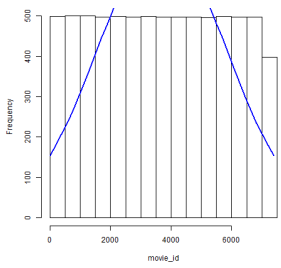
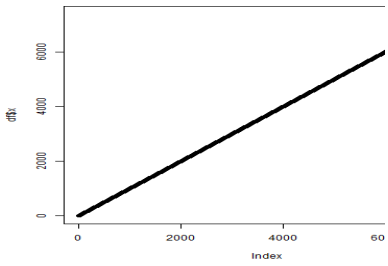
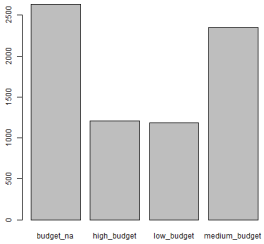
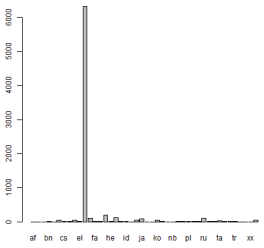
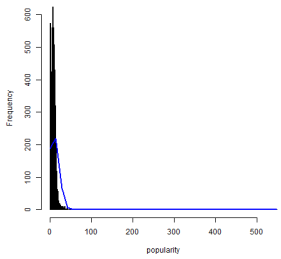
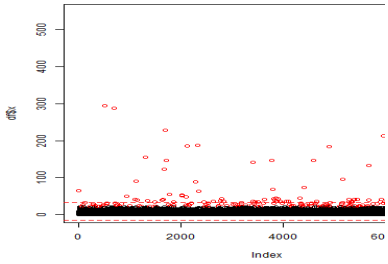
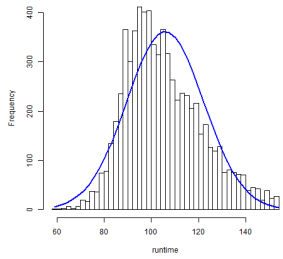
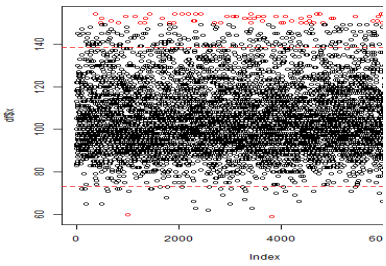
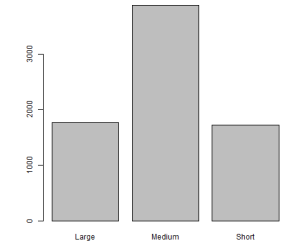
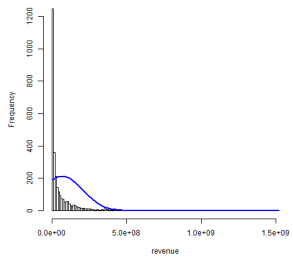
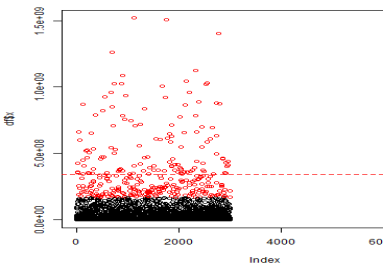
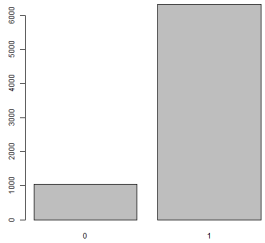
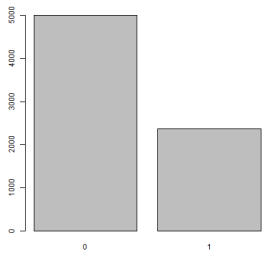
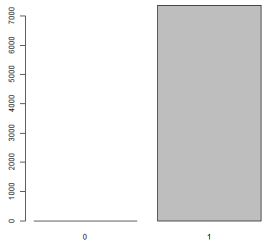
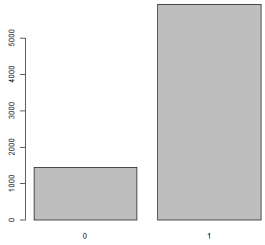
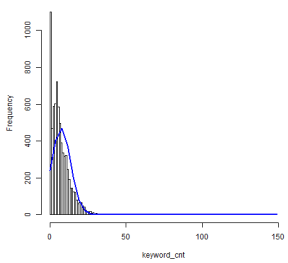
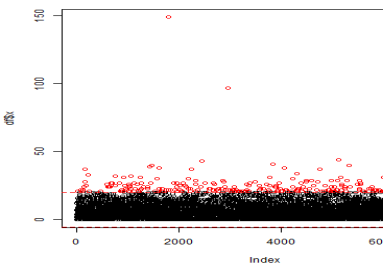
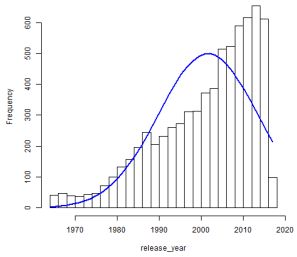
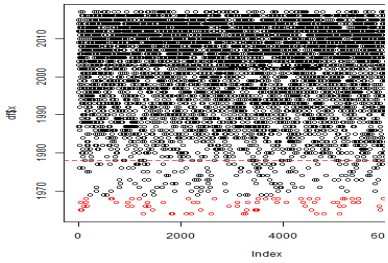
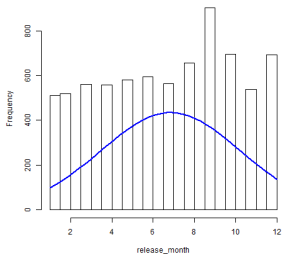
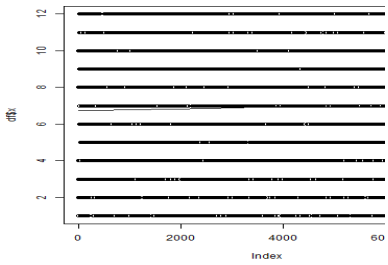
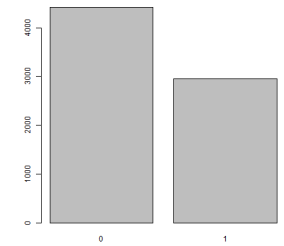
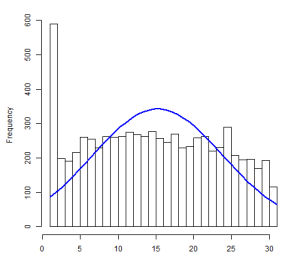
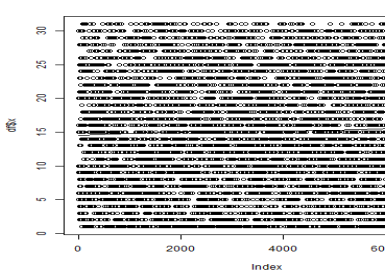
Data Visualization & Exploration

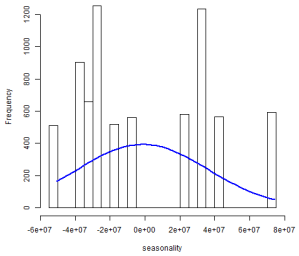
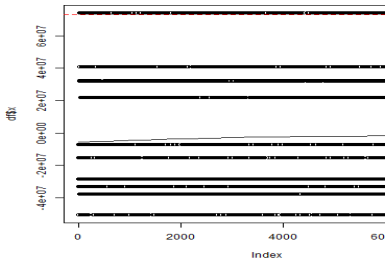
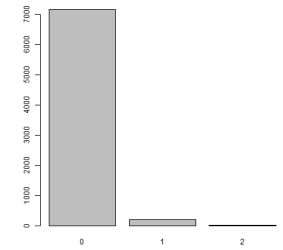
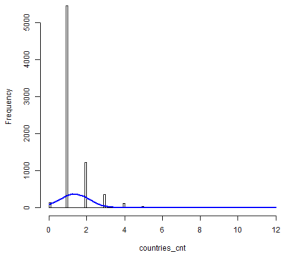
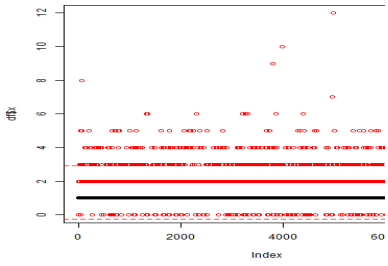
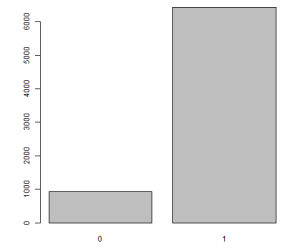
Show as Cards

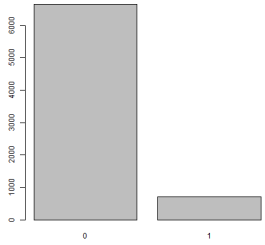
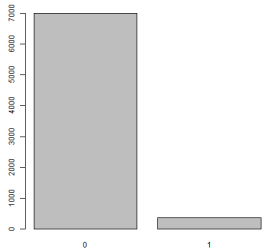
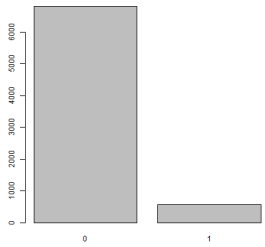
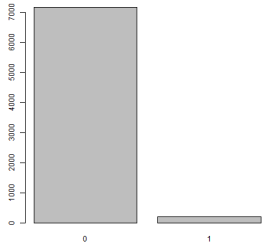
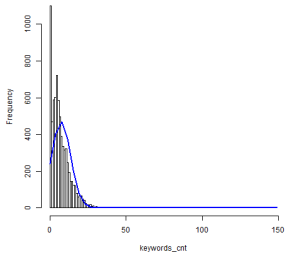
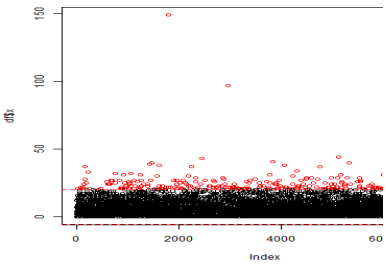
Variable	Distribution	Descriptive Statistics	Outliers
movie_id		Data type: Continuous Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Mean: 3698 StdDev: 2137 Median: 3696 IQR: 1848 - 5550 Min: 1 Max: 7398	 Outlier values: No outlier values found
budget		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 4 • budget_na : 2636 • high_budget : 1207 • low_budget : 1185 • medium_budget : 2347	
original_language		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 44	
popularity		Data type: Continuous Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Mean: 8.54 StdDev: 12.18 Median: 7.456 IQR:	

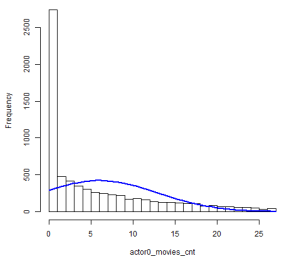
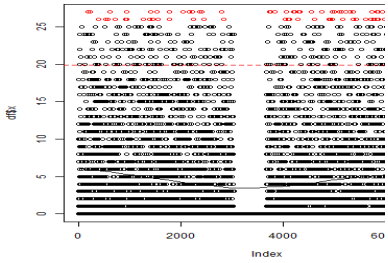
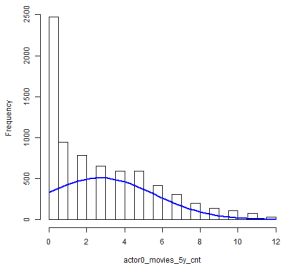
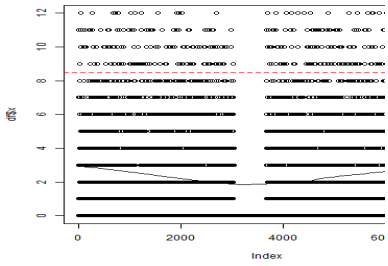
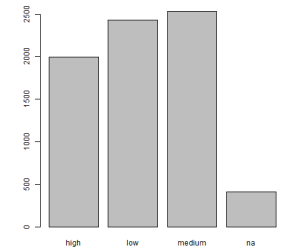
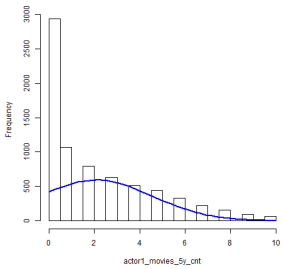
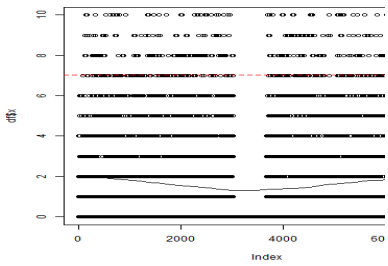
		3.97 - 10.93 Min: 1e-06 Max: 547.5	Outlier values: There are 201 outlier values
release_date		Data type: Date Data length: 7374 / 7375 (99.99 %) Missing: 1 (0.01356 %) Min date: 1918-08-01 Max date: 2017-08-11	
runtime		Data type: Continuous Data length: 7095 / 7375 (96.2 %) Missing: 280 (3.797 %) Mean: 105.9 StdDev: 16.27 Median: 103 IQR: 94 - 116 Min: 59 Max: 154	 Outlier values: 154, 150, 152, 151, 150, 153, 6 150, 150, 152, 154, 152, 150, 1 154, 154, 151, 151, 154, 154, 1 150, 153, 153, 153, 152, 152, 1 152, 153, 151, 153, 59, 153, 15 153, 151, 152, 150, 150, 151, 1 154, 154, 150, 152, 152, 153, 1 153, 154, 150, 154, 152, 150, 1 150, 151, 152, 150, 151, 150, 1 153
runtime_cat		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • Large : 1769 • Medium : 3884 • Short : 1722	
revenue		Data type: Continuous Data length: 2993 / 7375 (40.58 %) Missing: 4382 (59.42 %) Mean: 6.688e+07 StdDev: 1.377e+08 Median: 16928137 IQR: 2.405e+06 - 6.906e+07 Min: 1 Max: 1519557910	 Outlier values: There are 315 outlier values

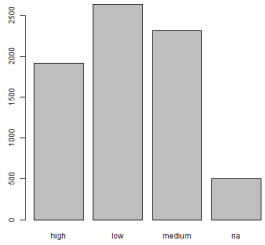
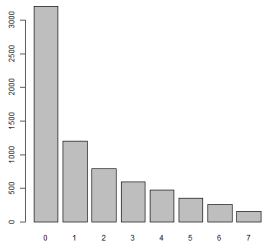
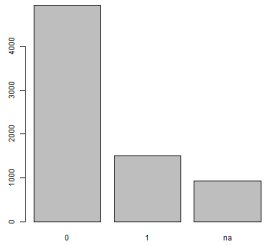
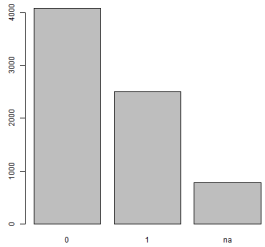
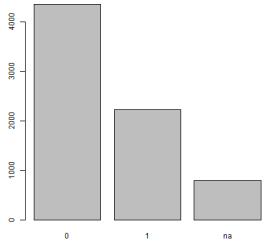
sw_lang_en		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 1039 • 1 : 6336	
sw_web_presence		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 5012 • 1 : 2363	
sw_has_poster		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 2 • 1 : 7373	
sw_tagline		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 1442 • 1 : 5933	
keyword_cnt		Data type: Continuous Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Mean: 7.265 StdDev: 6.252 Median: 6 IQR: 3 - 10 Min: 0 Max: 149	 Outlier values: There are 25 outlier values
release_year		Data type: Continuous	

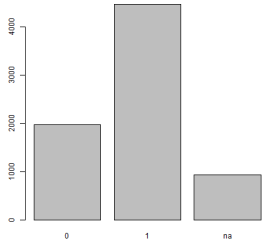
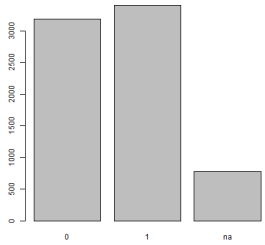
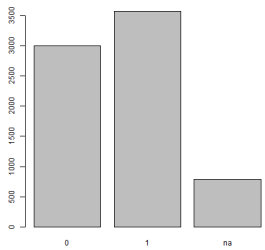
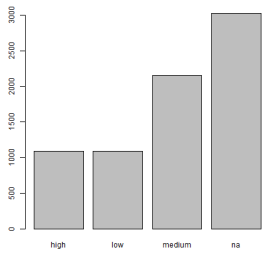
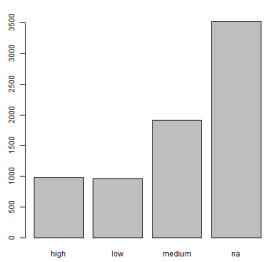
		Data length: 7105 / 7375 (96.34 %) Missing: 270 (3.661 %) Mean: 2002 StdDev: 11.79 Median: 2005 IQR: 1994 - 2011 Min: 1964 Max: 2017	 Outlier values: 1965, 1967, 1965, 1966, 1968, 1966, 1968, 1968, 1967, 1967, 1968, 1966, 1968, 1964, 1966, 1966, 1965, 1967, 1964, 1964, 1968, 1964, 1967, 1967, 1965, 1968, 1966, 1967, 1966, 1964, 1965, 1967, 1965, 1965, 1967, 1968, 1968, 1968, 1967, 1968, 1968, 1967, 1966, 1968, 1965, 1964, 1968, 1964, 1967, 1967, 1966, 1968, 1966, 1968, 1968, 1967, 1968, 1966, 1965, 1968, 1966, 1968, 1966, 1968, 1967, 1967, 1964, 1967, 1965, 1967, 1964, 1964, 1968, 1965
release_month		Data type: Continuous Data length: 7374 / 7375 (99.99 %) Missing: 1 (0.01356 %) Mean: 6.844 StdDev: 3.386 Median: 7 IQR: 4 - 10 Min: 1 Max: 12	 Outlier values: No outlier values found
high_release_month		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 4427 • 1 : 2948	
release_day		Data type: Continuous Data length: 7374 / 7375 (99.99 %) Missing: 1 (0.01356 %) Mean: 15.25 StdDev: 8.594 Median: 15 IQR: 8 -	 Outlier values: No outlier values found

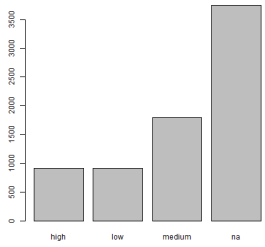
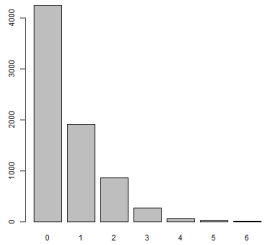
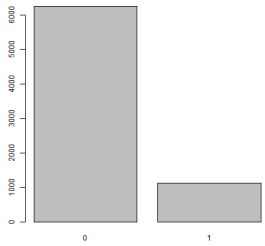
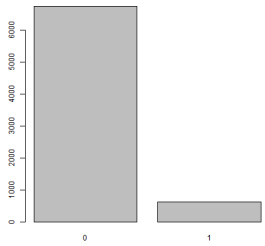
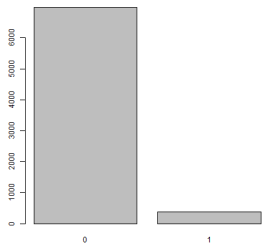
		22 Min: 1 Max: 31	
seasonality		Data type: Continuous Data length: 7374 / 7375 (99.99 %) Missing: 1 (0.01356 %) Mean: -1.337e+06 StdDev: 3.728e+07 Median: -1.531e+07 IQR: -3.327e+07 - 3.232e+07 Min: -50746591 Max: 74205073	 Outlier values: No outlier values found
sw_collection		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • 0 : 7169 • 1 : 195 • 2 : 11	
countries_cnt		Data type: Continuous Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Mean: 1.335 StdDev: 0.7899 Median: 1 IQR: 1 - 1 Min: 0 Max: 12	 Outlier values: There are 11 outlier values
lang_US		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 933 • 1 : 6442	
lang_FR		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %)	

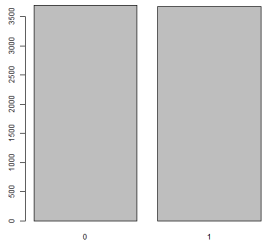
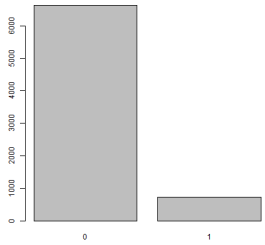
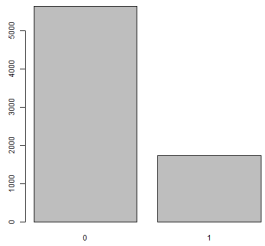
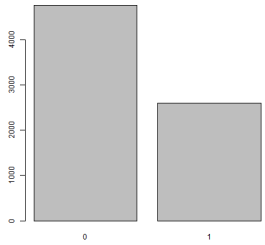
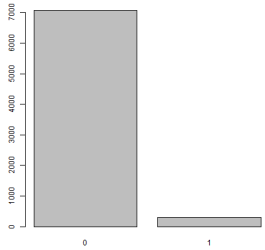
		Number of levels: 2 • 0 : 6664 • 1 : 711	
lang_RU		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7017 • 1 : 358	
lang_ES		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6816 • 1 : 559	
lang_JA		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7171 • 1 : 204	
keywords_cnt		Data type: Continuous Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Mean: 7.265 StdDev: 6.252 Median: 6 IQR: 3 - 10 Min: 0 Max: 149	 Outlier values: There are 25 outlier values
actor0_movies_cnt		Data type: Continuous Data length: 7069 / 7375 (95.85 %)	

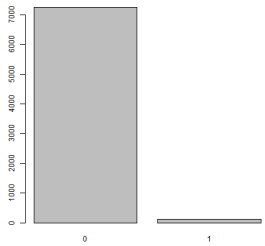
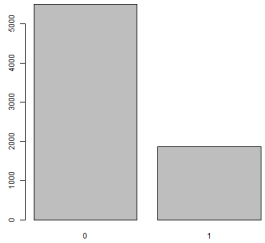
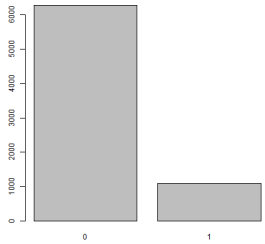
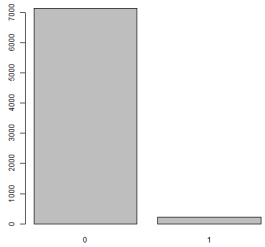
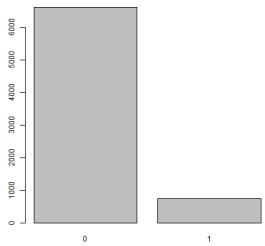
		Missing: 306 (4.149 %) Mean: 6.008 StdDev: 6.946 Median: 3 IQR: 0 - 10 Min: 0 Max: 27	 Outlier values: 27, 27, 26, 26, 26, 27, 26, 26, 2 26, 26, 27, 26, 27, 26, 26, 27, 2 27, 27, 26, 27, 27, 26, 27, 26, 2 27, 26, 27, 26, 26, 27, 27, 26, 2 27, 26, 27, 26, 26, 27, 26, 27, 2 27, 26, 27, 27, 27, 26, 26, 26, 2 26, 26, 27, 26, 27, 27, 26, 26, 2 27, 26, 26, 26, 26, 26, 26, 27, 2
actor0_movies_5y_cnt		Data type: Continuous Data length: 7323 / 7375 (99.29 %) Missing: 52 (0.7051 %) Mean: 2.691 StdDev: 2.884 Median: 2 IQR: 0 - 5 Min: 0 Max: 12	 Outlier values: No outlier values found
actor1_movies_cnt		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 4 • high : 1998 • low : 2429 • medium : 2537 • na : 411	
actor1_movies_5y_cnt		Data type: Continuous Data length: 7243 / 7375 (98.21 %) Missing: 132 (1.79 %) Mean: 2.079 StdDev: 2.48 Median: 1 IQR: 0 - 4 Min: 0 Max: 10	 Outlier values: No outlier values found
actor2_movies_cnt		Data type: Categorical Data Data length: 7375 / 7375 (100 %)	

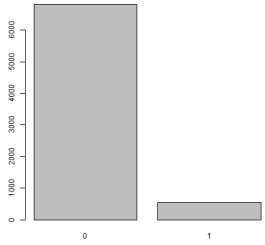
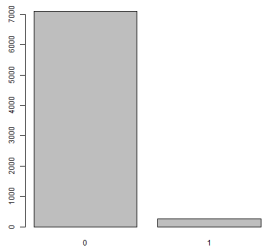
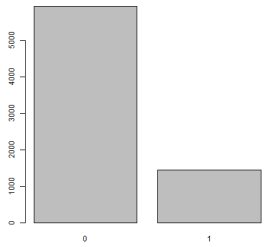
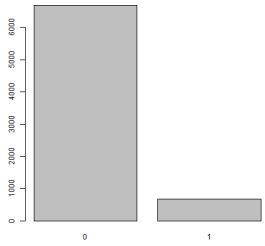
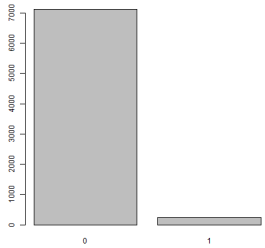
		Missing: 0 (0 %) Number of levels: 4 • high : 1916 • low : 2640 • medium : 2312 • na : 507	
actor2_movies_5y_cnt		Data type: Categorical Data Data length: 7030 / 7375 (95.32 %) Missing: 345 (4.678 %) Number of levels: 8	
sw_female_actor0		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • 0 : 4938 • 1 : 1503 • na : 934	
sw_female_actor1		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • 0 : 4093 • 1 : 2500 • na : 782	
sw_female_actor2		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • 0 : 4361 • 1 : 2226 • na : 788	
sw_male_actor0		Data type: Categorical Data	

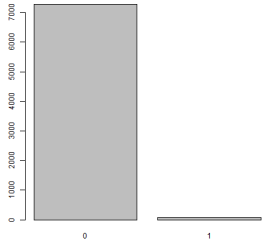
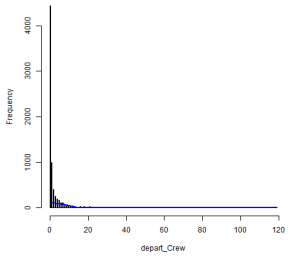
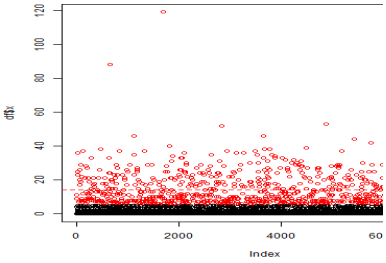
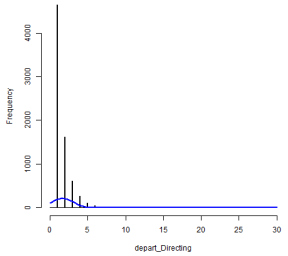
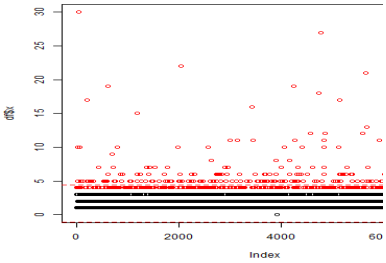
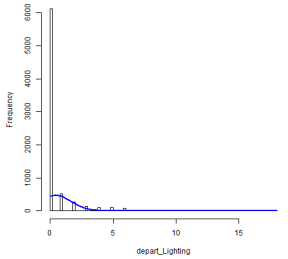
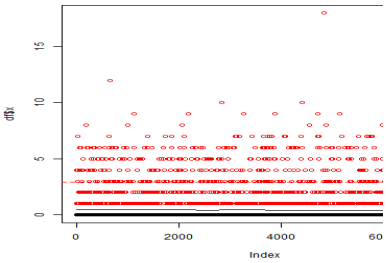
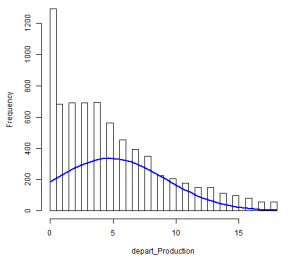
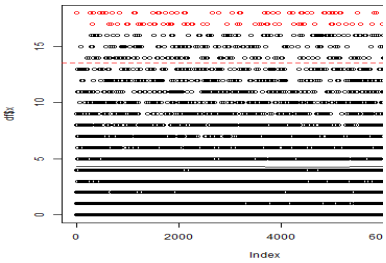
		Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • 0 : 1970 • 1 : 4471 • na : 934	
sw_male_actor1		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • 0 : 3184 • 1 : 3409 • na : 782	
sw_male_actor2		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 3 • 0 : 3006 • 1 : 3581 • na : 788	
actor0_prev_revenue		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 4 • high : 1091 • low : 1091 • medium : 2160 • na : 3033	
actor1_prev_revenue		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 4 • high : 976 • low : 964	

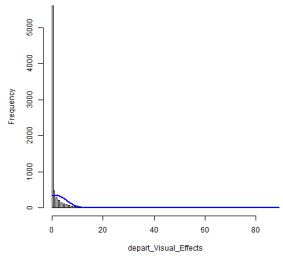
		• medium : 1908 • na : 3527	
actor2_prev_revenue		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 4 • high : 912 • low : 911 • medium : 1798 • na : 3754	
director_movies_5y_cnt		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 7	
genre_adventure		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6259 • 1 : 1116	
genre_fantasy		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6747 • 1 : 628	
genre_animation		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6993 • 1 : 382	

genre_drama		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 3699 • 1 : 3676	
genre_horror		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6640 • 1 : 735	
genre_action		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 5640 • 1 : 1735	
genre_comedy		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 4770 • 1 : 2605	
genre_history		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7080 • 1 : 295	
genre_western		Data type: Categorical	

		Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7258 • 1 : 117	
genre_thriller		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 5506 • 1 : 1869	
genre_crime		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6291 • 1 : 1084	
genre_documentary		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7154 • 1 : 221	
genre_science_fiction		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6631 • 1 : 744	
genre_mystery		Data type: Categorical Data	

		Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6825 • 1 : 550	
genre_music		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7108 • 1 : 267	
genre_romance		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 5940 • 1 : 1435	
genre_family		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 6700 • 1 : 675	
genre_war		Data type: Categorical Data Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7132 • 1 : 243	
genre_foreign		Data type: Categorical Data	

		Data length: 7375 / 7375 (100 %) Missing: 0 (0 %) Number of levels: 2 • 0 : 7291 • 1 : 84	
depart_Crew		Data type: Continuous Data length: 7341 / 7375 (99.54 %) Missing: 34 (0.461 %) Mean: 2.388 StdDev: 5.875 Median: 0 IQR: 0 - 2 Min: 0 Max: 119	 Outlier values: There are 44 outlier values
depart_Directing		Data type: Continuous Data length: 7341 / 7375 (99.54 %) Missing: 34 (0.461 %) Mean: 1.665 StdDev: 1.366 Median: 1 IQR: 1 - 2 Min: 0 Max: 30	 Outlier values: There are 19 outlier values
depart_Lighting		Data type: Continuous Data length: 7341 / 7375 (99.54 %) Missing: 34 (0.461 %) Mean: 0.4262 StdDev: 1.235 Median: 0 IQR: 0 - 0 Min: 0 Max: 18	 Outlier values: There are 12 outlier values
depart_Production		Data type: Continuous Data length: 7126 / 7375 (96.62 %) Missing: 249 (3.376 %) Mean: 4.785 StdDev: 4.371 Median: 4 IQR: 1 - 7 Min: 0 Max: 18	 Outlier values: 18, 18, 17, 18, 18, 18, 18, 17, 1 17, 18, 17, 18, 17, 17, 18, 18, 1 17, 17, 18, 17, 17, 18, 18, 17, 1 17, 17, 18, 18, 18, 17, 18, 18, 1 17, 17, 17, 17, 17, 18, 18, 17, 1 17, 18, 17, 18, 17, 17, 17, 18, 1 17, 18, 18, 17, 18, 17, 18, 18, 1

			18, 17, 18, 17, 18, 18, 17, 18, 1 17, 17, 18, 18, 17, 18, 18, 17, 1 17, 17, 18, 18, 18, 17, 18, 18, 1 17, 18, 17, 18, 18, 17, 18, 17, 1 18, 17
depart_Visual_Effects		Data type: Continuous Data length: 7341 / 7375 (99.54 %) Missing: 34 (0.461 %) Mean: 1.29 StdDev: 4.131 Median: 0 IQR: 0 - 0 Min: 0 Max: 89	