Taking Board Games Seriously

Andrew Charlton & Leon Liu
CIS3120 Group Project
December 20th, 2018

Abstract:

We selected a website that ranks board games by the website's Geek Members Rating, Average Ratings, and number of Voters that participated each game. Given the fact the ratings are neck to neck, such as the Geek Ratings are compared against by the 100th of decimal ratings. We want to test how reliable the ranking system by plotting the Geek Ratings against Avg Ratings and providing analysis with the founding standard deviation results. Conclusion shows us the Voter contribution is a main driving factor for reliability. Our PY Plot example also shows Avg Ratings as more lenient, giving higher ratings versus Geek Ratings's higher standard ratings. The statistics from overall Standard Deviation shows us, the most votes deserve more standing in their ranks meanwhile lesser votes are not as reliable because it has lesser data(voters) to work with.

Data:

We collected data from the website: "https://boardgamegeek.com/browse/boardgame"
Although there are 100 games on each page and 1035 pages total, we only selected the first page that shows top 100 to be our main focus. The main table we collected data contains ranking 1 to 100 as its first column, follow with the Board game's Name, Geek member Ratings, Average overall Ratings, and the Number of Voters participated per each game listed.

Programs:

- *Dic_of_games.py* – using *BEAUTIFULSOUP* from *bs4* and *REQUESTS*, we created 2 functions:
    - First function *dicOfGames* to scrape the top 100 board game Names, Geek Ratings, Avg Ratings, and Voters using lists and dictionary. The data scraped is placed into a dictionary and can be checked by entering a name of a game to retrieve the ratings and votes values.
    - Second function *ratingsList* conglomerates all ratings and voter counts to their proper orders.
    - We then use PICKLE to export data onto individual CSV files for dictionary, avg rating, geek ratings, and voters. Thereafter, the following programs will load the data as needed through the csv files created.

- *Statistics_by_columns* (imagine available under) – using *PANDA*, with *DATAFRAME*, we gather the MIN, MAX, MEDIAN, and STANDARD DEVIATIONS for each of Geek Ratings, Avg Ratings, and Voters

- *Statistics_overall* – using *PANDA*, with *DATAFRAME*, we collectively generated the standard deviation for each game, resulting in a complied list of true value of voter counts to their ratings. The main result that drives our conclusion.

- *Plot_charts* (image available under)– using *MATPLOTLIB* with *PYPLOT*, we created and compare against two dot charts (GeekRatings & AvgRatings)

Conclusions:

Typically, many users give more attention to the top ~30 of the list rather the bottom ~30. We draw our conclusion with first, the plot charts compared against to another shows Avg Ratings are noticeably higher than Geek Ratings, where they are more conservative in their ratings. Secondly, from the statistics overall results, the standard deviation is direct correlation to the number of voters that participated. What we found is, from the top 30 and bottom 30 is that some games are either overrated or underrated but seemly more popular.

Example #1: bottom 30 (image available below):
Rank # 75: "Pandemic"
Geek Rating: 7.554
Avg Rating: 7.65
Voters: 78,938
Standard deviation of 45,558 (the **highest** of bottom 30 and top 30).
We can say Pandemic is likely more popular, active, and honest than most games given the number of voters that participated.

Example #2: top 30 (image available below):
Rank # 15: "Twilight Imperium (Fourth Edition)"
Geek Rating: 7.976
Avg Rating: 8.77
Voters: 4,715
Standard Deviation of 2,712 (the **lowest** of bottom 30 and top 30)
The low voters count is possibly overrated and more prone to bias

Don't miss out on bottom ranked games because some of them could be just what you're looking for.

Contributions:

Web Scraping (dic_of_games.py)	(Initiated, Edit, Comments by Andrew, Review by Leon)
Statistics_overall.py	(Initiated by Leon, Edit and Comments by Andrew)
Statistics_by_columns.py	(Initiated by Leon, Edit and Comments by Andrew)
Plot_charts.py	(Initiated by Leon, Edit and Comments by Andrew)
Report	(Initiated, Draft & Edits by Leon, Review by Andrew)

Images:

```
Geek Rating Stats:
Min:  7.444
Max:  8.618
Median:  7.655
Standard Deviation:  0.23845951120651168

Average Rating Stats:
Min:  7.57
Max:  8.93
Median:  7.93
Standard Deviation:  0.29624928078163554

Voters Stats:
Min:  3280.0
Max:  78886.0
Median:  16166.5
Standard Deviation:  14508.152489399077
```
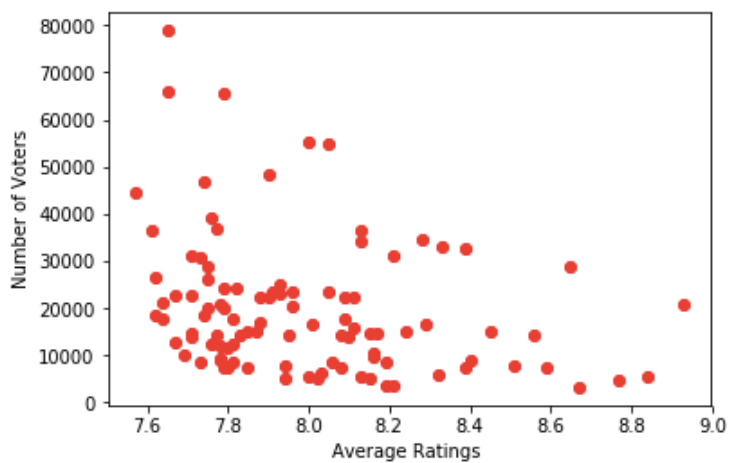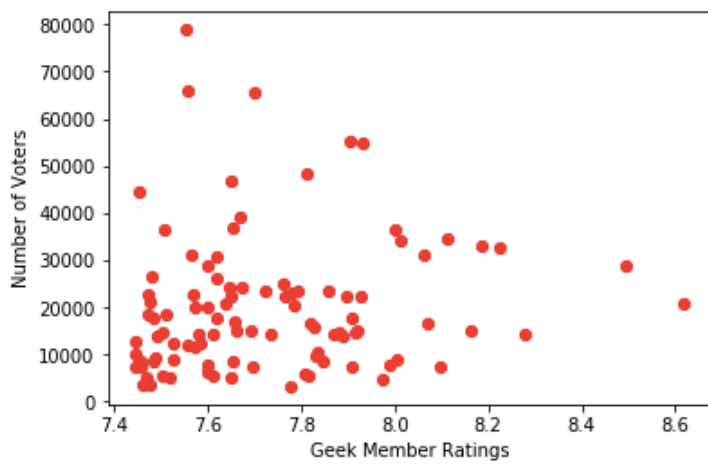
(Statistics for Geek Rating, Average Rating, and Voters)

(Plot Charts, Top: Geek Ratings, Bottom: Average Ratings)

```
Gloomhaven                                         11984.190020
Pandemic Legacy: Season 1                          16625.625746
Through the Ages: A New Story of Civilization       8199.286903
Terraforming Mars                                  18943.262725
Twilight Struggle                                  19028.739143
Star Wars: Rebellion                                8732.245576
Scythe                                             20082.436005
Gaia Project                                        4284.895963
Great Western Trail                                 9616.819800
Terra Mystica                                      17986.691478
The Castles of Burgundy                            19861.963546
War of the Ring (Second Edition)                    5235.872393
7 Wonders Duel                                     21173.706245
The 7th Continent                                   4613.462826
Twilight Imperium (Fourth Edition)                  2712.176820
Puerto Rico                                        31621.479440
Caverna: The Cave Farmers                          13004.810195
Arkham Horror: The Card Game                        8667.136632
Viticulture Essential Edition                       8600.190269
Spirit Island                                       4302.327899
Concordia                                          10280.299182
Agricola                                           32003.707546
Mage Knight Board Game                             13025.020053
Brass: Lancashire                                   7955.890463
Mansions of Madness: Second Edition                 8463.363452
Orléans                                             8366.396897
Blood Rage                                         13583.923403
Food Chain Magnate                                  5052.958878
Mechs vs. Minions                                   6017.145951
A Feast for Odin                                    5630.900065
```

Example # 2 (Standard Deviation for TOP 30 with example #15 highlighted)

```
Tigris & Euphrates                                 13030.426360
Dead of Winter: A Crossroads Game                  17914.233153
Trajan                                              6912.806166
Dominion                                           38002.001276
Pandemic                                           45558.939527
Crokinole                                           5251.200752
Kemet                                               7172.054334
Pandemic: Iberia                                    3086.071433
Descent: Journeys in the Dark (Second Edition)     10656.370425
Stone Age                                          21145.708798
Forbidden Stars                                     3271.410395
Santorini                                           8476.884706
Alchemists                                          7960.735875
War of the Ring (First Edition)                     5407.096872
Castles of Mad King Ludwig                         10338.864150
Legendary Encounters: An Alien Deck Building Game   5034.698443
Star Realms                                        15357.199907
Lisboa                                              2019.084569
The Resistance: Avalon                             12233.150095
A Game of Thrones: The Board Game (Second Edition) 13143.626273
Istanbul                                           10698.562605
Yokohama                                            2966.018949
Agricola (Revised Edition)                          2122.440368
Arcadia Quest                                       4265.663556
Ora et Labora                                       4990.273360
Ticket to Ride: Europe                             25836.706293
Champions of Midgard                                4412.905195
Chaos in the Old World                              7422.092900
Grand Austria Hotel                                 4302.057984
Railways of the World                               5904.237022
Length: 100, dtype: float64
```

Example # 1 (Standard Deviation for BOTTOM 30 with example #75 highlighted)