

ANÁLISIS PREDICTIVO

# TRATAMIENTO DE DATASET, ANÁLISIS Y GRÁFICOS

Sofía González del Solar

# DATASET

## AIRLINE\_PASSENGER\_SATISFACTION

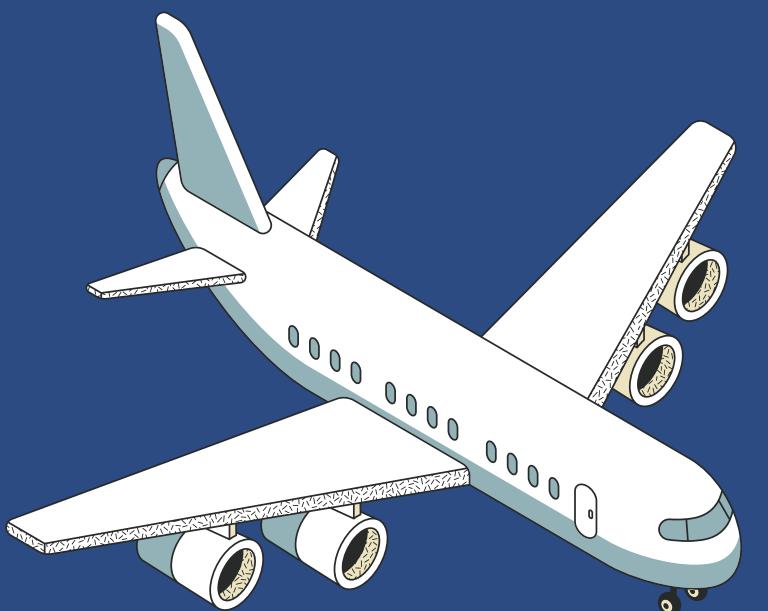
- Cantidad de Variables Totales: 23
- Cantidad de variables numéricas: 4
- Cantidad de variables categoricas: 19
- Cantidad de Registros: 103.905
- Origen del Dataset: Kaggle
- Frecuencia de Actualización: nunca

Gender	Customer Type	Age	Type of Travel Class	
Male	Loyal Customer	13	Personal Travel	Eco Plus
Male	disloyal Customer	25	Business travel	Business
Female	Loyal Customer	26	Business travel	Business
Female	Loyal Customer	25	Business travel	Business
Male	Loyal Customer	61	Business travel	Business
Female	Loyal Customer	26	Personal Travel	Eco
Male	Loyal Customer	47	Personal Travel	Eco
Female	Loyal Customer	52	Business travel	Business
Female	Loyal Customer	41	Business travel	Business
Male	disloyal Customer	20	Business travel	Eco
Female	disloyal Customer	24	Business travel	Eco
Female	Loyal Customer	12	Personal Travel	Eco Plus
Male	Loyal Customer	53	Business travel	Eco
Male	Loyal Customer	33	Personal Travel	Eco
Female	Loyal Customer	26	Personal Travel	Eco
Male	disloyal Customer	13	Business travel	Eco
Female	Loyal Customer	26	Business travel	Business
Male	Loyal Customer	41	Business travel	Business
Female	Loyal Customer	45	Business travel	Business
Male	Loyal Customer	38	Personal Travel	Eco
Male	Loyal Customer	9	Business travel	Eco
Female	Loyal Customer	17	Personal Travel	Eco
Female	Loyal Customer	43	Personal Travel	Eco
Female	Loyal Customer	58	Personal Travel	Eco
Female	disloyal Customer	23	Business travel	Eco
Male	Loyal Customer	57	Personal Travel	Eco
Female	Loyal Customer	22	Personal Travel	Eco

# CASO DE NEGOCIO

Se busca predecir la variable categórica llamada satisfacción.

El objetivo es que las aerolíneas puedan predecir si el cliente quedó satisfecho o neutral/desatisfecho para así poder mejorar el servicio o remediar una mala experiencia del cliente.



# ANÁLISIS DE MISSINGS

- Arrival Delay in Minutes: 310



Departure Delay in Minutes	Arrival Delay in Minutes	Atraso/Adelanto
25	18	7
1	6	-5
0	0	0
11	9	2
0	0	0
0	0	0
9	23	-14
4	0	4
0	0	0
0	0	0
0	0	0
0	0	0
28	8	20
0	0	0
43	35	8
1	0	1
49	51	-2
0	10	-10
7	5	2
17	18	-1
0	4	-4

Se agregó una variable



Promedio = -0.319

# ANÁLISIS DE OUTLIERS

Flight Distance: 2.291

Departure Delay: 14.529

Arrival Delay: 13.954



Flight.Distance
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4983
4963
4963
4963
4963
4963
4963
4963
4963
4963

Departure.Delay.in.Minutes	Arrival.Delay.in.Minutes
1592	1584
1305	1280
1017	1011
978	970
933	920
930	952
921	924
859	860
853	823
750	729
748	720
729	717
726	691
724	705
692	702
652	638
626	604
610	593

# VARIABLES CATEGÓRICAS



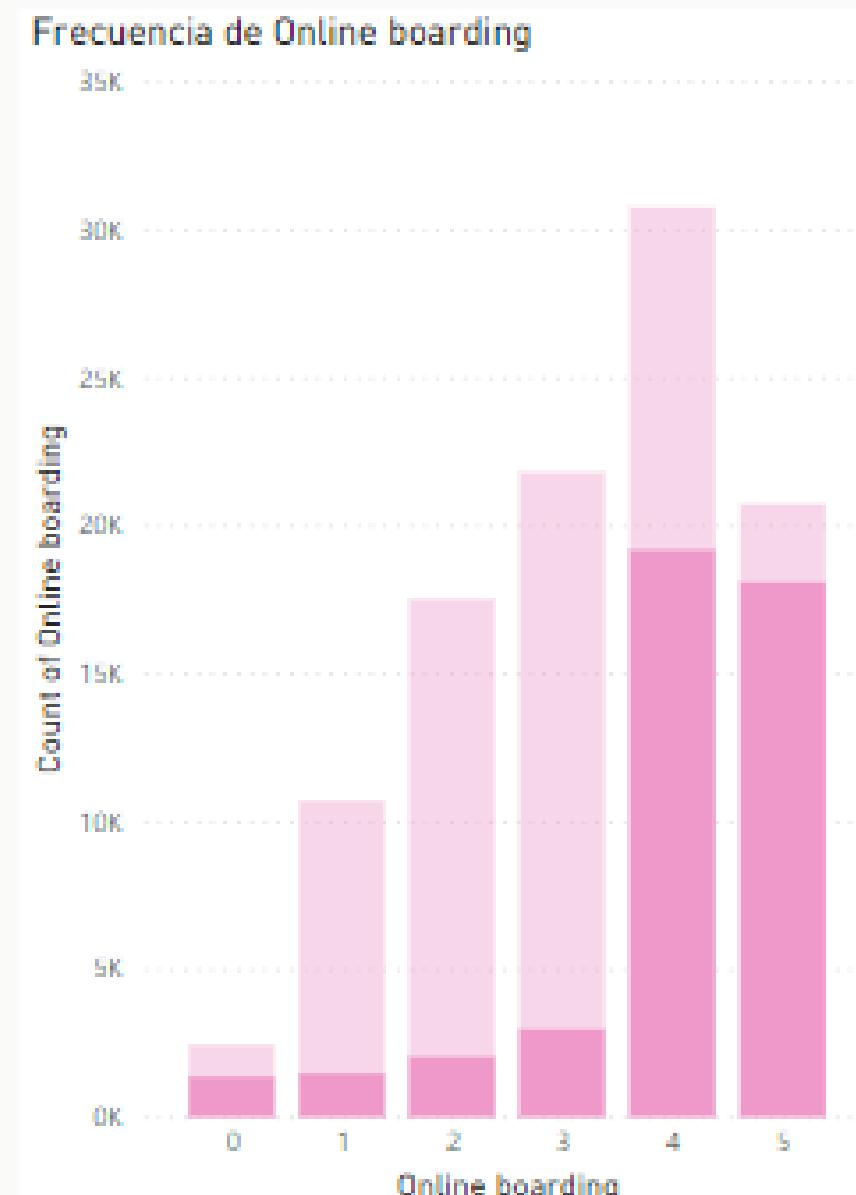
Dashboard Variables categóricas

# Ease of Online Boarding y Satisfaction

Tamaño del efecto

`correlation`

[1] 0.6185259



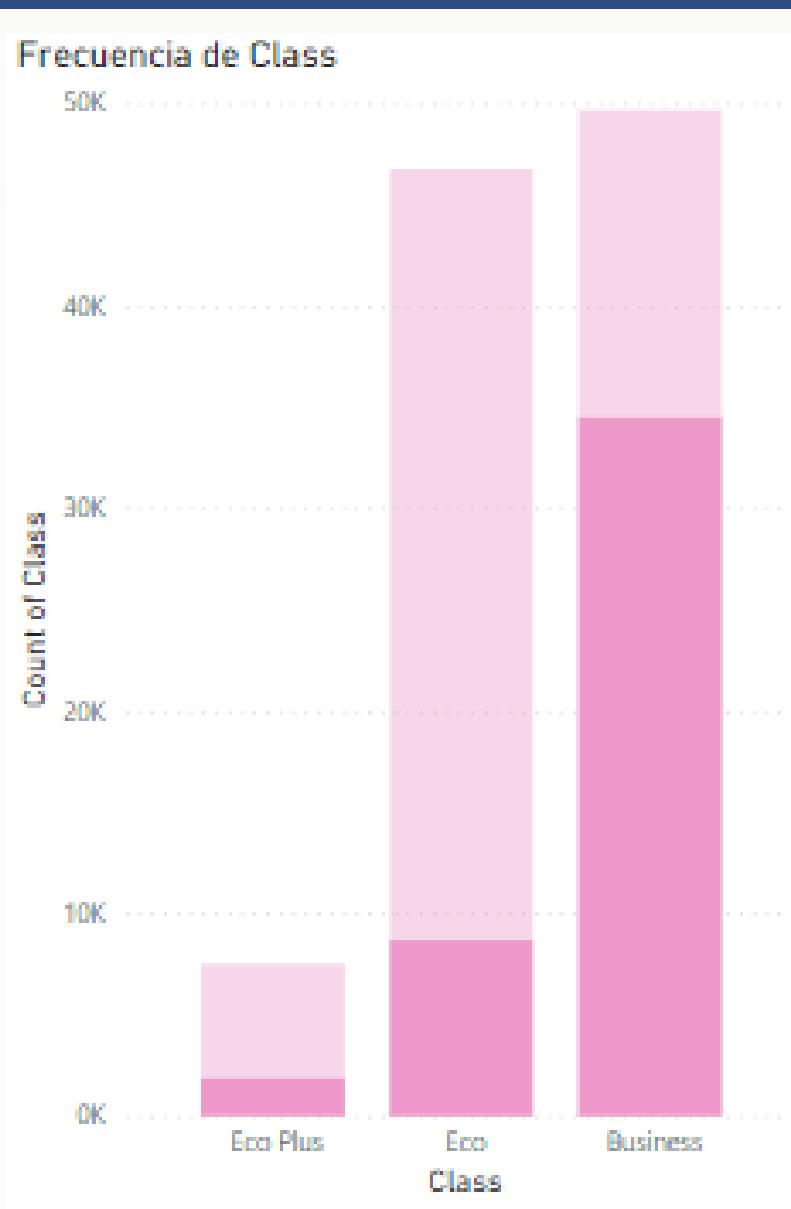
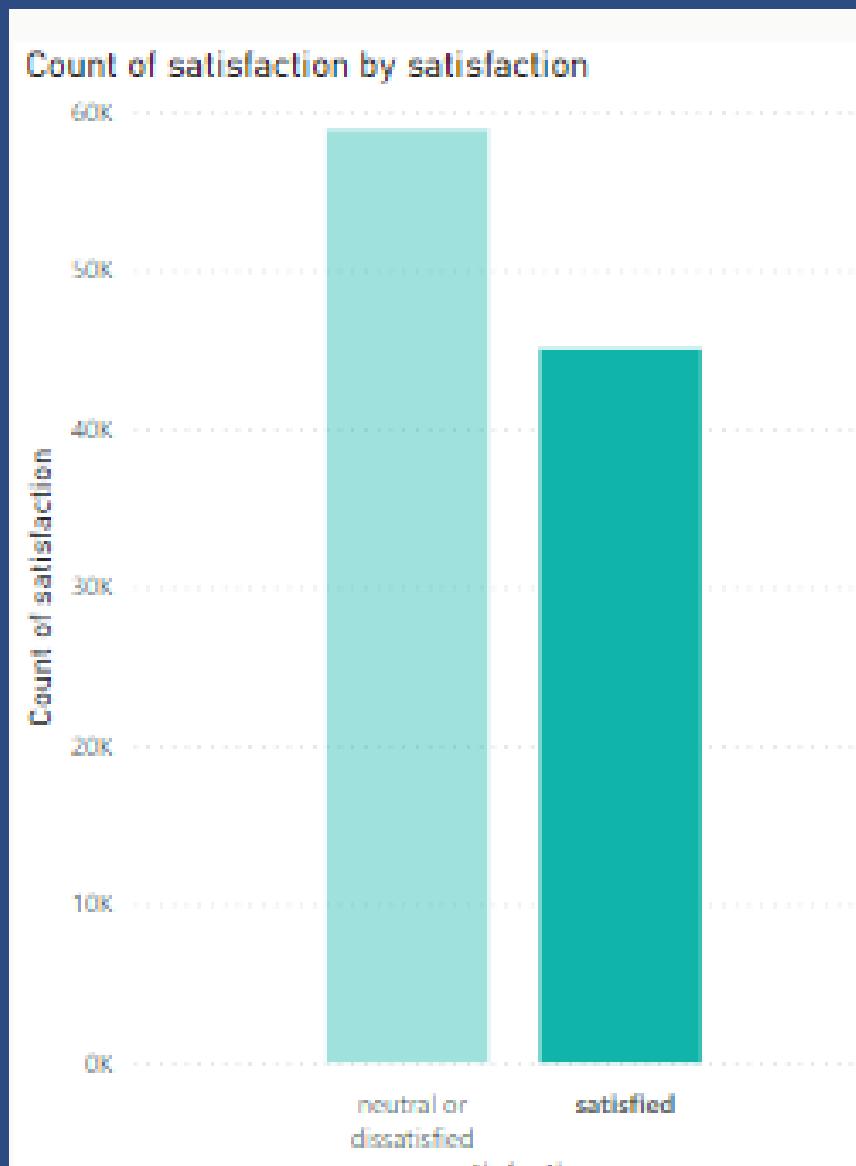
Online boarding	neutral or dissatisfied	satisfied
0	1077	1351
1	9219	1473
2	15486	2019
3	18845	2959
4	11596	19166
5	2656	18057



# Class y Satisfaction

Tamaño del efecto

[1] 0.5047498



Class	neutral or dissatisfied	satisfied
Business	15185	34480
Eco	38044	8701
Eco Plus	5650	1844



# Class y Type of Travel

Tamaño del efecto



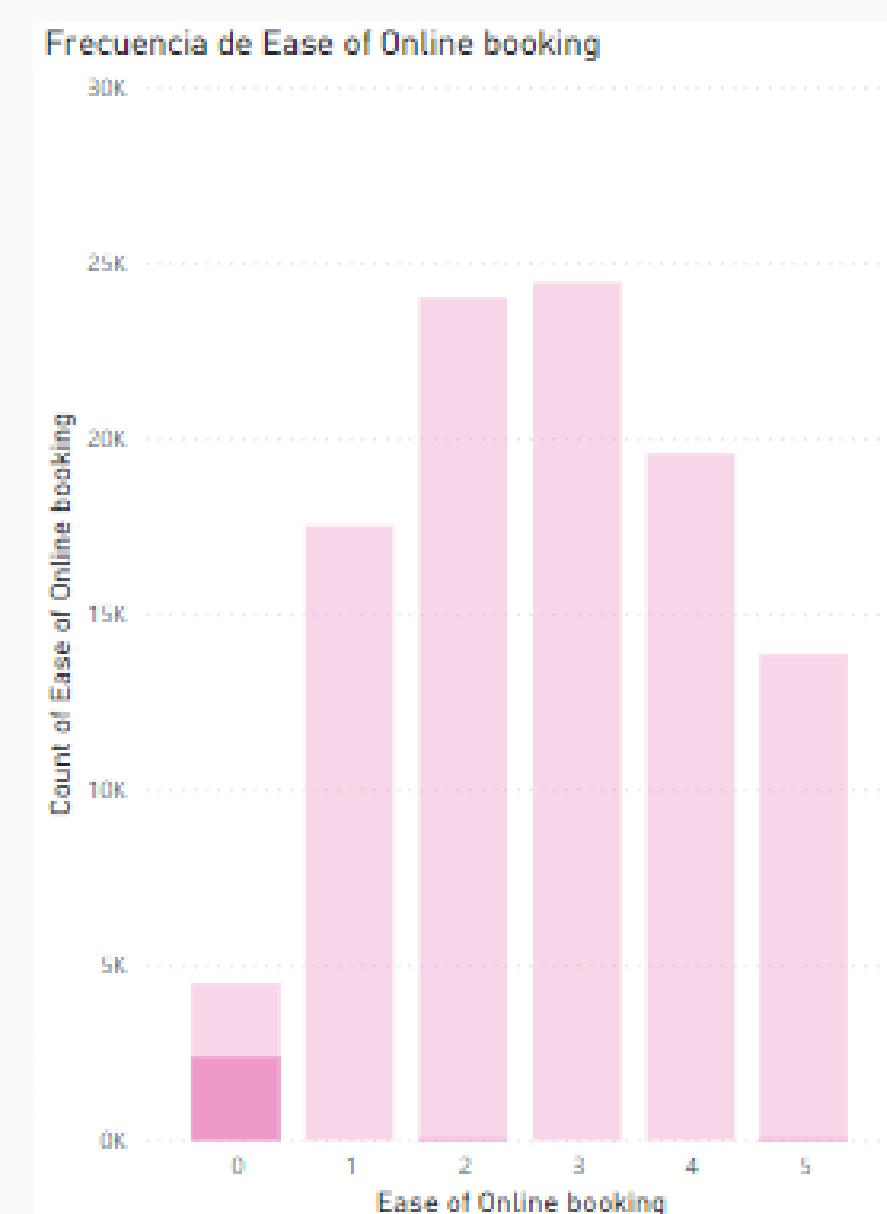
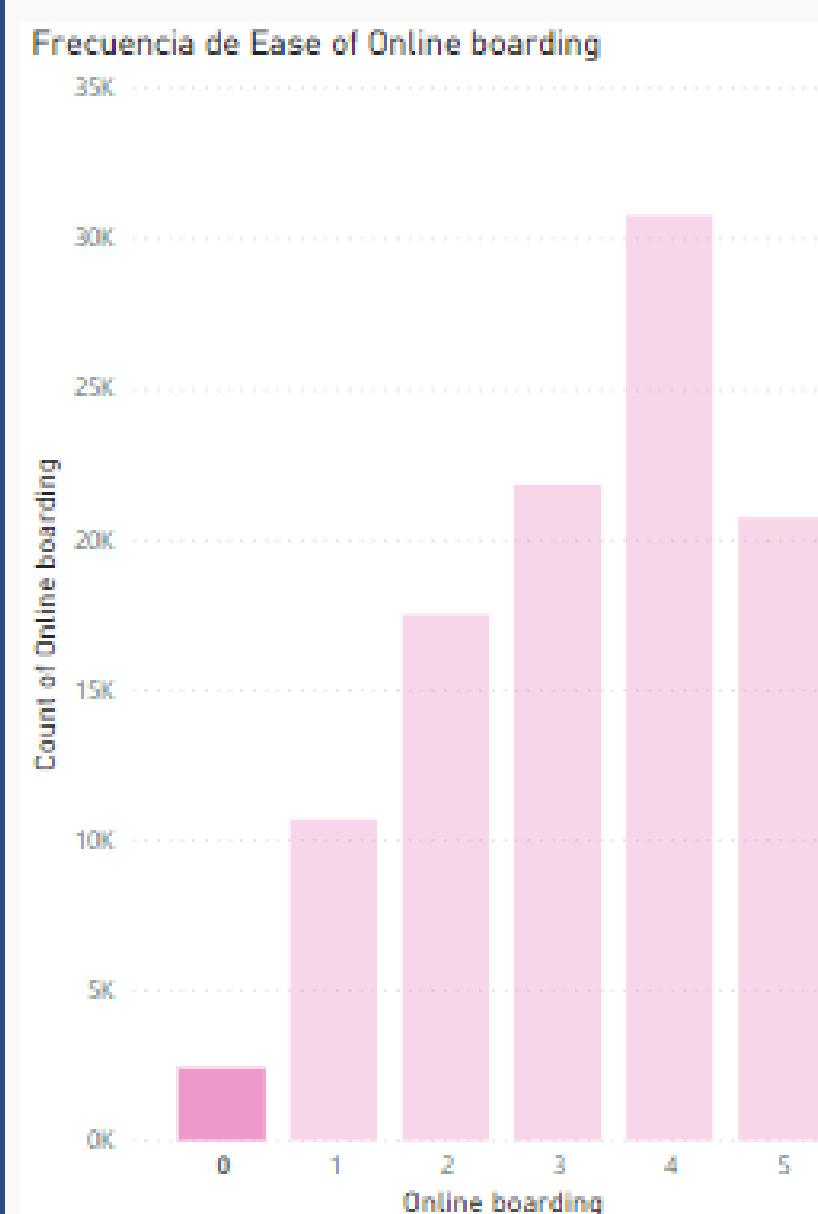
Class	Business travel	Personal Travel
Business	47508	2157
Eco	20257	26488
Eco Plus	3890	3604



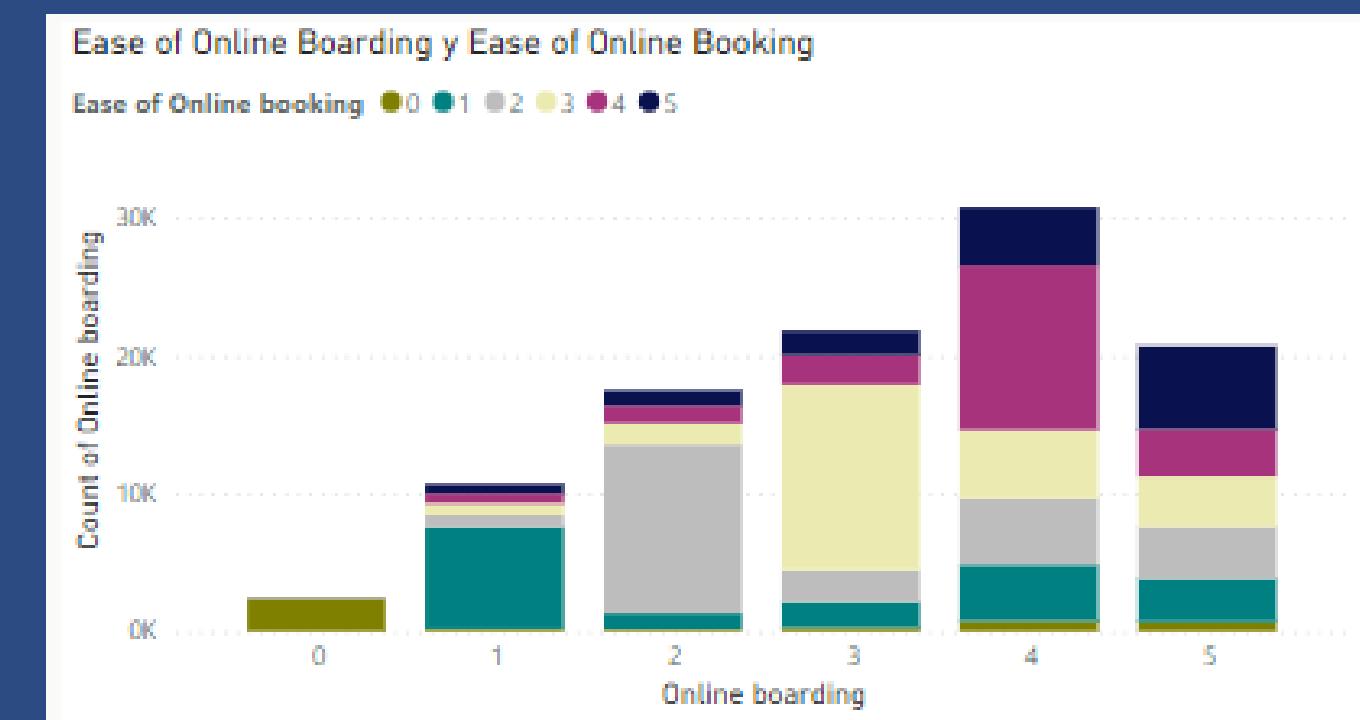
# Ease of Online Boarding y Ease of Online Booking

Tamaño del efecto

[1] 0.4897872



Online boarding	0	1	2	3	4	5
0	2387	190	7361	870	864	739
1	190	7361	870	864	739	668
2	7361	870	12230	1516	1305	1127
3	870	1516	12230	13424	2157	1766
4	864	13424	2157	290	1843	2324
5	739	2157	1766	290	1843	2324
6	668	1127	1305	1516	2157	290



# VARIABLES NUMÉRICAS



Dashboard Variables Numéricas

# Métricas



Departure Delay in Minutes

Mínimo	Máximo	Varianza	Mediana	Promedio
0	1592	1461.59	0	14.82

Arrival Delay in Minutes

Mínimo	Máximo	Varianza	Mediana	Promedio
0	1584	1497.57	0	15.18

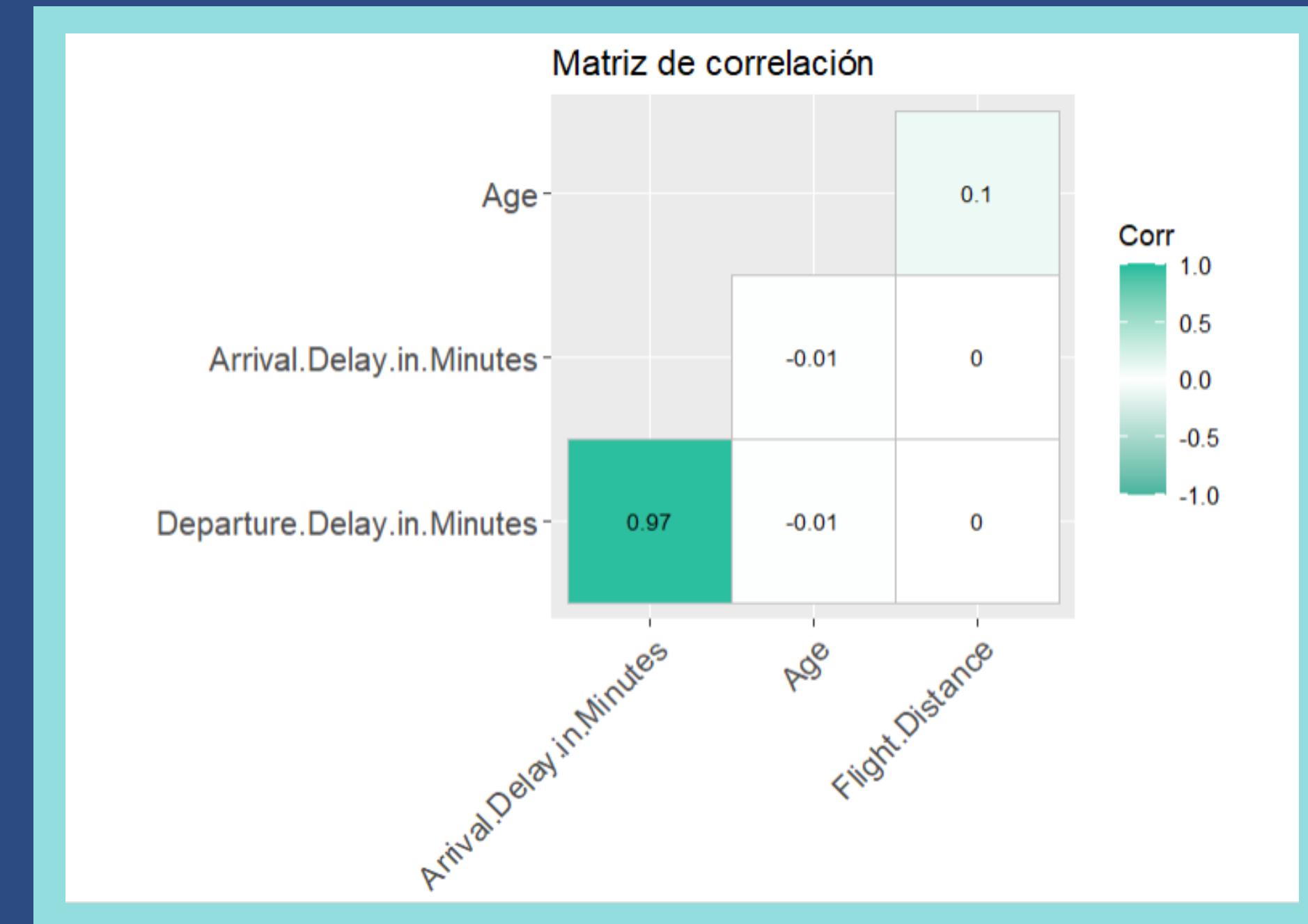
Age

Mínimo	Máximo	Varianza	Mediana	Promedio
7	85	228.46	40	39.38

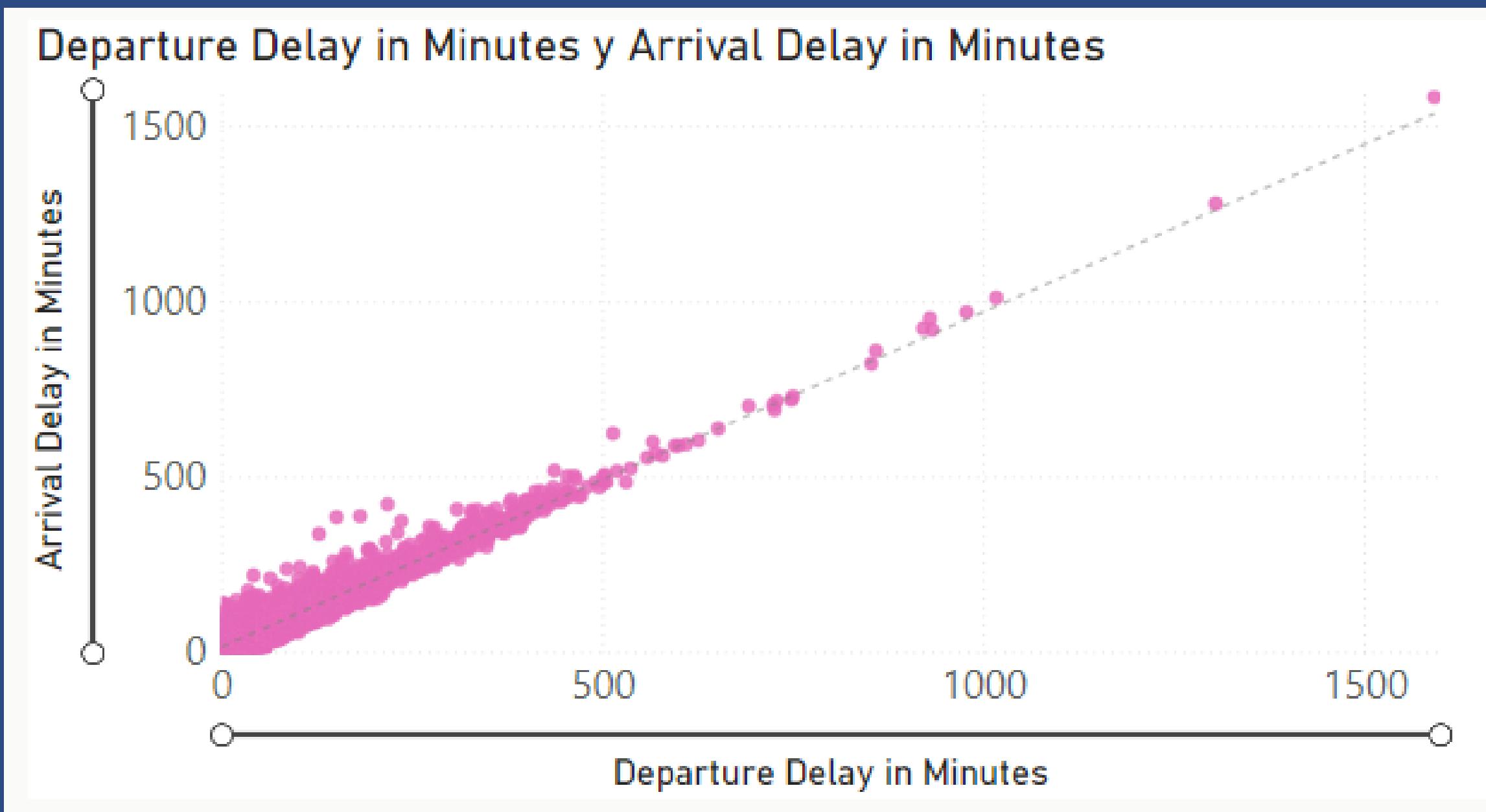
Flight Distance

Mínimo	Máximo	Varianza	Mediana	Promedio
31	4983	994293.13	843	1189.45

# Matriz de correlaciones



# Departure Delay in Minutes y Arrival Delay in Minutes

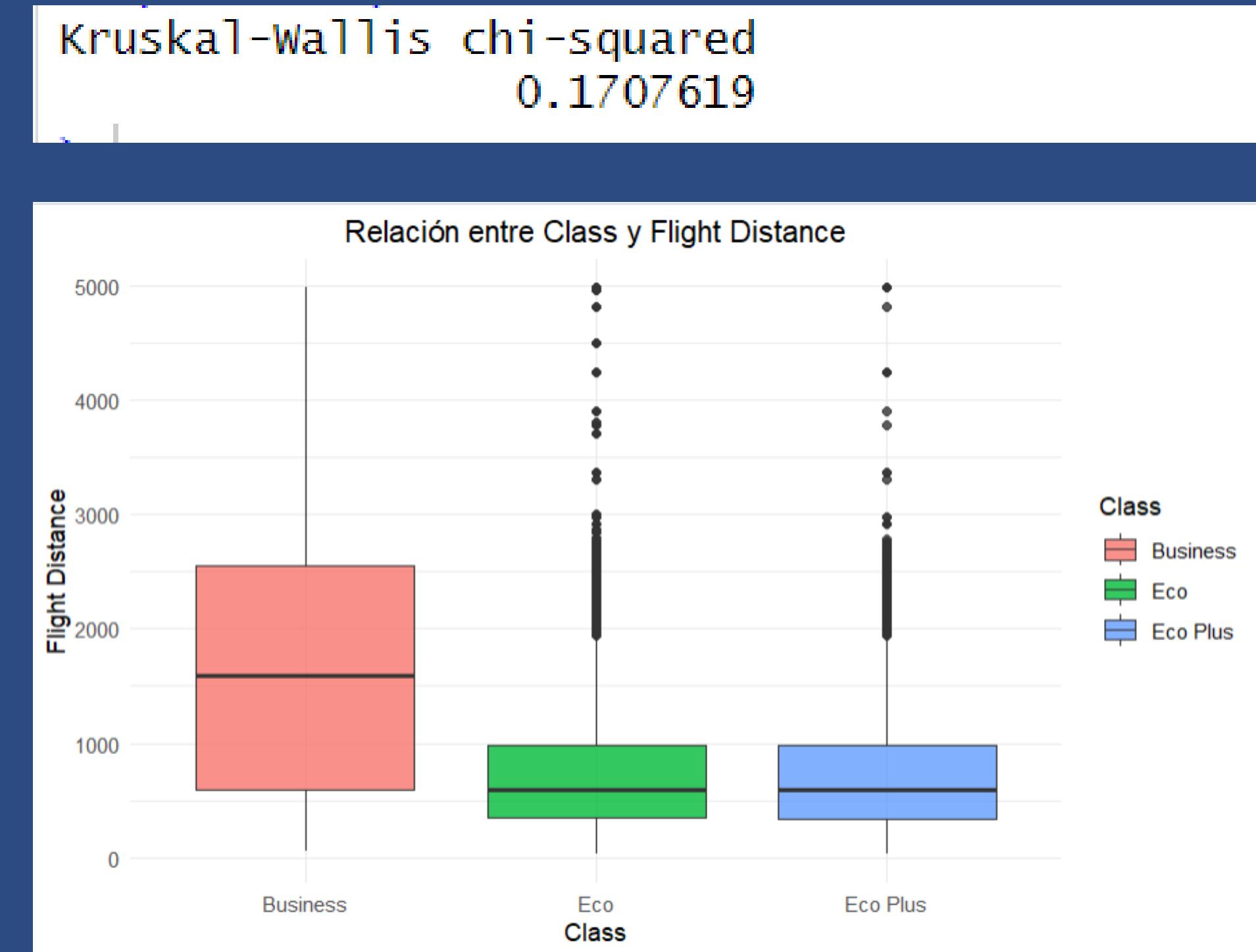


```
> correlacion  
[1] 0.9654809
```

# VARIABLES NÚMÉRICAS Y CATEGÓRICAS



# Flight distance y Class



# Conclusiones



Profesores:

- Ezequiel Martín Eliano Sombory
- Leonardo Andrés Caravaggio
- Francisco Valentini



1er cuatrimestre - 2023

Primer Examen



Sofía González del Solar