# Analysis of board games

András Bognár, Viktória Kónya

31 March 2022

**Description of the project**

In our final project we chose to analyze the characteristics of board games. The main source of our data is the GitHub repo of TidyTuesday from where we downloaded the dataset of the board games in csv format. The dataset contains rich information about the board games but lacks details about their prices. Therefore, we requested this data using the Board Game Atlas' API along with information about the learning and strategy complexity of the games.

The chart below summarizes our data pipeline with the corresponding notebook references.
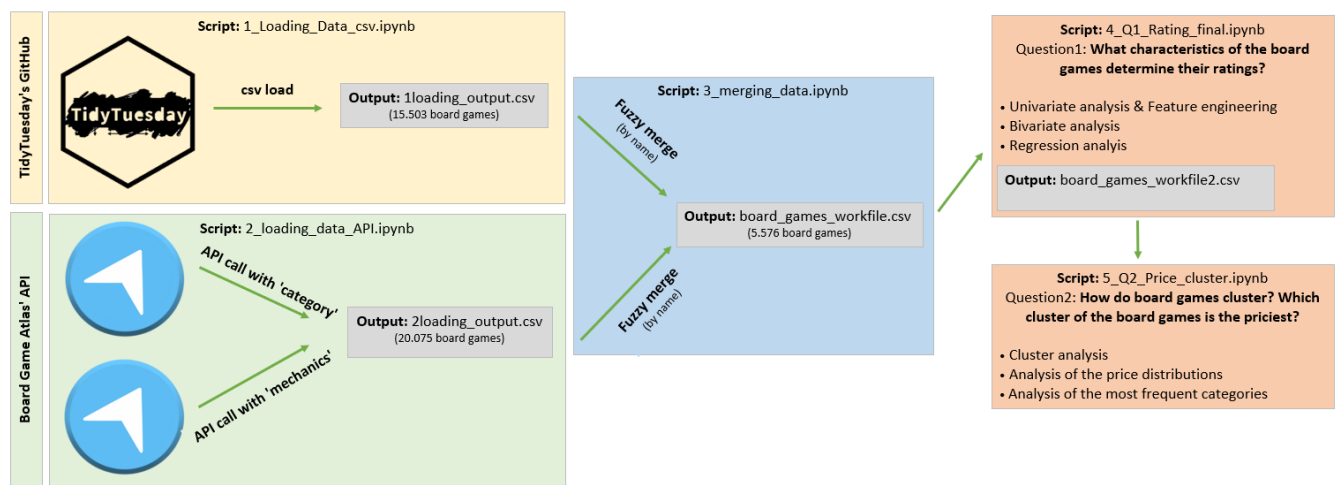


The **1_Loading_Data_csv.ipynb** contains the load of the TidyTuesday's dataset with the preliminary data cleaning (removing duplicates, examination of the fields with high missing rate and exclusion of the unnecessary fields). The dataset consists of 15.503 unique board games.

The **2_loading_data_API.ipynb** contains the API requests along with the same preliminary data cleaning steps as mentioned above. We had to create our dataset using two different API calls (one call using the category and another call using the mechanics as keys in the API call). The dataset consists of 20.075 unique board games.

The **3_merging_data.ipynb** contains the join of the two data sources. We joined the datasets on the name of the board games using fuzzy merge (the details of the fuzzy merge are discussed in the notebook). The final workfile consists of 5.576 unique board games. Note that we suffixed the fields from the TidyTuesday's dataset with '_base' and the ones from the API with '_api'.

We chose to explore two analytical questions about the board games:

- **What characteristics of the board games determine their ratings?**
- **How do board games cluster? Which cluster of the board games is the priciest?**

The **4_Q1_Rating_final.ipynb** contains the analysis related to the first question. It contains a detailed analysis of each variable of interest together with the additional feature engineering steps, elimination of extreme values and recoding of the variables. The analysis starts from the univariate analysis, then explores

the relationship of each predictor with the outcome and finally examines the pattern of association in regression framework.

The **5_Q2_Price_cluster.ipynb** contains our second analysis where we examined that based on their characteristics which board games create separate groups. We used cluster analysis to create group membership. Once we created separate clusters of the board games we examined how the distribution of the prices differ across the clusters. We also examined which categories are the most frequent in each cluster.

**Appendix**

**List of fields in the TidyTuesday's dataset:**

| variable | description |
|---|---|
| game_id | Unique game identifier |
| description | A paragraph of text describing the game |
| max_players | Maximum recommended players |
| max_playtime | Maximum recommended playtime (min) |
| min_age | Minimum recommended age |
| min_players | Minimum recommended players |
| min_playtime | Minimum recommended playtime (min) |
| playing_time | Average playtime |
| year_published | Year game was published |
| artist | Artist for game art |
| category | Categories for the game (separated by commas) |
| compilation | If part of a multi-compilation - name of compilation |
| family | Family of game - equivalent to a publisher |
| mechanic | Game mechanic - how game is played, separated by comma |
| publisher | Comoany/person who published the game, separated by comma |
| publisher | Average rating on Board Games Geek (1-10) |
| users_rated | Number of users that rated the game |
| name | Name of the game (PK) |

**List of used fields from the API:**

| variable | description |
|---|---|
| average_learning_complexity | Average learning complexity score |
| average_strategy_complexity | Average strategy complexity score |
| name | Name of the game (PK) |
| price | Price in USD |
| price_au | Price in AUD |
| price_ca | Price in CAD |
| price_uk | Price GBP |