

Satisfacción en Aerolínea

¿Qué variable/patrón determina la satisfacción del usuario?

Autor: Jhon R. / Gaston F.



AGENDA

- **01** | Contexto/Hipótesis/Audiencia
- **02** | Metadata
- **03** | Análisis Exploratorio
- **04** | Insights y Recomendaciones



CONTEXTO Y AUDIENCIA

Contexto:

El objetivo de este análisis es determinar el patrón/variable que conlleva a los usuarios de la aerolínea a estar o no satisfechos con el servicio recibido.

El dataset "Flights" cuenta con 23 variables y analiza más de 100.000 viajeros aéreos. Cuenta con variables de identificación de los mismos como: el género, que tipo de consumidor es (si cuenta con membresía Loyal o Disloyal), edad, motivo de viaje (viaje personal o por trabajo), el tipo de clase en la cual viajó (Eco, Eco Plus o Business) y la distancia del vuelo. El mismo dataset también cuenta con una encuesta de satisfacción con puntajes del 0 al 5 (14 variables)

Audiencia : Aerolinea

PREGUNTAS DE INTERÉS

Hipótesis/preguntas:

- - ¿El género es una variable que influye en la satisfacción del cliente?
- - ¿Las horas de vuelo tienen incidencia en la satisfacción del cliente?
- - La clase en la que vuelen será un factor importante en la satisfacción, a mejor clase mayor satisfacción?
- - Existe un servicio que determine directamente la satisfacción o no del cliente?

ANÁLISIS DESCRIPTIVO Y EXPLORATORIO

Análisis Univariado

Piechart de las variables target ('satisfaction') y Genero

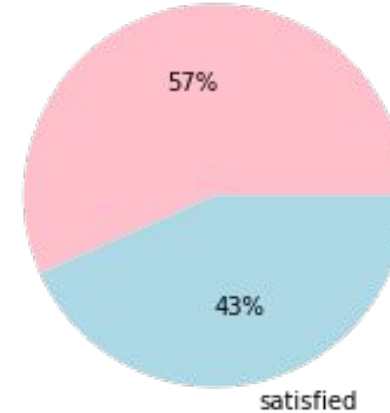
La distribución de la variable target "satisfaction" está balanceado (Insatisfechos o Neutro 57%, Satisfechos 43%)

La distribución de la variable "Gender" esta balanceada (51% Hombres, 49% mujeres)

V1: Piechart, satisfaction y genero.

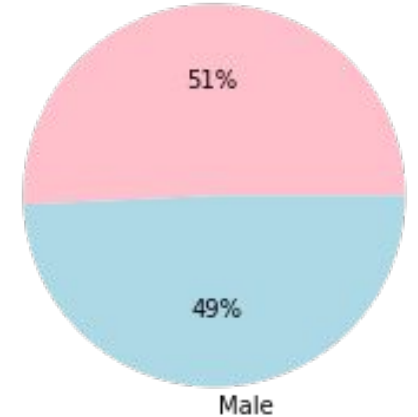
Distribucción - Satisfaction

neutral or dissatisfied



Distribucción - Gender

Female



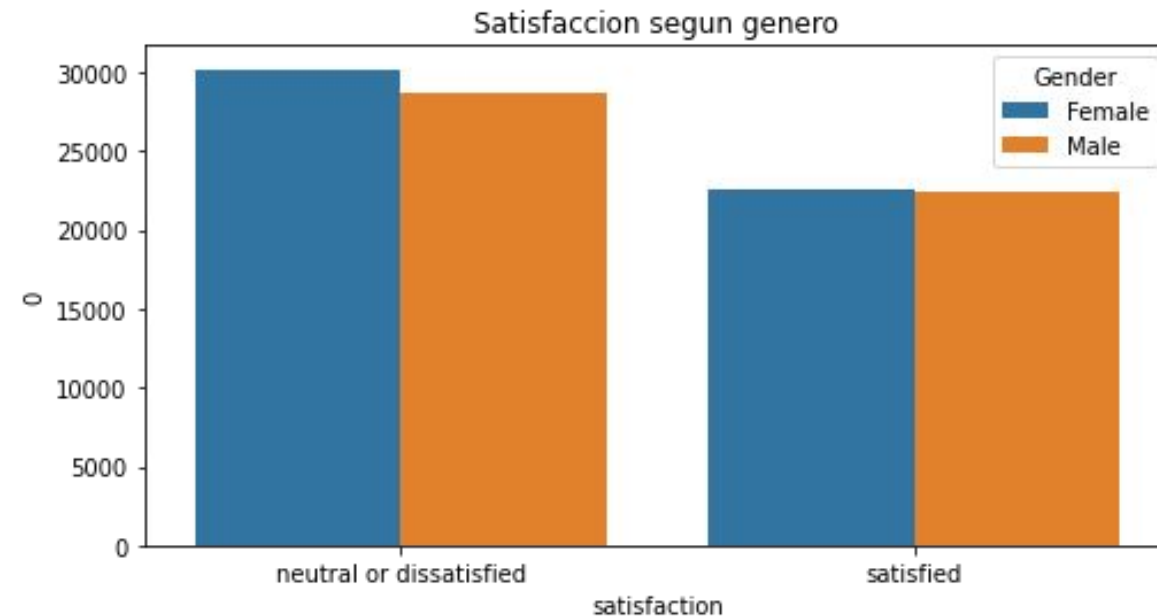
Análisis Bivariado

Barplot, se busca determinar si el género tiene incidencia sobre la satisfacción del usuario

No se observa incidencia del sexo sobre satisfacción del cliente.

Lo cual en base a V1 y V2 los datos no representa una diferencia con respecto al género y la satisfacción del vuelo

V2: Barplot, satisfaction según género.

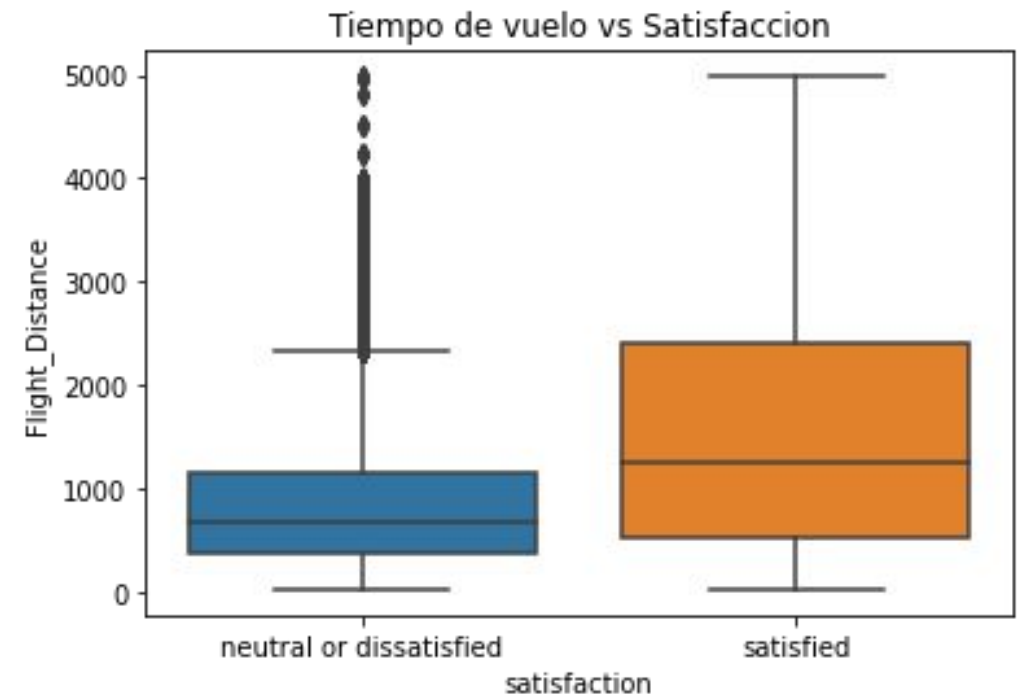


Análisis Bivariado

Boxplot: Se busca determinar si el tiempo de vuelo tienen incidencia en la satisfacción del usuario

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

V3: Boxplot, tiempo de vuelo y satisfacción

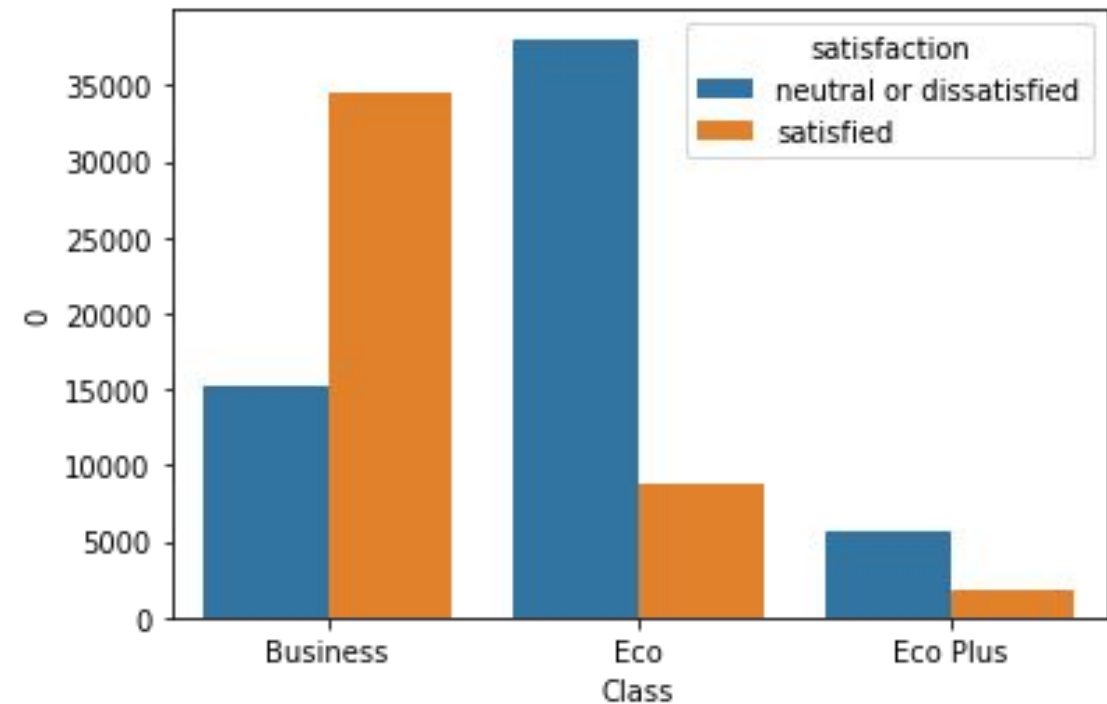


Análisis Bivariado

Barplot: Se busca determinar si la variable Class tiene incidencia sobre la satisfacción del usuario

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

V4: Boxplot, clase y satisfacción



MODELO MACHINE LEARNING

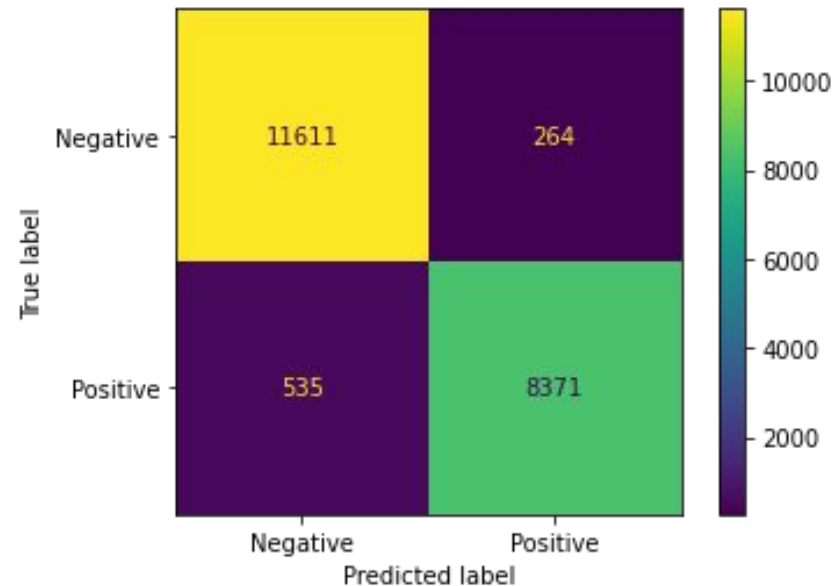
Modelo: Algoritmo Clasificación (RandomForest Classifier)

Procedimiento:

- **Generamos un dataframe con variables de tipo categóricas y luego mediante el uso del módulo LabelEncoder, convertimos las variables categóricas a numericas**
- **Dividimos el dataframe en dos, Variable Target 'y' (Serie) y variables explicativas 'X' (Dataframe)**
- **Mediante el uso del método Backward Elimination, realizamos feature selection (disminuir dimensionalidad del df)**
- **Aun manteniendo un nivel de significación bajo, el modelo no rechaza estadísticamente ninguna de las variables por lo que mantuvimos un modelo de 22 variables explicativas para explicar el nivel de satisfacción del viajero. Aunque no negamos tener problemas de multicolinealidad entre las variables por la homogeneidad de las mismas.**
- **Vamos a separar los conjuntos X,y en 4 subconjuntos: X_train, X_test, y_train, y_test**

Modelo: Algoritmo Clasificación (RandomForest Classifier)

El algoritmo explica correctamente 11.614 verdaderos negativos y 8.378 verdaderos positivos. Con solo 789 falsos negativos y positivos respectivamente.



El modelo predice correctamente el 96% de los casos de estudio, por lo cual es un muy buen modelo para predecir en base a los atributos de los coeficientes la satisfacción o falta de ella del viaje en avión. Nuestras futuras intenciones son en base a clustering caracterizar a los viajeros para determinar si, en base a las decisiones que tomen y características innatas de los mismos, si se va a sentir conforme o no con el viaje ofrecido.