

Rapport de Projet : Analyse des Biais de Notation de Films (Cas Fandango)

1. Introduction et Problématique

Contexte : Le projet s'inspire d'une enquête du site *FiveThirtyEight*. L'objectif est de vérifier si le site **Fandango**, qui vend des billets de cinéma, gonfle artificiellement les notes des films pour booster ses ventes. **Problématique :** Peut-on faire confiance aux notes d'une plateforme qui a un intérêt financier à ce que les films soient bien notés ?

2. Méthodologie et Outils

- **Langage :** Python
- **Bibliothèques :** * **Pandas & Numpy :** Manipulation, nettoyage et fusion de données.
 - **Matplotlib & Seaborn :** Visualisation avancée (KDE plots, Scatter plots).
- **Données :** Deux datasets comparant les notes de Fandango avec celles de Rotten Tomatoes, Metacritic et IMDB.

3. Analyse Étape par Étape

A. Analyse de Fandango (L'anomalie interne)

J'ai commencé par comparer les notes réellement calculées par les utilisateurs (RATING) et les étoiles affichées sur l'interface de Fandango (STARS).

- **Observation :** Le graphique de distribution (KDE Plot) a révélé un décalage systématique vers la droite.
- **Interprétation :** Fandango ne se contente pas d'arrondir à l'entier le plus proche, il arrondit presque toujours à l'unité **supérieure**. Un film noté 3.1 peut ainsi apparaître avec 3.5 ou 4 étoiles.

B. Comparaison Multi-plateformes (L'objectivité)

Pour valider le biais, j'ai normalisé toutes les notes (RT, IMDB, Metacritic) sur une échelle de 0 à 5.

- **Observation :** Alors que les autres sites présentent des distributions en cloche (avec des films très mal notés et d'autres très bien notés), Fandango ne descend quasiment jamais en dessous de 3 étoiles.
- **Interprétation :** Fandango affiche une vision du cinéma où "tout est bon", contrairement aux critiques professionnels de Rotten Tomatoes qui sont beaucoup plus sévères et nuancés.

C. Le test des "Pires Films"

J'ai isolé les 10 films les plus détestés par les critiques mondiales.

- **Résultat frappant :** Des films avec des scores proches de 0.5/5 sur les autres sites conservaient des notes de 3/5 ou plus sur Fandango.

4. Conclusion du Projet

Ce projet démontre une **corrélation entre l'intérêt commercial et la présentation des données**.

- **Compétences démontrées :** Nettoyage de données complexes, normalisation de variables, détection d'anomalies statistiques et communication visuelle d'un problème métier.