Movie Data Exploratory Data Analysis

A data science project that centers on extensive explorative data analysis of the IMDB ratings dataset.

About

This project begins with the creation of visualizations and calculations of descriptive statistics. The project seeks to identify whether a correlation exists between movie run time and movie performance scores. The work features a probability mass function (PMF) and a cumulative distribution function (CDF), along with test statistics. The project culminates with a regression analysis.

Link

Dataset: https://www.kaggle.com/datasets/omarhanyy/imdb-top-1000

Setup

This Python coding project was constructed on Jupyter Notebook. Please import the following libraries:

- Matplotlib
- Numpy
- Pandas
- Statsmodels
- Thinkstats2
- Thinkplot