Movie Revenue Data Analysis

Project Overview

This project aims to analyze the factors influencing the revenue and profitability of movies. By examining various attributes such as genre, lead studio, audience scores, and critical reception, the project seeks to uncover insights into what drives movie success. The analysis includes visualizations, statistical tests, and predictive modeling to provide a comprehensive understanding of movie revenue dynamics.

Dataset Description

The dataset used for this project contains the following fields:

Film: The title of the movie.
Genre: The genre of the movie (e.g., Action, Comedy, Drama).
Lead Studio: The studio that produced the movie.
Audience-score: The audience score for the movie.
% Profitability: The profitability percentage of the movie.
Rotten Tomatoes %: The percentage score from Rotten Tomatoes, representing critics' reviews.
Worldwide Gross: The worldwide gross revenue of the movie.
Year: The year the movie was released.

Project Structure

The repository contains the following files and directories:

data
— `movie_revenue_data.csv`: The dataset used for analysis.
— notebooks
— `data_cleaning.ipynb`: Jupyter notebook for data cleaning and preprocessing.
— `eda.ipynb`: Jupyter notebook for exploratory data analysis.
— `modeling.ipynb`: Jupyter notebook for predictive modeling.
— scripts/
— `data_cleaning.py`: Script for data cleaning.
— `eda.py`: Script for exploratory data analysis.
— `modeling.py`: Script for predictive modeling.
— **README.md:** Project overview and description (this file).
— **requirements.txt:** List of Python packages required to run the project.

Installation

To run this project, you need to have Python installed on your system. You can install the necessary packages by running:

bash
pip install -r requirements.txt
```

Usage

1. Data Cleaning:
   — Run the `data_cleaning.ipynb` notebook or execute the `data_cleaning.py` script to clean and preprocess the data.

2. Exploratory Data Analysis (EDA):
   — Open and run the `eda.ipynb` notebook to visualize and explore the dataset.
   — Alternatively, you can run the `eda.py` script to generate visualizations.

3. Predictive Modeling:
   — Use the `modeling.ipynb` notebook to build and evaluate predictive models.
   — The `modeling.py` script can also be used for this purpose.

**Key Insights**

Some key insights derived from the analysis include:

— The impact of different genres on movie profitability and revenue.
— The correlation between audience scores, Rotten Tomatoes scores, and movie revenue.
— The role of lead studios in the financial success of movies.
— Trends in movie profitability over the years.

**Contributing**

Contributions to this project are welcome. If you have any suggestions or improvements, please create an issue or submit a pull request.

**License**

This project is licensed under the MIT License. See the `LICENSE` file for more details.

**Contact**

If you have any questions or comments about the project, feel free to reach out.
Feel free to modify this README file according to your specific needs and details of your project. This template provides a comprehensive structure to present your data analysis project on GitHub effectively.
```