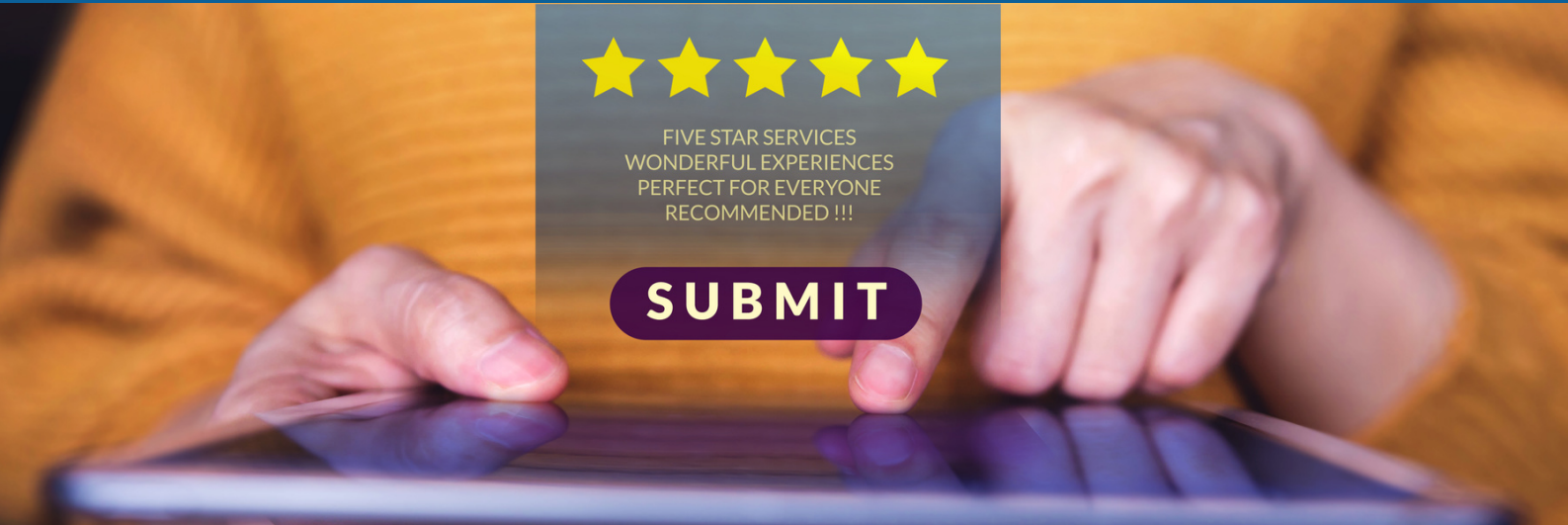


# RETO – SUPERMERCADO

CIENCIA DE DATOS | ANDREA ESCOBAR



Una cadena de supermercados en Ecuador tiene como componente importante de su modelo de negocio la componente de servicio. Es decir, el modelo de gestión de servicios esta soportado por la satisfacción, recomendación y lealtad de sus clientes. Un cliente insatisfecho es altamente probable que no recomiende a familiares y amigos y que pierda fidelidad con el supermercado; para la empresa le cuesta 4 veces captar un nuevo cliente que gestionar adecuadamente la satisfacción de sus clientes por el servicio de atención. Actualmente esta cadena de supermercados tiene presencia en las principales ciudades de Ecuador como Quito, Guayaquil, Machala y Ambato.

El gerente de Marketing de la empresa mediante estudios de investigación de mercados observó que el nivel de satisfacción en 2021 comenzó a disminuir. En enero del 2021 tenía un indicador de satisfacción del 83% y cerró ese año con un 17% de satisfacción. Como parte de su gestión hizo visita de campo a algunas tiendas de cada ciudad y encontró cosas en común en las tiendas:

- Se formaban largas colas en la sección de cajas
- Publicidad desactualizada
- Algunas tiendas estaban medias sucias
- Los productos no estaban ordenados





El Gerente de Marketing se reunió con el Gerente Comercial de la empresa para definir una estrategia que frene la caída en el indicador de satisfacción, definiéndose lo siguiente: “hay que brindarle al cliente lo que quiere”. Bajo este escenario la compañía adoptó la siguiente estrategia en junio del 2021:

- Redujo el precio de sus productos de mayor demanda, sacrificando el margen en un 10%.
- Realizó ofertas promocionales de venta cruzada.
- Mejoro el tiempo de ejecución de sistemas internos con aras de mejorar el tiempo de atención en la caja.

Sin embargo, la satisfacción siguió disminuyendo hasta llegar a cierre de año del 2021 al 17%. Para el 2012 tomo la siguiente definición:

- a.Hacer un estudio de investigación de mercados, para ello contrata a OpenLab, el cual lleva a cabo la investigación.
- b.Contratar a Ciencia de Datos Ecuador para que identifique los drivers e insight que impacten a la satisfacción de sus clientes durante el proceso de atención.

Supongan que ustedes son el equipo que Ciencia de Datos Ecuador ha designado para resolver este reto. A partir de la base de datos de la investigación y lo descrito líneas arriba, resuelva lo siguiente:

- 1.Defina claramente los objetivos de negocio y objetivos analíticos
- 2.Establezca la hipótesis que desea resolver usando el negocio descrito y los data que hay en la BD
- 3.Identifique los drivers que impactan a la satisfacción de los clientes y sus pesos de impactos.
- 4.Proponga estrategias y recomendaciones en función a lo desarrollado en el punto 3 y a los insight obtenidos de los datos



# Process

1. Lectura de la base de datos
2. Entender el negocio
3. Identificación de faltantes y duplicados
4. Identificación de outliers (valores fuera de rangos inter cuartiles):  
<https://www.aprendemachinelearning.com/deteccion-de-outliers-en-python-anomalia/> )
5. ¿Qué pregunta busco responder con mi modelo?
6. Agrupar la variable "p8" en :(preguntas de satisfacción {9,10} y otros) para construir la variable dependiente.
7. Balancear los datos de ser necesario (reducir clase positiva o aumentar clase negativa)
8. Analiza la correlación de las variables (correlación no implica causalidad pero se puede hacer un test para eliminar variables altamente correlacionadas)  
<https://www.cienciadedatos.net/documentos/pystats05-correlacion-lineal-python>
9. Aplica al menos un modelo de clasificación
10. Evalúa con la métrica correcta (MUY IMPORTANTE si es balanceado o desbalanceado)
11. Haz un plot de evaluación (Matriz de confusión, AUC-ROC o curva índice F1 depende de las clase )
12. Sobre el modelo de clasificación entrenado evalúa la importancia de las variables predictoras (Tree based :  
<https://towardsdatascience.com/explaining-feature-importance-by-example-of-a-random-forest-d9166011959e> o SHAP:  
<https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137>)
13. Realizar una presentación de máx. 3 slides con la explicación EJECUTIVA de este modelo