**Activitat DataFrames PySpark a partir dels datasets: flights, planes i airports**

# Estudi de la Petjada de Carboni en funció del temps a l’aire (Petjada de Carboni=airtime):

1. Quin és l’avió («tailnum») amb major petjada de carboni.

2. Quin és l’avió («tailnum») amb menor petjada de carboni. ¿ha estat temps sense volar?

3. Quina és la petjada promig i la petjada mitjana segons «tailnum». Si hi aprecieu diferència notable a què ho atribuiu?

4. Quina és l’aerolinia amb major petjada de carboni.

5. Quina és l’aerolinia amb menor petjada de carboni.

6. Quin és el fabricant d’avions amb major petjada de carboni.

7. Quin és el fabricants d’avions amb menor petjada de carboni.

8. Quin és el model d’avions amb major petjada de carboni.

9. Quin és el model d’avions amb menor petjada de carboni.

10. Fes una taula de petjada de carboni en funció dels anys de fabricació dels avions

11. Quina és l’aerolinia amb major petjada de carboni normalitzat al número de seients disponible a cada vol.

12. Quina és l’aerolinia amb menor petjada de carboni normalitzat al número de seients disponible a cada vol.

13. Quin és el fabricant d’avions amb major petjada de carboni normalitzat al número de seients disponible a cada vol.

14. Quin és el fabricants d’avions amb menor petjada de carboni normalitzat al número de seients disponible a cada vol.

15. Quin és el model d’avions amb major petjada de carboni normalitzat al número de seients disponible a cada vol.

16. Quin és el model d’avions amb menor petjada de carboni normalitzat al número de seients disponible a cada vol.

17. Fes una taula de petjada de carboni en funció dels anys de fabricació dels avions normalitzat al número de seients disponible a cada vol.

18. (Opcional) Fes un gràfic amb la taula anterior.

19. Surt a compte des del punt de vista de la petjada de cada passatger que els avions tinguin més seients, menys o no és una hipòtesi vàlida?

# Estudi de la Puntualitat

1. Quina és l’aerolinia amb més retard?

2. Quina és l’aerolinia amb més retard normalitzat al número de vols operats?

3. Quina és l’aerolinia amb més retard normalitzat al número d’avions disponibles? Afecta tenir més o menys avions (flota)?

4. Fes una taula: en una columna anys de posada en servei i en l’altra de retard. Es pot deduir que l’edat de l’avió afecta a la puntualitat?

5. (Opcional) Fes un gràfic amb la taula anterior.

6. Agafem la taula anterior i hi afegim la granularitat de l'aerolínia és a dir estudiem l’edat dels avions per cada aerolínia. Es pot fer alguna hipótesi relacionada amb la gestió d’alguna aerolínia?

7. (Opcional) Fes un gràfic amb la taula anterior.

8. Agafem la taula anterior i enlloc de granular per aerolínia granulem per «seats». Es pot deduir que a major nombre de seients disponibles més retard? Hi ha colinialitat?

9.(Opcional) Fes un gràfic amb la taula anterior.

10. Fes un ó dos passos més enllà. És a dir, si això fos una entrevista de treball per una aerolínia, presenta una o dues hipótesis més sobre la puntualitat a partir de les dades d’aquests Datasets.

# Estudi de la contaminació de l’aire

Agafarem els enlairaments com a unitat de contaminació en cada aeroport d’origen.

1. Quin és l’aeroport amb més contaminació?

2. Per al global dels aeroports quin és el dia de l’any amb més contaminació?

3. Per al global dels aeroports quin és la hora amb més contaminació? Coincideix en el dia de l’any amb més contaminació?

4. Si superposem els dies (dilluns, dimarts, etc..), quin és el dia amb més contaminació?

5. Si superposem les hores del dia (separem les dades en els agregats dels 24 slots), quina és la hora amb més contaminació?

6. Per a l’aeroport amb més contaminació quin és el dia de l’any amb més contaminació?

7. Per a l’aeroport amb més contaminació quina és la hora de l’any amb més contaminació? Coincideix en el dia de l’any amb més contaminació?

8. Per a l’aeroport amb més contaminació quin és el model d’avió que més contamina?

9. Per a l’aeroport amb més contaminació quina és l'aerolínia que més contamina?

10. Si haguessis d’assessorar les autoritats mediambientals de la ciutat on hi ha l’aeroport anterior i et facilitéssin un nou Dataset amb les dades mesurades d’emisions de CO2 de tots els avions («tailnum») en la darrera inspecció tècnica, quin pla els hi proposaries? Detalla a alt nivell, no cal que codifiquis.

# Ampliacions opcionals

1. Utilitza només «airports» i una llibreria gràfica on puguis representar sobre un mapa els punts de les localitzacions dels aeroports.

2. Intenta reproduir les tasques que has fet amb PySpark DataFrames també en Pandas DataFrames sempre que sigui possible.

3. Intenta refer tota la part de petjada en funció de la velocitat enlloc de temps a l’aire.