

Index of Scripts and Files

Script Name	Purpose	Input Files	Output Files	Library
CapStone 011 critic names list.R	<ul style="list-style-type: none"> 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_english.csv critics_urls_ds_corrected_for_english.csv critic_review_urls.csv page_count_vector.csv	rvest stringr readr
CapStone 03 Tomato Meter Score.R	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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_lettersgradescores.R	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Only data with letter grades Factor letter grades Plot out letter grade distributions 	newdata.RDS	only_letter_grades.RDS	dplyr stringr

Index of Scripts and Files

Script Name	Purpose	Input Files	Output Files	Library
Capstone 07c MedianScores_for_LetterGrades.R	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Split data set into training and test Associate cast with nominations by career nominations split between training and test 	tMeter_IMDB_Oscars.RDS	RT.Train.RDS RT.Test.RDS	plyr
Capstone 09b collate Genres.R	<ul style="list-style-type: none"> 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.RDS	NbClust dplyr
Capstone 09c Fix Duplicate Movies.R	<ul style="list-style-type: none"> Clean up dataset when duplicate movie titles discovered 	tMeter_IMDB_Oscars.RDS	tMeter_IMDB_Oscars.RDS movies.RDS	
Capstone 10a collate Genres NbClust.R	<ul style="list-style-type: none"> Cluster analysis by genre Confusion table ROC curve 	movies.RDS		NbClust