



Moovie Data Analysis

MOHAMED OLOU

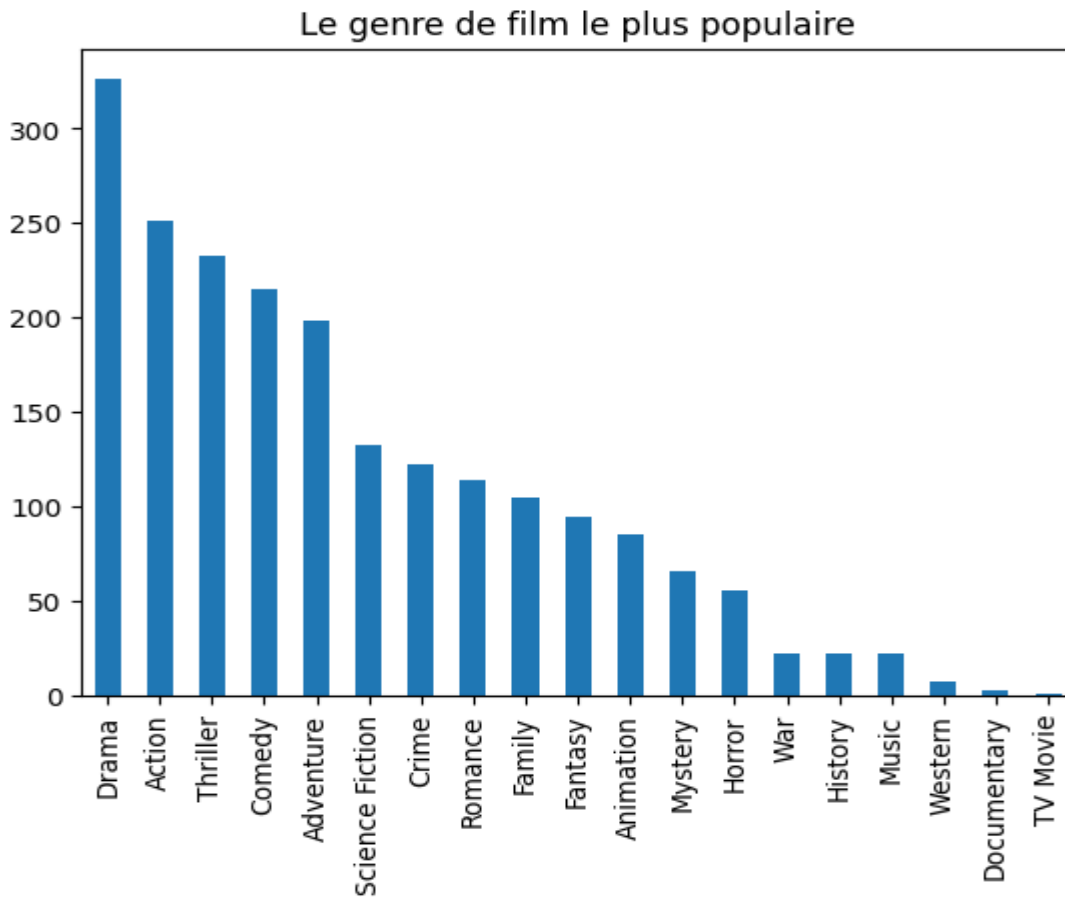
- For this analysis we had to use 1 dataset: tmdb-movies.csv

Durant l'analyse de ce dataset nous avons voulu répondre à certaines questions qui sont :

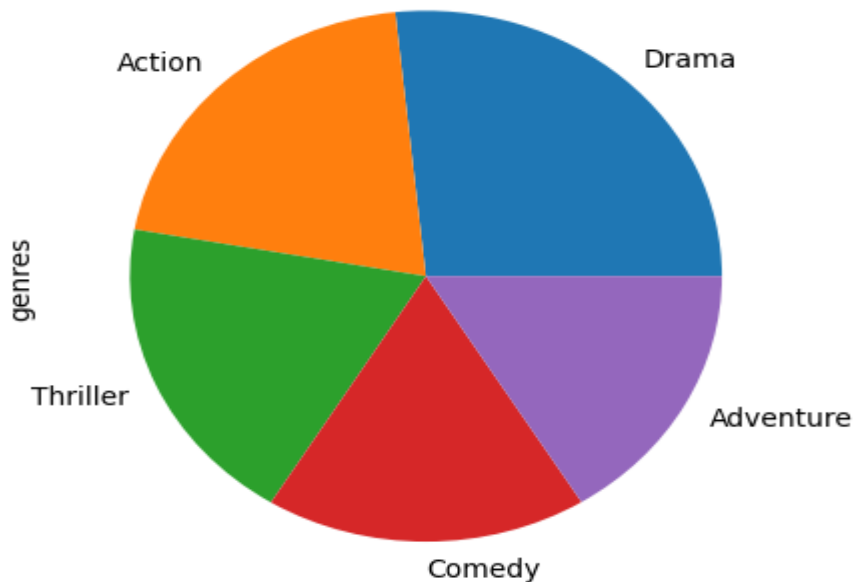
- Quel est le genres le plus populaire sur toutes les années ?
- Quel est le film le plus connu de tous les temps ?
- Quel sont les films au plus grand revenus ?
- Quelles sont les données qui sont associées au films aux plus gros revenus ?

Quel est le genre le plus populaire sur toutes les années ?

Pour répondre a cette question, nous avons groupé le dataset par genre afin d'avoir des données en fonction du genre de film. Grace à cela nous avons pu tracer un histogramme qui présente le nombre de film avec ce genre. On obtient au final comme genre le plus populaire le genre Drame.



Le genre de film le plus populaire



| Genre | Nombre de films |
|----------|-----------------|
| Drame | 326 |
| Action | 251 |
| Thriller | 233 |

Nombre de film suivant les genres

Quel est le film le plus connu de tous les temps ?

Pour répondre à cette question, nous avons d'abord filtré les données afin d'obtenir les films qui ont assez de vote et qui sont bien noté.

Après cela on a filtré ces données par le score de popularité. On obtient finalement comme film le plus connu : Jurassic World.

Quelles sont les données associées aux films au plus grand revenus ?

Pour répondre a cette question, nous avons tracé les histogrammes des données numériques pour voir celles qui sont plus semblable à celle du revenu.

Pour étayer ces résultats nous avons cherché les corrélations entre les données numériques du dataset. On remarque que les données en corrélation avec le revenu sont : le nombre de vote, le budget et la popularité.

| | Revenu |
|-----------------------|---------------|
| Nombre de vote | 0.804788 |
| Budget | 0.747273 |
| Popularité | 0.641346 |

Corrélation des données avec le revenu