Project proposal

A film investor want to gather information on what type of movie gain the most profit. This project will solve the problem with linear regression based on data from "Box Office Mojo". In the end, several insight trends will be provided as the foundamentals for future film invest reference.

Question/need:

- Question: What type of movie is the most valuable in the long term/short term market? Does COVID impact the profit pattern?

- Need: A filem investor would benifit from exploring this question for the purpose of planning a new investment in the current market.

Data Description:

- Dataset(s) plan to use: "Box Office Mojo" Movie data https://www.boxofficemojo.com/

- Individual sample/unit, characteristics/features: Use data for at least 100 movies. Metrics: Genres, Budget, Running Time, Global Gross.

- What to predict as a target: Select top 10 Genres, find out the trends between, Genres and Gross, Budget and Gross, Running Time and Gross.

Tools:

- Regression model

- Web scraping tools (BeautifulSoup, requests)

- SQL, python, pandas, matplotlib, seaborn

MVP Goal:

- Clarify the project question and goal.
- Describe the model/algorithm to use.
- Show one plot and a short conclusion: trends between Genres and gross revenue