

**Oscar Winning Movies and Rotten Tomato Ratings**

**ETL Project**

**Technical Report**

Group Members: Melissa Agruda, Zachary Steele, Iretemi Adelaja



## **Data Sources Extracted**

1. <https://www.kaggle.com/tmdb/tmdb-movie-metadata#tmdb_5000_movies.csv> (movies.csv file)
2. <https://data.world/crowdflower/academy-awards-demographics>. (oscar.csv file)
3. <https://www.rottentomatoes.com/browse/opening/>. Web scrapped Rotten tomatoes website in order to find movie titles as well as movie ratings. Approximately 2,500 movie titles were scrapped and loaded into a data frame using Pandas.

## **Data Transformation**

Cleaning, Filtering & Joining

1. Renamed 4 columns to **movie, award, birthplace and date of birth.** Set index to **movies**. Dropped duplicates from **movies** (Oscar.csv)
2. Deleted excess information and left required information such as **homepage, release\_date, title** and **popularity.** Set index to **title**. Dropped duplicates from **title** (movie.csv)
3. Used splinter and beautiful soup to web scrap rotten tomatoes website for the movie **title** and **ratings.** Transformed the data frame into a CSV file.

## **Data Loading**

1. Loaded **movies.csv, Oscar.csv** and **movie.df** into PgAdmin using the .to\_sql command. Created and joined table in SQL database
2. Joined to compare movies in common and Oscar winner!