



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Visualización de la Información

Proyecto de mitad del curso

- | | |
|---|--------------------------------------|
| ■ Lucas Cedric Cervantes
Beutelspacher | ■ Guillermo Gerardo Andrés
Urbano |
| ■ Hugo Didier Longines Tapia | ■ Elsy Camila Silva Velázquez |

Introducción

El Departamento de Transporte de los Estados Unidos abreviado como USDOT recibe informes de tráfico de las aerolíneas estadounidenses e internacionales que operan desde y hacia los aeropuertos estadounidenses. Cubren los años 1990 a 2020 (US Air Traffic 1990-2020).

Deseamos a conocer a partir de este conjunto información acerca de los vuelos locales y extranjeros, ver el comportamiento de las aerolineas a través del tiempo, conocer las aerolineas con más vuelos, ver el progreso en los años en cada aerolinea y ver el flujo de vuelos de las aerolineas.

Nombre de la columna	Descripción
usg_apt	Código de aeropuerto de entrada de EE.UU
fg_apt	Código de aeropuerto de puerta de enlace extranjera
Scheduled	Métrica volada por operaciones de servicio programadas extranjera
Charter	Métrica volada por operaciones chárter
Total	Métrica total volada por servicio programado y operaciones chárter

Nuestro conjunto de datos estará conformado por varias características, podemos resaltar de todas ellas cinco que serán los códigos de identificación de cada aeropuerto local y extranjero, como también las métricas sobre los vuelos programados. Con esto en mente deseamos lograr visualizaciones que nos ayuden a contestar estas interrogantes y descubrir diferentes hallazgos en este conjunto de datos.

Metodologías

Método para los aeropuertos fuera de EEUU con mas vuelos

Primero se obtuvieron los id's únicos de los aeropuertos, posteriormente hicimos un conteo agrupando nuestra información por id. Después obtuvimos los nombres a partir de los id's. Por último creamos una copia que solo tenga los aeropuertos con más vuelos registrados y graficamos con esta información.

Método para los aeropuertos en de EEUU con mas vuelos

Se hizo lo mismo que antes solo que ahora trabajaremos con los códigos correspondientes a Estados Unidos (usg_apt y usg_apt_id). Primero se obtuvieron los id's únicos de los aeropuertos, para posteriormente hacer un conteo agrupando nuestra información por id. Después obtuvimos los nombres a partir de los id's.

Por último creamos una copia que solo tenga los aeropuertos con más vuelos registrados y graficamos con esta información.

Método para los top 5 aeropuertos en EEUU con mas vuelos registrados

Primero definimos y cargamos el mapa donde vamos a marcar los aeropuertos y definimos las posiciones geográficas de cada uno de los aeropuertos con su longitud y latitud. Proyectamos nuestros puntos sobre nuestro mapa y definimos un grafo que va a tener como nodos a cada uno de nuestros aeropuertos. Por último graficamos nuestro grafo encima del mapa que definimos.

Método para las top 10 aerolíneas con mas vuelos

Tanto para este, como los dos siguientes métodos se procedió de forma similar ya que a final de cuentas es hacer conteos. Lo que hicimos fue tomar una columna de nuestro data frame, en este caso la columna de **airlineid** con todas sus filas.

Después a este data frame de una sola columna le agregamos una columna extra que consta de puros "1"s. Esto con la finalidad de después agrupar los elementos iguales en nuestra columna de interes (airlineid) y poder sumar las veces que aparece. Ya que tenemos los conteos (recordemos que las filas representan un vuelo) podemos graficar de acuerdo a la cantidad de veces que aparece una misma aerolínea. Decidimos hacerlo para las aerolíneas que tuvieran la mayor cantidad de apariciones en nuestro data frame. Para luego visualizar esto por medio de una gráfica de barras horizontal.

Método para la cantidad de vuelos por año

Realizamos un procedimiento similar al inciso pasado, nada mas que cambiamos en el código la columna que queremos. En este caso nos concentraremos en la columna que contiene el año de cada vuelo (**year**) y le agregamos contadores para luego sumar utilizando años iguales como llaves de un diccionario.

Método para la cantidad de vuelos por mes

Al igual que el ejemplo anterior, lo que hacemos ahora es enfocarnos en la columna de nuestro data frame que corresponde a los meses (**month**). Considerando que los meses están descritos de forma numérica. Es decir, 1 =enero, 2 =febrero, etc.

Algo que es bueno recalcar, es que antes de manipular de cualquier forma nuestra base de datos, creamos una copia del mismo, para evitar cualquier problema de borrado que no tenga solución mas que cargar la base de nuevo. Despues utilizamos el mismo método para la creación de un histograma con la particularidad de utilizar nada mas un color (con la finalidad de representar mejor la uniformidad de las barras).

Método para el mapa de correlaciones

Para entender mejor nuestro conjunto de datos, deseamos conocer las relaciones que existen entre las variables, por lo que usaremos la matriz de covariación para conocer que variables podrían estar fuertemente o débilmente relacionadas. Gracias que utilizamos los dataframes de pandas y la biblioteca de seaborn podemos llamar los método para encontrar la matriz y generar un heatmap para nuestra visualización.

Método para salidas totales realizadas por operaciones de servicios programados por cada aeropuerto

Para analizar el total de salidas totales de los aeropuertos agrupamos los registros a partir del código de los aeropuertos de Estados Unidos (usg apt) y extranjeros (fg apt), de esta forma podemos aplicar una función de agregación para sumar el total de vuelos de cada aeropuerto y ordenarlos de manera descendente.

Por último, visualizamos nuestros datos con un gráfico de barras con colores distintos y claros para diferenciar a cada aeropuerto y de esta forma hacer énfasis en cada instancia distinta.

Método para el comportamiento a lo largo del tiempo con el aeropuerto Miami

Para generar una visualización sobre el comportamiento de un aeropuerto en particular a lo largo del tiempo primero filtramos sobre los datos el aeropuerto de Miami,despues agrupamos nuevamente para aplicar una función de agregación que contará con los vuelos totales, también debemos considerar indexar las fechas y ordenarlas de manera descendente para que de esta manera podemos generar nuestra gráfico desde el inicio y fin de las fechas marcadas.

Por ultimo, hemos optado por utilizar una gráfica de puntos y líneas para nuestra serie de tiempo, ya que no ayudar visualizar de manera clara los vuelos a través de tiempo utilizando colores y marcas simples.Y con las marcas de símbolos de puntos podemos observar fácilmente las subidas y bajas en la gráfica.

Método para el Top 15 de aeropuertos con mayor número de vuelos programados y vuelos chárter.

Para visualizar el Top 15 de ambos tipos de vuelos, agrupamos los registros por el código único asociado a cada aeropuerto (*usg apt*) para así poder usar la función de agregación *sum()* y sumar los vuelos programados y chárter por aeropuerto, respectivamente. Ordenamos de mayor a menor y nos quedamos con los 15 registros que encabezan la lista.

Obtenido ya el subconjunto de datos necesarios, renombramos los ahora índices de nuestro dataframe (*usg apt*) usando el nombre completo del aeropuerto, esto para facilitar el entendimiento de la visualización (no se espera que el público conozca el código de cada aeropuerto de Estados Unidos). Finalmente, representamos el resultado mediante un gráfico de barras horizontal, usando un sólo color pastel para crear una visualización uniforme (sólo mostramos un "tipo" de dato).

Con el tamaño de imagen especificado, no se alcanza a visualizar el aeropuerto internacional 'Anchorage' ubicado en Alaska; sin embargo, no ampliamos la imagen para no perder calidad en la visualización de los demás aeropuertos (sabemos de antemano que, si bien Alaska forma parte de los estados de US, su ubicación se encuentra alejada de los demás estados).

Método para la ubicación de aeropuertos con más vuelos programados y chárter.

Usando el kit de herramientas de Basemap matplotlib, visualizaremos de manera conjunta la ubicación del Top 4 de vuelos programados y el de vuelos chárter (dado que dos aeropuertos aparecen en ambas listas, la visualización constará de 6 ubicaciones).

Para especificar la localización de los nodos(aeropuertos) buscamos manualmente las coordenadas (latitud y longitud) de cada aeropuerto. A su vez, nombramos de forma manual a cada nodo, según el aeropuerto al que representan.

Puesto que el mapa usa verdes y azules en tonalidades opacas, representamos cada aeropuerto mediante un nodo anaranjado con saturación y brillo medio, así como un etiquetado con su nombre principal en letras blancas.

Resultados y Discusión

Visualización de aeropuertos fuera de EEUU con mas vuelos

Para esta visualización usamos los atributos relacionados con Foreign Gateway Airport Code (fg_apt). Con estos datos podemos hacer un conteo de cuantos registros tienen asignado un mismo código.

En total contamos con 1304 aeropuertos diferentes fuera de EEUU en nuestra base de datos. Para poder tener una visualización clara vamos a seleccionar aquellos que tienen al menos 5000 vuelos registrados.

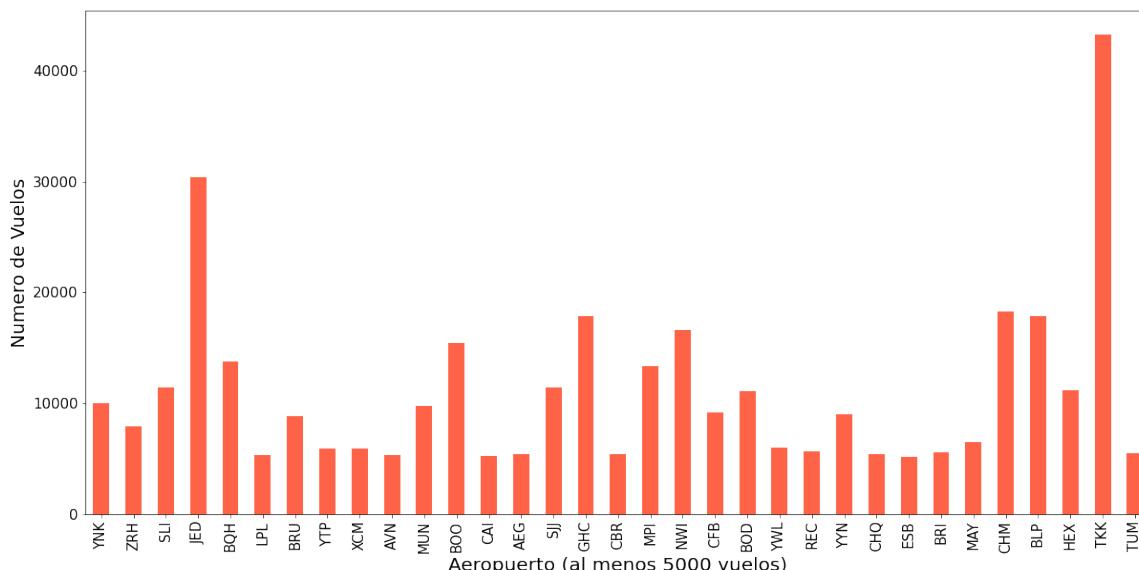


Figura 1: aeropuertos fuera de EEUU con mas vuelos

Visualización de aeropuertos en EEUU con mas vuelos

Para esta visualización usamos los atributos relacionados con US Gateway Airport Code (usg_apt). Con estos datos podemos hacer un conteo de cuantos registros tienen asignado un mismo código.

En total contamos con 832 aeropuertos diferentes dentro de EEUU en nuestra base de datos. Para poder tener una visualización clara vamos a seleccionar aquellos que tienen al menos 5000 vuelos registrados.

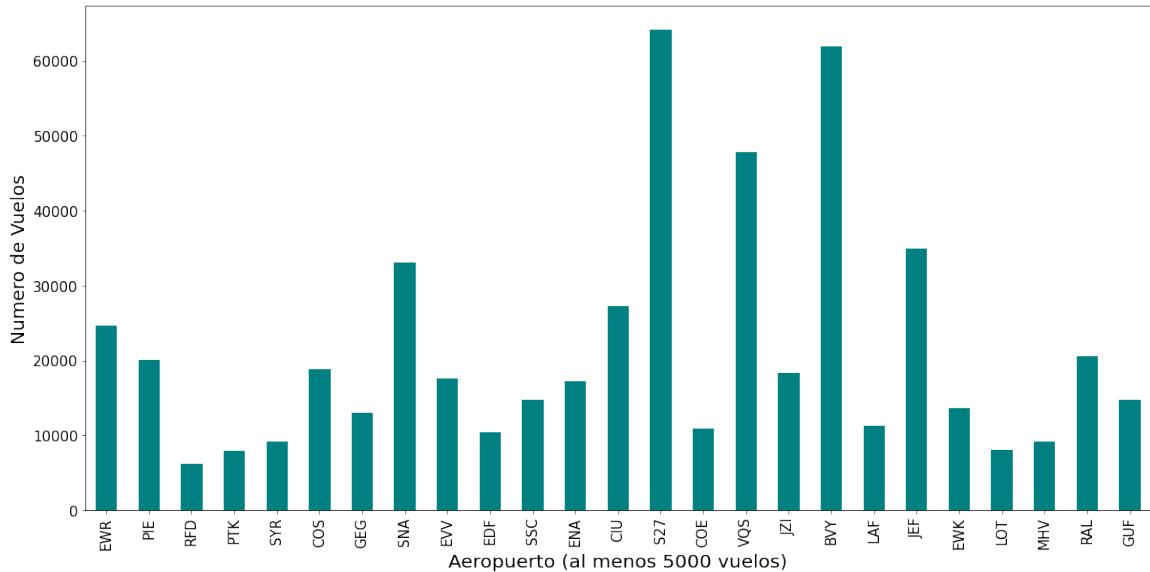


Figura 2: aeropuertos en EEUU con mas vuelos

Visualización geográfica de los 5 aeropuertos con mas vuelos en EEUU

Para esta visualización se tomaron los 5 aeropuertos en EEUU con más vuelos y se obtuvo sus coordenadas geográficas ya que en nuestra base de datos no se nos proporcionan. Una vez que ya tenemos sus coordenadas marcamos los puntos en donde se encuentra cada uno de los aeropuertos.

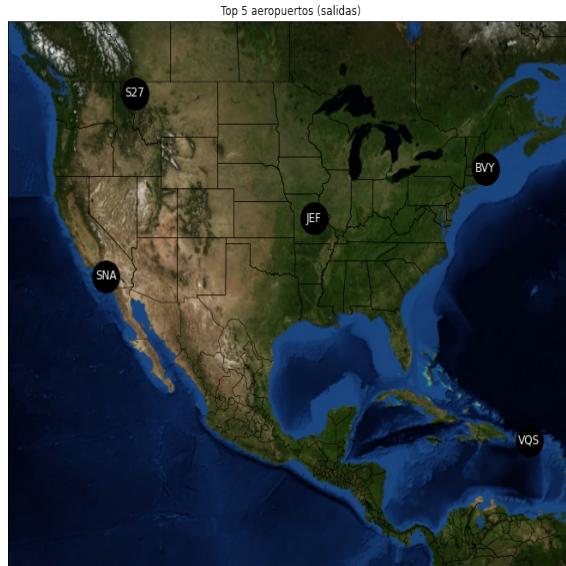


Figura 3: top 5 aeropuertos en EEUU con mas vuelos

Visualización de aerolíneas con mas vuelos

Los datos que tenemos contienen una columna que es el identificador de las aerolíneas. Por lo tanto, si tenemos en nuestra base de datos todos los registros de vuelos, algo que nos podría interesar es saber que aerolínea es la que aparece mas en nuestros registros.

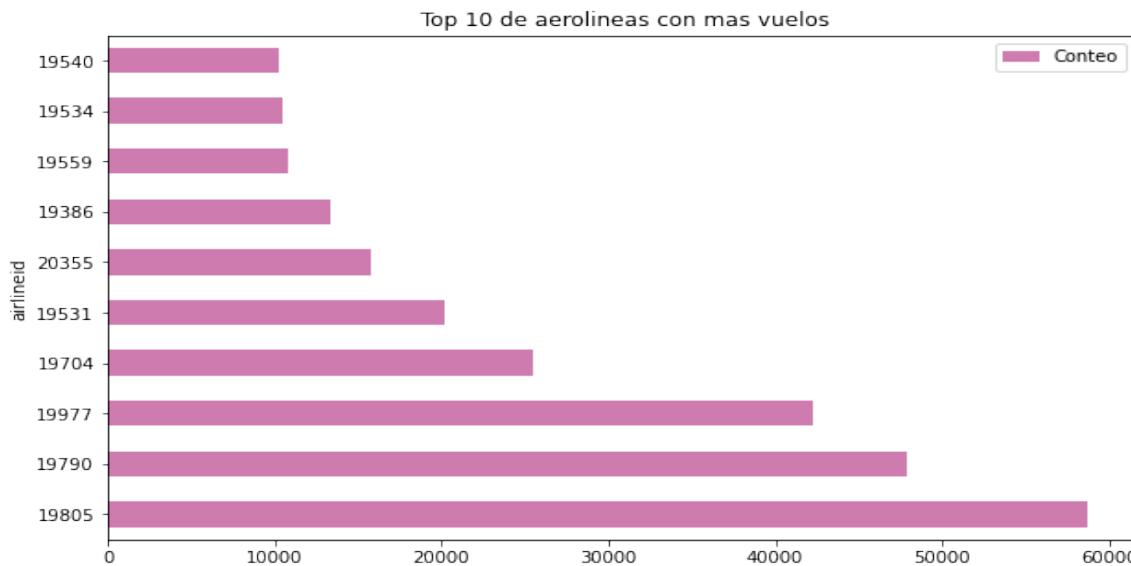


Figura 4: Top 10 aerolíneas de nuestros registros

Como podemos observar en la visualización, tenemos el top 10 de aerolíneas que mas aparecen en nuestra base de datos, cuando cada registro representa un vuelo. Para ello utilizamos los identificadores que vienen en nuestra base de datos los cuales representan alguna aerolínea en específico.

El problema es que estos identificadores no son los que mas se usan, por lo tanto encontrar a que aerolínea pertenece cada identificador es mas difícil de lo que parece. Después de buscar un rato por el internet, encontramos una página que te permite ingresar un código y te diga que aerolínea esta registrada con ese código. Sin embargo, a la hora de intentar buscar nuestra primera opción en la visualización ("19805") no nos aparecen resultados. Por otro lado, si solo utilizamos los últimos tres dígitos, si nos aparece una aerolínea de los Estados Unidos, lo cual hace completo sentido. Esto es porque si investigamos, normalmente se utiliza nada mas un código de tres dígitos para identificar a las aerolíneas a diferencia de como nos viene en nuestra base de datos.

Por lo mismo, podemos dar una lista de las seis primeras aerolíneas con sus nombres:

- Mercury Air Cargo, Inc.
- Amadeus IT Group SA
- PT. Sriwijaya Air
- No pudimos encontrar la aerolínea
- Piedmont Airlines, Inc. (U.S. Airways Express)
- Estafeta Carga Aerea, S.A. de C.V.

Como podemos observar, de acuerdo a nuestra base de datos, la aerolínea que mas vuelos realiza es Mercury Air Cargo, Inc.

Visualización de vuelos por año

Otra de las cosas que nos gustaría saber es la cantidad de vuelos que hay por año en nuestra base de datos. Aprovechando que no solo tenemos la fecha en formato de fecha, si no que también tenemos una columna específica donde está el año.

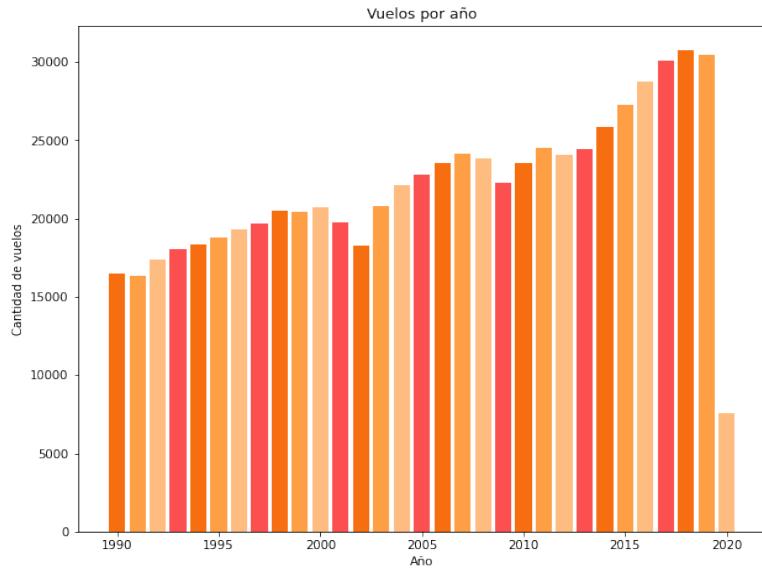


Figura 5: Vuelos por año

Aquí sucede algo interesante. Podemos observar como la cantidad de vuelos podríamos decir que es creciente a lo largo de los años, pero en el 2020 tiene una reducción bastante considerable.

A medida que la tecnología avanza se vuelve cada vez mas utilizada y barata. Cada vez se vuelve mas fácil de producir y vender a las masas. Podemos ver esto en la cantidad de personas que tienen un celular inteligente hoy en día o la cantidad de coches que existen en comparación a años anteriores. No es diferente para la aviación. Sin embargo, esta visualización nos permite confirmar la cantidad creciente de vuelos por año, pero sobre todo nos permite ver el caso particular de 2020.

Como sabemos, en el año 2020 comenzó la pandemia mundial causada por el COVID-19. Y una tendencia creciente en cuanto a la cantidad de vuelos que tenemos en nuestra base de datos dio un bajón importante. Podemos observar que por la pandemia mundial, la cantidad de vuelos fue menos de la mitad de la de 1990. Así que con esta visualización podemos observar como es que existe una anomalía y en caso hipotético de no saber que existió una pandemia en ese año, la misma visualización nos haría investigar la causa de la poca cantidad de vuelos.

Visualización de vuelos por mes

Este caso puede parecer muy similar al anterior, sin embargo nos da información completamente diferente. Ahora en vez de contar la cantidad de vuelos que hubieron por año, lo que hacemos es contar la cantidad que hubieron por mes en todos los años. De nuevo, esto es algo sencillo gracias a que nuestra base de datos contiene una columna la cual indica el mes nada mas, a parte de tener una columna la cual tenga la fecha.

Como podemos observar, en nuestra visualización los meses están indicados por números. Es decir, lo normal, enero es 1, febrero es 2, y así hasta el 12 que es diciembre.

En la visualización queda de una forma clara que todos los mes (de nuestra base de datos) cumplen con la característica de sobrepasar la cantidad de 5000 mil vuelos.

Teniendo en cuenta que esta visualización nos muestra la cantidad de vuelos por mes de todos los años, notamos como no existe una diferencia muy grande entre los meses. Es decir, no existe un valor muy abrupto como el de la visualización de vuelos por año. En este caso tampoco es que siga una tendencia aparente. De hecho si tuviéramos que comparar esta visualización a la de una variable aleatoria, lo primero que nos viene a la cabeza es la gráfica de una variable uniforme.

A partir de esta visualización podríamos decir que en general, los vuelos por mes no varían mucho, es decir que cada vez consta de una cantidad similar de vuelos.

Esto último es importante porque nos ayuda a tener una visión menos sesgada de las cosas. Ya que un sentimiento que se tiene en general, es que diciembre/ enero, o junio/julio, son meses donde hay mas vuelos porque mas personas salen de viaje. Y con esta gráfica podemos observar que de hecho eso no es cierto.

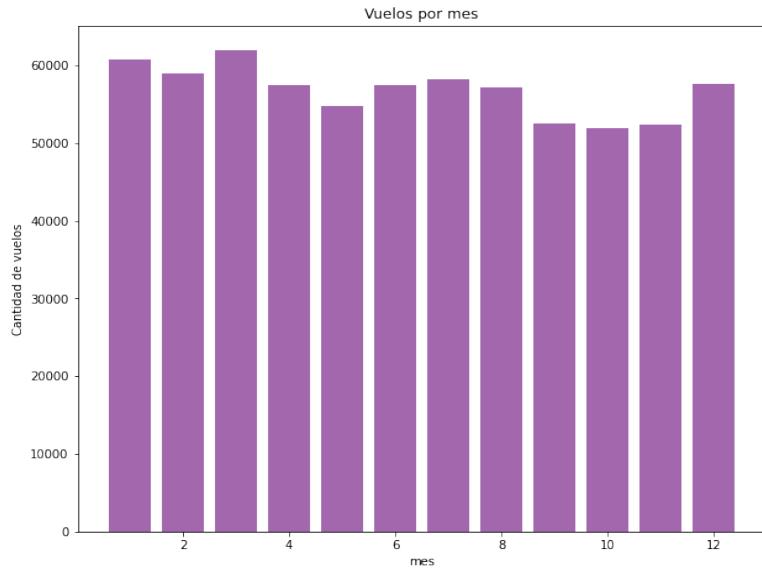


Figura 6: Cantidad de vuelos por mes

Visualización de correlaciones

Para entender mejor nuestro conjunto de datos, como primer paso hicimos una mapa de correlación entre las variables a través de su matriz de covarianza:

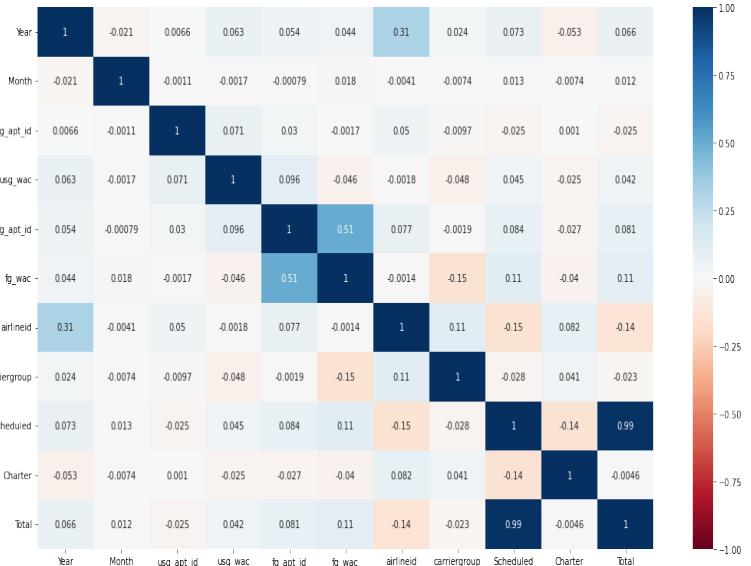


Figura 7: Mapa de correlación

Observamos a partir del mapa que las relaciones positivas más fuertes eran entre las variables Scheduled y Total. Las demás variables parecer ser que no hay una relación muy fuerte entre ellas. Para nuestro mapa escogimos una paleta de colores clara utilizando el contraste con los los colores fuertes azul y rojo, de esta manera es fácil de observar que variables están fuertemente relacionadas ya sea positiva o negativamente.

Visualización para salidas totales realizadas por operaciones de servicios programados por cada aeropuerto

Es importante conocer los vuelos totales de cada aeropuerto para de esta forma saber los más usados, y esto es posible gracias a los registros de que se ha tomado, utilizando los campos Scheduled, Charter y Total de nuestro conjunto de datos, podemos conocer la cantidad de viajes de las puertas Estados Unidos a hasta las puertas extranjeras de aeropuertos:

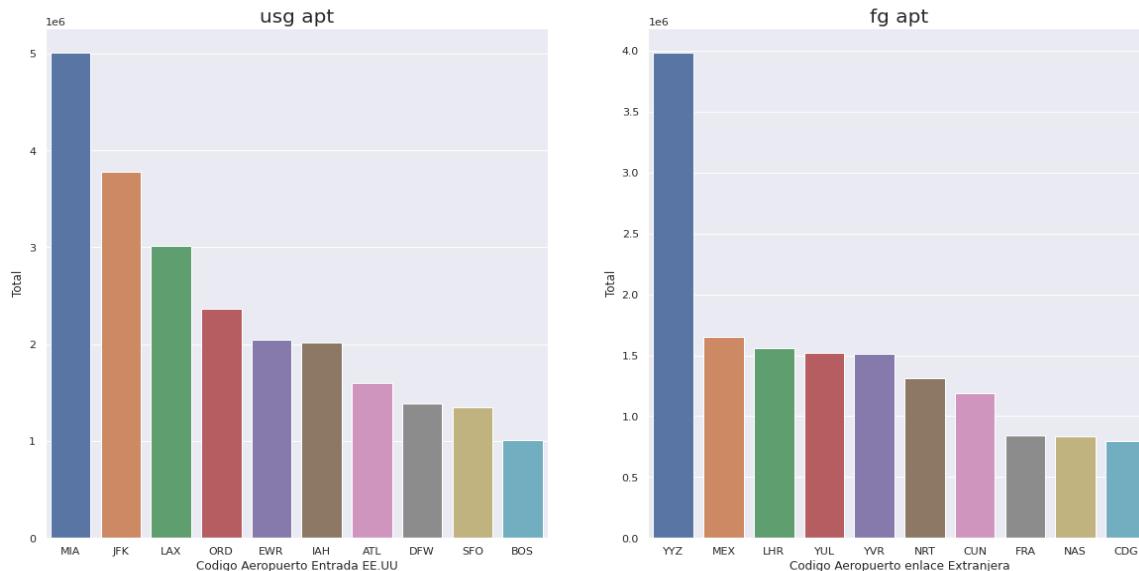


Figura 8: Top 10 de las aeropuertos con más viajes

A través de nuestra visualización podemos ver que el aeropuerto en Estados Unidos que más viajes hace es el aeropuerto de Miami international seguido del aeropuerto de John.F.Kennedy de New York. Por otra parte podemos también observar los aeropuertos extranjeros a los que viajan, en este caso, vemos que el número uno es el aeropuerto de Lester B Pearson International en la ciudad de Toronto en Canada y segundo lugar el aeropuerto Licenciado Benito Juarez International de la Ciudad de México. Para distinguir los diferentes aeropuertos en la gráfica utilizamos diferentes colores para cada contenedor, de esta forma, podemos hacer énfasis que son instancias distintas utilizando colores claros en cada contenedor.

Visualizando el comportamiento del aeropuerto de Miami a lo largo del tiempo

Ahora que sabemos que el aeropuerto que más viajes realiza es el de Miami podemos ver su éxito a lo largo del tiempo, es decir, ver si siempre ha sido elegido por los clientes o ha ido mejorando a lo largo del tiempo.



Figura 9: Miami international Airport a traves del tiempo

Vemos en nuestro visualización varias sorpresas, al parecer el aeropuerto de Miami no siempre tuvo una gran cantidad de vuelos, sino todo lo contrario, fue aumentando los vuelos a lo largo de tiempo llegando a un máximo entre los año de 1995 a 1998, también observamos subidas y bajadas en nuestra gráfica hasta que en el año 2020 existe un mayor decremente en los vuelos del aeropuerto, esto da la visión que estaba en el top 10 de los aeropuertos con mayores viajes por los vuelos realizados a través del tiempo, es decir, en actualidad podemos esperar otros aeropuertos. Por ultimo, hemos optado por esta visualizaciones utilizando colores y marcas simples, ya que estamos utilizando una serie de tiempo con las fechas registras. Con las marcas de puntos podemos observar fácilmente las subidas y bajas en la gráfica.

Visualización de los vuelos del aeropuerto Miami international

Para finalizar nuestro análisis sobre el aeropuerto de Miami decidimos visualizar los vuelos a través de una gráfica donde cada nodo son los aeropuertos extranjeros a donde se llegará, de esta podemos ver los viaje a través de las aristas de nuestro gráfica. Para una mejor visualización decidimos usar un mapa para proyectar los nodos y las aristas, esto nos ayudará ver el trafico de los vuelos en este aeropuerto.



Figura 10: Vuelos hacia el aeropuerto de Miami international

Visualización del Top 15 de Aeropuertos con mayor Número de Vuelos Programados.

Al momento de elegir un vuelo, queremos que éste sea en un aeropuerto de confianza y una de las maneras en que decidimos si es de confianza o no es revisando su historial: mientras más vuelos se hayan llevado a cabo, suponemos que el aeropuerto tiene más experiencia.

Con esto en mente, visualizamos los 15 aeropuertos con mayor número de vuelos programados (no chartér).

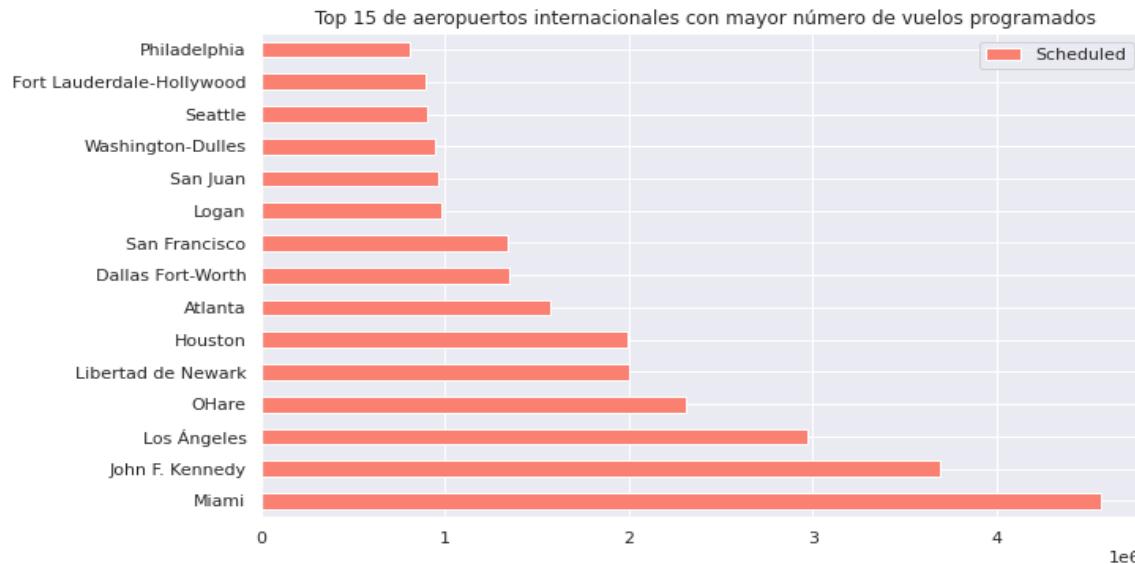


Figura 11: Top 15 de aeropuertos con mayor número de vuelos programados

Como podemos observar, el Top 3 de aeropuertos se encuentra bastante lejos de los demás, llegando o superando los 3 millones de vuelos programados (que se tengan registro). Por lo anterior, si queremos un vuelo en un aeropuerto confiable, nuestras primeras opciones deberían ser el: Aeropuerto Interdacional de Miami, Aeropuerto Interdacional John F. Kennedy, Aeropuerto Interdacional de Los Ángeles, en ese orden.

Visualización del Top 15 de Aeropuertos con mayor Número de Vuelos Chárter.

Al igual que en la visualización anterior, nos interesa conocer los aeropuertos con más experiencia (y más populares) en vuelos chárter aunque para estos, por ser vuelos privados, no debe ser muy necesario un top tan amplio ya que la mayor parte de vuelos seguro se concentra en unos cuantos aeropuertos.

Como habíamos pensado, la mayor parte de vuelos se concentra en un solo aeropuerto: el Aeropuerto Internacional de Miami, juntando más de 400,000 vuelos, mientras que la mayoría no pasa de los 50,000.

De esta manera, quedándonos con el Top 3 (que sobresalen un poco de los demás), habría que elegir entre el Aeropuerto Internacional John F. Kennedy con 79,759 vuelos programados, el Aeropuerto Internacional de Anchorage con 99,728 vuelos, y el Aeropuerto Internacional de Miami con 446,417.

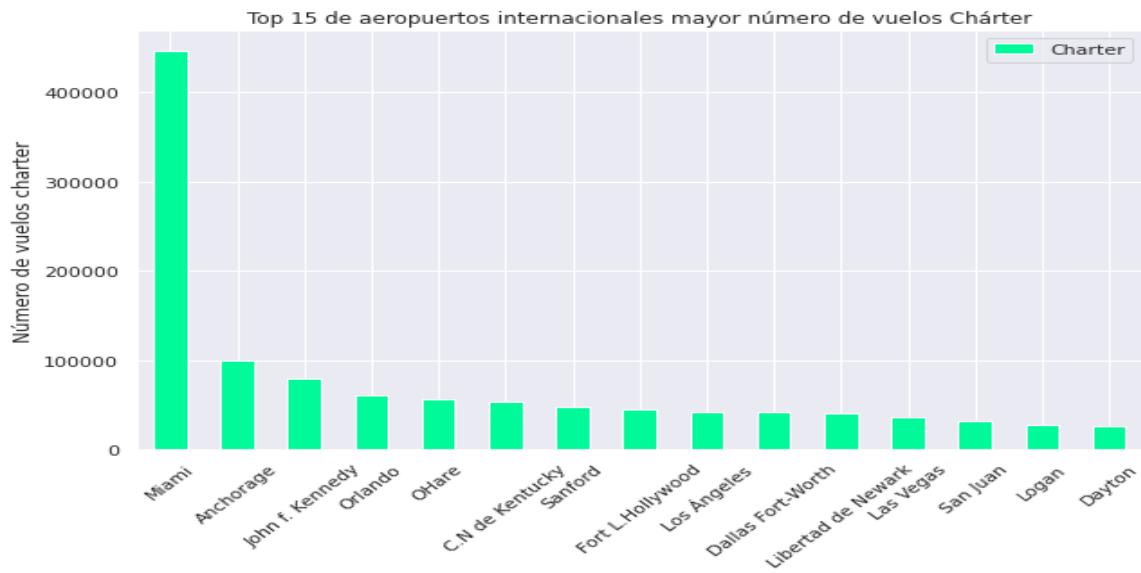


Figura 12: Top 15 de aeropuertos con mayor número de vuelos chárter

Visualización de la Ubicación de algunos Aeropuertos.

Finalmente visualizaremos la ubicación de los aeropuertos analizados anteriormente (únicamente el Top 4 de ambos tipos de vuelos) para determinar si tienen alguna relación clara dada su ubicación.

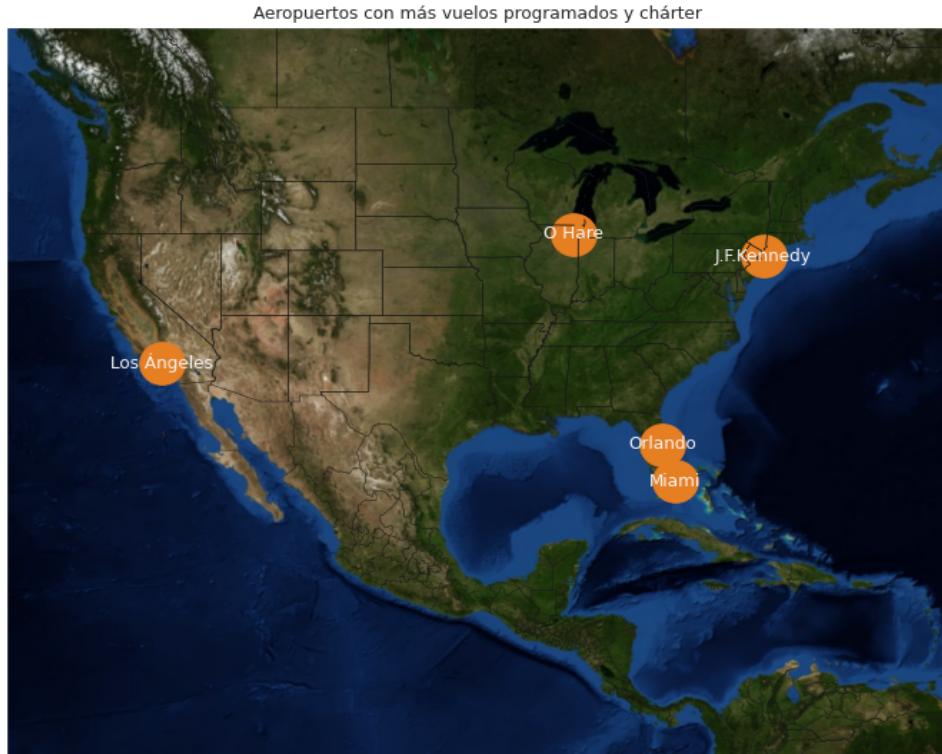


Figura 13: Ubicación de Aeropuertos con más Vuelos Programados y Chárter

Como podemos observar, únicamente se visualizan 5 de los 6 aeropuertos con más vuelos, esto se debe a que el aeropuerto internacional 'Anchorage' se ubica en Alaska. Ahora bien, vemos que los aeropuertos están distribuidos en todo el territorio Estadounidense (incluyendo a Alaska), por lo que concluimos los vuelos se concentran en un

sólo aeropuerto internacional por cada zona de los Estados Unidos.

Conclusiones

Tenemos diversas visualizaciones las cuales cumplen distintos objetivos. Sin embargo tenemos conclusiones las cuales son generales para la mayoría de las visualizaciones.

Por como es la base de datos, existen algunas columnas las cuales no nos sirven de mucho a la hora de hacer una visualización. Esto porque puede que solo sean identificadores o códigos de puertas de aeropuertos, etc. Pero esto puede ser por no tener un objetivo como tal con la base de datos. Por ejemplo, si alguien nos diera esta base de datos, y nos pidiera que buscáramos las puertas que mas aviones reciben, para darles mantenimiento, entonces esos datos nos serían útiles. Es decir, sin objetivos específicos, algunas de las columnas de nuestra base de datos no fueron tan utilizadas.

Por otro lado, varias de las visualizaciones resultaron ser gráficas de barras. Esto por la naturaleza de la base de datos y de la materia en sí. Ya que algo muy requerido es contar cosas. Es ver que aeropuertos son mas utilizados, que aerolíneas tienen mas viajes, etc. Contar es información básica que nos ayuda en casi cualquier base de datos. Sin embargo, las visualizaciones que mezclan un mapa con grafos son algo de interés para esta práctica en particular y por lo tanto cabe recalcar algo de cada una.

En la visualización geográfica de los 5 aeropuertos con mas vuelos en EEUU, podemos observar que están extrañamente esparcidos. Es decir, que están a lo largo de EEUU. Nos muestra como de cierta forma, estos aeropuertos con mas vuelos cubren en general todas las zonas de EEUU. No están aglomerados en una sola zona. Lo cual nos muestra que no hay una zona de personas de EEUU que viaje mas o que reciba mas. Esta balanceado.

En la visualización de ubicación de algunos aeropuertos sucede algo un poco diferente. Ya al considerar ambos tipos de vuelos (de carga y pasajeros), podemos ver como los 5 aeropuertos con mayor demanda ya se distribuyen de una forma no tan uniforme. Podemos observar como se acumulan mas aeropuertos en la zona este de EEUU. De hecho todos (sin contar Alaska) están en la mitad del este de EEUU a excepción del aeropuerto de Los Ángeles.

Al comparar las dos visualizaciones podemos concluir que al considerar nada mas vuelos con pasajeros, los aeropuertos mas utilizados se distribuyen de forma uniforme por EEUU. Sin embargo al considerar vuelos de carga, podemos ver como otros aeropuertos (sobre todo al este de EEUU) son los que toman protagonismo. Algo en lo que coinciden es en Los Angeles. Lo cual nos puede llevar a pensar que ese aeropuerto es importante tanto a nivel comercial (de pasajeros) como de carga. Lo mismo podemos decir del aeropuerto del J.F.Kennedy.

Referencias

- [1] Wilke, Claus O.: *Fundamentals of Data Visualization*, O'Reilly, 2019.
- [2] Seaborn: *User guide*. <https://seaborn.pydata.org/>. Accedido en Abril 24 del 2022.
- [3] Database, Airport: *User guide*. <https://www.partow.net/miscellaneous/airportdatabase/index.html>. Accedido en Abril 24 del 2022.

Código:

<https://deepnote.com/project/Proyecto1Viz-13853cea-fe70-48fe-93bf-666cb5f7137b/>