

Project Proposal- Web Scraping & Linear Regression

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?

Predict movie gross income based on the movie characteristics by linear regression model

- Who benefits from exploring this question or building this model/system?

Movie producers & investor, they will know what kind of movie catches people's eye the most which can have higher ratings and generate good profit.

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?

I will use data from Box Office Mojo & IMDB, web scraping

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

Observations: each movie

Features: 'title', 'genre', 'rank', 'lifetime gross', 'MPAA Rating', 'year', 'runtime minutes', 'release date'...

- If modeling, what will you predict as your target?

Target: 'domestic total gross'

Tools:

- How do you intend to meet the tools requirement of the project?

Python Jupyter Notebook, BeautifulSoup package

- Are you planning in advance to need or use additional tools beyond those required?

MVP Goal:

- What would a [minimum viable product \(MVP\)](#) look like for this project?

Be able to predict domestic gross