De la Donnée Brute à la Décision Stratégique

Portfolio de Projets Data

Piotr Grudzinski

NotebookLM

Une Approche Complète pour Transformer les Données en Valeur

Mon expertise couvre l'ensemble du cycle de vie de la donnée: de l'analyse exploratoire pour déceler des insights à l'ingénierie et la modélisation rigoureuse jusqu'à la création de dashboards décisionnels percutants.

Ingénierie & Transformation de Données

Construire des pipelines de données fiables et des modèles robustes pour des analyses fiables.

dbt SQL

Analyse Exploratoire & Science des Données

Interroger les données pour identifier les facteurs prédictifs et les tendances cachées.

plotly

Business Intelligence & Visualisation

Communiquer des informations complexes de manière claire et actionnable pour les décideurs.

NotebookLM

La base de toute analyse fiable : une ingénierie de données robuste avec dbt et Snowflake.

Le Contexte du Projet "Aircraft Performance Insights"

Le Défi Métier : Analyser en profondeur les performances des compagnies aériennes en croisant des données hétérogènes sur les avions, les aéroports et les revenus.

L'Objectif Technique : Mettre en place une configuration dbt complète pour transformer et modéliser les données brutes stockées sur Snowflake, afin de répondre à des questions business clés.

Stack Technologique

dbt

Snowflake

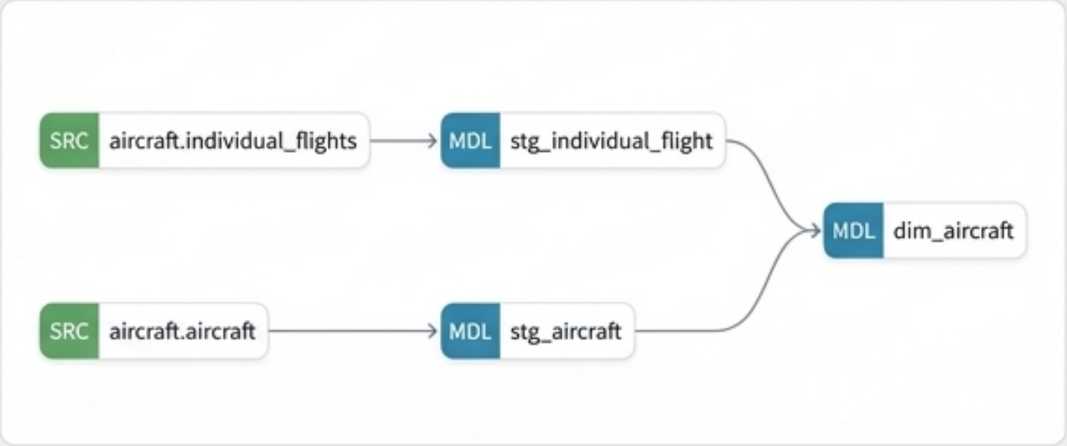
SQL

NotebookLM

D'une question simple à un modèle de données clair : quel avion a le plus volé ?

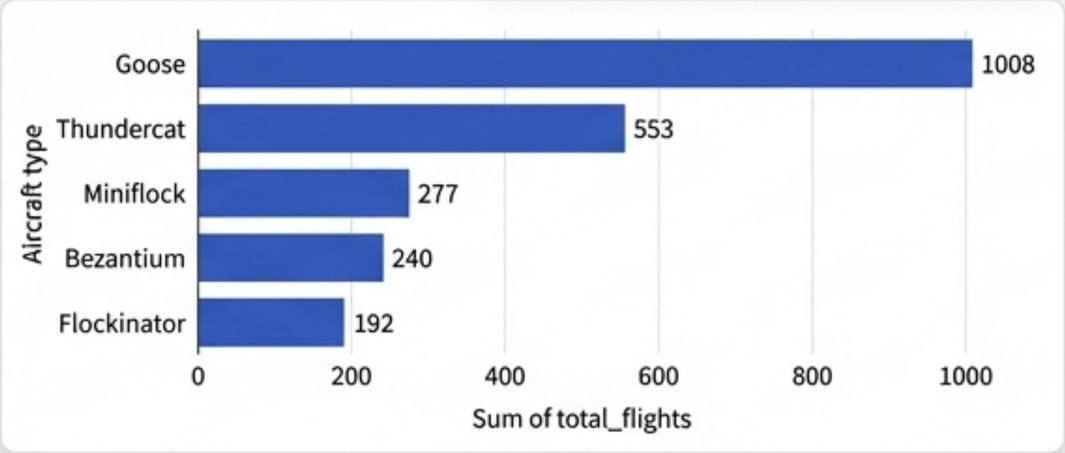
****Question Métier**:** Identifier l'appareil ayant effectué le plus grand nombre de vols.

Gauche - Lignage des Données



Droite - Résultat & Insight

Aircraft_Id	Aircraft_Type	total_flights
12d	Flockinator	192
b23	Bezantium	240
g72	Goose	1008
12a	Miniflock	277
t10	Thundercat	553



Les données sources (`individual_flights`, `aircraft`) sont préparées dans des modèles de staging avant d'être jointes et agrégées dans un modèle de dimension final.

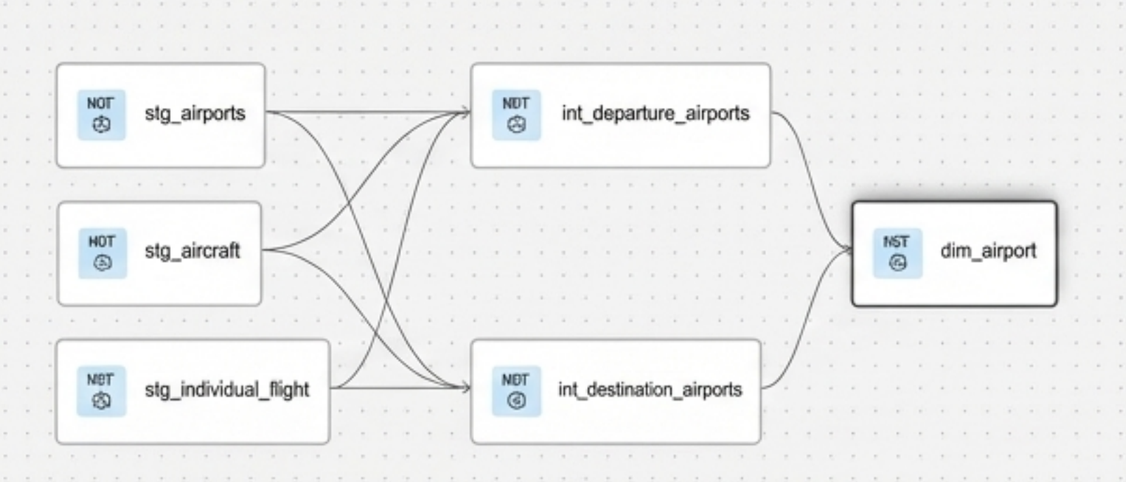
Insight Clé : Le modèle 'Goose' se détache avec 1008 vols, fournissant une réponse directe et vérifiable.

NotebookLM

Gérer la complexité logique : quel aéroport a vu transiter le plus de passagers ?

Question Métier : Calculer le nombre total de passagers transitant par chaque aéroport, en tenant compte à la fois des départs et des arrivées.

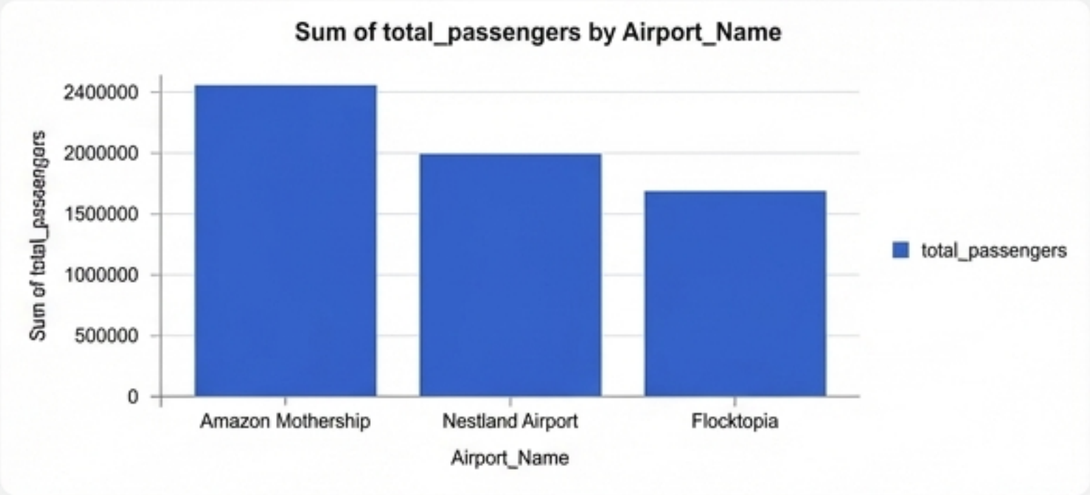
Gauche - Lignage des Données



La logique nécessite des modèles intermédiaires (`int_departure_airports`, `int_destination_airports`) pour traiter séparément les flux. Ces derniers sont ensuite unifiés et agrégés dans le modèle final.

Droite - Résultat & Insight

Airport_Name	total_passengers
Nestland Airport	1999700
Flocktopia	1685200
Amazon Mothership	2423400



Insight Clé: L'aéroport 'Amazon Mothership' domine le classement avec plus de 2.4 millions de passagers.

NotebookLM

Calculer des KPIs business complexes : Identifier la meilleure année de revenus et de croissance.

Questions Métier

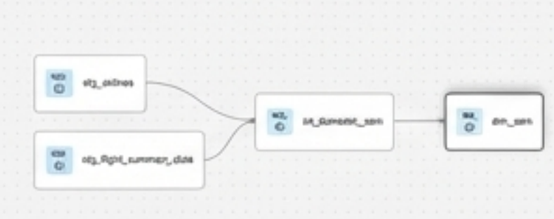
1. Quelle a été la meilleure année en termes de 'Revenue Passenger-Miles' (RPM) pour chaque compagnie ?
2. Quelle a été la meilleure année de croissance, mesurée par 'Available Seat Miles' (ASM) ?

Méthodologie

L'analyse implique la création de modèles intermédiaires pour isoler les métriques, et l'utilisation de fonctions de fenêtrage ("RANK","LAG") pour comparer les performances annuelles et identifier les pics.

Proof Visuelle: Meilleure année RPM

Proof Visuelle: Meilleure année de croissance ASM



Insights Clés :

Goose Airways a atteint son pic de RPM en 2016. Amazon Airlines a connu sa plus forte croissance ASM en 2015.

NotebookLM

Au-delà des chiffres : une analyse exploratoire rigoureuse pour prédire les ‘matchs’ sur Tinder.

Le Contexte du Projet

Face à une baisse du nombre de matchs, une analyse est menée sur les données d’un speed dating pour comprendre les facteurs de compatibilité.

Stack Technique

La Démarche

1. **Sélection des Indicateurs**: Identifier les variables pertinentes (`goal`, `income`, préférences) et écarter les données non exploitables.
2. **‘Sanity Checks’ Systématiques**: Évaluer la complétude et la variabilité des données pour assurer la fiabilité des conclusions. ***Cette étape est non-négociable.***
3. **Visualisation & Inférence**: Confronter la variable cible (`match`) aux indicateurs pour en dégager des corrélations significatives.

NotebookLM

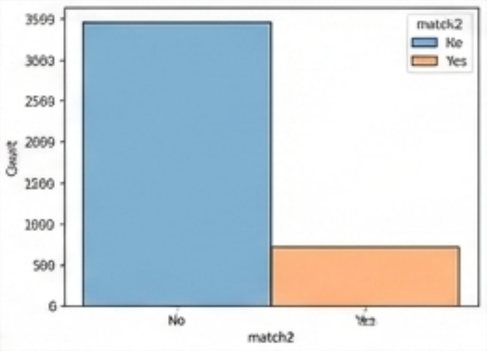
La validation des données est une étape critique : l’exemple de l’indicateur ‘revenu’.

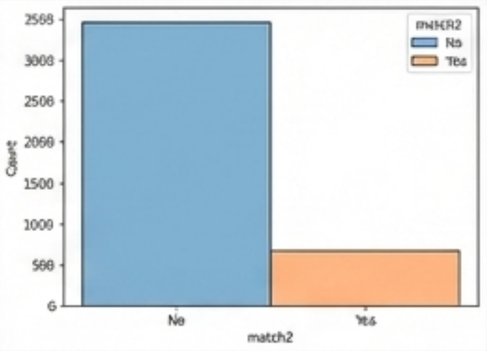
Étape 1 : Gérer les Données Manquantes (51% des données ‘income’)

Question de Validation

La population ayant renseigné son revenu est-elle biaisée par rapport à la variable cible (le 'match') ?

	countMatchYesIncomeNotNull	countMatchNoIncomeNotNull	countMatchYesIncomeNull	countMatchNoIncomeNull	percent_diff_Yes	percent_diff_No
0	644	3455	735	3543	-14.285714	-2.547023





Comparaison de la distribution des matchs entre les deux populations (renseignée vs. non-renseignée).

Les proportions sont 'comparables de manière satisfaisante'. L'indicateur est donc utilisable malgré les données manquantes.

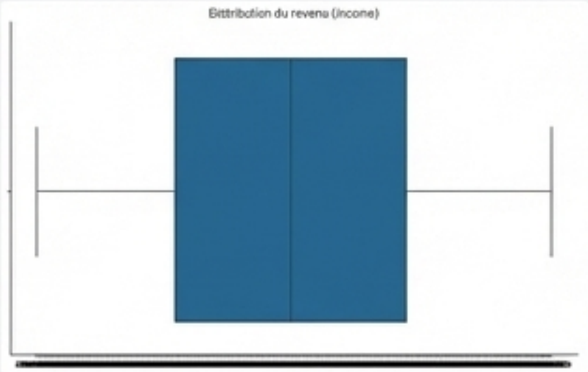
Étape 2 : Évaluer la Variabilité

Question de Validation

La diversité des données est-elle suffisante pour dresser des tendances ?

Test

Analyse de la distribution statistique.




```
count      4279.000000
mean      44887.606450
std       17206.920962
min        8607.000000
25%       31516.000000
50%       43185.000000
75%       54303.000000
max      189031.000000
Name: income, dtype: float64
```

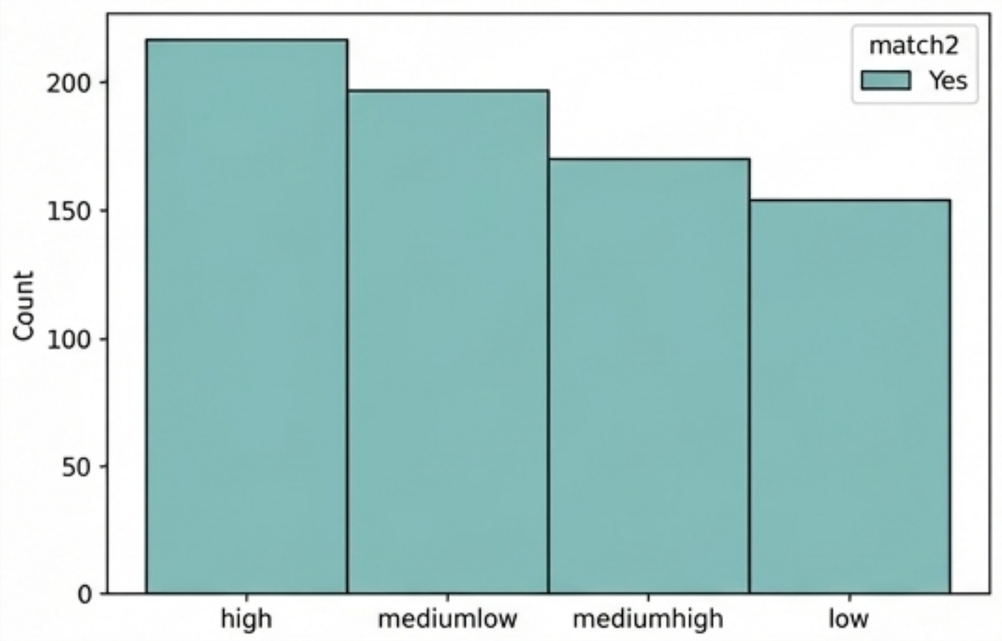
Variabilité suffisante (écart-type de 38% de la moyenne) et absence d’outliers. L'analyse peut procéder.

NotebookLM

Revenus élevés et préférences culturelles : des facteurs clés pour obtenir un ‘match’.

Partie 1 : Corrélation avec le Revenu

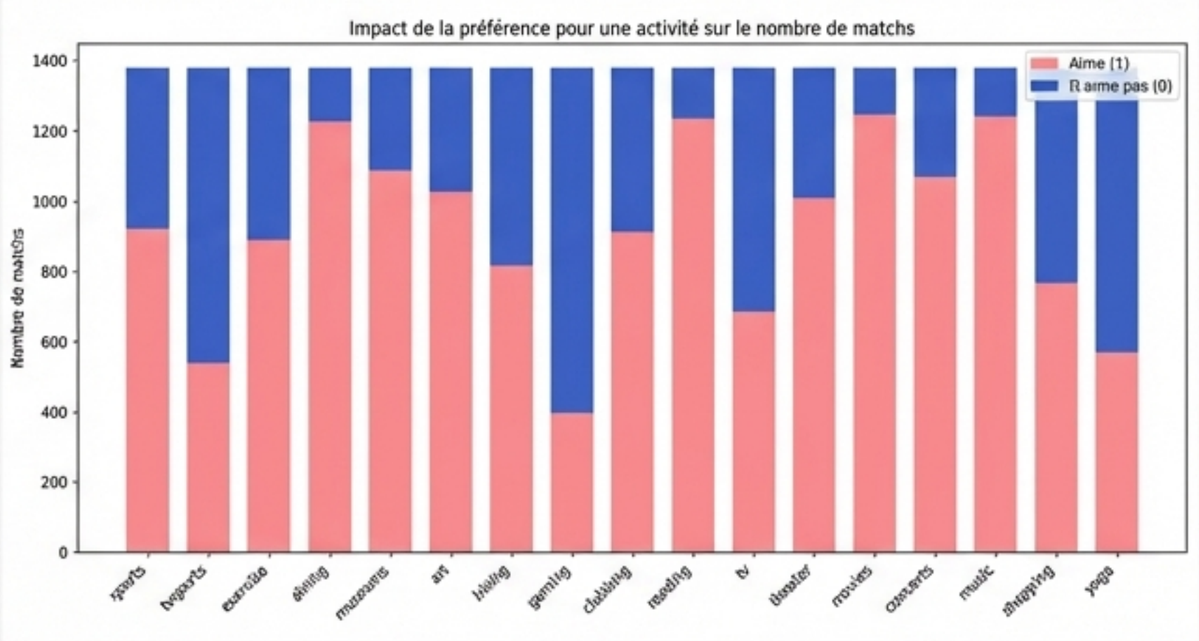
Méthode : Création de catégories de revenus et filtrage sur les 'match=Yes' pour une meilleure lisibilité.



Conclusion : Les revenus les plus élevés sont plus susceptibles d'obtenir un match que les revenus faibles.

Partie 2 : Corrélation avec les Préférences d’Activités

Méthode : Réarrangement des préférences en catégories binaires (Aime / N’aime pas) et visualisation via un histogramme empilé.



Conclusion : Les activités culturelles et sociales (sorties, lecture, films, musique) sont fortement associées à un nombre élevé de matchs, contrairement aux jeux vidéo ou au visionnage de sport à la TV.

NotebookLM

Rendre les données actionnables : des dashboards pour piloter la décision.

L'Objectif:

Transformer des analyses complexes en outils de pilotage interactifs et intuitifs, conçus pour les décideurs.

L'Approche:

Ingestion des données, transformation via des outils comme Power Query, et création de tableaux de bord qui synthétisent les KPIs et permettent une exploration dynamique.

Projets Présentés:

- Analyse du catalogue mondial de Netflix.
- Dashboard décisionnel sur mesure avec Power BI.

Stack Technique:

NotebookLM

Synthèse visuelle des KPIs : du catalogue Netflix à l'analyse de marché.

Analyse du Catalogue Netflix

- **8,809** titres au total (69.6% films / 30.4% séries).
- **Insight 1:** Les 10 premiers pays (sur 123) totalisent **75%** des titres, dont **40%** pour les USA seuls.
- **Insight 2**: Les **Drames** sont le genre dominant globalement, sauf en Afrique où la Comédie prévaut.
- **Insight 3**: La saisonnalité des sorties est marquée par continent (ex : S2 en Amériques, S4 en Europe).

Dashboard Décisionnel Power BI

Création d'un tableau de bord interactif de bout en bout : de l'ingestion et la transformation des données brutes avec **Power Query** jusqu'à la visualisation finale des indicateurs clés pour la prise de décision stratégique.

Suivi des ventes et performance opérationnelle.

NotebookLM

Une expertise polyvalente sur l'ensemble de la Modern Data Stack.

Résumé des Compétences Démontrées

1. **Ingénierie de Données**: Modélisation et transformation complexes avec dbt & Snowflake pour créer des fondations de données fiables.
2. **Analyse de Données**: Analyse exploratoire rigoureuse avec Python pour extraire des insights prédictifs à partir de données brutes.
3. **Business Intelligence**: Création de dashboards percutants avec Power BI pour faciliter la prise de décision.

Stack Technique Maîtrisée

dbt SQL plotly Flask
web development, one drop at a time

NotebookLM

Collaborons sur vos prochains défis data.

Informations de Contact:

Nom: Piotr Grudzinski

Email: pgrudzinski@gmail.com

06 71 68 63 05

NotebookLM