

SI 206 Final Project Plan

- a. What is your group's name?
API Explorers
- b. Who are the people in the group (first name, last name, umich email)?
Lia Du yutongdu@umich.edu
Eric Dorman krytt@umich.edu
Ismail Hazime ith@umich.edu
- c. What APIs/websites will you be gathering data from? The base URLs for the APIs/websites must be different for them to count as different APIs.
Tracking and analyzing trends in movie popularity using the following APIs:
 - IMDbOT: <https://github.com/TelegramPlayGround/Free-Movie-Series-DB-API>
 - Open Movie Database: <http://www.omdbapi.com/>
 - Simkl: <https://simkl.docs.apiary.io/#>
- d. What data will you collect from each API/website and store in a database? Be specific.
 - IMDbOT: Movie title, IMDb rating (numerical score out of 10), genres, year of release, director, language
 - Open Movie Database: movie title, main cast, runtime (duration in minutes), box office revenue (if available), award, Metascore (rating from Metacritic)
 - Simkl: Trending movie titles, user popularity score, viewing statistics, movie category
- e. What data will you be calculating from the data in the database? Be specific.
 1. Determine if there is a correlation between IMDb ratings and box office revenue
 2. Determine top genres based on average IMDb ratings and user popularity scores
 3. Determine if there is a relationship between runtime and genre → do certain genres tend to be longer or shorter?
- f. What visualization package will you be using (Matplotlib, Plotly, Seaborn, etc)?
 - Matplotlib, Seaborn, and Plotly
- g. What graphs/charts will you be creating?
 - Bar chart: average IMDb ratings for each movie genre
 - Scatter plot: top genres based on average IMDb ratings and user popularity scores
 - Box plot: distribution of runtimes across genres
 - Pie chart: genre distribution of trending movies
- h. Who is responsible for what? Please note that all team members should do an equal amount of programming and total work.

- As of now, we have distributed the work in the following way — this is subject to change
 1. Lia: Research and integrate Simkl API, extract data such as trending movie titles, user popularity scores, and viewing statistics, handle goal #2 (determine top genres based on average IMDb ratings and user popularity scores), using Plotly to create interactive graphs
 2. Eric: Research and integrate IMDbOT API, extract data such as movie titles, IMDb ratings, genres, release years, directors, and languages, create database tables, handle goal #1 (determine if there is a correlation between IMDb ratings and box office revenue)
 3. Ismail: Research and integrate the OMDb API for data collection, extract and clean data such as runtime, box office revenue, awards, and plot summaries, calculating correlation between IMDb ratings and box office revenue, using Seaborn to create visualizations, handle goal #3 (determine if there is a relationship between runtime and genre)
- We have assigned each member with an API to research and integrate and a goal from question E. However, shared responsibilities will include testing, debugging, and documenting processes when creating functions along the way.