ANALISIS PREDICTIVO

TRATAMIENTO DE DATASET, ANÁLISIS Y GRÁFICOS

Sofía González del Solar

|	A	B	C	D	E		Q	R
1	Gender	Customer Type	Age	Type of Trave	Class		Baggage hand	Checkin
2	Male	Loyal Customer	13	Personal Trav	Eco Plus		4	
3	Male	disloyal Customer	25	Business trave	Business		3	
4	Female	Loyal Customer	26	Business trave	Business		4	
5	Female	Loyal Customer	25	Business trave	Business		3	
6	Male	Loyal Customer	61	Business trave	Business		4	
7	Female	Loyal Customer	26	Personal Trav	Eco		4	
8	Male	Loyal Customer	47	Personal Trav	Eco		4	
9	Female	Loyal Customer	52	Business trave	Business		5	
10	Female	Loyal Customer	41	Business trave	Business		1	
11	Male	disloyal Customer	20	Business trave	Eco		4	
12	Female	disloyal Customer	24	Business trave	Eco		5	
13	Female	Loyal Customer	12	Personal Trav	Eco Plus		5	
14	Male	Loyal Customer	53	Business trave	Eco		3	
15	Male	Loyal Customer	33	Personal Trav	Eco		2	
16	Female	Loyal Customer	26	Personal Trav	Eco		2	
17	Male	disloyal Customer	13	Business trave	Eco		4	
18	Female	Loyal Customer	26	Business trave	Business		4	
19	Male	Loyal Customer	41	Business trave	Business		5	
20	Female	Loyal Customer	45	Business trave	Business		5	
21	Male	Loyal Customer	38	Personal Trav	Eco		4	
22	Male	Loyal Customer	9	Business trave	Eco		3	
23	Female	Loyal Customer	17	Personal Trav	Eco		3	
24	Female	Loyal Customer	43	Personal Trav	Eco		5	
25	Female	Loyal Customer	58	Personal Trav	Eco		4	
26	Female	disloyal Customer	23	Business trave	Eco		5	
27	Male	Loyal Customer	57	Personal Trav	Eco		4	

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

Airline_Passenger_Satisfaction

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.
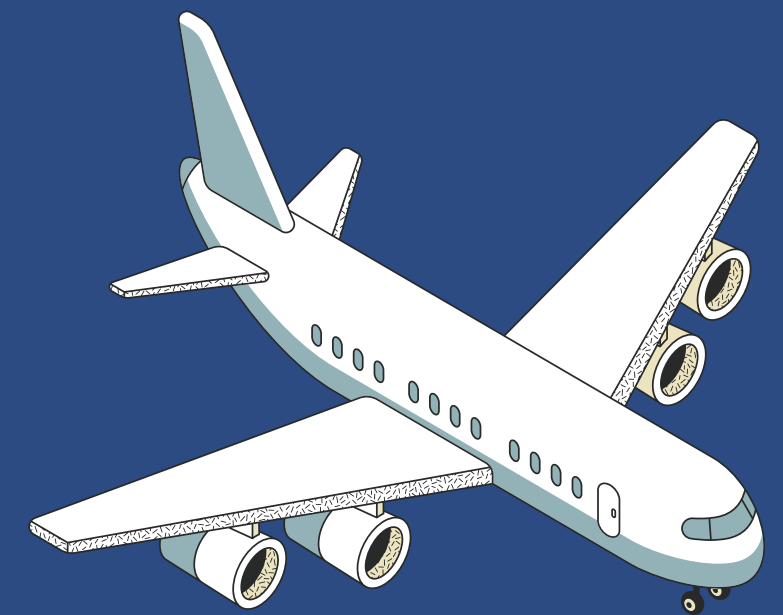
ANÁLISIS DE MISSINGS

- 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
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](#)

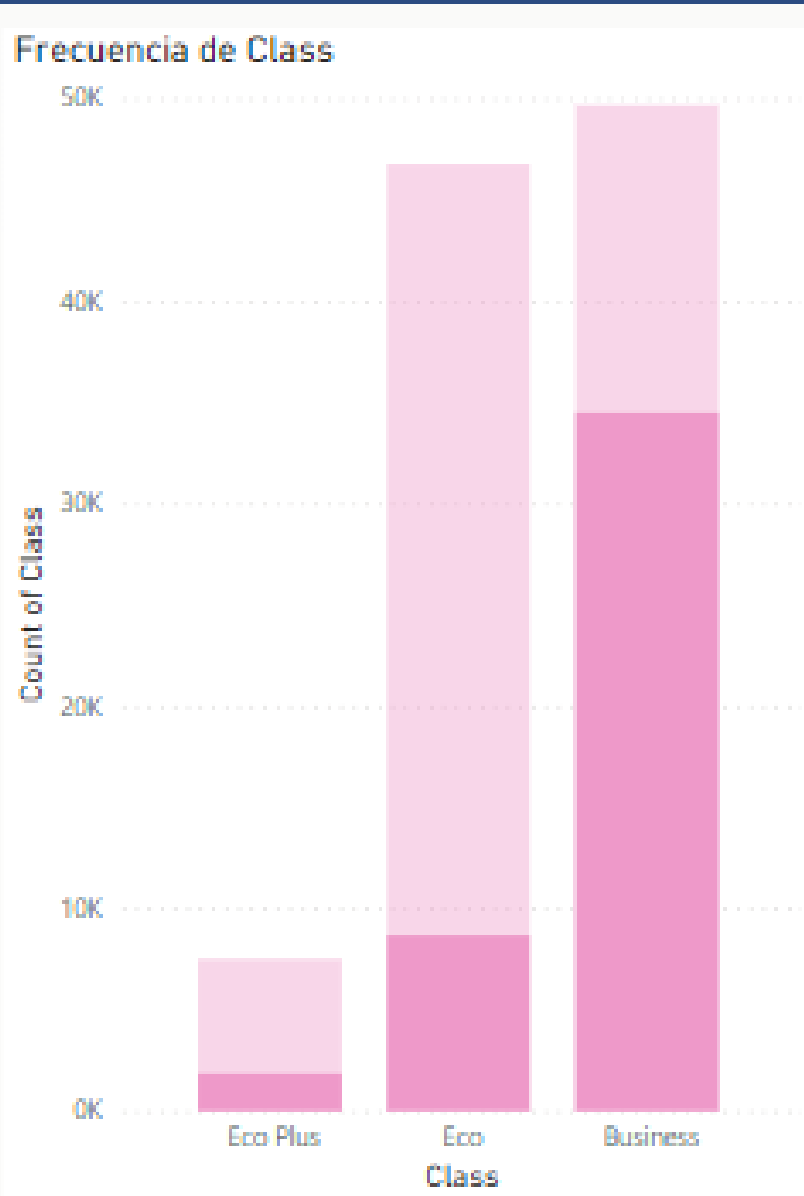# Ease of Online Boarding y Satisfaction

```
> correlacion4
[1] 0.6185259
```


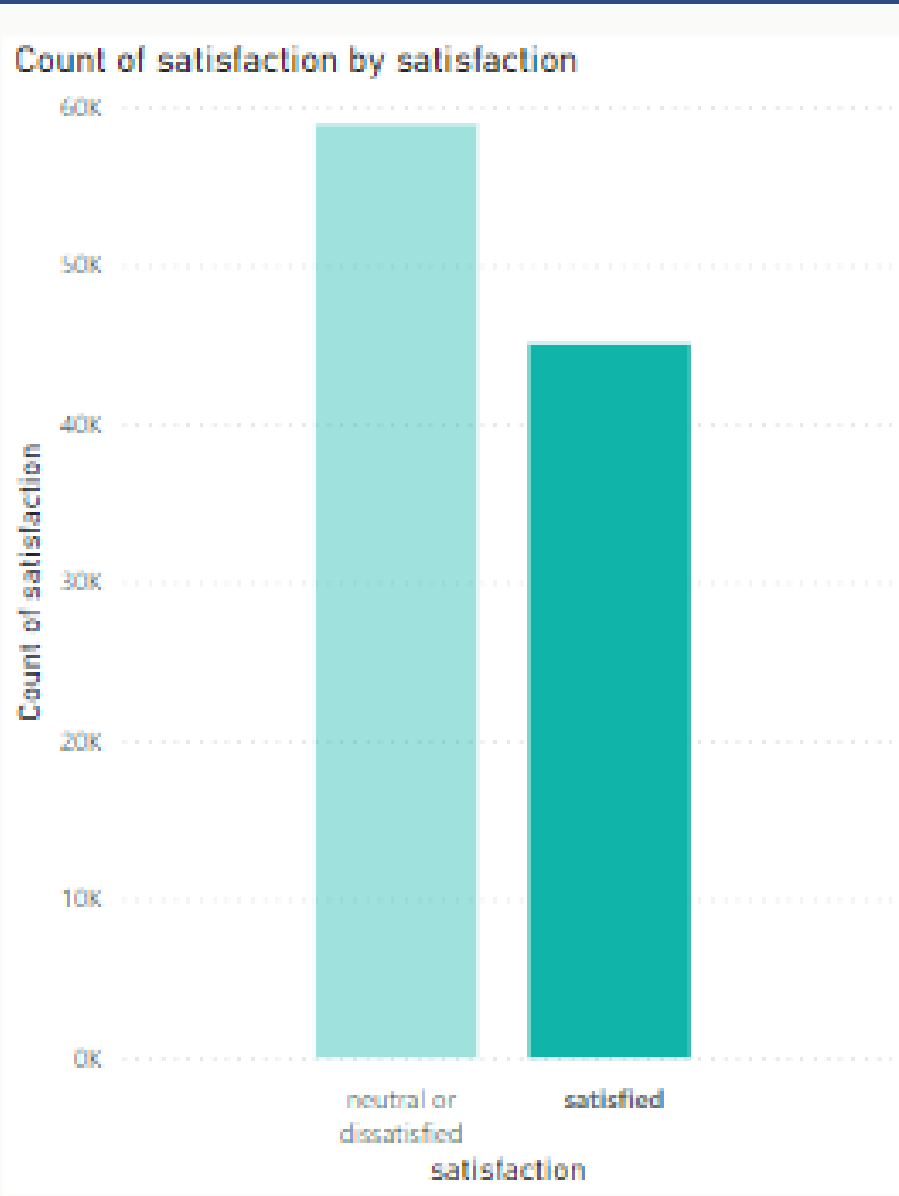Count of satisfaction by satisfaction

Frecuencia de Online boarding

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

Count of Online boarding by Online boarding and satisfaction

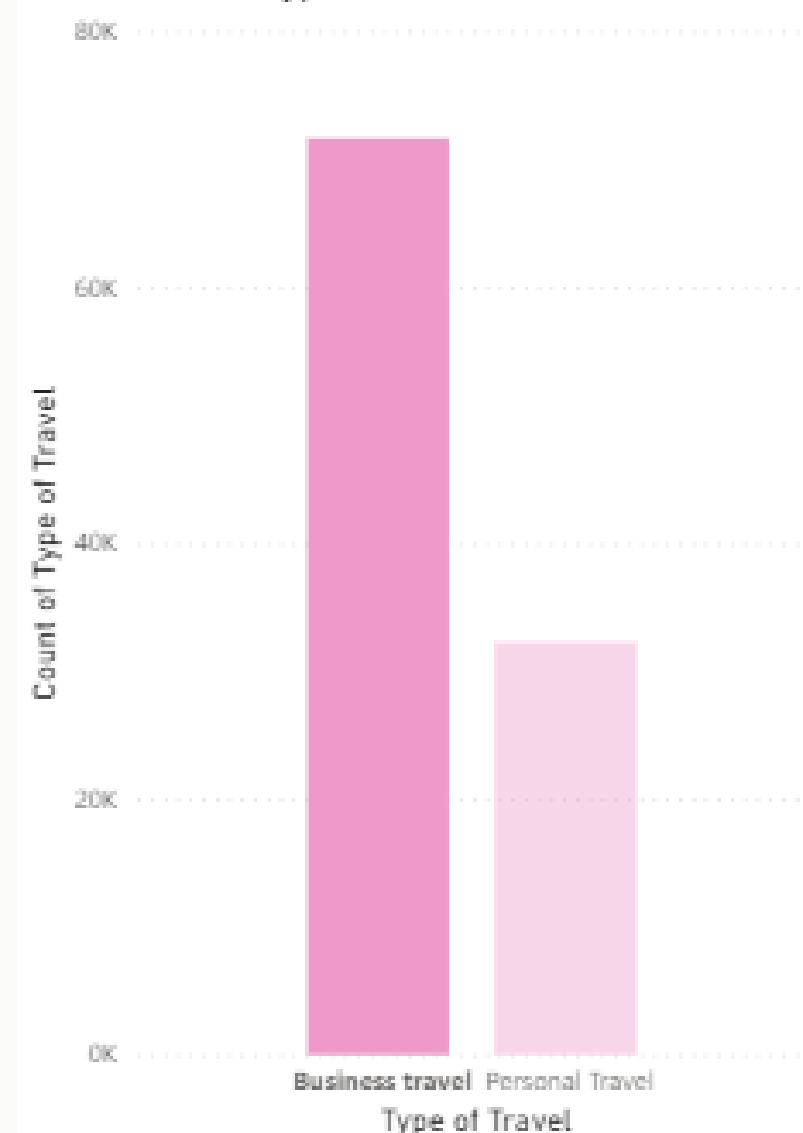satisfaction ● neutral or dissatisfied ● satisfied

Class y Satisfaction

Class y Type of Travel


```
> correlacion3
[1] 0.5540544
```

Frecuencia de Class
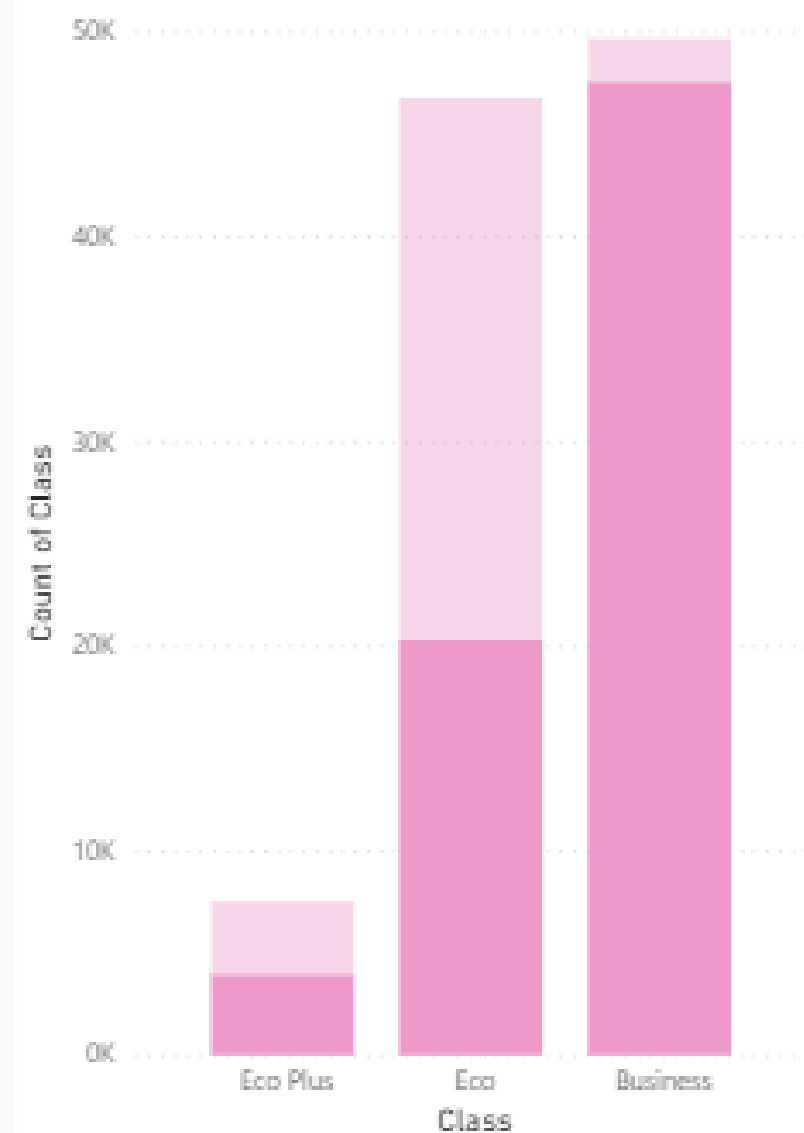
Frecuencia de Type of Travel

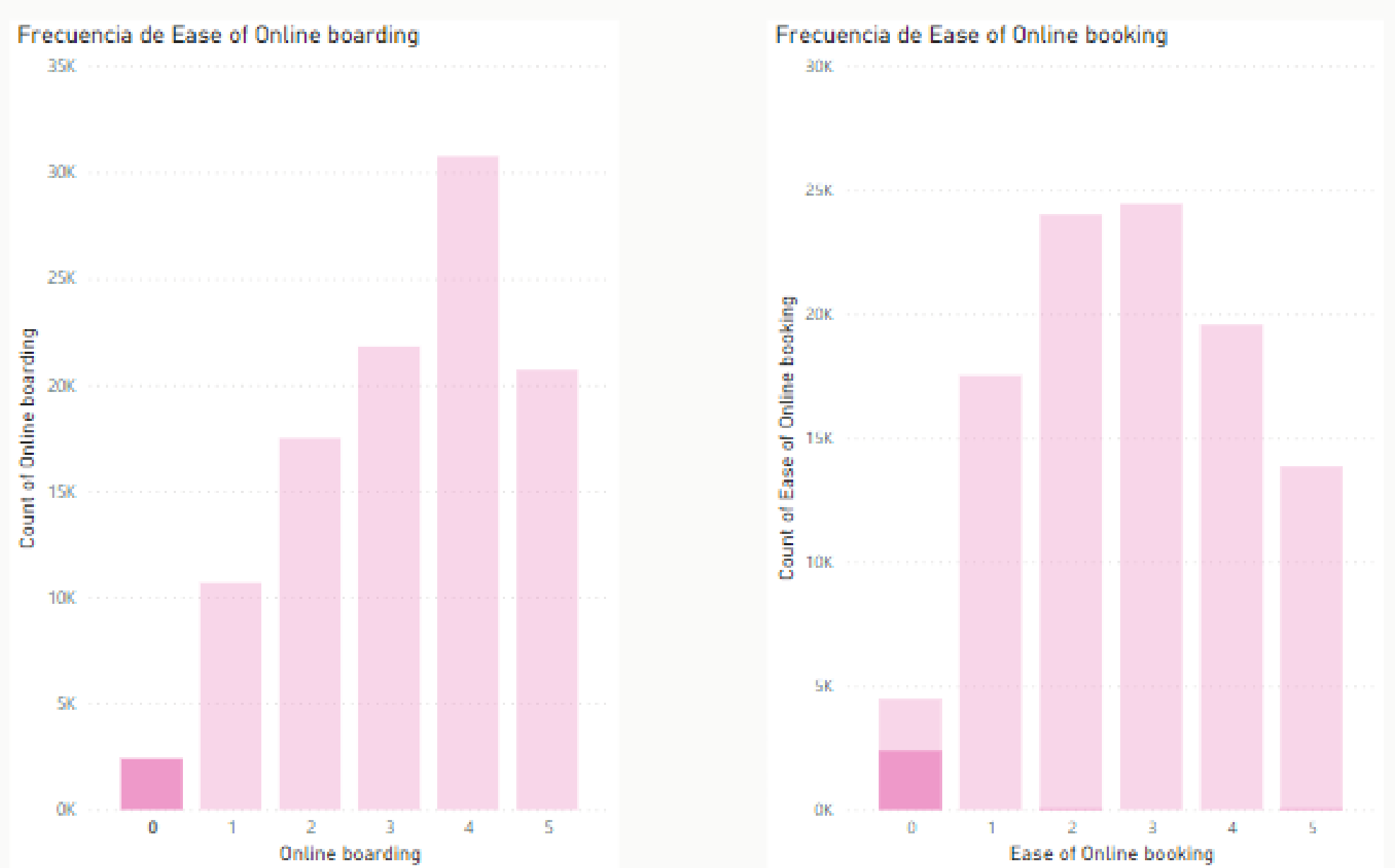| Class	Business travel	Personal Travel
Business	47508	2157
Eco	20257	26488
Eco Plus	3890	3604

Class y Type of Travel

Type of Travel ● Business travel ● Personal Travel
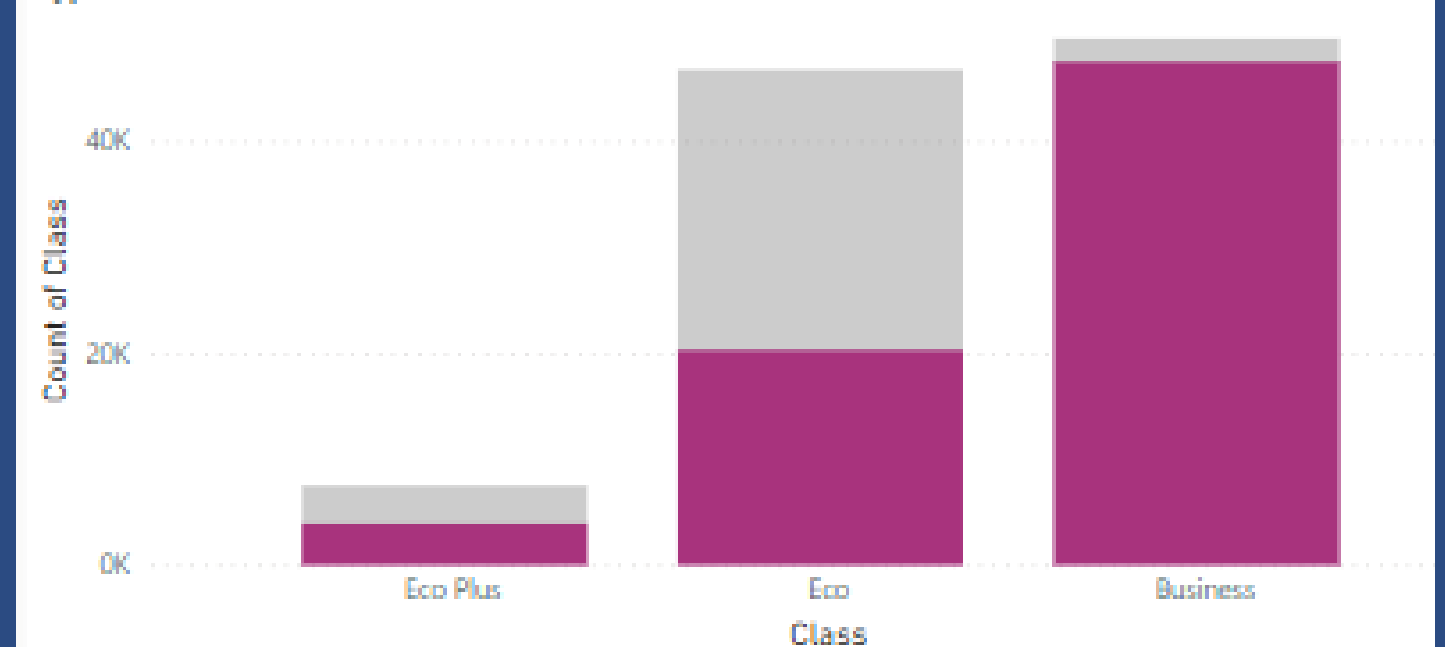
Ease of Online Boarding y Ease of Online Booking

```
> correlacion6
[1] 0.4897872
```


Frecuencia de Ease of Online boarding

Frecuencia de Ease of Online booking

Online boarding	0	1	2	3	4	5
0	2387		1			40
1	190	7361	870	864	739	668
2	162	1165	12230	1516	1305	1127
3	290	1843	2324	13424	2157	1766
4	760	4058	4823	4981	11915	4225
5	698	3098	3773	3664	3455	6025

Ease of Online Boarding y Ease of Online Booking

Ease of Online booking ●0 ●1 ●2 ●3 ●4 ●5
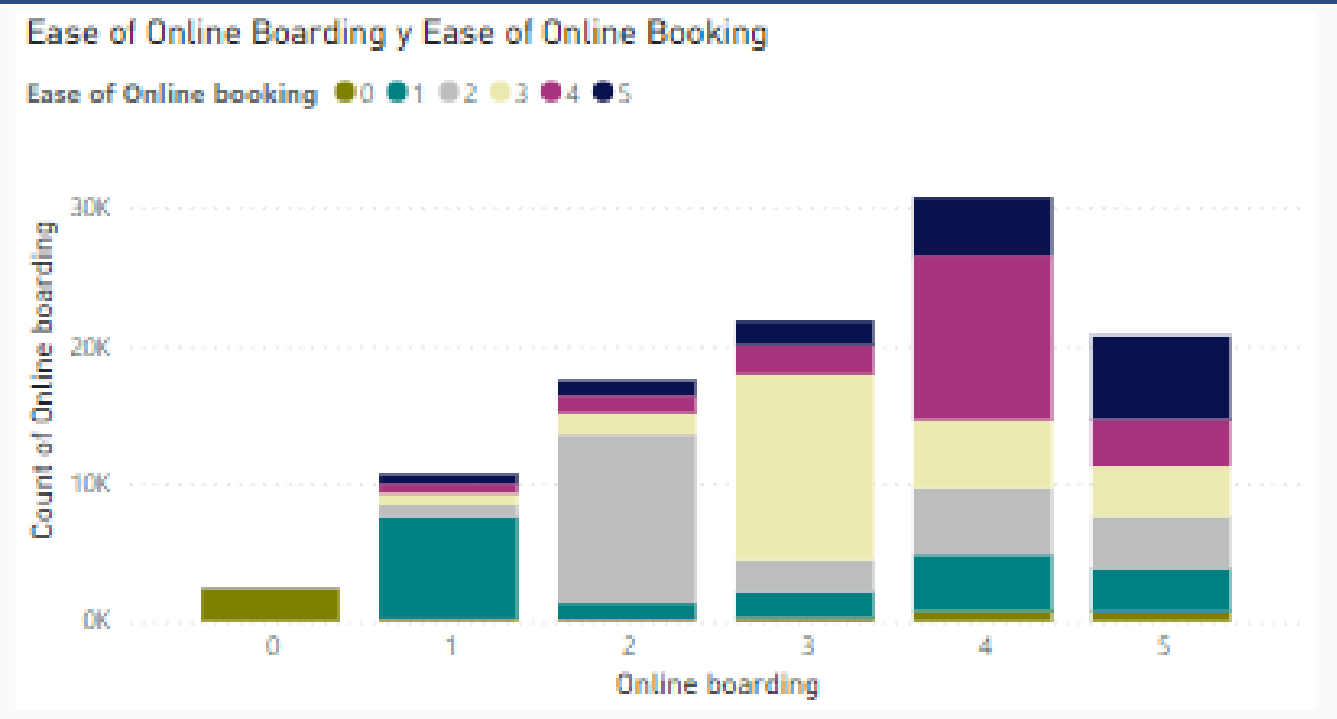
VARIABLES NUMÉRICAS

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
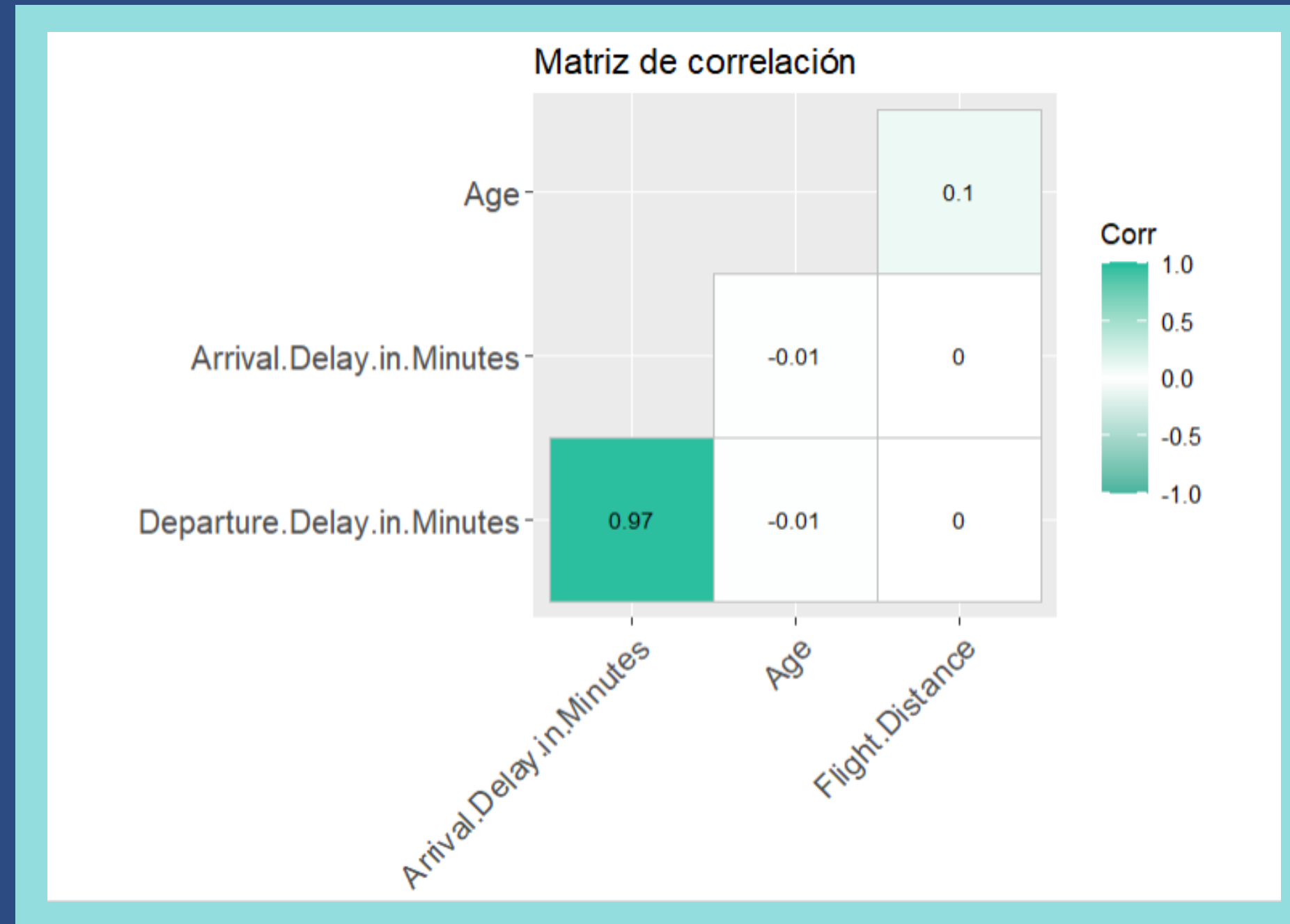
Departure Delay in Minutes y Arrival Delay in Minutes
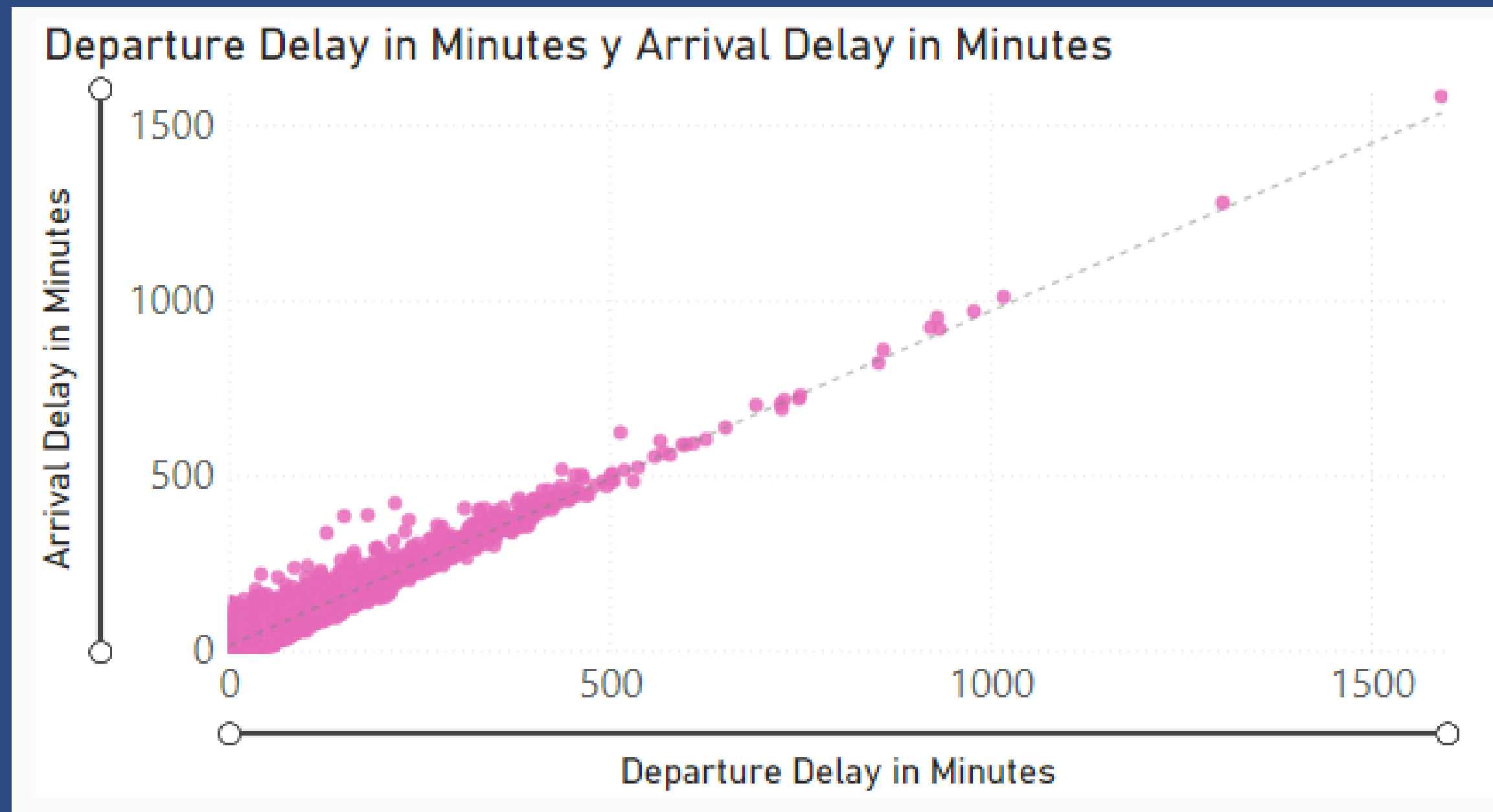
VARIABLES NUMÉRICAS Y CATEGÓRICAS

Flight distance y Class
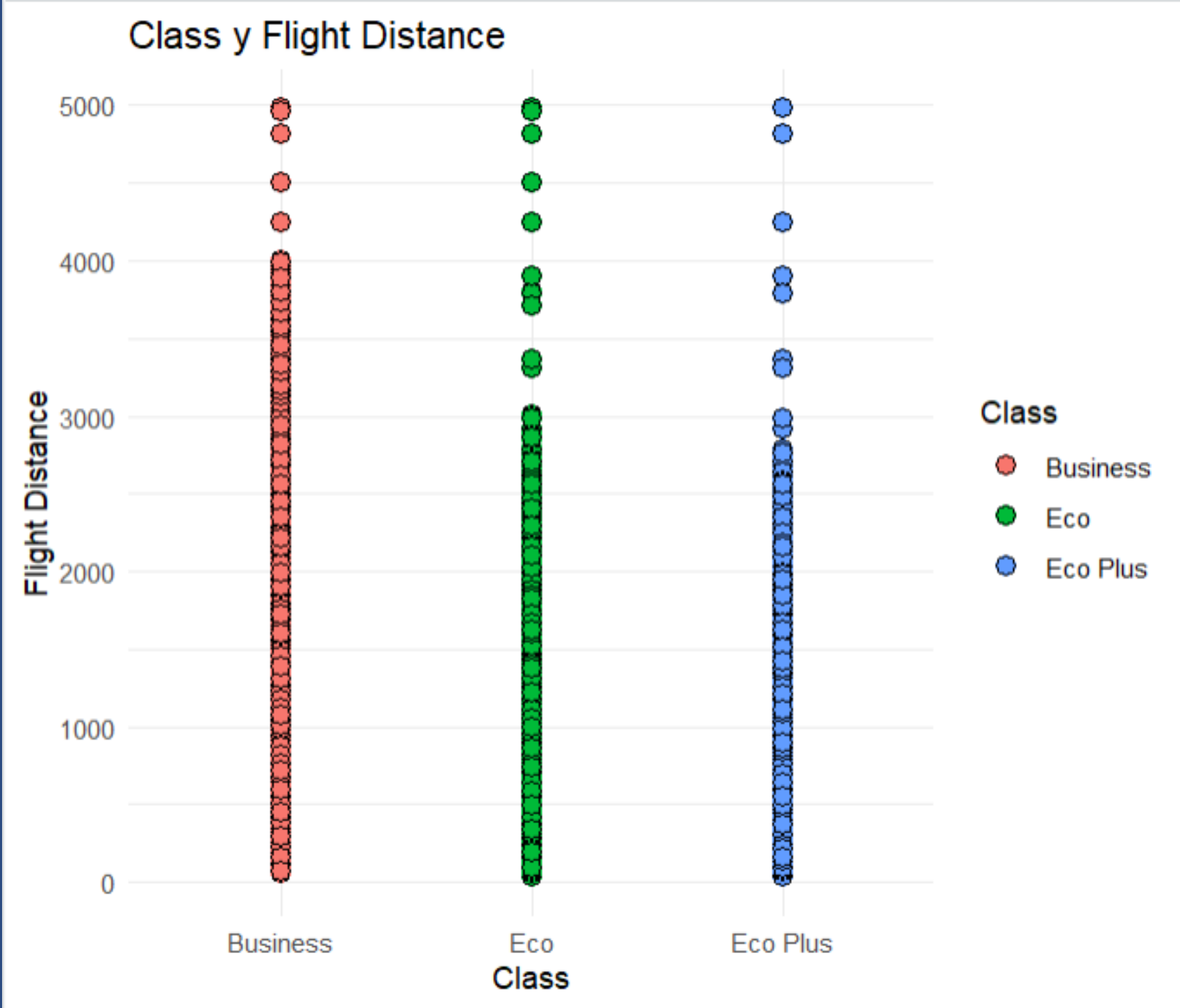
Kruskal-Wallis chi-squared
0.1707619