

Analyse de score de bonheur par pays

Basé sur l'enquête annuelle de 2021 dans plus de 160 pays

Objectifs :

- Quels facteurs influencent le plus le score de bonheur ?
- Quels sont les pays qui ont le meilleur score ?
- Le bonheur diminue-t-il à cause d'événements ?

Méthodologie :

- EDA : statistique, valeurs manquantes, normalisation
- Visualisation : heatmap, corrélation, barplot, cartographie
- Analyse prédictive : preprocessing, Régression linéaire
- Analyse critique des résultats et perspectives

Outils : Python, Tableau, VS Studio, Machine Learning

Analyse de Speed dating (Tinder)

Basé sur 8400 lignes et 195 colonnes entre 2002 et 2024

Objectifs :

- Analyse du jeu de données : qui sont les participants ?
- Quel facteur joue le plus dans un match ?

Méthodologie :

- EDA : statistique, valeurs manquantes, normalisation cat & num
- Visualisation (heatmap, corrélation, pie-charts, donut)
- Création de rapport d'analyse dans notebook ipynb

Outils : Python, Pandas, Matplotlib, Seaborn, VS Studio

Étude des données aéronautiques

Basé sur les données de vols (trois compagnies aériennes)

Objectifs :

- Quel aéroport et type d'avion ont le plus fort trafic ?
- Quelle a été la meilleure année en termes de RPM ?
- Meilleure année de croissance pour chaque compagnie ?

Méthodologie :

- Importation des données dans Snowflake via SQL
- Création de Warehouse
- Connexion à Deepnote et création de notebook
- Extraction et agrégation des données avec SQL
- Visualisation des extractions SQL (barchart, linechart)
- Création de rapport d'analyse dans notebook Deepnote

Outils : Snowflake, DBT, Deepnote, SQL

PORTFOLIO PROJETS DATA

Analyse du catalogue Netflix

Basé sur 8800 lignes du catalogue 2021

Objectifs :

- Quels genres sont les plus diffusés sur Netflix ?
- Quelle distribution mondiale des produits Netflix par pays ?
- Quelle est la saisonnalité des sorties Netflix ?

Méthodologie :

- Importation des données dans Power BI
- EDA : statistique, valeurs manquantes, normalisation
- Préparation en vue de la visualisation (split, measure..)
- Visualisation : graphiques, cartes, tableau de données, filtres
- Création et personnalisation de dashboard interactif multi pages

Outils : Power BI (Visualisations, Transform Data), Dax

Dataiku – Détection de fraude à la carte bancaire

Basé sur 300000 transactions en 2017-2018 de Elo merchant (Société brésilienne, spécialisée dans les cartes de paiement)

Objectifs :

- Comment prédire les transactions frauduleuses en minimisant les faux positifs ?
- Comment identifier les facteurs clés de la fraude ?

Méthodologie :

- Préparation et exploration des données sur Dataiku (feature engineering)
- Entraînement de modèles de ML supervisé : régression logistique, Random Forest
- Comparaison et évaluation des modèles

Outils : Dataiku, Python, Machine Learning, SQL

Clés du succès des jeux vidéo (3 plateformes)

Basé sur 131 000 jeux, 1M de joueurs, 250 pays, 6 datasets

Objectifs :

- Facteurs économiques : comment prix et ventes influencent le succès ?
- Facteurs d'adoption : où sont les joueurs et quels jeux s'imposent ?

Méthodologie :

- EDA : statistique, valeurs manquantes, concaténation
- Agrégation des données pour identifier les tendances
- Export des résultats SQL-like Python (Pandas) au format CSV
- Import dans Power BI pour les visualisations

Outils : Python, Power BI, Git, Git Hub