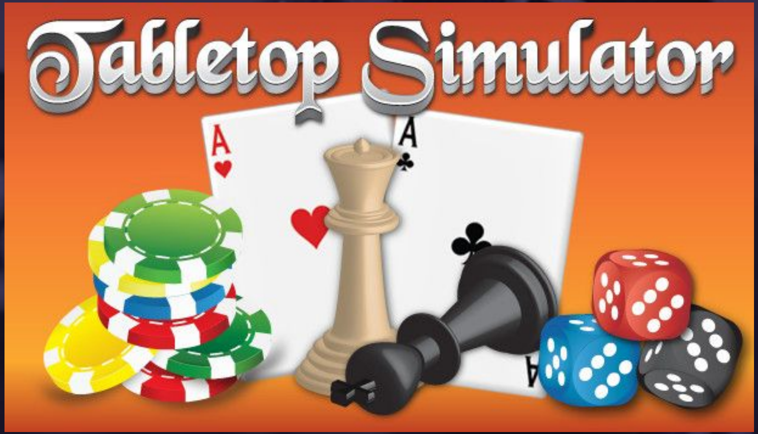
Recommender System Using BoardGameGeek

Presenter: Clifford Cho

KICKSTARTER

BOARD
GAME
GEEK


Tabletop Simulator

Fast-Dealing Prop ding Game ◈ Le jeu de transactions rapides de propriétés
MONOPOLY
AGE 8+
2-6 PLAYERS JOUEURS
Hasbro Gaming

Problem Statement

Board games are held back by issues in money and time. While basic party games range around $25 or less, highly rated board games can easily cost upwards of $50. Playing a new game can be difficult as you have to first read the rules, setup the game, explain the rules, and generally struggle for a bit. As such, choosing the right game can be highly difficult.

The main goal of this project is to use data gathered from BoardGameGeek in order to create a recommender system that provides board game recommendations based on personal interest. Users are recommended to have tried out a game or two from the top rankings of BoardGameGeek they find interesting and use this model to find recommendations.

Can this model create appropriate recommendations for new board games?

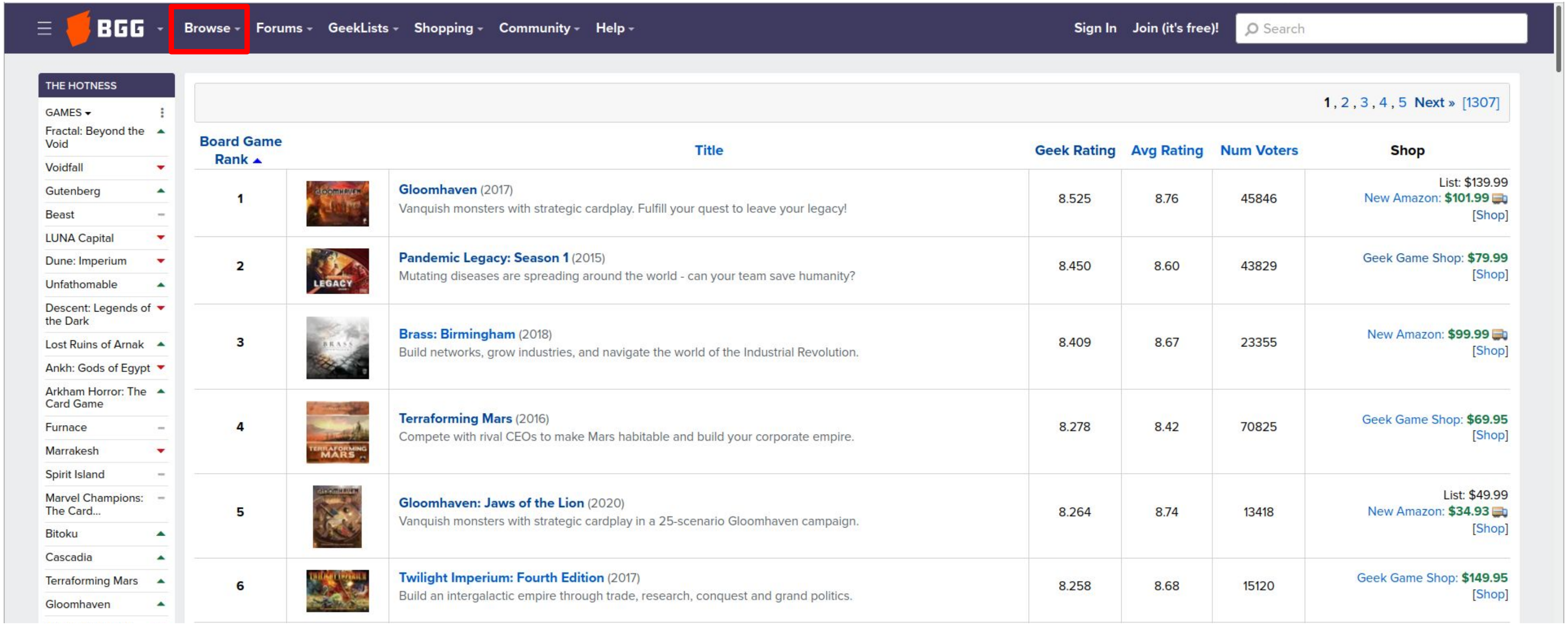
Success Evaluation and Limitations

Some limitations to the project involve a dependence on user interaction and static datasets. BoardGameGeek is a live service with constant updates, so the model is unable to provide current rankings and new games. As user interaction increases, the recommendations are stronger, but this also means that games with minimal user ratings will not function. For use along with the BGG website, using game IDs is recommended.

As this is a recommender system, this model's success metric is subjective to the user. Normally, this kind of model would use tangibles like web traffic, user reviews, or ad revenue as success metrics. For this project, I will act as my own evaluator and present observations on why it is successful. Some metrics include looking at the cosine similarity values, identifying connections between games and recommendations, and personal opinion.

Gathering Data Through Web Scraping and BGG XML API2

Collecting Ranked and Categorical Data



THE HOTNESS

GAMES

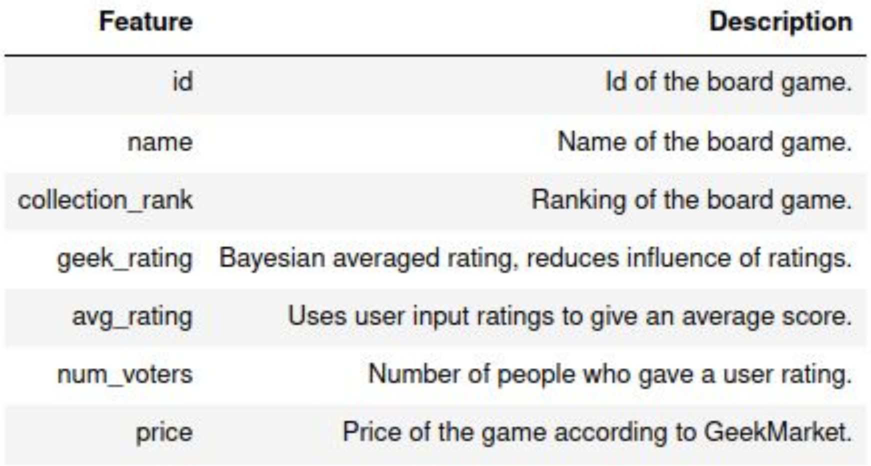
- Fractal: Beyond the Void ▲
- Voidfall ▼
- Gutenberg ▲
- Beast –
- LUNA Capital ▼
- Dune: Imperium ▼
- Unfathomable ▲
- Descent: Legends of the Dark ▼
- Lost Ruins of Arnak ▲
- Ankh: Gods of Egypt ▼
- Arkham Horror: The Card Game ▲
- Furnace –
- Marrakesh ▼
- Spirit Island –
- Marvel Champions: The Card... –
- Bitoku ▲
- Cascadia ▲
- Terraforming Mars ▲
- Gloomhaven ▲

1 , 2 , 3 , 4 , 5 Next » [1307]

Board Game Rank ▲	Title	Geek Rating	Avg Rating	Num Voters	Shop
1	**Gloomhaven** (2017) Vanquish monsters with strategic cardplay. Fulfill your quest to leave your legacy!	8.525	8.76	45846	List: $139.99 New Amazon: **$101.99** [Shop]
2	**Pandemic Legacy: Season 1** (2015) Mutating diseases are spreading around the world - can your team save humanity?	8.450	8.60	43829	Geek Game Shop: **$79.99** [Shop]
3	**Brass: Birmingham** (2018) Build networks, grow industries, and navigate the world of the Industrial Revolution.	8.409	8.67	23355	New Amazon: **$99.99** [Shop]
4	**Terraforming Mars** (2016) Compete with rival CEOs to make Mars habitable and build your corporate empire.	8.278	8.42	70825	Geek Game Shop: **$69.95** [Shop]
5	**Gloomhaven: Jaws of the Lion** (2020) Vanquish monsters with strategic cardplay in a 25-scenario Gloomhaven campaign.	8.264	8.74	13418	List: $49.99 New Amazon: **$34.93** [Shop]
6	**Twilight Imperium: Fourth Edition** (2017) Build an intergalactic empire through trade, research, conquest and grand politics.	8.258	8.68	15120	Geek Game Shop: **$149.95** [Shop]

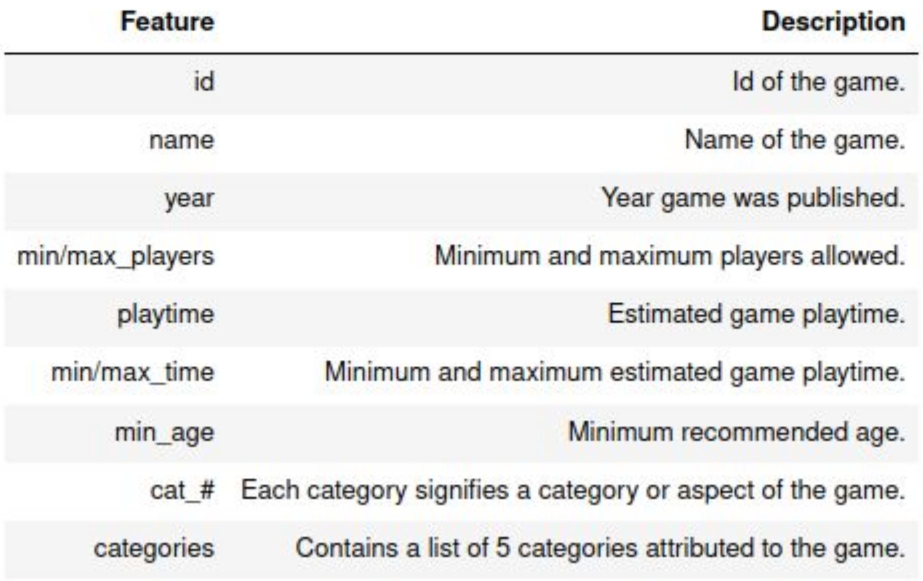
Collecting Ranked and Categorical Data

Features for Ranked Data



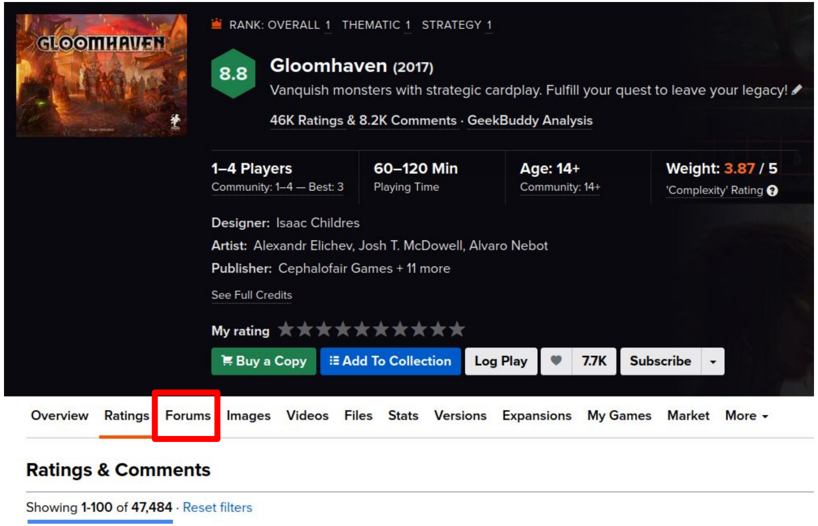
Feature	Description
id	Id of the board game.
name	Name of the board game.
collection_rank	Ranking of the board game.
geek_rating	Bayesian averaged rating, reduces influence of ratings.
avg_rating	Uses user input ratings to give an average score.
num_voters	Number of people who gave a user rating.
price	Price of the game according to GeekMarket.

Features for Categorical Data



Feature	Description
id	Id of the game.
name	Name of the game.
year	Year game was published.
min/max_players	Minimum and maximum players allowed.
playtime	Estimated game playtime.
min/max_time	Minimum and maximum estimated game playtime.
min_age	Minimum recommended age.
cat_#	Each category signifies a category or aspect of the game.
categories	Contains a list of 5 categories attributed to the game.

Collecting User Rankings



RANK: OVERALL 1 THEMATIC 1 STRATEGY 1

8.8 **Gloomhaven** (2017)

Vanquish monsters with strategic cardplay. Fulfill your quest to leave your legacy!

46K Ratings & 8.2K Comments · GeekBuddy Analysis

1–4 Players	60–120 Min	Age: 14+	Weight: 3.87 / 5
Community: 1–4 — Best: 3	Playing Time	Community: 14+	'Complexity' Rating

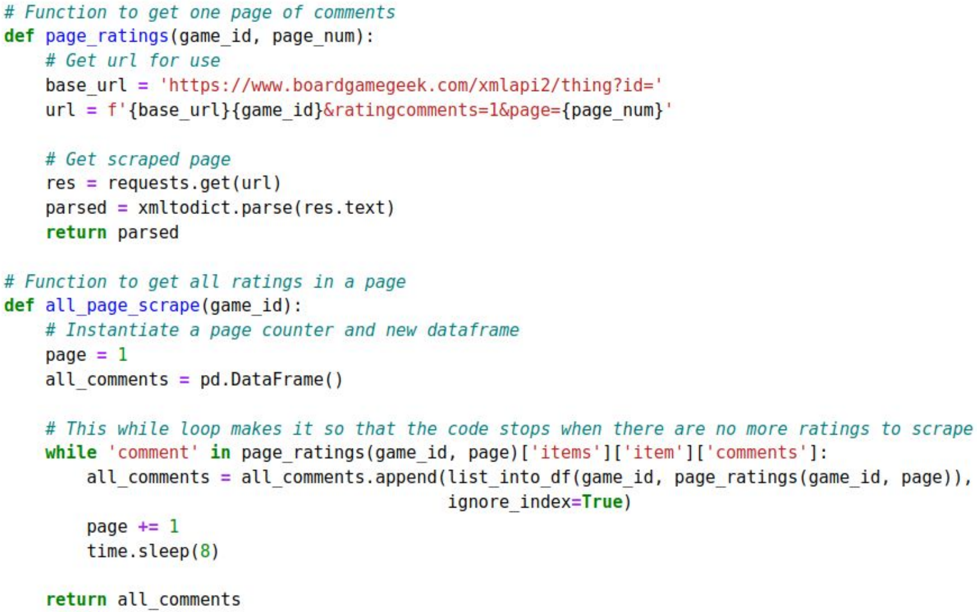
Designer: Isaac Childres
Artist: Alexandr Elichev, Josh T. McDowell, Alvaro Nebot
Publisher: Cephalofair Games + 11 more

See Full Credits

My rating ★★★★★★★★★★

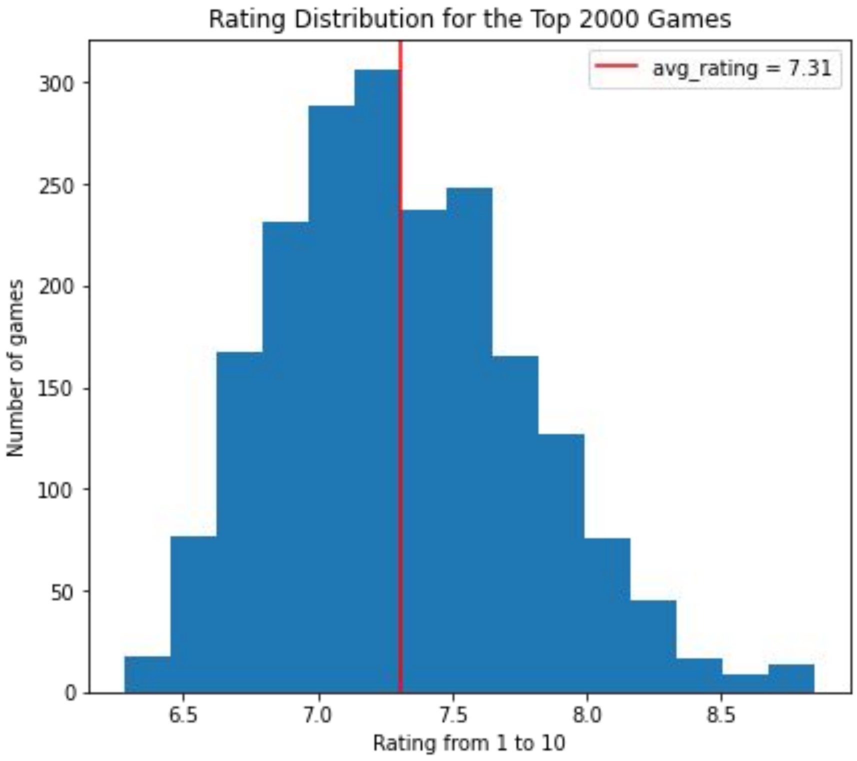
Ratings & Comments

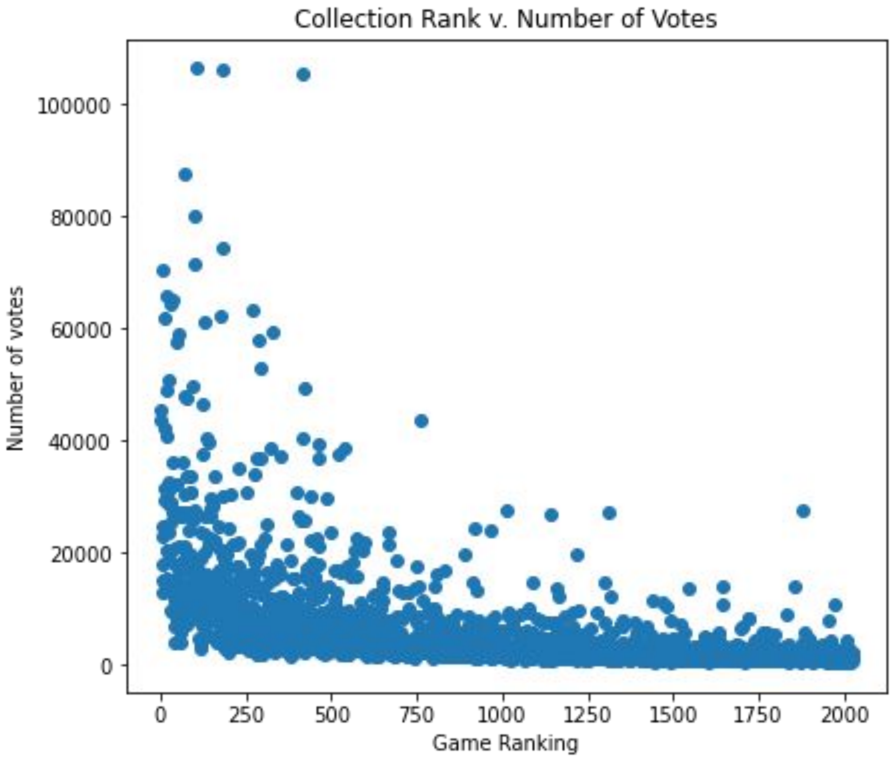
Showing 1-100 of 47,484 · Reset filters



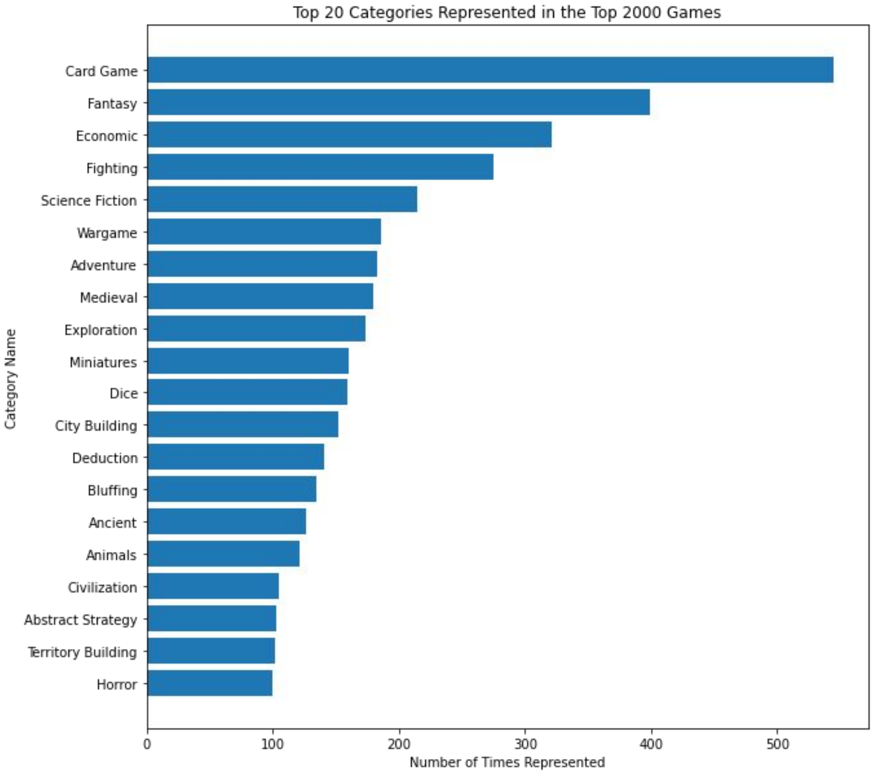
```python
# Function to get one page of comments
def page_ratings(game_id, page_num):
    # Get url for use
    base_url = 'https://www.boardgamegeek.com/xmlapi2/thing?id='
    url = f'{base_url}{game_id}&ratingcomments=1&page={page_num}'

    # Get scraped page
    res = requests.get(url)
    parsed = xmltodict.parse(res.text)
    return parsed

# Function to get all ratings in a page
def all_page_scrape(game_id):
    # Instantiate a page counter and new dataframe
    page = 1
    all_comments = pd.DataFrame()

    # This while loop makes it so that the code stops when there are no more ratings to scrape
    while 'comment' in page_ratings(game_id, page)['items']['item']['comments']:
        all_comments = all_comments.append(list_into_df(game_id, page_ratings(game_id, page)),
                                           ignore_index=True)
        page += 1
        time.sleep(8)

    return all_comments
```

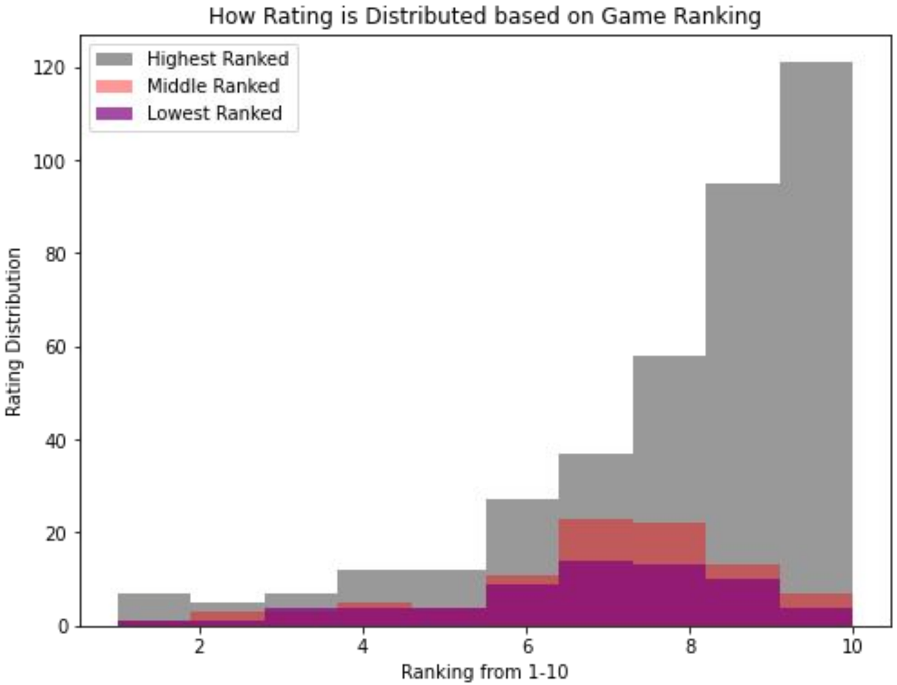
Exploratory Data Analysis



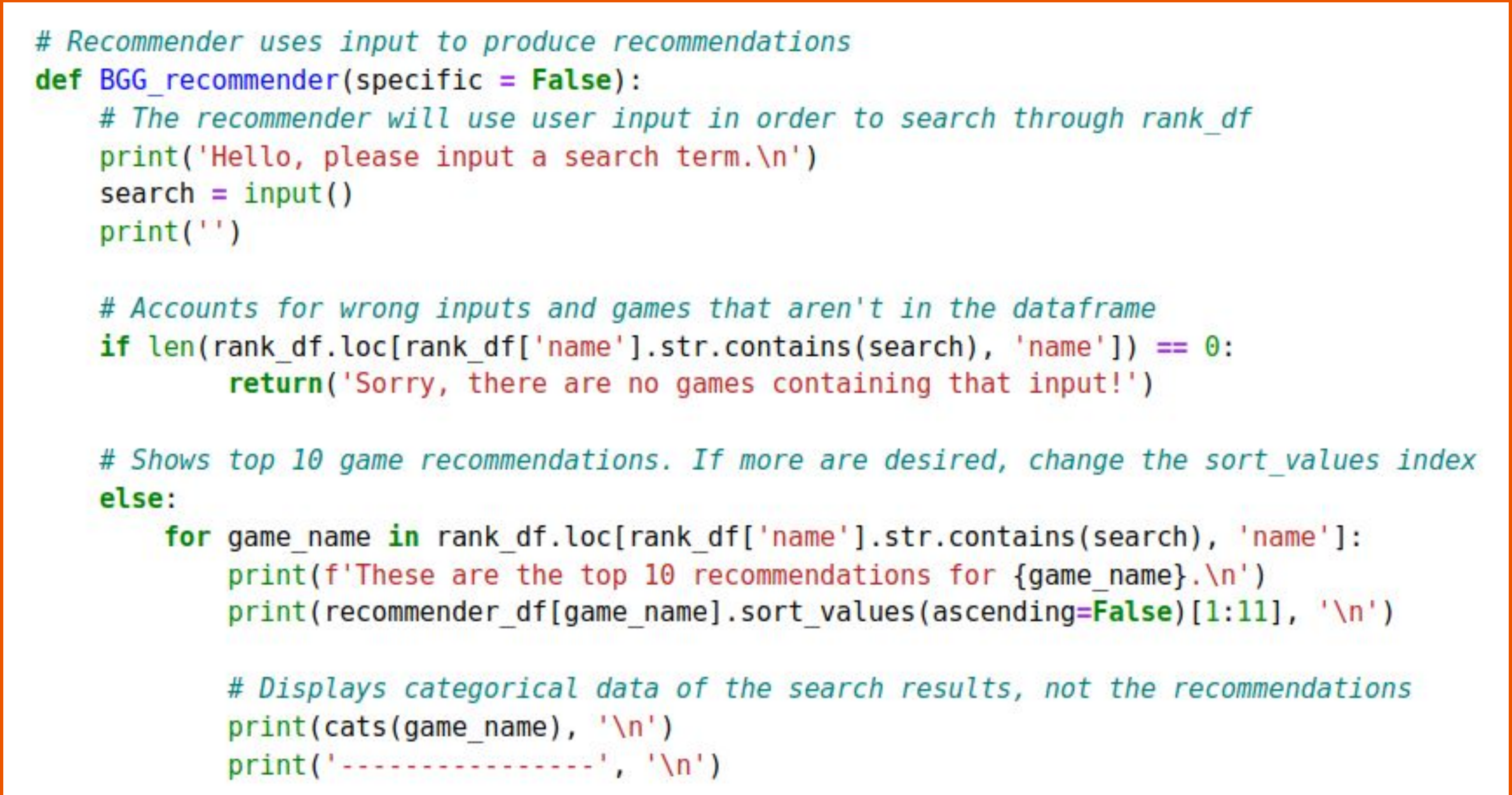


Exploratory Data Analysis

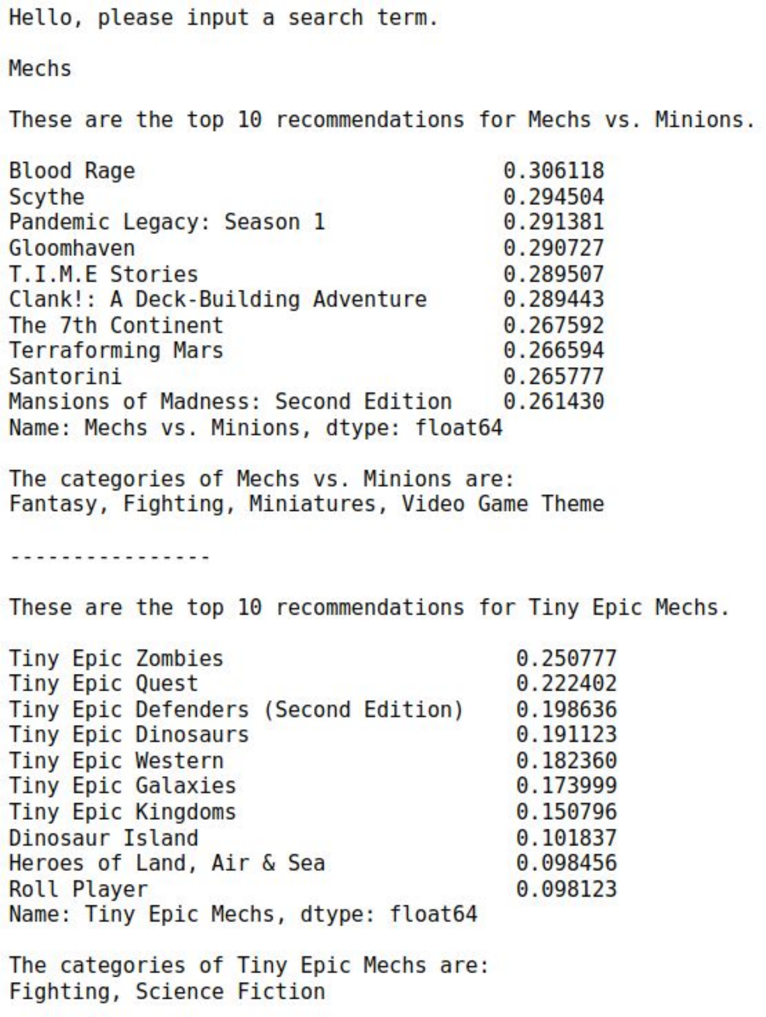




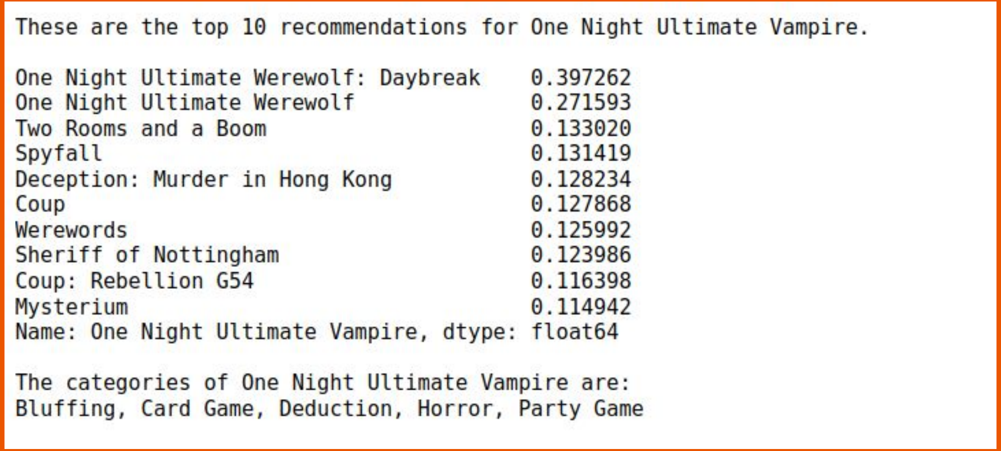
Model Output and Evaluation



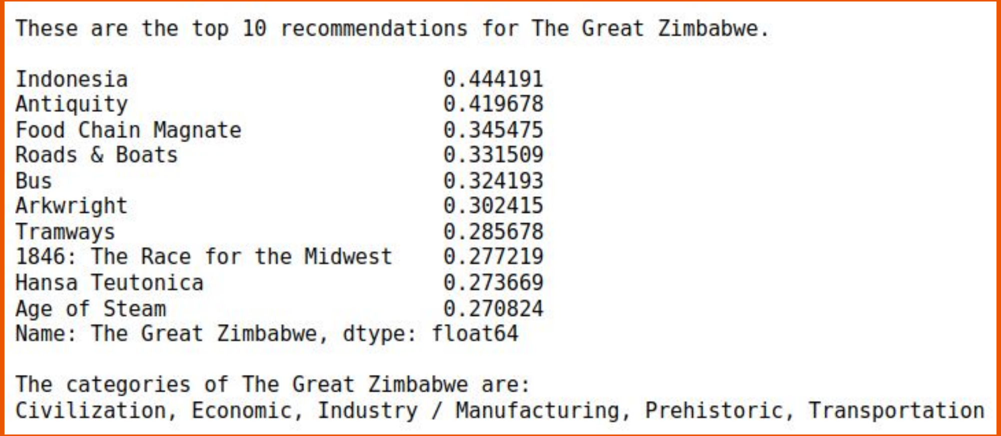
```python
# Recommender uses input to produce recommendations
def BGG_recommender(specific = False):
    # The recommender will use user input in order to search through rank_df
    print('Hello, please input a search term.\n')
    search = input()
    print('')

    # Accounts for wrong inputs and games that aren't in the dataframe
    if len(rank_df.loc[rank_df['name'].str.contains(search), 'name']) == 0:
            return('Sorry, there are no games containing that input!')

    # Shows top 10 game recommendations. If more are desired, change the sort_values index
    else:
        for game_name in rank_df.loc[rank_df['name'].str.contains(search), 'name']:
            print(f'These are the top 10 recommendations for {game_name}.\n')
            print(recommender_df[game_name].sort_values(ascending=False)[1:11], '\n')

            # Displays categorical data of the search results, not the recommendations
            print(cats(game_name), '\n')
            print('----------------', '\n')
```



```
Hello, please input a search term.

Mechs

These are the top 10 recommendations for Mechs vs. Minions.

Blood Rage                             0.306118
Scythe                                 0.294504
Pandemic Legacy: Season 1              0.291381
Gloomhaven                             0.290727
T.I.M.E Stories                        0.289507
Clank!: A Deck-Building Adventure      0.289443
The 7th Continent                      0.267592
Terraforming Mars                      0.266594
Santorini                              0.265777
Mansions of Madness: Second Edition    0.261430
Name: Mechs vs. Minions, dtype: float64

The categories of Mechs vs. Minions are:
Fantasy, Fighting, Miniatures, Video Game Theme


----------------

These are the top 10 recommendations for Tiny Epic Mechs.

Tiny Epic Zombies                       0.250777
Tiny Epic Quest                         0.222402
Tiny Epic Defenders (Second Edition)    0.198636
Tiny Epic Dinosaurs                     0.191123
Tiny Epic Western                       0.182360
Tiny Epic Galaxies                      0.173999
Tiny Epic Kingdoms                      0.150796
Dinosaur Island                         0.101837
Heroes of Land, Air & Sea               0.098456
Roll Player                             0.098123
Name: Tiny Epic Mechs, dtype: float64

The categories of Tiny Epic Mechs are:
Fighting, Science Fiction
```



```
These are the top 10 recommendations for One Night Ultimate Vampire.

One Night Ultimate Werewolf: Daybreak    0.397262
One Night Ultimate Werewolf              0.271593
Two Rooms and a Boom                     0.133020
Spyfall                                  0.131419
Deception: Murder in Hong Kong           0.128234
Coup                                     0.127868
Werewords                                0.125992
Sheriff of Nottingham                    0.123986
Coup: Rebellion G54                      0.116398
Mysterium                                0.114942
Name: One Night Ultimate Vampire, dtype: float64

The categories of One Night Ultimate Vampire are:
Bluffing, Card Game, Deduction, Horror, Party Game
```



```
These are the top 10 recommendations for The Great Zimbabwe.

Indonesia                         0.444191
Antiquity                         0.419678
Food Chain Magnate                0.345475
Roads & Boats                     0.331509
Bus                               0.324193
Arkwright                         0.302415
Tramways                          0.285678
1846: The Race for the Midwest    0.277219
Hansa Teutonica                   0.273669
Age of Steam                      0.270824
Name: The Great Zimbabwe, dtype: float64

The categories of The Great Zimbabwe are:
Civilization, Economic, Industry / Manufacturing, Prehistoric, Transportation
```

Conclusion

The model was successful based on several metrics. There are strong connections shown in the data as displayed by games in the same franchise and category relevance in the game recommendations.

The relevance of each previously mentioned metric was shown. There is clear correlation between games and user ratings. I believe the model increases the chance of picking a worthwhile game.

Recommendations and Future Work

The BGG website contains a strong advanced search tool. I recommend interested users to try out games and contribute if desired.

A recommendation is to gather more data on games outside of the top 2000 as entertainment is subjective, and the model utilizes collaboration. Having additional games and user connections strengthens the model.

Future work also includes creating an online version of the model using tools like Flask or Django.

Thank You! Any Questions?