



---

# Rapport de projet BI

---

Pierre BOULET, Dorian FUVEL, Théo HOPSORE, Arnaud  
BROSSAY

12 mars 2024

**AIRFRANCE** 

# Table des matières

I	Introduction . . . . .	2
I.i	Remise en contexte . . . . .	2
II	Structure de données . . . . .	2
III	Présentation du processus ETL . . . . .	3
IV	Présentation de l'OLAP . . . . .	4
V	Data Visualisation . . . . .	5
VI	Conclusion . . . . .	7

# I Introduction

La Business Intelligence (BI) est un domaine de l'informatique décisionnelle qui se concentre sur la collecte, l'analyse et la présentation de données pour soutenir la prise de décisions stratégiques au sein des organisations. Grâce à la BI, les entreprises peuvent transformer leurs données brutes en informations exploitables, leur permettant de comprendre les tendances, d'identifier les opportunités et de résoudre les problèmes. En utilisant des outils et des techniques spécialisés, la BI aide les entreprises à améliorer leur efficacité opérationnelle, à optimiser leurs processus métier et à rester compétitives sur le marché.

Dans le cadre de notre projet de BI, nous incarnons une équipe de data travaillant chez Air France, et notre objectif principal est de comprendre et d'analyser les données de vol des compagnies aériennes américaines. Notre but est d'utiliser ces informations pour étudier le marché de la concurrence et ainsi, améliorer l'implémentation stratégique d'Air France sur le long terme.

## I.i Remise en contexte

Air France, en tant que leader de l'industrie aéronautique, cherche continuellement à améliorer son positionnement stratégique sur le marché mondial. Dans cet objectif, l'entreprise est confrontée à un défi majeur : comprendre et optimiser sa position concurrentielle sur le marché américain, l'un des marchés les plus compétitifs et dynamiques du secteur.

Le problème auquel Air France est confrontée réside dans la nécessité d'analyser en profondeur les données de vol des compagnies aériennes américaines, y compris les statistiques de vol, les performances opérationnelles et les facteurs environnementaux tels que les conditions météorologiques. Ces données sont essentielles pour obtenir des insights précieux sur le marché américain et pour élaborer des stratégies compétitives qui permettront à Air France de se démarquer de la concurrence.

Le défi est double : premièrement, il est crucial de collecter, nettoyer et structurer efficacement les données provenant de différentes sources afin de garantir leur qualité et leur intégrité. Deuxièmement, une fois les données préparées, il est nécessaire de mener une analyse approfondie pour identifier les tendances, les opportunités et les menaces sur le marché américain, ainsi que pour évaluer la performance d'Air France par rapport à ses concurrents.

# II Structure de données

La base de données conçue pour l'analyse des données de vol des compagnies américaines est structurée de manière à capturer les différentes dimensions et mesures essentielles pour une analyse approfondie. Voici comment la base de données est représentée :

- **Table Fait\_de\_Vol** : Cette table contient les données de base sur les vols, telles que le nombre de vols, le nombre de vols annulés et la durée moyenne de retard. Les clés primaires sont définies pour identifier de manière unique chaque vol en fonction de la date, de la compagnie aérienne, de l'aéroport de départ et de l'aéroport d'arrivée.
- **Tables de Dimensions** : Plusieurs tables de dimensions sont utilisées pour représenter les différentes dimensions d'analyse. Cela inclut la dimension de la date (par jour, semaine, mois, etc.), la dimension de la compagnie aérienne, la dimension de l'aéroport (avec des informations sur la ville et la météo), la dimension de la météo, la dimension de la semaine, la dimension du mois, la dimension du trimestre, la dimension de l'année, la dimension de la ville et la dimension du pays. Ces tables sont liées à la table de faits via des clés étrangères pour permettre une analyse multidimensionnelle.
- **Relations entre les Tables** : Des contraintes de clés étrangères sont établies pour assurer l'intégrité référentielle entre les tables. Par exemple, les tables de dimensions sont liées à la table de faits par le biais de clés étrangères qui correspondent aux clés primaires dans la table de faits. De plus, des relations entre les dimensions sont établies là où cela est pertinent, comme entre la dimension de la météo et la dimension de l'aéroport.



FIGURE 1 – Représentation graphique de la structure de données

## III Présentation du processus ETL

Le processus ETL (Extract, Transform, Load) joue un rôle crucial dans la préparation des données pour une analyse efficace dans Tableau. Voici comment ce processus est appliqué dans notre contexte, où nous analysons les données de vol des compagnies américaines pour Air France :

- **Extraction (Extract)** : Dans cette phase, les données sont extraites de différentes sources telles que les systèmes de gestion des compagnies aériennes et les fournisseurs de données météorologiques. Nous utilisons des outils comme Apache NiFi pour extraire les données des sources variées, y compris des bases de données et des

API telles que The Weather Company API pour obtenir les données météorologiques.

- **Transformation (Transform)** : Une fois les données extraites, elles sont transformées pour les préparer à l'analyse. Cela inclut des activités telles que le nettoyage des données, la normalisation, l'enrichissement avec des données supplémentaires (comme les données météorologiques), et l'agrégation des données pour créer des mesures pertinentes. Nous utilisons des outils comme Python avec pandas pour effectuer des transformations complexes, ainsi que PySpark pour le traitement des données distribuées.
- **Chargement (Load)** : Dans cette phase, les données transformées sont chargées dans la base de données cible, qui est optimisée pour l'analyse et la visualisation dans Tableau. Nous utilisons des outils comme Apache Sqoop pour transférer les données vers Amazon Redshift, une base de données cloud conçue pour l'analyse de données à grande échelle. Les données sont organisées dans une structure appropriée pour une récupération rapide et efficace lors de l'analyse.

Une fois que les données sont chargées dans Amazon Redshift, elles peuvent être facilement connectées à Tableau pour la visualisation et l'analyse. Tableau offre une interface conviviale pour créer des tableaux de bord interactifs, des graphiques et des rapports qui permettent aux utilisateurs de explorer les données de vol de manière intuitive et de découvrir des insights significatifs.

## IV Présentation de l'OLAP

Dans notre approche d'analyse des compagnies aériennes, nous utilisons un cube OLAP, ou Online Analytical Processing, comme structure multidimensionnelle pour explorer et analyser nos données de manière interactive. Notre cube OLAP est configuré avec plusieurs dimensions clés, notamment les compagnies aériennes, les États des États-Unis, ainsi que les mesures de retard et d'annulation de vols.

- **Dimension Compagnie Aérienne** : Nous utilisons cette dimension pour analyser nos données en fonction des différentes compagnies aériennes. Cela nous permet de sélectionner une compagnie aérienne spécifique et d'examiner ses performances, telles que le nombre de vols effectués, les retards, les annulations, etc. Cette analyse comparative nous aide à identifier les leaders du marché et à comprendre les différences de performances entre les compagnies.
- **Dimension États des États-Unis** : Nous utilisons cette dimension pour représenter les États des États-Unis en tant qu'axe d'analyse. Cela nous permet d'explorer les données de vol en fonction de l'emplacement géographique, ce qui nous aide à identifier les tendances régionales, les principaux aéroports, les volumes de trafic, etc. Cette analyse géographique nous aide à mieux comprendre les variations dans les performances des compagnies aériennes à travers le pays.
- **Mesures de Retard et d'Annulation** : Bien que les raisons spécifiques de retard ou d'annulation ne soient pas incluses comme dimensions distinctes, nous captu-

rons ces informations en tant que mesures dans notre cube OLAP. Nous pouvons ainsi analyser les données de retard et d'annulation en fonction des compagnies aériennes et des États, ce qui nous permet de comprendre les motifs de retard et d'annulation les plus courants et d'identifier les domaines d'amélioration potentiels.

Grâce à notre cube OLAP configuré avec ces dimensions clés, nous sommes en mesure d'explorer nos données de vol de manière interactive. Cela nous permet de prendre des décisions stratégiques éclairées pour optimiser nos opérations et améliorer notre positionnement sur le marché concurrentiel des compagnies aériennes.

## V Data Visualisation

Nous avons créé une visualisation des données à l'aide de Tableau. Cette visualisation nous permet d'explorer et de présenter nos données de manière intuitive et interactive. En utilisant les fonctionnalités de Tableau, nous avons pu créer des tableaux de bord dynamiques, des graphiques et des cartes qui fournissent des insights précieux sur nos données de vol des compagnies aériennes. Cette approche nous permet de prendre des décisions stratégiques informées et de communiquer efficacement nos résultats aux parties prenantes internes et externes.

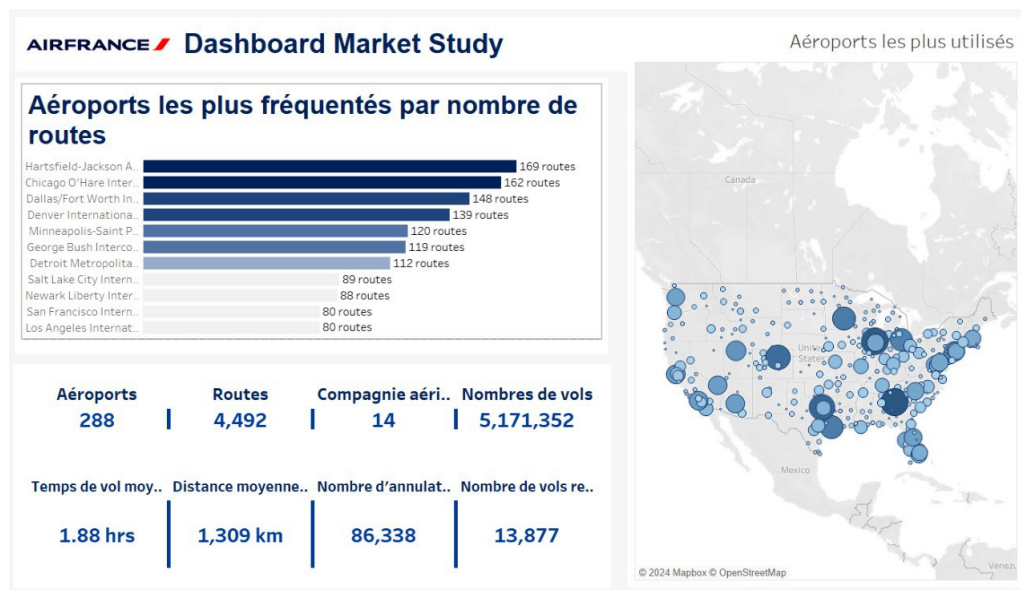


FIGURE 2 – Premier tableau de visualisation

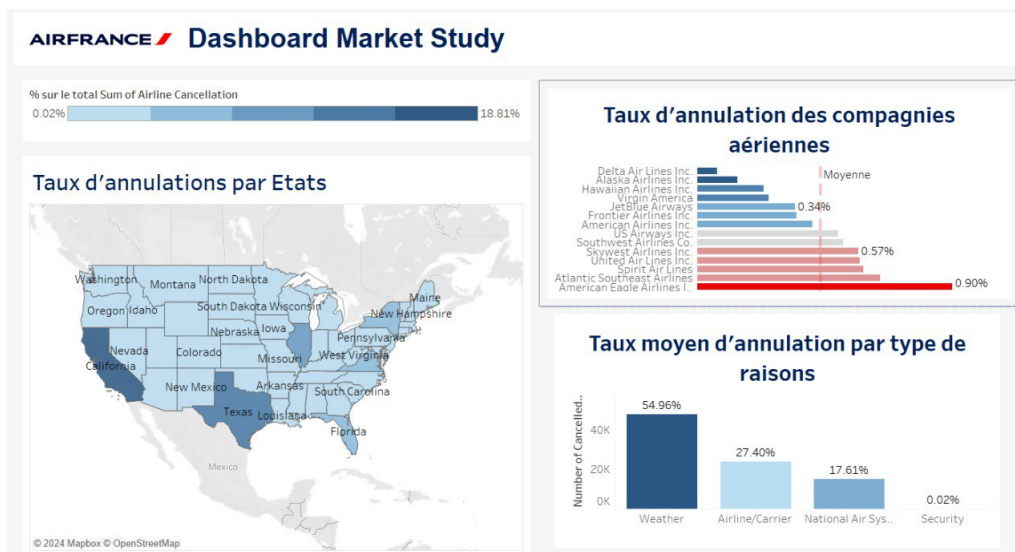


FIGURE 3 – Second tableau de visualisation

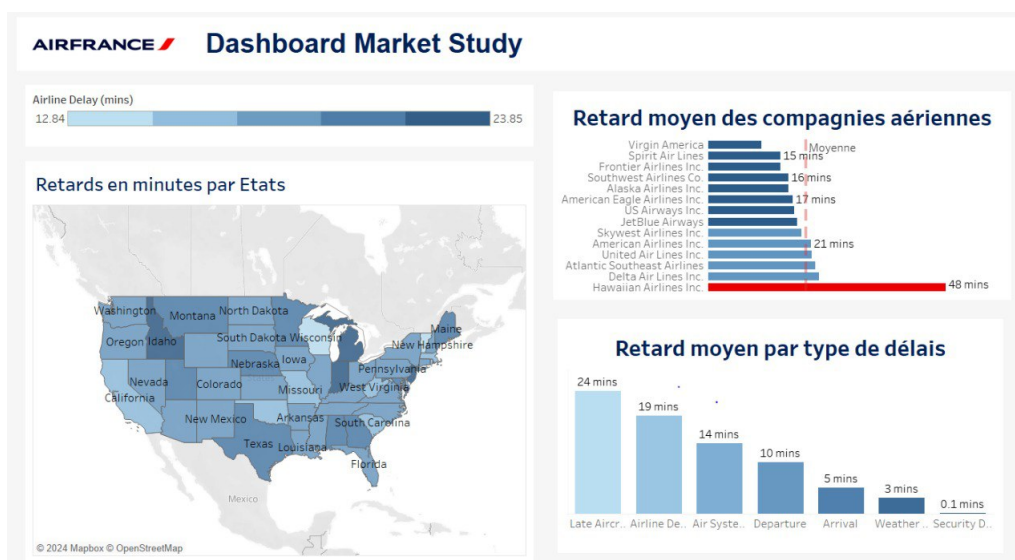


FIGURE 4 – Troisième tableau de visualisation

Dans notre analyse des données de vol des compagnies aériennes, nous avons identifié plusieurs indicateurs clés de performance (KPI) essentiels pour évaluer les performances opérationnelles et la qualité du service. Parmi ces KPI, nous incluons :

- **Aéroports les plus Fréquentés par Nombre de Routes** : Cet indicateur nous permet d'identifier les aéroports les plus fréquentés en fonction du nombre de routes desservies, ce qui nous donne un aperçu des principaux hubs de transport aérien.
- **Taux d'Annulation par États** : Nous surveillons le taux d'annulation des vols dans chaque État pour évaluer les variations régionales dans la fiabilité des services aériens.

- **Taux d'Annulation des Compagnies Aériennes** : Cet indicateur nous permet de mesurer la performance des différentes compagnies aériennes en termes de fiabilité de leurs services, en examinant le pourcentage de vols annulés par compagnie.
- **Taux d'Annulation Moyen par Type de Raisons** : Nous analysons les motifs d'annulation des vols et calculons le taux moyen d'annulation pour chaque type de raison, nous permettant ainsi de cibler les problèmes les plus fréquents et d'identifier les domaines d'amélioration potentiels.
- **Retards en Minutes par États** : Nous évaluons les retards des vols dans chaque État en mesurant le nombre total de minutes de retard accumulées, ce qui nous donne un aperçu des zones où les retards sont les plus fréquents.
- **Retard Moyen des Compagnies Aériennes** : Nous examinons le retard moyen des vols pour chaque compagnie aérienne afin de déterminer leur performance en termes de ponctualité.
- **Retard Moyen par Type de Délais** : En analysant les différents types de retards (météo, maintenance, trafic, etc.), nous pouvons identifier les principaux facteurs contribuant aux retards des vols et prendre des mesures pour les atténuer.

En utilisant ces KPI, nous sommes en mesure d'évaluer les performances des compagnies aériennes, d'identifier les tendances et les problèmes potentiels, et de prendre des décisions stratégiques pour améliorer notre efficacité opérationnelle et la satisfaction des passagers.

## VI Conclusion

En résumé, notre mission consiste à analyser les données de vol des compagnies aériennes américaines afin d'optimiser notre positionnement stratégique sur le marché. En utilisant un processus ETL robuste, nous avons extrait, transformé et chargé les données dans une base de données adaptée à l'analyse. À l'aide d'outils tels que Tableau, nous avons créé des visualisations interactives pour explorer nos données et avons identifié des KPI clés pour évaluer les performances opérationnelles. Ces efforts nous permettent de prendre des décisions éclairées et de continuer à améliorer notre compétitivité dans l'industrie aéronautique.