




PROYECTO **SATISFACCIÓN** **EN AEROLINEA**

Proyecto Data Science

AUTORES: JHON RODRIGUEZ
GASTÓN FIGUEREDO

- 
- 01** INTRODUCCIÓN
 - 02** HIPOTESIS/ PREGUNTAS
 - 03** METADATOS
 - 04** EDA
 - 05** MODELAMIENTO
 - 06** RECOMENDACIONES



INTRODUCCIÓN

En el siguiente proyecto presentamos el dataset "Flights" con 23 variables y analiza mas de 100.000 viajeros aereos. Cuenta con variables de identificacion de los mismos como: el genero, que tipo de consumidor es (si cuenta con membresia Loyal o Disloyal), edad, motivo de viaje (viaje personal o por trabajo), el tipo de clase en la cual viajo (Eco, Eco Plus o Business) y la distancia del vuelo.

Objetivo: determinar el patrón/variable que conlleva a los usuarios de la aerolínea a estar o no satisfechos con el servicio recibido.

Audiencia: Aerolineas, analistas con foco en crecimiento de la experiencia del usuario, etc.

HIPOTESIS / PREGUNTAS

Se intentara responder las siguientes hipotesis/preguntas:

1

¿EL GÉNERO ES UNA VARIABLE QUE INFLUYE EN LA SATISFACCIÓN DEL CLIENTE?

3

¿LAS HORAS DE VUELO TIENEN INCIDENCIA EN LA SATISFACCIÓN DEL CLIENTE?

2

¿LA CLASE EN LA QUE VUELEN SERÁ UN FACTOR IMPORTANTE EN LA SATISFACCIÓN, A MEJOR CLASE MAYOR SATISFACCIÓN?

4

- EXISTE UN SERVICIO QUE DETERMINE DIRECTAMENTE LA SATISFACCIÓN O NO DEL CLIENTE?

METADATOS

Se usa dataset ubicado en Kaggle.com, se conecta al mismo por medio de una API.

Consta de 23 columnas y mas de 100.000 filas.

El dataset brinda información sobre el usuario de la aerolinea como tambien de una encuesta de satisfaccion aplicada a él.



METADATOS

Encuesta de satisfacción con puntajes del 0 al 5 en donde el viajero evalúa los siguientes aspectos:

- Servicio de Wifi abordo
- Conveniencia del tiempo de arribo o partida del vuelo
- Facilidad al reservar de forma Online
- Ubicación de la puerta del vuelo
- Comida y bebida
- Abordaje del vuelo
- Comodidad del asiento
- Entretenimiento dentro del vuelo
- Servicio de abordaje al vuelo
- Espacio entre asientos (leg room)
- Manejo de equipaje
- Servicio de Check-in
- Servicio dentro del vuelo
- Limpieza

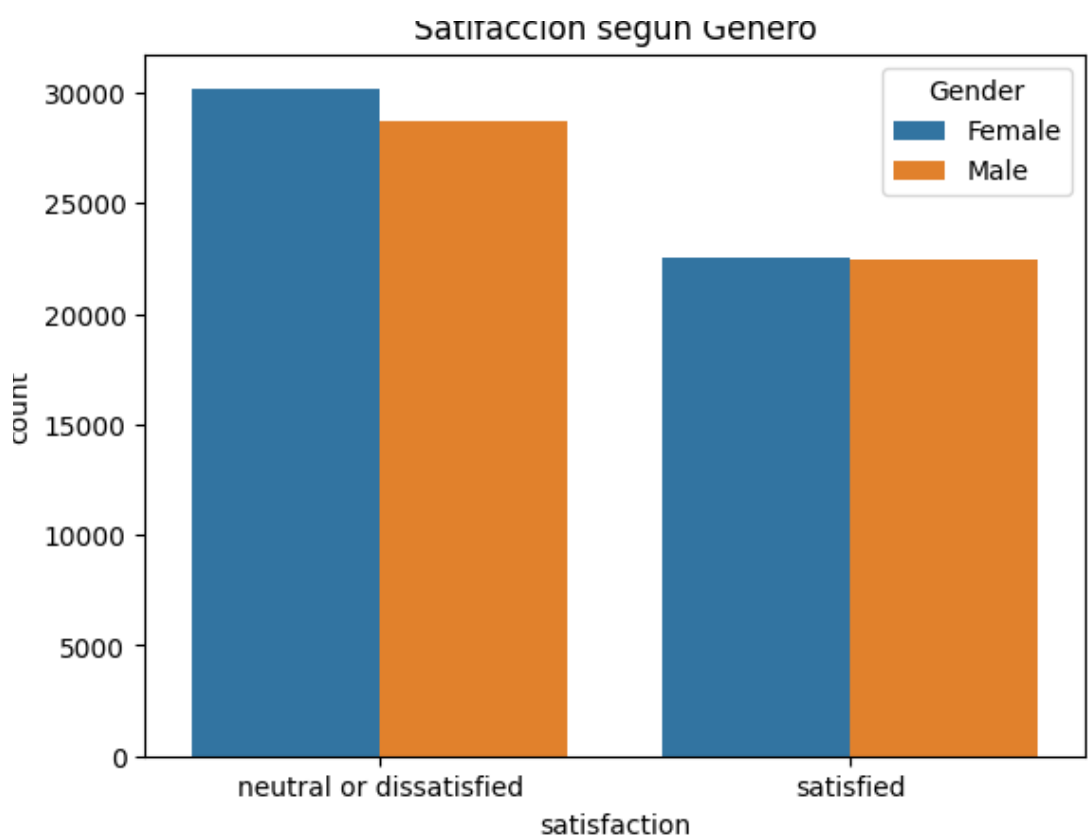


EXPLORATORY DATA ANALYSIS

1

IMPPLICANCIAS SOBRE LA VARIABLE GENERO

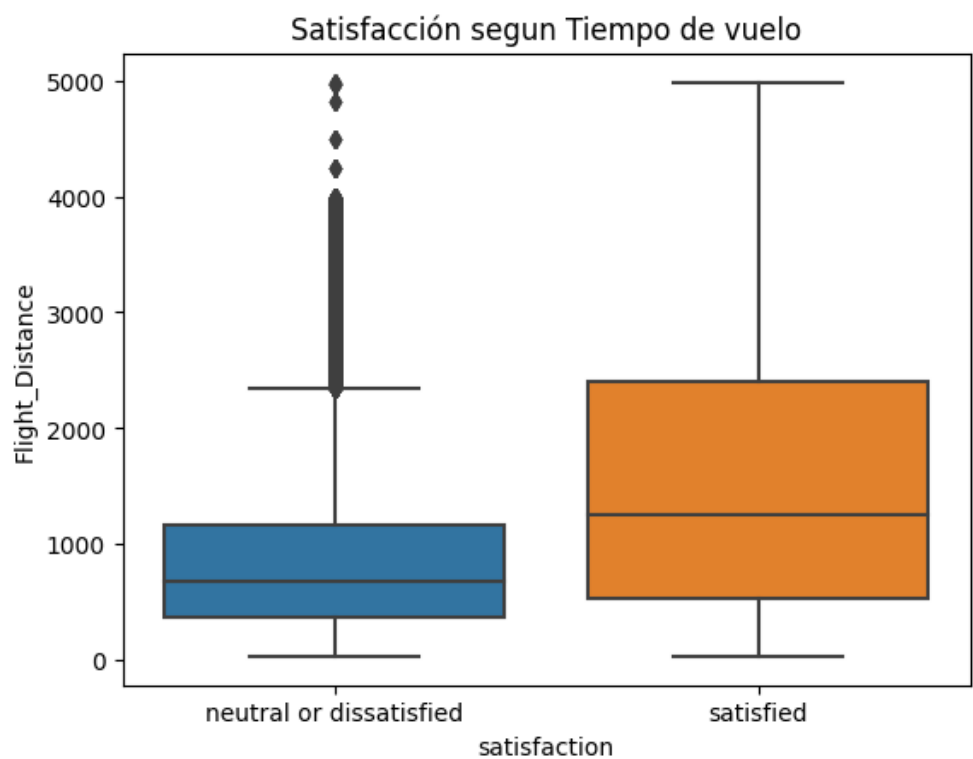
- La distribucion de la variable target "satisfaction" esta balanceado (Insatisfechos o Neutro 57%, Satisfechos 43%)
- La distribucion de la variable "Gender" esta balanceada (51% Hombres, 49% mujeres)
- No se observa incidencia del sexo sobre satisfaccion del cliente.



2

TIEMPO DE VUELO

Se observa que a mayor tiempo de vuelo, el usuario suele estar mas satisfecho con el servicio.

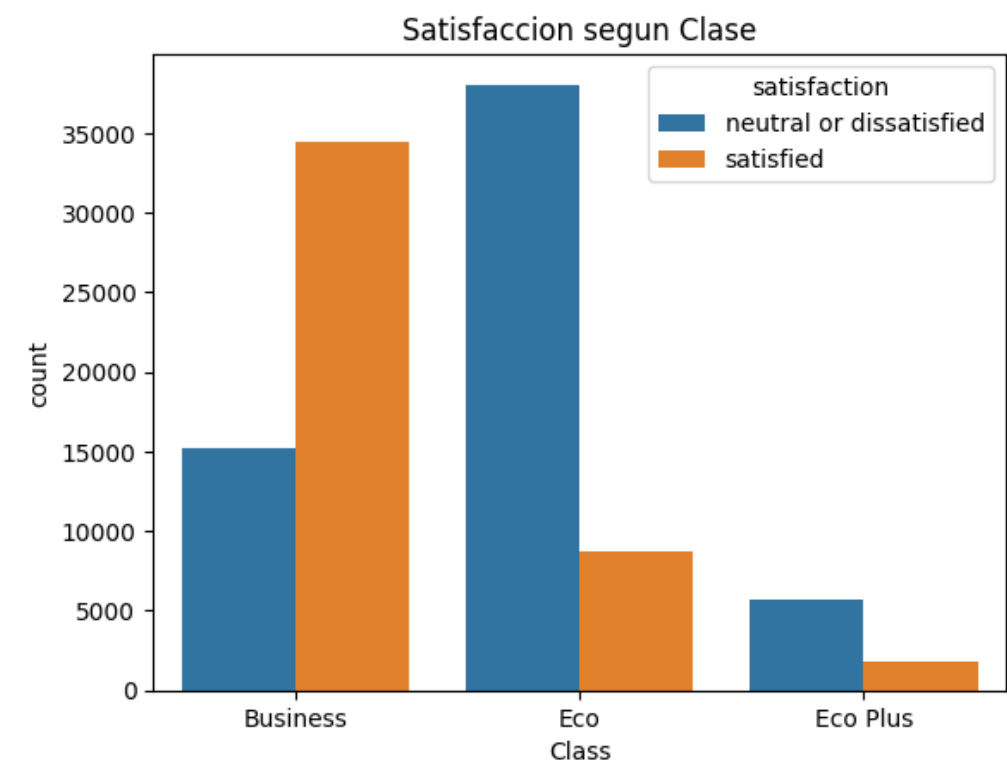


EXPLORATORY DATA ANALYSIS

3

TIPO DE CLIENTE

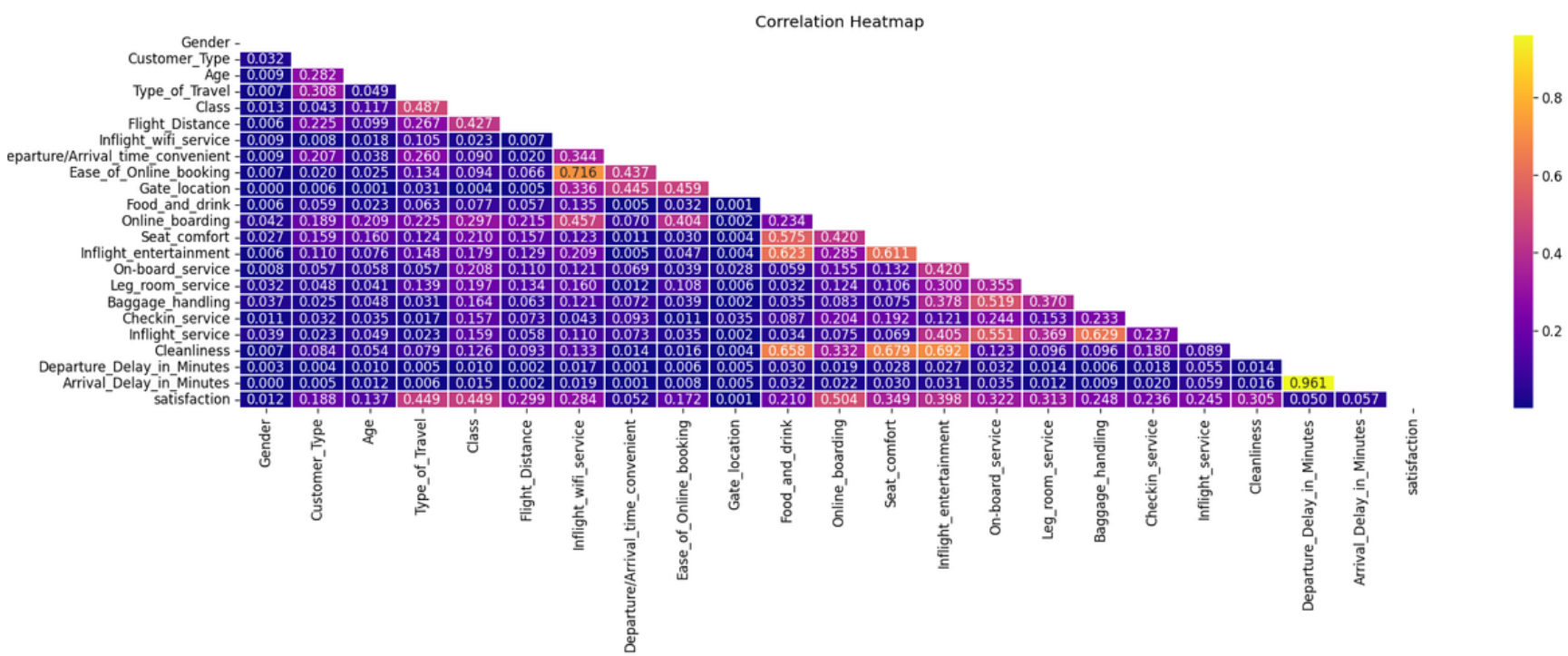
Se observa una clara incidencia de la variable "Class" sobre la satisfaccion del usuario. Usuarios que viajan en clase "Business" suelen estar mayormente satisfechos, por el contrario los usuarios que viajen en clase "Eco class" suelen estar mayormente insatisfechos o neutro.



4

CORRELACION

No se observa correlacion de la satisfaccion o no de algun servicio con respecto a la satisfacion total del viaje



MODELAMIENTO ETAPAS

1

Gestion de valores faltantes

De los 310 datos nulos de la variable "Arrival Delay in Minutes" se reemplazaran por su media correspondiente.

2

Selección de variables

Se eliminan variables de poca correlacion:

- Gender
- Departure/Arrival_time_convenient
- Gate_location
- Departure_Delay_in_Minutes
- Arrival_Delay_in_Minutes

3

Modelado

RFC

- Validación simple
- K-Fold
- Mejora Hiperparametros

SVM

- Validación simple
- K-Fold
- Mejora Hiperparametros

KNN

- Validación simple
- K-Fold
- Mejora Hiperparametros

METRICAS

RFC

- **Acurracy:** 0.9613
- **Acurracy after tuning:** 0.9595
- **Accuracy K-fold:** 0.9619
- **Bias:** 0.0014

KNN

- **Acurracy:** 0.9315
- **Acurracy after tuning:** 0.9303
- **Accuracy K-fold:** 0.9321
- **Bias:** 0.0046

SVM

- **Acurracy:** 0.9539
- **Acurracy after tuning:** 0.9508
- **Accuracy K-fold:** 0.9534
- **Bias:** 0.0021

*Varianzas estadísticamente despreciables (<0.0001)

CONCLUSIONES

- Los modelos aplicados presentan un buen performance, no se observan gran diferencias entre ellos. Por lo cual si tenemos que tomar una decision entre la eleccion de modelos, debemos ver todas las metricas como un conjunto y no tomar una decision basada en una unica variable de medida.
- Se corre el riesgo que con la alta exactitud, modelos insesgados y varianzas despreciables (bajo error) que presentaron los modelos, tiendan a presentar overfitting.



GRACIAS