

DATA SCIENCES

Proyecto Final

Analizar las preferencias de
los pasajeros de Aerolíneas
British Airways

Integrantes

Araujo Daniela
Mañaricua Aldana
Oxandaburu Gabriela
Sánchez Clara



Lo que se pretende con este proyecto es un análisis de de todas las elecciones de los pasajeros a la hora de reservar un vuelo, para desarrollar un plan de negocios que nos ayude a la toma de decisiones de una empresa.

Para esto usamos una base de datos de Reservas de la Aerolínea British Airways.

Usando como herramientas el Análisis Exploratorio, Modelo de Aprendizaje Supervisado y Modelo de Aprendizaje No Supervisado.



Características del DS

Tiene 50002 Filas y 14 Columnas.

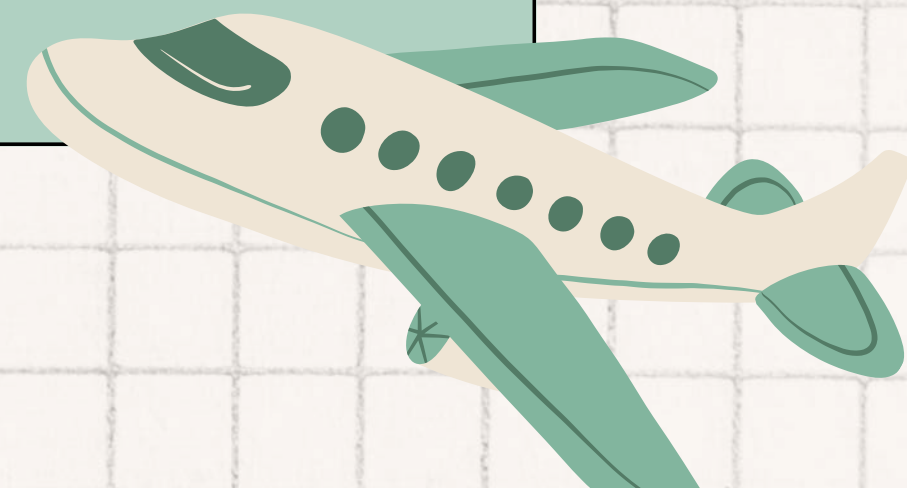
Variables:

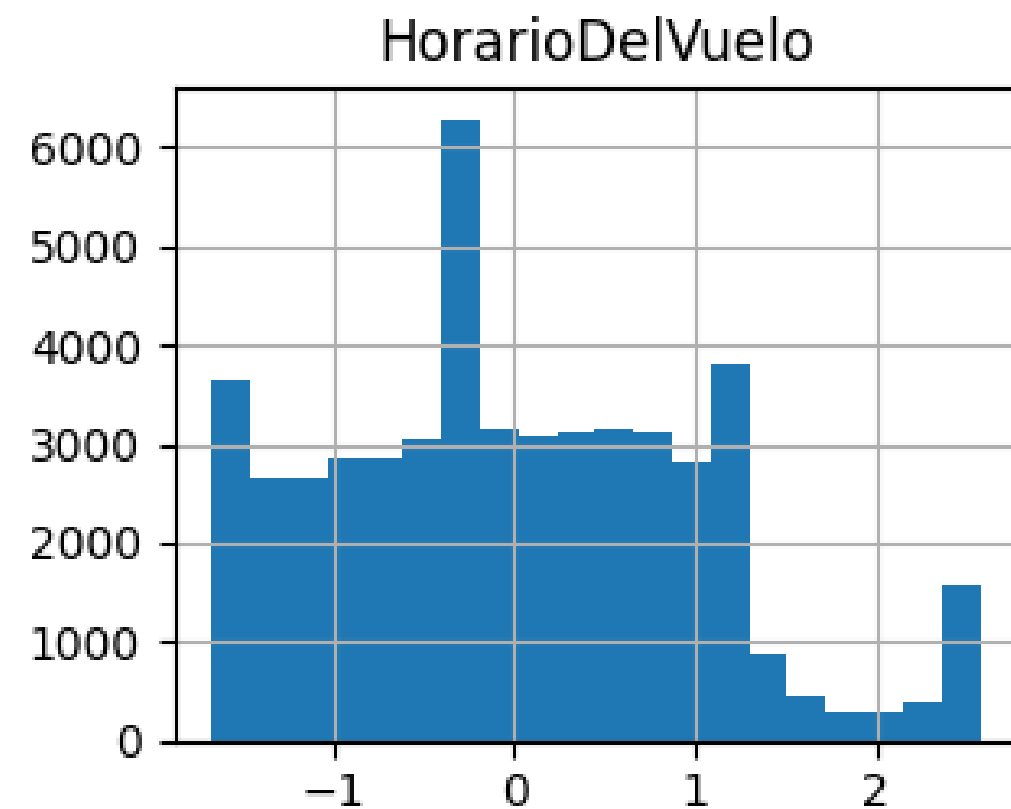
- Número de Pasajeros
- Canal de Venta
- Tipo de Viaje
- Cliente Potencial
- Estadía
- Horario del Vuelo
- Día del Vuelo
- Ruta
- Origen de la Reserva
- Equipaje Extra
- Asientos Preferidos
- Quiere Comida en Vuelo
- Duración del Vuelo
- Reserva Completa



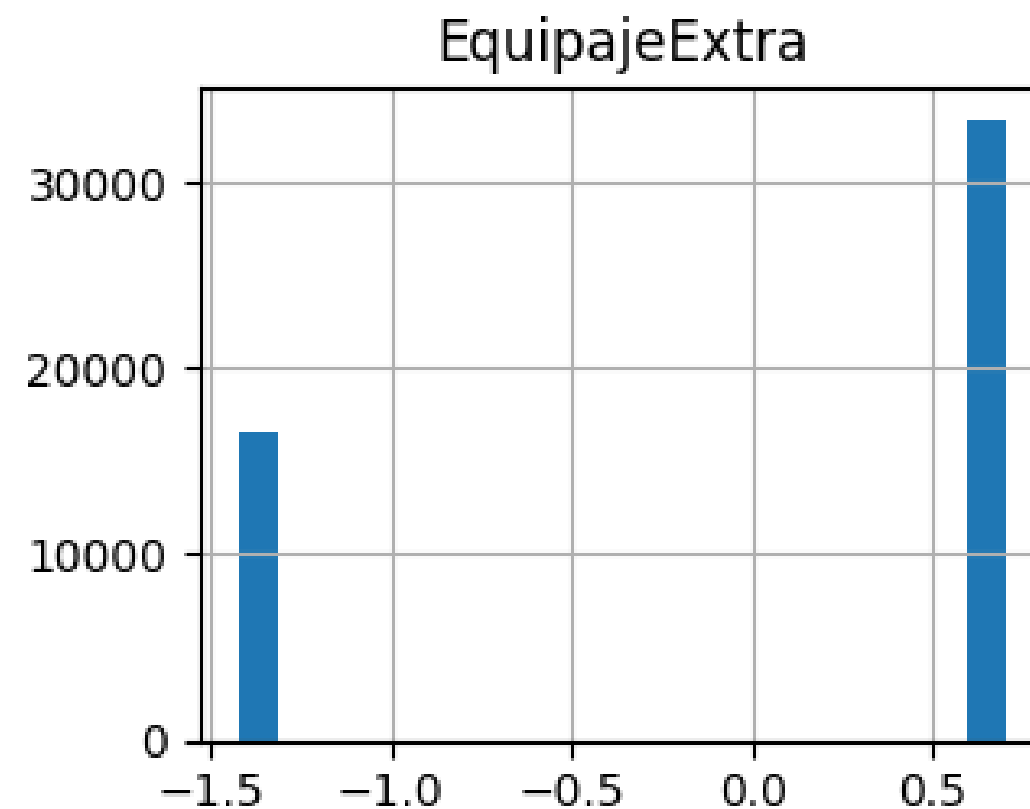
Del EDA obtuvimos que:

- El día mas elegido para volar es el dia lunes.
- El 35.74 % de las reservas se realizan desde Australia.
- Preferencia de los clientes es reservar pasajes de "Ida y Vuelta" / "RoundTrip".

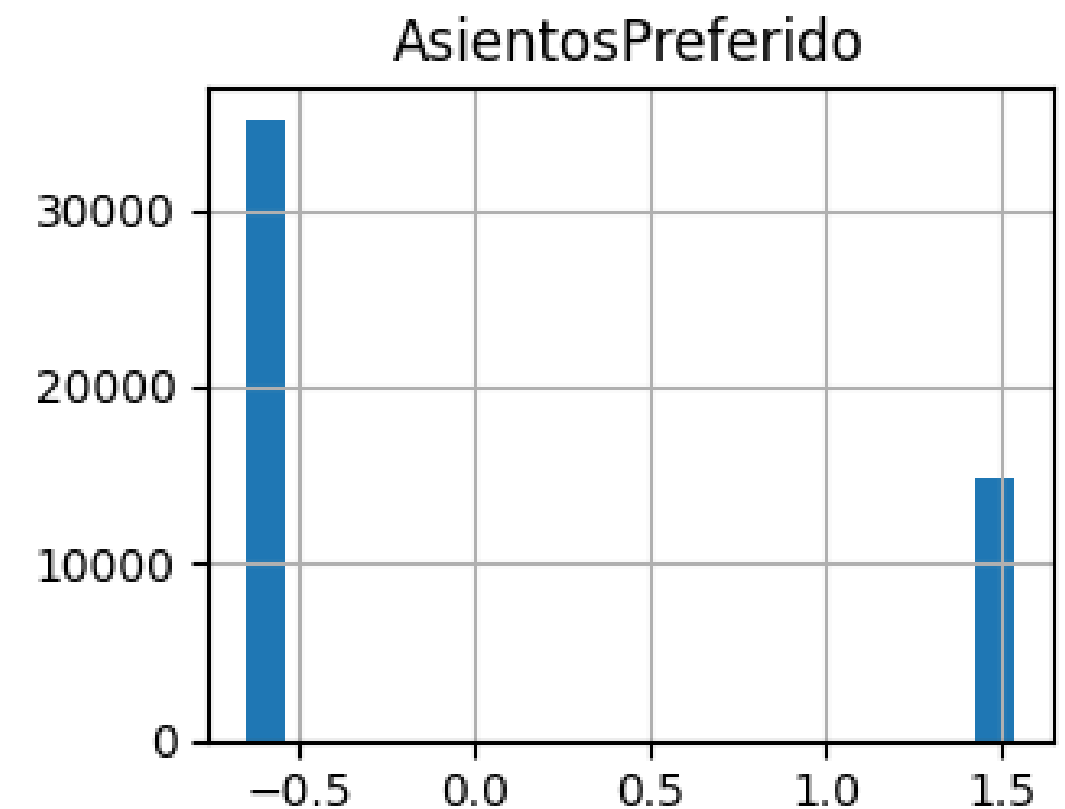




Los horarios de vuelo mas elegido son 8 AM, 12 AM y 9 AM.



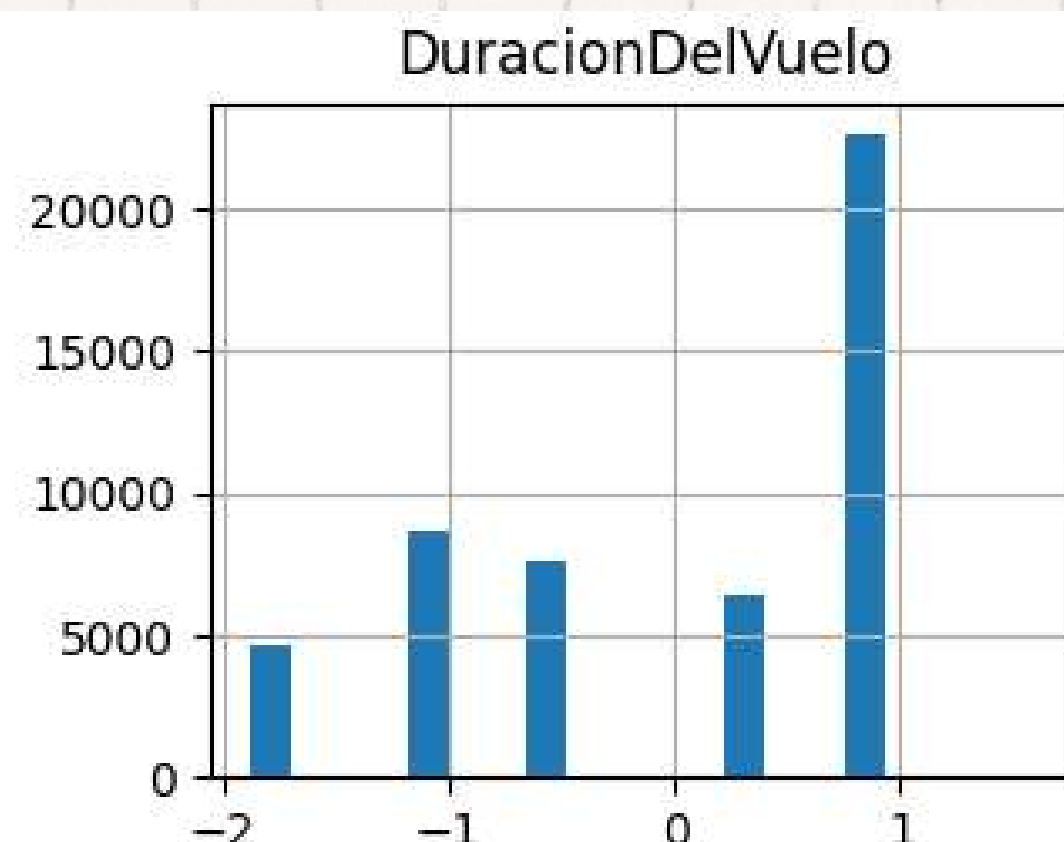
Mas de la mitad de los pasajeros eligen llevar equipaje extra.



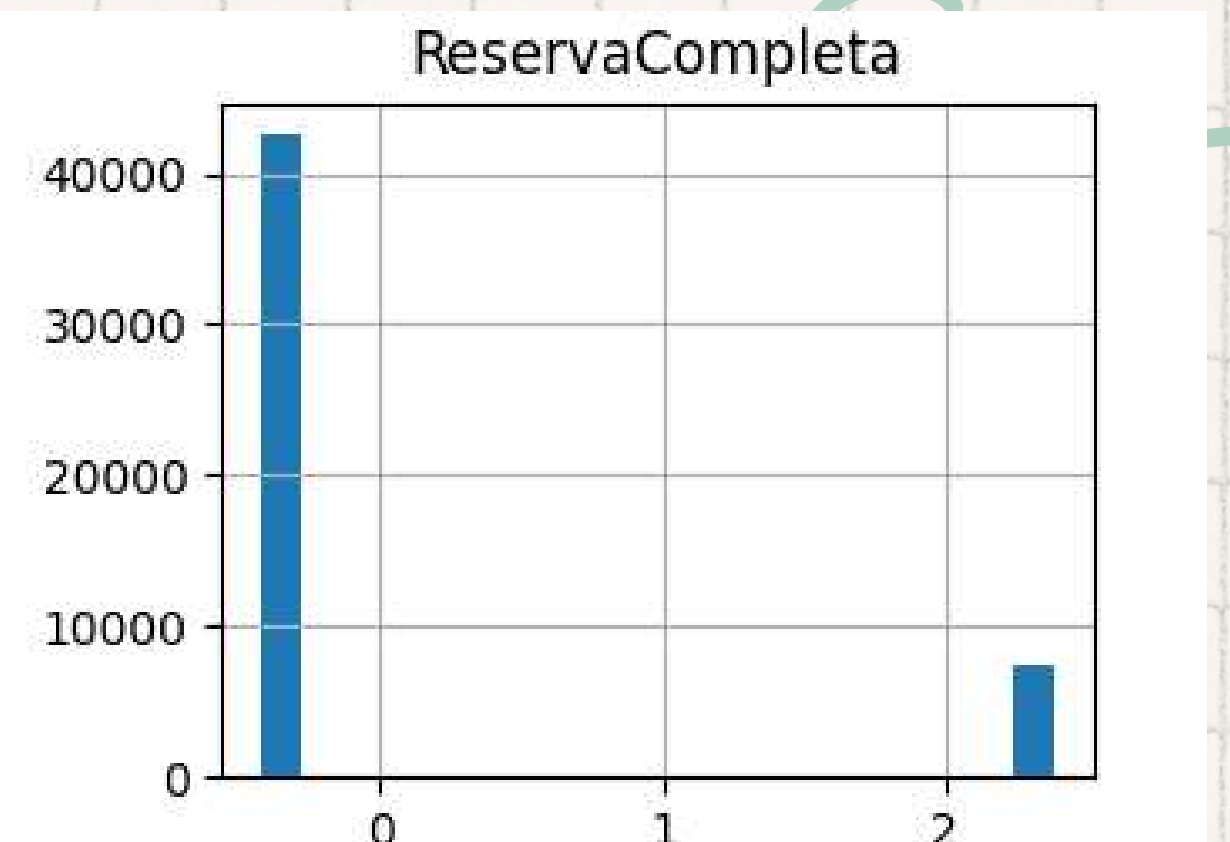
La mayoría no elige el asiento.



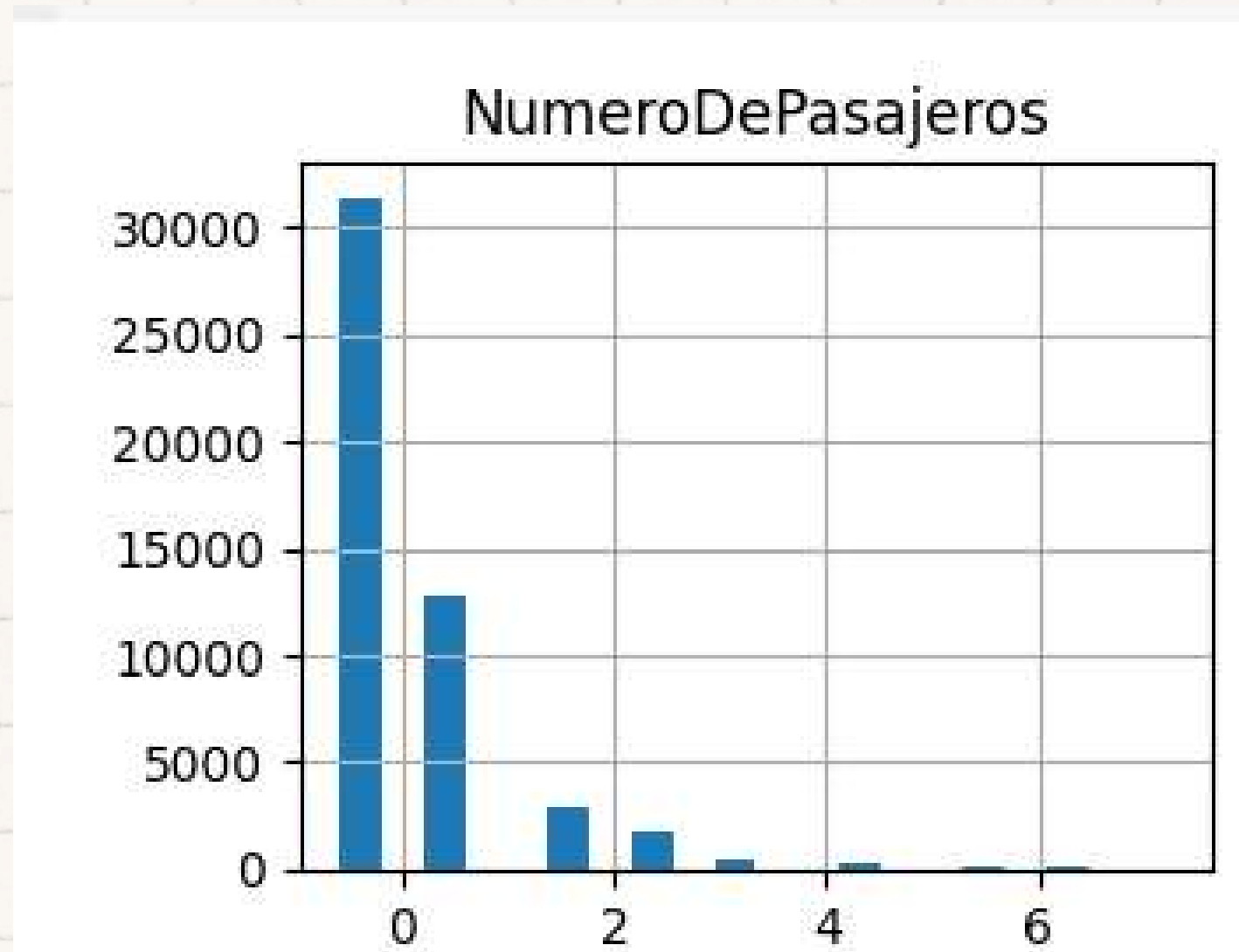
La mayoría de los pasajero prefiere no comer durante el vuelo.



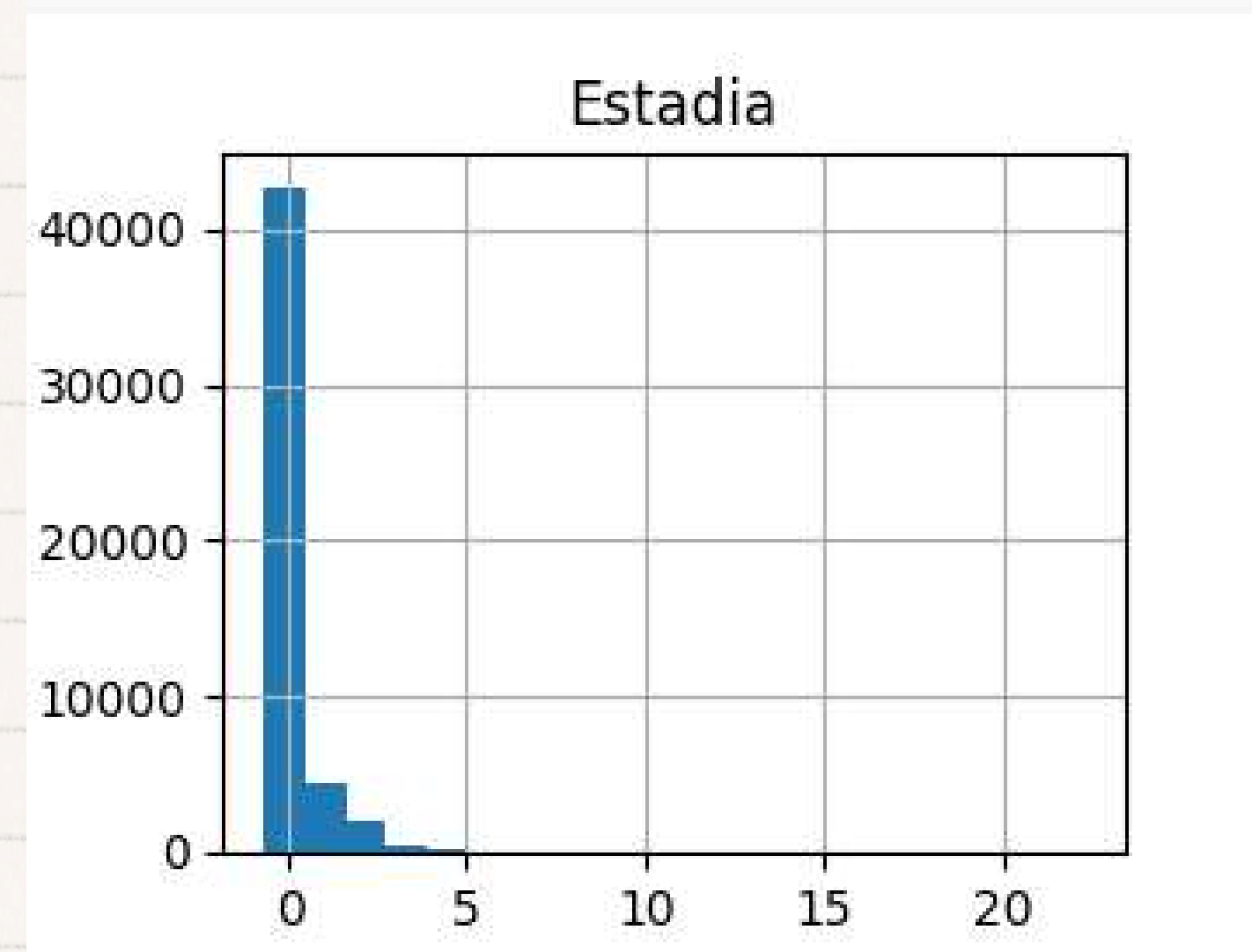
El promedio de duración de los vuelos es de 8 hs.



La mayoría no realiza la reserva completa.

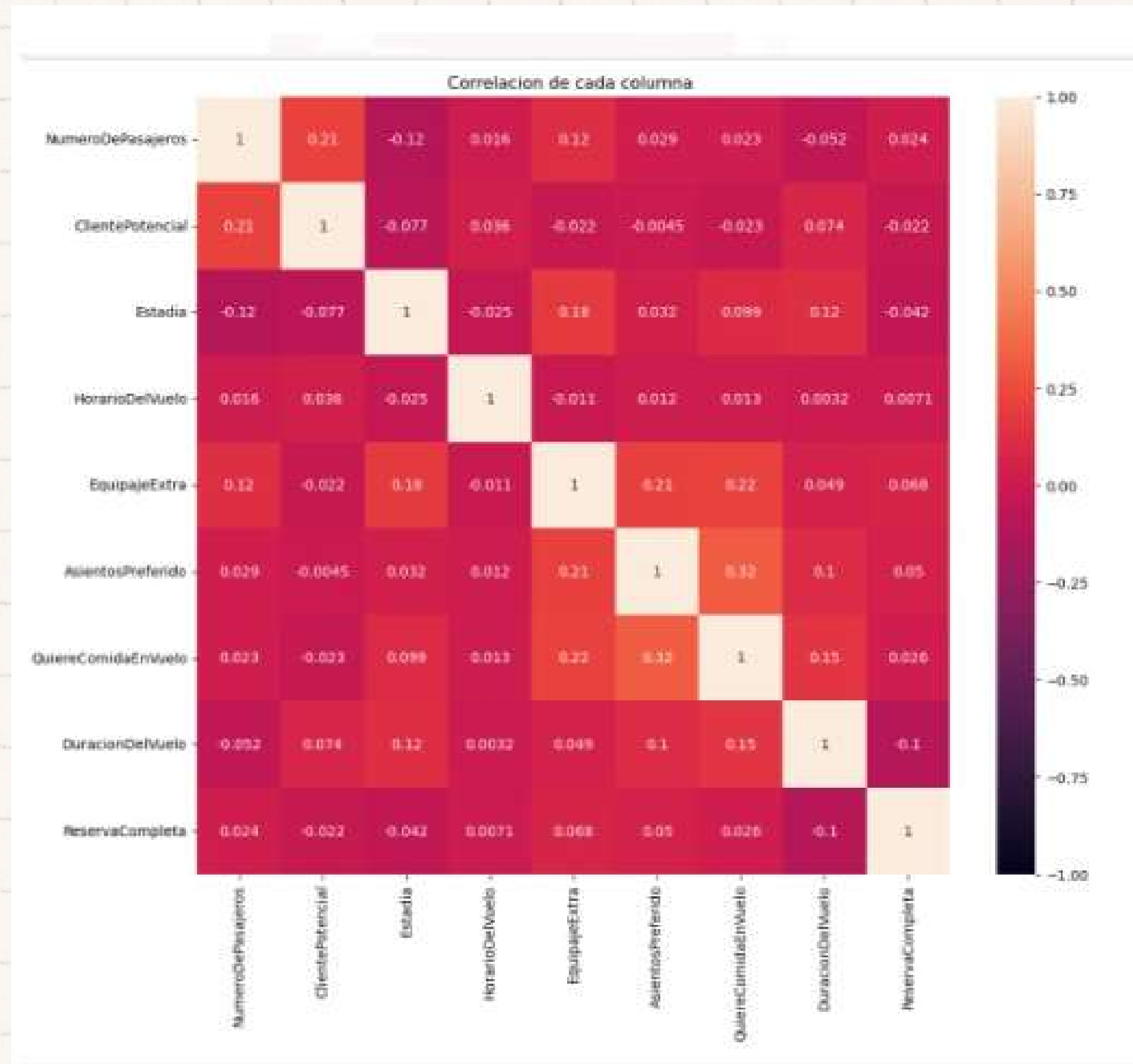


Observamos que la mayoría de los vuelos son de 1 solo pasajero.



La estadía mas frecuente es de 6 días de duración.

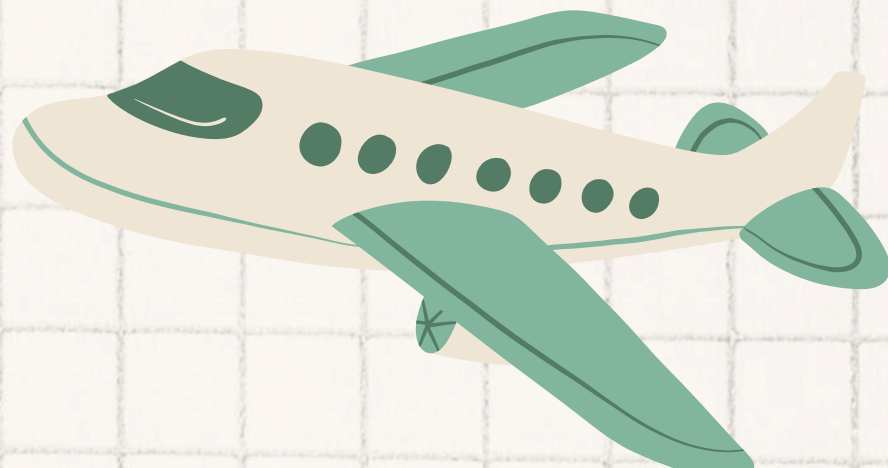
Gráfico de Correlación



Aprendizaje Supervisado

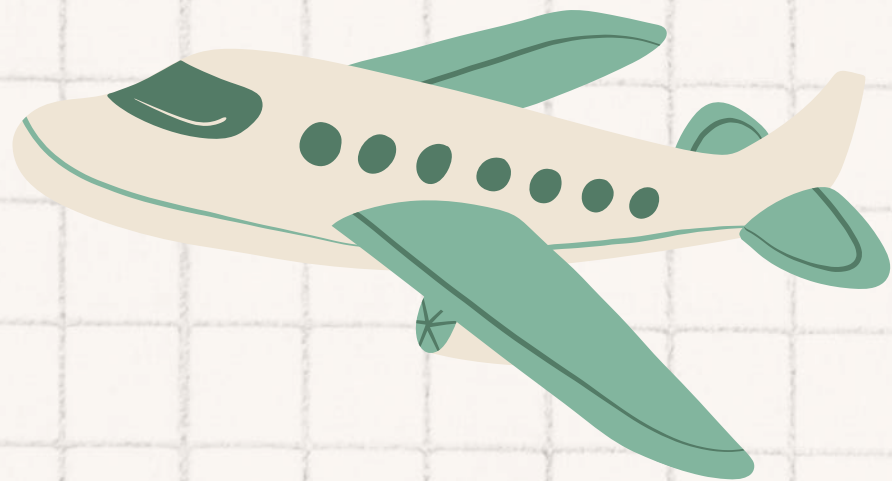
Clasificación

Objetivo: Predicción de la elección
de equipaje extra de aerolíneas
British Airways



Modelos Predictivos

- K - Nearest Neighbors
- Random Forest



Clasificador KNN

Primera Prueba

- Accuracy: 66.10%
- Precisión: 72.04%
- Recall: 80.57%
- F1 score: 76.07%



Clasificador KNN

Segunda Prueba

- Accuracy: 68.43%
- Precision: 74.43%
- Recall: 80.42%
- F1 score: 77.31%



Random Forest

- Accuracy: 73.04%
- Precision: 76.82%
- Recall: 85.46%
- F1 score: 80.91%



Técnica que busca mejorar los
hiperparámetros

Gridsearch

- Accuracy: 68.43%
- Precision: 74.43%
- Recall: 80.42%
- F1 score: 77.31%



Conclusión del AS

El Modelo que mejor performa es el Random Forest, ya que obtiene un Accuracy de 73.04%.

Esto representa la proporción de predicciones correctas realizadas por el modelo en relación con el total de predicciones.

Sin embargo dado que al analizar la correlación entre las variables mencionamos que hay muy baja correlación esto puede afectar el rendimiento del Modelo y por ello podría ser mas eficiente si hubieran features que cumplieran con esa característica. Incluyendo nuevas variables en el Modelo se podría obtener una mejor predicción de la elección de equipaje extra, para lo cual deberíamos solicitar a la Aerolínea nuevos datos

Modelo de Aprendizaje No Supervisado

Objetivo: clasificar y agrupar a los usuarios de British Airlines, según sus preferencias.

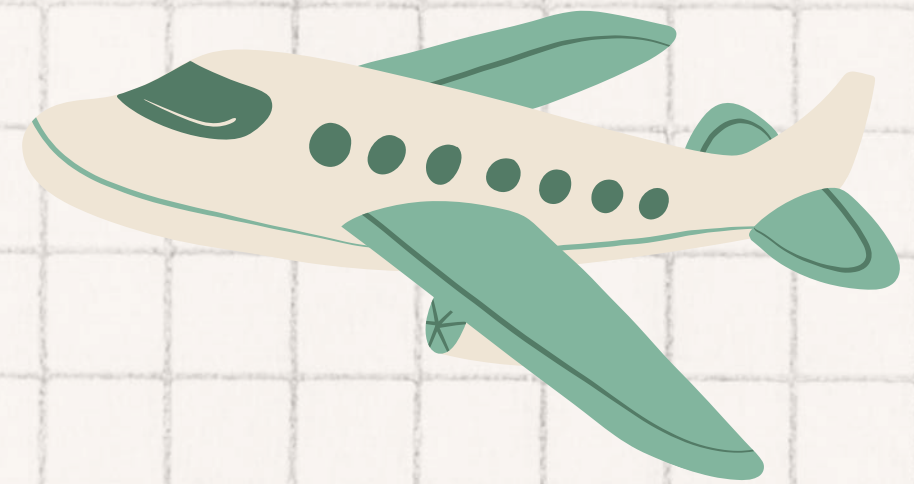


Modelo de Clustering

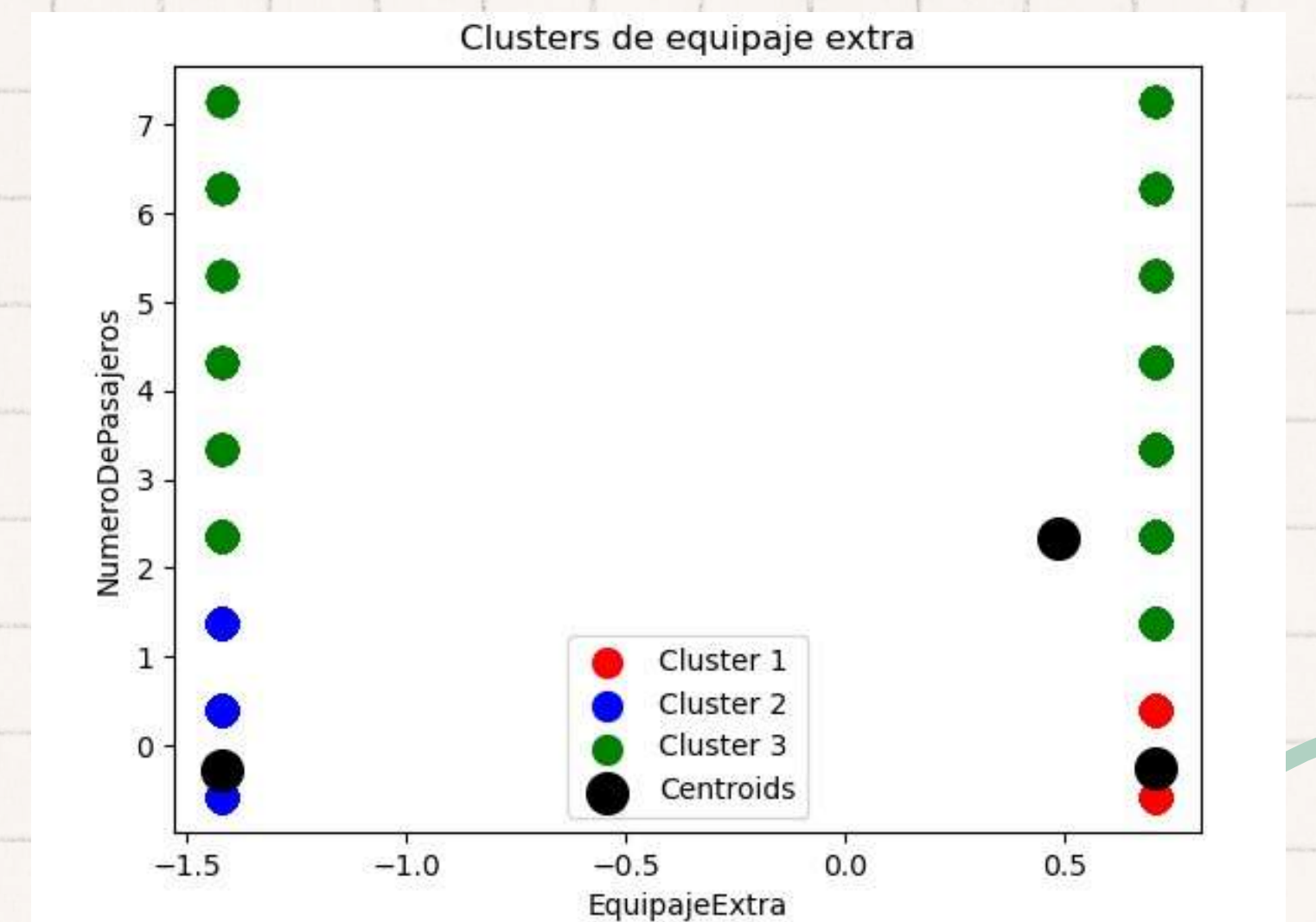
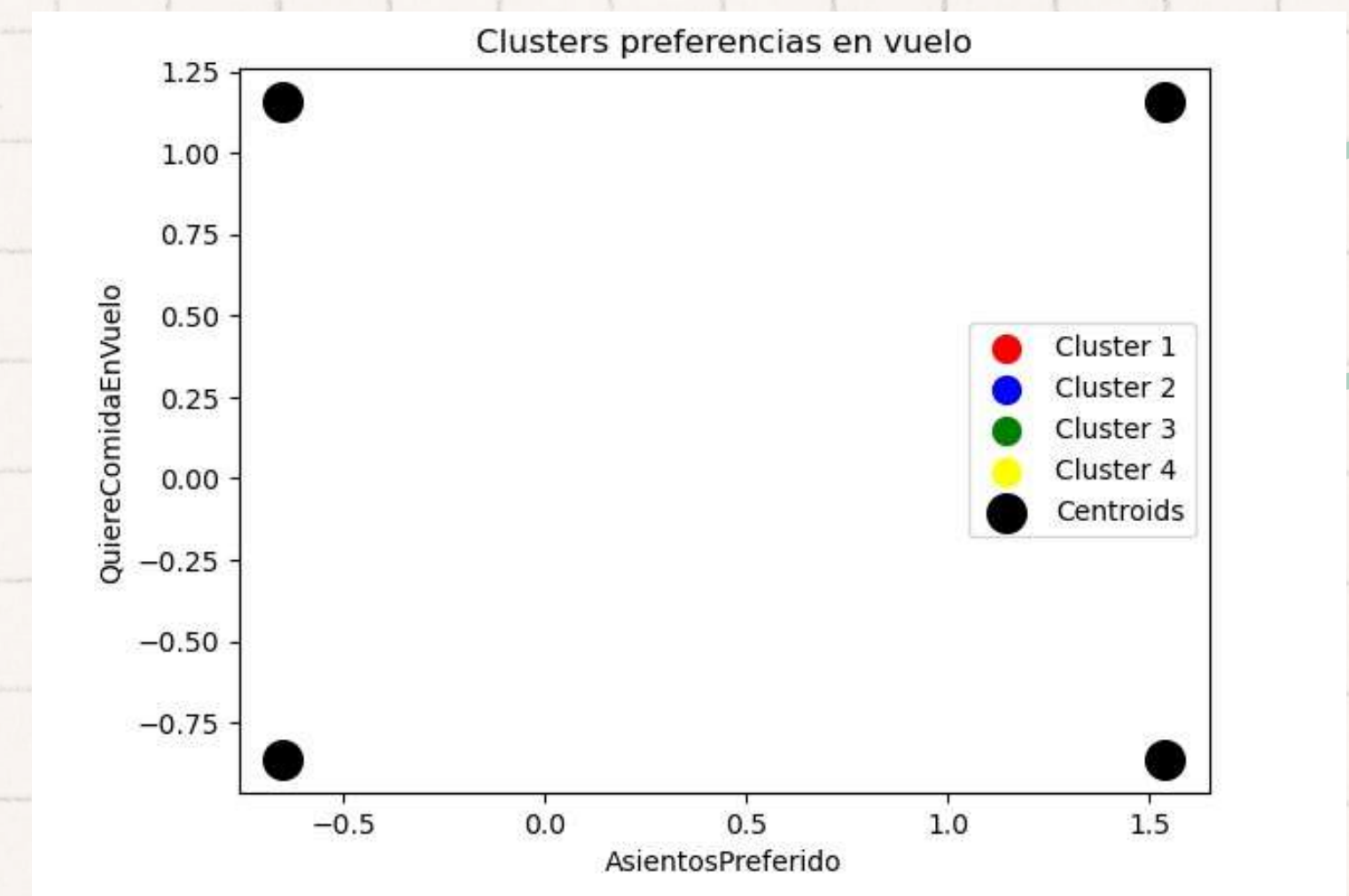
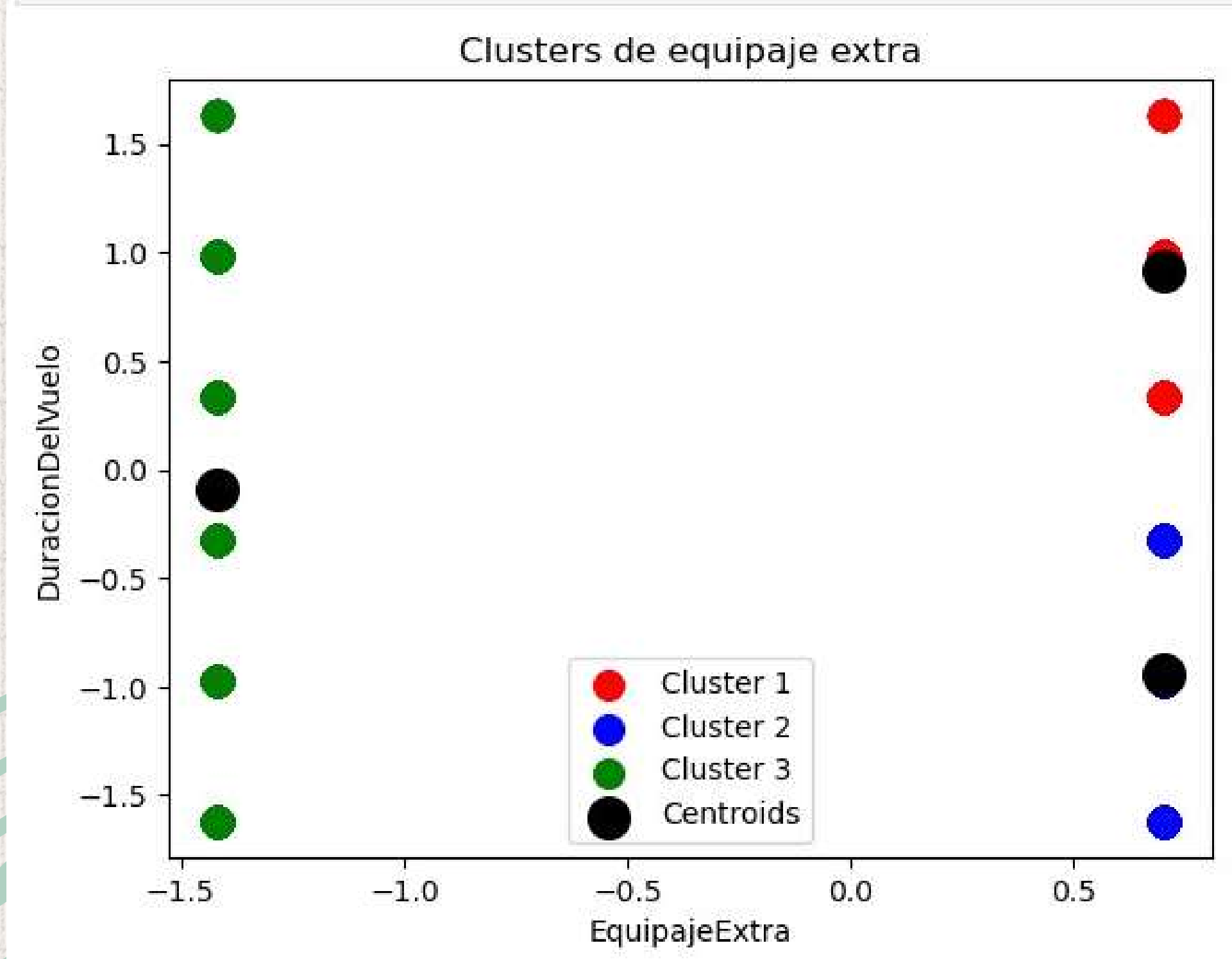
K-MEANS

Variables:

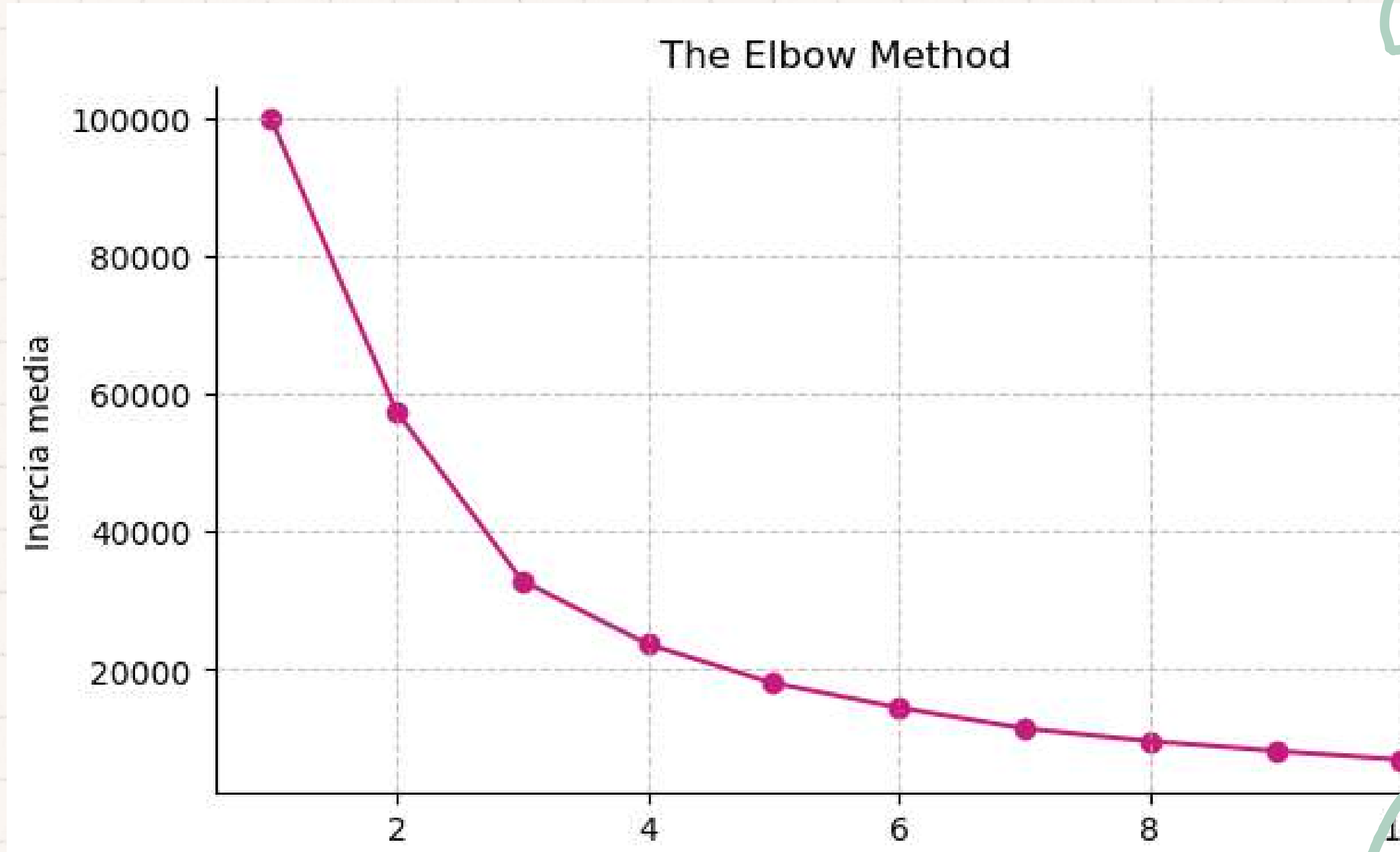
- Estadía
- Duración del Vuelo



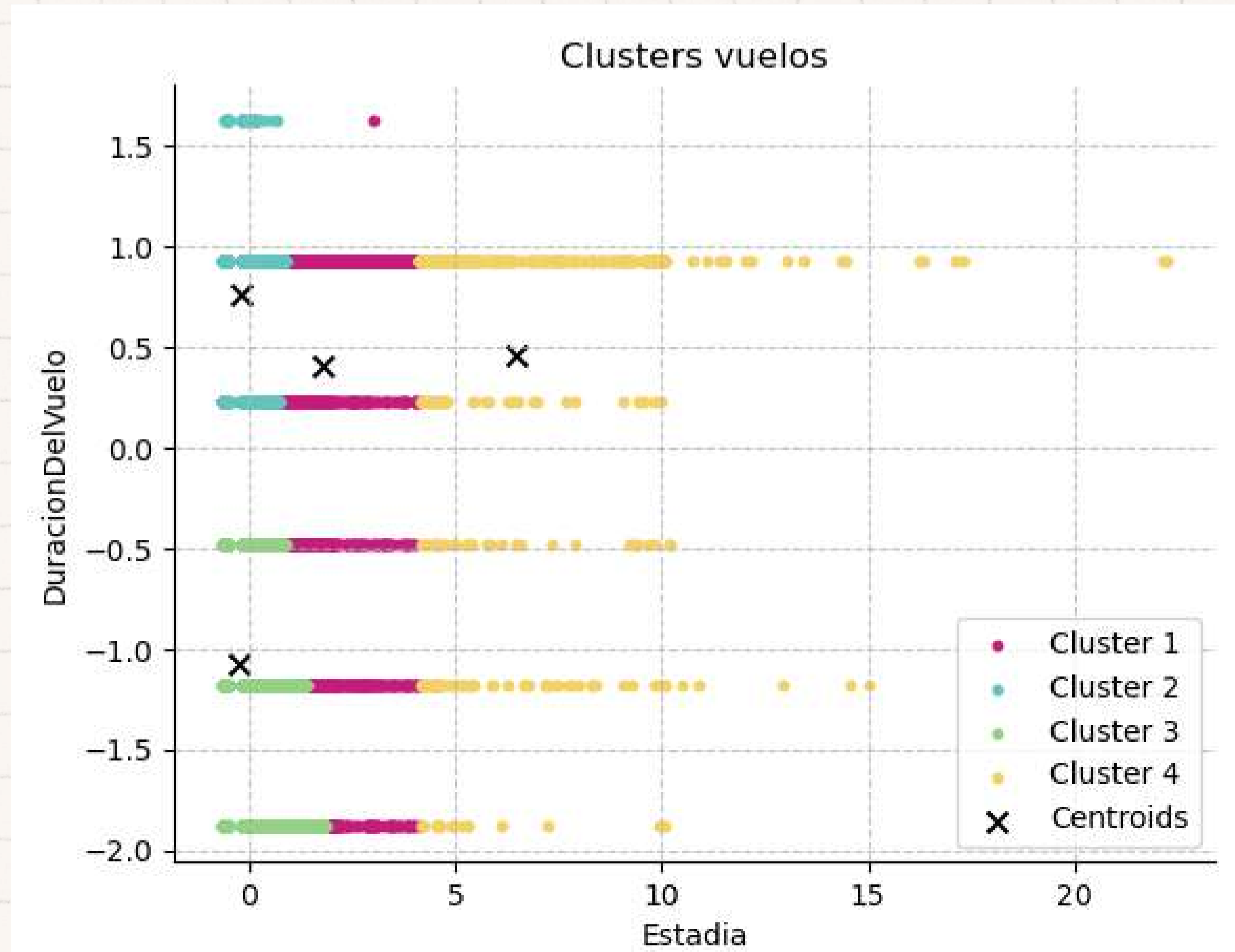
Pruebas previas ...



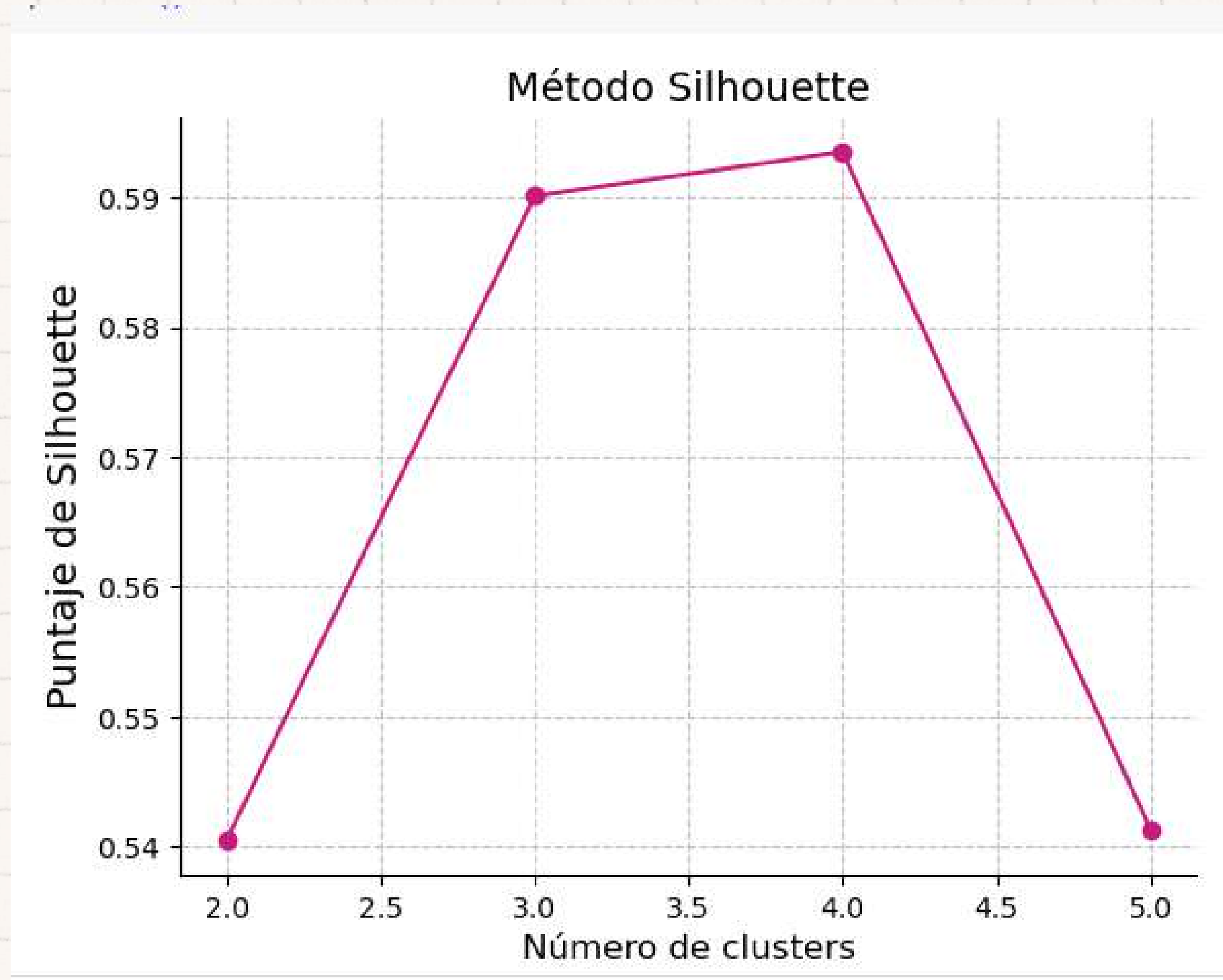
Método del Codo



Visualizamos los clusters



Método Silhouette



Conclusión del ANS

- Las características de las features del dataset impiden la realización de un modelo K- Means dado que la mayoría de las variables son categóricas y muchas de ellas binarias.
- Estadía tiene varios outliers
- Conviene Modelo de Aprendizaje Supervisado
- Solicitud de datos anexos que tengan las características necesarias, tales como por ejemplo: precio de los vuelos, distancia, edades de los pasajeros, etc.

Gracias!

