**Test technique Data Engineer AF**

Ce test rapide est à faire à la maison en environ 4H (temps à titre indicatif).

**Problématique**

Les métiers en charge de la performance opérationnelle d'Air France vous contactent car le COMEX leur demande d'investiguer les problèmes de ponctualité que nous avons.

Avec les fichiers que vous avez à votre disposition, vérifiez dans un premier temps les craintes du COMEX à l'aide des KPI clefs ci-dessous :

* % de vol réalisés avec un départ à l'heure sur l'année : D0
* Retard moyen à l'arrivée par type avion
* Le top 3 des causes retard sur les vols nationaux

Au-delà de ces premiers KPI, vous devrez approfondir l'analyse des données pour proposer un rapport visuel qui présente ces KPIs avec leur évolution dans le temps.

Techniquement, le problème devra être résolu avec une approche SQL. Si vous n’avez pas accès à un SGBD, nous vous suggérons d’utiliser DuckDB. La table principale, post-traitement et pré-dataviz devra être partagée au format CSV. Pour la partie visualisation, vous êtes libre du choix de l'outil.

Si vous avez des questionnements, n'hésitez pas à vous appuyer sur vos propres hypothèses, qui seront l'occasion de discussions lors de l'entretien.

**Fichiers mis à votre disposition :**

* Table principale : FLIGHTS\_2023.csv et FLIGHTS\_2024.csv :

Chaque enregistrement représente un vol. Les vols qui n’ont pas de réalisé (\*) sont considérés comme annulés.

| **Column name** | **Column title** | **Primary key** |
| --- | --- | --- |
| FLIGHT\_CARRIER\_CODE | Airline code | X |
| FLIGHT\_NUMBER | Flight number | X |
| LEG\_SCH\_DEP\_AIRPORT | Scheduled departure airport | X |
| LEG\_SCH\_DEP\_DATE\_UTC | Scheduled departure date | X |
| LEG\_SCH\_DEP\_TIME\_UTC | Scheduled departure time |  |
| LEG\_ACT\_DEP\_DATE\_UTC | Actual departure date (\*) |  |
| LEG\_ACT\_DEP\_TIME\_UTC | Actual departure time (\*) |  |
| LEG\_SCH\_ARR\_AIRPORT | Scheduled arrival airport |  |
| LEG\_SCH\_ARR\_DATE\_UTC | Scheduled arrival date |  |
| LEG\_SCH\_ARR\_TIME\_UTC | Scheduled arrival time |  |
| LEG\_ACT\_ARR\_DATE\_UTC | Actual arrival date (\*) |  |
| LEG\_ACT\_ARR\_TIME\_UTC | Actual arrival time (\*) |  |
| LEG\_DELAY\_CODE\_1 | Delay code 1 |  |
| LEG\_DELAY\_CODE\_2 | Delay code 2 |  |
| LEG\_DELAY\_CODE\_3 | Delay code 3 |  |
| LEG\_DELAY\_CODE\_4 | Delay code 4 |  |
| LEG\_DELAY\_CODE\_5 | Delay code 5 |  |

* Table de référentiel retard : REF\_DELAY.csv

| **Column name** | **Column title** | **Primary key** |
| --- | --- | --- |
| NUM\_CSE\_DLY | Delay id number | X |
| COD\_CSE\_DLY | Delay code |  |
| LIB\_CSE\_DLY\_FRA | Delay label (french) |  |
| LIB\_CSE\_DLY\_ANG | Delay label (english) |  |
| COD\_CLA\_DLY | Delay group |  |
| DAT\_DEB\_VLD | Start of validity period | X |
| DAT\_FIN\_VLD | End of validity period | X |

* Table de référentiel des avions : REF\_AIRCRAFT.csv

| **Column name** | **Column title** | **Primary key** |
| --- | --- | --- |
| FLIGHT\_CARRIER\_CODE | Airline code | X |
| FLIGHT\_NUMBER | Flight number | X |
| AIRCRAFT\_TYPE | Aircraft type |  |

* Table de référentiel géographique : airport\_ref.json

| **Column name** | **Column title** | **Primary key** |
| --- | --- | --- |
| AIRPORT\_CODE | Airport code | X |
| AIRPORT\_NAME | Aiport name |  |
| CITY\_CODE | City code |  |
| CITY\_NAME | City name |  |
| COUNTRY\_CODE | Country code |  |
| COUNTRY\_NAME | Country name |  |
| GEO\_CODE | Zone |  |

* Table de référentiel des compagnies : REF\_AIRLINES.csv

| **Column name** | **Column title** | **Primary key** |
| --- | --- | --- |
| COD\_CIE\_IAT\_2CA | Airline code | X |
| LIB\_CIE\_FRA | Delay label (french) |  |
| LIB\_CIE\_ANG | Delay label (english) |  |
| DAT\_DEB\_VLD | Start of validity period | X |
| DAT\_FIN\_VLD | End of validity period | X |