Microsoft-Movie-Analysis project

By Annastasiah Muchau

Business problem

▶ Create a new movie studio on movies that are currently doing the best.

The process/Steps followed

- Collect the necessary data required.

- Cleaning the data.

- Explore the data.

- Data visualization.

1. Collecting Data

▶ The data collected/ used were 3 datasets, namely:

    a)   mdb.title.basics.csv.gz (Referenced as "title")

    b)   bom.movie_gross.csv.gz (Referenced as "gross")
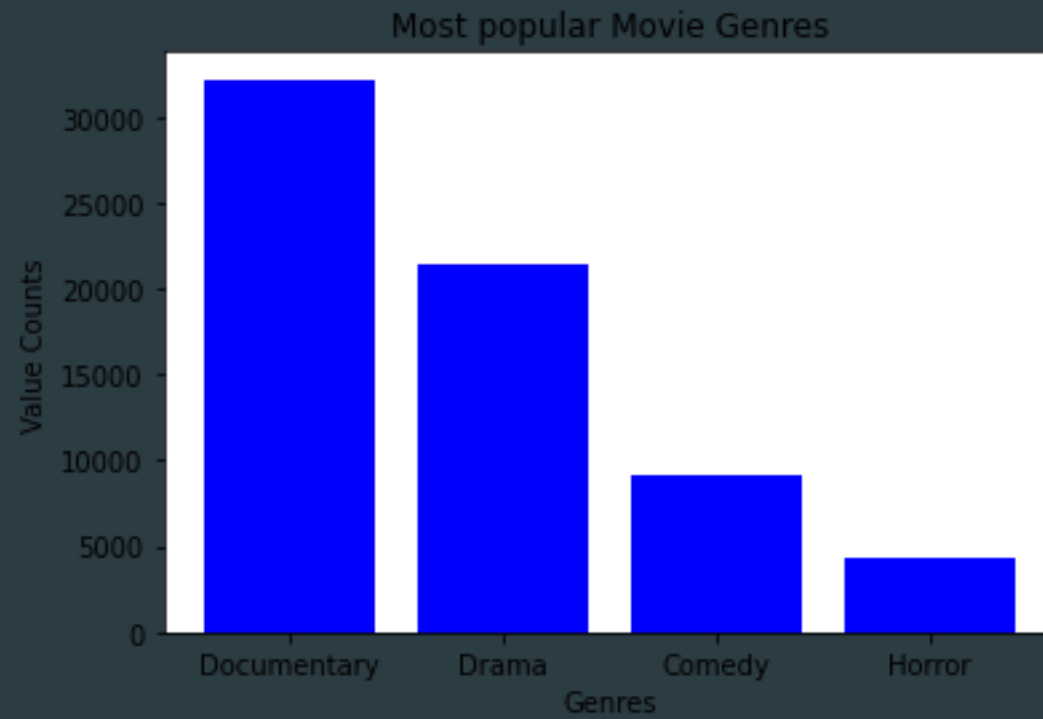
    c)   tmdb.movies.csv.gz (Referenced as "movies"

Data Cleaning

- The main purpose of this process is to remove "dirty" data that could hinder data analysis. The process involves checking for missing data, misspelled text data and invalid data.

- From our datasets, I checked and worked on missing values and duplicates

Data Visualization

- The data was represented using bar chart and correlation diagram

Bar Chart

Graph Interpretation

▶ The bar chart above shows the top 4 most popular movie genres. They are:

1. Documentary
2. Drama
3. Comedy
4. Horror movies

Recommendations

▶ I would recommend the company to:

1. To create Documentary Movies because it's the most popular movie genre.

2. Create movies in/using English language.

3. To sell their movies outside the production country since the foreign gross is more than the domestic gross

END