

# ¿Cómo predecir la satisfacción de pasajeros de una aerolínea?

...

Profesor: Jorge Ruiz

Tutor: Joaquín Armesto

Alumno: Javier Parigini

Enero 2024

# Contexto comercial

La industria aeronáutica tiene un contexto cada vez más competitivo, aún más luego de la pandemia. **La satisfacción del cliente respecto al vuelo es clave** a la hora de fidelizarlos y que vuelvan a elegir a la empresa.

La empresa ha recopilado detalles de un número grande de pasajeros respecto al vuelo que tomaron y características propias de las personas y una valoración de varios aspectos junto con el grado final de satisfacción de estas.

La empresa me ha contratado para construir un modelo analítico para anticipar si un usuario va estar satisfecho o no con su vuelo y cuáles son las variables más relevantes a tener en cuenta para asegurar un buen resultado.

# Objetivo

Predecir qué usuario va a estar satisfecho y qué usuario va a estar neutral/insatisfecho utilizando los parámetros definidos y provistos por la empresa

## Contexto analítico

Se me proporciona un archivo CSV con datos de los usuarios, datos del vuelo, valoraciones de los usuarios de aspectos del vuelo y clasificación de grado de satisfacción final de los usuarios (satisfecho o neutral/insatisfecho)

Con estos debo:

Leer, transformar y preparar datos para su visualización

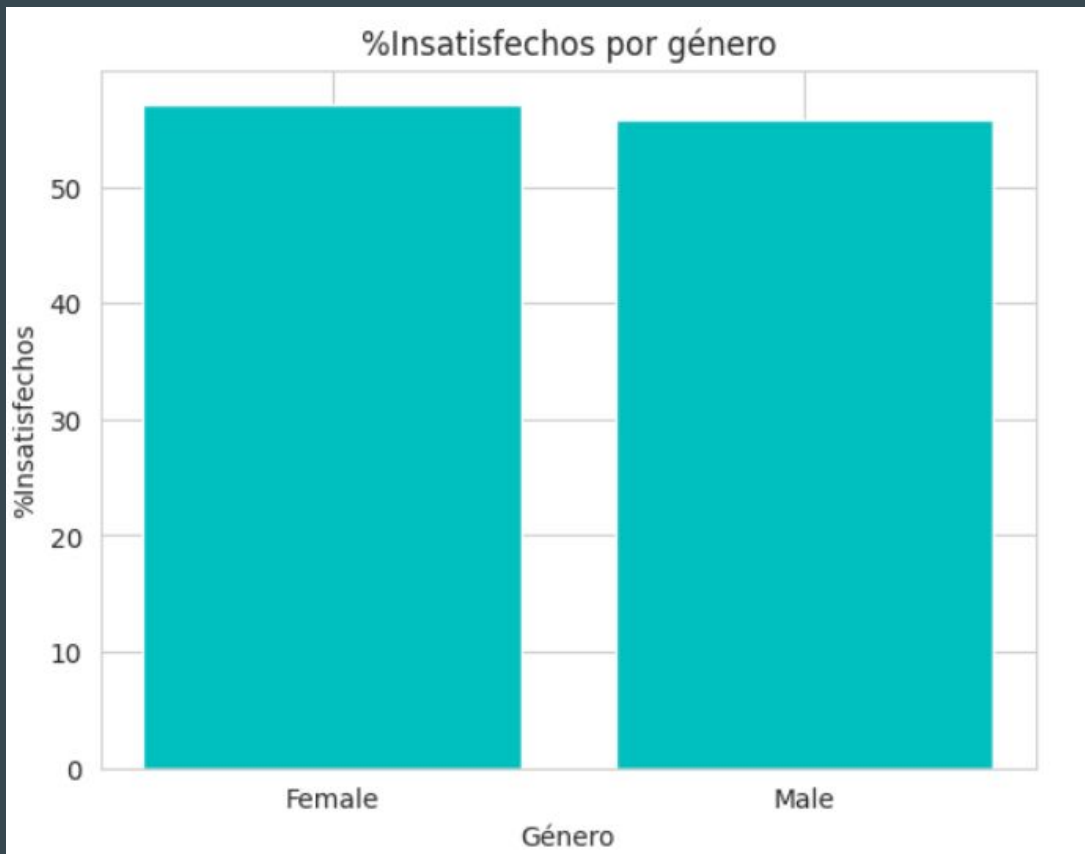
Realizar análisis y construir visualizaciones de los datos para identificar patrones en el conjunto de datos.

# Tareas realizadas y principales hallazgos

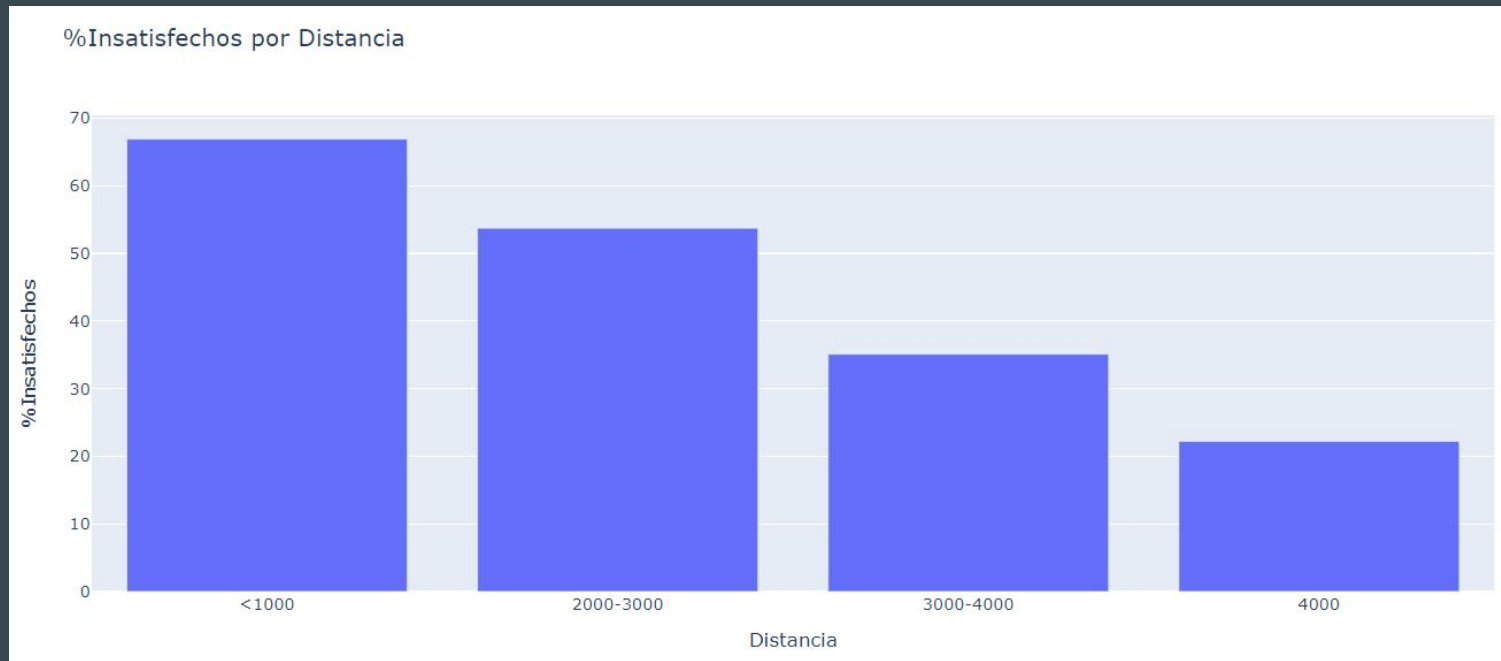
- Carga de Dataset
- Limpieza de datos
  - Pocos datos nulos concentrados en una sola columna. Se procede a eliminarlos
  - Eliminación de outliers que representan en el peor de los casos alrededor del 2% de los datos de la columna
- Análisis exploratorio de datos
- Modelado
  - Encoding
  - Definición de X e Y
  - Iteraciones 1,2 3
- Conclusiones

## EDA - ¿Hay algún grupo de personas que tienda más a la insatisfacción?

Tanto hombres como mujeres poseen un porcentaje similar de insatisfechos

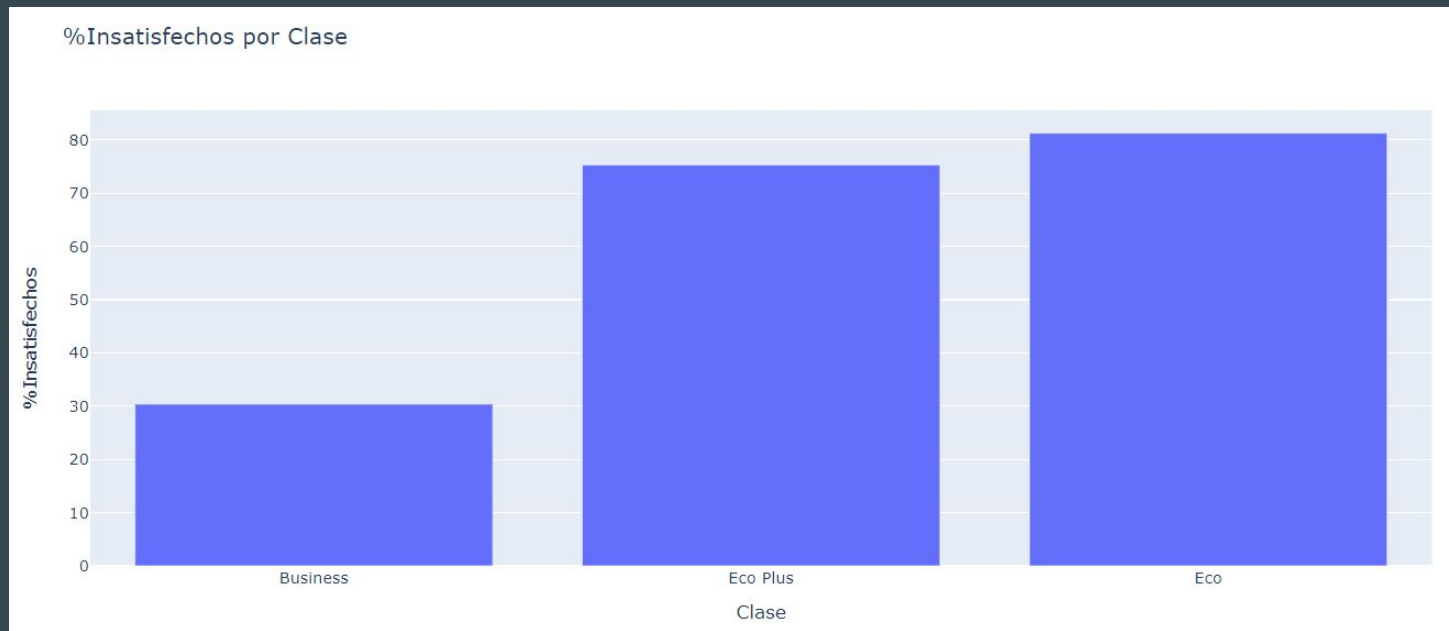


## EDA - ¿Cuánto influye la distancia de vuelo en la insatisfacción?



Contrario a lo esperado, la insatisfacción cae a medida que los vuelos son más largos.

# EDA - ¿Cuánto afecta la clase en la satisfacción?



Como era esperable el % de insatisfechos en clase business es bastante menor a las de clases Economy. Sin embargo, entre Eco y Eco Plus el % de insatisfechos es bastante similar y sorprende que sea tan alto

# Modelado - primera iteración

Se probaron con modelos y se eligieron los dos con accuracy más alta

Modelo	Accuracy
XG Boost	0,9618
Random Forest	0,9617
Árbol de decisión	0,9459
Regresión logística	0,8753
KNN	0,7392



# Modelado - segunda iteración

Se usó Halving Grid search para encontrar los mejores hiper parámetros de cada modelo  
XG Boost mejoró su accuracy  
En el caso de RF, no se logró una mejor accuracy

Modelo	Accuracy
XG Boost	0,963
Random Forest	0,957

---

# Modelado - tercera iteración

Se usó Grid Search sobre XGBoost y se determinaron los mejores parámetros

Modelo	Accuracy
XG Boost	0,9632

Parámetro	Valor
alpha	0,3
gamma	0,5
max_depth	10
n estimators	100

# Conclusiones

Se creó un modelo basado en el algoritmo XG Boost para determinar si un pasajero va a estar “satisfecho” o “neutral/insatisfecho” en base a variables cómo clase, edad, distancia de vuelo, con **una accuracy de 0,9632**.