Netflix Data Visualization Server

CS699:Software Lab Project Report

Instructor: **Prof. Bhaskaran Raman**

Submitted By:**Harikrishna (22m0759)**

Department of Computer Science and Engineering

Indian Institute of Technology Bombay,

Powai, Mumbai (Maharashtra)

Introduction:-

There are various OTT platforms functional in India like – Amazon Prime Video, Netflix, Hotstar, Voot etc, which are gaining more viewership day by day.

With the help of the project that we have intended to develop, we will analyze the data of Netflix and build a user interface to interact with analysis
.We have Netflix dataset which has the names of movies and shows that were released till mid 2021. Each movie or show has attributes like director, cast, release year etc.
Based on the data, we will try to analyze with visualization different aspects (director, cast, type etc.) that affected the rating and popularity of shows.

The technologies that we are going to use for the development of this project are – Python, Pyplot and HTML Pyplot to visualize the data in the form of graphs and HTML to showcase the inferences in the textual format.

The analysis that we have done are:-
1. Top 15 countries on Netflix
2. Total content added across all years
3. Distribution of Movies and TV Shows
4. Top 10 Directors on Netflix
5. Content Distribution by Rating
6. Top 10 actors in Movies and TV Shows
7. Search features based on actor and director are also implemented.
   1.Search the statistics by actors name
   2.Search the statistics by directors name.

Key Inferences:-

1. With the help of this analysis task, we were able to understand the trends in the production of Movies and TV Shows over the years by the Netflix.
2. How different actors and directors emerged with time.
3. The kind of ratings most number of TV Shows and Movies received.
4. Evolution of actor w.r.t to number of shows genres, movies etc.

LEARNINGS:-

1. We learnt how to use pandas for data analysis

2. We learnt how much the pyplot is useful to visualize different representations of data

3. We were able to built Web Server using Python for serving the HTML, CSS, JS and Image files.

4.Dynamics handling of web queries.

HIGH LEVEL DOCUMENTATION:-

Project folder has following folders-

1.resources: contains the html files to be served.
2.images: contains the dynamic and static images which will be served.
3:data: contains Netflix_titles.csv
4: server.py: consist of basic python http server, mainly focuses on routing the requests and handling the dynamic queries.
5: result_actor.py: consist of logic to create summary graphs based on a search query for actor.
6: result_director.py: consist of logic to create summary graphs based on a search query for actor
7: style.css: has the styling properties for the htmls served.

RUNNING INSTRUCTIONS:-

Before running the python code, one should have following modules installed on his/her machine-

1. numpy
2. pandas
3. matplotlib & matplotlib.pyplot
4. collections
5. http.server

Jupyter notebook contains the raw data explorating and generation of static image files for overall summary

o on cmdline type python3 server.py 8001
o On any browser navigate to localhost:8001

Current progress:

1.The summary section has been implemented .
2.Dynamic search feature for actor and director is being implemented.

POSSIBLE FUTURE WORK PENDING:-

1. Display the latest and trending titles for an actor.
2. Adding functionalities for search.

Github: https://github.com/master-hkshenoycse/MovieAnalyzerServer/tree/master

References:-

1. Netflix dataset

    https://www.kaggle.com/datasets/shivamb/netflix-shows

2. Matplotlib references

    https://matplotlib.org/stable/gallery/pyplots/index.html

    https://www.w3schools.com/python/matplotlib_pyplot.asp

3. Pandas references
    https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html

    https://stackoverflow.com/questions/16476924/how-to-iterate-over-rows-in-a-dataframe-in-pandas