

Reto | Caso Líneas Aéreas

Nombre del participante:

Indicaciones:

- **Genera una copia** de este documento y editarla con tu nombre de la siguiente manera:
C5SC3 Reto - Nombre completo del participante
- Una vez terminado el Reto deberás de **entregarlo** en la opción “Añadir publicación” en el apartado Reto de aplicación en el trabajo.

Los siguientes pasos te guiarán en el proceso de desarrollo del reto para que logres completarlo con éxito:

1. Descarga el archivo **pdf** llamado **Reto Material del Caso Aerolíneas**, ya que en éste se presentan estadísticas descriptivas y modelos de regresión que se deben analizar para determinar qué variables se consideran relevantes en el servicio.

2. Descarga, las siguientes bases de datos:

- **Descargar Caso Aerolíneas Datos.csv**
- **Descargar Datos Originales.xlsx**

3. Contesta cada uno de los puntos que se piden en el espacio asignado posterior a la tabla de actividades, basándote en los datos que se presentan en los recursos que previamente descargaste:

Actividad que realizar	Contesta lo siguiente
1.- Análisis descriptivo de las variables que describen el comportamiento del Servicio en las aerolíneas	Analiza la información proporcionada y destaca al menos 5 comportamientos importantes de las variables (explícalos utilizando medidas estadísticas).



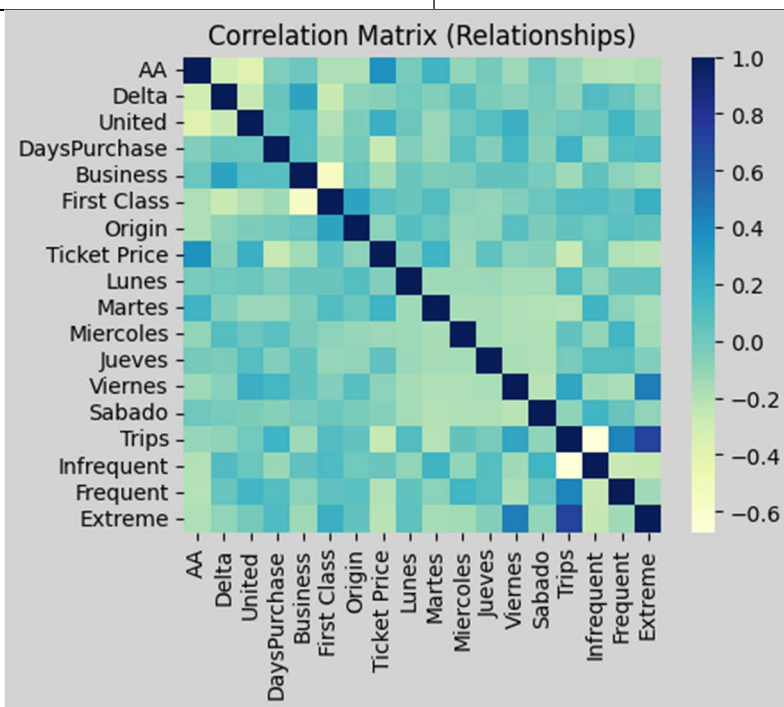
- Para el caso de la Variable Airline, donde la subclase SouthWest no presenta gran variabilidad, donde el rango de los precios asociados a ella es muy bajo, por lo tanto, no tendrá gran influencia en el regresor a generar por lo mismo.
- En el caso de la variable Route Point, no parece haber, en base a una estadística simple como es su media y varianza, gran diferencia en las mismas. Como resultado se espera que, aunque se encuentre correlacionada con los precios de los vuelos, no proporcione información significativa.
- Se puede observar mediante la variable Days from Purchase to Travel que los precios de los vuelos bajan proporcionalmente al número de días con lo cuales se compran los boletos, es decir, entre mayor anticipación menor será el costo.
- De manera similar al caso anterior, la variable Trips deja ver que entre mayor sea la frecuencia de viaje de los usuarios de las aerolíneas, menor será el costo de sus boletos.
- Se observa al analizar la duración de los vuelos, mediante la variable Route, que entre, más corta sea la duración de estos, menor será el costo de los boletos.

2.- Análisis de multicolinealidad

Utiliza la matriz de correlación para destacar qué variables independientes (X's) presentan multicolinealidad. Destaca **al menos 3** de las más importantes.

Además, en el **Modelo1** analiza el **VIF** que se proporciona para cada variable.

Escribe aquí



Trips Vs. Frequent, Trips Vs. Extreme, Trips Vs. Infrequent, Business Vs. First Class



	Coef	SE Coef	Corr. Coef	T-value	P-value(T)	VIF
const	239.724226	40.593357	NaN	5.905504	6.147089e-08	NaN
AA	167.774696	18.866045	0.362682	8.892945	5.749708e-14	2.724465
Delta	143.813828	21.160858	-0.070089	6.796219	1.133000e-09	2.063963
United	176.753591	19.088228	0.200270	9.259822	9.905651e-15	2.276292
DaysPurchase	-0.404415	0.214466	-0.258415	-1.885684	6.256305e-02	3.909147
Business	-33.687126	16.340401	-0.147254	-2.061585	4.213153e-02	2.071047
First Class	84.348041	16.735034	0.078112	5.040207	2.387354e-06	3.841915
Origin	14.197671	12.419772	-0.079771	1.143151	2.560077e-01	2.046764
Lunes	-26.652075	23.220720	-0.051963	-1.147771	2.541048e-01	1.795586
Martes	0.089123	21.302987	0.171748	0.004184	9.966712e-01	2.199649
Miercoles	-22.247835	22.058746	-0.126173	-1.008572	3.158838e-01	2.196485
Jueves	9.555218	22.818813	0.057597	0.418743	6.764016e-01	2.171437
Viernes	-20.844220	22.300784	-0.080451	-0.934686	3.524515e-01	2.672189
Sabado	-19.058353	20.840538	-0.068676	-0.914485	3.629058e-01	2.468030
Trips	2.848887	4.211125	-0.256126	0.676514	5.004491e-01	25.779984
Infrequent	0.603166	23.317638	0.029747	0.025867	9.794204e-01	3.749754
Frequent	-79.924947	31.985256	-0.193850	-2.498806	1.427592e-02	3.084898
Extreme	-54.374383	49.108093	-0.202628	-1.107239	2.711413e-01	6.455274

Se puede apreciar que la variable Trips presenta multicolinealidad, al tener un valor mayor a 10 en VIF. Además, no es congruente con el comportamiento en la matriz de correlación.

3.- Construir la ecuación o modelo matemático (copia y pega de la información proporcionada) Indica si el modelo es congruente, es decir, no presenta efectos de multicolinealidad

Utiliza el **Modelo 1** para hacer el análisis del modelo estimado. ¿Es congruente?

4.- Validación estadística del modelo:

- Medidas de calidad del ajuste
- Prueba de hipótesis para la ecuación (F)
- Prueba de hipótesis para cada una de las variables independientes (t)



Equation Found: $y = 239.7242 + 167.7747 \text{ AA} + 143.8138 \text{ Delta} + 176.7536 \text{ United} - 0.4044 \text{ DaysPurchase} - 33.6871 \text{ Business} + 84.348 \text{ First Class} + 14.1977 \text{ Origin} - 26.6521 \text{ Lunes} + 0.0891 \text{ Martes} - 22.2478 \text{ Miercoles} + 9.5552 \text{ Jueves} - 20.8442 \text{ Viernes} - 19.0584 \text{ Sabado} + 2.8489 \text{ Trips} + 0.6032 \text{ Infrequent} - 79.9249 \text{ Frequent} - 54.3744 \text{ Extreme}$

	DoF	F-Value	P-value(F)	R-Squared	Adj. R-squared	MAE	MSE
Scores	17	9.386843	1.563131e-13	63.938888	57.127345	41.105465	2670.729945

	Coef	SE Coef	Corr. Coef	T-value	P-value(T)	VIF
const	239.724226	40.593357	NaN	5.905504	6.147089e-08	NaN
AA	167.774696	18.866045	0.362682	8.892945	5.749708e-14	2.724465
Delta	143.813828	21.160858	-0.070089	6.796219	1.133000e-09	2.063963
United	176.753591	19.088228	0.200270	9.259822	9.905651e-15	2.276292
DaysPurchase	-0.404415	0.214466	-0.258415	-1.885684	6.256305e-02	3.909147
Business	-33.687126	16.340401	-0.147254	-2.061585	4.213153e-02	2.071047
First Class	84.348041	16.735034	0.078112	5.040207	2.387354e-06	3.841915
Origin	14.197671	12.419772	-0.079771	1.143151	2.560077e-01	2.046764
Lunes	-26.652075	23.220720	-0.051963	-1.147771	2.541048e-01	1.795586
Martes	0.089123	21.302987	0.171748	0.004184	9.966712e-01	2.199649
Miercoles	-22.247835	22.058746	-0.126173	-1.008572	3.158838e-01	2.196485
Jueves	9.555218	22.818813	0.057597	0.418743	6.764016e-01	2.171437
Viernes	-20.844220	22.300784	-0.080451	-0.934686	3.524515e-01	2.672189
Sabado	-19.058353	20.840538	-0.068676	-0.914485	3.629058e-01	2.468030
Trips	2.848887	4.211125	-0.256126	0.676514	5.004491e-01	25.779984
Infrequent	0.603166	23.317638	0.029747	0.025867	9.794204e-01	3.749754
Frequent	-79.924947	31.985256	-0.193850	-2.498806	1.427592e-02	3.084898
Extreme	-54.374383	49.108093	-0.202628	-1.107239	2.711413e-01	6.455274

Escribe aquí

El modelo no es congruente, se observa en la variable VIF y en la diferencia de signo entre los coeficientes y los coeficientes de correlacion con la variable Ticket Price.

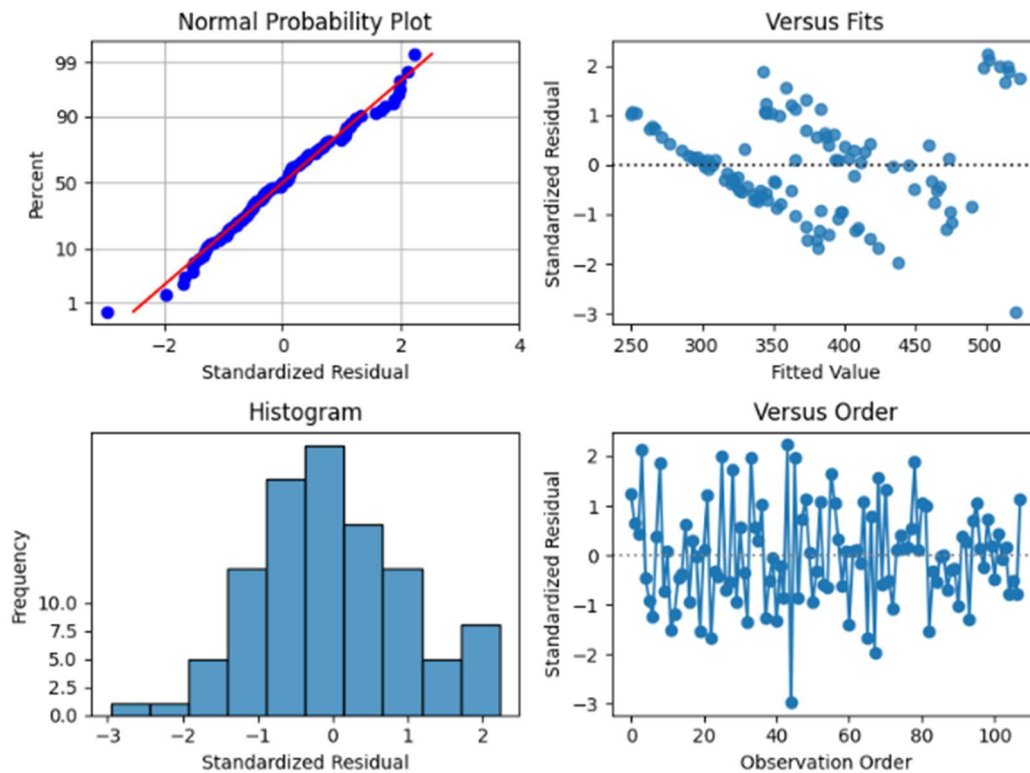


5.- Validación de supuestos:

- a) Normalidad en los residuales
- b) Errores con varianza constante
- c) Independencia de los errores

Analiza los supuestos Utilizando el **Modelo 1**

Escribe aquí



El modelo entrenado presenta normalidad en los residuales, tambien presenta independencia entre los errors, a excepcion de un valor atipico. Ademas, el modelo entrenado presenta una varianza independiente en base al comportamiento de sus residuales.

6.- Predicción del precio de venta

Con el **Modelo 2** describe cómo variaría el precio de venta según las características o variables relevantes.

Ecuación Modelo 2:

```
Equation Found: y = 190.5627 +  
158.2307 AA  
131.0599 Delta  
154.1663 United  
121.225 First Class  
-45.2514 Frequent  
61.0346 Medium_Haul
```




Resumen del desempeño del Modelo2 Entrenado:

	DoF	F-Value	P-value(F)	R-Squared	Adj. R-squared	MAE	MSE
Scores	6	35.582848	7.781329e-23	67.885235	65.977427	37.205497	2378.458685

	Coef	SE Coef	Corr. Coef	T-value	P-value(T)	VIF
const	190.562694	16.631778	NaN	11.457746	5.676258e-20	NaN
AA	158.230653	15.604740	0.362682	10.139910	4.413146e-17	1.506414
Delta	131.059878	17.852526	-0.070089	7.341251	5.519001e-11	1.169574
United	154.166345	15.994861	0.200270	9.638493	5.616972e-16	1.504011
First Class	121.225009	12.687636	0.078112	9.554578	8.596935e-16	1.423085
Frequent	-45.251350	11.904182	-0.302641	-3.801299	2.463998e-04	1.423307
Medium_Haul	61.034563	10.696773	0.378387	5.705886	1.161512e-07	1.944968

Por lo tanto:

El modelo 2 proporciona relaciones entre la variable Ticket Price y las seleccionadas en el modelo, donde, por ejemplo, una unidad de la variable AA aumentara en 190 el precio de los vuelos, cuando el resto de las variables seleccionadas se mantienen constantes. Por otro lado, cuando todas las variables se mantienen constantes, a excepción de la variable Frequent, el precio disminuirá en 45 por cada unidad de variable en cuestión. El efecto en conjunto del modelado de los precios en base al modelo 2 se puede apreciar mejor en la siguiente Figura:

