



Projet Aircraft

Dans le cadre de formation Data Analyst, voici une analyse de données concernant les différents critères de mesure au sujet du domaine aérien à l'aide de DBT, Snowflake et Deepnote.



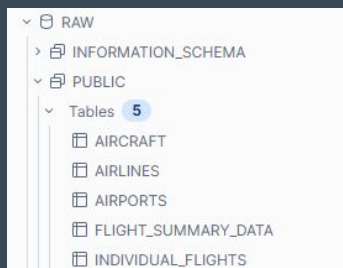


Création des tables dans Snowflake

Extrait du code de création des tables :

```
Home | aircraft_analytics.sql | +
My Workspace > aircraft_analytics.sql
ACCOUNTADMIN | TRANSFORMING (X-Small) | RAW | PUBLIC | Share | ...

1 CREATE OR REPLACE TABLE aircraft (
2   "index" INTEGER,
3   "Aircraft_Id" TEXT,
4   "Aircraft_Type" TEXT,
5   "Mass" INTEGER,
6   "Length" INTEGER,
7   "Cost" INTEGER,
8   "Capacity" INTEGER
9 );
10 INSERT INTO aircraft VALUES(0,'12d','Flockinator',1000000,10000,1000000000,1000),(1,'12a','Miniflock',50000,2000,2000000,200),
11 (2,'g72','Goose',200000,5000,3000000,500),(3,'t10','Thundercat',1500000,20000,400000000,2000),
12 (4,'b23','Bezantium',4000000,100000,1000000000,5000);
13 CREATE OR REPLACE TABLE airlines (
14   "index" INTEGER,
15   "AirLine_Code" TEXT,
16   "AirLine_Name" TEXT,
17   "Description" TEXT,
18   "Market_Cap" INTEGER,
19   "Employees" INTEGER,
20   "Age" INTEGER
```



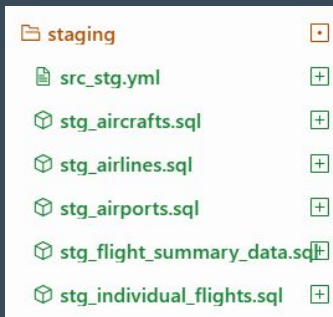
Dans la base de données RAW, le schéma PUBLIC, incrustation des tables Aircraft, Airlines, Airports, Flight_Summary_Data et Individual_Flights.



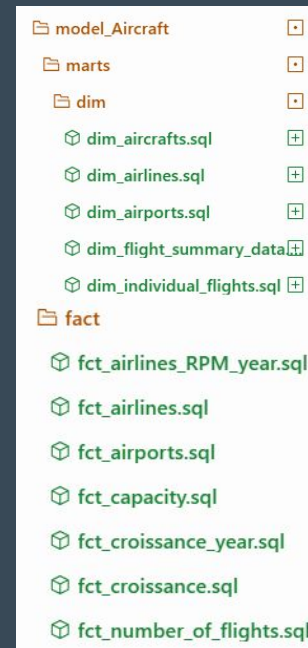


Répartition des documents SQL dans DBT

Mise en place des staging avec les données brutes avec le yaml :



Mise en place des dim et fact pour obtenir les tables de dimensions et les tables de faits :



Le Yaml :

```
model_Aircraft / staging / src_stg.yml

dbt_yaml_files (dbt_yaml_files-latest.json)
1  version : 2
2  sources:
3    - name: aircraft_analytics
4      database: raw
5      schema: public
6      tables:
7        Generate model
8        - name: aircraft
9        Generate model
10       - name: airlines
11       Generate model
12       - name: airports
13       Generate model
14       - name: flight_summary_data
15       Generate model
16       - name: individual_flights
```





Répartition des documents SQL dans DBT

Exemple de staging :

model_Aircraft / staging / stg_aircrafts.sql

```
1 with aircrafts as (  
2     select  
3         "Aircraft_Id" as aircraft_id,  
4         "Aircraft_Type" as aircraft_type,  
5         "Mass" as mass,  
6         "Length" as lenght,  
7         "Cost" as cost,  
8         "Capacity" as Capacity  
9     from {{ source('aircraft_analytics', 'aircraft')}}  
10 )  
11 select * from aircrafts
```

Exemple de fact :

model_Aircraft / marts / fact / fct_number_of_flights.sql

```
1 {{ config (  
2     materialized="table"  
3 )}}  
4  
5 WITH individual_flights_dim AS  
6 (SELECT * FROM {{ ref('dim_individual_flights')}}),  
7  
8 aircraft_dim AS  
9 (SELECT * FROM {{ "dim_aircrafts" }})  
10  
11 SELECT  
12     aircraft_dim.aircraft_type,  
13     individual_flights_dim.aircraft_id,  
14     COUNT(individual_flights_dim.flight_id) as number_of_flights  
15 FROM individual_flights_dim  
16 JOIN aircraft_dim ON aircraft_dim.aircraft_id = individual_flights_dim.aircraft_id  
17 GROUP BY individual_flights_dim.aircraft_id, aircraft_dim.aircraft_type  
18 ORDER BY COUNT(*) DESC
```





Liaison entre DBT et Snowflake



Les différents Staging, tables de dimension et tables de faits sont importés en Snowflake grâce à un dbt run.

Elles sont intégrées dans la base de données Analytics, le schéma DBT_SCARTIGNY.

Ce chemin nous servira pour l'analyse de données SQL effectuée dans Deepnote.

Les staging sont des vues alors que les tables de vues et de faits sont configurées comme des tables.





Problématiques rencontrées :

01

Quel type d'avion a le plus volé ?

02

Quel aéroport a transporté le plus de passagers ?

03

Quelle est la meilleure année pour le Passenger-Mile Payant de chaque compagnie aérienne ?

04

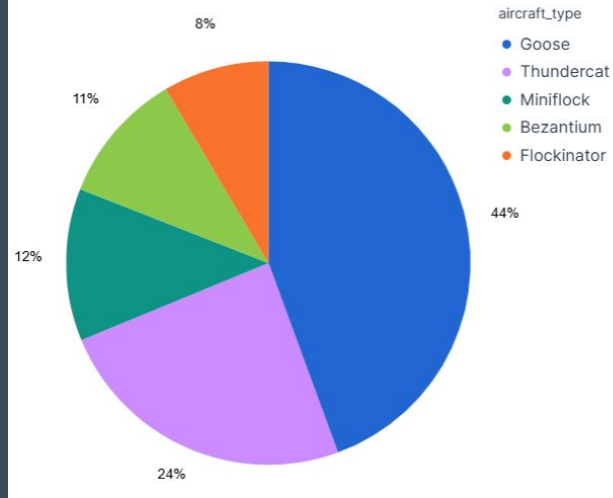
Quelle est la meilleure année de croissance pour chaque compagnie aérienne ?



Quel type d'avion a le plus volé ?



	aircraft_type obj...	aircraft_id object	number_of_flights i
0	Goose	g72	1008
1	Thundercat	t10	553
2	Miniflock	12a	277
3	Bezantium	b23	240
4	Flockinator	12d	192



Le type d'avion qui a le plus volé est l'aéronef Goose avec 1008 vols.

On remarque une certaine asymétrie dans l'utilisation de la flotte.

Le type d'avion Goose occupe quasiment la moitié des vols de l'ensemble des avions concernés (44%).

Désormais, il est de savoir si cette sur-utilisation de ce type d'avion se base sur sa disponibilité, son coût, sa capacité ou encore sa fiabilité.



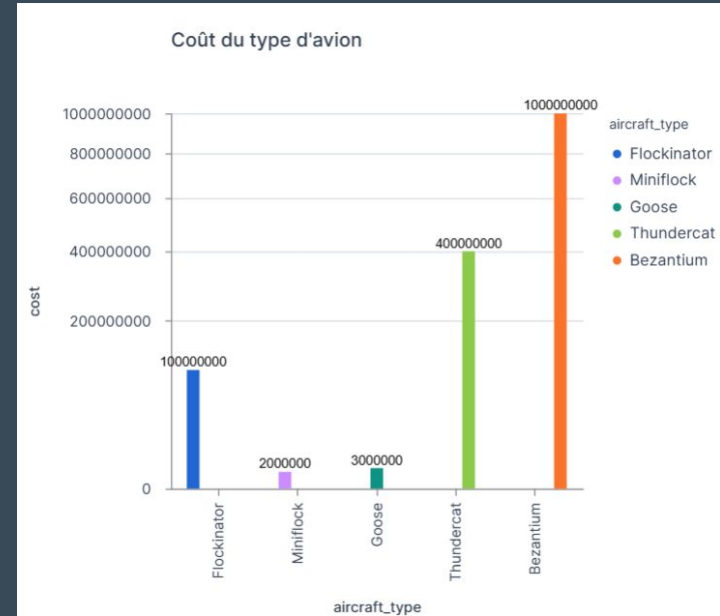


Quel type d'avion a le plus volé ?

Compte tenu du prix des différents types d'avions, on peut remarquer qu'il y a une stratégie de rationalisation des coûts.

Pour l'avion Goose, le ratio vols/coût relatif est très important.

On peut considérer que l'avion Goose est le pilier opérationnel de la flotte.



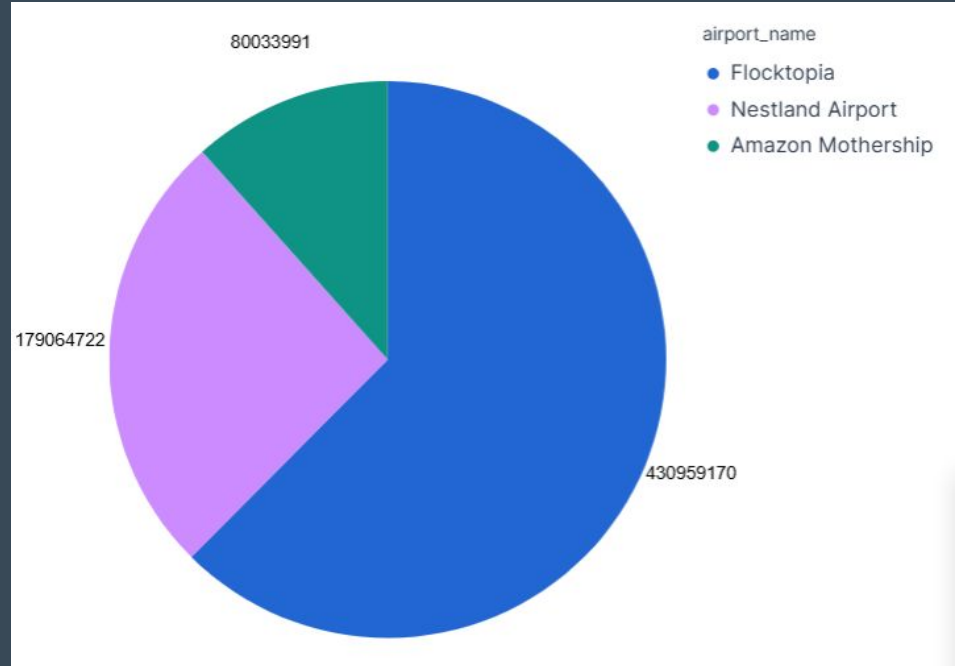


Quel aéroport a reçu le plus de passagers ?

On distingue une concentration majeure du trafic aérien sur l'aéroport de Flocktopia.

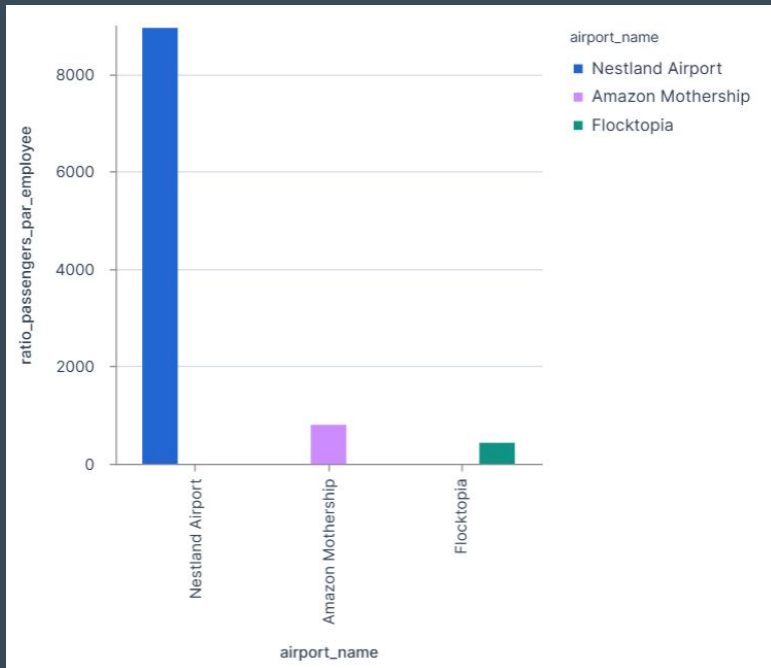
Les opérations de gestion des passagers doivent donc être essentiellement mises en oeuvre sur cet aéroport.

D'autre part, on pourra, par la suite, se poser des questions sur les raisons de l'attirance limitée du trafic sur les aéroports de Nestland et d'Amazon Mothership (capacité limitée, desserte limitée ?...)





Quel aéroport a reçu le plus de passagers ?



Cependant, on remarque que le Nestland Airport a un meilleur ratio passagers/employés.

Le ratio met en évidence un surinvestissement massif en personnel à Flocktopia ou alors ce champ ne comptabilise pas seulement les employés de l'aéroport, mais incluent l'ensemble des partenaires (sécurité, douanes, compagnies aériennes, etc...).



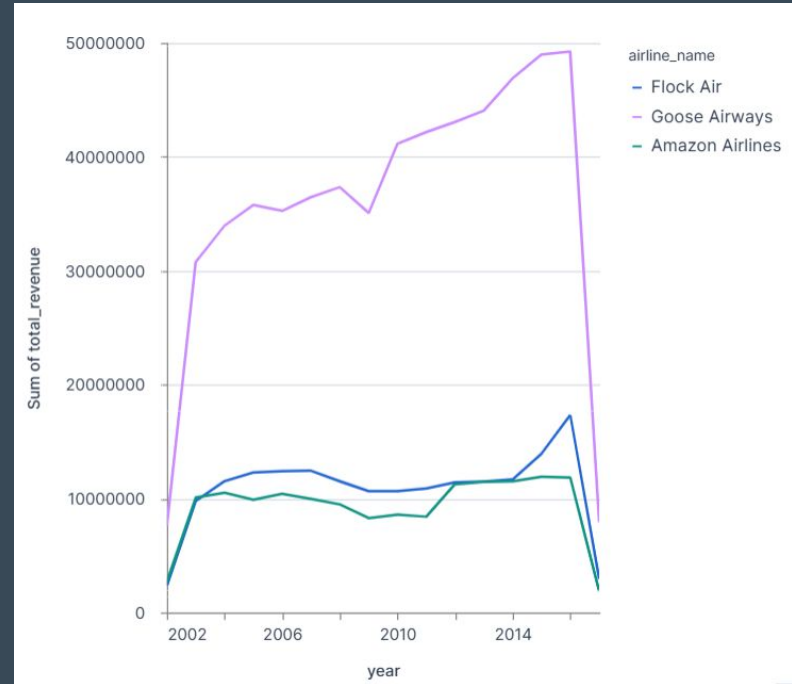
Quelle est la meilleure année pour le RPM pour chaque compagnie aérienne ?



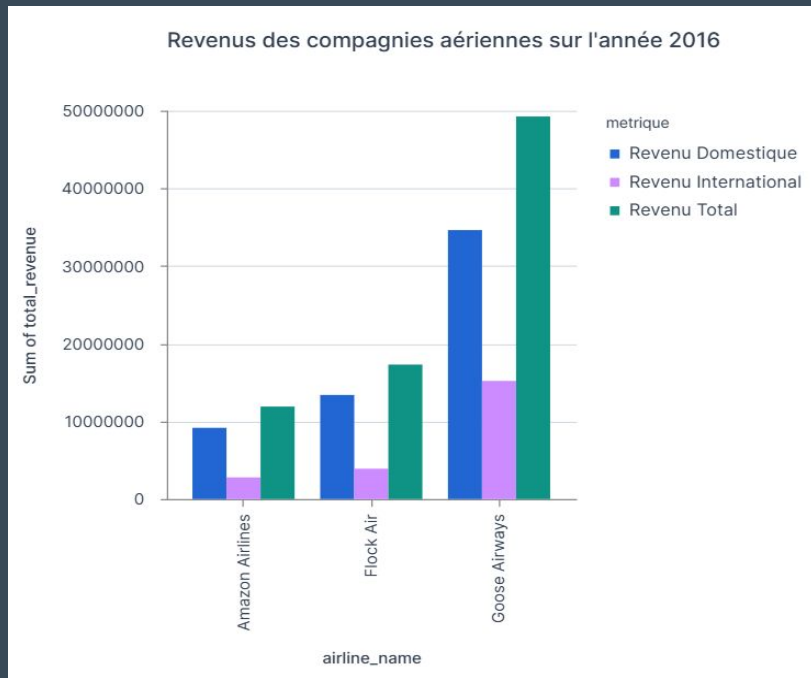
On peut remarquer une certaine croissance dans le montant des revenus par année pour la compagnie Goose Airways.

Du côté, des compagnies Flock Air et Amazon Airlines, on observe une certaine stagnation des revenus sur la période de 2003 à 2016. On observe un léger pic en 2016 pour la compagnie Flock Air.

De ce fait, on remarque les pics pour l'ensemble pour l'ensemble des compagnies se trouve sur l'année 2016.



Quelle est la meilleure année pour le RPM pour chaque compagnie aérienne ?



On peut remarquer que la Compagnie Goose Airways est leader incontesté sur la performance financière sur l'année 2016 avec des revenus frisant les 50M de dollars.

Les trois compagnies s'appuient essentiellement sur leurs revenus domestiques.

Doivent-ils s'attaquer au marché international pour ne pas trop dépendre du marché domestique ?



Quelle est la meilleure année de croissance pour chaque compagnie aérienne ?



airline_code	object	airline_name	obj...	best_year	int64	best_taux_croiss...
AA		Amazon Airlines		2012		46.02
FA		Flock Air		2016		16.04
GA		Goose Airways		2010		20.67



Le plus grand taux de croissance est pour la compagnie aérienne Amazon Airlines qui, pour l'année 2012, a eu un taux de croissance de 46.02%. Cependant, la compagnie connaît les chutes les plus sévères.

Néanmoins, Flock Air a connu des taux négatifs initialement. Néanmoins, elle est la seule compagnie à connaître un taux de croissance positif élevé (15%) en 2016.

La compagnie Goose Airways a plutôt un taux de croissance stagnant avec un pic en 2010 à 20%.



Conclusion



Les compagnies aériennes s'appuient principalement sur une flotte d'avions avec une capacité moyenne mais surtout avec une stratégie de réduction des coûts.

2 aéroports sur 3 ont un ratio intéressant dans la gestion des employés et la taille de leur structure.

Concernant leurs revenus, les compagnies aériennes trouvent essentiellement leurs revenus des vols domestiques (la conjoncture internationale et le contexte géopolitique peuvent parfois impacter le trafic).

Enfin, ces compagnies aériennes ont des fortunes diverses concernant leur taux de croissance. Elles contiennent généralement leur taux entre -15 et 20% sur la période de 2002 à 2016.

