

# 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 published 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

```
1 # Check data types
2 games_df.dtypes
```

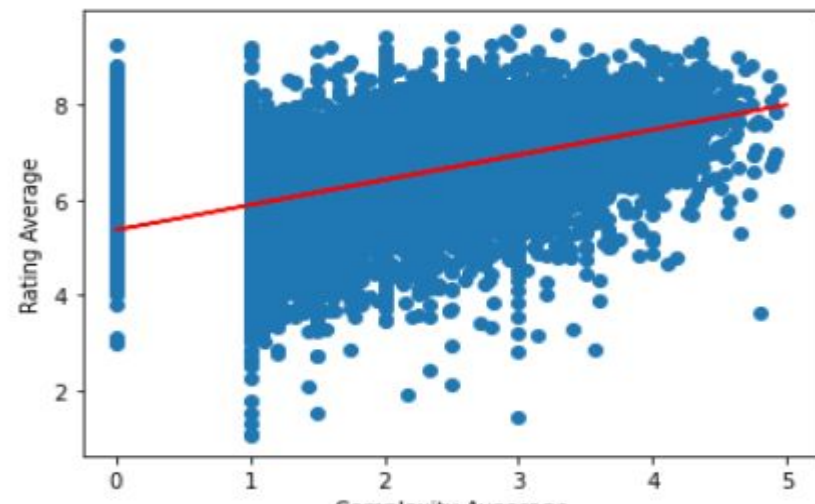
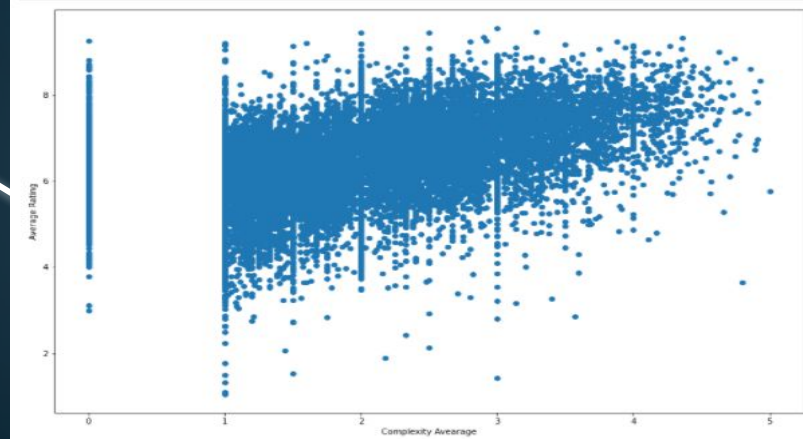
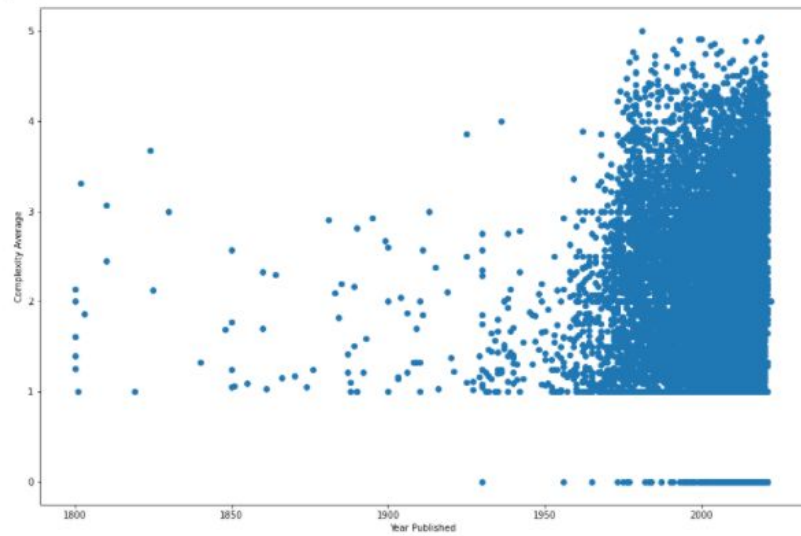
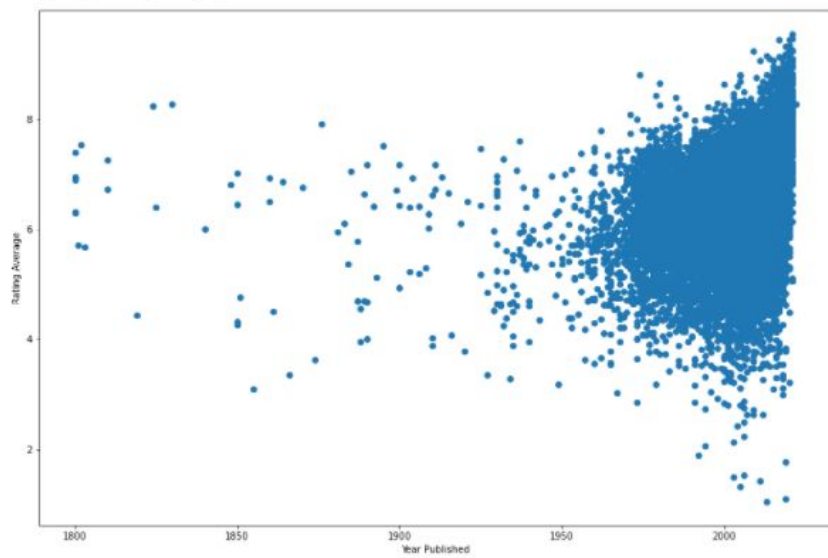
```
index          int64
ID              int64
Name           object
Year Published  int64
Min Players     int64
Max Players     int64
Play Time      int64
Min Age        int64
Users Rated    int64
Rating Average  object
BGG Rank       int64
Complexity Average  object
Owned Users    int64
Domains   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

```
1 # Change data types of Complexity object to Float64
2 games_df["Complexity Average"] = games_df["Complexity Average"].astype(float)
3 # Change data types of Rating object to Float64
4 games_df["Rating Average"] = games_df["Rating Average"].astype(float)
5 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
```

# Datatypes





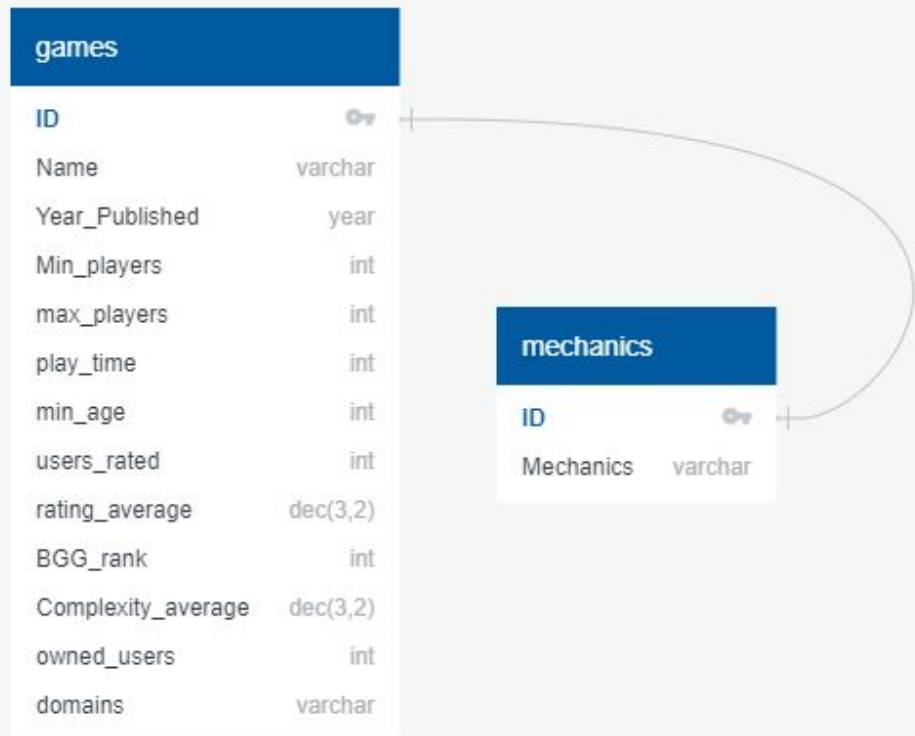
$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Multiple Linear  
Regression

# Database

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

www.quickdatabasediagrams.com



## 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





Thank  
You