1. **Abstract**: Resum del vostre projecte. Què conté, objectiu i resultat. Molt resumit

La puntualitat i el preu son les dos coses que els passatgers més valorem a l´hora de valorar una companyia aerea. En aques cast, tinc que esbrinar quin es la puntualitat d’una companyia aèria en lo que ells denomina’n el tercer salt, fent servir una base de dades proporcionada per ells.

1. **Introducció**: Aquí fem una introducció al nostre problema dins del marc de la Data Science. Així com objectius que tenim i els requisits que hem de complir.

La base de dades te varies columnes amb les dates de enlairament i aterrament de cada avió. Un avió en un dia fa varis “salts”. Primer tinc que adequar els camps temps a un format date per poder fer la resat de temps. Tinc que identificar el vols de cada avio per dia, ordenar-los per hora de sortida per poder identificar cada salt i crear una seqüencia. Això obligarà a treballar en un format de línia perquè el temps calculats de cada salt estiguin en una sola línia per després poder filtrar el fitxer per tenir tota la informació de un avió per dia, el tres salts. Amb això reduirem el temps de càlcul al simplificar el tamany del arxiu en un 80% accelerant el procés. Pesem de 500.000 Líneas a 80.000.

Una vegada el arxiu està muntat podrem fer els principals càlculs i poder fer estudis de capacitat, calculant el Cp y Cpk. Obtenint el valor de % de vols que la seva puntualitat es més gran de 0 o 10 minuts.

*Definicions de Cp i Cpk*

*Tant Cp com Cpk són Índexs de la Capacitat Potencial d'un procés i es mostren en un histograma que arreplega la disponibilitat d'oferir les prestacions exigides a aquell. Els índexs formen part de les eines de control estadístic de processos.*

Tinc que esbrinar si les dades de enlairament aterrament, temps de boarding i temps de vols estàn relacionades entre salts. Tinc que crear una regressió logística en que una sortida es 0 que l’avió arriba a l’hora o 1 que arriba 10 minuts tard del temps planificat.

La base de dades està molt des-balancejada perquè el 9% del avions no arriben a l’hora per lo que si volem tenir un resultat creïble esté que balancejar. Farem servir el mètode ....... per tenir un resultat mes o menys del 50% per el 0 i 1.

1. **State of Art**: Aquí comentem com s'ha resolt en el nostre problema amb anterioritat, els mètodes ja utilitzats i els resultats ja extrets. D'aquesta manera donem una perspectiva de quins objectius volem arribar a batre així com intentar no tornar a fer servir elements ja utilitzats.

No em consta haver vist una base de dades semblant a aquesta perquè son dades bastant confidencials del 2022 i 2019. Segur que les companyies fan anàlisis interns per esbrinar el seu comportament amb la puntualitat. Però intentaré arribar una mica més juny comparant la puntualitat amb un altre companyia per exemple.

Principalment la idea es saber si la puntualitat que tenen es deguda a una raó i esbrinar quina és aquesta raó.

Les bases de dades d’aviació a Europa està molt tancades cosa que als EEUU es poden consultar lliurament, i casi online.

1. **Metodologia**: Aquí explicarem el preprocessat, els models, les mètriques així com llibreries si s'escau.

Farem servir la llibreria ….. Dividirem el fitxer en dades de train i dades de test al 80/20% respectivament. Crearem un model de regressió logística.

La regressió logística múltiple, desenvolupada per David Cox el 1958, és una extensió de la regressió logística simple. És un mètode de regressió que permet estimar la probabilitat d’una variable qualitativa binària en funció d’una variable quantitativa. Una de les principals aplicacions de la regressió logística és la de classificació binària, en què les observacions es classifiquen en un grup o un altre depenent del valor que prengui la variable emprada com a predictor. Els predictors poden ser tant continus com a categòrics

Per què regressió logística i no lineal? Si una variable qualitativa amb dos nivells es codifica com a 1 i 0, matemàticament és possible ajustar un model de regressió lineal per mínims quadrats β0+β1x. El problema d'aquesta aproximació és que, com que es tracta d'una recta, per a valors extrems del predictor, s'obtenen valors de Y menors que 0 o més grans que 1, cosa que entra en contradicció amb el fet que les probabilitats sempre estan dins del rang [0,1].

En aquet cast es podria haver fet servir una regressió lineal però que tenim com variable de sortida un valor continuo, però com un avió arriba tard independent del temps de retard si sobrepassa el 10 minuts es millor veure que predictors son els mes influents.

1. **Resultats**: Aquí és hora de mostrar resultats, taules i gràfiques. Així com remarcar els millors resultats obtinguts.

Com resultat veiem que el tercer salt esta influenciat únicament per el temps d’enlairament, duració de vol i no del temps a terra. No afecta el retards dels salts anterior i la raó es que entre salt i salt està calculat temps de sobra per fer una funció de ‘molla’ que evita que el soroll que pot influencia en un vol/salt no afecta al següent.

Es important de dir que aquesta asseveració s’arriba quan s’eliminen tots el temps superiors a 30 minuts, perquè 30 minuts podem considerar que es un temps dintre de lo normal i que si un avió surt més tard es per una raó que influirà en el vols de tot el dia. Exemples, avió espallat, vaga de controladors, un aeroport tancat, etc, i aquest outliers estàn fora del estudi.

1. **Conclusió**: Aquí tornarem a portar els objectius i determinar si els hem complert o no. Comparar l'state of art així com presentar totes les conclusions extretes.
2. Entre salts existeix una molla que esmorteeix l’impacta entre retards de salts.
3. Haver donat mes tems de boarding en els aeroports grans com Barcelona afavoreix a minimitzar el impacta en el retard.
4. El concepte de puntualitat tindria que canviar i no és No arribar tard 10 minuts, sinó sortir a l’hora. Això es demostra amb el primer salt perquè es el salt amb menys elements exògens per sortir amb retard i veiem que un 10%\_\_\_\_ dels avions ja surten amb retard. En la resta vols ja surten més element exògens com pot ser els anterior, més passatgers que arriben tard al transbords d’un altre vol, un passatger que no puja al avió i està la maleta que este que baixar, o persones que no respecten la mida de les maletes, pujant grans maletes a cabina, que enredereix molt el temps de boarding.
5. Fer benchmarking amb altres companyies es molt bo perque aquí em vist que el comportament en el primer salt tomant com referencia el aeroport de Barcelona en millor per la competència.... I si ells podem ¿per què aquesta companyia no?
6. **Referències**: És important documentar totes les vostres referències!!!