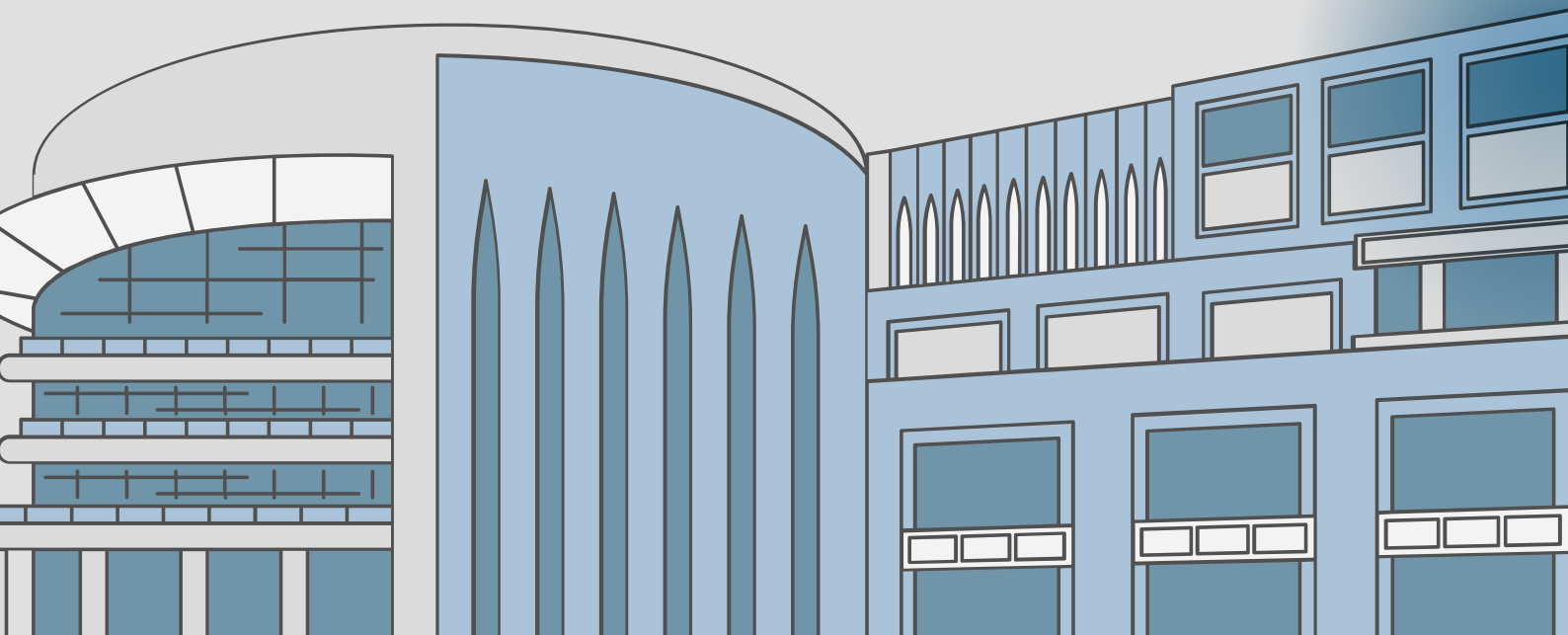


Mall Customers

Análisis de resultados

Prepared by

**Ana M. Correa, Sofía Carolina Muñoz & Maria
José Vallejo**



Data Brief

PROJECT NAME: Customer Generation Project

BASE DE DATOS	<p>La base de datos encontrada en Kaggle con el nombre de “Customer Segmentation”, tiene como finalidad la segmentación de nuevos clientes potenciales que visitan un mall. Esta conformada por distintas variables como ID, Gender (Género), Ever Married (Ha estado casado), Age (Edad), Graduated (Si se ha graduado o no), Profession (Profesión), Work Experience (Experiencia laboral), Family Size (Miembros en la familia), y la Categoría de mall.</p> <p>Adicionalmente se encuentran variables como <i>Spending Score (Puntaje de gastos) y finalmente si pertenece o no al loyalty program de este mall.</i></p>																																							
RESTRUCTURACIÓN DE VARIABLES:	<div><table><tr><td rowspan="2">Gender</td><td>Female</td><td>1</td></tr><tr><td>Male</td><td>0</td></tr><tr><td rowspan="2">Ever_Married</td><td>Yes</td><td>1</td></tr><tr><td>No</td><td>0</td></tr><tr><td rowspan="2">Graduated</td><td>Yes</td><td>1</td></tr><tr><td>No</td><td>0</td></tr></table><p><u>Tomamos la base de datos y cambiamos los yes y no por 0 y 1's.</u></p></div> <div><table><tr><th colspan="3">Mall Categories</th></tr><tr><td>Cat_7</td><td>7</td><td>Highest Level</td></tr><tr><td>Cat_6</td><td>6</td><td></td></tr><tr><td>Cat_5</td><td>5</td><td></td></tr><tr><td>Cat_4</td><td>4</td><td></td></tr><tr><td>Cat_3</td><td>3</td><td></td></tr><tr><td>Cat_2</td><td>2</td><td></td></tr><tr><td>Cat_1</td><td>1</td><td>Lowest Level</td></tr></table></div>	Gender	Female	1	Male	0	Ever_Married	Yes	1	No	0	Graduated	Yes	1	No	0	Mall Categories			Cat_7	7	Highest Level	Cat_6	6		Cat_5	5		Cat_4	4		Cat_3	3		Cat_2	2		Cat_1	1	Lowest Level
Gender	Female		1																																					
	Male	0																																						
Ever_Married	Yes	1																																						
	No	0																																						
Graduated	Yes	1																																						
	No	0																																						
Mall Categories																																								
Cat_7	7	Highest Level																																						
Cat_6	6																																							
Cat_5	5																																							
Cat_4	4																																							
Cat_3	3																																							
Cat_2	2																																							
Cat_1	1	Lowest Level																																						
VARIABLES DEPENDIENTES	<div><div><h3>KNN Model</h3><div><div>Loyalty Program</div><div>Programa de lealtad del centro comercial para participar por premios y obtener descuentos.</div><div>Yes (1) // No (0)</div></div></div><div><div>Naive Bayes Model</div><div><div>Mall Category</div><div>Categorías del mall para compradores y clientes, se sube de nivel considerando el gasto promedio en el mall.</div><div>Categorías: 1-2-3-4-5-6-7</div></div></div><div><div>Spenditure</div><div>Muestra los gastos promedio en el mall de los individuos.</div></div><div><h3>Madeline Model</h3></div></div>																																							
VARIABLES INDEPENDIENTES	<div><ul style="list-style-type: none">• Gender: Género del cliente• Ever_Married: Estado marital del cliente• Age: Edad del cliente• Graduated: Si el cliente es graduado o no• Work_Experience: Experiencia laboral del cliente en años</div> <div><ul style="list-style-type: none">• Spenditure: Gastos promedio del cliente en el mall• Spending_Score: Puntaje de gastos del cliente• Family_Size: Número de miembros de la familia del cliente</div>																																							

Hallazgos

- * Al momento de llevar a cabo la construcción del modelo de agrupación KNN a un total de 2 clusters, para el pronóstico de lealtad al programa del mall para los nuevos clientes. Se determinó que las dos variables que tienen la mayor correlación frente al pronóstico de lealtad al programa son: el *spenditure* (0.7548), el *spenditure_Score* (0.6819) y el *Mall_Category* (0.9268), es decir que tienen un impacto sobre la variable de pronóstico. En cuanto a la precisión del modelo, se obtuvo que es del 99,25%.
- * De acuerdo con las métricas de sensibilidad y especificidad, el modelo KNN frente al pronóstico de los valores de referencia (no pertenece al programa de Lealtad (0) – Pertenece al programa de Lealtad (1)) obtuvieron un porcentaje de comportamiento del 98,34% y 99,66% respectivamente.
- * Siguiendo el modelo KNN logramos determinar que una persona con las siguientes variables socioeconómicas: hombre(0), casado (1), edad 45 años, graduado (1), experiencia laboral de 7 años, gastos de 4.000, con un puntaje de gasto de 4.5, con 4 miembros en su familia y perteneciente a la categoría del mall número 4, como resultado se obtuvo que este cliente no pertenece al programa de lealtad del mall (0).
- * Los porcentajes de pertenencia y no pertenencia para el programa de lealtad según lo pronosticado correctamente por el modelo son: el 30% (615/2088) de los clientes del mall no pertenecen al programa de lealtad mientras que el 70% (1473/2088) de los clientes pertenecen al programa de lealtad.

propósito DEL MODELO

Este modelo nos permite saber si un individuo pertenece o no al programa de lealtad del centro comercial.

El modelo #1 implementado con la base de datos de Customer Segmentation se basa en la variable dependiente **Loyalty Program**, el cual permite pronosticar **si el cliente pertenece al programa de lealtad del mall el cual ofrece descuentos, sorteos y beneficios.**



KNN
MODELO #1

propósito DEL MODELO

Este modelo nos permite saber a cual categoría pertenece un cliente del mall.



El modelo #2 implementado con la base de datos de Customer Segmentation se basa la variable dependiente **Mall_Category**, el cual indica a cual **categoría o segmento del mall pertenece el cliente de acuerdo con sus variables socioeconómicas** tales como: edad, género, casado o soltero, graduado o no, gastos promedio en el mall y miembros de su familia.



NAIVE BAYES

MODELO #2

Hallazgos



De acuerdo el análisis, la categoría del mall que presento la mayor cantidad de clientes fue la categoría 6 con 1421 clientes. De esta misma manera, la categoría del mall que tiene la menor cantidad de clientes, es la categoría 5 con 23 clientes. Esto nos permite concluir que los clientes que visitan el mall poseen variables socioeconómicas que favorecen su categoría de compra.



Al momento de analizar las categorías con las variables socioeconómicas más altas, se identificó que: la categoría del mall que tiene la edad más alta es la categoría 6 con 46 años de edad, en cuanto a la experiencia laboral, la categoría 7 posee el nivel más alto con 2,8 años. De manera similar, la categoría con el gasto promedio más alto es la número 6 con 6,475.



Posteriormente, se analizó las categorías con las variables socioeconómicas más bajas y se identificó que: la categoría del mall que posee la menor edad es la categoría 2 con 36 años. En cuanto a experiencia labora, la categoría con los menores años es la número 4 con 2,18 años. Finalmente la categoría con el gasto promedio más bajo es la número 7 con 1,168.



De acuerdo con la matriz de confusión, la categoría que tuvo la mayor cantidad de aciertos fue la categoría 6 con 1,412 aciertos. La exactitud de este modelo fue de 99,48%, esto significa que el modelo tuvo un comportamiento general muy bueno, y el coeficiente de especificidad fue de 99,53%. Finalmente, la categoría del mall a la que pertenece el individuo que pose los siguientes valores socioeconómicos: hombre(0), casado (1), edad 45 años, graduado (1), experiencia laboral de 7 años, gastos de 4.000 y con 3 miembros en su familia pertenece a la categoría del mall número 4.

propósito DEL MODELO

Este modelo nos permite saber los gastos promedio de un cliente en el mall.



El modelo #3 implementado con la base de datos de Customer Segmentation se basa en la variable dependiente **Spenditure**, la cual permite identificar **los gastos promedio de los clientes del mall de acuerdo con sus variables socioeconómicas** tales como: género, edad, si está casado o soltero, graduado o no, experiencia laboral, categoría o segmento del mall al que pertenece el cliente y miembros de su familia.



MADALINE MODELO #3

Hallazgos



Teniendo en cuenta la base de datos **Customer Segmentation** se utilizó el modelo MADALINE que nos permite conocer los gastos promedio de un cliente en el mall de acuerdo con las siguientes variables socioeconómicas: **Gender, Ever_Married, Age, Graduated, Work Experience, Family_Size y Mall_Category**, obteniendo una variable de salida **Spenditure**.



Posteriormente, se ejecutó el modelo MADALINE, teniendo en cuenta las variables anteriores y los métodos de normalización; por el máximo para las variables de entrada y el método de centralización para la variable de salida. Se ejecutó el modelo en 5 oportunidades para un total de 250 iteraciones, tomando para cada iteración un batch del 10% del total de los datos, es decir, 614. Luego de ejecutarlo 5 veces, se identificó que tuvo un error cuadrático de **0.0170**. De igual manera, al analizar las 5 ejecuciones se obtuvo un promedio del **coeficiente de correlación de 0,796%**, para las **discrepancias de las medias y las dispersiones**, se obtuvo un promedio de **-2,707% y 31,15%** respectivamente.



Por otro lado, con el objetivo de obtener un promedio de los efectos independientes, se ejecutó el modelo considerando 250 iteraciones, para el cual se obtuvieron los siguientes resultados para las 7 variables: Gender (-0,47), Ever_Married (0,38), Age (-2,67), Graduated (-0,34), Work_Experience (-1,13), Family_Size (-1,20) y Mall_Category (2,75), lo que nos permite concluir que el efecto independiente que tuvo un mayor impacto fue **Mall_Category** y el que tuvo un menor impacto fue **Age**.



Finalmente, se calculó el gasto promedio de un cliente en el mall que tenga las siguientes variables socioeconómicas: Género: Mujer, Casado, Edad: 48 años, Graduado, Experiencia laboral: 10 años, Número de miembros de la familia: 5 y Categoría a la que pertenece el cliente: 6. Luego de ejecutar el modelo MADALINE mediante el método de normalización de máximo, se obtuvo que el cliente con las variables socioeconómicas anteriores tiene un gasto promedio en el mall de **6.659 dólares**.

HALLAZGOS GENERALES

KNN Model

La sensibilidad es: 98,04%
La especificidad es: 99,53%

Naive Bayes Model

La sensibilidad es: 99,07%
La especificidad es: 99,13%

MADALINE Model

	MADALINE		
Iteraciones	Correlación	Discrepancia de la Medias	Discrepancia de las dispersiones
1	0,8833	-0,759	19,37
2	0,7838	-3,058	33,46
3	0,7841	-2,895	33,04
4	0,7510	-3,298	36,78
5	0,7768	-3,523	33,09
Promedio	0,796	-2,707	31,15

MADALINE						
Efectos Independientes						
	1	2	3	4	5	PROMEDIO
GENDER	-0,04	-0,16	-0,16	-1,83	-0,16	-0,47
EVER_MARRIED	0,01	0,04	0,05	1,73	0,05	0,38
AGE	-0,22	-1,06	-1,09	-9,86	-1,11	-2,67
GRADUATED	-0,04	-0,11	-0,11	-1,31	-0,11	-0,34
WORK_EXPERIENCE	-0,07	-0,41	-0,41	-4,32	-0,42	-1,13
FAMILY_SIZE	-0,26	-1,39	-1,38	-1,52	-1,43	-1,20
MALL_CATEGORY	3,69	2,58	2,61	2,31	2,57	2,75