

## Contents ↻ ⚙

- 1 Business Case and Goals
- ▼ 2 Library and Data Imports
  - 2.1 IMDB Data Set Imports
  - 2.2 Rotten Tomatoes Data Set Imports
- ▼ 3 IMDB Data Merging and Cleaning
  - 3.1 Finding Top 5 Genres
  - ▼ 3.2 Merging and Subsetting Data for Analysis
    - 3.2.1 Creating Subset of main data to clean
    - 3.2.2 Checking if crew member is deceased or likely retired and handling NA values
    - 3.2.3 Cleaning Runtime Minutes
- ▼ 4 IMDB Analysis and Visualization
  - 4.1 Creating functions to extract data
  - 4.2 Creating dataframes for each category in each genre
  - 4.3 Biography Visualization
  - 4.4 History Visualizations
  - 4.5 Music Visualizations
  - 4.6 Sports Visualizations
  - 4.7 Drama Visualizations
  - ▼ 4.8 Run Time Visualizations
    - 4.8.1 Creating functions and dataframes to plot
    - 4.8.2 Plotting
  - 4.9 **\*Findings**
- ▼ 5 Rotten Tomatoes Data Merging and Cleaning
  - 5.1 Merging
  - ▼ 5.2 Cleaning ratings\_x containing the rating values I want
    - 5.2.1 Building a function to clean ratings
    - 5.2.2 Handling Letter Ratings
  - ▼ 5.3 Cleaning Reviews
    - 5.3.1 Importing stopwords from Natural Language Toolkit and removing them from reviews
    - 5.3.2 Create new column with a dataframe containing words and their rating
    - ▼ 5.3.3 Creating new dataframe with all words and ratings by appending each dataframe in the new column
      - 5.3.3.1 **\*\* This cell takes long to run**
    - 5.3.4 Grouping data frame values, sorting and grabbing top/least 25 words
  - 5.4 Positive Words Visualization
  - 5.5 Negative Words Visualization
  - 5.6 **\*Findings**
- ▼ 6 Summary
  - 6.1 Insight 1 - Cast and Crew Recommendations:
  - 6.2 Insight 2 - Runtime Recommendations:
  - 6.3 Insight 3 - Review Words To Aim For or Avoid: