

Rapport Sky Data Lab

Résumé

Ce rapport présente l'ensemble de ma démarche pour traiter et analyser le jeu de données lié aux vols aériens, de la phase de réflexion jusqu'aux KPI.

Dans un premier temps, j'expose ma réflexion pour répondre au sujet puis j'explique les différentes étapes de traitement et de préparation des données réalisées dans Azure Synapse Analytics.

Enfin, je présente la restitution et l'évolution des KPI dans le temps à l'aide de Power BI.

Août 2025

Réflexion et démarche

Pour traiter les données brutes, j'ai choisi de suivre une approche en plusieurs étapes :

1. **Mise en place de l'architecture** : Création du "Storage Account" et d'un "container" dédié pour héberger les données sources et du "Synapse Workspace" pour centraliser le traitement ;
2. **Ingestion** : Chargement des fichiers CSV/JSON dans Azure Synapse pour centraliser les données ;
3. **Phase de découverte des données** : Via des requêtes simples, cela m'a permis d'identifier des anomalies comme des espaces superflus ou des formats incohérents ;
4. **Nettoyage/Transformation SQL** : Création de vues intermédiaires pour chaque fichier source en les nettoyant (suppression d'espaces, correction de formats,...) ;
5. **Construction et export de la table Gold** : Création de la table principale qui sera utilisée pour construire les KPI et test des trois KPI pour m'assurer de la validité de la table ;
6. **Construction du rapport sous Power BI** : Construction des trois KPI.

Traitement des données – Azure Synapse

1. Mise en place de l'architecture

Avant de commencer l'ingestion, j'ai mis en place l'architecture Azure nécessaire au projet. J'ai donc créé les éléments suivants :

- un Storage Account ;
- un container à l'intérieur de ce Storage Account ;
- un Synapse Workspace.

2. Ingestion

Après la mise en place de l'architecture, j'ai procédé au chargement des fichiers sources (CSV et JSON) directement dans le container créé au sein du Storage Account.

3. Phase de découverte des données

J'ai commencé par créer une base de données dédiée dans Azure Synapse, destinée à centraliser toutes les vues issues des fichiers sources. Ensuite, j'ai exploré chaque fichier un par un via des requêtes SQL simples afin de mieux comprendre la structure et le contenu des données.

Cette étape m'a permis d'identifier d'éventuelles anomalies (valeurs manquantes, doublons, incohérences de format) et de préparer le nettoyage nécessaire.

4. Nettoyage/Transformation SQL

Pour chaque fichier source, j'ai créé une vue dédiée dans Azure Synapse afin d'appliquer les opérations de nettoyage nécessaires. Ces traitements ont consisté à supprimer les espaces superflus, harmoniser les formats de texte ou encore corriger certaines incohérences de données.

Cette approche par vues intermédiaires m'a permis de garder une structure claire et modulaire avant de passer à l'étape de consolidation dans la table Gold.

5. Construction de la table Gold

À partir des vues nettoyées, j'ai consolidé l'ensemble des vues sources dans une table Gold, prête pour l'analyse. Cette table agrège les données de vols (avions, compagnies, aéroports, retards), applique les colonnes calculées (indicateurs de départ à l'heure, retard à l'arrivée, vols nationaux) et harmonise les libellés des causes de retard.

Une fois la table construite, j'ai validé sa qualité en la testant en SQL sur les trois KPI (le taux de départ à l'heure, le retard moyen à l'arrivée par type d'avion et le top 3 des causes de retard sur les vols nationaux). En validant la fiabilité de la table Gold, je l'ai par la suite exportée au format CSV pour réaliser la data visualisation.

Restitution des KPI

1. Construction du rapport sous Power BI

J'ai connecté Power BI Desktop directement à la table Gold puis j'ai conçu un rapport en quatre pages :

- **Page d'accueil** : présentation d'un résumé des indicateurs clés.
- **Trois pages dédiées** : une par KPI, chacune intégrant les visualisations les plus adaptées et permettant une analyse globale ainsi qu'une évolution dans le temps.

Page d'accueil du rapport



Sky Data Lab
Analyse des problèmes de ponctualité

Taux de vols au départ à l'heure – Annuel (%)

1

Retard moyen à l'arrivée (min) par type avion

2

Top 3 causes de retard – vols nationaux

3



Source : Base vols | Période : 2023-2024

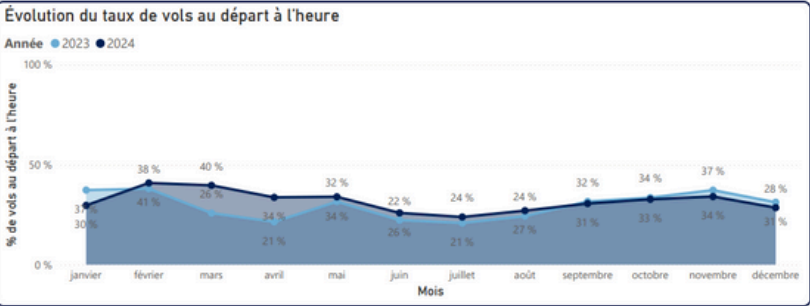
KPI 1 : Taux de vols au départ à l'heure – Annuel (%)



Sky Data Lab
Analyse des problèmes de ponctualité

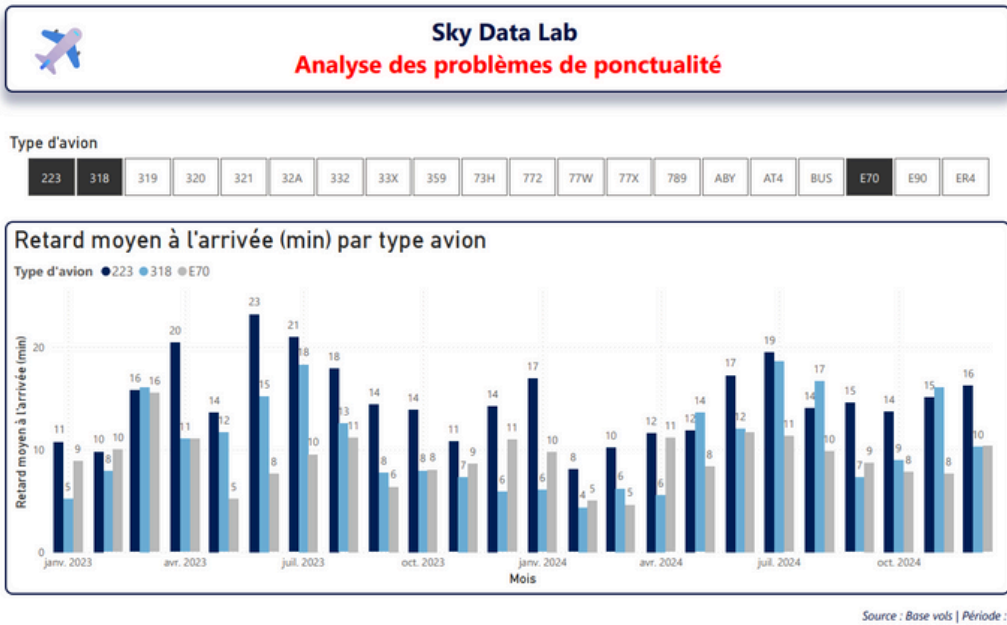
Taux de vols au départ à l'heure – Annuel (%)

2023	29,4 %
2024	31,4 %

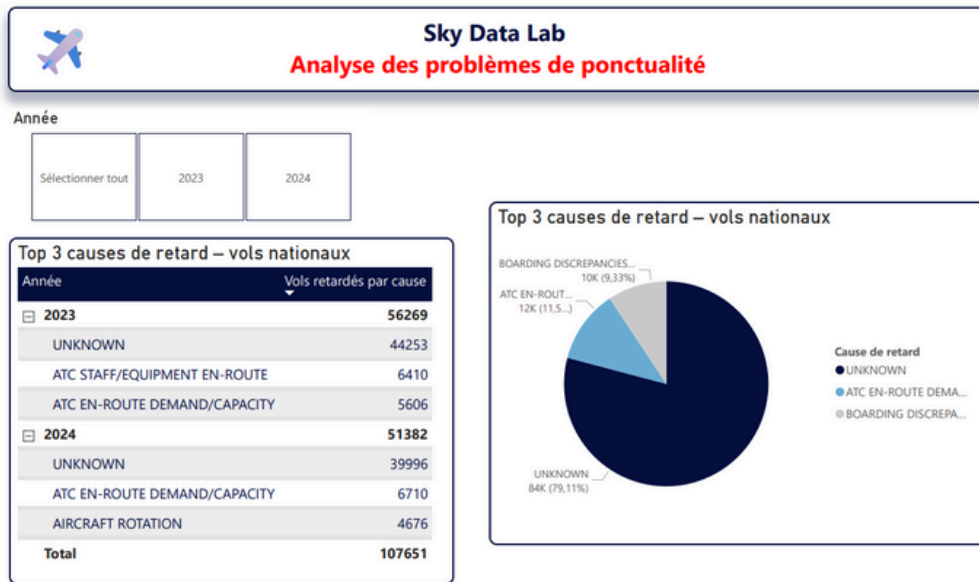


Source : Base vols | Période : 2023-2024

KPI 2 : Retard moyen à l'arrivée (min) par type avion



KPI 3 : Top 3 causes de retard – vols nationaux



2. Conclusion

L'analyse menée m'a permis de dégager plusieurs enseignements clés sur la ponctualité et les causes de retard.

En moyenne, seulement 30 % des vols partent à l'heure, ce qui met en évidence un important potentiel d'amélioration opérationnelle.

Les retards constatés à l'arrivée restent notables, avec une moyenne autour de 15 minutes, impactant la fluidité des opérations.

Enfin, l'étude des causes montre que la première position du Top 3 est occupée par la catégorie "UNKNOWN", ce qui traduit un manque de précision dans le suivi des motifs de retard. Ce point pourrait constituer une priorité pour améliorer la qualité des données et orienter plus efficacement les actions correctives.