Logo

Description automatically generated with low confidence

* Questions/Problem statement
  + List of Questions (IMDB) .
    - What is the most genre of movies that people prefer?
    - Who’s the director that has a highest rating of movies ?
    - Top 5 best movies of the 20’s?
    - Top 5 movies directed by the same person?
    - What is the movie that has a highest revenue ?
  + And other questions depend on:
    - Ratings
    - Specific genre
    - Time of release
    - And how many viewers rate this movie
* Data Description
  + Columns:
    - Title (text)
    - Actors (text)
    - Year (int)
    - Runtime in minutes (int)
    - Rating (float)
    - Votes (int)
    - Revenue (float)
    - Metascore (int)
  + Rows:
    - 1001 rows and 8 columns.
* Tools
  + Programs: Spyder, GoogleColab, Pycharm , MS Word, MS PowerPoint, GitHub
  + Libraries: Statsmodels, BeautifulSoup, lxml, html5lib, JSON, Patsy, Sklearn, Matplotlib, Seaborn, Pandas, NumPy.
  + Functions: fit(), score(), predict(), intercept(), Prettify, dump(), Pairplot(), OLS(), .head(), read\_csv(), find(), find\_all(), mean() groupby(), sum().
  + Plot: Heatmap, pairplot, Bar graph, Histogram, Scatter plot, Area plot, Pie plot.
* MVP Goal
* Eventually, represent the outputs of the data as plots
* Answering the questions that we’ve selected before
* Increase the accuracy of the data (Model).
* Reduce the amount of errors and the complexity of the data (Model).