Game Boards

By: Sheena, Laura, & Aleena

Project Overview

- Board game data is interesting and fun
- The different ways in which people win and have opinions on these games.
- Board game data is raw
- Online games data is created through technology itself
- Board games data don’t have technology as a way to show who won

Purpose

- Analyze the various data points in the board game dataset
- See if we can find any correlations between data points
  - Games' ratings
  - Games' complexity
- Model that could predict what makes a good game.

Questions To Answer

- How complex the games are according to people's ratings of the games?
- What are the games' maximum amount of players?
- What are the games' average play time?
- What are the games' average complexity?
- Can we reasonably and accurately predict the average game rating?

Challenge Summary

Things we had trouble with:

- Finding a dataset
- Getting our data into pgAdmin (“psychopg2” does not exist)
- Produced predictions that were not within a normal range of the other predictions

Dashboard

Tableau

Purpose

Data table

Preprocessing

Models and outcomes

```
# Games pusblished before 1800 removed.
games_df[games_df['Year Published'] < 1800].index
games_df.drop(games_df[games_df['Year Published'] < 1800].index, inplace = True)
# Games with Max Players of 0 removed.
games_df[games_df['Max Players'] == 0].index
games_df.drop(games_df[games_df['Max Players'] == 0].index, inplace = True)
# Games with Play Time of 0 removed.
games_df[games_df['Play Time'] == 0].index
games_df.drop(games_df[games_df['Play Time'] == 0].index, inplace = True)
games_df.describe()
```

Pre-Processing

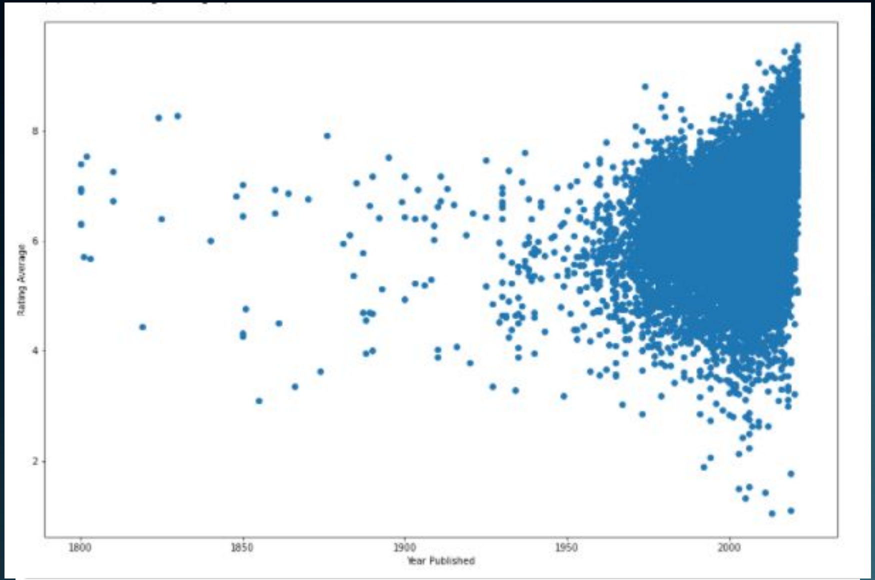
Datatypes

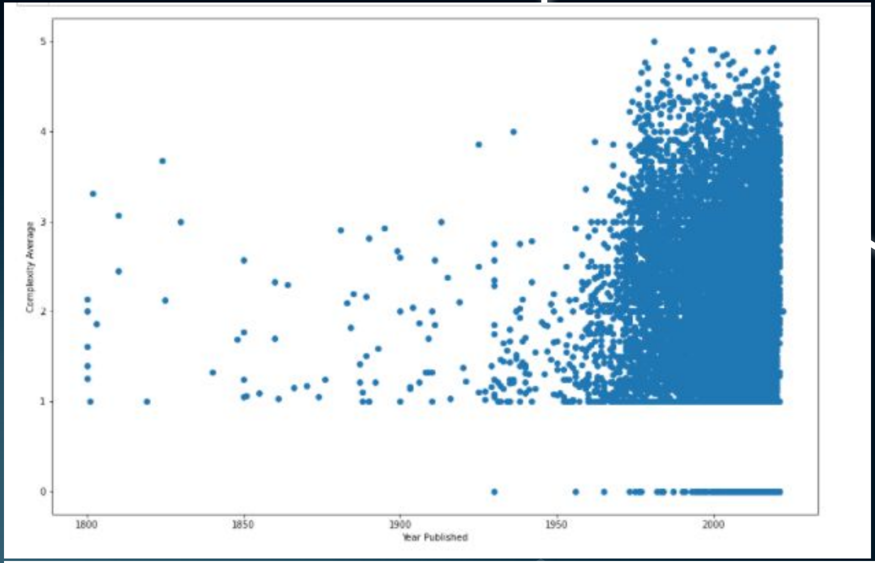
```python
# Check data types
games_df.dtypes
```

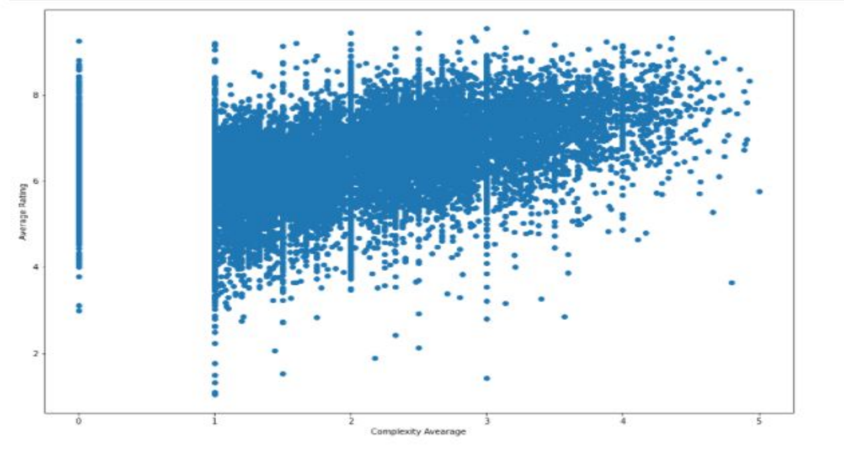
```
index                  int64
ID                     int64
Name                  object
Year Published         int64
Min Players            int64
Max Players            int64
Play Time              int64
Min Age                int64
Users Rated            int64
Rating Average        object
BGG Rank               int64
Complexity Average    object
Owned Users            int64
Domains               object
dtype: object
```

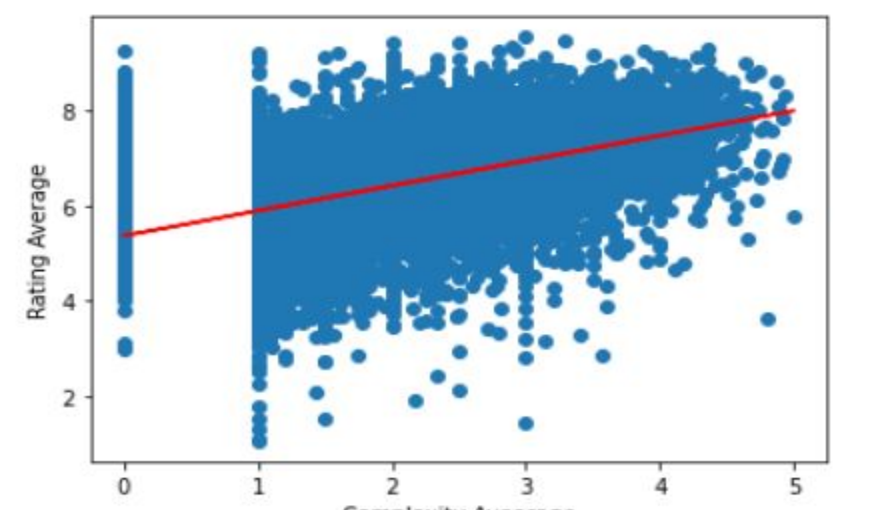
```python
# Change data types of Complexity object to Float64
games_df["Complexity Average"] = games_df["Complexity Average"].astype(float)
# Change data types of Complexity object to Float64
games_df["Rating Average"] = games_df["Rating Average"].astype(float)
games_df.dtypes
```

```
index                   int64
ID                      int64
Name                   object
Year Published          int64
Min Players             int64
Max Players             int64
Play Time               int64
Min Age                 int64
Users Rated             int64
Rating Average        float64
BGG Rank                int64
Complexity Average    float64
Owned Users             int64
Domains                object
dtype: object
```


Rating Average
Year Published
1800
1850
1900
1950
2000
2
4
6
8


Complexity Average
Year Published
1800
1850
1900
1950
2000
0
1
2
3
4
5


Average Rating
Complexity Avearage
0
1
2
3
4
5
2
4
6
8

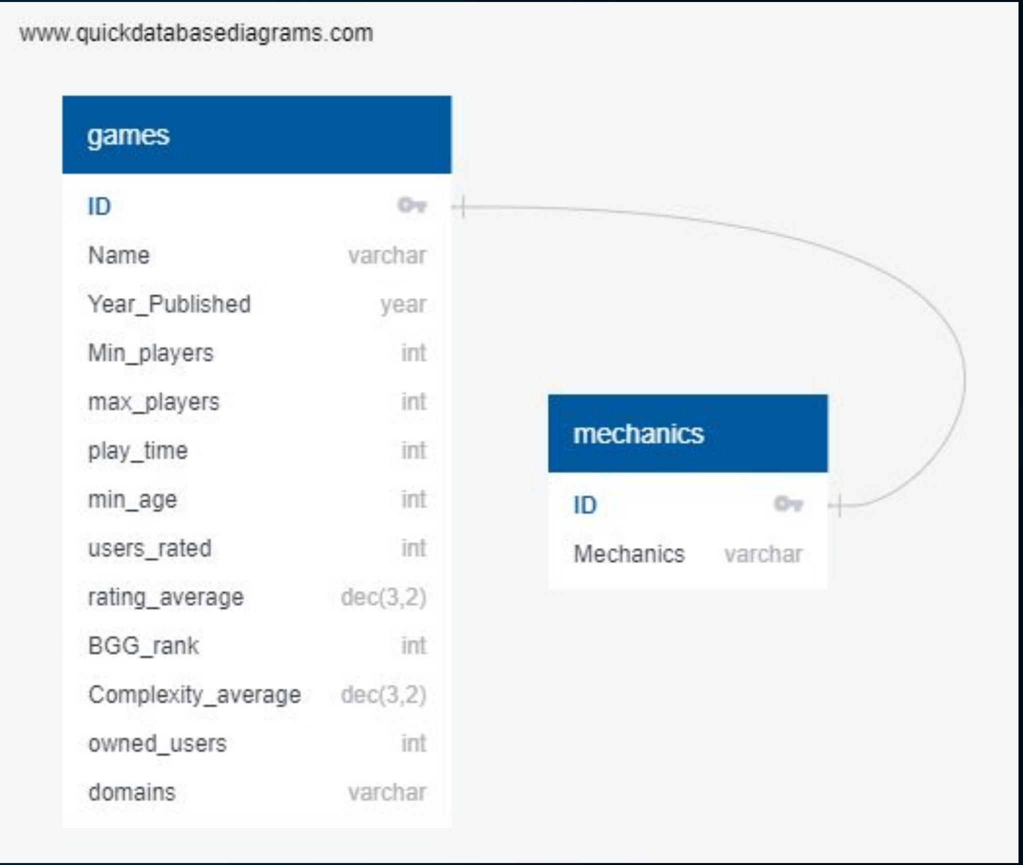

Rating Average
0
1
2
3
4
5
2
4
6
8

$$y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$$

Multiple Linear Regression

Database

- Used a PostgreSQL database
- Two tables
  - Mechanics
  - All other features



Description & Links of Datasource

BoardGameGeek.com

- Allow users to rank board games on a scale of 1-10,
- Presents an average rating based on individual ratings.
- A collection of all ranked games in the BGG database (raw dataset)

Resources

- Raw Data: https://www.kaggle.com/andrewmvd/board-games/version/2
- Data Source: BoardGameGeek.com
- Software: Excel 365, Python 3.6.1, pgAdmin 4

Communication Protocols

- Slack
- Decided on a day and time that we agreed to meet each week from now on (3PM on Fridays)
- When a question arises...

  Ask each other
- If we all don't know the answer…

  Ask our captain, Savannah.

Next Week

- Visualizations from Tableau
- Building models
- Refine ML model

Segment 2
MLR

Segment 2 - Project Overview

Initially Laura took the dataframe trying to form a model with four files:

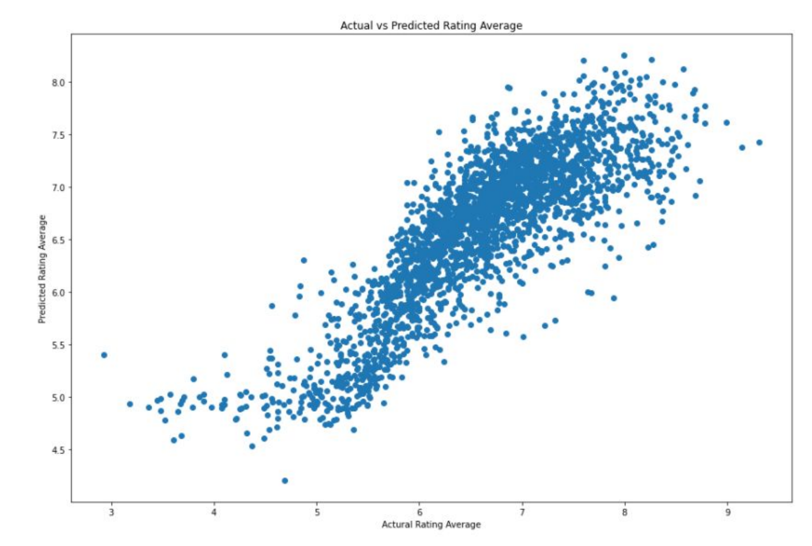
- mlr_after_1800_wtih_db.ipynb
- mlr_getdummies_predict_rating.ipynb
- mlr_predict_average_rating_with_db.ipynb
- mlr_predict_complexity_wtih_db.ipynb

The accuracy scores of the these files were not high enough and so she ended up making another file with a model that we will be talking about for this segment:

- Multiple_linear_regression_model.ipynb

Segment 2 - Description of Model

This is a scatterplot of the correlation between actual rating averages from our board game data and the predicted rating averages. The way that this model works is that it compares actual and predicted rating averages. The reason why this model was initially chosen was because its accuracy is the best out of all the others that were done. This model has an accuracy of 70% and the statistics included in analysis is mainly the rating averages from our data. The reason why our statistic is average ratings is because it is one of the only statistical data that seems best to analyze so far together with complexity.



Pre-Processing

The Processes Involved for Pre-Processing:

1. Checking data types for numerical values
2. Change data types of complexity object to Float
3. Finding the null values
4. Dropping the null rows
5. Finding duplicate entries
6. Using get_dummies on domain categories to get new columns with columns values
7. Checking the dataframe
8. Dropping the ID and Name
9. Transforming the values of owned users and users rated to smaller numbers to give the computer easier numbers to compare. This was done by dividing Users Rated by 1000 and Owned Users by 1000.

Accuracy

Initially Laura took the dataframe trying to form a model with four files:

- mlr_after_1800_wtih_db.ipynb
- mlr_getdummies_predict_rating.ipynb
- mlr_predict_average_rating_with_db.ipynb
- mlr_predict_complexity_wtih_db.ipynb

The accuracy scores of the these files were not high enough and so she ended up making another file with a model that we will be talking about for this segment:

- Multiple_linear_regression_model.ipynb

If There Was More Time

Something that I would have done was help with finding more stories from our data. What I mean by this is I would help Laura investigate other data that may have a possible correlation with average ratings that may have a high enough accuracy to analyze.

Segment 3 Visualizations

Project Overview

This week, we focused on visualizations and how to make it interactive on Tableau. The data that we are using for these visualizations is the correlation between the games' rating averages and complexity averages of our domain and comparing it with the other domains that we have. When we say domains, we mean the types of games that will be shown in our interactive visualization:

- Abstract Games
- Children's games
- Customizable Games
- Family Games
- Party Games
- Strategy Games
- Thematic Games
- Wargames

Purpose

The whole point of the work for this week is making sure that our visualization data has any correlation and figuring out what we need to separate and/or put together in order to see this correlation between average ratings, complexity averages, and our domains clearly. On top of that of all that, we needed to figure a way to make it all interactive.

Questions to Answer

Questions that we hope to answer from our visualizations:

- What is the correlation between complexity averages and rating averages in Strategy Games?
- What is the correlation between complexity averages and rating averages in Abstract Games?
- What is the difference between the the correlation in Strategy Games and Abstract Games?
- Do people prefer Strategy Games or Abstract Games?
- Which type of game do people like the most out of all the games that we have?

If There Was More Time

Something that I would have done was help with finding more stories from our data. What I mean by this is I would help Laura investigate other data that may have a possible correlation with average ratings that may have a high enough accuracy to analyze.

Challenge Summary

In the beginning, we had trouble thinking about how to turn our visualization with so many different specific game names attached together with their game types. Since this was the case, Sheena was trying to figure out how to condense these games into their own domains. To put it in simple words, how to separate these game types and condense them into their own domain. This was something that was solved through the grouping function on Tableau where you would select the games with the same game types and group it altogether. After that, we were able to select the different game types and look at the relationship between rating averages and complexity averages in these different domains and compare these domains with the other domains that we have. The only problem that we have now is downloading the desktop version of Tableau before Sheena's version of Tableau ends to show these visualizations when we present.

If There Was More Time

If there was more time, I would try to get the pictures for these visualization. The only problem with this is that I need to gain access to the Tableau visualizations that Sheena has made to further explain what these visualizations look like.

Thank You