.9%
(30%,40%]	4,772,799	7.7%
(40%,50%]	3,717,582	6.0%
(50%,60%]	3,039,554	4.9%
(60%,70%]	2,411,330	3.9%
(70%,80%]	1,920,990	3.1%
(80%,90%]	793,657	1.3%
(90%,100%]	85,395	0.1%
Total	61,901,131	100%

La tabla anterior en forma gráfica se puede apreciar a continuación en la gráfica 29, en donde claramente se muestra un sesgo a la izquierda, validándose que existe una mayor cantidad de

usuario con probabilidades bajas de realizar una compra por medio de internet y conforme es mas alto el rango de las probabilidades de compra se aminora la cantidad de individuos.



Gráfica 29. Distribución de usuario de internet por probabilidad de realizar una compra por medio del comercio electrónico.

Fuente: Elaboración propia con datos de la ENDUTIH 2021.

4.3. Resultados y Discusión Del Modelo Logit

El objetivo del que se parte en la presente investigación gira en torno a la comprensión del consumidor dentro del mercado, ahondando específicamente en el comportamiento respecto a la compra por medio del comercio electrónico, utilizando datos de la encuesta ENDUTIH de 2021, con la cual se identifica el uso que le dan a las tecnologías de la información y comunicación, considerando si es para realizar compras en línea, así como las características de sociodemográficas, económicas, gustos y preferencias, cuantificado por medio del modelo econométrico Logístico.

En la investigación se aborda la teoría del consumidor, en la que se menciona que los individuos son libres de elegir lo que quieren comprar en el mercado, basados en sus

preferencias y considerando la restricción del ingreso monetario con el que cuentan. Las preferencias marcan la diferencia en el momento de tomar una decisión sobre un bien u otro, las cuales son heterogéneas entre cada consumidor, sin embargo, puede existir un patrón de consumo de acuerdo con Cue (2024).

Retomando el párrafo anterior, en cuanto a los patrones que se pueden generar entre los individuos resulta ser importante dentro de la compra en Ecommerce, dado a que una primera similitud es precisamente la elección de adquirir bienes y servicios por este tipo de comercio. Basados en esto, el modelo logit que se desarrolla en la investigación detecta los patrones que existen entre los usuarios de internet, validando la relación que existe entre cada variable con respecto a la compra por internet, llegando de esta manera a tener una visión de cómo interactúan las características de los consumidores en sus decisiones de consumo.

Gracias a la información proporcionada por la ENDUTIH 2021 se idéntica que 75.6% de la población de México equivalente a 88.56 millones de personas cuenta con acceso a internet de los cuales solo el 30% afirma que realizó una compra por este medio, lo que implica que existe un 70% de usuarios que aún no realizan compras por este medio, por lo tanto, el tamaño del mercado es de 61.9 millones de personas que navegan por la red y con los estímulos adecuados tienen potencial para poder formar parte los compradores de Ecommerce.

Para las empresas de este giro, la presente investigación les brinda información acerca del mercado potencial que existe, con este conocimiento tienen noción de qué tipo de estrategias se pueden implementar para atraer clientes. Por tanto, resulta relevante conocer a los usuarios a los que se están dirigiendo, de tal forma que en el análisis se seleccionan 24 variables principales, que hacen alusión a las características sociodemográficas, económicas, gustos y preferencias de los individuos.

```

glm(formula = compra_ecommerce ~ horas_uso_prom_x_dia_internet +
  dispone_celular + celular_smatphone + uso_compu_lap_tablet +
  asiste_escuela + conexion_por_red_cel + `Tiempo_utilizando_internet_Entre 01 y hasta 2 años` +
  `Tiempo_utilizando_internet_Más de 5 años` + `Tiempo_utilizando_internet_Menos de 1 año` +
  `tipo_actividad_jornalero(a) o peón` + `tipo_actividad_patrón(a) o empleador(a)` +
  `tipo_actividad_Sin actividad` + `tipo_actividad_trabajador(a) por cuenta propia` +
  tipo_conexion_internet_hogar_Móvil + `tipo_conexion_internet_hogar_Sin conexión en el hogar` +
  ESTRATO_Alto + `ESTRATO_Medio alto` + `ESTRATO_Medio bajo` +
  `PAREN_Esposo(a) o compañero(a)` + `PAREN_Jefe(a) del hogar` +
  nivel_estudio_Doctorado + `nivel_estudio_Lic o Ing` + nivel_estudio_Maestría +
  `nivel_estudio_Preescola o kínder` + nivel_estudio_Primeria +
  nivel_estudio_secundaria + `grupos_edad_15 a 24 años` + `grupos_edad_25 a 34 años` +
  `grupos_edad_35 a 44 años` + `grupos_edad_55 años y más` +
  `TLOC_100,000 y más habitantes` + `TLOC_15,000 a 99,999 habitantes` +
  `TLOC_2,500 a 14,999 habitantes` + uso_redes_sociales_Sí +
  utiliza_facebook_Sí + utiliza_Twitter_Sí + utiliza_Instagram_Sí +
  utiliza_Youtube_Sí + utiliza_Messenger_Sí, family = binomial(link = "logit"),
  data = data_usuarios)

```

Las 24 variables seleccionadas se consideran de tipo continuas y categóricas. Las variables categóricas fueron transformadas, separando y convirtiendo en variables dummy, resultando un nuevo total de 79 variables, con lo cual, ante la primera prueba de multicolinealidad facilita la comparación de cada variable en la matriz de correlación, permitiendo detectar la existencia de correlación perfecta entre los diferentes niveles que componen a las variables categóricas.

Si bien, entre más información exista acerca de los usuarios, no necesariamente toda es relevante para poder explicar la probabilidad de compra en Ecommerce, es por esto, que la implementación de los algoritmos para la selección de variables dentro de un modelo como lo es el método de Step Wise, el cual parte de un primer modelo en donde se incluyen las 79 variables y tras realizar todas las iteraciones combinando los métodos de Backward y Forward permitiendo llegar a un mejor modelo con 37 variables relevantes para explicar a la variable en cuestión. Sin embargo, se evalúan ambos modelos para poder validar que se elija al mejor.

```
seleccion<-step(model1)
```

El aporte de los modelos logit se ve sustancial en el análisis de problemáticas cuyos resultados se limitan a éxito o fracaso, ausencia o presencia de un fenómeno, en otras palabras, permiten explicar variables dicótomas, en término de probabilidades. En el caso de la presente se consideraron un total de 37 variables resultantes de aplicar el método de Step Wise, las cuales en su totalidad resultaron ser estadísticamente significativas, obteniendo así un modelo parsimonioso.

En un proceso de modelado, uno de los pasos relevantes es la aplicación de pruebas de bondad de ajuste, éstas permiten validar si efectivamente el modelo captura la información necesaria para explicar a la variable dependiente. Para efectos de la investigación se estima el pseudo R cuadrado de McFadden, el cual fue aplicado al primer modelo que contempla todas las variables y al segundo modelo que contiene solo las variables elegidas por el algoritmo de Step Wise.

Con la metodología antes mencionada, ambos modelos evaluados y presentan un buen ajuste dado que el valor de la R cuadrada de McFadden se encuentra en 0.2 y 0.4, sin embargo, con esta prueba se verifica que el modelo que sólo contiene las variables resultantes del algoritmo de selección tiene un valor ligeramente más alto, lo que indica un mejor ajuste comparado con el modelo que considera todas las variables, respaldando la afirmación de que no toda la información disponible es necesaria para poder tener un buen modelo y explicar a la variable dependiente.

Dado el buen ajuste del modelo con variables seleccionadas, los resultados que proporciona permiten entender e inferir sobre el comportamiento de los usuarios de internet ante la compra por medio del Ecommerce, es decir, el modelo permite establecer cuáles son las variables que intervienen al momento de realizar una compra electrónica, y mejor aún, indica en términos probabilísticos cada factor.

Una de las variables relevantes es el nivel de estudios de los usuarios, de manera particular, el nivel doctorado es el que más favorece la compra en Ecommerce, en segundo lugar, se posiciona el grado de maestría y licenciatura o ingeniería. Por el lado contrario, se identifica que existen niveles de estudios que reducen la probabilidad de compra, como lo es el grado de secundaria, primaria y preescolar.

Con las observaciones anteriores, se argumenta que entre mayor nivel de estudios tenga el individuo la compra por internet puede ser una de las alternativas que utilice para adquirir sus bienes y servicios, además esto se relaciona con que cuenta con mayor preparación académica y se deduce que cuentan con la mayoría de edad, estos elementos les permiten que indaguen en la diversidad de formas que pueden adquirir un bien y las adopten tras averiguar el funcionamiento; mientras que, en los niveles básicos de educación, los

individuos generalmente tiene menos experiencia en estas prácticas, además de que algunos de ellos aún son menores de edad y por ende este tipo de comercio no resulte en una alternativa viable para su consumo.

Se considera la relación que tienen las redes sociales sobre la compra en Ecommerce, en donde, se verifica que tienen un efecto positivo, específicamente el uso de Instagram es la segunda variable que más favorece la compra. Si bien esta red social no fue la más usada en México de acuerdo con el estudio de Statista (2023), resulta ser un promotor de las compras por internet. Además, se identifica que Twitter, Facebook, YouTube y Messenger incentivan la compra en diferente medida, por lo que estrategias de marketing en redes sociales son una opción con altas probabilidades de convertir a un usuario en consumidor.

La tercera variable de mayor impacto positivo en la compra por comercio electrónico es el tipo de dispositivo al que tienen acceso, en este caso el que se evalúa y es de relevancia para explicar la variable dependiente es el acceso a una computadora o Tablet, esto se relaciona con que la calidad de imagen por las pulgadas del display de estos aparatos es mejor que otros dispositivos. Por lo tanto, con este hallazgo se puede concluir que es importante el diseño de las páginas web para una mejor experiencia frente a los clientes potenciales, logrando que búsqueda se culmine en una compra.

Además de considerar el uso de laptop o Tablet se integra en el modelo la posesión de un celular tipo smartphone, un teléfono inteligente por el que el individuo puede acceder a internet. Si el usuario tiene este dispositivo móvil, las probabilidades de compra en línea aumentan, ya que permite acceder a las aplicaciones de las empresas dedicadas al comercio electrónico donde estén siempre y cuando tengan la conexión a internet.

Dentro de la investigación se buscan las variables que favorecen en mayor medida la compra en línea y con esto poder identificar las variables que aportan información valiosa aunque no sean las que favorezcan mayormente la variable dependiente, como lo es la edad, que al dividirla en intervalos se identifica que los usuarios de 25 a 34 años son los que presentan una mayor probabilidad de compra, el segundo rango de edad con mayor probabilidad de es de 15 a 24 años, quienes son individuos que están más en contacto con las tecnologías

haciéndolos más hábiles comparado con usuarios de más 50 años los cuales presentan un coeficiente negativo, es decir que tienen menos probabilidad de hacer compras por internet.

En relación con el párrafo anterior, la edad es un factor relevante dado a que en la época de la llegada del internet en 1990 las generaciones cercanas a estas fechas crecieron o se adaptaron al uso este servicio y actualmente se encuentran en los rangos de edad que resultan ser relevantes para la compra de Ecommerce, por lo que, el tiempo que llevan utilizando internet basado en los resultados del modelo indica que es una variable significativa con la cual se puede decir que si un usuario tiene más de 5 años utilizando el servicio las probabilidades de compra se aumentan, mientras que si se observa el rango de menos de 1 año y de entre 1 y 2 años se ve afectada la probabilidad de compra de forma negativa. Dejando así en evidencia que entre mayor sea la experiencia en cuanto al uso del internet mayor es la posibilidad de que compre por este medio.

En la investigación se considera la variable Estrato la cual hace referencia al nivel socioeconómico de los usuarios, en donde se destacan 3 que son: nivel Alto el cual tiene el un coeficiente más mayor que favorece la compra en Ecommerce y los niveles Medio Alto y Medio bajo que de igual forma tienen un efecto positivo en la compra en línea. Por lo que el pertenecer a un estrato económico más alto, aumenta las probabilidades de realizar una compra electrónica.

Además de considerar el nivel socioeconómico, el tener en cuenta el tipo de actividad que realiza el usuario es relevante para seguir construyendo el perfil con mayores probabilidades de realizar compras por internet. Dicho lo anterior, se identifica que si el individuo es patrón o empleador tiene mayor probabilidad hacer una compra por internet, relacionando esta actividad con un mayor cargo y con mejores ingresos económicos. Otro tipo de actividad que se relaciona de forma positiva con la compra es ser trabajador por cuenta propia, que también se puede relacionar con un mejor ingreso dado a negocios, emprendimientos propios en donde las ganancias pueden ser mayores y generalmente se asocia con habilidades de autogestión, por lo que una compra electrónica puede facilitar su actividad.

Si bien ya se mencionaron los factores que favorecen la compra, se detecta que existen ocupaciones desempeñadas por los usuarios que reduce la probabilidad de compra en

Ecommerce, como lo es el no tener una actividad, dado a que esto se relaciona con la falta de ingresos económicos para la obtención de bienes y servicios. Por último, el ser jornalero o peón reduce las probabilidades de compra dado a que es un empleo con remuneraciones bajas, además de que se relaciona con un nivel de escolaridad básico, con lo cual la compra en línea no se considera como opción de para consumo dada la falta de preparación para poder manipular las tecnologías.

Con la regresión logística se tiene visibilidad del efecto que tiene cada variable exógena sobre la variable dependiente, por consiguiente, con la aplicación de modelo econométrico a la base de datos completa, se consigue la probabilidad de compra para cada individuo, de acuerdo con sus características, evidenciando que aquellos usuarios que no realizaron una compra por medio del comercio electrónico durante 2021 se aglomeran en probabilidades bajas menores a 10%. Conforme se analizan los intervalos de probabilidades más elevados se observa como reduce la cantidad de individuos, dejando solo 8.25 millones de individuos con probabilidades compra mayores o iguales a 50% de formar parte del Ecommerce.

Como ya se ha hecho mención, la presente investigación busca entender el comportamiento de los usuarios de internet ante la compra en Ecommerce, por lo que resulta relevante para las empresas pertenecientes a esta industria, primero permite conocer el tamaño del mercado que aún pueden capturar, sabiendo que no todo usuario de internet realiza compras por este medio y segundo, el modelo aplicado ayudan en la toma de decisiones de estas empresas, ya que los resultados de dicho modelo permiten entender el tipo de audiencias que presentan mayores probabilidades de realizar compras en línea; entender sus características como edad, nivel socioeconómico, escolaridad entre otras, con las cuales se pueden diseñar estrategias de marketing que permitan atraer a los usuarios con potencial para adoptar este tipo de comercio en la rutina habitual.

Por otro lado, la cobertura de internet en todo México se ha visto en aumento en los últimos años, gracias a la llegada de diferentes empresas privadas que proporcionan este servicio por medio de cuotas mensuales. Además de programas de cobertura social implementados por el gobierno para dar acceso y promover el uso de las TIC's en todo México en donde tienen como objetivo establecer las bases para promover el incremento en la cobertura de las redes y la penetración de servicios de telecomunicaciones teniendo en cuenta a aquellas

comunidades en donde aún no con alto y muy alto grado de marginación (Secretaría de Comunicaciones y Transporte, 2019).

Como parte de los beneficios que proporciona la investigación, resulta de utilidad a los estudiantes de carreras como Actuaría, Economía, Relaciones Económicas Internacionales, por mencionar algunas, que se relacionan con temas econométricos, dada la metodología que se implementa sobre modelos logísticos, el desarrollo teórico sobre esta y la implementación del modelo resultan como material de apoyo en problemáticas a los que se pueda dar respuesta bajo el mismo método.

Por otro lado, la investigación que se desarrolla tiene futuras aplicaciones sobre la misma base de datos proporcionada por la ENDUTIH ya que dicha encuesta tiene una actualización anual, con lo que se puede monitorear si al paso del tiempo las características que se detectan en el presente modelo como relevantes para incrementar la probabilidad de compra en Ecommerce continúan siendo relevantes y comprobar la existencia de nuevas características que comiencen a ser significativas.

Conclusiones

La presente investigación, se desarrolla en cuatro capítulos, en los cuales se da contexto, fundamento y repuesta al objetivo general de analizar el comportamiento y hábitos de compra de los consumidores mexicanos en el comercio electrónico por medio de variables sociodemográficas, gustos y preferencias y económicas, bajo la perspectiva microeconómica y a través de un modelo logístico. Con esto en un inicio se indaga sobre la teoría del consumidor, la cual es fundamental dado el objetivo que se plantea sobre entender a los individuos entorno a las compras que lleva a cabo, para satisfacer necesidades de consumo.

La importancia de profundizar en el consumidor radica en que la economía de un país está integrada por miles de ellos, los cuales toman decisiones de consumo buscando maximizar su beneficio. La conducta que adopta el consumidor ante sus decisiones se ve afectada por sus preferencias, la restricción presupuestal y elecciones de el mismo consumidor en cuanto a cómo combinar sus bienes para maximizar su utilidad. A pesar de que cada individuo tiene gustos particulares, es posible establecer axiomas de comportamiento.

Como lo establece la teoría económica, un individuo es libre de elegir lo que quiere consumir en el mercado según sus preferencias y bajo su ingreso monetario, lo que lo convierte en un consumidor soberano, capaz de decidir los bienes y servicios que desea consumir. Con este entendimiento de libertad que poseen los individuos, se da a entender que con tal soberanía también pueden optar qué tipo de comercio es el adecuado (comercio en tienda en físico o comercio electrónico), es decir, cuál es el que ofrece una mejor variedad de bienes y servicios a precios que puedan ser cubiertos por el presupuesto.

Dicho lo anterior, el Ecommerce tras su evolución, de pasar a ventas por catálogos digitales a la creación de los actuales Marketplace, los cuales tiene el potencial para ser el proveedor de bienes y servicios dado sus portafolios de estos y con facilidad de compra sin necesidad de trasladarse. Por esta razón durante la pandemia por COVID-19 en el año 2020 tras las restricciones de circulación, este tipo de comercio tuvo un crecimiento exponencial en poco tiempo, debido a que se convirtió en la forma eficiente de poder satisfacer necesidades, dada la accesibilidad y la posibilidad de recibir su comprar a domicilio, lo que generó una sinergia, ya que quienes ya realizaban compras electrónicas las aumentaron y se sumaron muchos nuevos compradores digitales.

La pandemia resultó ser un parteaguas para el comercio electrónico, mientras las tiendas físicas por medidas de sanidad se tuvieron que mantener cerradas la forma de continuar con sus operaciones de ventas fue hacer uso de las herramientas electrónicas, haciendo que el comercio electrónico en México se expandiera de una forma acelerada, llegando a crecimientos de 300% de un año a otro, en lo que se estimaba llevaría en 5 o 10 años, resultando uno de los primeros cinco países con mayor auge en Ecommerce a nivel mundial.

Ante el acelerado crecimiento de las compras en línea surge la necesidad de tener información acerca de la industria y sobre los consumidores que forman parte de ella. Por lo que, basados en la literatura se explora la teoría económica, con el fin de entender el comportamiento de un individuo ante diferentes factores; como son los cambios de precios, el tamaño y eficacia del portafolio de bienes y servicios que ofertan dentro del comercio electrónico, teniendo así la oportunidad de conocer al cliente y de esta forma identificar qué le resulta de interés para poder mantenerse activo en las compras electrónicas.

Este tipo de estudios son de relevancia no solo para las empresas dedicadas al Ecommerce, sino también para aquellas que usan canales físicos y digitales para realizar ventas, así como para las empresas que están evaluando formar parte de este tipo de comercio, ya que ayuda a dar una noción de marketing, en donde una vez sabiendo que tipo de clientes hay en el mercado es más fácil poder estructurar estrategias para atraer y conservar a los compradores.

Al examinar contexto histórico del comercio electrónico y la importancia del consumidor dentro de este, así como su relevancia para las empresas, se indaga en las diferentes fuentes de información existentes, para recolectar información sobre las compras por internet, en donde una fuente confiable que se tiene es el Instituto Nacional de Estadística y Geografía (INEGI), el cual proporciona información estadística, geográfica y económica mediante encuestas aplicadas a muestras representativas dentro de la república mexicana, en donde una de estas encuestas es la Encuesta Nacional de Distribución y Uso de las Tecnologías de la Información en los Hogares (ENDUTIH) para el ejercicio de 2021, la cual investiga en qué medida y por qué medios la población tiene acceso a las tecnologías de la información además de averiguar para que la utilizan tanto los hogares como los individuos.

Por la estructura de las preguntas de la ENDUTIH que permiten saber en qué utilizan las tecnologías la población y si tienen acceso a internet, además de datos específicos de los usuarios como lo es su edad, escolaridad, nivel socioeconómico, tipo de actividad que desempeña, género, edad, por mencionar algunas. La encuesta cuenta con la información que permite conocer la disponibilidad y uso del internet, en donde destaca si el usuario de internet realizó compras por este medio.

La hipótesis que guía la presente investigación plantea que el consumidor de acuerdo con sus características sociodemográficas, gustos y preferencias; y económicas decide efectuar o no una compra por medio de internet o Ecommerce. La validación de la hipótesis se lleva a cabo partiendo de un análisis teórico, un estudio estadístico y la presentación de modelos econométricos.

La pregunta con la que se obtiene la información sobre si realizó una compra por medio de internet tiene una respuesta tipo binaria, por lo tanto para poder realizar el análisis se recurre a la metodología econométrica de modelos logísticos también llamado modelos logit, los

cuales sirven para explicar variables dicótomas y sus resultados se interpretan en términos de probabilidades por lo que permite responder a la pregunta de ¿Cuál es la probabilidad de que un usuario de internet realice una compra por medio de Ecommerce?

Los modelos logit forman parte de los modelos lineales generalizados, los cuales son caracterizados por que la distribución es binomial y la función de enlace es el logaritmo de probabilidades. En estos modelos no se suponen linealidad y puede existir heterocedasticidad, lo que permite que los resultados se adapten mejor a la realidad bajo el método de máxima verosimilitud, con el cual se elige el valor estimado del parámetro aquel que tenga una mayor probabilidad de suceder con base en lo observado.

Para aplicar el modelo en primer lugar las variables seleccionadas se someten a un análisis de correlación, en donde se identifican aquellas que tienen una relación lineal perfecta por lo que se suprimen del modelo, para así eliminar problemas de multicolinealidad a futuro. Como un primer modelo se toma el que contiene la totalidad de variables del que se obtiene una bondad de ajuste con la prueba de McFadden un Pseudo-R cuadrado 0.298375, el cual funge como punto de comparación para poder iniciar con la selección del mejor modelo.

El primer modelo se somete a la metodología de Step Wise para la selección de variables y tener un modelo con mayor verosimilitud y pocos parámetros, resultando como mejor modelo aquel con un criterio de Akaike más bajo. Al detectarlo se valida que su bondad de ajuste en donde el Pseudo-R cuadrado es de 0.298211. Si bien este número es menor que el primer modelo es importante considerar que es un buen ajuste dado que está dentro de los parámetros, pero con una menor cantidad de variables, es decir, que con el segundo modelo se tiene un buen ajuste con las variables que realmente aportan a la variable explicada, convirtiéndose en un modelo parsimonioso.

Las variables seleccionadas en el mejor modelo son edad, escolaridad, nivel socioeconómico, tipo de actividad que realizan los usuarios, uso de redes sociales y el tipo de red social, tiempo utilizando internet en años, horas promedio de uso de internet al día, tipo de conexión a internet, uso de computadora, laptop y Tablet, acceso a celular y smartphone, tamaño de localidad en la que radica, ser estudiante y el parentesco con el jefe del hogar. Todas resultando estadísticamente significativas al aplicar el modelo.

Considerando las variables que se preservan en el modelo y de acuerdo con lo observado tras la ejecución de este, se prueba la hipótesis que plantea que la decisión de un consumidor en cuanto realizar o no una compra por medio del Ecommerce se ve influenciada por sus características demográfica que en este caso lo fueron la edad, escolaridad, nivel socioeconómico y la relación con el jefe del hogar la cual funge como el rol familiar del individuo; gustos y preferencias en donde entra el tipo de red social que usan, dispositivos que utilizan, el tipo de acceso a internet; y económicas siendo el tipo de actividad que desempeñan la variable significativa.

Los resultados mostrados permiten comprender mejor cómo funciona el mercado electrónico. Esta investigación proporciona información relevante sobre las variables de mayor impacto en el consumidor, de tal forma que los usuarios que tengan ciertas características (edad, escolaridad, género, etc.) serán más propensos que otros a concretar una compra electrónica. La actualidad nos obliga a tener más conocimiento sobre las actividades realizadas en este nicho de mercado, de tal suerte que la experiencia sea cada vez más grata para el usuario.

Bibliografía

1. AMVO (2021). Estudio de venta online 2021 Disponible en: <https://www.amvo.org.mx/estudios/estudio-sobre-venta-online-en-mexico-2021/>
2. AMVO (2023) Estudio de Venta Online 2023. Obtenido de <https://www.amvo.org.mx/> consultado el 01-04-2023
3. Anaya, A. de J. Microeconomía intermedia. Conductas del consumidor y productor en los diferentes mercados. Santa Marta: Siglo del Hombre Editores, 2017. ISBN 9789587460971. Disponible en: <https://search-ebshost-com.pbidi.unam.mx:2443/login.aspx?direct=true&db=nlebk&AN=1838134&site=ehost-live&scope=site>. Acceso em: 12 out. 2022.
4. Basantes A. & et. al (2016). Comercio electrónico. Disponible en: <http://repositorio.utn.edu.ec/handle/123456789/6793> consultado el 3-02-2022
5. BlackSip (2020). Reporte de industria del ecommerce de México. Disponible en: <https://clustertic.org/wp-content/uploads/2020/09/Reporte-de-industria-del-Ecommerce-de-Mexico-2019-2020-1.pdf> Consultado el 8/07/2021
6. Blacksip (2022). Las cifras clave del eCommerce en el mundo en 2021. Disponible en <https://content.blacksip.com/cifras-del-ecommerce-en-el-mundo-en-2021> consultado el 07-03-2023
7. Case K. et al (2012). Principios de Microeconomía, México, editorial Pearson disponible en http://aulavirtual.iberoamericana.edu.co/recursosel/documentos_para_descarga/MICROE_p81nc1p105_m1c803c0n0m14_c4s3_f418_www.economiadigitals.blogspot.pe.pdf consultado el 27-12-2022
8. Cervantes L. et al (2016). Fundamentos de Microeconomía. Perú, editorial Fondo editorial disponible en: <http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/426/fundamentos%20de%20microeconomia.pdf?sequence=1> consultado el 22-12-2022
9. Colell A, Whinston, M.D. y Green, J.R., Microeconomic Theory, Oxford University Press, New York, 1995, cap. 3. disponible en: <http://ebour.com.ar/pdfs/MICROECONOMIA%20II%20->

[%20Nota%20de%20clase-%20consumidores-Parte%20A.pdf](#) consultado el 25 octubre 2022

10. Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD) (2021). <https://unctad.org/es/news/el-comercio-electronico-mundial-alcanza-los-267-billones-de-dolares-mientras-covid-19-impulsa> Consultado el 21-03-2023
11. Conferencia De Las Naciones Unidas Sobre Comercio Y Desarrollo (UNCAD) (2021). Informe Sobre La Economía Digital 2021 Flujos de datos transfronterizos y desarrollo: Para quién fluyen los datos. United Nations Publications. Estados Unidos de América
12. Cuadros J. et al (2012) Elementos conceptuales y aplicaciones de microeconomía para la evaluación de proyectos, Impreso en Naciones Unidas, Santiago de Chile disponible en: https://repositorio.cepal.org/bitstream/handle/11362/5519/1/S1200582_es.pdf consultado el 31-01-2023
13. Cue Mancera, A. y Quintana Romero (2014), L. Introducción a la Microeconomía: un enfoque integral para México. México D.F: Grupo Editorial Patria.
14. Díaz, B. (2020). El ecommerce en tiempos de COVID-19, Conferencia REDLAS Disponible en: https://comunidades.cepal.org/redlas/sites/redlas/files/2020-12/REDLAS_Ecommerce_astarloa_compressed.pdf Consultado el 23-03-2023
15. El economista (2020). El comercio electrónico en México creció 28.6% en 2019, según la Asociación de Internet MX Disponible en: <https://www.eleconomista.com.mx/tecnologia/El-comercio-electronico-en-Mexico-crecio-28.6-en-2019-segun-la-Asociacion-de-Internet-MX-20200311-0106.html> consultado el 28-03-2023
16. El economista (2021). E-commerce y su crecimiento en México. Disponible en: <https://www.eleconomista.com.mx/opinion/E-commerce-y-su-crecimiento-en-Mexico-20210702-0045.html> consultado el 20-04-2023
17. eMarketer (2022). Retail ecommerce sales México 2022-2026 obtenido de <https://forecasts-na1.emarketer.com/5a4fff58d8690c0c28d1f4ca/5a4ffa3bd8690c0c28d1f488> consultado el 31-03-2023

18. EMarketer, 2023. Global Retail Ecommerce Forecast 2023. <https://content-na1.emarketer.com/global-retail-ecommerce-forecast-2023> Consultado el 21-03-2023
19. Encuesta Nacional de Disponibilidad y Uso de las Tecnologías de la Información y Comunicación en los Hogares (ENDUTIH) (2022). Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares 2021. Diseño Muestral. México. INEGI. Obtenido de https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvini/egi/productos/nueva_estruc/889463904632.pdf consultado el 09-04-2023
20. Ferre, M. (2014). Regresión logística. Apuntes del curso FEIR3. Disponible en https://gauss.inf.um.es/feir/45/#225_odds_odds-ratio_y_coeficientes consultado el 7-02-2024
21. García J. (2020) Comercio electrónico en China y México: Surgimiento, evolución y perspectivas. Guadalajara Obtenido de: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-53082020000300053#B4 Consultado el 28-03-2023
22. Garson, D. (2014). Logistic Regression: Binary And Multinomial <https://es.scribd.com/document/379584439/G-David-Garson-Logistic-Regression-Binary-and-Multinomial-Statistical-Associates-Publishing-2014>
23. Gobierno de Navarra. Acercate a las TIC. Uso de dispositivos móviles (“teléfonos móviles”, “smartphone”, “ebooks”, GPS, y “tablets” Disponible en <https://www.navarra.es/NR/rdonlyres/48F9746B-080C-4DEA-BD95-A5B6E01797E1/315641/7Usodedispositivosmoviles.pdf> consultado el 5-Agosto-2024
24. González A. (2015). Selección de variables: Una revisión de métodos existentes http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1263.pdf consultado el 14-01-2024
25. Gregory N. (2012). Principios de economía, Editorial Cengage Learning disponible en: <https://clea.edu.mx/biblioteca/files/original/bd2711c3969d92b67fcf71d844bcbaed.pdf> consultado el 22-12-2022

26. Gujarati, D. (2003). *Econometría*. Cuarta edición. México: McGraw-Hill.
27. Hernández D. & Mendoza G. (2018). *El funcionamiento del Comercio Electrónico, Categorías Seguridad Para Usuarios y Demografía De Usos Habituales Disponible*. (Tesis de licenciatura) en: <http://ri.uaemex.mx/bitstream/handle/20.500.11799/95210/TESIS-COMERCIO-ELECTRONICO.-Definitivo...pdf?sequence=1&isAllowed=y> Consultado el 8/01/2022
28. INEGI (2022). *Cuenta de México. Población*. Disponible en: https://cuentame.inegi.org.mx/poblacion/rur_urb.aspx?tema=P consultado el 10-01-2024
29. INEGI. *Glosario* <https://www.inegi.org.mx/app/glosario/default.html?p=CPV1990> consultado el 25-05-2023
30. Instituto Nacional de Estadística y Grafía INEGI (2021) obtenido de <https://www.inegi.org.mx/inegi/contenido/infoest.html> consultado el 09-04-2023
31. Kassambara A. (2017). *Machine Learning Essentials Practical Guide in R*. Editorial STHDA Capítulo 24.
32. Laudon K. y Guercio C. (2013). *Ecommerce Negocios, tecnología, sociedad*. Editorial Pearson
33. Market Intellingece (AMI), (2022). *DATOS E-COMMERCE MÉXICO [ACTUALIZACIÓN 2022]* <https://americasmi.com/insights/lo-que-mas-compran-los-mexicanos-por-internet/>
34. Marketing4ecommerce (2023). *El número de usuarios de internet en el mundo crece un 1,9% y alcanza los 5.160 millones (2023)*. Disponible en: <https://marketing4ecommerce.net/usuarios-de-internet-mundo/> Consultado el 08-03-2023
35. Martínez, X. (2008). “Microeconomía Avanzada”, en Universidad Autónoma de Barcelona.
36. Medel J. (2020). *E-commerce en México. La estructura de mercado en compranet*. (Tesis de licenciatura) Disponible en: <http://132.248.9.195/ptd2020/agosto/0802724/Index.html> Consultado el 24/11/2021

37. Molinero, L. (2003) <https://www.alceingenieria.net/bioestadistica/maxverosim.pdf>
consultado el 21-08-2023
38. Molineros, P. (2021)
<https://bibdigital.epn.edu.ec/bitstream/15000/21433/1/CD%2010925.pdf>
39. Mora, J. (2013). “Introducción a la teoría del consumidor”, en Universidad ICESI.
https://mpr.ub.uni-muenchen.de/48129/1/MPPA_paper_48129.pdf
40. Nicholson W. (2005). Teoría Microeconómica: Principios Básicos y ampliaciones
<https://elvisjgblog.files.wordpress.com/2019/04/teorc3ada-microeconc3b3mica-9c2b0-edicic3b3n-walter-nicholson.pdf>
41. Nicholson W. (2005). Teoría Microeconómica: Principios Básicos y ampliaciones
<https://elvisjgblog.files.wordpress.com/2019/04/teorc3ada-microeconc3b3mica-9c2b0-edicic3b3n-walter-nicholson.pdf>
42. Nielsen (2022). La evolución del comercio electrónico a nivel mundial. Disponible en:
<https://nielseniq.com/global/es/insights/analysis/2022/the-evolution-of-Ecommerce-globally/#:~:text=Hoy%20en%20d%C3%ADa%2C%20el%20comercio,Medio%20Oriente%20y%20Am%C3%A9rica%20Latina>. Consultado el 14-03-2023
43. Paladino, M. (2017), Modelos Logit en R, Disponible en
https://www.institutomora.edu.mx/testU/SitePages/martinpaladino/modelos_logit_c_on_R.html#la-ordenada-al-origen-de-un-modelo-logit. Consultado el 01-05-2024
44. PARKIN, M. et al.(2010) Microeconomía. Versión para Latinoamérica Novena edición. [s. l.]: Pearson Educación de México, Disponible en:
https://www.ecotec.edu.ec/material/material_2017X1_ECO513_01_84479.pdf
direct=true&db=cat02025a&AN=lib.MX001001976568&lang=es&site=eds-live.
Acesso em: 11 out. 2022
45. Pindyck R. & Rubinfeld D. (2009). Microeconomía. Séptima edición, Madrid. Editorial: Pearson. Disponible en: <https://vicamswitch.mx/wp-content/uploads/2019/06/Pindyck-y-Rubinfeld-2009-MICROECONOM%C3%8DA.pdf> Consultado el: 16-10-2022
46. Quintana, L. y Mendoza, M. (2016). “Econometría Aplicada utilizando R”. Capítulo 13. México: Universidad Nacional Autónoma de México.

http://saree.com.mx/econometriaR/sites/default/files/Ebook_econometriaR.pdf

consultado el 7-12-2022

47. Saussure C., (2016). El uso del smartphone como herramienta para la búsqueda de información en los estudiantes de pregrado de educación de una universidad de Lima Metropolitana. Universidad Católica Sedes Sapientiae. Perú
http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1019-94032016000200002#:~:text=Entre%20las%20principales%20definiciones%20acerca,m%C3%B3vil%20como%20el%20smartphone%20tiene consultado el 17-05-2023
48. Secretaría de Comunicaciones y Transportes (2019). Disponible en: <https://www.gob.mx/sct/es/articulos/conectividad-a-internet-a-todo-el-pais> consultado el 16-03-2024
49. Soloaga I. et al. (2022). Lo rural y lo urbano en México Una nueva caracterización a partir de estadísticas nacionales. CEPAL
https://mexico.un.org/sites/default/files/2022-06/S2200396_es_0.pdf consultado el 31-05-2023
50. Statista (2021). Comercio electrónico en el mundo - Datos estadísticos. Disponible en: <https://es.statista.com/temas/9072/comercio-electronico-en-el-mundo/> consultado el 07-03-2023
51. Statista (2023) Redes sociales con el mayor porcentaje de usuarios en México en 2022
<https://es.statista.com/estadisticas/1035031/mexico-porcentaje-de-usuarios-por-red-social/> consultado el 04-07-2023
52. Saussure, C. (2016). El uso del smartphone como herramienta para la búsqueda de información en los estudiantes de pregrado de educación de una universidad de Lima Metropolitana. disponible en <https://dialnet.unirioja.es/descarga/articulo/5676374.pdf> consultado el 14-12-2023
53. Tordable, A. (2016). La insolvencia en el sector español del transporte: una regresión logística. Universidad de Valladolid Facultad de ciencias económicas y empresariales Disponible en: <https://uvadoc.uva.es/bitstream/handle/10324/21563/TFG-E-235.pdf;jsessionid=57B064DE1E62B2C3B1294562BAA64A5B?sequence=1> consultado el 5-02-2024

54. Universidad de California en Los Ángeles (UCLA). Advanced Research Computing. Statistical Methods and data Analytics, Los Angeles, 2021. <https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/> consultado el 13-01-2024
55. VARIAN, H. R. et al. (2010) Microeconomía intermedia: un enfoque actual. Octava edición. [s. l.]: Antoni Bosch Editor, 2010. ISBN 9781449280628. Disponible en: <https://ebookcentral.proquest.com/lib/bibliodgbsp/reader.action?docID=3202839> Acceso em: 11 out. 2022.
56. Vega J. (2022) E-commerce: desde la perspectiva del cliente de la ciudad de Pasaje. Digital Publisher CEIT, 7(5-3), 66-77 disponible en <https://doi.org/10.33386/593dp.2022.5-3.1381> consultado el 10-08-2024.
57. Velásquez A. (2021) Implementación de las redes sociales para captar clientes en la empresa Verde Constructora e Inmobiliaria S.R.L. Universidad de Piura obtenido de https://pirhua.udep.edu.pe/bitstream/handle/11042/5156/TSP_AE_2113.pdf?sequence=1&isAllowed=y consultado el 04-07-2023
58. Vial B. & Zurita F.(2011) Microeconomía, Primera edición, editorial edicionesuc disponible en: <https://www-digitaliapublishing-com.pbidi.unam.mx:2443/visor/15178> consultado el 18-10-2022
59. Vinuesa, P. (2016). Curso fundamental: Análisis estadístico en ciencias biológicas utilizado R. Universidad Nacional Autónoma de México. Disponible en https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.pdf consultado el 14-01-2024

Anexos

Anexo 1. Lista de variables del modelo con el coeficiente resultante y su conversión en Odd ratios y probabilidades

Variable	Estimate	odd ratios	probabilidad
(Intercept)	-4.47	0.01	1.13%
horas_uso_prom_x_dia_internet	0.06	1.06	51.6%
dispone_celular1	0.26	1.29	56.3%
celular_smatphone1	0.37	1.44	59.1%
uso_compu_lap_tablet1	0.67	1.95	66.2%
asiste_escuela1	-0.23	0.79	44.2%
conexion_por_red_cell	0.44	1.55	60.7%
Tiempo_utilizando_internet_Entre 01 y hasta 2 años`1	-0.23	0.79	44.2%
`Tiempo_utilizando_internet_Más de 5 años`1	0.41	1.50	60.0%
`Tiempo_utilizando_internet_Menos de 1 año`1	-0.43	0.65	39.5%
`tipo_actividad_jornalero(a) o peón`1	-0.31	0.74	42.4%
`tipo_actividad_patrón(a) o empleador(a)`1	0.63	1.89	65.4%
`tipo_actividad_Sin actividad`1	-0.22	0.80	44.6%
`tipo_actividad_trabajador(a) por cuenta propia`1	0.29	1.34	57.3%
tipo_conexion_internet_hogar_Móvil1	-0.34	0.71	41.6%
`tipo_conexion_internet_hogar_Sin conexión en el hogar`1	-0.52	0.59	37.3%
ESTRATO_Alto1	0.45	1.57	61.0%
`ESTRATO_Medio alto`1	0.28	1.33	57.0%
`ESTRATO_Medio bajo`1	0.17	1.18	54.2%
`PAREN_Esposo(a) o compañero(a)`1	0.26	1.30	56.6%
`PAREN_Jefe(a) del hogar`1	0.26	1.29	56.4%
nivel_estudio_Doctorado1	1.24	3.44	77.5%
`nivel_estudio_Lic o Ing`1	0.31	1.37	57.7%
nivel_estudio_Maestría1	0.56	1.75	63.7%
`nivel_estudio_Prescolar o kínder`1	-2.57	0.08	7.1%

nivel_estudio_Primeria1	-0.56	0.57	36.4%
nivel_estudio_secundaria1	-0.17	0.84	45.7%
`grupos_edad_15 a 24 años`1	0.62	1.86	65.0%
`grupos_edad_25 a 34 años`1	0.65	1.92	65.7%
`grupos_edad_35 a 44 años`1	0.41	1.50	60.1%
`grupos_edad_55 años y más`1	-0.42	0.66	39.7%
`TLOC_100,000 y más habitantes`1	0.27	1.32	56.8%
`TLOC_15,000 a 99,999 habitantes`1	0.19	1.21	54.8%
`TLOC_2,500 a 14,999 habitantes`1	0.17	1.19	54.2%
uso_redes_sociales_Sí1	0.39	1.48	59.7%
utiliza_facebook_Sí1	0.49	1.63	61.9%
utiliza_Twitter_Sí1	0.51	1.66	62.4%
utiliza_Instagram_Sí1	0.72	2.06	67.4%
utiliza_Youtube_Sí1	0.12	1.12	52.9%
utiliza_Messenger_Sí1	0.11	1.12	52.7%

Anexo 2. Código en R para implementación del modelo

```
#### Cargar librerías #####
```

```
library(magrittr)
```

```
if("dplyr" %nin% installed.packages()){install.packages("dplyr")}
```

```
library(dplyr)
```

```
if("Hmisc" %nin% installed.packages()){install.packages("Hmisc")}
```

```
library(Hmisc)
```

```
if("networkD3" %nin% installed.packages()){install.packages("networkD3")}
```

```
library(networkD3)
```

```
if("curl" %nin% installed.packages()){install.packages("curl")}
```

```
library(curl)

if("webshot" %nin% installed.packages()){install.packages("webshot")}

library(webshot)

if("tidyverse" %nin% installed.packages()){install.packages("tidyverse")}

library(stringr)

library(foreign)

library(ggplot2)

library(stats)

library(corrplot)

library(base)

library(forecast)

library(graphics)

library(tibble)

library(Hmisc)

library(graphics)

library(readxl)

library(tidyr)

library(ggplot2)

library(stats)

library(corrplot)

library(base)
```

```
library(graphics)
```

```
library(tibble)
```

```
library(fastDummies)
```

```
options("scipen"=100,digits = 6,width = 120)
```

```
##### Importamos datos #####
```

```
#conjunto de datos disponible en:
```

```
https://www.inegi.org.mx/programas/dutih/2021/#microdatos
```

```
Usuarios1<-read.dbf("C:/Colocar ruta de acceso/Base de datos  
2021/tic_2021_usuarios.dbf",as.is = T)
```

```
Usuarios2<-read.dbf("C:/ Colocar ruta de acceso /Base de datos  
2021/tic_2021_usuarios2.dbf",as.is = F)
```

```
hogares<-read.dbf("C:/ Colocar ruta de acceso /Base de datos  
2021/tic_2021_hogares.dbf",as.is = T)
```

```
vivienda<-read.dbf("C:/ Colocar ruta de acceso Base de datos  
2021/tic_2021_viviendas.dbf",as.is = T)
```

```
residentes<-read.dbf("C:/ Colocar ruta de acceso /Base de datos  
2021/tic_2021_residentes.dbf",as.is = T)
```

```
#----- Quitamos las variables que se repiten en las bases de datos
```

```
Usuarios2<-select(Usuarios2,-ESTRATO,-FAC_PER,-SEXO ,-EDAD,-TLOC,-  
DOMINIO,-NIVEL,-ENT)
```

```
hogares<-select(hogares,-ESTRATO,-TLOC,-DOMINIO,-ENT)
```

```
vivienda<-select(vivienda,-ESTRATO,-TLOC,-DOMINIO,-ENT)
```

```
residentes<-select(residentes,-ESTRATO,-SEXO ,-EDAD,-TLOC,-DOMINIO,-NIVEL,-  
ENT)
```

```
#----- Creamos el id del hogar unificado y el id de usuario unificado para las bases de datos
```

```
#--Tab usuarios 1
```

```
Usuarios1$ID_hog_modif<-as.character(str_c((Usuarios1$UPM),  
      (Usuarios1$VIV_SEL),  
      (Usuarios1$HOGAR)))
```

```
Usuarios1$ID_hog_us_modif<-as.character(str_c((Usuarios1$UPM),  
      (Usuarios1$VIV_SEL),  
      (Usuarios1$HOGAR),  
      (Usuarios1$NUM_REN)))
```

```
unicos<-as.data.frame(unique(Usuarios1$ID_hog_modif))
```

```
rm(unicos)
```

```
#--Tab usuarios 2
```

```
Usuarios2$ID_hog_us_modif<-as.character(str_c((Usuarios2$UPM),  
      (Usuarios2$VIV_SEL),  
      (Usuarios2$HOGAR),  
      (Usuarios2$NUM_REN)))
```

```
unicos<-as.data.frame(unique(Usuarios2$ID_hog_us_modif))
```

```
rm(unicos)
```

```
#write.csv(Usuarios2,"usuarios 2.csv",row.names = F)
```

```
#--Tab residentes
```

```
residentes$ID_hog_us_modif<-as.character(str_c((residentes$UPM),  
      (residentes$VIV_SEL),  
      (residentes$HOGAR),  
      (residentes$NUM_REN)))
```

```

unicos<-as.data.frame(unique(residentes$ID_hog_us_modif))

rm(unicos)

#--Tab hogares

hogares$ID_hog_modif<-str_c((hogares$UPM),
                             (hogares$VIV_SEL),
                             (hogares$HOGAR))

unicos<-as.data.frame(unique(hogares$ID_hog_modif))

rm(unicos)

#----- unimos las bases de datos de usuario1, usuario2, residentes, hogar

Consolidado<-merge( Usuarios1,Usuarios2,by=c("ID_hog_us_modif"),all.x = T)

Consolidado<-merge( Consolidado,residentes,by=c("ID_hog_us_modif"),all.x = T)

Consolidado<-merge( Consolidado,hogares,by=c("ID_hog_modif"),all.x = T)

##### Creamos la tabla con las variables seleccionadas #####

#----- Hacemos la selección de variables

data<-select(Consolidado,
              ID_hog_modif, #id del usuario
              EDAD,
              P7_1, # Ha usado internet dentro o fuera del hogar si/no
              P7_19, # Ha realizado compras por internet si/no
              P7_4, # Horas promedio en que utiliza internet al día
              P7_6, # Tiempo que lleva utilizando internet
              P8_1, # Dispone de celular si/no
              P8_4_2, # El celular es smartphone si/no
              P6_1, # Hizo uso de computadora, Tablet o laptop
              P6_5, # Cuantas horas al día usa computadora, Tablet o laptop

```

P3_7, # Asiste actualmente a la escuela si/no

P3_10, # La semana pasada... trabajó, estudió, busco trabajo etc.

P3_12, # En su actividad fue. Empleado, obrero etc

P4_4, # Dispone de conexión a internet en su hogar sí/no

P4_5, # Tipo de conexión que tiene el hogar fija/movil/ambas

P8_11_1, # Se conecta a internet mediante Wifi

P8_11_2, # Se conecta a internet mediante red celular (conexión a datos)

P7_13, # Ha utilizado redes sociales

P7_14_1, # Utiliza facebook

P7_14_2, # Utiliza Twitter

P7_14_3, # Utiliza Instagram

P7_14_4, # Utiliza LinkedIn

P7_14_5, # Utiliza Snapchat

P7_14_6, # Utiliza Whatsapp

P7_14_7, # Utiliza Youtube

P7_14_8, # Utiliza Piterest

P7_14_9, # Utiliza Messenger

P7_14_10, # Utiliza Tumblr

P7_14_11, # Utiliza Otras redes

FAC_PER, # factor de expansión

DOMINIO, # Es urbano o rural

TLOC, # Tamaño localidad

ESTRATO, # Estrato socioeconómico

PAREN, # Parentesco con la jef@ de familia

SEXO, # Masculino o femenino

```

    NIVEL, # Hasta que año aprobó en la escuela
    ENT    # Entidad federativa del usuario)

#----- Liberamos espacio en la consola

rm(Consolidado,hogares,residentes,Usuarios1,Usuarios2,vivienda)

##### Ajuste del dataset #####

#----- Cambiamos el nombre de las variables

data=rename(data,c(uso_internet="P7_1"))

data=rename(data,c(compra_ecommerce="P7_19"))

data=rename(data,c(horas_uso_prom_x_dia_internet="P7_4"))

data=rename(data,c(Tiempo_utilizando_internet="P7_6"))

data=rename(data,c(dispone_celular="P8_1"))

data=rename(data,c(celular_smartphone="P8_4_2"))

data=rename(data,c(uso_compu_lap_tablet="P6_1"))

data=rename(data,c(horas_uso_compu_lap_tablet="P6_5"))

data=rename(data,c(asiste_escuela="P3_7"))

data=rename(data,c(economicamente_activo="P3_10"))

data=rename(data,c(tipo_actividad="P3_12"))

data=rename(data,c(conexion_internet_hogar="P4_4"))

data=rename(data,c(tipo_conexion_internet_hogar="P4_5"))

data=rename(data,c(nivel_estudio="NIVEL"))

data=rename(data,c(GENERO="SEXO"))

data=rename(data,c( conexion_por_Wifi ="P8_11_1"))

data=rename(data,c( conexion_por_red_cel="P8_11_2"))

data=rename(data,c( uso_redes_sociales="P7_13"))

data=rename(data,c( utiliza_facebook="P7_14_1"))

```



```

data=rename(data,c( utiliza_Twitter="P7_14_2"))
data=rename(data,c( utiliza_Instagram="P7_14_3"))
data=rename(data,c( utiliza_LinkedIn="P7_14_4"))
data=rename(data,c( utiliza_Snapchat="P7_14_5"))
data=rename(data,c( utiliza_Whatsapp="P7_14_6"))
data=rename(data,c( utiliza_Youtube="P7_14_7"))
data=rename(data,c( utiliza_Piterest="P7_14_8"))
data=rename(data,c( utiliza_Messenger="P7_14_9"))
data=rename(data,c( utiliza_Tumblr="P7_14_10"))
data=rename(data,c( utiliza_Otras_redes="P7_14_11"))
data$celular_smatphone<-as.character(data$celular_smatphone)
data$dispone_celular<-as.character(data$dispone_celular)
data$conexion_por_Wifi<-as.character(data$conexion_por_Wifi)
data$conexion_por_red_cel<-as.character(data$conexion_por_red_cel)
#----- Revisamos cuales son los avlores unicos de cada variable para ajustarlos
unique(data$uso_internet)
unique(data$dispone_celular)
unique(data$celular_smatphone)
unique(data$uso_compu_lap_tablet)
unique(data$economicamente_activo)
unique(data$asiste_escuela)
unique(data$tipo_actividad)
unique(data$conexion_internet_hogar)
unique(data$tipo_conexion_internet_hogar)
unique(data$nivel_estudio)

```

```

unique(data$Tiempo_utilizando_internet)

unique(data$utiliza_facebook)

#----- Recodificar las columnas para que sean 0 (No) y 1 (sí)

data$uso_internet[data$uso_internet == "2"] <- "0" #el 2 significa No

data$compra_ecommerce[data$compra_ecommerce == "2"] <- "0" #el 2 significa No

data$compra_ecommerce[is.na(data$compra_ecommerce)] <- "0" #el 2 significa No

data$dispone_celular[data$dispone_celular == "2"] <- "0" #el 2 significa No

data$celular_smatphone[as.character(data$celular_smatphone) == "2"] <- "0" #el 2
significa No

data$celular_smatphone[is.na(data$celular_smatphone)] <- "0"

data$conexion_por_red_cel[is.na(data$conexion_por_red_cel)]->"sin respuesta"

data$conexion_por_red_cel[data$conexion_por_red_cel == "2"] <- "0" #el 2 significa No

data$conexion_por_Wifi[data$conexion_por_Wifi == "2"] <- "0" #el 2 significa No

data$conexion_por_Wifi[is.na(data$conexion_por_Wifi)] <- "sin respuesta"

data$uso_redes_sociales[data$uso_redes_sociales == "2"] <- "0" #el 2 significa No

data$uso_redes_sociales[is.na(data$uso_redes_sociales)] <- "sin respuesta"

data$utiliza_facebook[data$utiliza_facebook == "2"] <- "0" #el 2 significa No

data$utiliza_facebook[is.na(data$utiliza_facebook)] <- "sin respuesta"

```

```
data$utiliza_Instagram[data$utiliza_Instagram == "2"] <- "0" #el 2 significa No  
data$utiliza_Instagram[is.na(data$utiliza_Instagram)] <- "sin respuesta"
```

```
data$utiliza_Twitter[data$utiliza_Twitter == "2"] <- "0" #el 2 significa No  
data$utiliza_Twitter[is.na(data$utiliza_Twitter)] <- "sin respuesta"
```

```
data$utiliza_Whatsapp[data$utiliza_Whatsapp == "2"] <- "0" #el 2 significa No  
data$utiliza_Whatsapp[is.na(data$utiliza_Whatsapp)] <- "sin respuesta"
```

```
data$utiliza_LinkedIn[data$utiliza_LinkedIn == "2"] <- "0" #el 2 significa No  
data$utiliza_LinkedIn[is.na(data$utiliza_LinkedIn)] <- "sin respuesta"
```

```
data$utiliza_Piterest[data$utiliza_Piterest == "2"] <- "0" #el 2 significa No  
data$utiliza_Piterest[is.na(data$utiliza_Piterest)] <- "sin respuesta"
```

```
data$utiliza_Youtube[data$utiliza_Youtube == "2"] <- "0" #el 2 significa No  
data$utiliza_Youtube[is.na(data$utiliza_Youtube)] <- "sin respuesta"
```

```
data$utiliza_Messenger[data$utiliza_Messenger== "2"] <- "0" #el 2 significa No  
data$utiliza_Messenger[is.na(data$utiliza_Messenger)] <- "sin respuesta"
```

```
data$utiliza_Snapchat[data$utiliza_Snapchat== "2"] <- "0" #el 2 significa No  
data$utiliza_Snapchat[is.na(data$utiliza_Snapchat)] <- "sin respuesta"
```

```
data$utiliza_Tumblr[data$utiliza_Tumblr == "2"] <- "0" #el 2 significa No
```

```
data$utiliza_Tumblr[is.na(data$utiliza_Tumblr)] <- "sin respuesta"
```

```
data$utiliza_Otras_redes[data$utiliza_Otras_redes == "2"] <- "0" #el 2 significa No
```

```
data$utiliza_Otras_redes[is.na(data$utiliza_Otras_redes)] <- "sin respuesta"
```

```
data$uso_compu_lap_tablet[data$uso_compu_lap_tablet == "2"] <- "0" #el 2 significa No
```

```
data$asiste_escuela[data$asiste_escuela == "2"] <- "0" #el 2 significa No
```

```
data$economicamente_activo[data$economicamente_activo == "1"] <- 1 #el 1 trabajó
```

```
data$economicamente_activo[data$economicamente_activo == "2"] <- 1 #el 2 tenía trabajo  
pero no trabajó
```

```
data$economicamente_activo[data$economicamente_activo == "3"] <- 1 # buscó trabajo
```

```
data$economicamente_activo[data$economicamente_activo == "4"] <- 0 # jubilado o  
pensionado
```

```
data$economicamente_activo[data$economicamente_activo == "5"] <- 0 # estudia
```

```
data$economicamente_activo[data$economicamente_activo == "6"] <- 0 # se dedico a lo  
quehaceres del hogar
```

```
data$economicamente_activo[data$economicamente_activo == "7"] <- 0 # es persona con  
limitaciones fisica o metal
```

```
data$economicamente_activo[data$economicamente_activo == "8"] <- 0 # no trabajó
```

```
data$economicamente_activo[is.na(data$economicamente_activo)] <- 0 # lo contamos  
como no trabajó
```

```
data$GENERO[data$GENERO == "2"] <- "0" #el 2 es mujer
```

```

data$nivel_estudio[data$nivel_estudio == "00"] <- "Ninguno"
data$nivel_estudio[data$nivel_estudio == "01"] <- "Preescola o kínder"
data$nivel_estudio[data$nivel_estudio == "02"] <- "Primaria"
data$nivel_estudio[data$nivel_estudio == "03"] <- "secundaria"
data$nivel_estudio[data$nivel_estudio == "04"] <- "Normal básica"
data$nivel_estudio[data$nivel_estudio == "05"] <- "Estudio técnico terminal con
secundaria"
data$nivel_estudio[data$nivel_estudio == "06"] <- "Preparatorio o bachillerato"
data$nivel_estudio[data$nivel_estudio == "07"] <- "Técnico superior con prepa terminada"
data$nivel_estudio[data$nivel_estudio == "08"] <- "Lic o Ing"
data$nivel_estudio[data$nivel_estudio == "09"] <- "Especialidad"
data$nivel_estudio[data$nivel_estudio == "10"] <- "Maestría"
data$nivel_estudio[data$nivel_estudio == "11"] <- "Doctorado"
data$nivel_estudio[data$nivel_estudio == "99"] <- "No_sabe"

data$Tiempo_utilizando_internet[data$Tiempo_utilizando_internet == "1"] <- "Menos de 1
año"
data$Tiempo_utilizando_internet[data$Tiempo_utilizando_internet == "2"] <- "Entre 01 y
hasta 2 años"
data$Tiempo_utilizando_internet[data$Tiempo_utilizando_internet == "3"] <- "Más de 2 y
hasta 5 años"
data$Tiempo_utilizando_internet[data$Tiempo_utilizando_internet == "4"] <- "Más de 5
años"
data$Tiempo_utilizando_internet[data$Tiempo_utilizando_internet == "5"] <- "No
recuerda"
data$Tiempo_utilizando_internet[is.na(data$Tiempo_utilizando_internet)] <- "No usa
internet"

```

```
data$PAREN[data$PAREN == "1"] <- "Jefe(a) del hogar"  
data$PAREN[data$PAREN == "2"] <- "Esposo(a) o compañero(a)"  
data$PAREN[data$PAREN == "3"] <- "Hijo(a)"  
data$PAREN[data$PAREN == "4"] <- "Nieto(a)"  
data$PAREN[data$PAREN == "5"] <- "Otro parentezco"  
data$PAREN[data$PAREN == "6"] <- "Ningun parentesco"
```

```
data$ESTRATO[data$ESTRATO == "1"] <- "Bajo"  
data$ESTRATO[data$ESTRATO == "2"] <- "Medio bajo"  
data$ESTRATO[data$ESTRATO == "3"] <- "Medio alto"  
data$ESTRATO[data$ESTRATO == "4"] <- "Alto"
```

```
data$TLOC[data$TLOC == "1"] <- "100,000 y más habitantes"  
data$TLOC[data$TLOC == "2"] <- "15,000 a 99,999 habitantes"  
data$TLOC[data$TLOC == "3"] <- "2,500 a 14,999 habitantes"  
data$TLOC[data$TLOC == "4"] <- "menor a 2500 habitantes"
```

```
data$DOMINIO[data$DOMINIO == "U"] <- "1"  
data$DOMINIO[data$DOMINIO == "R"] <- "0"
```

```
data$conexion_internet_hogar[data$conexion_internet_hogar == "2"] <- "0"
```

```
data$tipo_conexion_internet_hogar[data$tipo_conexion_internet_hogar == "1"] <- "Fija"  
data$tipo_conexion_internet_hogar[data$tipo_conexion_internet_hogar == "2"] <- "Móvil"
```

```

data$tipo_conexion_internet_hogar[data$tipo_conexion_internet_hogar == "3"] <-
"Ambas"

data$tipo_conexion_internet_hogar[data$tipo_conexion_internet_hogar == "9"] <- "No
sabe"

data$tipo_conexion_internet_hogar[is.na(data$tipo_conexion_internet_hogar)] <- "Sin
conexión en el hogar"

data$tipo_actividad[data$tipo_actividad == "1"] <- "empleado(a)"
data$tipo_actividad[data$tipo_actividad == "2"] <- "obrero(a)"
data$tipo_actividad[data$tipo_actividad == "3"] <- "jornalero(a) o peón"
data$tipo_actividad[data$tipo_actividad == "4"] <- "trabajador(a) por cuenta propia "
data$tipo_actividad[data$tipo_actividad == "5"] <- "patrón(a) o empleador(a)"
data$tipo_actividad[data$tipo_actividad == "6"] <- "trabajador(a) sin pago en un negocio
familiar o no familiar"
data$tipo_actividad[is.na(data$tipo_actividad)] <- "Sin actividad"

data$grupos_edad <- ifelse(data$EDAD <= 14, "06 a 15 años",
  ifelse(data$EDAD >= 15 & data$EDAD <= 24, "15 a 24 años",
    ifelse(data$EDAD >= 25 & data$EDAD <= 34, "25 a 34 años",
      ifelse(data$EDAD >= 35 & data$EDAD <= 44, "35 a 44 años",
        ifelse(data$EDAD >= 45 & data$EDAD <= 54, "45 a 54 años",
          ifelse(data$EDAD >= 55, "55 años y más", ""))))))

#----- Exportamos la data en un archivo csv
#write.csv(data, "data ENDUTIH 2021 variables seleccionadas.csv", row.names = F)
#write.csv(Consolidado, "consolidado ENDUTIH 2021.csv", row.names = F)
#----- Filtramos el data set para quedarnos con la población de interes que son
#----- las personas que declararon tener acceso a internet

```

```

data_usuarios<-filter(data,uso_internet=="1")

sum(data_usuarios$FAC_PER)

#----- Convertimos las variables en factores

data_usuarios<- mutate_if(data_usuarios, is.character, as.factor)

#----- Modificamos las variables en numero

data_usuarios$horas_uso_compu_lap_tablet<-
as.integer(data_usuarios$horas_uso_compu_lap_tablet)

data_usuarios$horas_uso_prom_x_dia_internet<-
as.integer(data_usuarios$horas_uso_prom_x_dia_internet)

#----- Quitamos NA de la tabla

data_usuarios[is.na(data_usuarios)] <- 0

#----- Total de Usuarios de internet

sum(data_usuarios$FAC_PER)

#----- Modificamos las avariables multicategoricas en dummies

colnames(data_usuarios)

data_usuarios<-dummy_cols(data_usuarios,select_columns =
c('Tiempo_utilizando_internet','tipo_actividad','tipo_conexion_internet_hogar','ESTRATO',
'PAREN','nivel_estudio','grupos_edad','TLOC','uso_redes_sociales',"utiliza_facebook","util
iza_Twitter","utiliza_Instagram","utiliza_LinkedIn","utiliza_Snapchat"
,"utiliza_Whatsapp","utiliza_Youtube","utiliza_Piterest","utiliza_Messenger","utiliza_Tum
blr" ,"utiliza_Otras_redes"),remove_selected_columns = FALSE)

#-----modificamos nombres de variables del modelo

colnames(data_usuarios)

new_names<-c("ID_hog_modif", "EDAD", "uso_internet", "compra_ecommerce",
"horas_uso_prom_x_dia_internet", "Tiempo_utilizando_internet", "dispone_celular",
"celular_smatphone", "uso_compu_lap_tablet", "horas_uso_compu_lap_tablet",
"asiste_escuela", "economicamente_activo", "tipo_actividad", "conexion_internet_hogar",
"tipo_conexion_internet_hogar", "conexion_por_Wifi", "conexion_por_red_cel",
"uso_redes_sociales", "utiliza_facebook", "utiliza_Twitter", "utiliza_Instagram",

```


"utiliza_LinkedIn", "utiliza_Snapchat", "utiliza_Whatsapp", "utiliza_Youtube",
 "utiliza_Piterest", "utiliza_Messenger", "utiliza_Tumblr", "utiliza_Otras_redes",
 "FAC_PER", "DOMINIO", "TLOC", "ESTRATO", "PAREN", "GENERO",
 "nivel_estudio", "ENT", "grupos_edad", "Tiempo_utilizando_internet_Entre 01 y hasta 2
 años", "Tiempo_utilizando_internet_Más de 2 y hasta 5 años",
 "Tiempo_utilizando_internet_Más de 5 años", "Tiempo_utilizando_internet_Menos de 1
 año", "Tiempo_utilizando_internet_No recuerda", "tipo_actividad_empleado(a)",
 "tipo_actividad_jornalero(a) o peón", "tipo_actividad_obrero(a)", "tipo_actividad_patrón(a)
 o empleador(a)", "tipo_actividad_Sin actividad", "tipo_actividad_trabajador(a) por cuenta
 propia", "tipo_actividad_trabajador(a) sin pago en un negocio familiar o no familiar",
 "tipo_conexion_internet_hogar_Ambas", "tipo_conexion_internet_hogar_Fija",
 "tipo_conexion_internet_hogar_Móvil", "tipo_conexion_internet_hogar_No sabe",
 "tipo_conexion_internet_hogar_Sin conexión en el hogar", "ESTRATO_Alto",
 "ESTRATO_Bajo", "ESTRATO_Medio alto", "ESTRATO_Medio bajo",
 "PAREN_Esposo(a) o compañero(a)", "PAREN_Hijo(a)", "PAREN_Jefe(a) del hogar",
 "PAREN_Nieto(a)", "PAREN_Ningun parentesco", "PAREN_Otro parentezco",
 "nivel_estudio_Doctorado", "nivel_estudio_Especialidad", "nivel_estudio_Estudio técnico
 terminal con secundaria", "nivel_estudio_Lic o Ing", "nivel_estudio_Maestría",
 "nivel_estudio_Ninguno", "nivel_estudio_No sabe", "nivel_estudio_Normal básica",
 "nivel_estudio_Preescola o kínder", "nivel_estudio_Preparatorio o bachillerato",
 "nivel_estudio Primaria", "nivel_estudio_secundaria", "nivel_estudio_Técnico superior
 con prepa terminada", "grupos_edad_06 a 15 años", "grupos_edad_15 a 24 años",
 "grupos_edad_25 a 34 años", "grupos_edad_35 a 44 años", "grupos_edad_45 a 54 años",
 "grupos_edad_55 años y más", "TLOC_100,000 y más habitantes", "TLOC_15,000 a
 99,999 habitantes", "TLOC_2,500 a 14,999 habitantes", "TLOC_menor a 2500 habitantes",
 "uso_redes_sociales_No", "uso_redes_sociales_Sí", "utiliza_facebook_No",
 "utiliza_facebook_Sí", "utiliza_facebook_sin respuesta", "utiliza_Twitter_No",
 "utiliza_Twitter_Sí", "utiliza_Twitter_sin respuesta", "utiliza_Instagram_No",
 "utiliza_Instagram_Sí", "utiliza_Instagram_sin respuesta", "utiliza_LinkedIn_No",
 "utiliza_LinkedIn_Sí", "utiliza_LinkedIn_sin respuesta", "utiliza_Snapchat_No",
 "utiliza_Snapchat_Sí", "utiliza_Snapchat_sin respuesta", "utiliza_Whatsapp_No",
 "utiliza_Whatsapp_Sí", "utiliza_Whatsapp_sin respuesta", "utiliza_Youtube_No",
 "utiliza_Youtube_Sí", "utiliza_Youtube_sin respuesta", "utiliza_Piterest_No",
 "utiliza_Piterest_Sí", "utiliza_Piterest_sin respuesta", "utiliza_Messenger_No",
 "utiliza_Messenger_Sí", "utiliza_Messenger_sin respuesta", "utiliza_Tumblr_No",
 "utiliza_Tumblr_Sí", "utiliza_Tumblr_sin respuesta", "utiliza_Otras_redes_No",
 "utiliza_Otras_redes_Sí", "utiliza_Otras_redes_sin respuesta")

colnames(data_usuarios)[<-new_names

```

#-----seleccionamos las variables para el modelo
colnames(data_usuarios)
d<-select(data_usuarios,
           horas_uso_prom_x_dia_internet
           ,dispone_celular
           ,celular_smatphone
           ,uso_compu_lap_tablet
           ,horas_uso_compu_lap_tablet
           ,asiste_escuela
           ,economicamente_activo
           ,conexion_por_red_cel
           ,DOMINIO
           ,`Tiempo_utilizando_internet_Entre 01 y hasta 2 años`
           ,`Tiempo_utilizando_internet_Más de 2 y hasta 5 años`
           ,`Tiempo_utilizando_internet_Más de 5 años`
           ,`Tiempo_utilizando_internet_Menos de 1 año`
           ,`Tiempo_utilizando_internet_No recuerda`
           ,`tipo_actividad_empleado(a)`
           ,`tipo_actividad_jornalero(a) o peón`
           ,`tipo_actividad_obrero(a)`
           ,`tipo_actividad_patrón(a) o empleador(a)`
           ,`tipo_actividad_Sin actividad`
           ,`tipo_actividad_trabajador(a) por cuenta propia `
           ,`tipo_actividad_trabajador(a) sin pago en un negocio familiar o no familiar`
           ,tipo_conexion_internet_hogar_Ambas

```

,tipo_conexion_internet_hogar_Fija
,tipo_conexion_internet_hogar_Móvil
,`tipo_conexion_internet_hogar_No sabe`
,`tipo_conexion_internet_hogar_Sin conexión en el hogar`
,ESTRATO_Alto
,ESTRATO_Bajo
,`ESTRATO_Medio alto`
,`ESTRATO_Medio bajo`
,`PAREN_Esposo(a) o compañero(a)`
,`PAREN_Hijo(a)`
,`PAREN_Jefe(a) del hogar`
,`PAREN_Nieto(a)`
,`PAREN_Ningun parentesco`
,`PAREN_Otro parentezco`
,nivel_estudio_Doctorado
,nivel_estudio_Especialidad
,`nivel_estudio_Estudio técnico terminal con secundaria`
,`nivel_estudio_Lic o Ing`
,nivel_estudio_Maestría
,nivel_estudio_Ninguno
,nivel_estudio_No_sabe
,`nivel_estudio_Normal básica`
,`nivel_estudio_Preescola o kínder`
,`nivel_estudio_Preparatorio o bachillerato`
,nivel_estudio Primaria

,nivel_estudio_secundaria
,`nivel_estudio_Técnico superior con prepa terminada`
,`grupos_edad_06 a 15 años`
,`grupos_edad_15 a 24 años`
,`grupos_edad_25 a 34 años`
,`grupos_edad_35 a 44 años`
,`grupos_edad_45 a 54 años`
,`grupos_edad_55 años y más`
,`TLOC_100,000 y más habitantes`
,`TLOC_15,000 a 99,999 habitantes`
,`TLOC_2,500 a 14,999 habitantes`
,`TLOC_menor a 2500 habitantes`
#,conexion_por_Wifi_No
#,conexion_por_Wifi_1
#`,`conexion_por_Wifi_sin respuesta`
,uso_redes_sociales_No
,uso_redes_sociales_Sí
,utiliza_facebook_No
,utiliza_facebook_Sí
,`utiliza_facebook_sin respuesta`
,utiliza_Twitter_No
,utiliza_Twitter_Sí
,`utiliza_Twitter_sin respuesta`
,utiliza_Instagram_No
,utiliza_Instagram_Sí

```

    ,`utiliza_Instagram_sin respuesta`
    ,utiliza_Whatsapp_No
    ,utiliza_Whatsapp_Sí
    ,`utiliza_Whatsapp_sin respuesta`
    ,utiliza_Youtube_No
    ,utiliza_Youtube_Sí
    ,`utiliza_Youtube_sin respuesta`
    ,utiliza_Messenger_No
    ,utiliza_Messenger_Sí
    ,`utiliza_Messenger_sin respuesta`)
d<- mutate_if(d, is.factor, as.numeric)
#-----creamos la matriz de correlación entre las variables seleccionadas
correlacion<-round(cor(d),1)
#write.csv(correlacion,"matriz de correlación 2.csv")
#corrplot(correlacion,method = "number")
sum(data_usuarios$FAC_PER)
#----- Convertimos las variables en factores
data_usuarios<- mutate_if(data_usuarios, is.integer, as.factor)
#----- Modificamos las variables que salieron de convertir en dummies a factores
data_usuarios$horas_uso_compu_lap_tablet<-
as.integer(data_usuarios$horas_uso_compu_lap_tablet)
data_usuarios$horas_uso_prom_x_dia_internet<-
as.integer(data_usuarios$horas_uso_prom_x_dia_internet)
data_usuarios$EDAD<-as.integer(data_usuarios$EDAD)
#----- Eliminamos el factor de expansion dado que no se requiere ahora
data_usuarios<-select(data_usuarios,-FAC_PER)

```

```

#####
##### INICIAMOS CON MODELO LOGIT #####
#####
#-----revisamos el total de variables que tiene data_usuarios

colnames(data_usuarios)

#-----creamos un primer modelo en donde se incluyan todas las variables que pueden
explicar al Churn

model1<-glm( compra_ecommerce~

    horas_uso_prom_x_dia_internet

+dispone_celular

+celular_smatphone

+uso_compu_lap_tablet

+horas_uso_compu_lap_tablet

+asiste_escuela

+economicamente_activo

+conexion_por_red_cel

+DOMINIO

+`Tiempo_utilizando_internet_Entre 01 y hasta 2 años`

+`Tiempo_utilizando_internet_Más de 2 y hasta 5 años`

+`Tiempo_utilizando_internet_Más de 5 años`

+`Tiempo_utilizando_internet_Menos de 1 año`

#+`Tiempo_utilizando_internet_No recuerda` #----- dummy penalizada

+`tipo_actividad_empleado(a)`

+`tipo_actividad_jornalero(a) o peón`

+`tipo_actividad_obrero(a)`

```

+`tipo_actividad_patrón(a) o empleador(a)`
 +`tipo_actividad_Sin actividad`
 +`tipo_actividad_trabajador(a) por cuenta propia`
 #+`tipo_actividad_trabajador(a) sin pago en un negocio familiar o no familiar`
 +tipo_conexion_internet_hogar_Ambas
 +tipo_conexion_internet_hogar_Fija
 +tipo_conexion_internet_hogar_Móvil
 #+`tipo_conexion_internet_hogar_No sabe` #----- dummy penalizada
 +`tipo_conexion_internet_hogar_Sin conexión en el hogar`
 +ESTRATO_Alto
 #+ESTRATO_Bajo #----- dummy penalizada
 +`ESTRATO_Medio alto`
 +`ESTRATO_Medio bajo`
 +`PAREN_Esposo(a) o compañero(a)`
 +`PAREN_Hijo(a)`
 +`PAREN_Jefe(a) del hogar`
 +`PAREN_Nieto(a)`
 #+`PAREN_Ningun parentesco` #----- dummy penalizada
 #+`PAREN_Otro parentezco` #----- dummy penalizada
 +nivel_estudio_Doctorado
 #+nivel_estudio_Especialidad #----- dummy penalizada
 #+`nivel_estudio_Estudio técnico terminal con secundaria` #---- dummy penalizada
 +`nivel_estudio_Lic o Ing`
 +nivel_estudio_Maestría
 #+nivel_estudio_Ninguno #----- dummy penalizada

#+nivel_estudio_No_sabe #----- dummy penalizada
#+`nivel_estudio_Normal básica` #----- dummy penalizada
+`nivel_estudio_Preescola o kínder`
+`nivel_estudio_Preparatorio o bachillerato`
+nivel_estudio_Primeria
+nivel_estudio_secundaria
#+`nivel_estudio_Técnico superior con prepa terminada` #---- dummy penalizada
#+`grupos_edad_06 a 15 años` #----- dummy penalizada
+`grupos_edad_15 a 24 años`
+`grupos_edad_25 a 34 años`
+`grupos_edad_35 a 44 años`
+`grupos_edad_45 a 54 años`
+`grupos_edad_55 años y más`
+`TLOC_100,000 y más habitantes`
+`TLOC_15,000 a 99,999 habitantes`
+`TLOC_2,500 a 14,999 habitantes`
#+`TLOC_menor a 2500 habitantes` #----- dummy penalizada
#+conexion_por_Wifi_No
#+conexion_por_Wifi_Sí
#+`conexion_por_Wifi_sin respuesta` #----- dummy penalizada
#+uso_redes_sociales_No #----- dummy penalizada
+uso_redes_sociales_Sí
#+utiliza_facebook_No
+utiliza_facebook_Sí
#+`utiliza_facebook_sin respuesta` #----- dummy penalizada


```

#+utiliza_Twitter_No
+utiliza_Twitter_Sí
#+`utiliza_Twitter_sin respuesta` #----- dummy penalizada
#+utiliza_Instagram_No
+utiliza_Instagram_Sí
#+`utiliza_Instagram_sin respuesta`#----- dummy penalizada
#+utiliza_Whatsapp_No
+utiliza_Whatsapp_Sí
#+`utiliza_Whatsapp_sin respuesta` #----- dummy penalizada
#+utiliza_Youtube_No
+utiliza_Youtube_Sí
#+`utiliza_Youtube_sin respuesta` #----- dummy penalizada
#+utiliza_Messenger_No
+utiliza_Messenger_Sí,
#+`utiliza_Messenger_sin respuesta`#----- dummy penalizada
family = binomial(link="logit"),data=data_usuarios)

```

```
summary(model1)
```

```
#-----Calculamos el pseudo R cuadrado de McFadden (“residual deviance” / “null deviance”):
```

```
#-----los resultados entre 0,2 y 0,4 indican un excelente ajuste del modelo.
```

```
prueba_bondad<- 1 -(model1$deviance / model1$null.deviance)
```

```
prueba_bondad
```

```
DescTools::PseudoR2(model1, which = "McFadden")
```

#-----Realizamos un segundo modelo con la selección de variables con el modelo anterior utilizando la metodología StepWise

```
seleccion<-step(model1)
```

```
modelo2<-glm(compra_ecommerce ~ horas_uso_prom_x_dia_internet + dispone_celular +  
celular_smatphone + uso_compu_lap_tablet + asiste_escuela + conexion_por_red_cel +  
`Tiempo_utilizando_internet_Entre 01 y hasta 2 años` + `Tiempo_utilizando_internet_Más  
de 5 años` + `Tiempo_utilizando_internet_Menos de 1 año` + `tipo_actividad_jornalero(a)  
o peón` + `tipo_actividad_patrón(a) o empleador(a)` + `tipo_actividad_Sin actividad` +  
`tipo_actividad_trabajador(a) por cuenta propia` + tipo_conexion_internet_hogar_Móvil +  
`tipo_conexion_internet_hogar_Sin conexión en el hogar` + ESTRATO_Alto +  
`ESTRATO_Medio alto` + `ESTRATO_Medio bajo` + `PAREN_Espos(a) o  
compañero(a)` + `PAREN_Jefe(a) del hogar` + nivel_estudio_Doctorado +  
`nivel_estudio_Lic o Ing` + nivel_estudio_Maestría + `nivel_estudio_Preescola o kínder` +  
nivel_estudio_Primeria + nivel_estudio_secundaria + `grupos_edad_15 a 24 años` +  
`grupos_edad_25 a 34 años` + `grupos_edad_35 a 44 años` + `grupos_edad_55 años y más`  
+ `TLOC_100,000 y más habitantes` + `TLOC_15,000 a 99,999 habitantes` +  
`TLOC_2,500 a 14,999 habitantes` + #conexion_por_Wifi_No + se elimina por tener la  
variable sin conexion en el hogar
```

```
uso_redes_sociales_Sí + utiliza_facebook_Sí + utiliza_Twitter_Sí + utiliza_Instagram_Sí +  
utiliza_Youtube_Sí + utiliza_Messenger_Sí, family =  
binomial(link="logit"), data=data_usuarios)
```

```
prueba_bondad_model2<- 1 -(modelo2$deviance / modelo2$null.deviance)
```

```
prueba_bondad_model2
```

```
modelo_step_wise<-modelo2
```

```
DescTools::PseudoR2(modelo_step_wise, which = "McFadden")
```

```
p<-as.data.frame(colnames(as.data.frame(modelo2$model)))
```

```
#write.csv(p,"p.csv",row.names=F)
```

```

summary(model1)

summary(modelo2)

resumen_m1<-summary(model1)

resumen_m2<-summary(modelo2)

#-----Calculamos el pseudo R cuadrado de McFadden (“residual deviance” / “null
deviance”):

#-----los resultados entre 0,2 y 0,4 indican un excelente ajuste del modelo.

prueba_bondad #-- tiene un mejor ajuste

prueba_bondad_model2

#----- Criterio de Akaike

resumen_m1$aic

resumen_m2$aic #mejor ajuste

#----- Convertimos en un dataframe los resultados obtenidos del modelo.

coeficientes_resultado_modelo_2<-as.data.frame(resumen_m2$coefficients)

coeficientes_resultado_modelo_2$`Pr(>|z|)_2`<-
round(coeficientes_resultado_modelo_2$`Pr(>|z|)` ,7)

#----- Exportamos los coeficientes del modelo en un archivo csv

#write.csv(coeficientes_resultado_modelo_2, "coeficientes modelo 2_.csv",row.names = F)

#----- Aplicamos del modelo a toda la base de usuarios que tenemos

resultados_model2<-as.data.frame(predict(modelo2,type="response"))

#----- Se asignan a cada usuario su probabilidad de compra en ecommerce

data_usuarios_predict <-cbind(data_usuarios,resultados_model2)

#----- Ajustamos nombre de la columna

colnames(data_usuarios_predict)[123]<-"Probabilidad de compra en Ecommerce"

```

```

#----- Se realiza una resumen de las probabilidades

summary(data_usuarios_predict$`Probabilidad de compra en Ecommerce`)

#----- Añadimos una variable para dar rangos de probabilidad

data_usuarios_predict$intervalo<-if_else(data_usuarios_predict$`Probabilidad de compra
en Ecommerce`<=0.10, "[0%,10%]",

      if_else(data_usuarios_predict$`Probabilidad de compra en
Ecommerce`>0.10 & data_usuarios_predict$`Probabilidad de compra en
Ecommerce`<=0.20, "(10%,20%]",

      if_else(data_usuarios_predict$`Probabilidad de compra en
Ecommerce`>0.20 & data_usuarios_predict$`Probabilidad de compra en
Ecommerce`<=0.30, "(20%,30%]",

      if_else(data_usuarios_predict$`Probabilidad de compra
en Ecommerce`>0.30 & data_usuarios_predict$`Probabilidad de compra en
Ecommerce`<=0.40, "(30%,40%]",

      if_else(data_usuarios_predict$`Probabilidad de
compra en Ecommerce`>0.40 & data_usuarios_predict$`Probabilidad de compra en
Ecommerce`<=0.50, "(40%,50%]",

      if_else(data_usuarios_predict$`Probabilidad
de compra en Ecommerce`>0.50 & data_usuarios_predict$`Probabilidad de compra en
Ecommerce`<=0.60, "(50%,60%]",

if_else(data_usuarios_predict$`Probabilidad de compra en Ecommerce`>0.60 &
data_usuarios_predict$`Probabilidad de compra en Ecommerce`<=0.70, "(60%,70%]",

if_else(data_usuarios_predict$`Probabilidad de compra en Ecommerce`>0.70 &
data_usuarios_predict$`Probabilidad de compra en Ecommerce`<=0.80, "(70%,80%]",

if_else(data_usuarios_predict$`Probabilidad de compra en Ecommerce`>0.80 &
data_usuarios_predict$`Probabilidad de compra en Ecommerce`<=0.90, "(80%,90%]",

if_else(data_usuarios_predict$`Probabilidad de compra en Ecommerce`>0.90 &
data_usuarios_predict$`Probabilidad de compra en Ecommerce`<=1, "(90%,100%]", "sin
rango"))))))))

```

```

#----- Pegamos el factor de expansión en la base
factor<-select(data,ID_hog_modif,FAC_PER)
data_usuarios_predict<-merge(data_usuarios_predict,factor,by=c("ID_hog_modif"))
sum(data_usuarios_predict$FAC_PER)

#----- se filtra la base de datos para poder analizar a los usuarios que aún no realizan
compras en ecommerce

sin_compra_ecommerce<-filter(data_usuarios_predict,compra_ecommerce==0)
sum(sin_compra_ecommerce$FAC_PER)

#----- se crea una tabla de frecuencias con los usuarios que aún no realizan compras en
ecommerce

población_por_intervalo_de_proabilidad<-sin_compra_ecommerce %>%
  group_by(intervalo) %>%
  summarise(poblacion=sum(FAC_PER))
sum(población_por_intervalo_de_proabilidad$poblacion)

#----- Añadimos variables para tener en porcentaje los datos

población_por_intervalo_de_proabilidad$porcentaje<-
round((población_por_intervalo_de_proabilidad$poblacion/(sum(población_por_intervalo_
de_proabilidad$poblacion)))*100,1)

población_por_intervalo_de_proabilidad$porcentaje_str<-
as.factor(str_c(as.character(población_por_intervalo_de_proabilidad$porcentaje),"%"))

#----- Comprobamos que la data este completa con la suma de las variables

sum(población_por_intervalo_de_proabilidad$poblacion)
sum(población_por_intervalo_de_proabilidad$porcentaje)

#----- Ajustamos el data frame para que el orden se mantenga creciente en los rangos
ajust<-población_por_intervalo_de_proabilidad[10,]

```

```

población_por_intervalo_de_proabilidad<-población_por_intervalo_de_proabilidad[-c(10),
]

common_cols <- intersect(colnames(ajust),
colnames(población_por_intervalo_de_proabilidad))

población_por_intervalo_de_proabilidad<-rbind(subset(ajust, select = common_cols),
subset(población_por_intervalo_de_proabilidad, select =
common_cols))

bar_probabilidades<-barplot(población_por_intervalo_de_proabilidad$poblacion,
names=población_por_intervalo_de_proabilidad$intervalo,
xlab = "Intervalo de probabilidad de realizar una compra por internet",
ylab = "Número de personas",
col=c("#9EDAE5", "#F7B6D2", "#98DF8A", "#FF9896", "#C5B0D5", "#C7C7C7", "#DBDB
8D", "#C49C94", "#1A9993", "#C80813", "#1A5354", "#EC407A") ,ylim=c(0,30000000))

```