

TRABAJO DE GRADO
Opción Seminario-Diplomado.

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES
EN DATOS DE PEDIDOS DE ALIMENTOS EN LINEA, UTILIZANDO ESTRATEGIAS
DE MACHINE LEARNING**

Corporación Universitaria Remington.

Nombre de la facultad: Facultad de Ingenierías

Nombre del programa académico:

Seminario de Machine Learning en Tiempos de Datos

Estudiante: Daryl Giselle Leguizamo Rodriguez

Tutor: Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.

2024

Dedicatoria

Dedico este trabajo de grado a Jehová Dios porque ha sido mi fortaleza en este proceso de formación como Desarrolladora de Software y en todo el camino conté con su guía para lograr culminar este meta satisfactoriamente. Por otra parte, quiero agradecer a mi familia todo el apoyo que me brindaron en todo momento, siendo pacientes y comprensivos en este proceso de formación y quiero dedicar este trabajo a mis amigos incondicionales que me motivaron día tras día a continuar y ver el futuro con optimismo.

Tabla de Contenidos

Resumen	4
Marco conceptual y contextual	6
Objetivos	
Objetivo general	9
Objetivos Específicos	9
Desarrollo e implementación del aprendizaje	
Análisis de la base de datos	10
Graficas	10
Implementación en contextos reales	24
Conclusiones	30
Referencias	32

Resumen

Actualmente nos encontramos en la era de los datos, muchas empresas por bastante tiempo contaron con un excelente manejo de sus procesos, pero debido a que el mundo sigue en constante cambio las entidades se han visto obligadas a utilizar nuevas técnicas e implementar métodos digitales para que sus empresas puedan ampliarse y llegar a mayor público. Sin embargo, todos esos datos recolectados con el tiempo pueden aportar información clave y relevante que brinde ayuda para mejorar o comprender como se maneja la entidad y con que utilidades puede disponer para brindar un servicio de calidad.

Por esa razón, el objetivo de este trabajo de grado es dejar en manifiesto como el aprendizaje automático más conocido también como el Machine Learning puede contribuir a la toma de decisiones en estos tiempos actuales de datos. Se quiere demostrar que podemos sacar provecho de la era digital para nuestro beneficio, con ayuda de la inteligencia artificial, el Machine Learning usaremos todas las variables y los datos integrando las herramientas o algoritmos de esta inteligencia para sacar posibles conclusiones y resultados que aporten a la entidad de manera constructiva.

Se utilizará una base de datos para realizar el desglose de la información y tener un análisis detallado de las variables para tomar una decisión en base a los resultados. Esta base de datos es de una empresa que tiene como objetivo entregar pedidos de alimentos en línea, usaremos técnicas del Machine Learning para consultar la información, obtener como se dividen los datos y que aspectos se toman en cuenta en la entidad de entrega de pedidos de alimentos en línea.

Además, se evaluará como estas variables se relacionan entre sí y que puede aportarnos del comportamiento que ejerce la entidad en su lugar de operación, al analizar estos datos a profundidad podemos implementar la técnica de Machine Learning mencionada anteriormente. Esta técnica denominada “aprendizaje no supervisado (agrupación)” tiene

como objetivo brindar resultados que estén orientados a mejorar la producción de la entidad o relacionar como las variables se parecen entre sí con miras a la toma de futuras decisiones. Por ende, se visualizará como al integrar nuevos registros a la base de datos después de usar un método de segmentación, los datos se reparten y nos genera una predicción de en qué agrupación de la información corresponde y como esto nos da una idea del manejo de la entidad y su relación entre sí.

Palabras clave

Base de datos: conjunto de información que pertenece a un mismo contexto, este puede estar ordenado en modo físico o en modo digital con el fin de preservar la información y llevar un orden de la misma.

Machine Learning: el aprendizaje automático conocido como Machine Learning es un subcampo de la ciencia de la computación que tiene como objetivo reconocer patrones complejos de un conjunto de datos y así aprenda a brindar una ejecución mejorada.

Algoritmo: es una secuencia de instrucciones paso a paso que realiza alguna operación o tarea específica, este permite controlar determinados procesos.

Inteligencia Artificial: es la habilidad de que las computadoras o los ordenadores puedan realizar ciertas actividades que normalmente realizaríamos los seres humanos.

Técnica: es un conjunto de procedimientos o reglas que permiten cumplir con un objetivo concreto o determinado.

Marco conceptual y contextual

Machine learning es la capacidad de permitir que la computadora tome decisiones en base de varios datos. Esta ha permitido que en la actualidad muchas empresas estén migrando sus datos a la digitalización, esto está contribuyendo a una gran ventaja a nivel económico para dichas entidades. Implementar el machine learning en la digitalización de su información esta contribuyendo a un mejor manejo de los mismos y como estos pueden usarse o evaluarse a plenitud, logrando obtener grandes beneficios y decisiones acertadas.

Sin embargo, se requiere usar las técnicas adecuadas y contar con el conocimiento para usar esta inteligencia artificial. El machine Learning surge útil cuando se manejan datos extensos de información, esto permite obtener una vista mas detallada de como el manejo de cierta empresa puede mejorar y aprovechar ciertas acciones que se dejan en evidencia al usar Machine Learning en los datos.

En el uso de machine Learning nos encontramos con varias técnicas que permiten hacer un desglose de la información y aporten excelentes resultados permitiendo tener un mejor control y manejo de laguna decisión. Para hacer uso de alguna se debe tener presente cual es el enfoque que se le quiere dar a la investigación, podemos usar algunos de los siguientes métodos.

- Linear regressions o regresiones: si deseamos realizar una estimación o predicción a futuro de una empresa podemos usar esta técnica de regresión. Tiene como objetivo establecer en base a datos de entrada y salida un futuro estimado de la empresa.
- Aprendizaje no supervisado (Agrupación): En el caso de que nuestra investigación este orientada a mejorar la producción de una empresa o detallar

como los datos se parecen entre si podemos usar el método de “aprendizaje no supervisado (agrupación)” el cual nos permite buscar una relación entre la información y en base a ello tomar decisiones adecuadas para un mejor beneficio a la entidad.

- Aprendizaje supervisado (Clasificación): este método podemos usarlo cuando nuestra investigación tiene como objetivo usar datos que en el pasado ya tuvieron un estudio o en base a ellos se tomaron decisiones. Se usa para buscar mayores utilidades.

Todos estos métodos de machine Learning cumplen con un objetivo preciso, por ende, el machine Learning se ha convertido en una gran herramienta para el uso de los datos en las empresas como se mencionó al principio de este informe. Para demostrar como el Machine Learning es una herramienta de inteligencia artificial muy provechosa tomaremos de ejemplo una empresa “OnlineFoods” que entrega pedidos de alimentos en línea.

Estas plataformas de entrega de alimentos en línea están generando un incremento en las ventas y ahorros para dichas empresas, al aumentar su uso es importante realizar un análisis de la información y contar con estrategias que permitan mejorar los servicios de esa entidad. En el caso de la investigación que pretendemos enfocar por medio de este trabajo vamos a usar diferentes algoritmos y métodos mencionados antes, con todos esos datos podemos tomar decisiones acertadas y ver a detalle cómo nos arrojan información que pueda contribuir a un buen avance de la empresa de pedidos de alimentos en línea.

¿Qué información tenemos de la plataforma de pedidos de comida en línea?

Se tienen datos de información recolectada de esta plataforma por un determinado tiempo. Se abarcan muchos aspectos en la información como los siguientes:

- La ocupación del cliente.

- Edad del cliente
- El género.
- Estado civil.
- Ingreso mensual del cliente.
- Tamaño de la familia.
- Cualificaciones educativas del cliente.
- Latitud y longitud para la ubicación del cliente.
- Código para ubicar el cliente.
- Estado actual del pedido.
- Comentarios proporcionados por los clientes.

Con esta información podemos implementar varias técnicas de machine Learning que nos pueden arrojar datos claves acerca de cómo la ubicación geográfica afecta o contribuye a la frecuencia con que los clientes solicitan el servicio. Sin embargo, de estos datos también podemos analizar como la ocupación de los clientes se relaciona con la ubicación geográfica y contribuye o no a un aumento en el servicio y por ende variables como la edad y el estado civil influye en el uso del servicio de pedidos que se realizan en la plataforma.

Estos datos nos van a proporcionar una orientación para entender como la empresa esta llevando el manejo de la entidad y como se puede tomar decisiones en base a la información para tener una mejora en su productividad.

Pregunta problema:

¿En que aspectos debe tener una mejoría la empresa de pedidos de alimentos en línea mediante el uso de técnicas u algoritmos de machine Learning?

Objetivos:**Objetivo general.**

Implementar algoritmos de machine learning que permitan obtener un análisis de los datos de pedidos de alimentos en línea y se pueda llegar a una toma de decisiones acertada para mejorar y comprender el funcionamiento de la entidad.

Objetivos específicos.

- Conocer y procesar los datos necesarios para proceder a la toma de decisiones de mejoras en una empresa de pedidos de alimentos en línea “OnlineFoods”.
- Implementar un algoritmo de Machine learning para realizar un análisis de los datos y proceder a la toma de decisiones a partir de los datos de pedidos de alimentos en línea “OnlineFoods”.
- Evaluar y analizar el desempeño de los algoritmos que se implementaron para la toma de decisiones en una empresa de pedidos de alimentos en línea “OnlineFoods”.
- Validar el funcionamiento del modelo utilizado para la toma de decisiones a partir de datos nuevos ingresados en la base de datos de la entidad de pedidos de alimentos en línea “OnlineFoods”.

Desarrollo e implementación del aprendizaje

La empresa de pedidos de alimentos en línea tiene como información los datos de sus clientes, se tiene registro de su información personal como nombre, edad, estado civil, salario, entre otros aspectos.

Vamos a realizar un algoritmo para conocer y procesar estos datos de forma mas detallada y empezar a dar un análisis con miras a decisiones.

- **Conocer la base de datos de la empresa de pedidos de alimentos en línea.**

(Tabla 1. Base de datos de la empresa de entrega de pedidos de alimentos en línea)

	Age	Gender	Marital Status	Occupation	Monthly Income	Educational Qualifications	Family size	latitude	longitude	Pin code	Output	Feedback	Unnamed: 12
0	20	Female	Single	Student	No Income	Post Graduate	4	12.9766	77.5993	560001	Yes	Positive	Yes
1	24	Female	Single	Student	Below Rs.10000	Graduate	3	12.9770	77.5773	560009	Yes	Positive	Yes
2	22	Male	Single	Student	Below Rs.10000	Post Graduate	3	12.9551	77.6593	560017	Yes	Negative	Yes
3	22	Female	Single	Student	No Income	Graduate	6	12.9473	77.5616	560019	Yes	Positive	Yes
4	22	Male	Single	Student	Below Rs.10000	Post Graduate	4	12.9850	77.5533	560010	Yes	Positive	Yes
5	27	Female	Married	Employee	More than 50000	Post Graduate	2	12.9299	77.6848	560103	Yes	Positive	Yes
6	22	Male	Single	Student	No Income	Graduate	3	12.9770	77.5773	560009	Yes	Positive	Yes
7	24	Female	Single	Student	No Income	Post Graduate	3	12.9828	77.6131	560042	Yes	Positive	Yes
8	23	Female	Single	Student	No Income	Post Graduate	2	12.9766	77.5993	560001	Yes	Positive	Yes
9	23	Female	Single	Student	No Income	Post Graduate	4	12.9854	77.7081	560048	Yes	Positive	Yes

- En la tabla 1 podemos encontrar los datos de la empresa de entrega de pedidos en línea, se describen las variables correspondientes de la empresa. Entre ellas podemos destacar el ámbito feedback (comentarios) de los clientes después de recibir el pedido, la cual se describe como si el comentario fue positivo o negativo sin mencionar los detalles. También podemos destacar la parte de la columna de

famiily size (tamaño de la familia) porque nos aportaran información para realizar el análisis de cómo influye este aspecto en las solicitudes de comida.

- **Algunas columnas las retiramos porque no serán de caso de análisis en este informe.**
- Las columnas que vamos a retirar son las de “output”, si se encuentra listo el pedido o no. También vamos a retirar la del tipo de “educación” que maneja el cliente ya que no es motivo de análisis en este trabajo y la sección del “Pin code” la retiramos de la misma manera y la de “Unnamed: 12” la cual no es necesario los nombres de los clientes.

Por lo tanto, la tabla queda de la siguiente manera:

(Tabla 2. Base de datos de onlineFoods actualizada)

	Age	Gender	Marital Status	Occupation	Monthly Income	Family size	latitude	longitude	Feedback
0	20	Female	Single	Student	No Income	4	12.9766	77.5993	Positive
1	24	Female	Single	Student	Below Rs.10000	3	12.9770	77.5773	Positive
2	22	Male	Single	Student	Below Rs.10000	3	12.9551	77.6593	Negative
3	22	Female	Single	Student	No Income	6	12.9473	77.5616	Positive
4	22	Male	Single	Student	Below Rs.10000	4	12.9850	77.5533	Positive
5	27	Female	Married	Employee	More than 50000	2	12.9299	77.6848	Positive
6	22	Male	Single	Student	No Income	3	12.9770	77.5773	Positive
7	24	Female	Single	Student	No Income	3	12.9828	77.6131	Positive
8	23	Female	Single	Student	No Income	2	12.9766	77.5993	Positive
9	23	Female	Single	Student	No Income	4	12.9854	77.7081	Positive

Como se puede evidenciar en la tabla 2 dejamos las columnas que serán de análisis en este informe. Contaremos con la edad, estado civil, ocupación, salario, tamaño de la familia, latitud y longitud de la zona y el comentario del cliente al recibir el pedido.

- **Convertir las variables de texto a numéricas para realizar el análisis mas detallado.**
- Aunque algunos de los datos que manejamos son variables en tipo string o texto, para realizar un correcto análisis debemos convertirlas en tipo numérico para identificar mejor los resultados.

Los cambios que se realizaran son los siguientes:

(Tabla 3. Datos que se actualizaron de texto a numéricos en la base de datos)

Variables de la base de datos y los cambios de texto a numerico.		
Variable	Tipo String	Tipo numerico
Gender	Female	1
	Male	2
Marital Status	Single	1
	Married	2
	Prefer not to say	0
Occupation	Student	0
	Employee	1
	Selft Employee	2
	House wife	3
Monthly Income	No Income	0
	Below RS. 10000	1
	More than 50000	2
	10001 to 25000	3
	25001 to 50000	4
Feedback	Positive	1
	Negative	0

Actualizando la tabla de la base de datos nuevamente en base a la tabla 3 mencionada anteriormente, la información procede a quedar de la siguiente manera:

(Tabla 4. Datos actualizados en tipo numérico de la base de datos)

	Age	Gender	Marital Status	Occupation	Monthly Income	Family size	latitude	longitude	Feedback
0	20	1	1	0.0	0	4	12.9766	77.5993	1
1	24	1	1	0.0	1	3	12.9770	77.5773	1
2	22	2	1	0.0	1	3	12.9551	77.6593	0
3	22	1	1	0.0	0	6	12.9473	77.5616	1
4	22	2	1	0.0	1	4	12.9850	77.5533	1
5	27	1	2	1.0	2	2	12.9299	77.6848	1
6	22	2	1	0.0	0	3	12.9770	77.5773	1
7	24	1	1	0.0	0	3	12.9828	77.6131	1
8	23	1	1	0.0	0	2	12.9766	77.5993	1
9	23	1	1	0.0	0	4	12.9854	77.7081	1

Además, eliminamos los datos duplicados para tener un resultado mas detallado.

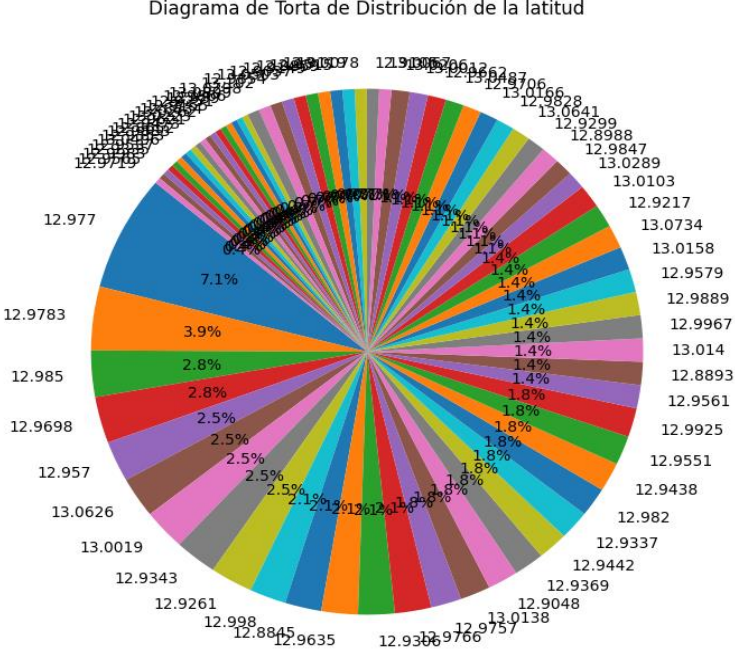
(Tabla 5. Datos duplicados eliminados de la base de datos)

```
Número de filas antes de eliminar duplicados: 388
Número de filas después de eliminar duplicados: 281
```

Teniendo la información organizada y establecida correctamente procedemos a realizar el siguiente paso, implementar un algoritmo de machine learning para el análisis y la toma de decisiones a partir de los datos de pedidos de alimentos en línea “OnlineFoods”.

Empezamos a profundizar en los valores que encontramos de la ubicación geográfica, como esta contribuye o aporta datos claves en la zona donde mas se reparten las entregas de comida y como puede mejorar la entidad. Debemos recordar que los datos de estos están es por latitud y longitud. Realizáremos un algoritmo que permita ver como se reparten en este aspecto las zonas y un balance de ellas mismas. Utilizaremos la técnica de “aprendizaje no supervisado (Agrupación)”.

Realizamos dos graficas de los datos de latitud y longitud:



(Imagen 1. Grafica de Latitud en la base de datos)

Podemos analizar en la imagen 1 que de los valores que se presentan para la ubicación geográfica de longitud el mayor cuenta con un 7.1% el cual corresponde a la longitud de 12.977. Las siguientes longitudes que se presentan en porcentaje siguientes como el de 3.9% equivale a una longitud muy cercana a la mayor, la cual es 12.9783.

Tomando valores de la gráfica de latitudes como las siguientes tenemos:

(Tabla 6. Latitudes a considerar de la grafica)

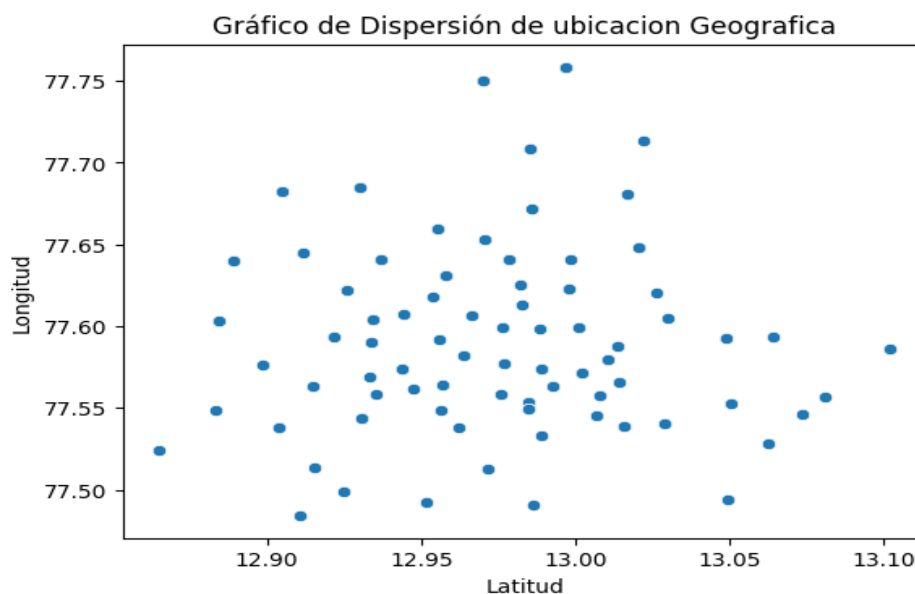
Latitudes a considerar de la grafica	
Latitud	Porcentaje
12.977	9.30%
12.9783	4.10%
12.985	3.60%
12.8845	3.10%

(Tabla 7. Longitudes a considerar de la gráfica)

Longitudes a considerar de la grafica	
Longitud	Porcentaje
77.5773	9.30%
77.6408	4.10%
77.5533	3.60%
77.6036	3.10%

En la tabla 7 podemos ver la relación que se encuentra entre las longitudes con mayor porcentaje, todas corresponden a un cuadro de zona geográfico donde se realizan con frecuencia entregas de pedidos de alimentos.

Si usamos una gráfica de dispersión para comparar los datos de longitud y latitud podemos encontrar ahora una relación entre ellos mismo.

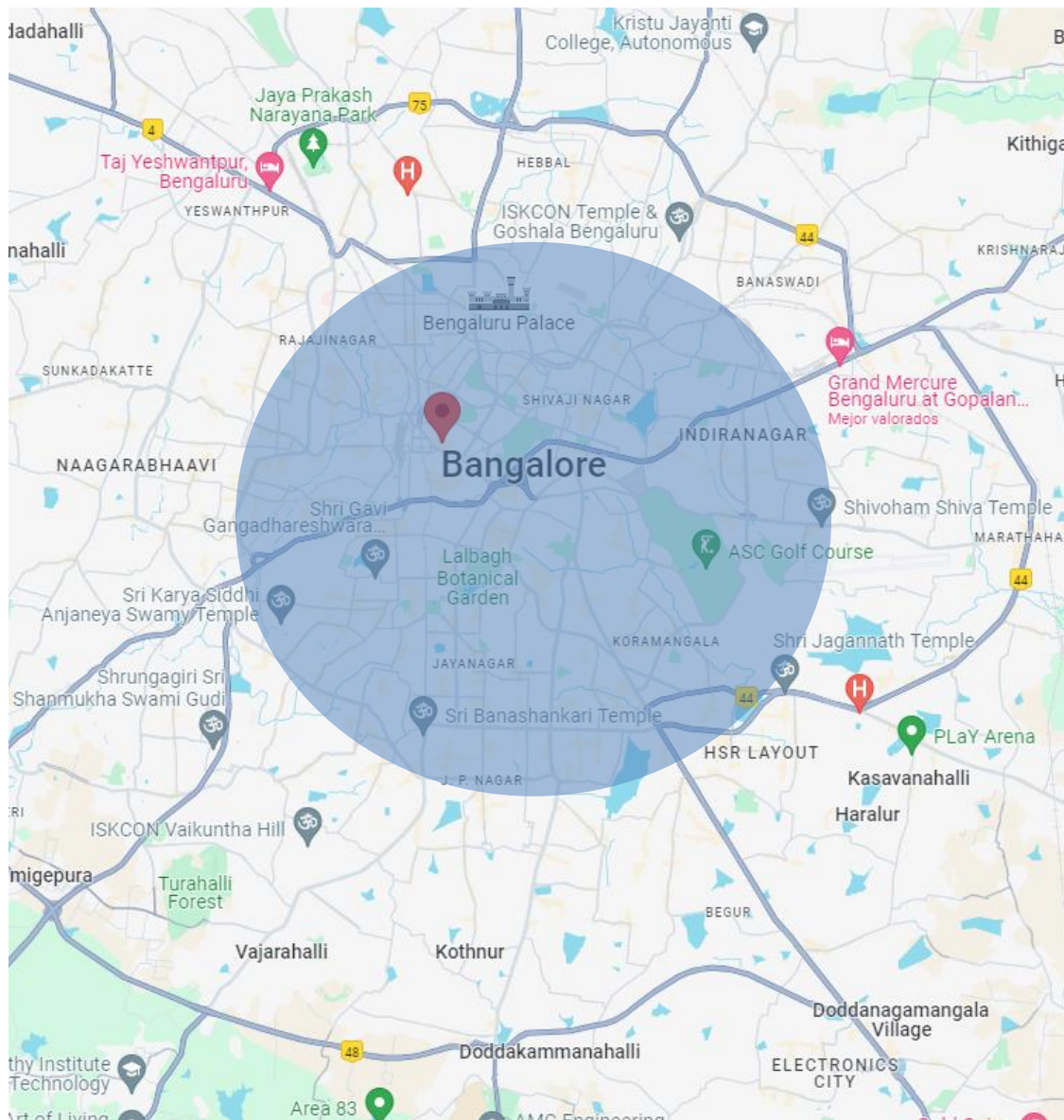


(Imagen 3. Grafica de dispersión de la ubicación geográfica)

Según la gráfica de dispersión podemos encontrar que la mayor concentración de ubicación geográfica es en longitud de 77.55 a 77.65 y en cuanto a latitud esta corresponde desde 12.93 aproximadamente a 13.01 aproximadamente.

Usando un mapa podemos ubicar estas direcciones y de finir que zona de la ciudad donde opera esta empresa de entrega de pedidos de alimentos en línea suele ser mas frecuente.

Nos ayudaremos de la herramienta de “Google maps”.



(Imagen 3. Mapa de la zona frecuente de entrega de alimentos en línea)

Ubicando las coordenadas de longitud y latitud en el mapa podemos encontrar que la circunferencia en la imagen corresponde a la zona con mas frecuencia de la ciudad de Bangalore (India) donde se realizan entregas de pedidos de alimentos en línea. Esta zona corresponde al centro de la ciudad donde se encuentra mucho comercio y zona de lugares

de trabajo, encontramos edificaciones y centros de educación. Además, de tener lugares como turismo para los residentes y extranjeros.

Este análisis de la ubicación geográfica demuestra que la empresa realiza una mayor entrega de alimentos en la zona principal de la ciudad como se menciona anteriormente. Pero, se debe tener presente que encontramos otras razones por las cuales es una zona muy apetecida por la entrega en línea.

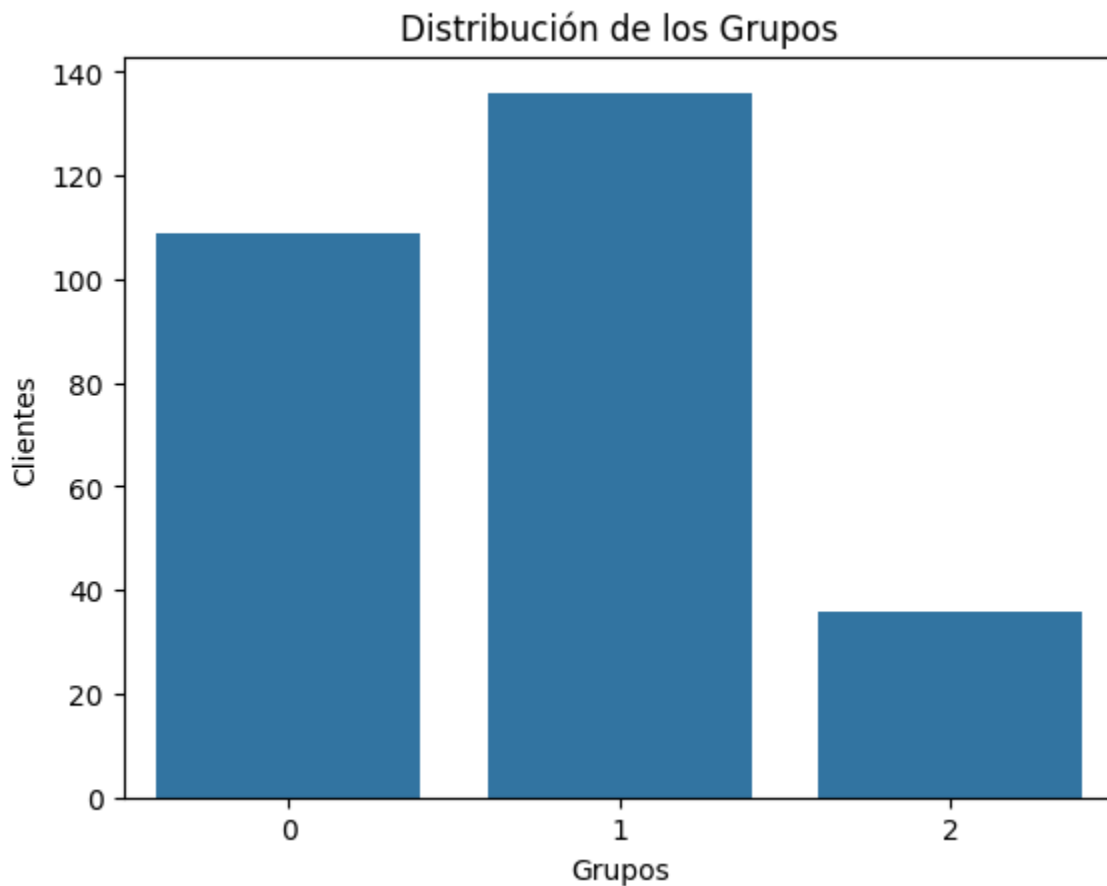
Evaluemos otras variables como la edad de quienes solicitan el servicio y el tamaño de las familias teniendo presente también su ocupación y estado civil, factores importantes que nos darán un cuadro mas amplio para comprender como la empresa de alimentos opera y puede mejorar en base a lo encontrado.

Para encontrar una relación entre estas variables y la ubicación geográfica, ahora si hacemos un uso mayor de la técnica de “aprendizaje no supervisado, (Agrupación)”. Haremos uso de un algoritmo de machine Learning para proceder a dividir los datos en 3 grupos, estos nos darán mayor claridad.

Grupo:						
	Age	Gender	Marital Status	Occupation	Monthly Income	\
0	25.467890	1.550459	1.385321	1.155963	2.889908	
1	22.514706	1.595588	1.044118	0.088235	0.286765	
2	30.694444	1.555556	1.833333	1.638889	2.083333	
	Family size	latitude	longitude	Feedback		
0	3.119266	12.972655	77.604606	0.788991		
1	3.102941	12.973035	77.592270	0.867647		
2	4.305556	12.976106	77.599200	0.638889		

(Imagen 4. Datos de la base de datos divididos en grupos)

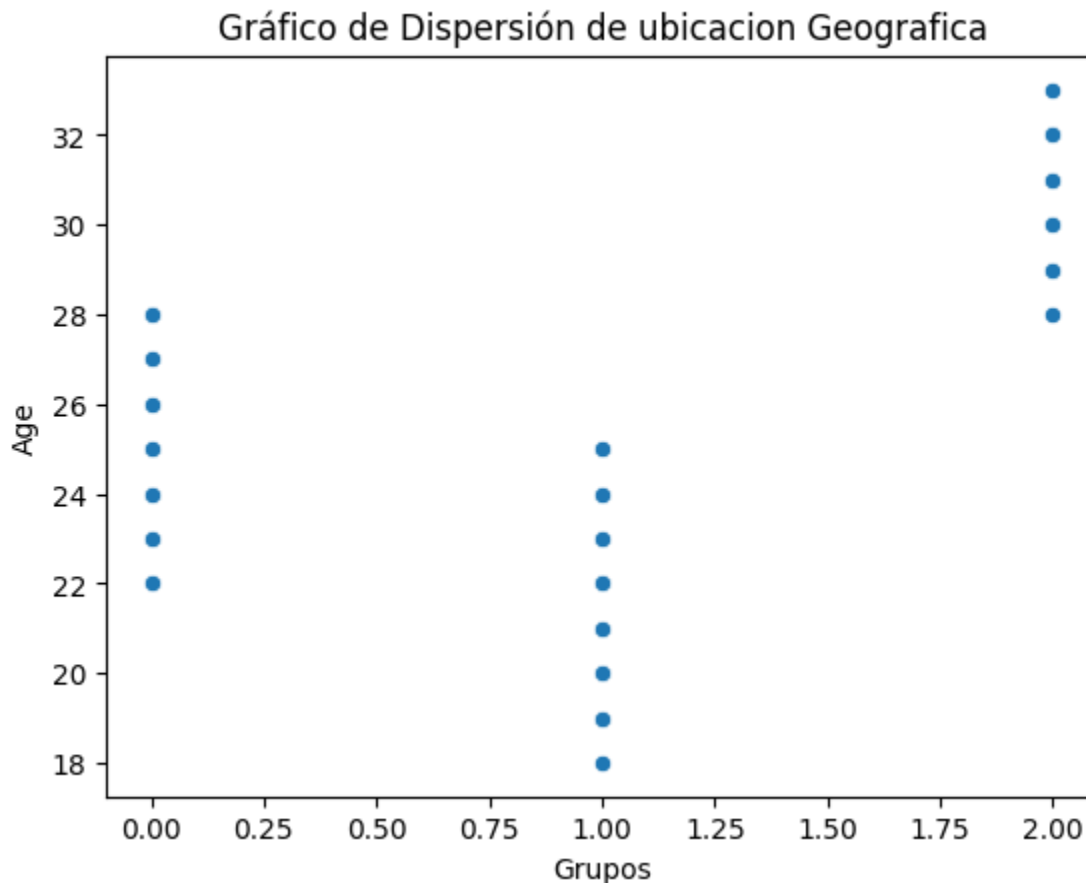
Usamos un grafico de barras para representar a estos tres grupos:



(Imagen 5. Grafica de barras de los grupos)

En estos datos encontramos varias similitudes entre los grupos creados, entre ellos podemos notar que la ubicación geográfica esta muy relacionada con lo mencionado anteriormente, mostrando o dejando en evidencia donde opera con frecuencia la entidad. Haciendo uso de estos datos agrupados procedemos a buscar relación con las otras variables.

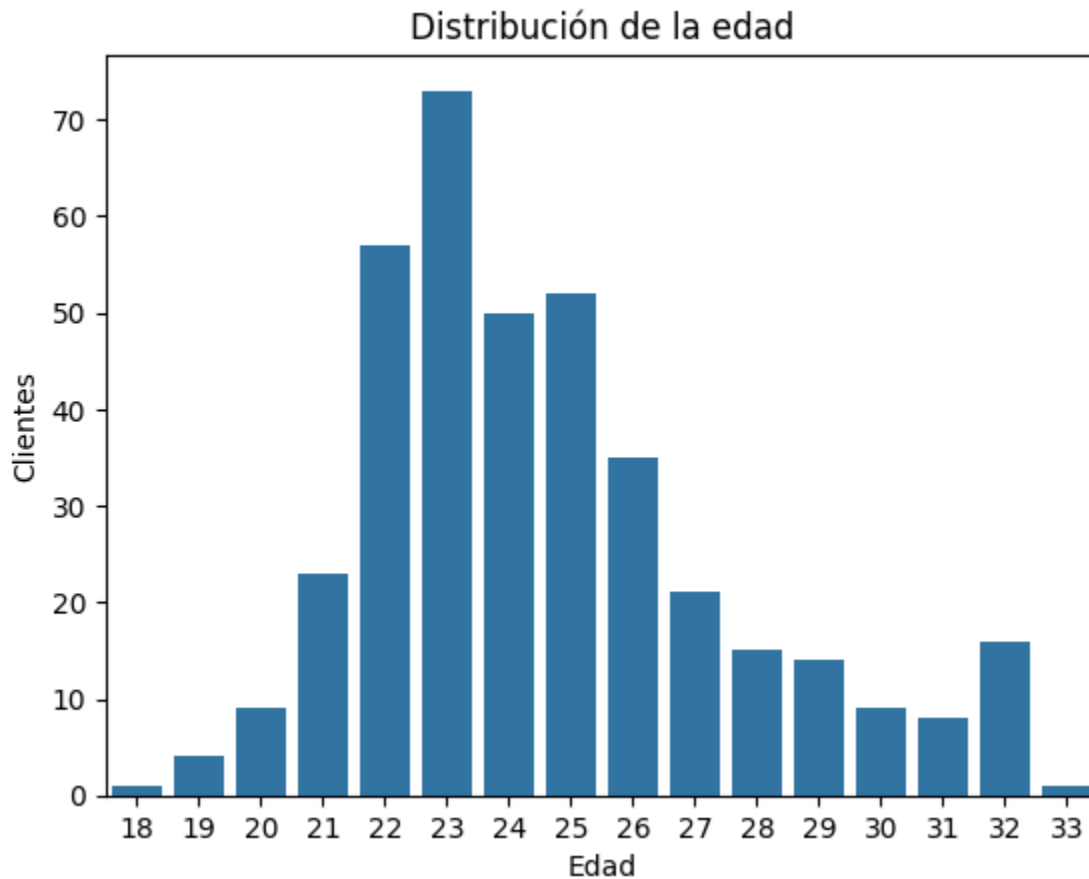
En la siguiente grafica de dispersión podemos ver que personas en un rango de edad son las que más solicitan el servicio y como se relaciona esto con la ubicación geográfica.



(Imagen 6. Grafica de dispersión de la edad con relación a la ubicación geográfica)

En la grafica encontramos que el grupo con mayor concentración es el “1”, este grupo según la grafica esta en un rango de edad de 18 a 25 años de edad y el cual también nos aporta que es el mayor grupo que reside en la zona centro de la ciudad donde se solicitan los pedidos de comida. En cuanto a los demás grupos como el “0” se encuentra en un rango de 22 a 28 años de edad y el grupo “2” se encuentra en un rango de 28 a 32 en delante de años de edad. Lo que se puede destacar que estos tres grupos cuentan con un solo grado de diferencia en la dispersión, el grupo “0” cuenta con 7 puntos de dispersión, el grupo “1” cuenta con 8 puntos de dispersión y el grupo “2” cuenta con solo 6 puntos de dispersión lo cual demuestra que no existe gran diferencia entre los rangos de edad, pero si una mayor concentración.

Aunque para ser mas exactos y no tomar en cuenta solo los rangos podemos hacer uso de un grafico de barras para encontrar cual es la edad que usa con frecuencia el servicio.

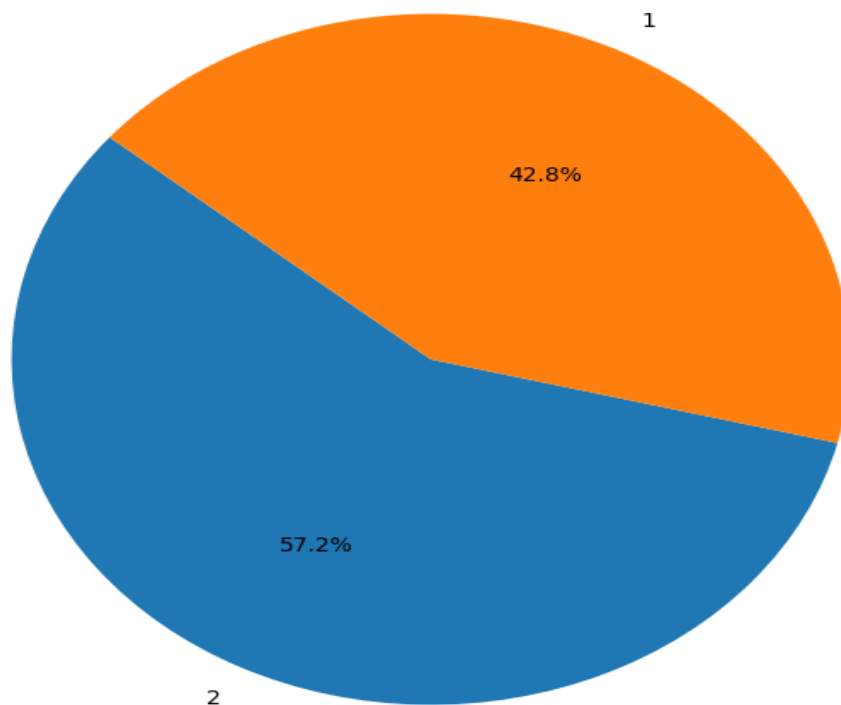


(Imagen 7. Grafica de barras de la edad)

La edad es de 23, estos jóvenes son los que más solicitan el servicio y la edad de 33 y 18 son los que menos solicitan el servicio. Sin embargo, el genero puede aportar un dato importante.

La siguiente grafica muestra este publico a que genero pertenece:

Diagrama de Torta de Distribución del Genero



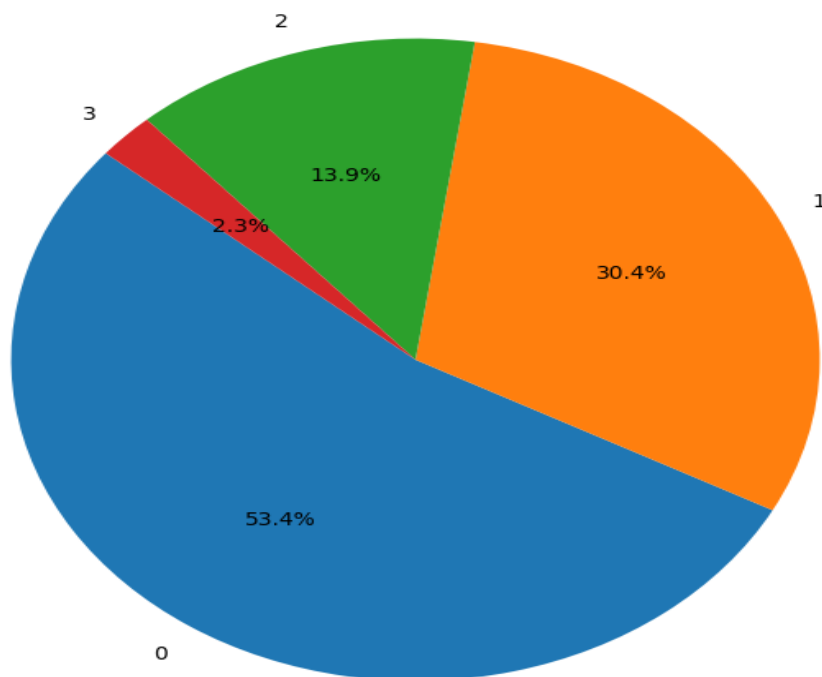
(Imagen 8. Diagrama de torta de Distribución de genero)

Según el diagrama podemos encontrar que el género Masculino, es el que presenta un porcentaje del 57.2% en relación con el genero femenino. Esto podemos relacionarlo con la zona geográfica ya que es lugar de comercio y entidades.

Sin embargo, podemos relacionar esto de la edad con otra variable como la ocupación para obtener más detalles de cómo son las personas que más solicitan un pedido y las que no solicitan con frecuencia el servicio.

La siguiente grafica nos muestra a detalle cual de los grupos se relaciona con la ocupación.

Diagrama de Torta de Distribución de la ocupacion

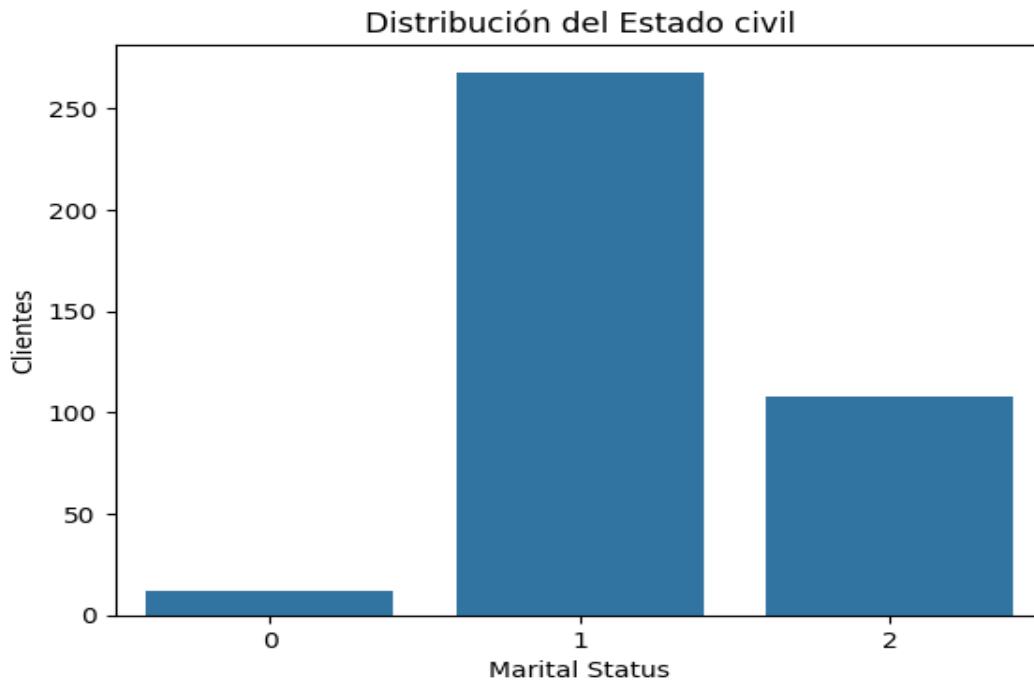


(Imagen 9. Diagrama de distribución de la ocupación)

Teniendo presente la información, encontramos que el mayor porcentaje de 53.4% corresponde a estudiantes y el siguiente porcentaje que le sigue de 30.4% corresponde a empleados. Esto demuestra que los jóvenes de 23 años que viven o estudian en el centro de la ciudad son lo que mas prefieren usar el servicio de la empresa de entrega de comida en línea.

Sin embargo, como buscamos tomar una decisión con miras a proporcionar una mejora a la entidad podemos tomar en cuenta que las amas de casa que corresponden al 2.3% y los desempleados que corresponden al 13.9% que estarían en el rango de edad de 18 y 33 años pueden ser un publico con potencial que genere incrementos en la entidad.

La siguiente grafica de barras demuestra que el estado civil de esta población de 23 años se encuentra en la siguiente:



(Imagen 10. Grafica de barras de distribución del estado civil)

La mayoría de jóvenes de 23 años que solicitan el servicio son solteros y se encuentran estudiando, en cambio los que menos solicitan el servicio se encuentran casados porque el otro porcentaje prefiere no mencionar su estado civil.

En base a estos análisis podemos encontrar que es importante para la empresa tomar estrategias que le brinden un incremento al publico que se encuentra en la zona de entregas y sus alrededores para que se motiven a frecuentar el uso de la empresa.

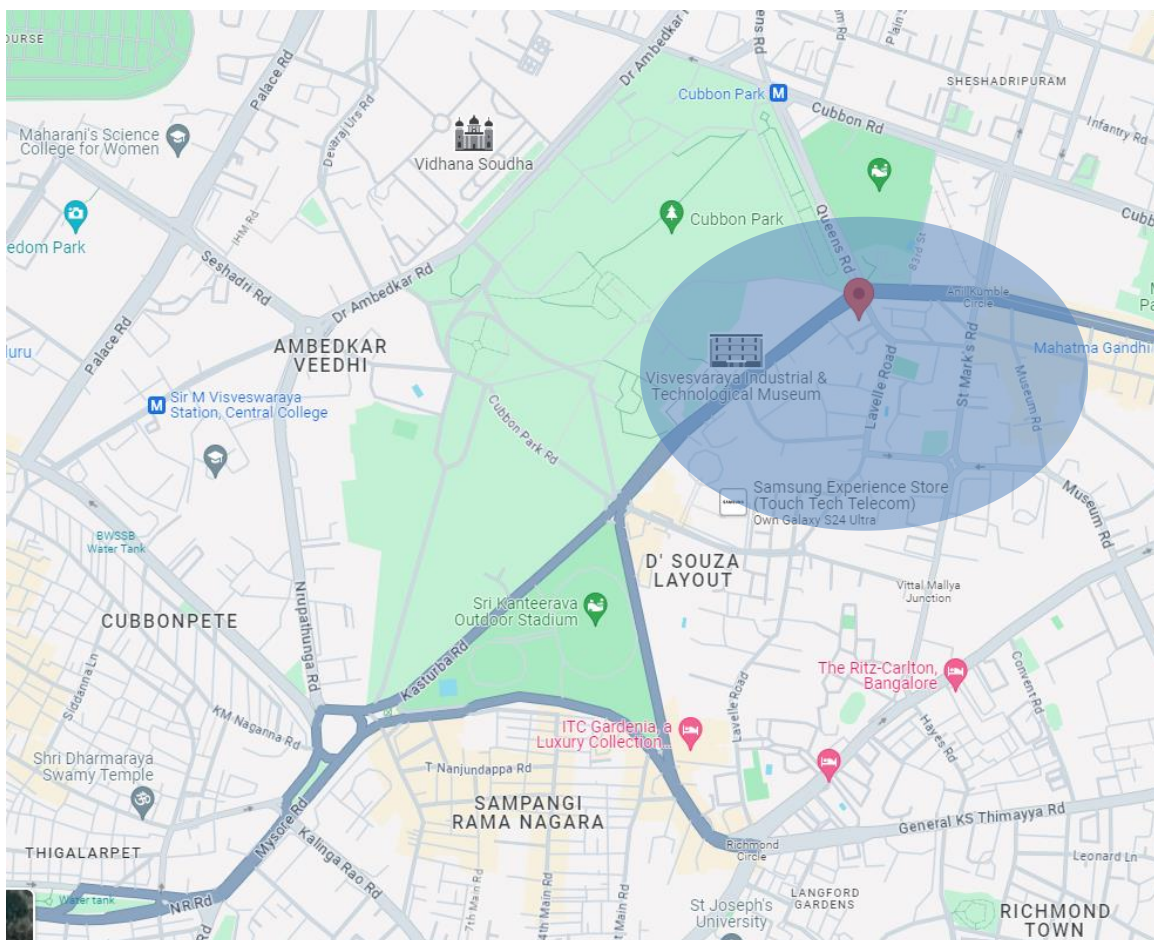
Implementación en contextos reales

Por esa razón vamos a utilizar el método y declarar si el nuevo cliente ingresado corresponde a alguno de los tres grupos obtendrá algún beneficio para que se anime a seguir solicitando el servicio.

Queremos que el grupo “2” que es el que menor rango tiene obtenga un incremento en la empresa por esa razón le brindaremos algunos beneficios a esa población. Recordemos de quienes está compuesto este grupo:

- Grupo 2: está en el rango de edad de 30 años, los cuales esta población la mayoría este casado y tiene una ocupación de empleado, empleado independiente o ama de casa.

Beneficios que se pueden otorgar en vista de la ubicación geográfica que maneja el grupo, parte de la zona residencial y sus alrededores, pero sin dejar de ser tan aislado de la ciudad podemos brindar un descuento a la compra de este grupo.

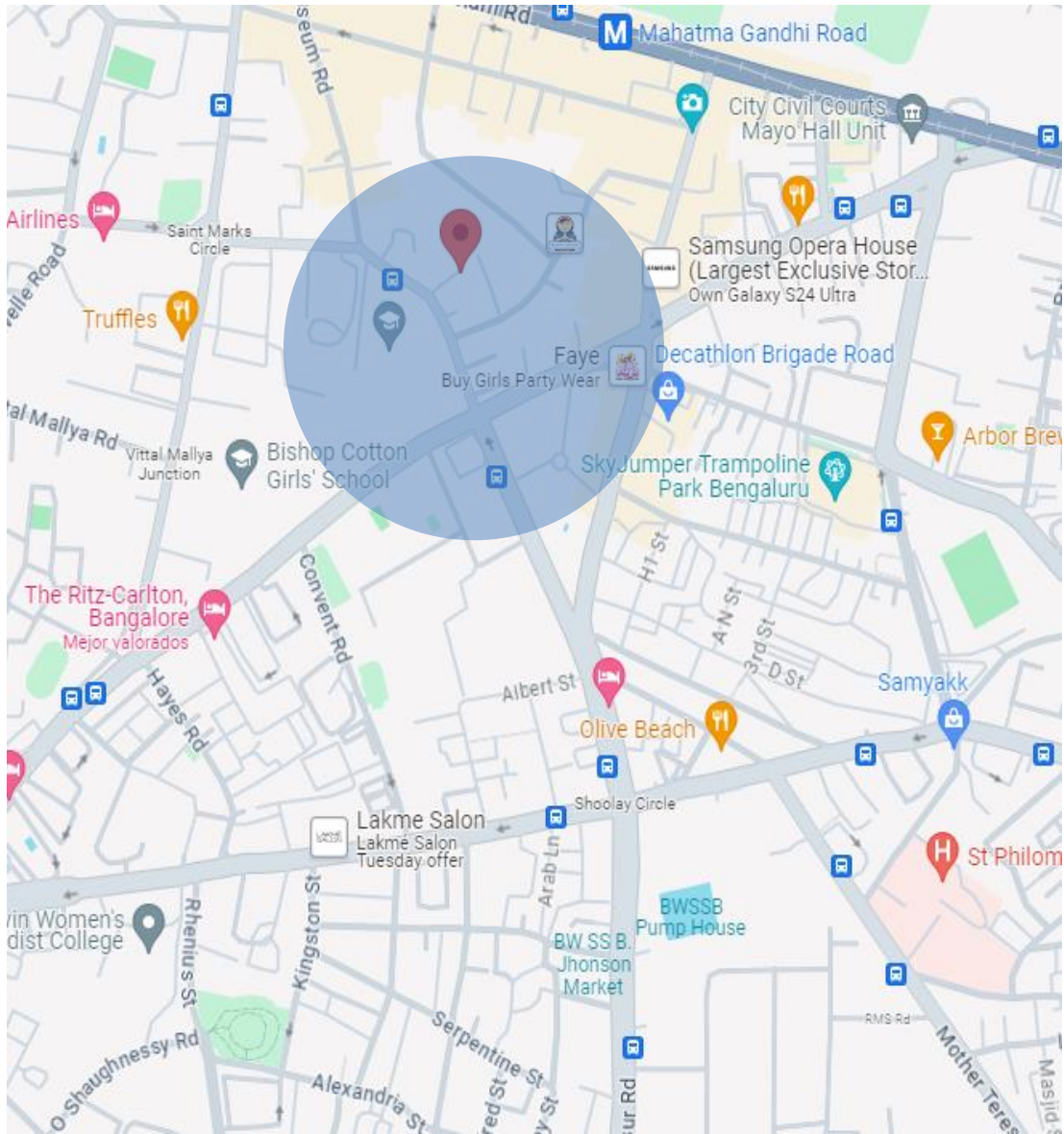


(Imagen 11. Mapa de ubicación geográfica del grupo 2 de la base de datos)

Tomamos como referencia la imagen 11 para esa decisión ya que este grupo se encuentra ubicado en una zona más residencial que de comercio por ende se puede deducir que las personas no suelen utilizar este servicio porque prefieren hacer los alimentos ellos mismos, esta hipótesis tiene referencia o relación que una parte de quienes conforman este grupo son las amas de casa. Como es público en su mayoría femenino podemos ofrecer descuentos o cupones en los meses donde se de honra a la mujer para motivarlas y agradecer su labor, eso las ayudara a sentirse valoradas y desearan solicitar nuevamente nuestros servicios.

- Grupo 1: para este grupo podemos recordar que este hace mayor parte de los jóvenes que estudian en el centro de la ciudad y otros que trabajan en la zona, en preferencia un publico masculino.
Para este grupo por ser frecuente en el servicio se le puede brindar promociones con tal de captar su atención y acceder a mas servicios con el fin de recibir ese beneficio.
- Grupo 0: es un grupo que se encuentra en un nivel medio, tiene parte de zona residencial pero también parte de zona comercial. Por ende, se mueve bastante, a este grupo podemos ofrecer un cupón por su primera compra o registro en la plataforma con eso fomentamos la competencia con otras entidades encargadas de entrega alimentos y motivamos al cliente a redimir ese cupón para obtener otro de los productos.

Obtenemos los siguientes resultados:



(Imagen 12. Mapa de ubicación geográfica del grupo 0 de la base de datos)

Obtenemos los siguientes resultados:

```
# Ingresar datos nuevos del cliente (9 variables)
Cliente_Nuevo=np.zeros((1,9))
Cliente_Nuevo[0,0]=23 #Edad
Cliente_Nuevo[0,1]=1 #Genero
Cliente_Nuevo[0,2]=1 #Estado civil
Cliente_Nuevo[0,3]=1 #Ocupacion
Cliente_Nuevo[0,4]=3 #IngresoMensual
Cliente_Nuevo[0,5]=1 #Tamaño familia
Cliente_Nuevo[0,5]=12.9644 #latitud
Cliente_Nuevo[0,5]=77.5698 #longitud
Cliente_Nuevo[0,5]=1 #comentarios

# Realizar la predicción del cluster del nuevo almacen
datos_nuevos = modelo_cargado.predict(Cliente_Nuevo)
print('')
print("El nuevo cliente pertenece al Grupo:", datos_nuevos)
if datos_nuevos == 0:
    print('Pertenece al grupo 0 y recibira un cupon por su primera compra')

if datos_nuevos == 1:
    print('Pertenece al grupo 1 y recibe una de las promociones de nuestros mejores productos')

if datos_nuevos == 2:
    print('Pertenece al grupo 2 y recibe descuento del 5% por su primera compra y cupones en fecha especiales')

El nuevo cliente pertenece al Grupo: [0]
Pertenece al grupo 0 y recibira un cupon por su primera compra
```

(Imagen 12. Implementación del modelo en el grupo 0)

En la imagen 12 al llenar los datos del nuevo cliente el modelo nos verifica a que grupo puede pertenecer este cliente y nos lanza una predicción y en base al grupo asignado se le otorgan los beneficios al cliente.

Lo mismo sucede si ingresamos un cliente que pueda ingresar o pertenecer a otro grupo diferente del grupo 0.

```
# Ingresar datos nuevos del cliente (9 variables)
Cliente_Nuevo=np.zeros((1,9))
Cliente_Nuevo[0,0]=20 #Edad
Cliente_Nuevo[0,1]=2 #Genero
Cliente_Nuevo[0,2]=2 #Estado civil
Cliente_Nuevo[0,3]=2 #Ocupacion
Cliente_Nuevo[0,4]=0 #IngresoMensual
Cliente_Nuevo[0,5]=3 #Tamaño familia
Cliente_Nuevo[0,5]=12.9700 #latitud
Cliente_Nuevo[0,5]=77.5895 #longitud
Cliente_Nuevo[0,5]=1 #comentarios

# Realizar la predicción del cluster del nuevo almacén
datos_nuevos = modelo_cargado.predict(Cliente_Nuevo)
print('')
print("El nuevo cliente pertenece al Grupo:", datos_nuevos)
if datos_nuevos == 0:
    print('Pertenece al grupo 0 y recibirá un cupón por su primera compra')

if datos_nuevos == 1:
    print('Pertenece al grupo 1 y recibe una de las promociones de nuestros mejores productos')

if datos_nuevos == 2:
    print('Pertenece al grupo 2 y recibe descuento del 5% por su primera compra y cupones en fecha especiales')

El nuevo cliente pertenece al Grupo: [1]
Pertenece al grupo 1 y recibe una de las promociones de nuestros mejores productos
```

(Imagen 13. Implementación del modelo en el grupo 1)

En la imagen 13 podemos ver el registro de un nuevo cliente el cual este asignado por el modelo en el grupo 1 contando con los beneficios adecuados para el mismo.

```
# Ingresar datos nuevos del cliente (9 variables)
Cliente_Nuevo=np.zeros((1,9))
Cliente_Nuevo[0,0]=31 #Edad
Cliente_Nuevo[0,1]=1 #Genero
Cliente_Nuevo[0,2]=2 #Estado civil
Cliente_Nuevo[0,3]=3 #Ocupacion
Cliente_Nuevo[0,4]=2 #IngresoMensual
Cliente_Nuevo[0,5]=4 #Tamaño familia
Cliente_Nuevo[0,5]=12.5821 #latitud
Cliente_Nuevo[0,5]=77.3616 #longitud
Cliente_Nuevo[0,5]=1 #comentarios

# Realizar la predicción del cluster del nuevo almacén
datos_nuevos = modelo_cargado.predict(Cliente_Nuevo)
print('')
print("El nuevo cliente pertenece al Grupo:", datos_nuevos)
if datos_nuevos == 0:
    print('Pertenece al grupo 0 y recibirá un cupón por su primera compra')

if datos_nuevos == 1:
    print('Pertenece al grupo 1 y recibe una de las promociones de nuestros mejores productos')

if datos_nuevos == 2:
    print('Pertenece al grupo 2 y recibe descuento del 5% por su primera compra y cupones en fecha especiales')

El nuevo cliente pertenece al Grupo: [2]
Pertenece al grupo 2 y recibe descuento del 5% por su primera compra y cupones en fecha especiales
```

(Imagen 14. Implementación del modelo en el grupo 2)

En la imagen 14 podemos encontrar que el nuevo cliente pertenece según el modelo al grupo 2. Por lo tanto, recibe un beneficio por su primera compra y cupones en fechas especiales.

Con este análisis validamos el funcionamiento del modelo en nuestra base de datos y tomamos decisiones con los resultados establecidos.

Conclusiones

La información presentada anteriormente nos muestra varias conclusiones recibidas del algoritmo implementado de machine learning.

Podemos resaltar las siguientes:

- La base de datos consta de variables muy detalladas de la empresa que labora entregando pedidos de alimentos en línea, variables como la edad y la ubicación geográfica fueron relevantes en la investigación. Se puede concluir que la mayoría de clientes son jóvenes masculinos de 23 años que solicitan el servicio porque se encuentran en una zona de comercio.
- Podemos entender que las personas mayores de 27 años aproximadamente no usan con frecuencia estos servicios ya que la mayoría vive o se encuentra en una zona mas residencial que comercial, pero algo a destacar es que parte de ese grupo son personas o que trabajan independiente o son amas de casa, por lo que sacamos la conclusión que son familias que prefieren cocinar en casa.
- Otro aspecto a destacar es que la empresa labora en un perímetro que contiene comercio, residencias pero también entidades educativas, esto lo decimos porque la mayoría de los jóvenes que solicitan el servicio son estudiantes y no cuentan con ingresos mensuales.

- Usando la herramienta de Google maps podemos obtener también que la zona donde labora esta empresa a parte de ser comercial cuenta con zonas turísticas a sus alrededores, esto lo decimos porque hablamos de una ciudad de la india llamada Bangalore según describen las coordenadas en los mapas.
- Otro aspecto a concluir es que la empresa cuenta con un gran manejo de sus labores. Sin embargo, gracias al método implementado nos damos cuenta por la ubicación geográfica que hay potenciales zonas que pueden aportar y ser mas frecuentes en el servicio con la entidad. Por eso, se tomaron decisiones que benefician al publico menos frecuente para usar el servicio, pero a su vez sin descuidar el público que genera más ganancias dándoles la oportunidad de tener beneficios por su constancia.
- Así también como buscando oportunidades de atraer al publico femenino para acceder a los servicios otorgados con la confianza de que sientan que la entidad piensa en todos sus clientes y demuestra un interés personal satisfaciendo sus necesidades con calidad.
- Las decisiones que se demuestran en este informe dejan en evidencia la calidad del aprendizaje automático y como se logra resaltar las variables otorgadas relacionándolas con las demás para obtener un resultado preciso y satisfactorio.

Gracias a la implementación de estos algoritmos podemos dar una toma decisiones más precisa que toma en cuenta a profundidad el manejo de la entidad y como esta puede mejorar sus procesos a partir de los datos recolectados y estudiados.

Referencias

- "Base de datos". Autor: Equipo editorial, Etecé. De: Argentina. Para: Concepto.de.
Disponible en: <https://concepto.de/base-de-datos/>.
Última edición: 19 de noviembre de 2023. Consultado: 05 de abril de 2024
- Vázquez, A. M. (2018). Introducción a machine learning.
Disponible en el sitio web:
<https://unidad.gdl.cinvestav.mx/doc/investigacion/computacion/Introduccion-Machine-Learning.pdf>
- "Algoritmo en informática". Autor: Equipo editorial, Etecé. De: Argentina.
Para: Concepto.de. Disponible en: <https://concepto.de/algoritmo-en-informatica/>.
Última edición: 21 de febrero de 2024. Consultado: 05 de abril de 2024
- Rouhiainen, L. (2018). Inteligencia artificial. *Madrid: Alienta Editorial*, 20-21.
Disponible en el sitio web:
https://planetadelibrosec0.cdnstatics.com/libros_contenido_extra/40/39308_Inteligencia_artificial.pdf
- "Técnica". Autor: Equipo editorial, Etecé. De: Argentina. Para: Concepto.de.
Disponible en: <https://concepto.de/tecnica/>.
Última edición: 5 de agosto de 2020. Consultado: 05 de abril de 2024
- Imágenes y tablas.
Autor: Fuente Propia. Herramienta: Google Colab. 05 de abril de 2024.
Disponible en:
https://drive.google.com/file/d/118IB_9FclwDfhtMrTfYeqF5nPH5r19O/view?usp=sharing
- Mapas
Autoría Propia. Herramienta: Google Maps. 05 de abril de 2024.
Disponible en:
<https://www.google.com/maps/place/Bangalore,+Karnataka,+India/@13.0238689,77.4786654,11.25z/data=!4m6!3m5!1s0x3bae1670c9b44e6d:0xf8dfc3e8517e4fe0!8m2!3d12.9715987!4d77.5945627!16zL20vMDIjMTc?entry=ttu>