

# Kata

## Data Analyse



• Click icon to add picture

01.

**Présentation du kata**

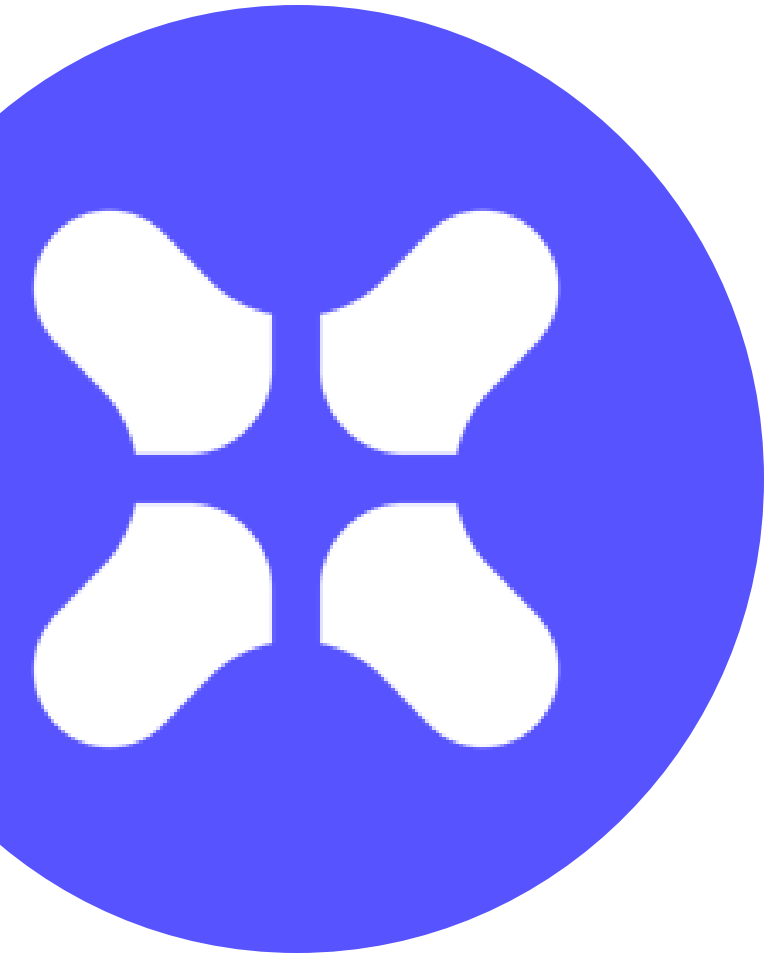
02.

**Test SQL / Python**

03.

**Test Business Intelligence**





## Présentation du kata

Ce kata de data analyse évalue la capacité des candidats à manipuler des données aériennes via **SQL ou Python**, et à créer une **visualisation dans PowerBI**.

L'accent est mis sur la capacité à formuler des **requêtes efficaces** et à résoudre des problèmes en exploitant des sources de données variées.

Enfin, les candidats doivent démontrer leur maîtrise de PowerBI en créant un **tableau de bord interactif**, pour présenter leurs analyses de manière claire et exploitable pour la **prise de décision**.

# Test SQL / Python

À l'aide des 5 tables (flights, airports, planes, airlines et flight\_returns) du modèle de données présenté dans la slide précédente, répondez aux questions suivantes en fournissant des **requêtes SQL** ou **Python** à l'aide de bibliothèques comme Pandas.



# Modèle de données

flights (Table des Vols)

Colonne	Type	Description
flight_id	INTEGER	Identifiant unique du vol
flight_number	STRING	Numéro de vol
departure_airport_id	INTEGER	Référence vers l'aéroport de départ
arrival_airport_id	INTEGER	Référence vers l'aéroport d'arrivée
airline_id	INTEGER	Référence vers la compagnie aérienne
plane_id	INTEGER	Référence vers le modèle d'avion
departure_time	DATETIME	Date et heure de départ
arrival_time	DATETIME	Date et heure d'arrivée
distance	FLOAT	Distance entre l'aéroport de départ et d'arrivée
passenger_count	INTEGER	Nombre de passagers
status	STRING	Statut du vol (en cours, annulé, terminé)
delay	FLOAT	Retard du vol (en minutes)

airports (Table des Aéroports)

Colonne	Type	Description
airport_id	INTEGER	Identifiant unique de l'aéroport
airport_name	STRING	Nom de l'aéroport
city	STRING	Ville où se trouve l'aéroport
country	STRING	Pays où se trouve l'aéroport
latitude	FLOAT	Latitude de l'aéroport
longitude	FLOAT	Longitude de l'aéroport

planes (Table des Modèles d'Avion)

Colonne	Type	Description
plane_id	INTEGER	Identifiant unique de l'avion
model	STRING	Modèle de l'avion
manufacturer	STRING	Constructeur de l'avion (ex : Boeing, Airbus)
seating_capacity	INTEGER	Capacité en nombre de sièges

airlines (Table des Compagnies Aériennes)

Colonne	Type	Description
airline_id	INTEGER	Identifiant unique de la compagnie
airline_name	STRING	Nom de la compagnie aérienne
country	STRING	Pays de la compagnie aérienne

flight\_returns (Table des Retours de Vols)

Colonne	Type	Description
flight_return_id	INTEGER	Identifiant unique du retour
flight_id	INTEGER	Référence vers le vol Annulé ou retourné
return_reason	STRING	Motif de retour ou d'annulation

# Liste des questions

1. Combien y a-t-il de vols en cours et combien de destinations différentes sont desservies ?
2. Quelle est la compagnie aérienne avec le plus grand nombre de vols pour une journée donnée ?
3. Quels sont les trois avions les plus utilisés (modèle d'avion) et combien d'heures de vol ont-ils effectuées ?
4. Quel est le vol avec la distance la plus longue entre deux aéroports ?
5. Quelle est la moyenne du nombre de passagers par vol pour chaque compagnie ?
6. Quel est le numéro du dernier vol pour chaque compagnie aérienne ?
7. Quel est le retard moyen pour chaque compagnie aérienne ?
8. Quel aéroport a enregistré le plus de vols entrants et sortants ?
9. Pour chaque compagnie aérienne, calculez le pourcentage de vols annulés par rapport au nombre total de vols.
10. Quelle est la compagnie aérienne ayant réalisé le plus de vols en interne (vols domestiques) et celle ayant réalisé le plus de vols en externe (vols internationaux) le 7 juin 2024 ?





# Test Business Intelligence



Vous avez le jeu de données "DealData". Ce fichier contient 2 onglets de données.

L'onglet « data » contient des données sur les contrats avec les colonnes suivantes :

- Asset Code : l'identifiant du centre commercial
- Contract code : numéro de contrat unique
- Contract Name : nom du contrat
- Contract Start Date : la date début du contrat
- Contract End Date : la date fin du contrat
- Guaranteed rents : le loyer du contrat

L'onglet « ref » contient des données sur les centres commerciaux avec les colonnes suivantes :

- Asset Code : l'identifiant du centre commercial
- Asset Name : Nom du centre commercial
- Asset City : la ville du centre commercial
- Asset Country : le pays du centre commercial
- Asset Longitude
- Asset Latitude



## Test Business Intelligence

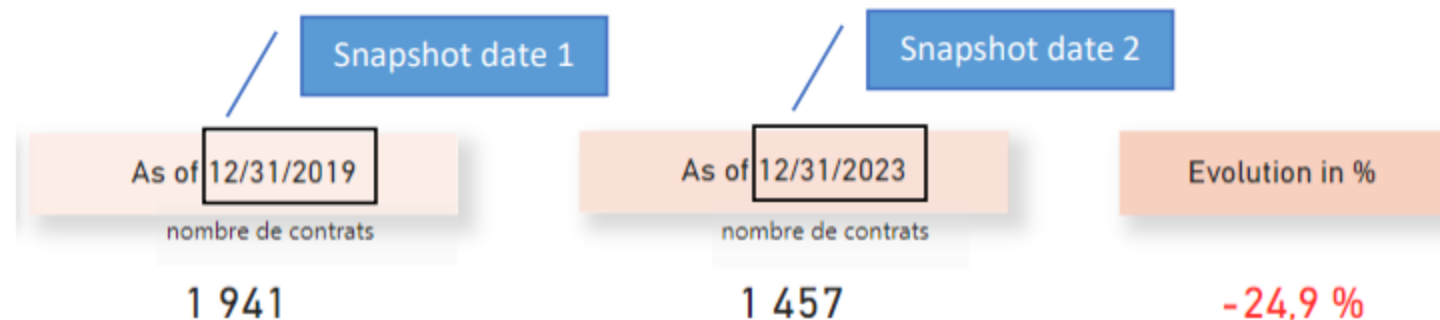
1/Importez ce jeu de données dans PowerBi.

2/Introduire 2 snapshot dates, ensuite calculer le nombre de *contrats en cours* selon la snapshot date choisie par l'utilisateur.

(\*Contrat en cours : la date de fin du contrat > la snapshot date choisie par l'utilisateur)

3/Calculer une variation en nombre de contrats en % entre les 2 snapshot dates

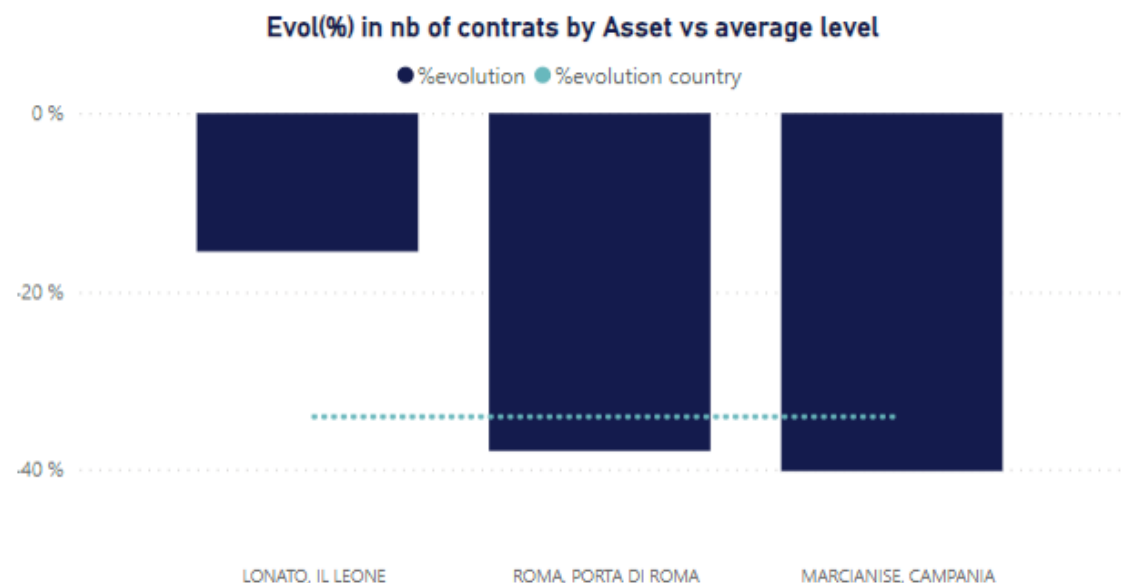
Exemple :



*\*Ces 2 snapshot dates doivent être choisies librement par l'utilisateur*

4/ Etablir un graphique qui montre cette variation en % avec **le nom du centre commercial**, dans ce même graphique, montrer la variation moyenne au niveau du pays (Italie)

Exemple :



5/Rassembler tous ces éléments (résultats) dans un Dashboard.

Dans ce dashboard, il faut avoir 4 filtres qui sont ci-dessous pour l'utilisateur :

- les 2 snapshot dates
- code du centre commercial
- nom du centre commercial

Les résultats varient en fonction du choix de l'utilisateur.

# Bon courage

