

EDA à l'aide de Snowflake et Deepnote sur Aircraft 

<<aircraft_db.sql>>

L'objectif de ce projet est d'utiliser Deepnote pour effectuer une EDA sur les données Aircraft chargées dans Snowflake.

Chargez les données dans Snowflake à l'aide de l'IDE web

Tout d'abord, procédez comme suit : créez une base de données / un schéma / un petit entrepôt. Sélectionnez le schéma et l'entrepôt dans une feuille de calcul active.

Copiez le contenu de src/aircraft_db.sql et collez-le dans une feuille de calcul Snowflake pour le charger dans la base de données.

Créez un nouveau projet dans Deepnote pour récupérer ces données.

Vérifiez que vous pouvez vous connecter à cette nouvelle base de données et récupérer correctement toutes les données.

Préparez la visualisation pour répondre aux questions suivantes :

Quel avion a le plus volé ?

Par « effectué le plus de vols », nous faisons référence au nombre de vols et non à la distance cumulée parcourue.

Consultez les données du tableau individual_flights, comptez le nombre de vols individuels par avion, puis améliorez le résultat en indiquant le nom de l'avion au lieu du code. Nous vous fournirons le nombre de vols ainsi que le nom de l'avion.

Quel aéroport a transporté le plus de passagers ?

Calculez le nombre de passagers transportés en multipliant le nombre de vols de chaque type d'avion par sa capacité, le tout regroupé par aéroport.

Il convient de mentionner que chaque vol individuel transportera N passagers, mais que cela entraînera un double comptage du nombre de personnes, car il sera attribué aux deux aéroports (départ et destination) par vol. Par conséquent, N sera attribué à l'aéroport A et à l'aéroport B.

Quelle a été la meilleure année en termes de recettes par passager-kilomètre pour chaque compagnie aérienne ?

Les champs qui suivent les valeurs internationales contiennent souvent des valeurs nulles. Pour simplifier, nous les considérerons comme des zéros, mais cela doit être pris en compte avant toute analyse, car cela pourrait biaiser les résultats.

Processus Calculez la SOMME de tous les RPM (domestiques, internationaux ou les deux à la fois, donc le total), puis sélectionnez le MAX par compagnie aérienne, ce qui nous donnera également

l'année. En sélectionnant un type spécifique de RPM ou les deux, nous obtiendrons des réponses différentes, c'est pourquoi les trois doivent être fournis.

Quelle a été la meilleure année en termes de croissance pour chaque compagnie aérienne ?

Nous disposons des données ASM et RPM, ainsi que du nombre de vols nationaux et, pour certains vols, internationaux, par compagnie aérienne et par aéroport. Voici quelques-unes des façons possibles d'évaluer la croissance : Réduire la différence entre ASM et RPM, ce qui permet d'attirer plus de passagers que la capacité disponible. Augmenter l'ASM, ce qui implique d'augmenter la flotte ou la taille des avions, ou le nombre de vols. Augmenter uniquement le nombre de vols.

Pour simplifier et parce que cela semble être la meilleure réponse, nous utiliserons l'ASM comme indicateur de croissance. Plus une compagnie aérienne aura d'ASM au fil du temps, plus nous dirons qu'elle a connu une croissance.

Pour calculer la croissance, nous prendrons la moyenne (AVG(ASM_Domestic)) par compagnie aérienne et par an.

N'oubliez pas de supprimer ce que vous avez créé !
Supprimez la base de données et l'entrepôt associés.