

Évaluation

Python (orienté data)

Évaluation

Vous avez tout juste décroché un poste pour le ministère de la Culture 🕶️

Ils vous ont chargé d'étudier l'**impact des infrastructures cinématographiques (écrans et fauteuils) sur les entrées annuelles dans les cinémas français.**

👉 Objectif :

- Cette étude permettra de déterminer si l'ajout d'écrans ou de fauteuils influence significativement la fréquentation.
- Vous devrez aussi construire un modèle prédictif pour estimer les entrées futures en fonction des infrastructures.

Ce devoir vous permettra de valider tout ce qui a été appris depuis le début de ce cours.

Objectifs de l'évaluation

- Manipuler et explorer un dataset avec **Pandas**.
- Réaliser des visualisations claires et pertinentes avec **Matplotlib**.
- Effectuer des analyses statistiques et interpréter les corrélations.
- Construire un modèle prédictif simple avec **Scikit-learn** pour prédire les entrées.

Dataset fourni

Un fichier CSV officiel [salle de cinema ile-de-france.csv](#) (à nommer cinemas.csv) contenant des données, dont :

- `region` : Région administrative du cinéma.
- `commune` : Nom de la commune.
- `population_commune` : Population totale de la commune.
- `ecrans` : Nombre d'écrans dans le cinéma.
- `fauteuils` : Nombre total de fauteuils.
- `annee` : Année (2018 à 2022).
- `entrees_annuelles` : Nombre total d'entrées pour cette année.
- `label_art_et_essai` : Indique si le cinéma a un label Art et Essai (oui/non).

Exercice 1 : Nettoyage et exploration des données (5 pts)

Nettoyage :

1. Chargez le fichier CSV dans un DataFrame.
2. Identifiez les valeurs manquantes ou incohérentes et traitez-les dans votre programme.
3. Laissez un court commentaire de code expliquant pourquoi vous avez gardé ces valeurs.

Exploration :

1. Affichez les premières lignes du dataset dans le terminal pour visualiser son contenu.
2. Affichez des statistiques descriptives des colonnes numériques principales (fauteuils, écrans, entrées annuelles).

Exercice 2 : Analyse des données (4 pts)

1. Calculez les entrées moyennes par fauteuil pour chaque région en 2022.
2. Identifiez les 3 régions ayant :
 - Les meilleurs résultats en termes d'entrées moyennes par fauteuil.
 - Les pires résultats.
3. Créez un graphique à barres représentant les entrées moyennes par fauteuil pour les 10 régions.

Exercice 3 : Corrélation entre infrastructures et fréquentation (4 pts)

1. Filtrez les données pour ne garder que l'année 2022.
2. Calculez et affichez la corrélation entre :
 - Le nombre d'écrans et les entrées annuelles.
 - Le nombre de fauteuils et les entrées annuelles.
3. Créez un nuage de points pour chacune des relations ci-dessus, avec une régression linéaire superposée (utilisez Seaborn ou Matplotlib).
4. Question : Selon vous, quelle est la variable ayant le plus d'impact sur les entrées annuelles ? Justifiez votre réponse avec les visualisations. (Réponse à écrire dans le Readme)

Exercice 4 : Modèle prédictif des entrées annuelles (5 pts)

1. Divisez les données de 2018 à 2021 en deux sous-ensembles :
 - Variables explicatives : ecrans, fauteuils, population_commune.
 - Variable cible : entrees_annuelles.
2. Entraînez un modèle de régression linéaire pour prédire les entrées annuelles en fonction des variables explicatives.
3. Évaluez les performances du modèle avec un split en train/test (80%/20%).
 - Affichez le coefficient de détermination (R^2) et l'erreur moyenne absolue (MAE).
4. Testez votre modèle sur les données de 2022 et comparez les prédictions avec les valeurs réelles.
5. Question : Selon les performances du modèle, le nombre d'écrans ou de fauteuils est-il un bon prédicteur des entrées ? Expliquez.

Exercice 5 : Recommandations stratégiques (2 pts)

1. Imaginez qu'un cinéma d'une commune fictive (20 000 habitants) souhaite augmenter ses entrées :
 - Il dispose actuellement de 2 écrans et 120 fauteuils.
 - Proposez une stratégie (en augmentant les écrans ou fauteuils) basée sur les résultats de votre analyse.
2. Justifiez votre recommandation avec des calculs simples.

Evaluation - Remarques

N'oubliez pas les bonnes pratiques, ça compte dans le barème !

- Un README clair et propre décrivant ce que fait le projet et comment l'utiliser
- Les bonnes pratiques Git : **des commits effectués à chaque étape de votre avancement** sont essentiels pour que je comprenne votre raisonnement et pour que je vous note correctement. Sans cela, vous perdrez beaucoup de points.
- Un code plus propre est mieux récompensé : bien rangé, séparation des responsabilités de chaque fonction ou classe...

Où répondre aux questions d'exercices sans code ?

Répondez-y dans le README, dans une section claire dédiée

L'évaluation est notée sur 20.

Comment rendre le devoir ? (Évaluation globale)

Envoyez-moi un email à josselin.jtech@gmail.com contenant :

Le lien vers votre repository en ligne **accessibles publiquement**, nommé CinemaDataAnalysis.

Attention :

- C'est un exercice individuel. Si le rendu est celui d'un camarade, je le remarquerai.
- Si le repository n'est pas en **public**, je ne peux pas y accéder et c'est 0.
Pour vérifier si le projet est bien public, allez sur une nouvelle fenêtre en navigation privée de votre navigateur.

Si vous avez des questions, n'hésitez pas à me les poser.

Date de rendu : vous avez 12 jours pour me rendre le devoir.