NETFLIX DATA ANALYTICS PROJECT SUMMARY

BY YASHASVI DEWANGAN

DATA ANALYTICS | PYTHON

This project performs **exploratory data analysis (EDA)** on a Netflix movie dataset using Python. The goal of the analysis is to examine genre distribution, audience engagement through vote counts and popularity, and yearly trends in movie releases.

Project Workflow

Introduction

The analysis follows a structured workflow including data loading, initial exploration, data cleaning, preprocessing, exploratory analysis, visualization, and finally answering key analytical questions derived from the dataset.

01

Data Loading

Import and read the Netflix dataset

02

Initial Exploration

Examine structure and columns

03

Data Cleaning

Handle missing values and duplicates

04

Preprocessing

Prepare data for analysis

05

Exploratory Analysis

Analyze patterns and trends

06

Visualization

Create graphs and charts

07

Key Questions

Answer analytical questions

Objectives of the Project

The objectives of this project are:

Dataset Understanding

To understand the structure and columns of the Netflix dataset

Data Loading & Exploration

To perform data loading and initial exploration

Data Preparation

To clean and preprocess the dataset

Genre Analysis

To analyze genre-wise distribution of movies

Audience Engagement

To study audience engagement using vote counts and popularity

Popularity Insights

To identify movies with the highest and lowest popularity

Yearly Trends

To determine the year with the highest number of movie releases

Visual Support

To support analytical findings using visualizations

Tools and Libraries Used

The following Python libraries are used in this project:

Pandas

for data loading, cleaning, and manipulation

NumPy

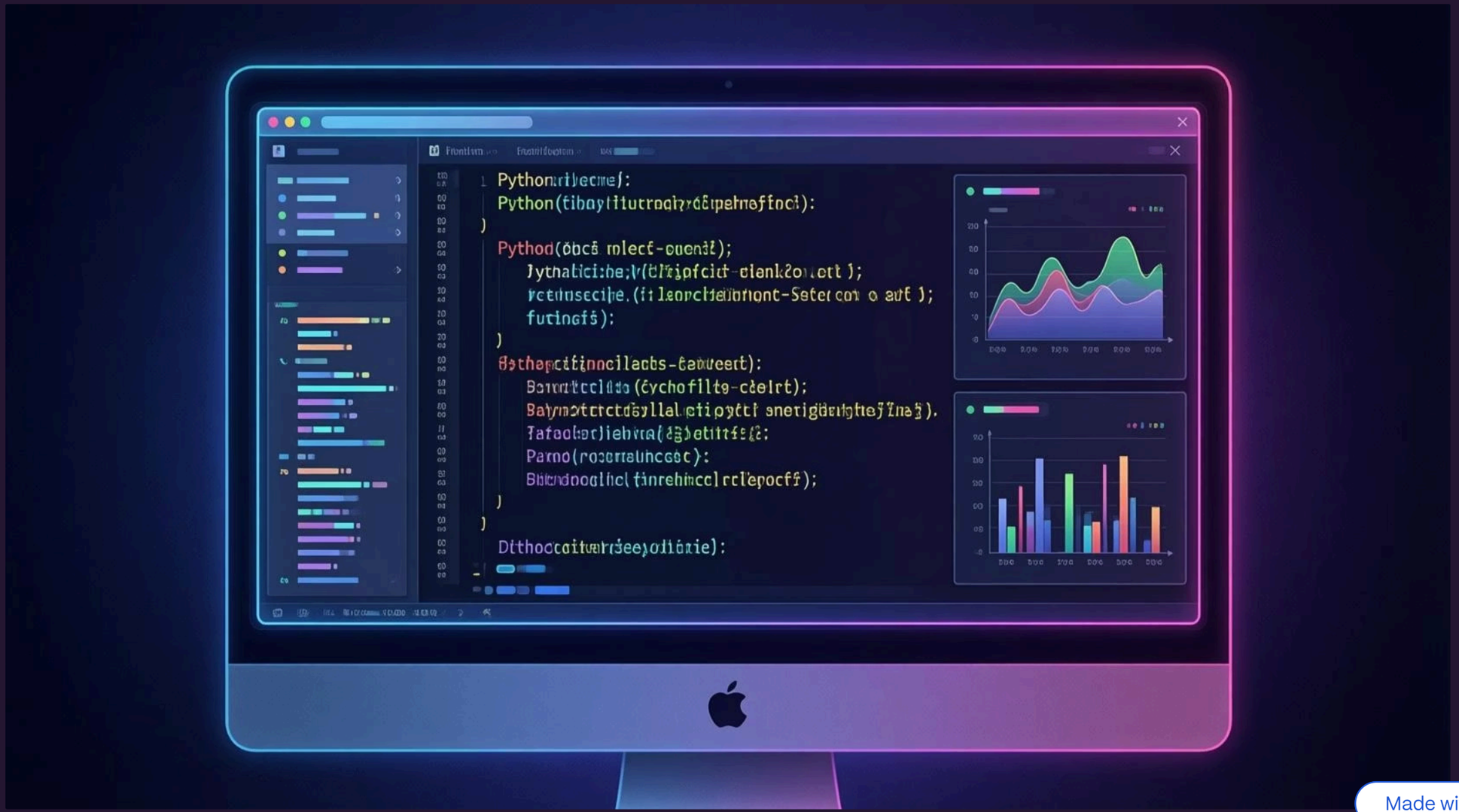
for numerical operations

Matplotlib

for plotting graphs

Seaborn

for enhanced data visualizations



Dataset Description

The dataset used in this project is loaded from a CSV file and initially inspected using functions such as df.head(), df.info(), and df.columns().

Dataset Columns (As per Original Dataset)

Based on the original dataset structure shown in the CSV file, the dataset contains the following columns:

- **Release_Date** – Official release date of the movie
- **Title** – Name of the movie
- **Overview** – Short description or summary of the movie
- **Popularity** – Popularity score indicating audience interest
- **Vote_Count** – Total number of votes received by the movie
- **Vote_Average** – Average rating based on votes
- **Original_Language** – Original language of the movie
- **Genre** – Genre(s) associated with the movie
- **Poster_Url** – URL link to the movie poster

The dataset includes both **categorical** and **numerical** features, making it suitable for exploratory data analysis and visualization.

Data Loading and Initial Exploration

The dataset is loaded into a Pandas DataFrame using read_csv().

Initial exploration is carried out using:

df.head()

to view the first few rows and understand column values

df.shape

to determine the number of rows and columns

df.columns

to list all column names

df.info()

to inspect data types and identify missing values

This step provides a clear overview of the dataset and confirms successful data loading.

Data Cleaning and Preprocessing

Before performing analysis, the following preprocessing steps are carried out:

Duplicate Record Check

- Duplicate rows are checked using df.duplicated()
- **Result:** No duplicate records are found

Missing Value Analysis

- Missing values are identified using isnull().sum()
- Columns with missing values are reviewed to assess their impact on analysis

Data Preparation

- Relevant columns such as **Genre**, **Popularity**, **Vote_Count**, and **Release_Date** are selected for analysis
- Data types are verified for consistency
- The dataset is prepared for exploratory analysis and visualization

After these steps, the dataset is considered clean and ready for analysis.

Key Findings

Exploratory Data Analysis with Visual Insights

To make the analysis more intuitive and engaging, graphs are plotted at each stage. These visualizations support numerical findings and help in identifying patterns within the dataset.

Genre Distribution

Drama is the most frequent genre

Vote Analysis

Drama has highest vote counts

Popularity Trends

Action movies dominate rankings

Yearly Patterns

2022 had most releases

Detailed Analysis Questions

Question 1: What is the most frequent genre of movies released on Netflix?

Answer: Drama is identified as the most frequent genre among the 19 different genre categories present in the dataset.

Insight: The visualization clearly shows Drama dominating other genres, indicating Netflix's strong focus on drama-based content.

Question 2: Which genre has the highest number of votes?

Answer: Drama has the highest total number of votes among all genres, indicating strong audience engagement.

Insight: Genres with higher representation, particularly Drama, tend to receive more audience votes.

Question 3: Which movie has the highest popularity, and what is its genre?

Answer: The movie with the highest popularity is *Spider-Man: No Way Home*. Its genres include **Action**, **Adventure**, and **Science Fiction**.

Insight: Highly popular movies are often action-oriented and franchise-based, attracting a larger audience.

Question 4: Which movie has the lowest popularity, and what is its genre?

Answer: The movies with the lowest popularity are:

- *The United States vs. Billie Holiday* – Genre: Music, Drama, History
- *Threads* – Genre: Science Fiction, War, Drama

Insight: Movies with less common genre combinations show lower popularity compared to highly popular, action-oriented movies.

Question 5: Which year has the most filmed movies?

Answer: Based on the analysis of movie releases by year (derived from the release date), **2022** has the highest number of filmed movies.

Insight: The visualization shows a clear peak in movie production in 2022, indicating increased content creation during that year.

Overall Summary of Findings

Based on the complete exploratory analysis and visualizations:

Drama Dominates

Drama is the most frequent and most voted genre

Action Popularity

Action-oriented movies dominate popularity rankings

Niche Genres

Lower Popularity for Certain Genre Combinations

2022 Peak

Movie production peaked in 2022

Visual Insights

Visualizations strongly support numerical insights

Conclusion

This Netflix Data Analytics Project Summary demonstrates a complete exploratory data analysis workflow using Python. By integrating data loading, initial exploration, data cleaning, preprocessing, analytical questioning, and visualization, meaningful insights are extracted from a real-world entertainment dataset.

The project effectively showcases foundational data analytics skills and is suitable for portfolio presentation and entry-level data analytics roles.