| Script Name | Purpose | Input Files | Output Files | Library |
| --- | --- | --- | --- | --- |
| Capstone 011 critic names list.R | * List of Critic Name URLs * Number of review pages for each critic * Dataset of critic review page URLs (>12,000) * Fix non-english characters * Only critics with more than 25 reviews |  | critic\_urls\_ds.csv  critics\_ds\_corrected\_for\_enlish.csv critics\_urls\_ds\_corrected\_for\_enlish.csv  critic\_review\_urls.csv page\_count\_vector.csv | rvest  stringr readr |
| ****CapStone 03 Tomato Meter Score.R**** | * Dataset of every critics review for every movie for every year = 600,000 records | critics\_urls\_pages.csv  critics\_urls\_ds\_corrected\_english.csv | list\_of\_review1.csv list\_of\_review2.csv list\_of\_review3.csv list\_of\_review4.csv list\_of\_review5.csv list\_of\_review6.csv list\_of\_review7.csv unique\_reviews.RDS | rvest  readr |
| CapStone 05 cleaning unique\_reviews.R | * Cleanup data to create movie year when year appears twice in review title (translated movie titles) | unique\_reviews.RDS | clean\_tMeter\_score\_ds.RDS |  |
| Capstone 06a IMDB merge | * review score # * merge IMDB dataset & reviews * spread out genre to 21 columns | IMDB meta data 5000 movies.csv  tMeter96plus.RDS | merged\_df.RDS  newdata.RDS | stringr plyr |
| ****Capstone 07a CleanUp\_lettersgradecores.R**** | * clean up newdata * convert all mistyped letter grades into the same format * convert all 4/5 "stars" codes to just 4/5 * deduce and place in missing denominators where possible | newdata.RDS | newdata.RDS |  |
| ****Capstone 07b collate Letter grade scores.R**** | * Only data with letter grades * Factor letter grades * Plot out letter grade distributions | newdata.RDS | only\_letter\_grades.RDS | dplyr stringr |
| ****Capstone 07c MedianScores\_for\_LetterGrades.R**** | * Aggregate medians of scores of movies with Letters A through F scores * Lookup Table | movies\_after\_2006.RDS | AtoF\_convert.RDS  Only\_letter\_grades.RDS |  |
| **Capstone 08a Merge IMDB Oscars and regression analysis.R** | * Merge actors and director to dataset * Summarize career totals * Add Fresh rating | AcademyAwards1996\_2017ActorDirector.csv | tMeter\_Oscars\_IMDB.RDS |  |
| **Capstone 08b Oscars\_removing\_data\_2015\_16.R** | * Split data set into training and test * Associate cast with nominations by career nominations split between training and test | tMeter\_IMDB\_Oscars.R | RT.Train.RDS RT.Test.RDS | plyr |
| **Capstone 09b collate Genres.R** | * which genres are the most common * which genres cluster together * model linear regression * model logistic regression | RT.Train.RDS  RT.Test.RDS | RT.train\_clustered |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |