



Projet Aircraft

Analyse exploratoire, descriptive et inférentielle des données

Bloc 2 - RNCP 35288 CDSD

Stéphanie BRETTON



L'objectif de ce projet :

Utiliser Deepnote pour effectuer une EDA sur les données Aircraft chargées dans Snowflake.

- Création d'un entrepôt dans Snowflake avec le chargement des données aircraft_db.
- Utilisation de l'outil open source DBT (Data Build Tool), pour effectuer l'étape de Transformation "T" dans le modèle ETL.
- Création d'un nouveau projet dans deepnote pour récupérer les données et préparer la visualisation. Requêtes avec SQL.
- Visualisations réalisées également sur Visual Code Studio avec python en récupérant les données "Raw" dans Snowflake.

1- Quel avion a effectué le plus de vols ?

aircraft_analytics ▾ Variable: DataFrame ▾ df_1 Hide schema

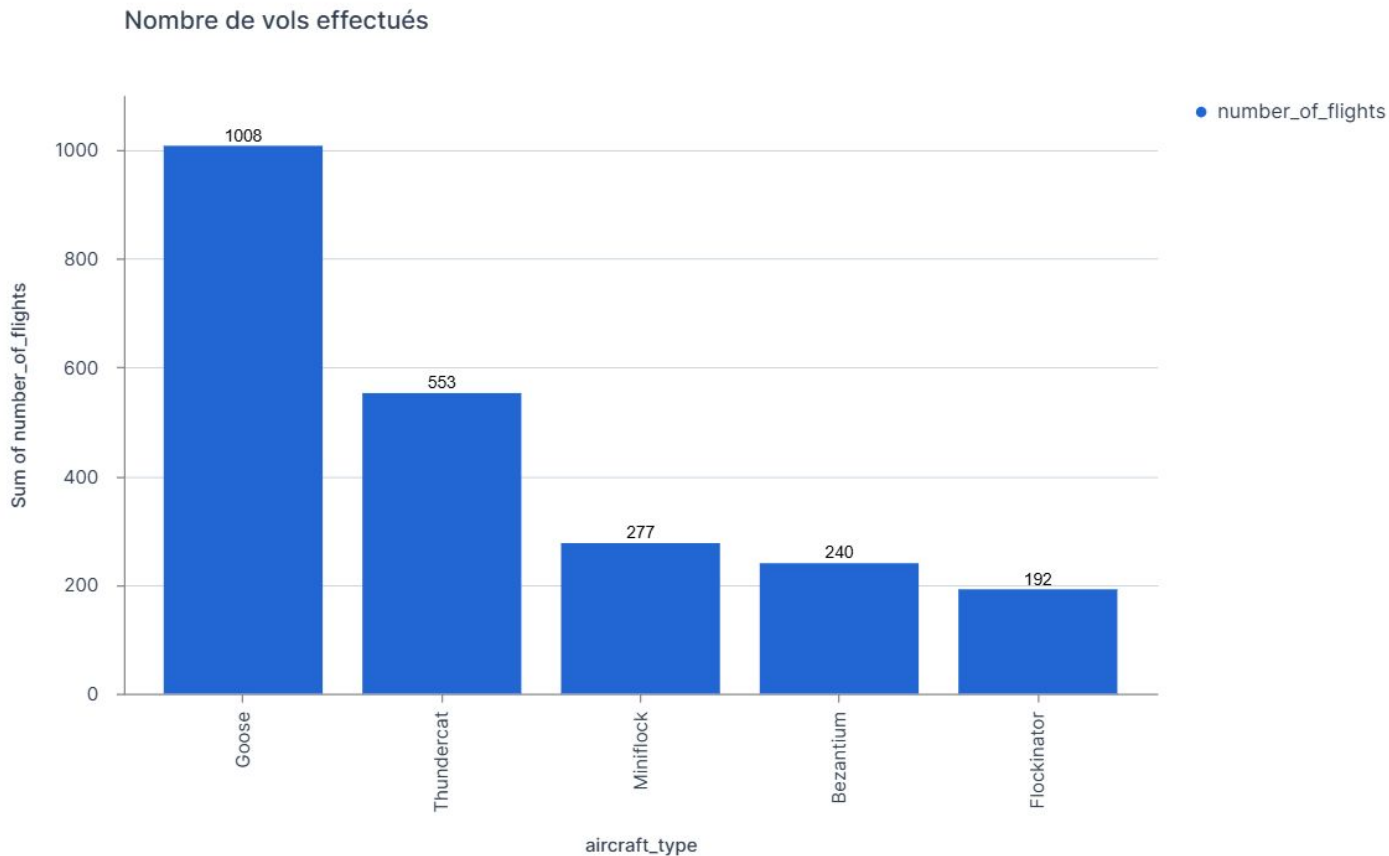
```
1 -- QUEL AVION A EFFECTUE LE PLUS DE VOLS ? --  
2 select * from ANALYTICS.DBT_SBRETTON.FCT_NUMBER_OF_FLIGHTS
```

Visualize

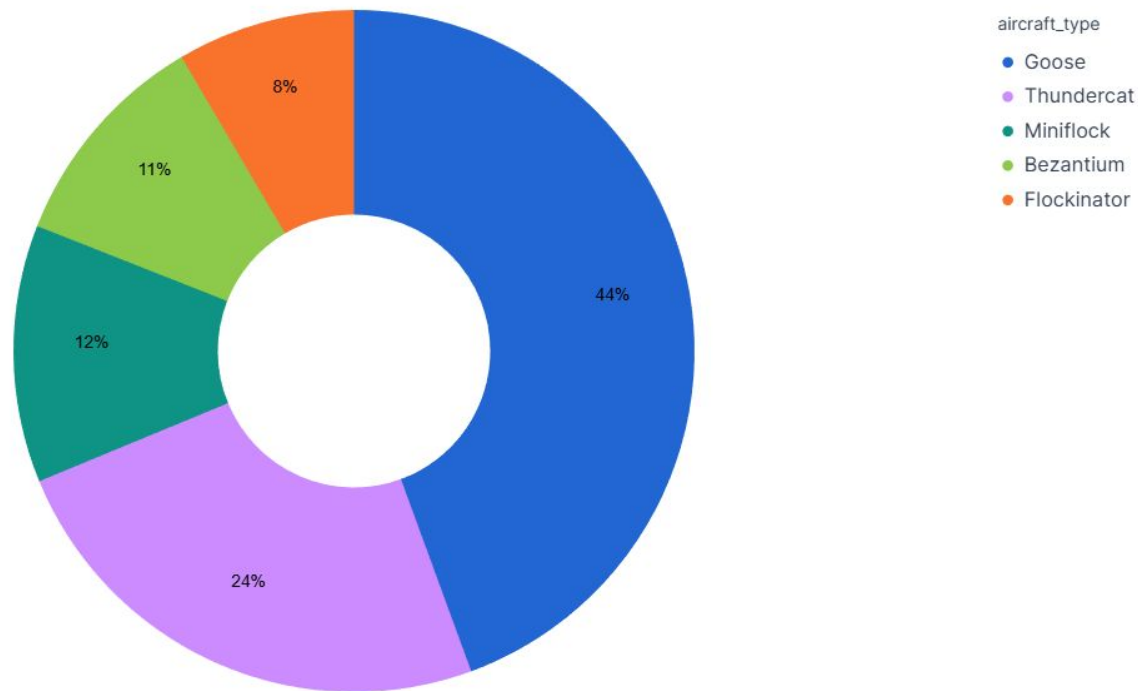
	aircraft_type obj...	aircraft_id object	number_of_flights i	
0	Goose	g72	1008	
1	Thundercat	t10	553	
2	Miniflock	12a	277	
3	Bezantium	b23	240	
4	Flockinator	12d	192	

5 rows, 3 cols 10 ▾ / page << < Page 1 of 1 > >> Format 📄 ⬇

L'appareil
GOOSE a
effectué le plus
grand nombre
de vols (1008)
et la flotte
repose
majoritairement
sur ce modèle.



Répartition en pourcentage des vols effectués



Avec une part de 44%, l'appareil GOOSE domine le marché très nettement.



2- Quel aéroport a accueilli le plus grand nombre de passagers ?

aircraft_analytics ▾ Variable: DataFrame ▾ df_1 Hide schema

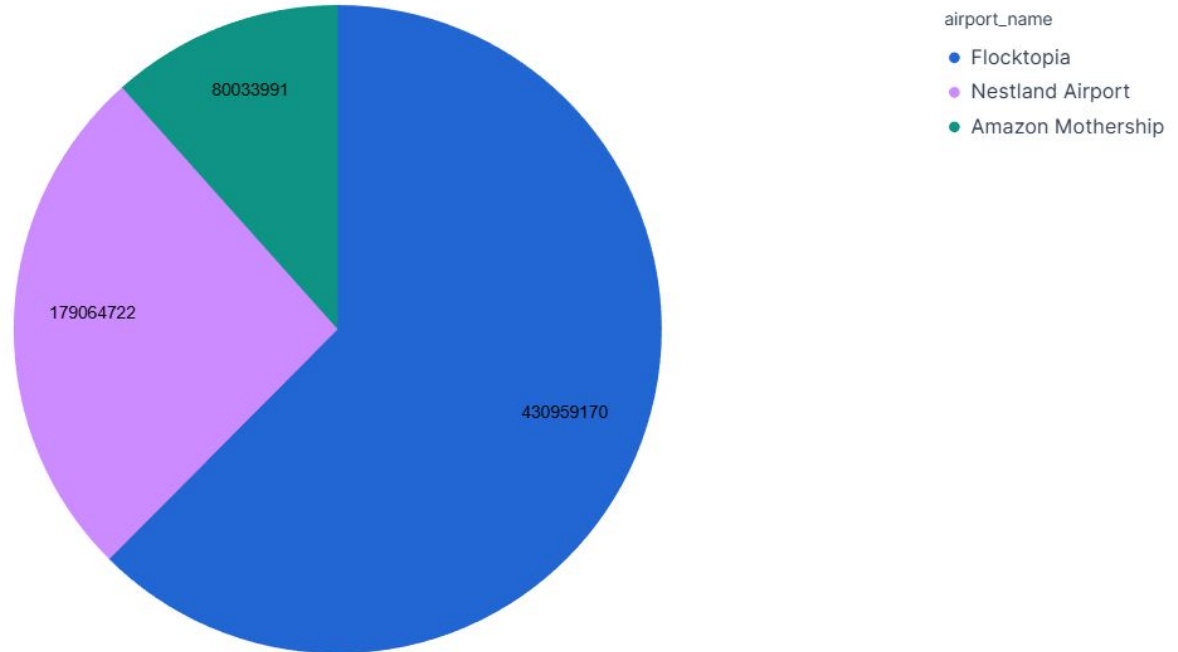
```
1 -- QUEL AEROPORT A TRANSPORTE LE PLUS DE PASSAGERS ?  
2 SELECT * FROM ANALYTICS.DBT_SBRETTON.FCT_AIRPORTS
```

Visualize

	airport_code obj...	airport_name obj...	passengers_nb fl...	
0	FKT	Flocktopia	430959170	
1	NSA	Nestland Airport	179064722	
2	AMP	Amazon Mothers...	80033991	

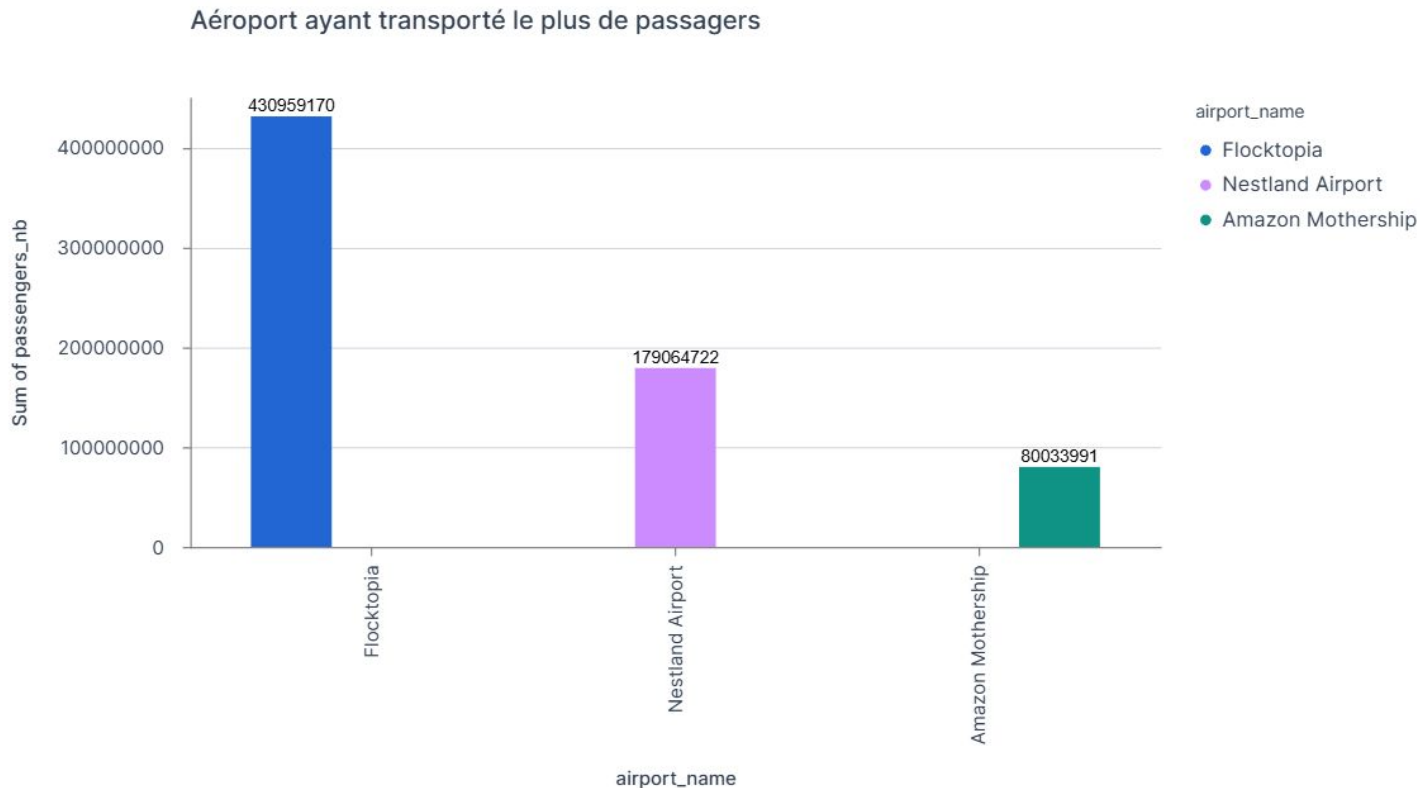
3 rows, 3 cols 10 ▾ / page << < Page 1 of 1 > >> Format 📄 ⬇️

Quel aéroport a transporté le plus de passagers ?



Flocktopia est l'aéroport ayant accueilli le plus de passagers.

A lui seul,
l'aéroport
Flocktopia
génère un
trafic
nettement
supérieur aux
2 autres
réunis.



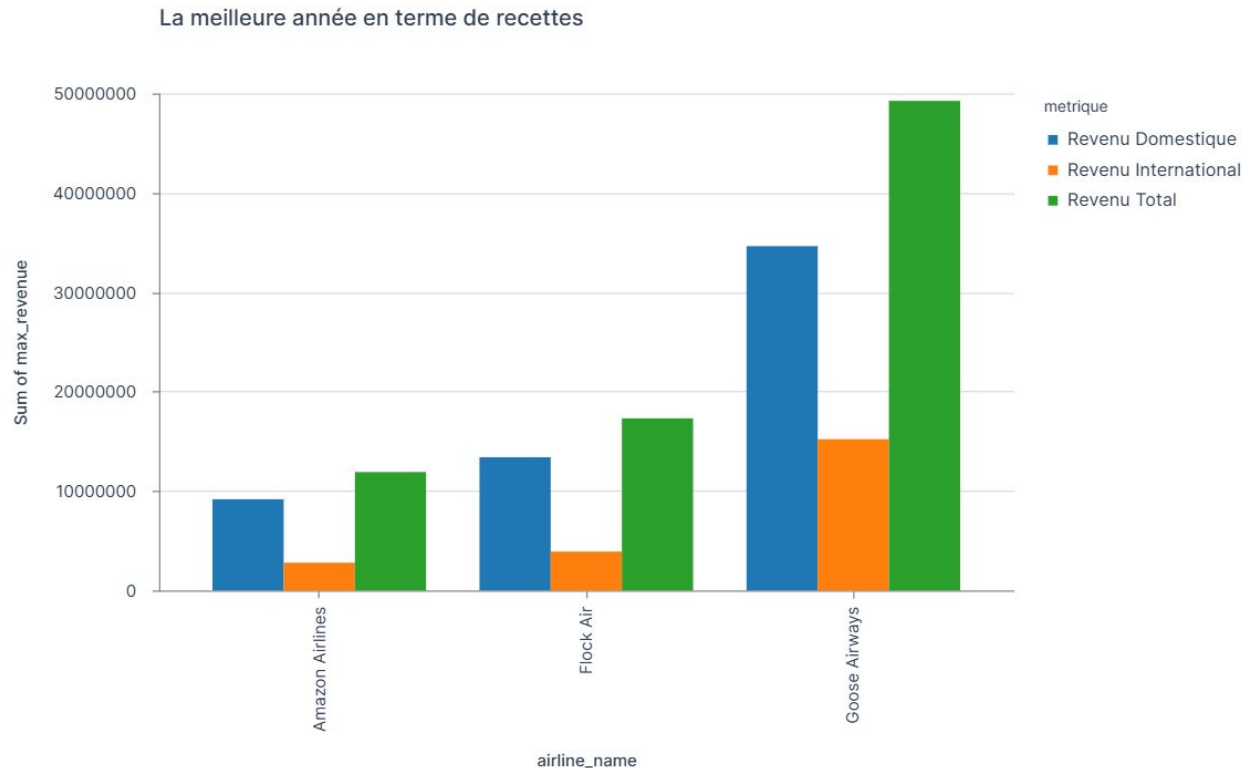
3- Quelle a été la meilleure année en terme de recettes par passager-kilomètre pour chaque compagnie aérienne ?



Le maximum de recettes est généré en 2015 pour Amazon Airlines, 2016 pour Flock Air et Gosse Airways.

Cette dernière génère également le plus de revenus.

Pour les 3 compagnies, les revenus des vols domestiques représentent la part la plus importante du revenu total.



4- Quelle a été la meilleure année en terme de croissance pour chaque compagnie aérienne ?

aircraft_analytics Variable: DataFrame df_3 Hide schema

1 --- QUELLE A ETE LA MEILLEURE ANNEE EN TERME DE CROISSANCE PAR COMPAGNIE AERIENNE ?
2 SELECT * FROM ANALYTICS.DBT_SBRETTON.FCT_CROISSANCE_PER_YEAR

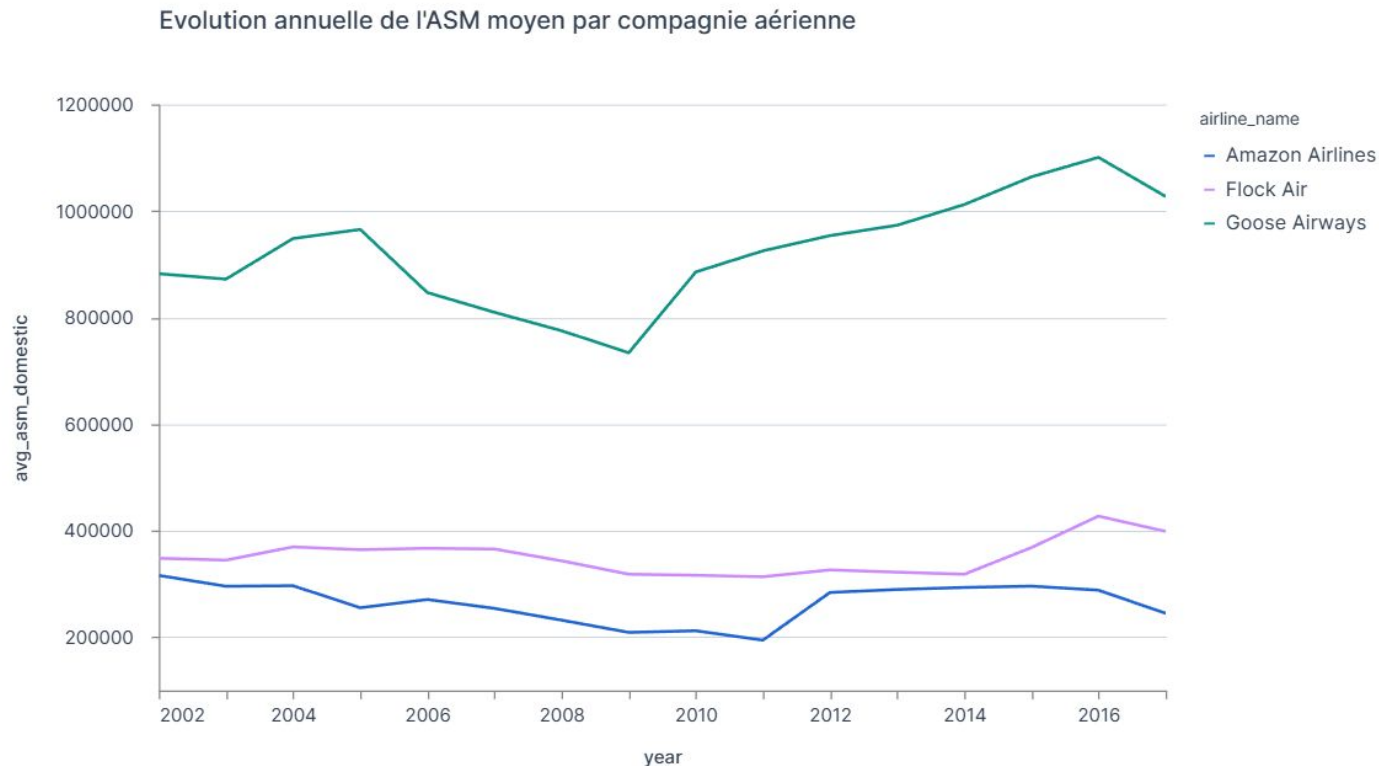
Visualize

	airline_code object	airline_name obj...	year int64	avg_asm_domestic	previous_year flo...	taux_croissance f...
	AA 33.3%	Amazon ... 33.3%	2002 - 2017	194470.666667 - ...	194470.666667 - ...	-15.1 - 46.02
	FA 33.3%	Flock Air 33.3%				
	GA 33.3%	Goose Air... 33.3%				
0	AA	Amazon Airlines	2002	315931	nan	nan
1	AA	Amazon Airlines	2003	295645.5278	315931	-6.42
2	AA	Amazon Airlines	2004	296641.3889	295645.5278	0.34
3	AA	Amazon Airlines	2005	255289.0833	296641.3889	-13.94
4	AA	Amazon Airlines	2006	270788.3611	255289.0833	6.07
5	AA	Amazon Airlines	2007	254119.0556	270788.3611	-6.16
6	AA	Amazon Airlines	2008	232021.2778	254119.0556	-8.7
7	AA	Amazon Airlines	2009	209025.5556	232021.2778	-9.91
8	AA	Amazon Airlines	2010	212045.7222	209025.5556	1.44
9	AA	Amazon Airlines	2011	194470.6667	212045.7222	-8.29

48 rows, 6 cols 10 / page << < Page 1 of 5 > >> Format

L'ASM moyen (Available Seat Miles) est largement dominé en terme de capacité par Goose Airways. Cette dernière à une capacité en en hausse.

Les deux autres compagnies ont une capacité plutôt stable au fil des années.

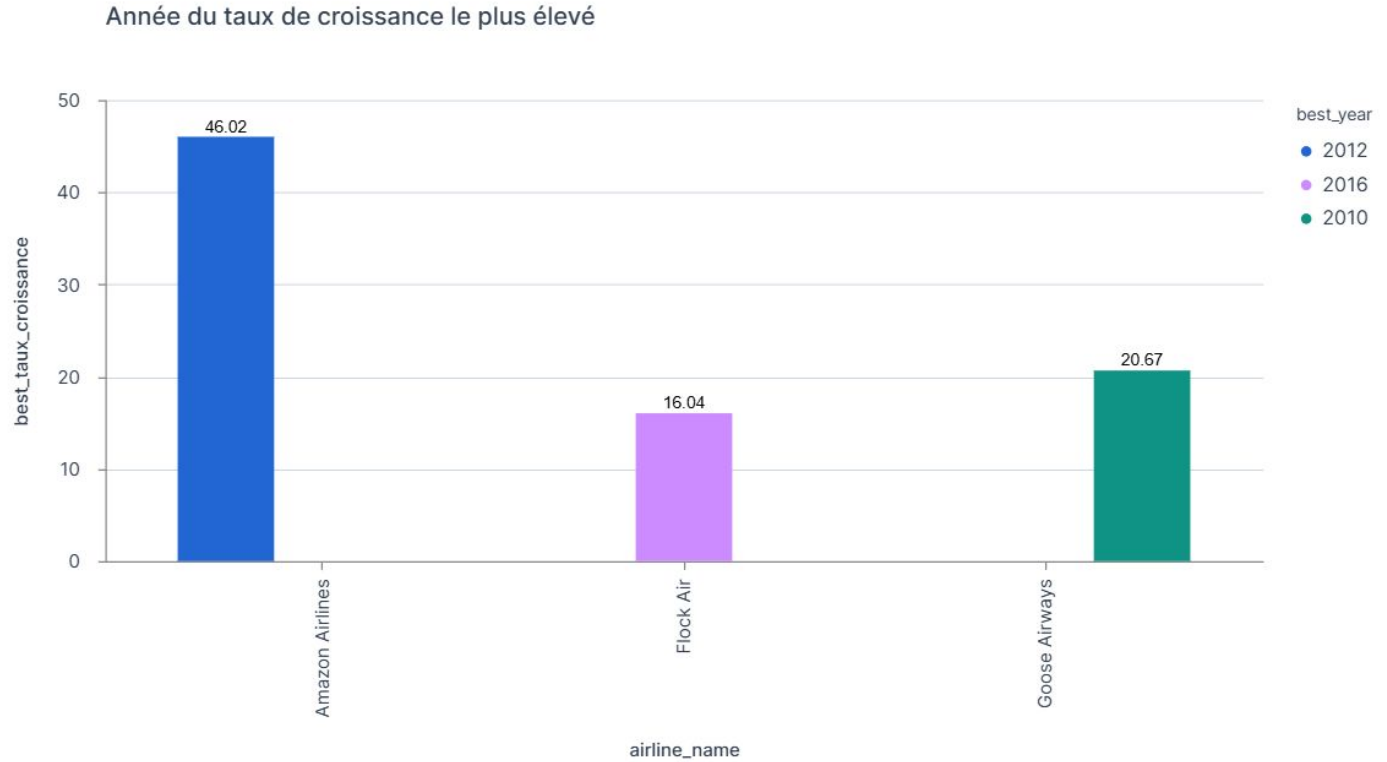


Meilleur taux de croissance :

-AA en 2012

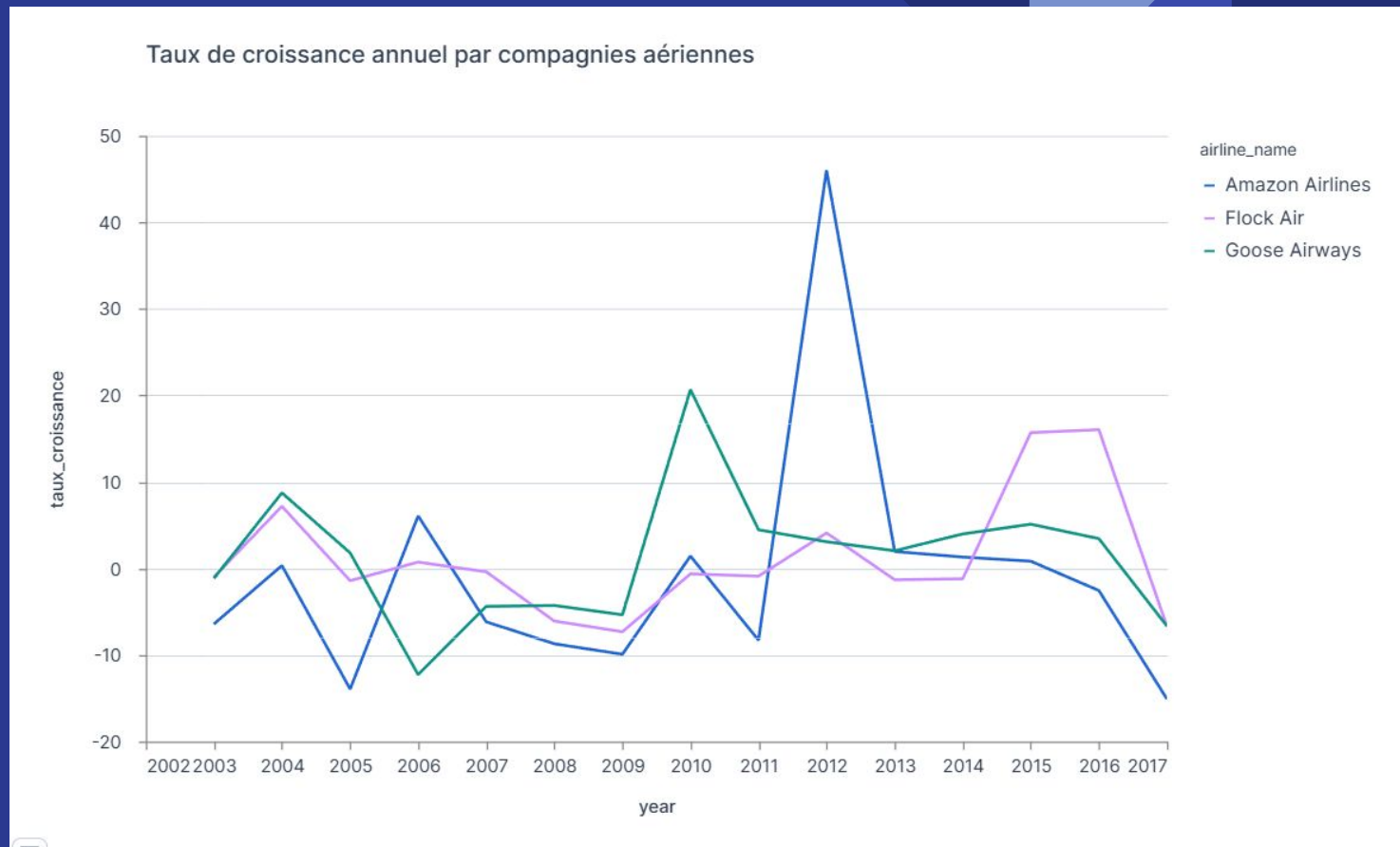
-FA en 2016

-GA en 2010



Le taux de croissance est faible et instable de 2002 à 2009 avec des taux de croissance souvent négatifs.

A partir de 2010 on note des progressions spectaculaires mais le taux de croissance reste instable pour les 3 compagnies.



Conclusion :

- Les données sont issues d'un jeu de données à but pédagogique.
- L'estimation des passagers est basée sur la capacité théorique des avions.
- L'analyse est principalement descriptive, sans modélisation prédictive.

Perspectives d'amélioration :

- Ajout d'indicateurs de remplissage (RPM / ASM)
- Analyse comparative des trafics (Domestic Vs International)
- Extension vers des analyses prédictives