

Inciso 2: Unión de Listas de Pasajeros

Proceso realizado:

Se leen los archivos CSV "pasajeros2016.csv" y "pasajeros2017.csv".

Se concatenan los dos DataFrames en uno solo, combinando los datos de ambos años.

	ID_Pasajero	Pasajero	Edad
0	576	Danielle Thompson	60
1	579	Natalie Cuevas	49
2	683	John Murray	28
3	681	Michael Jacobs	24
4	592	Brian Hunt	40
...
150	729	Kendra Daniel	59
162	562	Zachary White	49
174	658	April Wiggins	21
176	402	Julia Barnett	33
189	715	Justin Martin	48

Anomalías y solución:

Anomalía: Una posible anomalía que prevé fue la de datos duplicados

Cómo se resuelve: Se usa el método `drop_duplicates()` especificando la columna "ID_Pasajero". En caso de haber un duplicado, solo se conserva la primera aparición

Inciso 3: Unir Pasajeros y Vuelos

Proceso utilizado: Se fusionan los DataFrames de vuelos y pasajeros usando la columna "Cve_Cliente" del DataFrame de vuelos y "ID_Pasajero" del DataFrame de pasajeros.

Tipo de relación y justificación: Utilicé un INNER JOIN, esto para poder recolectar solo la información que tenga coincidencias en ambos DataFrames.

Vuelos y Pasajeros:								
	Cve_LA	Viaje	Clase	Precio	Ruta	Cve_Cliente	ID_Pasajero	Pasajero
0	SW	9/10/2016	Economy	60	DAL-MDW	553	553	Nancy Lomonaco
1	AA	6/11/2016	Economy	150	DAL-SLC	554	554	Naresj Patel
2	DA	9/21/2016	Economy	68	DAL-AMA	554	554	Naresj Patel
3	UA	6/28/2016	Business	160	DAL-SLC	556	556	Nat Gilpin
4	SW	6/25/2016	Economy	65	DAL-OKC	557	557	Natalie DeCherney
...
395	SW	8/16/2017	Economy	229	DAL-TUL	627	627	Jordan English
396	AM	6/5/2017	First Class	85	DAL-AMA	669	669	Jason Crane
397	AM	6/2/2017	First Class	55	DAL-TUL	693	693	Shahid Shariari
398	AV	2/9/2017	Business	140	DAL-MDW	637	637	Rob Beeghly
399	AA	11/27/2017	Business	232	DAL-HOU	723	723	Henry Griffin

Inciso 4: Unir Datos de Líneas Aéreas

¿Qué tipo de proceso consideraste para unir los datos que se piden?

Una unión entre el DataFrame que salió del inciso 3 y el de las líneas aéreas.

¿Qué columnas utilizaste para lograr esa relación?

Se utiliza la columna "Cve_LA" del DataFrame del inciso 2 y la columna "Code" del DataFrame de líneas aéreas.

¿Qué tipo de unión utilizaste para unir los datos?

Utilicé un LEFT JOIN para que, en si no existe coincidencia en las líneas aéreas, se puedan mantener los registros y luego asignar un valor por defecto a los que no coincidan.

¿Qué tipo de proceso utilizaste para dejar únicamente las columnas que se piden?

Después de la unión, se filtran las columnas solo con las necesarias (Viaje, Clase, Precio, Ruta, Edad y Linea_Aerea).

```
[400 rows x 9 columns]
```

Consolidado de Vuelos, Pasajeros y Líneas Aéreas:						
	Viaje	Clase	Precio	Ruta	Edad	Linea_Aerea
0	9/10/2016	Economy	60	DAL-MDW	34	Southwest
1	6/11/2016	Economy	150	DAL-SLC	25	American Airlines
2	9/21/2016	Economy	68	DAL-AMA	25	Otra
3	6/28/2016	Business	160	DAL-SLC	55	Otra
4	6/25/2016	Economy	65	DAL-OKC	63	Southwest
..
395	8/16/2017	Economy	229	DAL-TUL	52	Southwest
396	6/5/2017	First Class	85	DAL-AMA	32	Aeromexico
397	6/2/2017	First Class	55	DAL-TUL	67	Aeromexico
398	2/9/2017	Business	140	DAL-MDW	29	Avianca
399	11/27/2017	Business	232	DAL-HOU	68	American Airlines

Lo último fue el promedio:

```
[400 rows x 6 columns]
```

Promedio semestral del precio:					
Linea_Aerea	Año	Clase	Ruta	Semestre	
0	2016	Business	DAL-AMA	1	NaN
1	2016	Business	DAL-AMA	2	NaN
2	2016	Business	DAL-ATL	1	218.0
3	2016	Business	DAL-ATL	2	NaN
4	2016	Business	DAL-AUS	1	250.0
..
128	2017	First Class	DAL-PHX	2	119.0
129	2017	First Class	DAL-SLC	1	NaN
130	2017	First Class	DAL-SLC	2	NaN
131	2017	First Class	DAL-TUL	1	55.0
132	2017	First Class	DAL-TUL	2	208.5

NaN siendo que esa aerolínea no tuvo vuelos de esa ruta en ese semestre.