

## Trabajo Práctico Big Data

### Análisis y Modelado de Datos con Orange Data Mining

#### **Dataset seleccionado:**

Airline Passenger Satisfaction

<https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

#### **Informe**

##### **Introducción**

En la industria aérea, la satisfacción del cliente es un indicador esencial que afecta directamente la lealtad de los pasajeros y la reputación de la aerolínea. Este proyecto utiliza Big Data para identificar y analizar los factores que influyen en la satisfacción de los pasajeros, con el fin de ayudar a las aerolíneas a mejorar sus servicios y retener clientes.

##### Problema Específico de Negocio:

Queremos predecir la satisfacción del pasajero (satisfecho o no satisfecho) en función de las características del vuelo y del servicio recibido. Esto nos ayudará a entender mejor qué factores influyen en la experiencia del cliente y a identificar áreas de mejora clave para aumentar la satisfacción.

##### Objetivo:

El objetivo de este proyecto es analizar y modelar los datos de satisfacción de pasajeros para identificar los factores críticos que influyen en la experiencia del cliente en vuelos aéreos. A través de esta investigación, se busca ofrecer recomendaciones que las aerolíneas puedan implementar para optimizar sus servicios y aumentar la retención de clientes.

##### **Variables e Hipótesis**

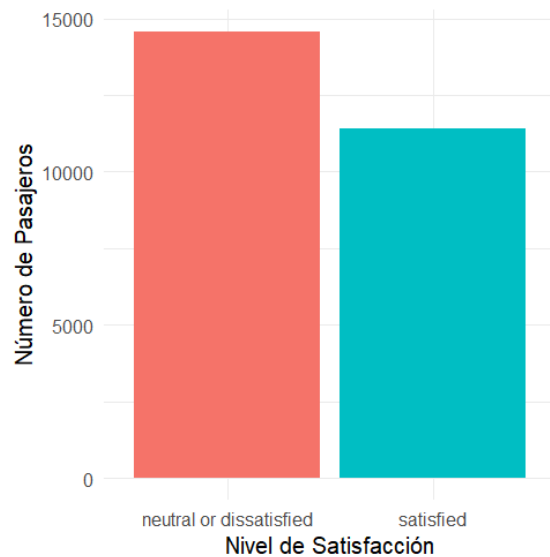
###### Variables Categóricas:

**Satisfaction, Gender, Customer Type, Type of Travel, Class**

###### Variables Numéricas:

**Age, Flight Distance, Inflight wifi service, Arrival Delay in Minutes, Departure/Arrival time convenient, Ease of Online booking, Food and drink, Seat comfort, Inflight entertainment, On-board service, Leg room service, Baggage handling, Checkin service, Cleanliness, Online boarding,**

### Distribucion General de Satisfaccion



Este gráfico inicial muestra que la mayoría de los pasajeros se clasifica en la categoría de no satisfechos, lo que sugiere que existen oportunidades significativas para mejorar el servicio. Este hallazgo refuerza la relevancia de este análisis, ya que detectar áreas de mejora podría revertir esta tendencia y elevar la satisfacción general.

### Hipótesis principales

**Calidad del Servicio:** Variables como **Seat comfort**, **Inflight entertainment**, **Food and drink**, y **On-board service** probablemente tengan un fuerte impacto en la satisfacción, ya que reflejan la calidad del servicio que el pasajero experimenta durante el vuelo.

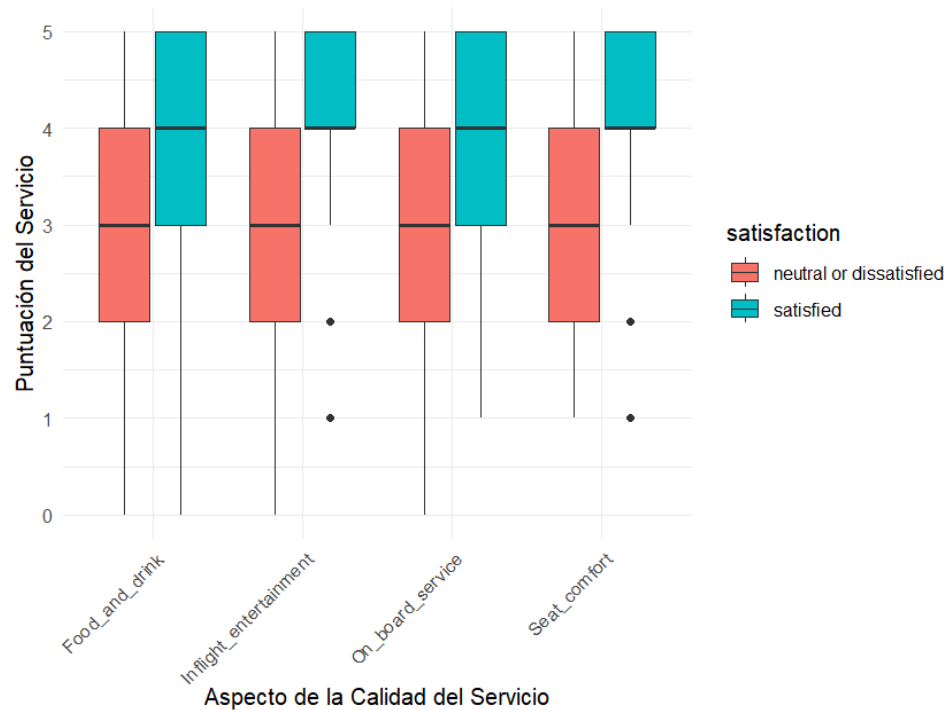
**Tipo de Cliente y Tipo de Viaje:** Se podría esperar que los **"Loyal Customers"** (clientes leales) y **"Business Travel"** (viaje de negocios) tengan una mayor expectativa de calidad, lo que puede influir en la satisfacción.

**Clase de Vuelo:** Pasajeros en clase **"Business"** pueden tener una mayor satisfacción que aquellos en **"Eco"** o **"Eco Plus"**.

**Factores Operacionales:** Variables como el retraso en la llegada **"Arrival Delay in Minutes"** y la conveniencia de los tiempos de salida y llegada **"Departure/Arrival time convenient"** podrían tener un impacto negativo en la satisfacción si el servicio es percibido como ineficiente.

### Gráficos y estadísticas que soportan las hipótesis

#### **Impacto de la calidad del servicio en la satisfacción**

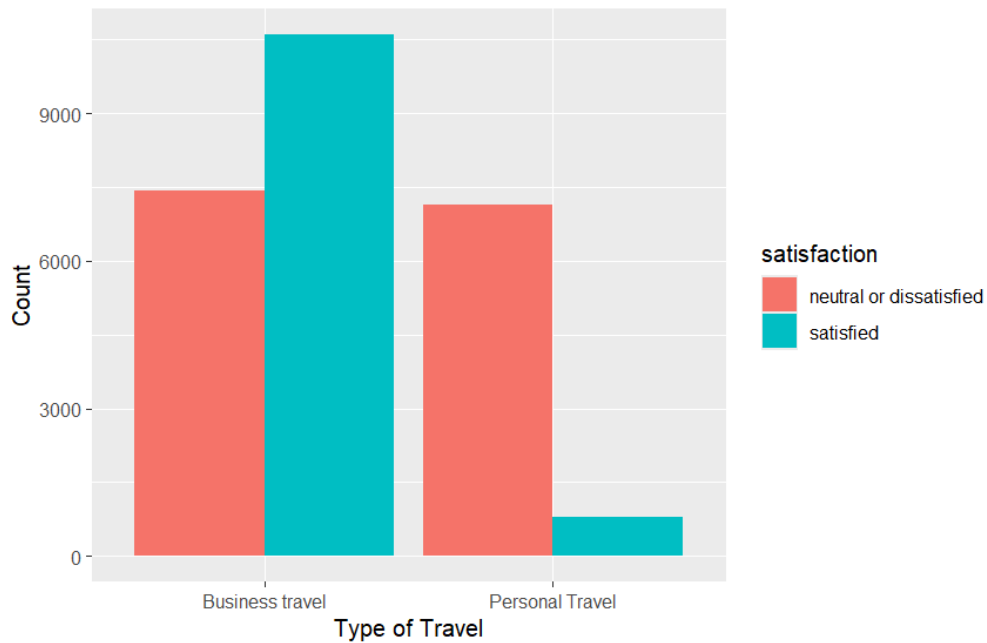


Observamos que los pasajeros más satisfechos tienden a otorgar puntajes más altos en estas categorías de servicio, mientras que los pasajeros menos satisfechos califican más bajo en estas áreas.

A partir de esto, vemos como cada componente de la calidad del servicio se relaciona directamente con la satisfacción general, sugiriendo que una mejora en estos servicios a bordo podría tener un impacto positivo en la percepción del cliente.

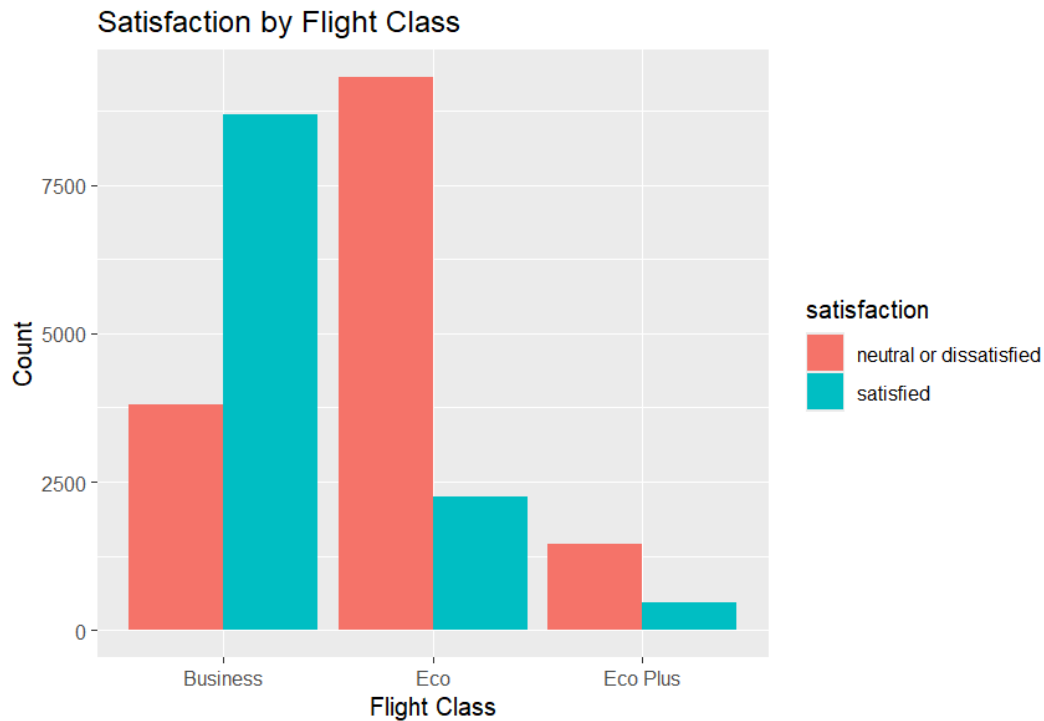
Cada aspecto evaluado contribuye a una experiencia de vuelo positiva, y una calificación alta en estas áreas se asocia con pasajeros satisfechos.

## Impacto del tipo de viaje en la satisfacción



En este caso, vemos que los pasajeros de *Business Travel* muestran un porcentaje más alto de satisfacción en comparación con los de *Personal Travel*. Esto podría deberse a que los pasajeros en viajes de negocios suelen tener expectativas específicas de calidad en el servicio, lo cual influye en su percepción del viaje. Este hallazgo sugiere que la aerolínea podría optimizar los servicios para adaptarse mejor a las necesidades de cada grupo, elevando la satisfacción de ambos tipos de viajeros.”

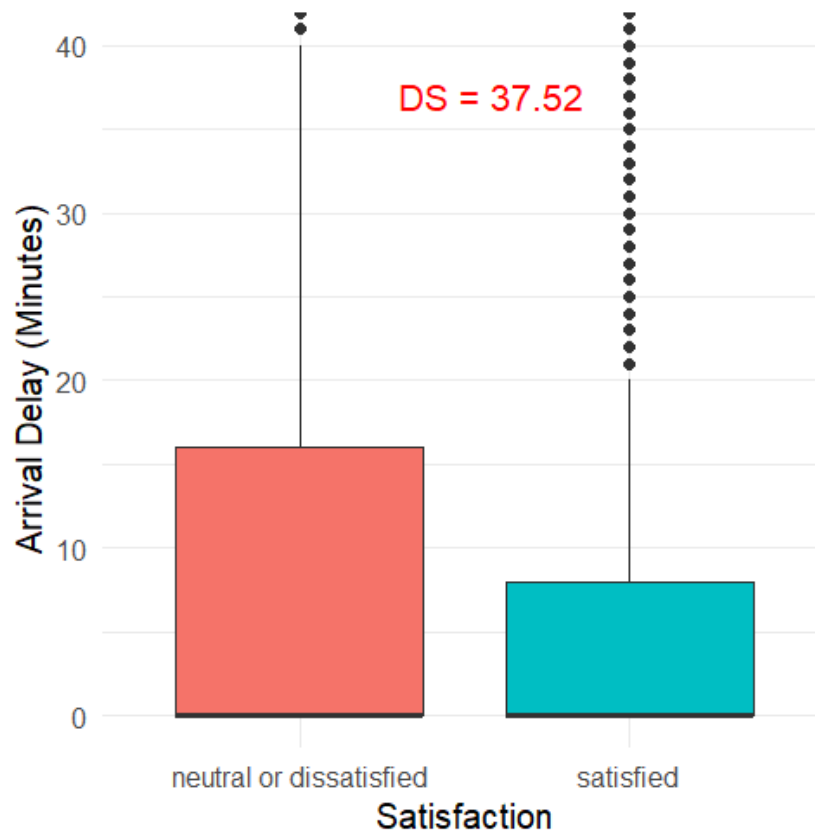
## Impacto de la clase de vuelo en la satisfacción



Este gráfico analiza la relación entre la satisfacción del pasajero y la clase de vuelo, comparando niveles de satisfacción entre las clases *Business*, *Eco Plus*, y *Eco*.

Se observa que los pasajeros de la clase *Business* tienden a mostrar una mayor satisfacción en comparación con los de *Eco Plus* y *Eco*. Esto respalda la hipótesis de que la clase de vuelo tiene un impacto en la satisfacción del cliente, probablemente debido a la diferencia en la calidad del servicio, el espacio y las comodidades disponibles en cada clase. La clase *Business*, que ofrece más beneficios, logra mayores niveles de satisfacción, mientras que la clase *Eco*, al ser la opción más económica, muestra menores niveles de satisfacción. Esto indica que el nivel de confort y los servicios adicionales de las clases superiores contribuyen a mejorar la experiencia del pasajero y, en consecuencia, su satisfacción.

### Impacto del retraso en la llegada en la satisfacción



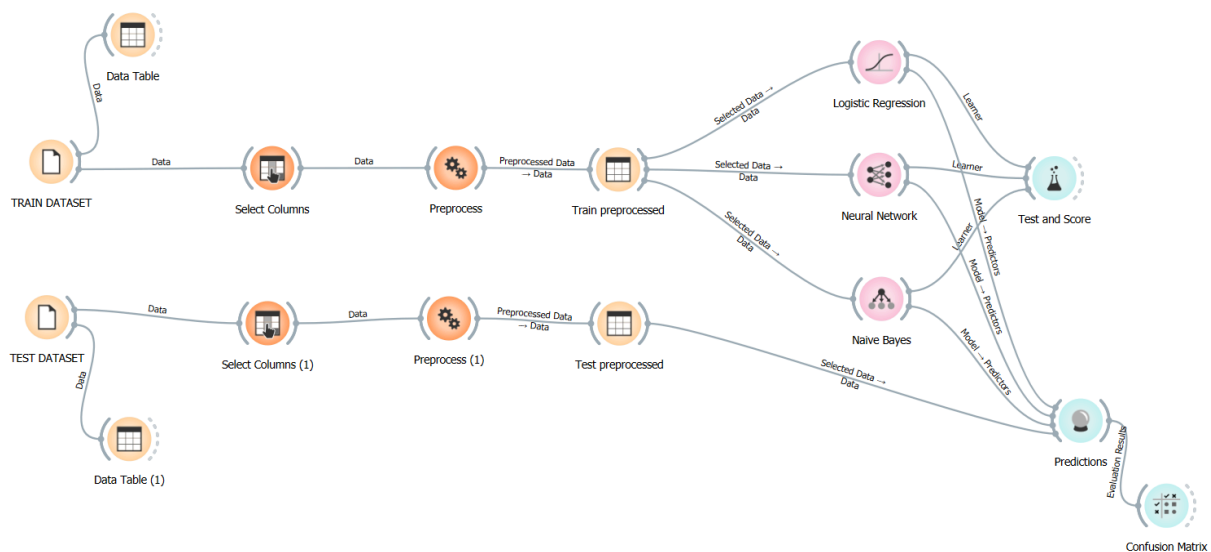
Este gráfico muestra la relación entre el retraso en la llegada de los vuelos (*Arrival Delay in Minutes*) y el nivel de satisfacción de los pasajeros. Para hacer más clara la visualización, se ha ajustado el gráfico de caja y bigotes (*boxplot*) para priorizar los datos centrales y minimizar el impacto de los casos extremos (outliers).

Los pasajeros que experimentaron un retraso menor en la llegada tienden a tener un nivel de satisfacción más alto. Esto respalda la hipótesis de que los factores operacionales, como los retrasos, afectan negativamente la satisfacción. Una llegada puntual o con poco retraso contribuye a una experiencia de viaje percibida como eficiente, mejorando la satisfacción general.

Por otro lado, la desviación estándar de 37,52 minutos en los retrasos de llegada sugiere una alta variabilidad. Este nivel de dispersión indica que los tiempos de retraso son inconsistentes, lo que puede aumentar la insatisfacción del cliente debido a la incertidumbre y la imprevisibilidad de la llegada.

## Pasos realizados en Orange:

1. Se selecciona el dataset para entrenar el modelo.
2. Se limpia el dataset eliminando las columnas de ID, Feature 1 y Gender. Estas columnas no afectan a la satisfacción del vuelo y se eliminan para no recargar el modelo con columnas de más.
3. Se entrena con 3 modelos diferentes, Regresión Logística, Red Neuronal y Naive bayes.
4. Una vez entrenado los modelos, se le envía un dataset para testear los modelos.
5. Con predictions, se pueden ver las métricas para cada modelo.
6. Con la matriz de confusión se puede ver los errores del modelo, los falsos positivos y falsos negativos.



## Conclusión:

Al correr el programa y ver los resultados, nos encontramos con que el método de Machine Learning más eficiente para nuestro caso es el de Red Neuronal, esto se debe a que este modelo logra tomar en cuenta múltiples factores y relaciones entre los mismos. La Red Neuronal es capaz de captar relaciones complejas y no lineales en los datos y con ello pueden aprender patrones en estos grandes conjuntos de datos y realizar predicciones con alta precisión.

Además, cuando analizamos la Classification Accuracy dentro de nuestro sistema, se destaca muy por sobre el resto la clasificación lograda por la Red Neuronal, la cuál alcanza el 92% de efectividad a la hora de predecir.

Nuestro objetivo con este modelo era lograr identificar los factores críticos que tiene mayor influencia en la satisfacción de los pasajeros de la aerolínea.

Para la prueba, separamos los factores que creíamos más importantes en 3 grupos:

- **Variables de Calidad de Servicio:** *Leg room service, Seat comfort, inflight entertainment e inflight wifi service, Cleanliness.*
- **Factores Operacionales:** *Departure delay in mins, Arrival delay in mins, Departure / Arrival time convenient, Ease of Online Booking.*
- **Tipos de Cliente y Clase:** *Customer type, Type of travel, Class*

Las variables, separadas en grupos, resultaron positivas a la hora de predecir la satisfacción. El grupo “Variables de Calidad de Servicio” fue el grupo que mayor Classification Accuracy obtuvo con un 86%, muy por encima del resto (69% y 78% respectivamente). Con esto, podemos concluir que nuestra hipótesis es acertada y que estos factores específicos tienen un impacto directo en la experiencia de los pasajeros. A partir de esto, le recomendaremos a la aerolínea que inviertan en mejorar la comodidad, limpieza y entretenimiento a bordo antes que en cualquier otra área. Invertir en las Variables de Calidad Servicio debería incrementar la satisfacción de sus clientes más que cualquier otro factor tomado en cuenta en este caso.