1. **What data was used to enrich the client’s data?**

In this project, we used the movie rank of movies released between 2018 and 2020 from the IMDb official webpage to enrich our client’s dataset. The movies on the web page have been sorted by votes in descending order, so we just pulled movie\_id, rank, votes, title, originalTitle, year, rating, titleType, isAdult, runtime, and genres for the top 500 movies from the web page and merged them with our client’s data.

1. **Describe the data cleaning and transformation that was implemented.**

In part A of the project, first we pulled the information we needed from the IMDb web page. We changed the data type of movie\_year, movie\_rank, movie\_runtime, and movie\_votes to integer, then we used this dataset to enrich our client’s dataset. We used “left merge” to merge df2(IMDB data) to df1(client’s data) to ensure all 500 rows are included. The one problem we met is since there are some null values after merging, we cannot change the data type of some columns back to integer. And since we did not think there’s a right way to fill in those null values, we just left them as the original data type. Then we rearranged the columns in our new dataset and exported it as a csv file.