En aquesta **TASCA** has d'entregar un document PDF d'un màxim de 4 pàgines en el qual aportis la següent informació sobre el teu projecte final:

* **Presentació del conjunt de dades escollit:** explicació i observacions generals del conjunt de dades a utilitzar per a fer el projecte final.
* **Característiques generals:** sense necessitat d'entrar molt en detall, explicar les principals característiques que defineixen aquest conjunt de dades. Tipologia, sector, tipus de dades, font, context, etc.
* **Definició de les variables:** explicació teòrica de les principals variables que conté el conjunt de dades.
* **Presentació dels objectius:** detallar els objectius inicials marcats de cara a extreure informació rellevant del conjunt de dades.

Base de datos:

Para el proyecto utilizaré una base de datos de una línea aérea española con datos del 2022 hasta Mayo. Los datos son reales y la estructura del fichero es esta:

A esta base de datos se complementa con otras bases de datos del 2019 y del 2020 de la misma compañía. El principal problema es que la operativa de la compañía en este periodo causado por el COVID por lo que la información es difícil poder contrastarla porque el periodo es muy distinto, para minimizar el impacto de los retrasos lo que han hecho es poner un buffer de 1 hora después de 3 saltos que reduje el impacto del retraso.

Otra base de datos será de otra compañía que opera en Barcelona pero que principalmente opera en la terminal 2 para poder hacer un bench marking.

Las columnas son muy semejantes entre cada base de datos en el contenido de la columna, pero no iguales. Una tiene la columna fecha y en otro tiene el campo hora. Hay otras bases que en la columna tiene el campo fecha y hora juntos que ayuda mucho en los cálculos.

El retraso es motivado porque el avión sale tarde, y se retrasa más porque pierde el SLOT dado por el controlador aéreo (ATC) lo que obliga al avión a esperar hasta que el ATC dé otro SLOT para minimizar el impacto de sobrecarga por aviones volando.

Las fases de vuelo son embarque pasajeros y maletas, cierre puertas, taxi a cabecera pita, despegue, vuelo, aterrizaje, taxi hasta terminal y apertura de puertas.

En el vuelo de un avión hay unos tiempos predefinidos que son el STD y STA, (la hora de despegue y hora de aterrizaje), que se puede ver en las pantallas de los aeropuertos o páginas de compra de billetes. Estos horarios son fijos y son las referencias que hay de tiempo.

Tenemos que analizar si un avión sale antes de su STD, que lo puede hacer si tiene permiso del ATC y todos los pasajeros7maletas están dentro del avión. Esta situación siempre ocurre en el primer salto que el avión hace en el día. En los saltos posteriores existe el impacto en el retraso que el avión acumula de los asaltos anteriores.

Hay que analizar si el tiempo de vuelo es posible reducir el retraso yendo el avión más rápido o acortando en la aproximación al no tener que hacer rodeos (con autorización del ATC), o porque el tiempo calculado es más largo que la media del vuelo normal.

Los tiempos de boarding, en teoría son iguales independientes del avión y el aeropuerto que esté el avión, pero sabemos que por ejemplo en el aeropuerto de Barcelona estos tiempos son mayores (aeropuertos grande) que aeropuertos más pequeños porque el impacto en el control de los SLOT por congestiones del tráfico aéreo es menor y el personal de tierra puede controlar mejor el tiempo de boarding y landed

Una cosa que nos encontraremos es que en los retrasos existe muchos outliers con respecto al modo de operar norma de un avión. Puede darse el caso que el avión sufra un retraso importante porque el motor se ha estropeado, decido que estas situaciones se escapan al estudio de normal de trabajo y eliminaré todos los vuelos que su retraso y adelanto del vuelo sea superior a 30 minutos. Estos outliers solo dan ruido y muy poca información. Analizaré que porcentaje de vuelos tengo estén afectados por este motivo y si es inferior a 5% consideraré mi asumtion ok.

Sobre la calidad de la información (MSA, measure sistema analysis) tengo que decir que los pilotos no todos proceden del mismo modo, porque como están muy controlados en sus objetivos personales el tiempo de retraso, controlan el momento 0 quitando el freno e inmediatamente cambia de estatus el avión e imputa este retraso a otro motivo, y eso hay que tenerlo en cuenta a la hora de crear el modelo.

Un tema importante es saber cuál es la puerta de embarque porque el tiempo de operativa de embarque desembarque no es igual si se realiza con finguer o en la pista a que se tiene que coger un autobús.

Un avión como elemento individual tiene poco margen de maniobra tiene y se ve supeditado a las órdenes del ATC y slots. Por ejemplo, con un coche podemos salir tarde, y podemos escoger ir por autopista o por carreteas, o escoger las calles más anchas o buscar atajos, o parar si queremos o necesitamos. En la aviación no se puede. Si hay muchos aviones en circulación por seguridad hay que separarlos en el espacio 3D y eso se hace por tiempo. Los aviones siempre están volando lo que implica desplazamiento, y consumo de combustible. En conclusión, que el nº de aviones volando en un sector aéreo nunca puede ser infinito, sino que finito.

Otro elemento que distorsiona mucho, es el tiempo atmosférico, si hay niebla, lluvia día noche, visibilidad etc. En situación extremas se obliga los aviones a aumentar esta distancia de separación lo que fuerza haber más retrasos. Estos retrasos impacto en otros aeropuertos que pueden tener unas condiciones de sol y visibilidad superior a 30 km.

Está claro que en meses de invierno las condiciones ambientales son peores en general en Europa, pero no en los países de otras latitudes. Y en periodos de verano y en fechas muy señaladas como puede ser principio de semana, fin de semana, Navidades los aviones se saturan más hay muchos más SLOTs por hora lo que provoca interferencia en las puntualidades entre compañías. Es importante hace run mapa de calor de retrasos para ver si hay periodos (mes/semana/día especifico) con más impacto.

Lo que está claro que en el primer salto, el piloto tiene la capacidad de poder salir 10 minutos antes que es el tiempo que obligan al pasaje a estar antes del embarque y no se ve influenciado por ejemplo con pasajeros que van tarde por retrasos en otros vuelos en saltos posteriores. y que el avión debe esperar a que todo el pasaje/ equipaje esté abordo.

La estructura de la base de datos por cada línea esta información principal de tiempos de un salto de un avión. Por la referencia podemos saber en ese día que vuelos (saltos) ha hecho. Si filtramos por fecha, matrícula de avión, tenemos todos los saltos que ha hecho un avión ese día.

Como está ordenado por STD o por el ATD, podemos saber que el que tiene las STD más baja (hora más temprana) es el primer vuelo del día, el segundo es el siguiente vuelo, y así sucesivamente.

Es importante identificar que un avión puede llegar al día siguiente y eso hay que tenerlo en cuenta porque la columna fecha solo da el día de despegue. Pueden salir tiempo de vuelo absurdos.

**Objetivos:**

Analizar las causas generales del retraso de los aviones, y focalizándonos en el 3 salto, (tercer vuelo) si el retraso es superior a 10 minutos que es un tiempo acordado por las compañías como No retraso, saber qué porcentaje de aviones llegan a la hora.

Me quiero focalizar especialmente en el aeropuerto de Barcelona y ver el impacto en el retraso que hay.

