

PROYECTO FINAL

Chris Mérida 18013387 - PAPD

ENLACE VIDEO

<https://youtu.be/OBFbRGY-41A>



PROCESO

Exploración Estandatización

1. Revisión manual de archivos .csv.
2. Estandarizar las columnas (python script)

Estructura

1. Definir la estructura de cada tabla (dataframes dtypes).
2. Documentos sql.

Creación de tablas

1. Conexión a la base de datos
2. Python script ejecuta query de SQL para creación de tablas (5).



PROCESO

Validación de datos

1. Verificar tipos de datos por columna.
2. Directorio de valid data.

Ingesta de datos

1. Lectura y carga de archivo valido a la base de datos (Python script).

Notebook

1. Conexión a la base de datos.
2. Query de prueba, verificando la cada tabla.
3. Responder preguntas planteadas.

PROCEDIMIENTO - ARCHIVOS

- db_config.py
 - db_connection.py
 - run_tables_creation.py
 - normalize_columns.py
 - validate_data.py
 - populate_tables.py
 - research.ipynb
-
- **README.md**
 - **venv**
 - **requirements.txt**
 - **Presentación - Enlace al video**
 - **Folders (raw, normalized, valid, tables)**

PREGUNTAS

1. ¿EN QUE PAÍS Y QUE AVIÓN SE ENCUENTRA ENTRE EL 85% Y EL 70% DE LA CANTIDAD DE ATERRISAJES? PROPORCIONE EL NOMBRE DEL PAÍS Y EL NOMBRE DEL AVIÓN.

✓ 0.2s

	pais	cantidad_vuelos	porcentaje_del_total
0	US	336776	100.0

	tailnum	cantidad_vuelos	porcentaje_del_total	manufacturer	model
0	N725MQ	575	0.170737	NaN	NaN
1	N722MQ	513	0.152327	NaN	NaN
2	N723MQ	507	0.150545	NaN	NaN
3	N711MQ	486	0.144310	GULFSTREAM AEROSPACE	G1159B
4	N713MQ	483	0.143419	NaN	NaN

2. ¿INDIQUE CUAL ES EL AREOPUERTO CON LA TEMPERATURA MÁS ALTA REGISTRADA EN LOS DATOS?

✓ 0.0s

Aeropuerto con la temperatura mas alta: Newark Liberty Intl
Temperatura máxima: 100.04 °F

3. ¿CUÁL ES LA AEROLÍNEA CON LA MENOR CANTIDAD DE VUELOS REGISTRADOS, INDIQUE CUANTOS VUELOS, EL CÓDIGO DE LA ÁREAOLINEA, EL NOMBRE COMPLETO?

✓ 0.0s

	carrier	name	cantidad_vuelos
0	00	SkyWest Airlines Inc.	32

4. INDIQUE LA MEDIA, MEDIAN, MINIMO, MAXIMO Y DESVIACIÓN ESTÁNDAR DE LAS MILLAS RECORRIDAS POR CADA AVIÓN, DEBE MOSTRAR EL NOMBRE DEL AVIÓN Y LA INFORMACIÓN ESTADÍSTICA EN COLUMNAS ADICIONALES

✓ 0.0s

	tailnum	media	mediana	mínimo	máximo	desviación_estándar	manufacturer	model
0	D942DN	854.500000	853.0	762	950	106.837883	NaN	NaN
1	N0EGMQ	676.188679	719.0	213	1147	199.956176	NaN	NaN
2	N10156	757.947712	746.0	116	1325	332.057299	EMBRAER	EMB-145XR
3	N102UW	535.875000	541.0	529	544	6.749704	AIRBUS INDUSTRIE	A320-214
4	N103US	535.195652	529.0	529	544	6.615200	AIRBUS INDUSTRIE	A320-214
...
4038	N997DL	867.761905	950.0	187	1096	223.907218	MCDONNELL DOUGLAS AIRCRAFT CO	MD-88
4039	N998AT	593.538462	762.0	397	762	185.562762	BOEING	717-200
4040	N998DL	857.818182	950.0	187	1096	233.728797	MCDONNELL DOUGLAS CORPORATION	MD-88
4041	N999DN	895.459016	1020.0	187	1096	242.964028	MCDONNELL DOUGLAS CORPORATION	MD-88
4042	N9EAMQ	674.665323	719.0	213	1020	183.691490	NaN	NaN

4043 rows × 8 columns

5. MUESTRE UN CUBO DE INFORMACIÓN INCLUYENDO LA INFORMACIÓN DE TODAS LAS TABLAS PROPORCIONADAS

- INDIQUE LA CANTIDAD DE FILAS Y COLUMNAS

✓ 1.1s

Número de filas: 336776

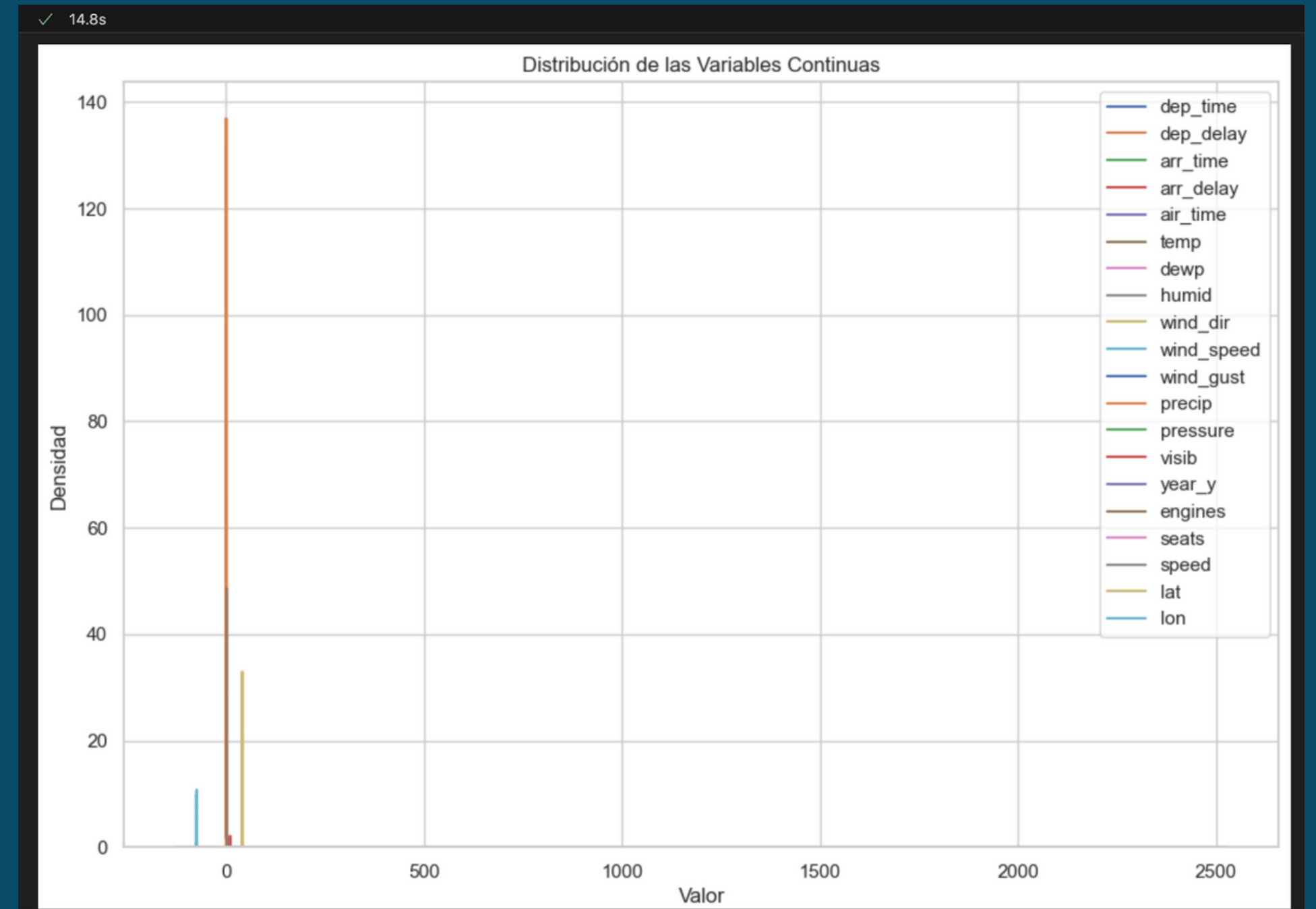
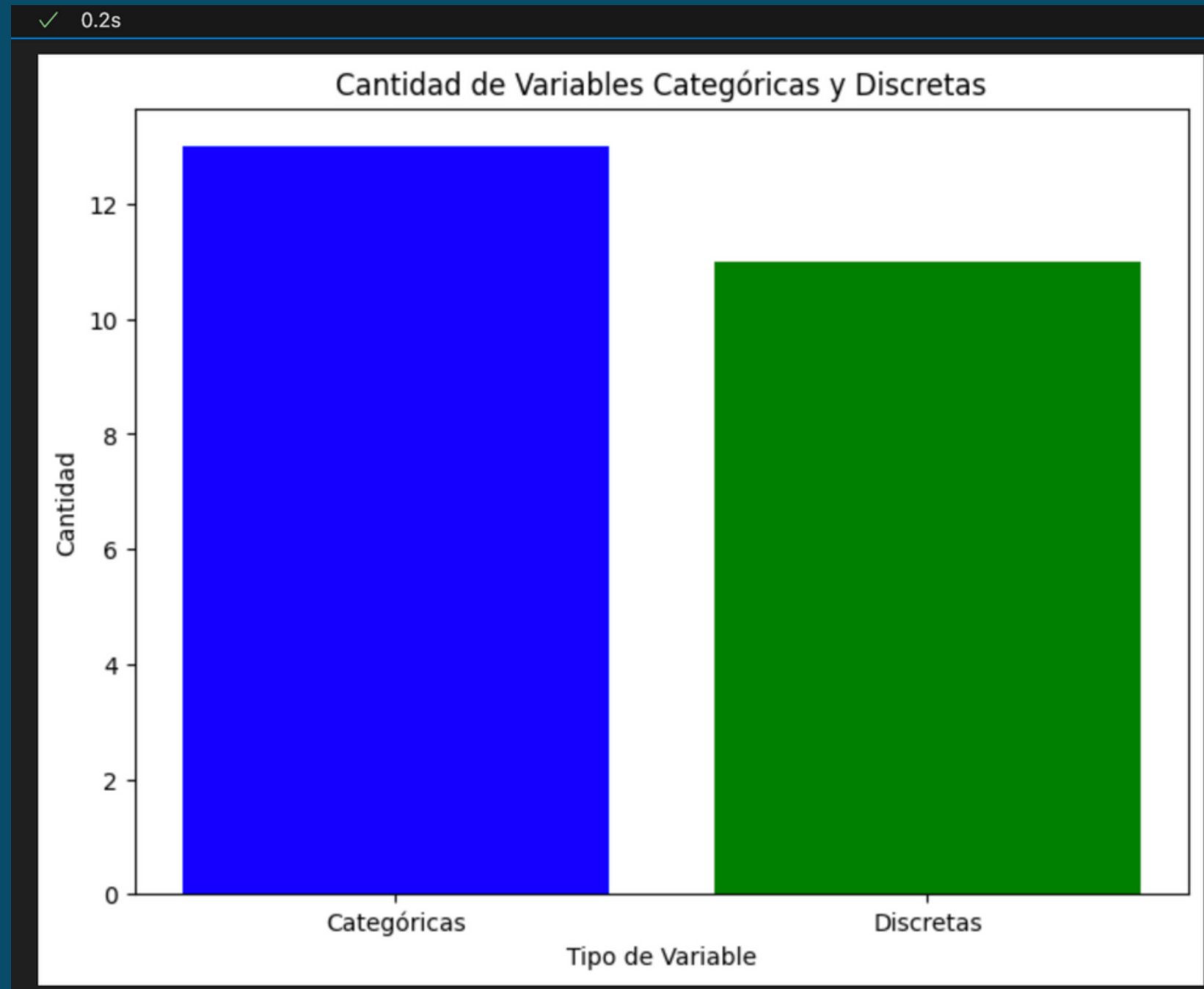
Número de columnas: 46

- INDIQUE CUANTAS Y CUALES SON LAS VARIABLES CATEGORICAS, CONTINUAS, DISCRETAS Y DE FECHA Y HORA.

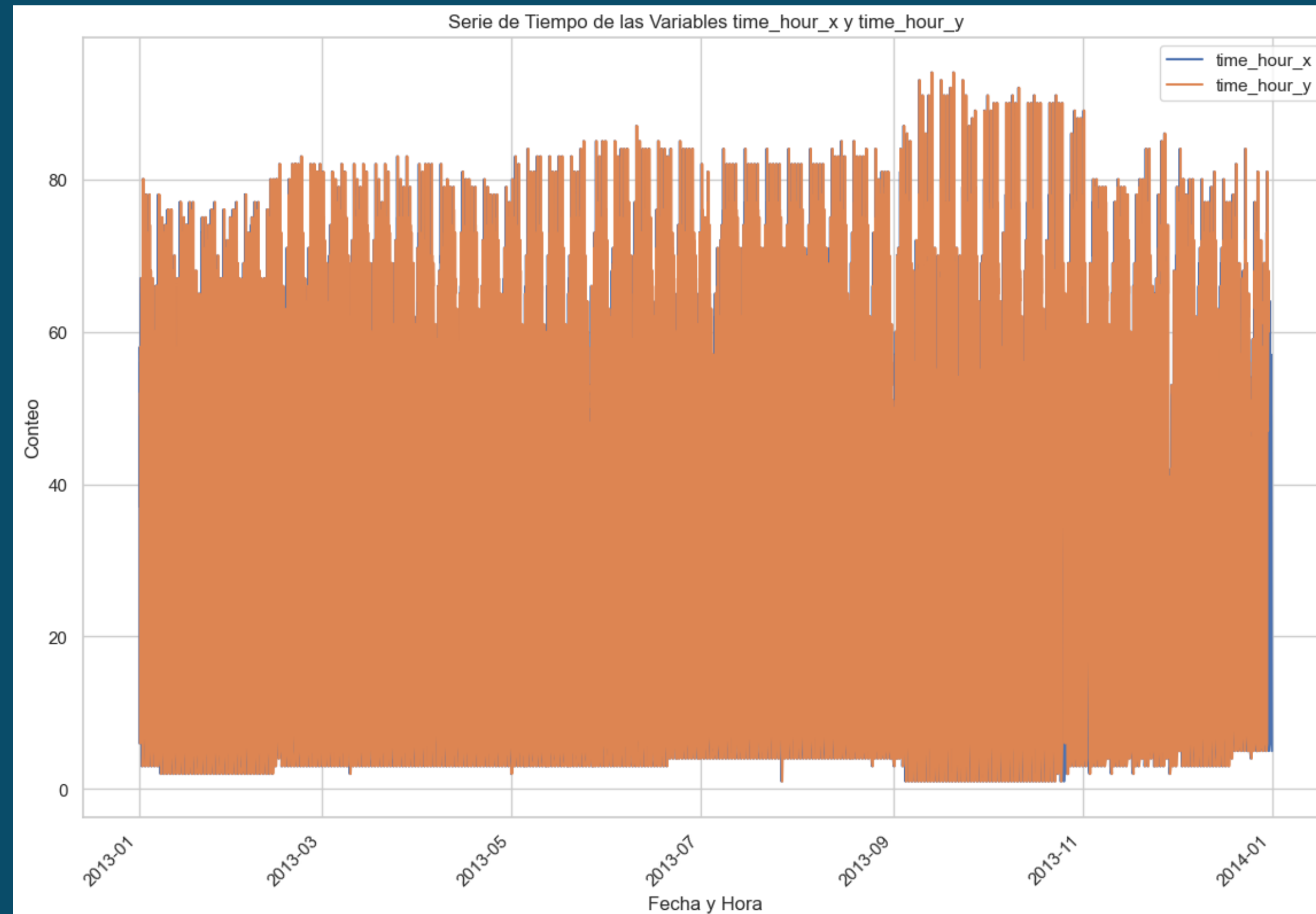
✓ 0.1s

```
Variables categóricas: 13 ['carrier', 'tailnum', 'origin', 'dest', 'type', 'manufacturer', 'model', 'e  
Variables continuas: 20 ['dep_time', 'dep_delay', 'arr_time', 'arr_delay', 'air_time', 'temp', 'dewp',  
Variables discretas: 11 ['year_x', 'month', 'day', 'sched_dep_time', 'sched_arr_time', 'flight', 'dist  
Variables de fecha y hora: 2 ['time_hour_x', 'time_hour_y']
```

- GRAFICAS



- GRAFICAS



PREGUNTAS DE NEGOCIO PLANTEADAS

1. ¿CUÁLES SON LOS DESTINOS DE VUELO MÁS FRECUENTES?

Destinos de vuelo más frecuentes:

dest

ORD 17283

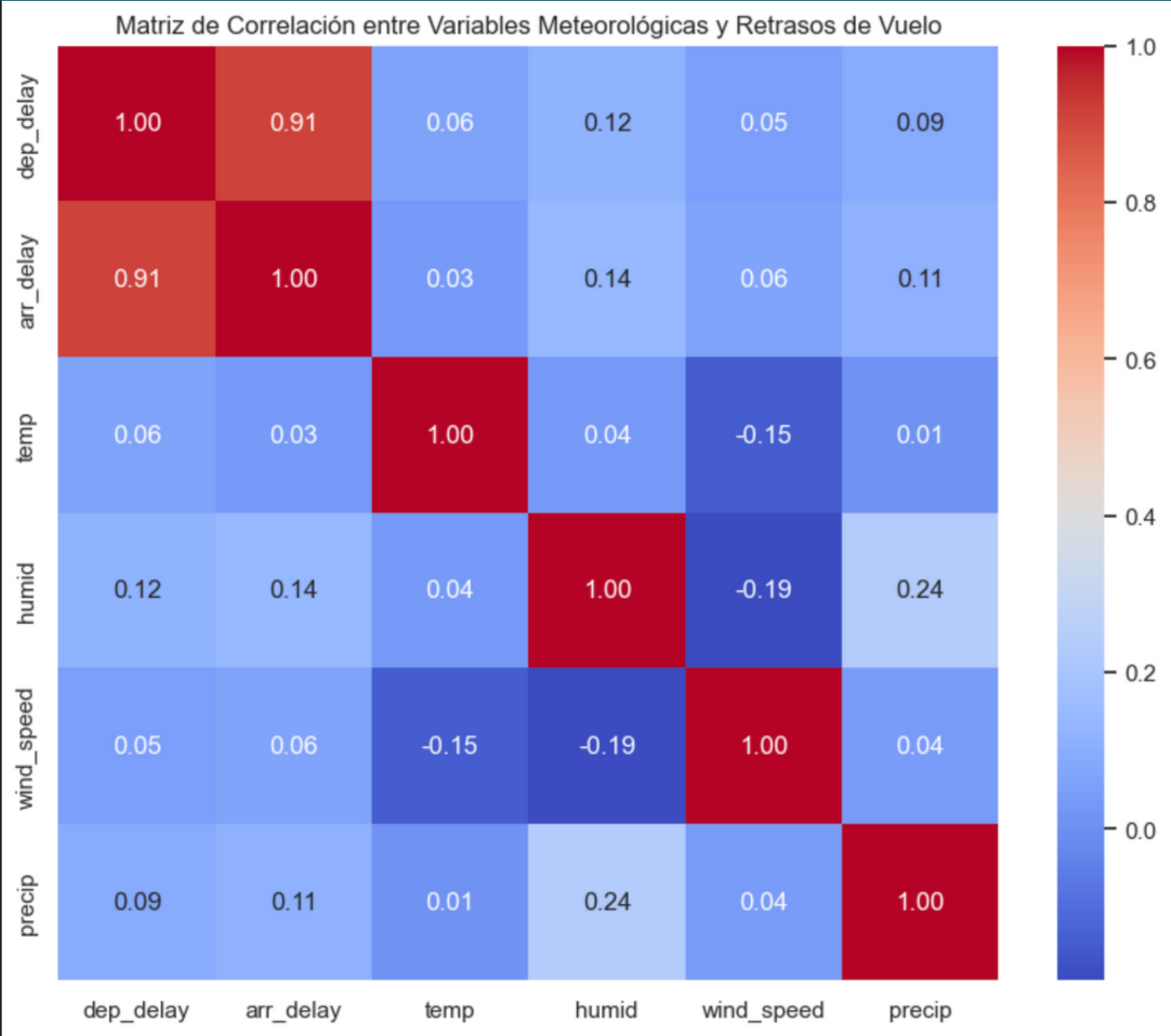
ATL 17215

LAX 16174

2. ¿QUÉ AEROLÍNEA TIENE EL MAYOR NÚMERO DE VUELOS?

La aerolínea con el mayor número de vuelos es: UA

3. ¿CÓMO AFECTA EL CLIMA A LOS RETRASOS DE LOS VUELOS?



4. ¿QUÉ AEROPUERTO TIENE LA MAYOR DIFERENCIA ENTRE EL NÚMERO DE SALIDAS Y LLEGADAS?

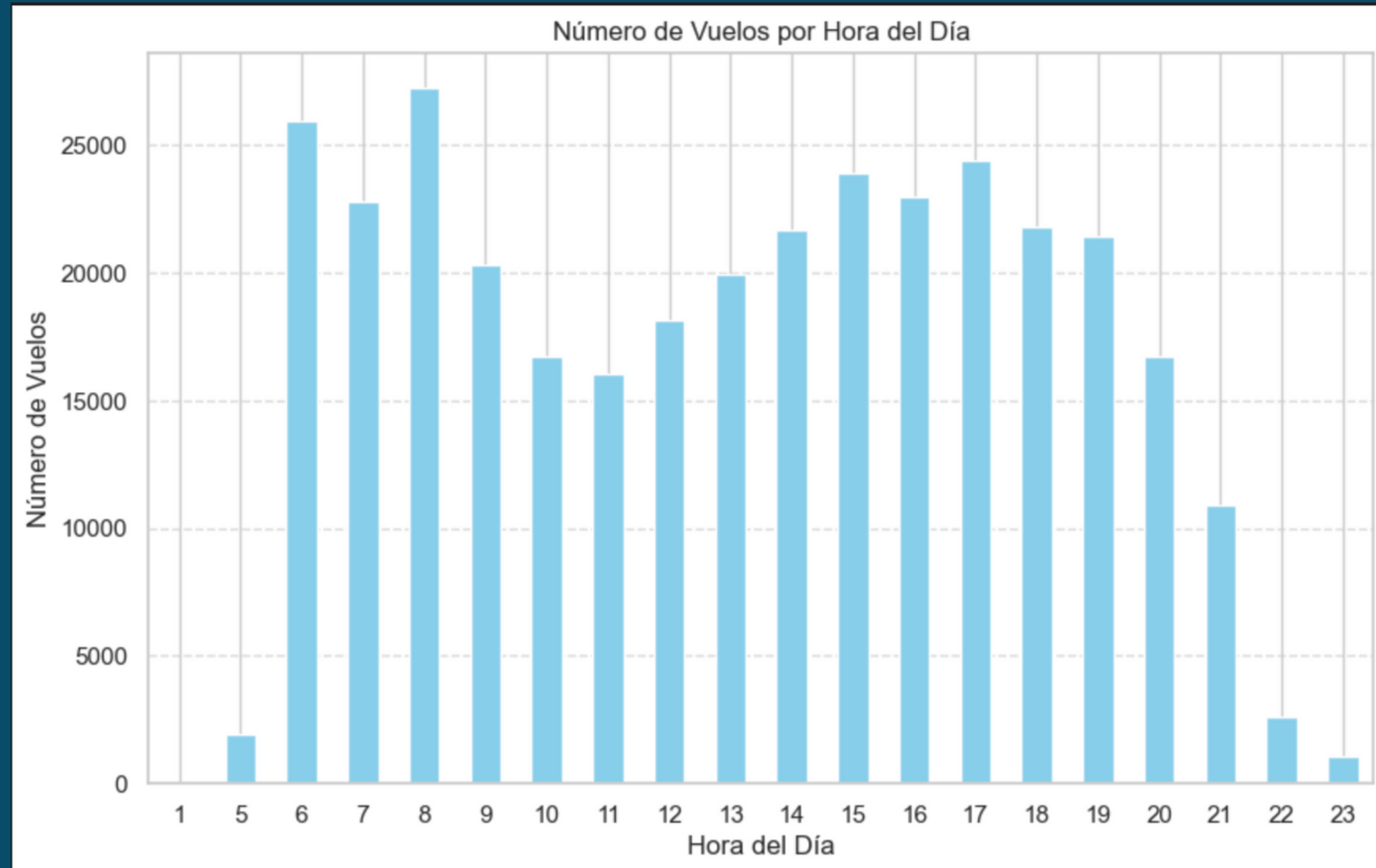
El aeropuerto con la mayor diferencia entre el número de salidas y llegadas es: LGA

5. ¿QUÉ AEROPUERTO TIENE EL MAYOR NÚMERO DE SALIDAS Y LLEGADAS?

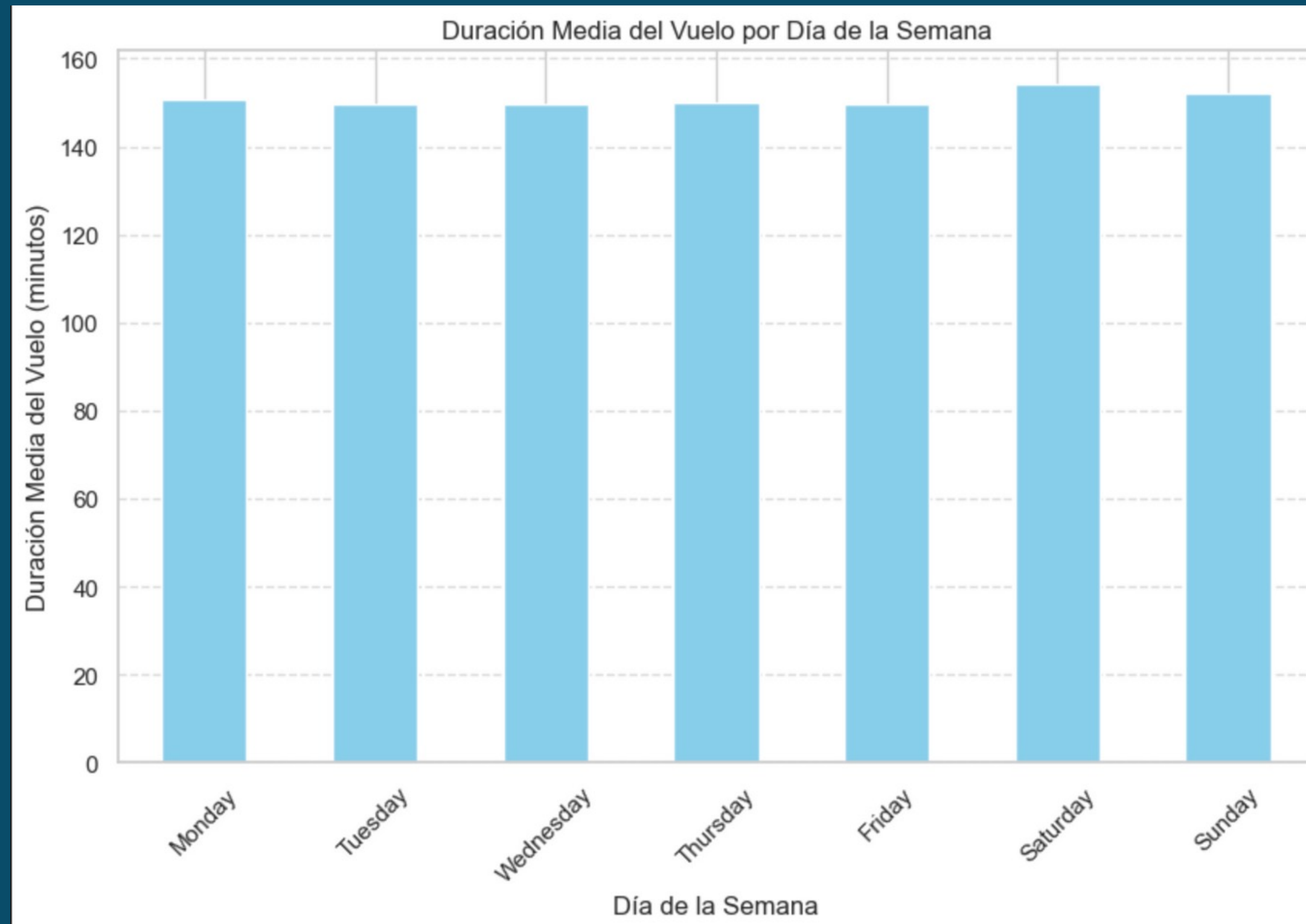
El aeropuerto con el mayor número de salidas es: Newark Liberty Intl

El aeropuerto con el mayor número de llegadas es: Chicago Ohare Intl

6. ¿CÓMO VARÍA EL NÚMERO DE VUELOS POR HORA DEL DÍA?



7. ¿CÓMO VARÍA LA DURACIÓN MEDIA DEL VUELO DURANTE DIFERENTES DÍAS DE LA SEMANA?



8. ¿CUÁL ES EL TIEMPO PROMEDIO DE RETRASO POR AEROLÍNEA?

	Tiempo Promedio de Retraso en Salida	Tiempo Promedio de Retraso en Llegada
carrier		
9E	16.725769	7.379669
AA	8.586016	0.364291
AS	5.804775	-9.930889
B6	13.022522	9.457973
DL	9.264505	1.644341

9. ¿CUÁL ES EL TIEMPO PROMEDIO DE RETRASO POR TIPO DE AERONAVE?

	Tiempo Promedio de Retraso en Salida por Tipo de Aeronave	Tiempo Promedio de Retraso en Llegada por Tipo de Aeronave
tailnum		
D942DN	31.500000	31.500000
N0EGMQ	8.491525	9.982955
N10156	17.815068	12.717241
N102UW	8.000000	2.937500
N103US	-3.195652	-6.934783

10. ¿EN QUÉ ÉPOCAS DEL AÑO HAY UN MAYOR NÚMERO DE VUELOS?

