

# UNIVERSIDAD DE COSTA RICA

ESCUELA DE MATEMÁTICA

DEPARTAMENTO DE MATEMÁTICA PURA Y CIENCIAS

ACTUARIALES

ANÁLISIS DE DATOS I, CA0411

---

## Proyecto de investigación

---

**Impacto de la calidad del servicio y las condiciones operativas  
en la satisfacción de pasajeros en aerolíneas.**

Realizado por

---

Edwin Amador, B80429

Amel Cáceres, B51313

Johan Castaño, B71744

Diana Ramírez, B76187



UNIVERSIDAD DE  
COSTA RICA

EMat

Escuela de  
Matemática

# Índice general

<b>1. Bitácora 1</b>	<b>4</b>
1.1. Análisis de los Datos . . . . .	4
1.1.1. Características de la tabla . . . . .	4
1.1.2. Población de estudio: . . . . .	6
1.1.3. Muestra observada . . . . .	6
1.1.4. Unidad estadística o individuos . . . . .	6
1.2. UVE de Gowin . . . . .	8
1.3. Pregunta de investigación . . . . .	9
1.4. Objeto de estudio . . . . .	9
1.5. Conceptos fundamentales . . . . .	9
1.6. Teorías . . . . .	10
1.6.1. Teoría de Confirmación y Desconfirmación de Expectativas: . . . . .	10
1.6.2. Modelo Servqual: . . . . .	10
1.6.3. Metodología Kano: . . . . .	11
1.7. Tabla de datos . . . . .	12
1.8. Resumen de 5 números de las variables cuantitativas y analizar el mismo. . . . .	14
1.9. Gráficos . . . . .	16
1.9.1. Distribución de las variables cuantitativas: . . . . .	16
1.9.2. Relación entre las variables: . . . . .	19
1.9.3. Distribución de las variables categóricas: . . . . .	22

ÍNDICE GENERAL

2

1.10. Valores faltantes y posibles outliers . . . . .

24

1.10.1. Identificar valores faltantes y posibles outliers . . . . .

24

1.10.2. Técnicas que permitan subsanar los valores perdidos y outliers . . . . .

25

A. Repositorio

27

Referencias . . . . .

28

# Índice de figuras

1.1. Pirámide poblacional . . . . .	16
1.2. Distancia recorrida en millas náuticas por persona. . . . .	17
1.3. Retraso en hora de salida del vuelo en minutos por persona. . . . .	18
1.4. Retraso en hora de llegada del vuelo en minutos por persona. . . . .	19
1.5. Proporción de satisfacción por Clase . . . . .	20
1.6. Correlación de Pearson entre las variables . . . . .	21
1.7. Distribución de calificaciones para diferentes servicios. . . . .	22
1.8. Distribución de calificaciones para diferentes servicios. . . . .	23
1.9. Boxplot de las variables. . . . .	25

# Capítulo 1

## Bitácora 1

### 1.1. Análisis de los Datos

#### 1.1.1. Características de la tabla

- **Nombre del archivo:** Airline Passenger Satisfaction.
- **Fuente:** Kaggle.
- **Número de registros:** 103903.
- **Número de variables:** 24.
- **Variables:**
  1. Gender: Género de los pasajeros (Female, Male).
  2. Customer Type: Tipo de cliente (Loyal customer, disloyal customer).
  3. Age: Edad real de los pasajeros (en años).
  4. Type of Travel: Propósito del vuelo (Personal Travel, Business Travel).

5. Class: Clase de viaje en el avión (Business, Eco, Eco Plus).
6. Flight distance: Distancia del vuelo (en millas).
7. Inflight wifi service: Satisfacción con el servicio de wifi a bordo.
  - 0: No aplica.
  - 1-5: Escala de satisfacción (1=Muy insatisfecho, 5=Muy satisfecho).
8. Departure/Arrival time convenient: Satisfacción con horarios (escala 0-5).
9. Ease of Online booking: Facilidad de reserva en línea (escala 0-5).
10. Gate location: Satisfacción con ubicación de puerta (escala 0-5).
11. Food and drink: Satisfacción con comida/bebida (escala 0-5).
12. Online boarding: Satisfacción con embarque en línea (escala 0-5).
13. Seat comfort: Comodidad del asiento (escala 0-5).
14. Inflight entertainment: Entretenimiento a bordo (escala 0-5).
15. On-board service: Servicio durante el vuelo (escala 0-5).
16. Leg room service: Espacio para piernas (escala 0-5).
17. Baggage handling: Manejo de equipaje (escala 0-5).
18. Check-in service: Servicio de check-in (escala 0-5).
19. Inflight service: Servicio general en vuelo (escala 0-5).
20. Cleanliness: Limpieza del avión (escala 0-5).
21. Departure Delay in Minutes: Minutos de retraso en salida.

22. Arrival Delay in Minutes: Minutos de retraso en llegada.

23. Satisfaction: Satisfacción general con la aerolínea.

- Satisfaction.
- Neutral or dissatisfaction.

### 1.1.2. Población de estudio:

Todos los pasajeros que han volado con la aerolínea y han completado una encuesta de satisfacción.

### 1.1.3. Muestra observada

- **Tamaño de la muestra:** 103903.
- **Cobertura:** Pasajeros de diferentes edades, géneros, tipos de viaje y clases de vuelo.
- **Limitaciones:** No se especifica el período de tiempo ni la aleatorización de la muestra, lo que podría afectar la generalización.

### 1.1.4. Unidad estadística o individuos

- **Unidad de análisis:** Cada registro representa un pasajero único identificado por id.
- **Atributos por individuo:**
  - Demográficos (género, edad).
  - De viaje (tipo de cliente, clase, distancia).
  - Evaluaciones de servicios (wifi, comida, entretenimiento).
  - Retrasos y satisfacción.

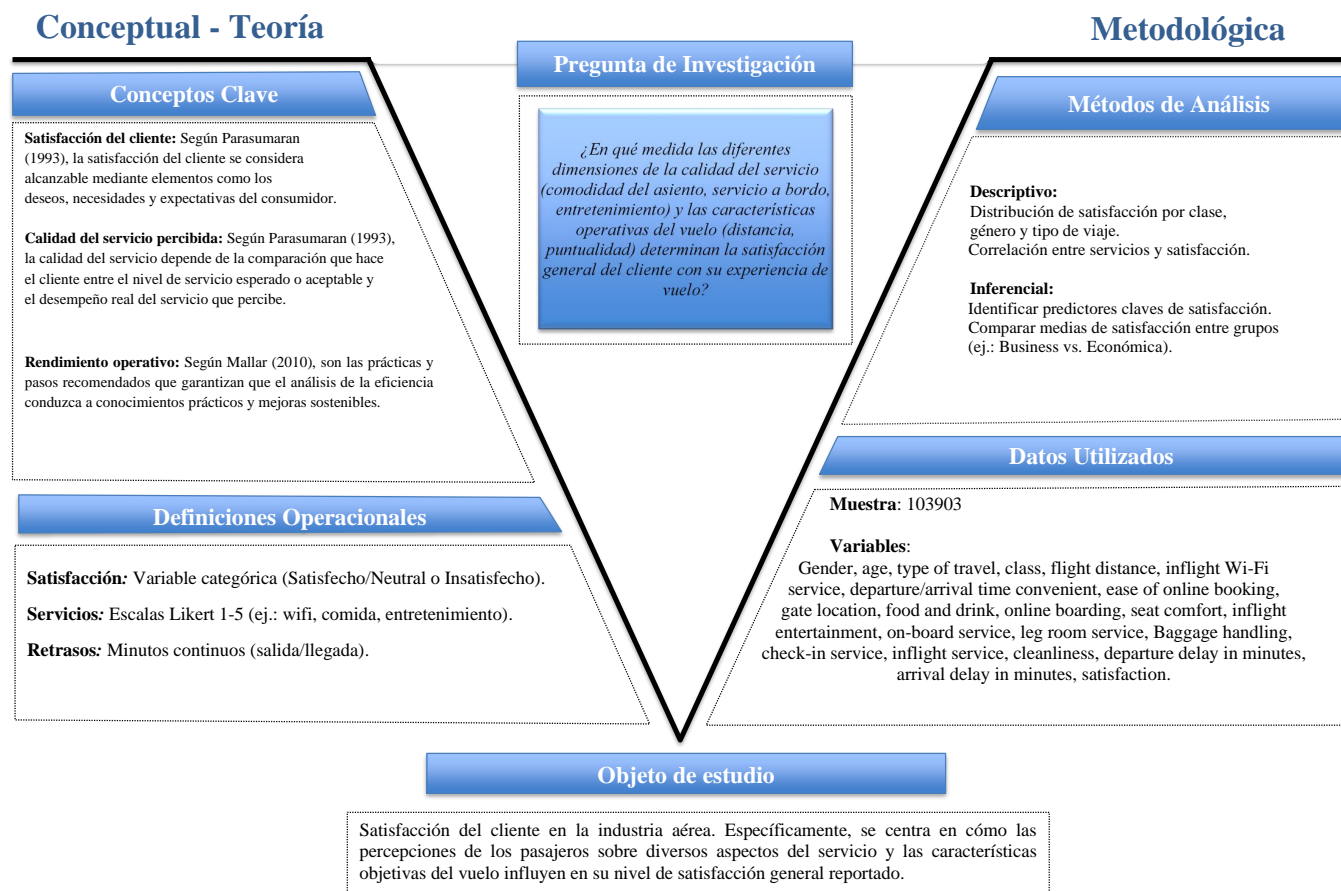
Cuadro 1.1: Variables del estudio

Variable	Tipo	Escala	Descripción
Gender	Cualitativa	Nominal	Género (Masculino/Femenino).
Customer Type	Cualitativa	Nominal	Tipo de cliente (Leal/No leal).
Age	Cuantitativa	Razón	Edad en años
Type of Travel	Cualitativa	Nominal	Motivo del viaje (Negocios/Personal).
Class	Cualitativa	Ordinal	Clase (Económica/Económica Plus/Business).
Flight Distance	Cuantitativa	Razón	Distancia del vuelo en millas.
Retrasos	Cuantitativa	Razón	Minutos de demora.
Satisfaction	Cualitativa	Nominal	Nivel de satisfacción (Satisfecho/Insatisfecho).
Inflight wifi	Cuantitativa	Intervalo (0-5)	Satisfacción con wifi a bordo (0=No usado).
Departure/Arrival time	Cuantitativa	Intervalo (1-5)	Conveniencia horarios salida/llegada.
Ease of Online booking	Cuantitativa	Intervalo (1-5)	Facilidad de reserva en línea.
Gate location	Cuantitativa	Intervalo (1-5)	Ubicación de puerta de embarque.
Food and drink	Cuantitativa	Intervalo (1-5)	Calidad comida/bebida.
Online boarding	Cuantitativa	Intervalo (1-5)	Proceso de embarque en línea.
Seat comfort	Cuantitativa	Intervalo (1-5)	Comodidad del asiento.
Inflight entertainment	Cuantitativa	Intervalo (1-5)	Entretenimiento a bordo.
On-board service	Cuantitativa	Intervalo (1-5)	Servicio durante el vuelo.
Leg room service	Cuantitativa	Intervalo (1-5)	Espacio para piernas.
Baggage handling	Cuantitativa	Intervalo (1-5)	Manejo de equipaje.
Check-in service	Cuantitativa	Intervalo (1-5)	Servicio de check-in.
Inflight service	Cuantitativa	Intervalo (1-5)	Servicio general en vuelo.
Cleanliness	Cuantitativa	Intervalo (1-5)	Limpieza del avión.
Departure Delay	Cuantitativa	Razón	Retraso salida (minutos).
Arrival Delay	Cuantitativa	Razón	Retraso llegada (minutos).

Fuente: Elaboración propia, con datos de (Klein, 2021).



## 1.2. UVE de Gowin



## 1.3. Pregunta de investigación

¿En qué medida las diferentes dimensiones de la calidad del servicio (comodidad del asiento, servicio a bordo, entretenimiento) y las características operativas del vuelo (distancia, puntualidad) determinan la satisfacción general del cliente con su experiencia de vuelo?

## 1.4. Objeto de estudio

El objeto de estudio es la satisfacción del cliente en la industria aérea. Específicamente, se centra en cómo las percepciones de los pasajeros sobre diversos aspectos del servicio y las características objetivas del vuelo influyen en su nivel de satisfacción general reportado.

## 1.5. Conceptos fundamentales

- **Satisfacción del cliente:** Según Parasumaran, A. y Zeithaml, V. (1993), la satisfacción del cliente se considera alcanzable mediante elementos como los deseos, necesidades y expectativas del consumidor. Estos principios se derivan de la Teoría de la Elección del Consumidor, la cual plantea que los compradores tienen un presupuesto limitado para adquirir bienes y servicios. Por ello, al realizar una compra, los consumidores evalúan los precios y adquieren una cantidad de productos que, dentro de sus posibilidades económicas, les permita satisfacer de manera óptima sus deseos y necesidades.
- **Calidad del servicio percibida:** Según Parasumaran, A. y Zeithaml, V. (1993), la calidad del servicio depende de la comparación que hace el cliente entre el nivel de servicio esperado o aceptable y el desempeño real del servicio que percibe. El nivel aceptable es aquel que el cliente considera satisfactorio, basándose en su propia evaluación anticipada de cómo debería ser el servicio.
- **Rendimiento operativo:** Según Mallar, M. (2010), son las prácticas y pasos recomendados que garantizan que el análisis de la eficiencia conduzca a conocimientos prácticos y mejoras

sostenibles.

## **1.6. Teorías**

### **1.6.1. Teoría de Confirmación y Desconfirmación de Expectativas:**

Según Ramos, Unda, y Pantoja (2016), la Teoría de Confirmación y Desconfirmación de Expectativas se refiere a cómo las expectativas que los individuos tienen sobre un resultado influyen en su satisfacción después de experimentar ese resultado. Esta teoría se basa en tres variables fundamentales:

1. Expectativas: Las creencias o suposiciones que una persona tiene antes de experimentar un servicio o producto (en el contexto educativo, esto se refiere a las expectativas previas al inicio de un curso).
2. Desempeño del Producto: Se refiere a la calidad y efectividad del servicio o producto (en educación, esto implica la calidad de la formación impartida).
3. Desconfirmación de Expectativas: Es el proceso de comparar las expectativas iniciales con los resultados finales. Esto puede resultar en:
  - a) Desconfirmación Positiva: Cuando el resultado final es mejor de lo esperado.
  - b) Desconfirmación Negativa: Cuando el resultado final es peor de lo esperado.

Dicha teoría refuerza la idea de que estudiar los datos observables en los vuelos es un buen punto de partida para estudiar la satisfacción que podrían tener los clientes que planeen tomar vuelos similares, generando expectativas específicas.

### **1.6.2. Modelo Servqual:**

El Modelo Servqual es una técnica de investigación comercial que se utiliza para medir la calidad del servicio a través de las expectativas y percepciones de los clientes. Se centra en detectar

las brechas entre lo que los clientes esperan de un servicio y lo que realmente reciben.

”El modelo Servqual se publicó por primera vez en el año 1988, y ha experimentado numerosas mejoras y revisiones desde entonces. El modelo Servqual es una técnica de investigación comercial, que permite realizar la medición de la calidad del servicio, conocer las expectativas de los clientes, y cómo ellos aprecian el servicio” (Matsumoto Nishizawa, R., 2014).

Este modelo incluye cinco dimensiones: fiabilidad, sensibilidad, seguridad, empatía y elementos tangibles, que se utilizan para evaluar la calidad del servicio en diversas industrias.

De manera que se pueden utilizar de manera práctica los resultados del estudio actual, utilizándolos como expectativas o percepciones de los clientes, los cuáles pueden ser trabajados para mejorar la calidad de los servicios de los vuelos.

### **1.6.3. Metodología Kano:**

La metodología Kano es un enfoque utilizado para entender y categorizar las expectativas de los clientes en relación con los productos y servicios. Se basa en el concepto de que no todos los atributos de un producto o servicio tienen el mismo impacto en la satisfacción del cliente. Los diferentes atributos se dividen en cinco categorías:

1. Obligatorios: Elementos que el cliente espera que estén presentes; su ausencia genera insatisfacción.
2. Desempeño: Atributos que influyen directamente en la satisfacción; cuanto mejor se cumplen, mayor es la satisfacción.
3. Atractivos: Características que, si están presentes, sorprenden positivamente al cliente; su ausencia no causa insatisfacción.
4. Indiferentes: Aspectos que no afectan la satisfacción del cliente, sin importar su estado.
5. Inversos: Elementos que, si están presentes, generan insatisfacción.

Según Ramírez Ruiz, C. y Mejía Trejo, J. (2014), se considera que no todos los atributos del producto o servicio poseen el mismo peso ante la perspectiva de los clientes, ya que algunos generan mejores niveles de fidelización.

Por lo tanto, es razonable pensar que cada una de las características de los vuelos tienen efectos diferentes en la satisfacción de las personas, lo que da peso a la pregunta de investigación.

## 1.7. Tabla de datos

Tabla 1.2

Dataframe - Parte 1.

X	id	Gender	Customer.Type	Age	Type.of.Travel	Class
0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus
1	5047	Male	disloyal Customer	25	Business travel	Business
2	110028	Female	Loyal Customer	26	Business travel	Business
3	24026	Female	Loyal Customer	25	Business travel	Business
4	119299	Male	Loyal Customer	61	Business travel	Business

Fuente: Elaboración propia, con datos de (Klein, 2021).

Tabla 1.3

Dataframe - Parte 2.

Flight.Distance	Inflight.wifi.service	Departure.Arrival.time.convenient	Ease.of.Online.booking	Gate.location	Food.and.drink
460	3	4	3	1	5
235	3	2	3	3	1
1142	2	2	2	2	5
562	2	5	5	5	2
214	3	3	3	3	4

Fuente: Elaboración propia, con datos de (Klein, 2021).

Tabla 1.4

Dataframe - Parte 3.

Online.boarding	Seat.comfort	Inflight.entertainment	On.board.service	Leg.room.service	Baggage.handling
3	5	5	4	3	4
3	1	1	1	5	3
5	5	5	4	3	4
2	2	2	2	5	3
5	5	3	3	4	4

Fuente: Elaboración propia, con datos de (Klein, 2021).

Tabla 1.5

Dataframe - Parte 4.

Checkin.service	Inflight.service	Cleanliness	Departure.Delay.in.Minutes	Arrival.Delay.in.Minutes	satisfaction
4	5	5	25	18	neutral or dissatisfied
1	4	1	1	6	neutral or dissatisfied
4	4	5	0	0	satisfied
1	4	2	11	9	neutral or dissatisfied
3	3	3	0	0	satisfied

Fuente: Elaboración propia, con datos de (Klein, 2021).

## 1.8. Resumen de 5 números de las variables cuantitativas y analizar el mismo.

A continuación se presentará una serie de estadísticos importantes respecto a las variables cuantitativas presentes en la base de datos seleccionada, la cual incluye como datos cuantitativos la edad, la distancia del vuelo y el retraso en tiempo de salida y llegada del vuelo, lo que proporcionará una idea general del comportamiento de estas variables.

Tabla 1.6

Resumen de 5 números y media de las variables cuantitativas.

Variables	Mínimo	Q1	Mediana	Q3	Máximo	Promedio
Edad	7	27	40	51	85	39,3797
Distancia del vuelo	31	414	843	1743	4983	1189,448
Retraso en tiempo de salida	0	0	0	12	1592	14,8156
Retraso en tiempo de llegada	0	0	0	13	1584	15,1787

*Fuente: Elaboración propia, con datos de (Klein, 2021).*

En primera instancia, se puede notar en la tabla 1.6 que, con respecto a la edad, se puede visualizar cómo el 75% de los viajes los realizan personas por debajo de los 51 años, con un promedio de edad de viaje de 37 años. Por otro lado, se observa que solo el 25% de los viajes los realizan personas por encima de 51 años.

Con respecto a la distancia, se puede visualizar que el 50% de los vuelos son a distancias menores a 843 millas náuticas, lo cual muestra que la mayoría de las personas recorren distancias cortas. Por otro lado, se nota que la media está por encima de la mediana, lo que indica que hay un poco más de peso en el lado derecho de la distribución, lo cual se evidencia al tener un valor máximo tan alto.

En cuanto al tiempo de retraso en la salida del vuelo, se muestra que el 50% de los vuelos no

#### 1.8. RESUMEN DE 5 NÚMEROS DE LAS VARIABLES CUANTITATIVAS Y ANALIZAR EL MISMO.15

tienen retraso en la hora de salida. Por otro lado, se muestra que el promedio es de 14.81, a pesar de que la mitad de los datos no tiene tiempo de espera y para el 75 % este no es mayor a 12, lo cual sugiere valores muy altos con respecto al tiempo de espera, que afectan un aumento de la media, y se evidencia en el valor máximo tan alto.

Por último, de manera similar a lo que ocurre con el tiempo de retraso en la salida del vuelo, el tiempo de retraso en la llegada muestra que el 50 % de este no tiene tiempo de retraso. Sin embargo, posee una media incluso mayor que el tiempo de salida, lo cual sugiere que, considerando que el 75 % de los datos no posee un tiempo mayor a 12 minutos, hay valores extremos o muy altos que están afectando la media.



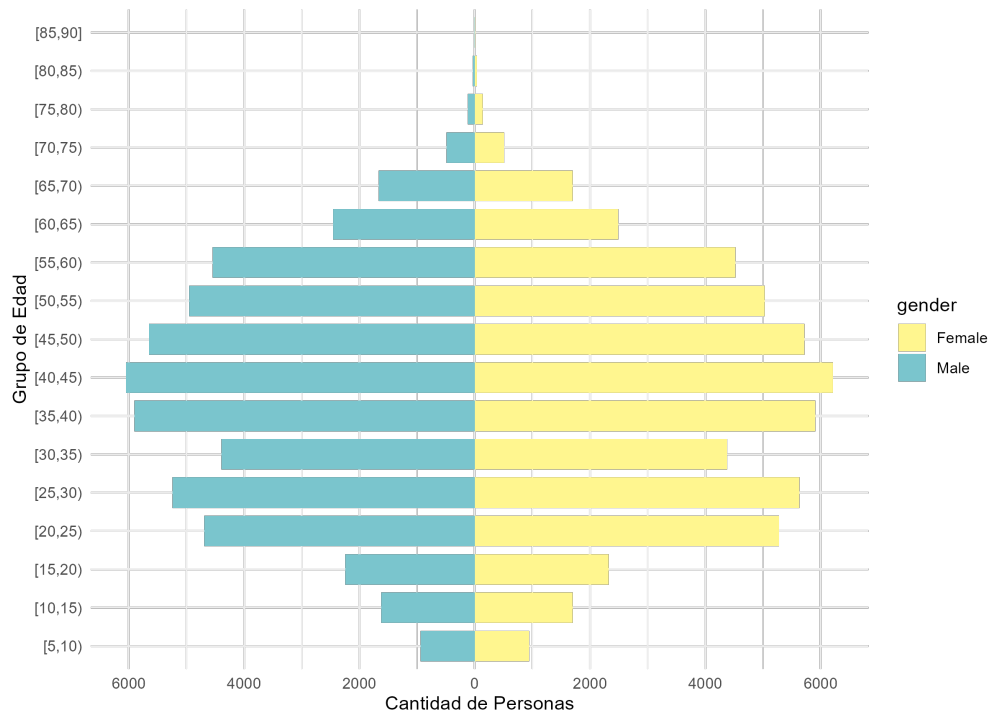
## 1.9. Gráficos

### 1.9.1. Distribución de las variables cuantitativas:

A continuación, se presentará una serie de gráficos que visualizan las variables cuantitativas presentes en la base de datos, entre las que se encuentran la edad, la distancia recorrida y los retrasos del vuelo en la hora de llegada y salida.

Figura 1.1

Pirámide poblacional



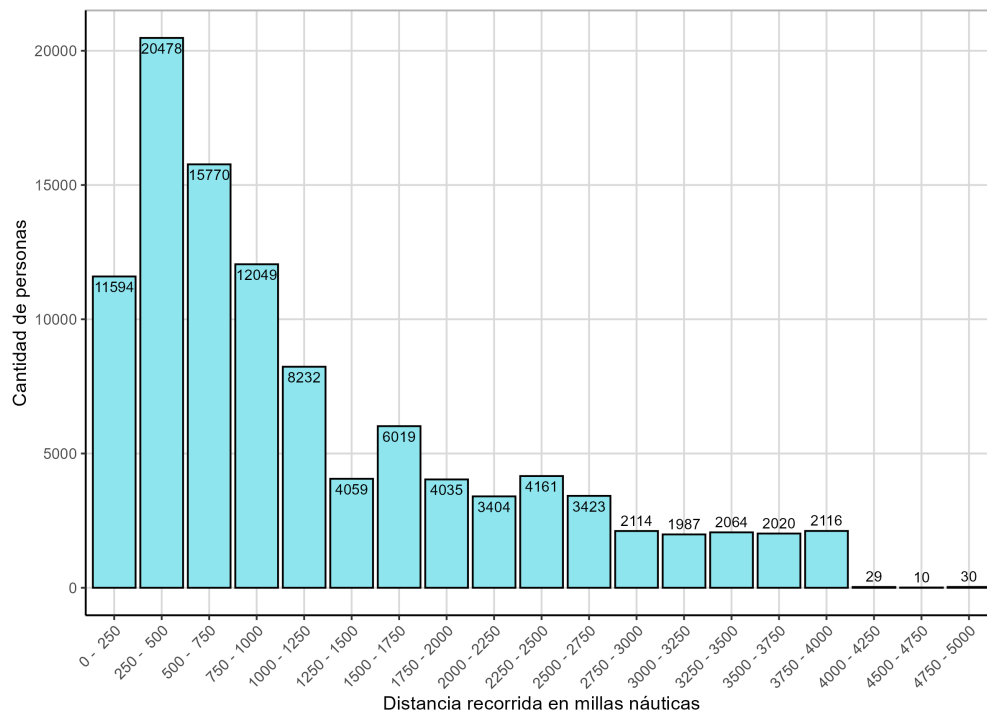
**Fuente:** Elaboración propia, con datos de (Klein, 2021).

Se puede notar que la mayor cantidad de personas en el gráfico 1.1 se encuentra en el rango de edad de 40 a 45 años, tanto para mujeres como para hombres, lo cual tiene mucho sentido debido a los recursos y posibilidades que este rango de edad ofrece, tanto física como económicamente. En comparación con las edades muy jóvenes, donde la posibilidad de viajar depende en su mayoría de

las oportunidades de los padres, se puede visualizar que la población de entre 35 y 50 años es, en general, el grupo con mayor cantidad de vuelos. Por otro lado, como era de esperarse, las edades muy jóvenes de 0 a 20 años tienen una cantidad baja de vuelos, así como las edades muy avanzadas, de entre 65 y 90 años.

Figura 1.2

Distancia recorrida en millas náuticas por persona.

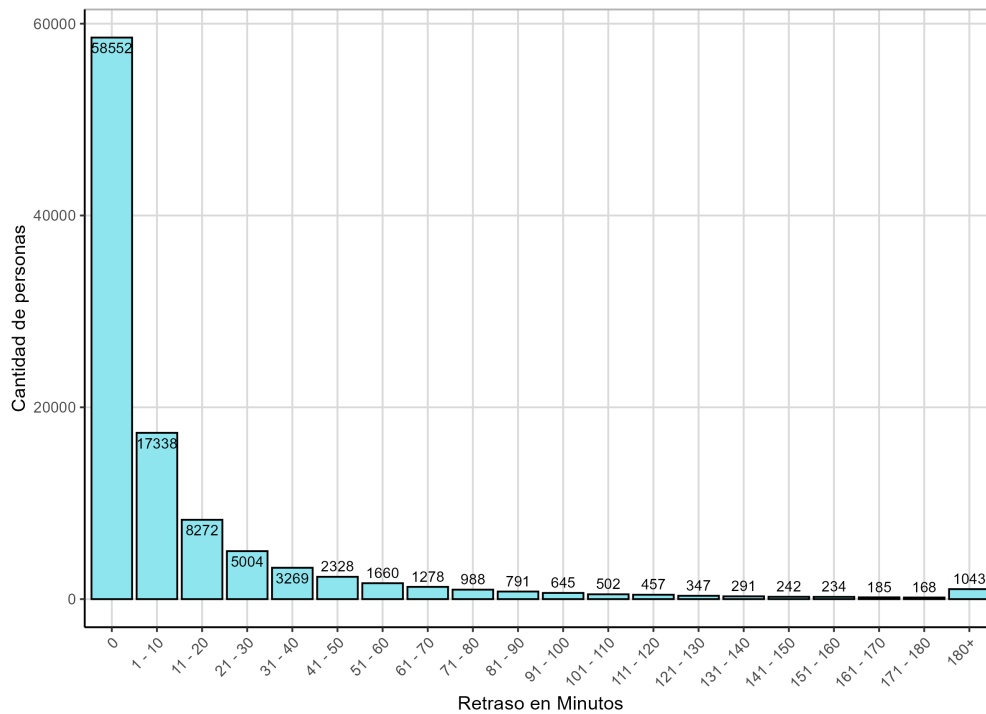


**Fuente:** Elaboración propia, con datos de (Klein, 2021).

Por otro lado, se puede visualizar en 1.9, que la distancia más recorrida es entre 250 y 500 millas náuticas. Asimismo, se puede determinar que la concentración de distancias recorridas se encuentra en el rango de 0 a 1250 millas náuticas, lo que muestra que los vuelos de distancias cortas son los más concurridos y frecuentes. Por otro lado, los vuelos de distancias medianas tienen una discrepancia muy considerable con respecto a la cantidad de personas que recorren poca distancia en vuelos, y por último, los vuelos que recorren mucha distancia son elegidos por pocas personas.

Figura 1.3

Retraso en hora de salida del vuelo en minutos por persona.

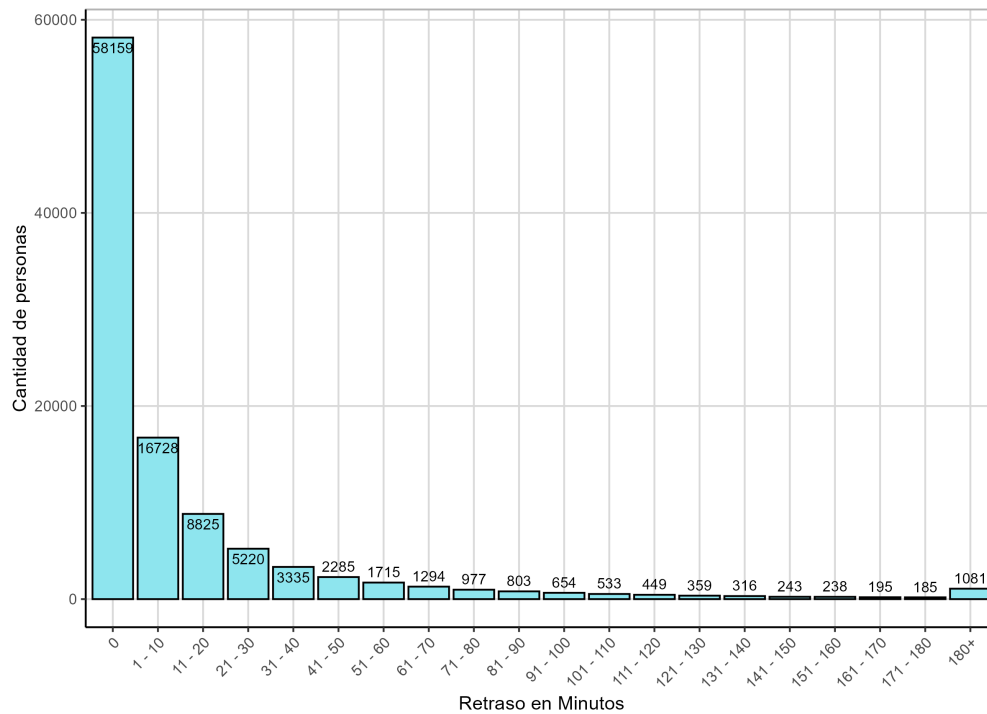


**Fuente:** Elaboración propia, con datos de (Klein, 2021).

Además, a partir de 1.3, se puede visualizar cómo con respecto al retraso en la hora de salida de los vuelos en minutos, un valor muy alto de estos corresponde a un retraso de 0 minutos, lo cual implica que no hubo retraso. Esto coincide con lo visto en 1.6, donde se muestra que el 50% de los vuelos no presenta retraso en salida o que su retraso es 0. Por otro lado, se puede visualizar que la mayor concentración del retraso se presenta en el intervalo de 1 a 40 minutos.

Figura 1.4

Retraso en hora de llegada del vuelo en minutos por persona.



**Fuente:** Elaboración propia con datos de (Klein, 2021)

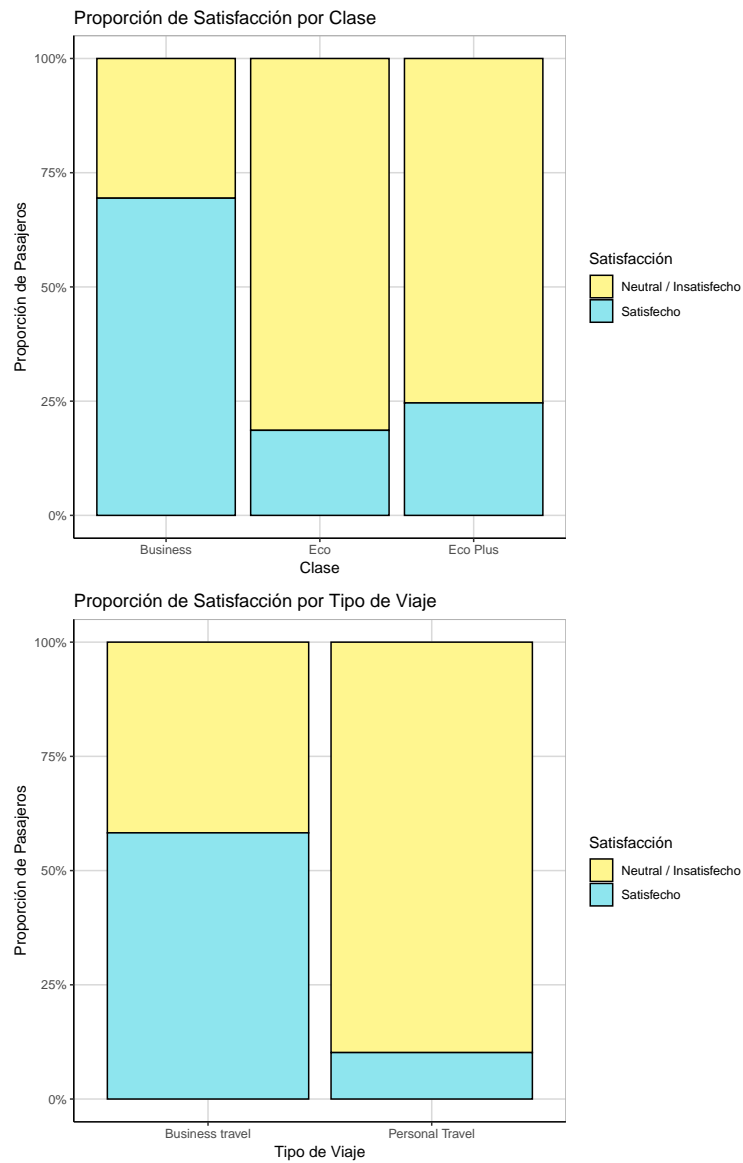
Con respecto al tiempo de llegada visto en 1.4, se puede visualizar que, al igual que el tiempo de salida, más del 50 % de los vuelos no presentan retraso en el tiempo de llegada. Por otro lado , se observa que en los tiempos de llegada positivos, estos se concentran en 1-70 minutos, mientras que, en los tiempos de llegada largos, se reduce la cantidad de personas.

### 1.9.2. Relación entre las variables:

A continuación, se presentará una serie de gráficos que visualizan las relaciones existentes entre la satisfacción del pasajero y diferentes factores clave del vuelo, en donde se revela que en los casos en donde la mayoría de clientes se encuentran satisfechos es en los viajes de trabajo.

Figura 1.5

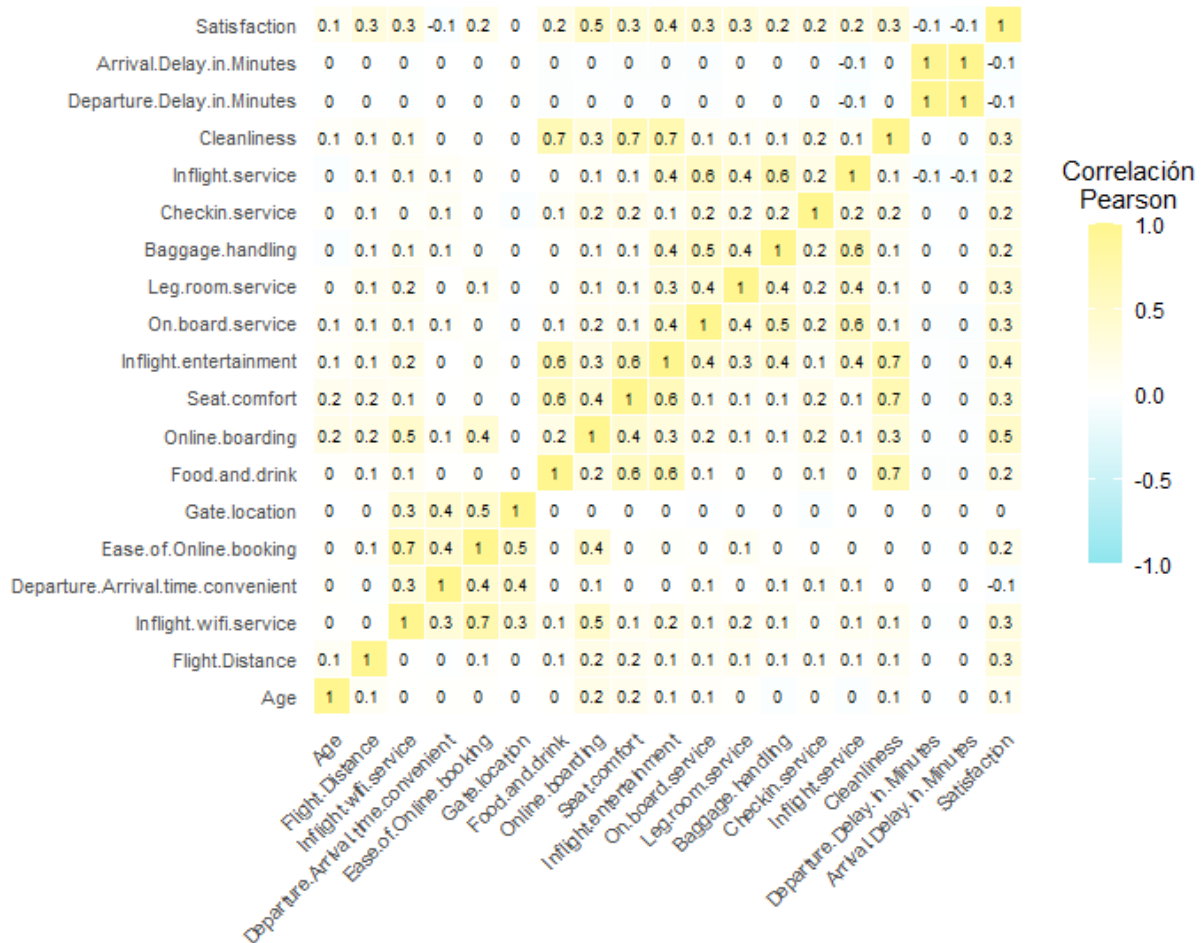
Proporción de satisfacción por Clase

**Fuente:** Elaboración propia con datos de (Klein, 2021)

Además, se presenta un mapa de calor en donde se muestra la correlación de Pearson entre las variables de la base de datos:

Figura 1.6

Correlación de Pearson entre las variables



**Fuente:** Elaboración propia con datos de (Klein, 2021)

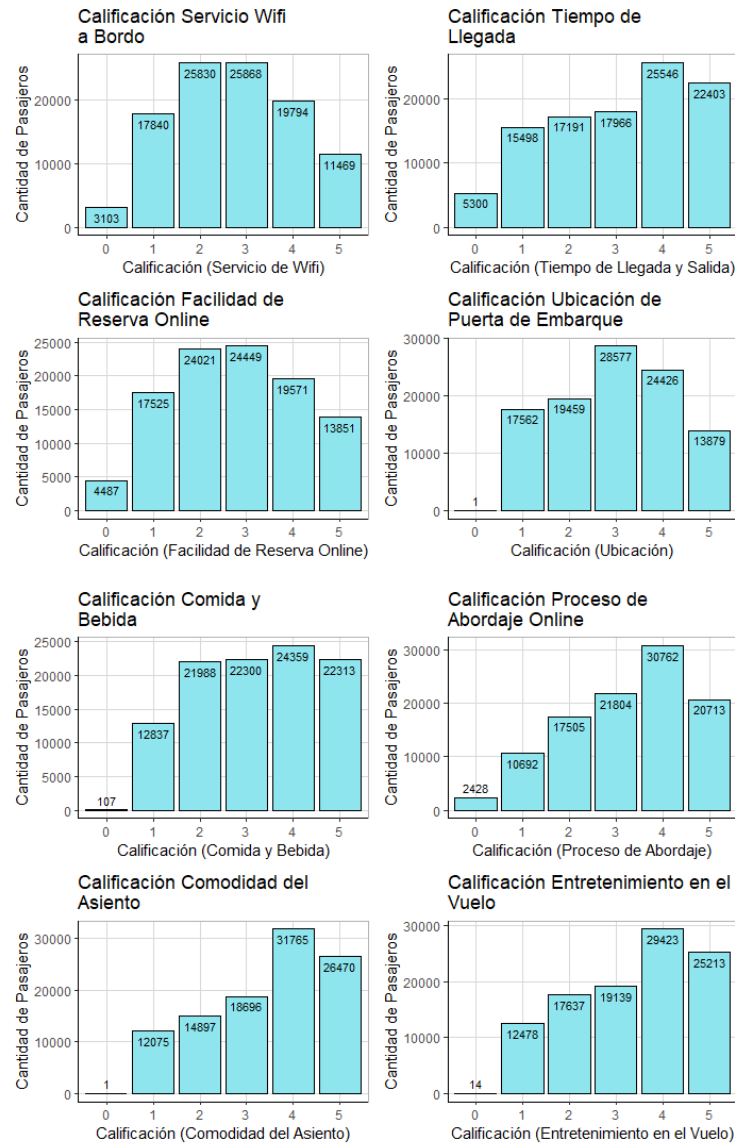
En donde podemos notar que la mayoría de las relaciones son positivas, lo cuál tiene sentido debido a la naturaleza de las variables, ya que mientras mayor sea su calificación, mayor será la satisfacción del cliente, con la excepción de los atrasos en minutos, que mientras mayor es el atraso, menor será la satisfacción.

### 1.9.3. Distribución de las variables categóricas:

A continuación se verán las distribuciones de las variables categóricas, reflejando la cantidad de veces que un servicio recibió una calificación específica:

Figura 1.7

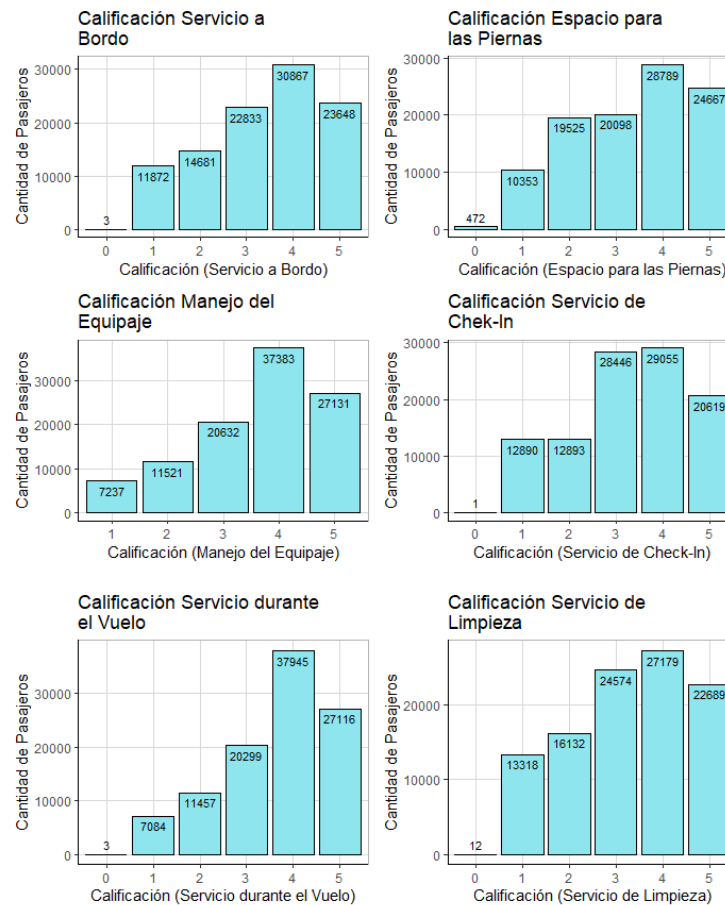
Distribución de calificaciones para diferentes servicios.



**Fuente:** Elaboración propia con datos de (Klein, 2021)

Figura 1.8

Distribución de calificaciones para diferentes servicios.



**Fuente:** Elaboración propia con datos de (Klein, 2021)

Se puede observar que la mayoría de las distribuciones muestran una tendencia hacia calificaciones altas (4 y 5), lo que sugiere una satisfacción general relativamente buena en varios de los servicios evaluados.

La mayoría de los servicios evaluados reciben calificaciones altas de forma consistente (distribuciones asimétricas a la izquierda). Sin embargo, algunos como el servicio Wifi, destacan por tener calificaciones más bajas y divididas (distribución bimodal/multimodal), y la comida y bebida



muestran una mayor dispersión en las opiniones, aunque con tendencia a la satisfacción.

## 1.10. Valores faltantes y posibles outliers

### 1.10.1. Identificar valores faltantes y posibles outliers

Aquí se muestra un fragmento de código R para verificar valores faltantes:

```
1 # Ver si hay valores faltantes en el conjunto de datos
2 missing_values <- is.na(data)
3
4 # Ver la cantidad de valores faltantes por columna
5 colSums(missing_values)
```

Del cual se logra identificar que no existen datos faltantes, a excepción de la variable "Arrival Delay in Minutes", la cual cuenta con 310 registros faltantes.

Para las variables categóricas, se analiza como se muestra en este fragmento de código R para trabajar con las variables no numéricas, identificar posibles outliers y encontrar valores faltantes:

```
1 # Una forma de poder identificar outliers
2 table_results <- lapply(no_numeric_data, table)
3
4 # Mostrar resultados de la tabla
5 table_results
6
7 # Encontrar valores faltantes
8 na_results <- lapply(no_numeric_data, function(x) table(is.na(x)))
9
10 # Mostrar resultados de valores faltantes
11 na_results
```

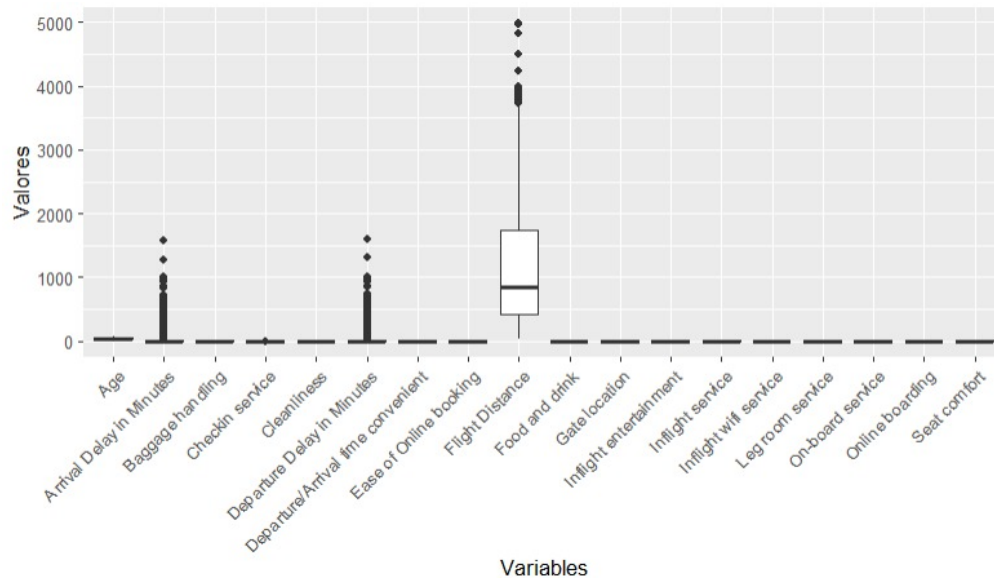
Lo cual indica que no existen posibles outliers y no hay datos faltantes.

El boxplot es una forma sencilla y visual de identificar los outliers. Los valores que se encuentran fuera de los "bigotes" del boxplot se consideran outliers.

A continuación, se presenta un gráfico de tipo boxplot:

Figura 1.9

Boxplot de las variables.



**Fuente:** Elaboración propia, con datos de (Klein, 2021).

Donde se pueden apreciar los posibles outliers principalmente en la variable de Flight Distance. Pero, dependiendo de la interpretación, estos datos de esta variable en particular, muestra los vuelos con las distancias más grandes de toda la base de datos.

### 1.10.2. Técnicas que permitan subsanar los valores perdidos y outliers

Existen diversas maneras de manejar los valores faltantes. Dependiendo del tipo de variable (numérica o categórica) y la cantidad de valores faltantes, puedes optar por diferentes enfoques.

Si la cantidad de datos faltantes es pequeña, puedes eliminar las filas o columnas con valores faltantes. Sin embargo, si la proporción de datos faltantes es alta, este enfoque puede generar un

sesgo en los datos, y puede resultar mejor realizar una imputación de datos mediante medias, k-means, regresión, interpolación o extrapolación.

Por otro lado, los outliers son observaciones que se desvían significativamente del resto de los datos. Detectarlos y manejarlos adecuadamente es crucial para evitar que distorsionen los resultados del análisis. Por ello, las técnicas que permiten subsanar estos valores son los siguientes:

**Eliminación de outliers:** Si los outliers son errores de medición o datos claramente erróneos, puedes eliminar esas observaciones del conjunto de datos.

**Reemplazo de outlier:** Puedes reemplazar los outliers por un valor más razonable, como la media o mediana de la variable.

**Winsorización:** Consiste en reemplazar los valores más extremos por un percentil más cercano. Este método es útil cuando no quieres perder información al eliminar los outliers, pero tampoco quieres que afecten excesivamente el análisis.

**Transformaciones de datos:** Algunas veces, los outliers pueden ser atípicos debido a la escala de los datos. En este caso, puedes aplicar transformaciones como logaritmo, raíz cuadrada o Box-Cox para reducir el impacto de los outliers.

# **Apéndice A**

## **Repositorio**

Se puede encontrar el repositorio en el siguiente enlace:

`https://github.com/leitondiana/ProyectoCA0411.git`

## Referencias

- Klein, T. (2021). *Airline passenger satisfaction*. <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>. (Accessed: 2025-04-08)
- Mallar, M. (2010). La gestión por procesos: un enfoque de gestión eficiente. *Revista Científica Visión del Futuro*.
- Matsumoto Nishizawa, R. (2014). Desarrollo del modelo servqual para la medición de la calidad del servicio en la empresa de publicidad ayuda experto. *PERSPECTIVAS*(34), 181–209.
- Parasumaran, A., y Zeithaml, V. (1993). Calidad total en la gestión de servicios.
- Ramos, V., Unda, X., y Pantoja, O. (2016). La teoría de confirmación - desconfirmación expectativas para gestionar la motivación en el aula. *Congresos CLABES*. Descargado de <https://revistas.utp.ac.pa/index.php/clabes/article/view/1409>
- Ramírez, J. R. (2020). *Tests de equivalencia para comparar medias* (Tesis Doctoral no publicada). Universidades de Santiago de Compostela.
- Ramírez Ruiz, C., y Mejía Trejo, J. (2014). Metodología kano para el diseño de un sitio web de turismo de aventura. *Mercados y Negocios*, 15(1), 92–110.