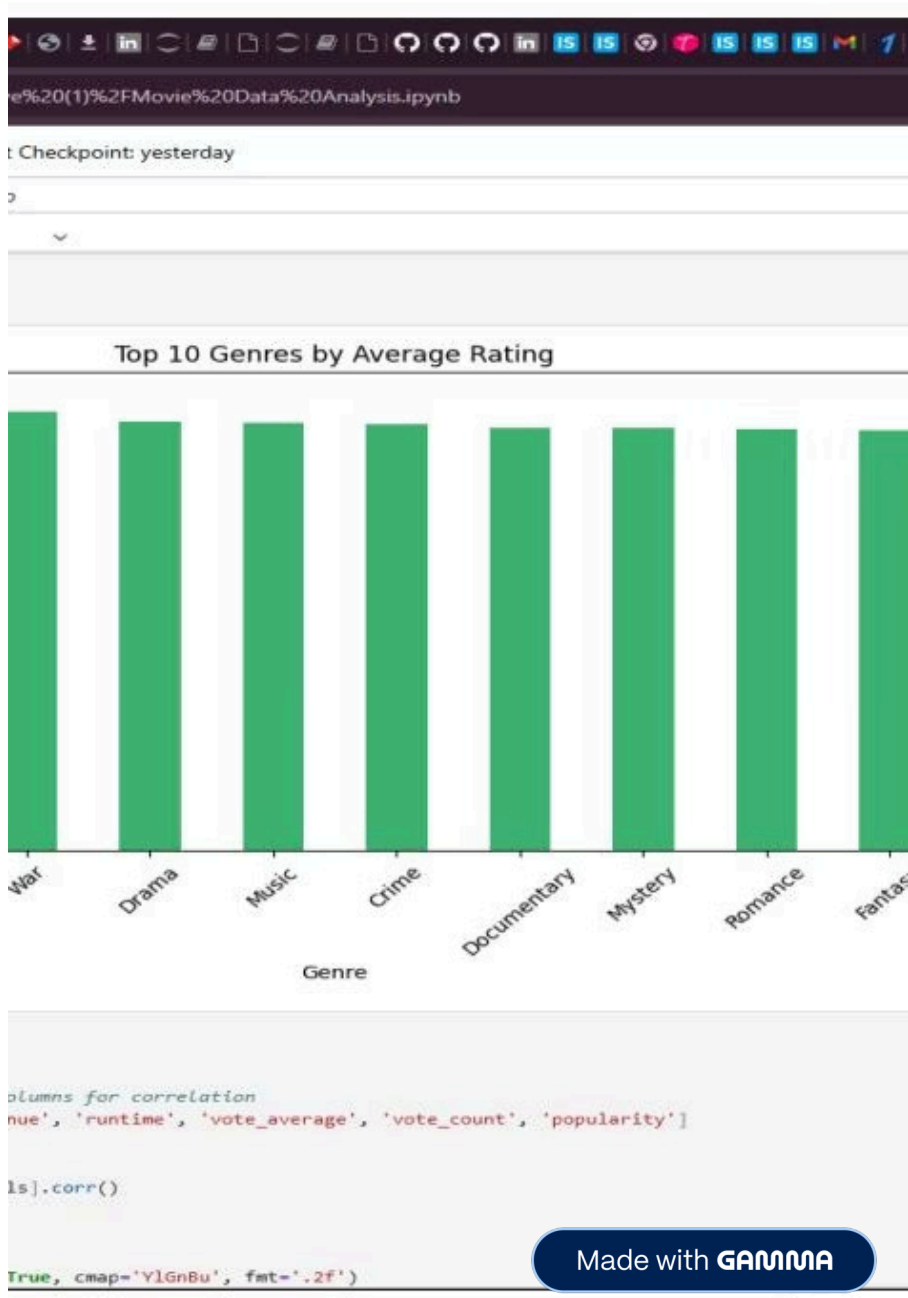
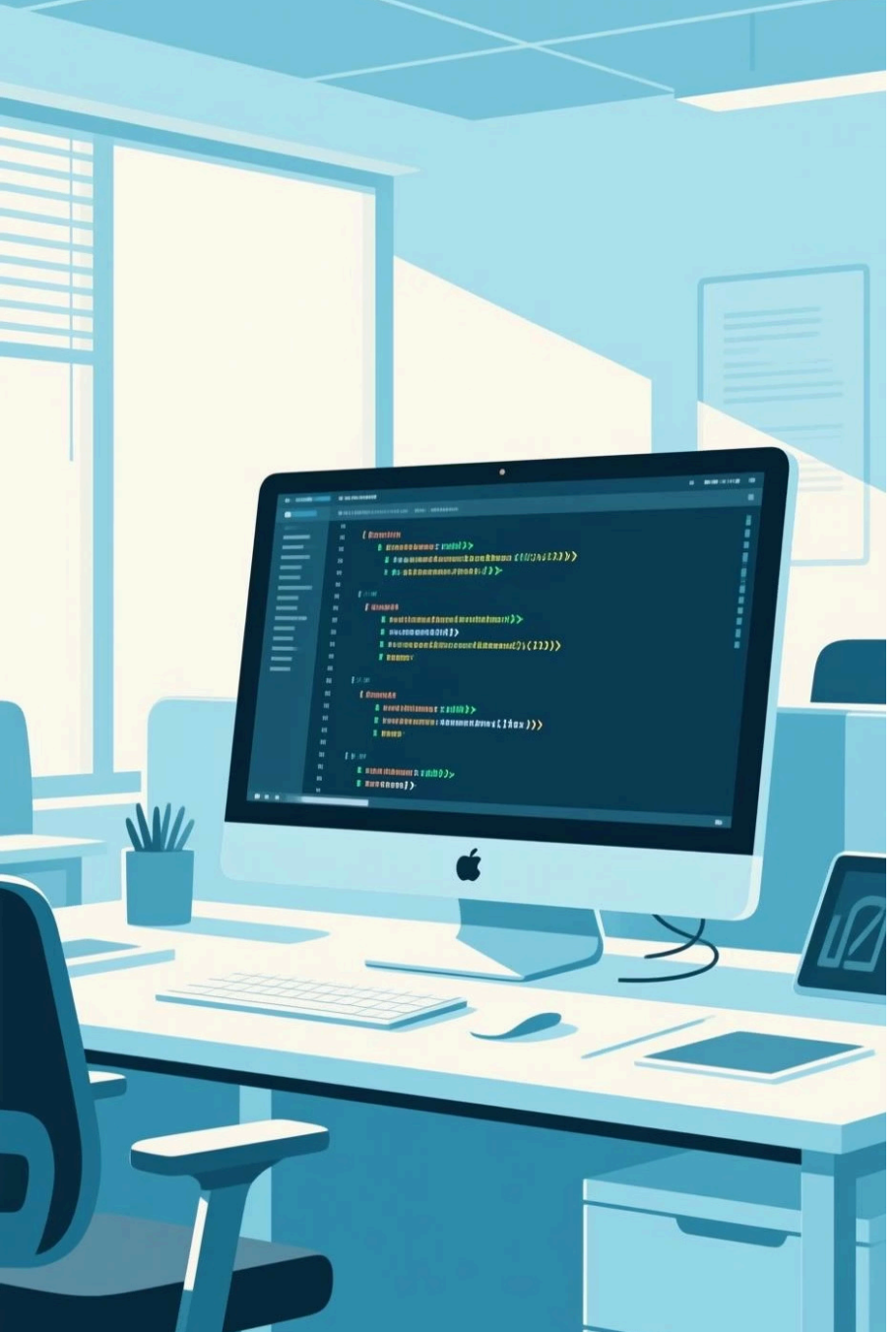
Movie Data Analysis using Python

Uncovering insights about revenue, ratings, and genres through exploratory data analysis





Project Overview

Objective

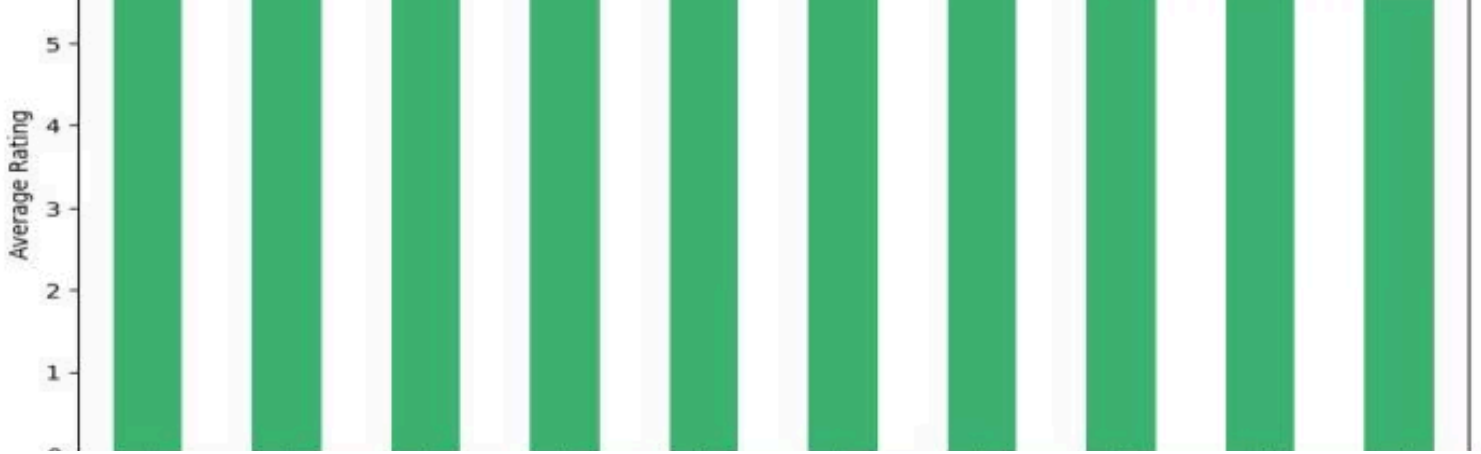
Perform exploratory data analysis on movie datasets to identify top-performing movies and trends

Tools Used

Python libraries including Pandas, NumPy, Matplotlib, and Seaborn for data manipulation and visualization

Approach

Data cleaning, statistical analysis, and visual storytelling to reveal patterns in movie performance

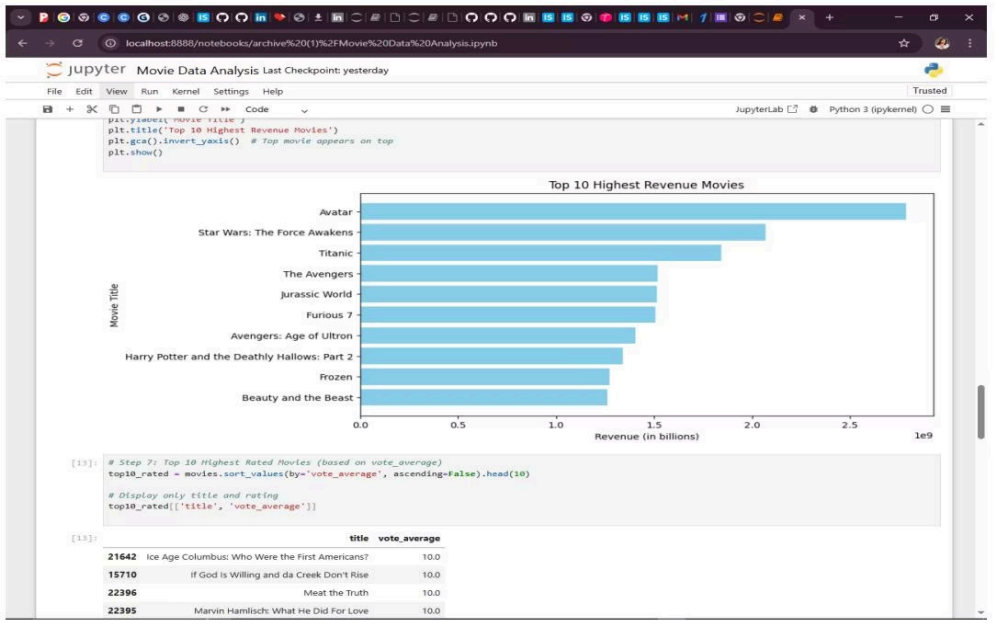


Top 10 Highest Revenue Movies

This chart highlights the top-grossing movies such as **Avatar**, **Titanic**, and **The Avengers**, showing their total revenue performance and dominance in the box office.

Avatar, Titanic, and The Avengers lead the pack

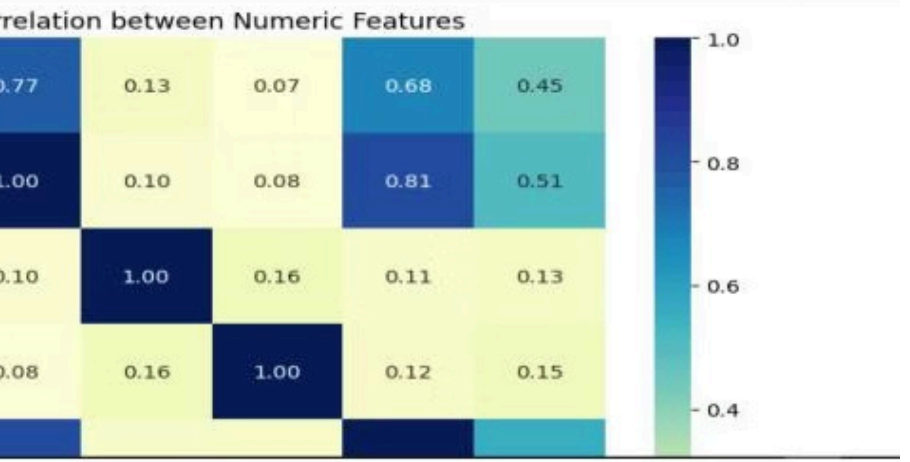
Top 10 Genres by Average Rating



Viewer Favorites

This bar chart displays the top genres based on viewer ratings, with Animation, History, and Drama performing the best among audiences.

These genres consistently deliver high-quality content that resonates with viewers.



Correlation between Numeric Features

Budget vs Revenue

Strong correlation (≈ 0.77) indicates higher investments often lead to higher earnings

Popularity Impact

Popularity and vote count show strong correlation with revenue performance

Key Insights

Budget Drives Revenue

Movies with higher budgets tend to generate more revenue, with a correlation of approximately 0.77

Popularity Matters

Popularity and vote count show strong correlation with revenue, indicating audience engagement is crucial

Genre Excellence

Animation, History, and Drama maintain the highest average ratings among all genres

Box Office Champions

Top movies such as Avatar, Titanic, and The Avengers dominate in total revenue

The Budget–Revenue Connection

0.77

Correlation Coefficient

Strong positive relationship between budget and revenue

3

Top Genres

Animation, History, and Drama lead in average ratings

10

Movies Analyzed

Top revenue-generating films identified and visualized

Technical Approach

01

Data Cleaning

Prepared and cleaned movie datasets for accurate analysis

02

Visualization

Created compelling charts using Matplotlib and Seaborn

03

Analytical Storytelling

Transformed data into actionable insights and narratives

Pandas

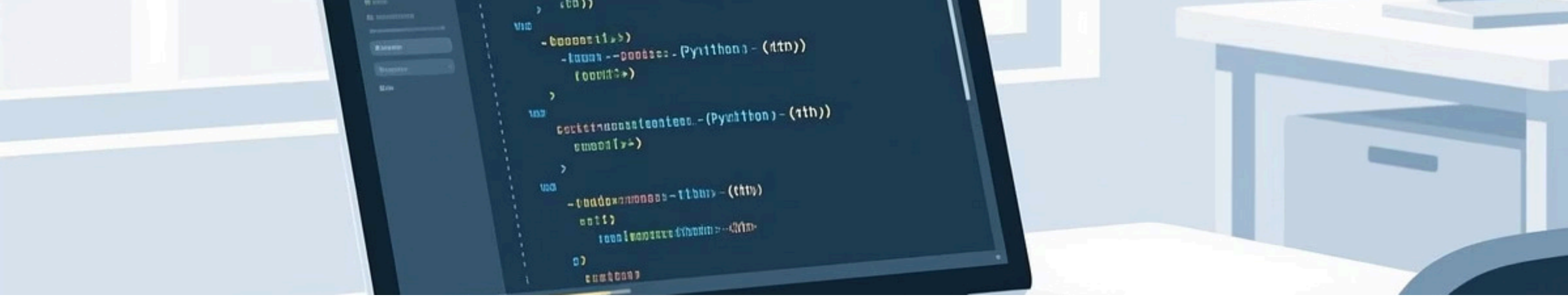
Data manipulation and analysis

NumPy

Numerical computing and statistics

Matplotlib & Seaborn

Data visualization and plotting



Project Showcase

Created by

Aanchal Pandey

Aspiring Data Analyst

Purpose

Portfolio and LinkedIn showcase demonstrating Python data analysis skills

Skills Demonstrated

Data cleaning, visualization, statistical analysis, and storytelling