# KAGGLE-Board Games

**Repository**: https://github.com/DmitriSajutinski/SA-project

2. Business understanding

One of the goals of this project is to get broader understanding of what makes one board game popular than the others. For that we are using our dataset to find if there are certain indicators in that data which can be used to anticipate which board games are more popular and/or have higher user ratings.

Our dataset itself is from KAGGLE website but is originally collected from BoardGameGeek, which is an online forum and a game database that holds reviews and other type of information about different board games and card games. Collected dataset contains information around 20,000 different board game. Our team has three members and we don't use any other type of relevant software than Jupyter Notebook.

Useful terminology:

- average weight - board game complexity rating (from 1 to 5)
- board game honor count - how many awards board game has received
- board game category count - how many categories are related to specific board game
- board game mechanic count - how many different game mechanics board game uses
- board game version count - availability in different languages

3. Data understanding

For our project to be successful we need a dataset that has many different parameters about every board game and which are comparable to each other. Our first dataset didn't contain sufficient amount of data for our project to be successful. We searched KAGGLE and fortunately, we found another dataset which is quite similar to ours (also contains around 20, 000 different board games) and has more comparable data that we can use in our project. We also tried to find if there is a possibility to collect the data directly from BoardGameGeek but we didn't find any practical solutions to do that.

Like mentioned previously ourdataset has around 20,000 records and 52 fields, from which 24 are usable inour project. Most of these records are also numerical, so they don't need somekind of conversion and we can use them directly.We think if we describe fields thatwe are going to use , describe why we decide use that fields, we will best waydescribe our data.In our analsys we are using:name - we are going to find outif first letter of the board game name can effect on its score, yearpublished -this show us which yearˇgame was published, it helps us answer the questionthat year can effect on the score or not. Minplayers and maxplayers showhow many players are recommended for comfortable game, also we can figure whichrecommendation player count is most good if you are going to make high rated boardgame. In a same way we planned to use minplaytime and maxplaytime, which showuse average game play time. Minage

show us age rating, so we think we have touse it, because dependencies between it and game rating. Usersrated show us how many players/users have rated games and this will help us to understand how fair is assessment of the game, than more than better. Average is more important because, it show the average game rating and this way we will compare is game good or not. Bavarage is rating from the website. So we have two results that can help us to improve our reseach. Both of them are from 1 to 10. Avgweight, numweights show how hard is a game for the users(1-5 system) and numweights show how fair result is. Numgeeklists show us how many users have added this board game is a like list  and siteviews show us how many users have watched this game. We think that is part of success(popularity) so we can use it. We also think that news , blogs, weblink, podcast can be part of advertising and we want to understand how advertising effect on a rating of game(more of it is good or not).