EDA

1. Can I predict a board game’s rating from metrics with low risk of input bias?

2. All of my data comes from [BoardGameGeek](https://boardgamegeek.com/) (BGG) through the BGG XML API. As outlined in the [terms of use](https://boardgamegeek.com/wiki/page/XML_API_Terms_of_Use), I can use this data as long as I credit BGG as the source

3. Variables in data:

Logistic:

* id: Internal BGG id number, will not be used in the model
* name: Name of the boardgame, will not be used in the model

Target:

* bayesaverage: Weighted average rating on a game.

Factors:

* yearpublished: Year that the game was published.
* description: Paragraph of text describing the game.
* minplayers: The fewest number of players you can play the game with.
* maxplayers: The maximum number of players.
* minplaytime: Expected minimum playtime.
* maxplaytime: Expected maximum playtime.
* complexity: How difficult a game is to learn, play, and play well.
* categories: A list of categories that a game belongs to.
* mechanics: A list of mechanics that a game contains.

4. Numeric Variables:

* yearpublished:
  + Sample Size: 764
  + Mean: 1993.021
  + Std: 176.190
  + 5 number summary: [-2200, 2004, 2011, 2016, 2023]
* minplayers:
  + Sample Size: 764
  + Mean: 1.959
  + Std: 0.748
  + 5 number summary: [1., 2., 2., 2., 8.]
* maxplayers:
  + Sample Size: 764
  + Mean: 6.115
  + Std: 9.336
  + 5 number summary: [1., 4., 5., 6., 100.]
* minplaytime:
  + Sample Size: 764
  + Mean: 46.398
  + Std: 34.961
  + 5 number summary: [ 0., 30., 30., 60., 360.]
* maxplaytime:
  + Sample Size: 764
  + Mean: 65.568
  + Std: 61.721
  + 5 number summary: [ 0., 30., 45., 90., 1000.]
* complexity:
  + Sample Size: 764
  + Mean: 2.125
  + Std: 0.794
  + 5 number summary: [1.0158 , 1.47715, 2.00455, 2.6145 , 4.4365 ]

Categorical Variables:

* description:
  + Sample Size: 764
  + Category Counts: N/A
* categories:
  + Sample Size: 764
  + Category Counts: 84 (A game can have multiple categories)
* mechanics: A list of mechanics that a game contains.
  + Sample Size: 764
  + Category Counts: 192 (A game can have multiple mechanics)

5.

A comparison of blue and white dots

Description automatically generated

6. Questions:

* Were there any challenges in finding the right dataset for your project?
  + No, this is a data source that I have worked with before. I consider it one of the best places for information on boardgames.
* Are there any other concerns that I have with the project?
  + I am a little worried that I wont find anything significant with my data, my intuition tells me that most of the numeric factors won’t contribute much to the models, so I will have a few lists of text data to carry the model’s weight.