Technical Report

1. Extraction
   * We used the following CSV and Website to source our data
     1. CSV From Kaggle - <https://www.kaggle.com/axeltorbenson/top-4000-movies>
     2. Scraped IMDb Top 250 Movies - <https://www.imdb.com/chart/top/>
   * The fields of interest were:
     1. Movie Title
     2. Ranking
     3. Year
     4. Rating
2. Transforming
   * Scrapping IMDb top 250 movies
     1. Used Beautiful Soup, splinter, webdriver\_manager to conduct scraping of IMDB website
     2. Used comprehensive list to gather movie titles, year, and rating, ranking
     3. Loaded scrapped data into pandas dataframe
   * Cleaning the CSV
     1. Used Pandas to load CSV
     2. Dropped rows with null values
     3. Dropped rows with 0s
   * Merged dataframes
     1. Merged imdb with Top\_4k on “Movie Title”
3. Loading
   * Connected to PostgreSQL via sqlalchemy
     1. Decided to use a SQL database instead of Mongodb because all of our data was structured data
   * Checked for existing database and created new if did not exist
   * Loaded our 3 tables into PostgreSQL
     1. Imdb
     2. top\_4k
     3. merged\_df
   * Confirmed all data was passed as intended to PostgreSQL