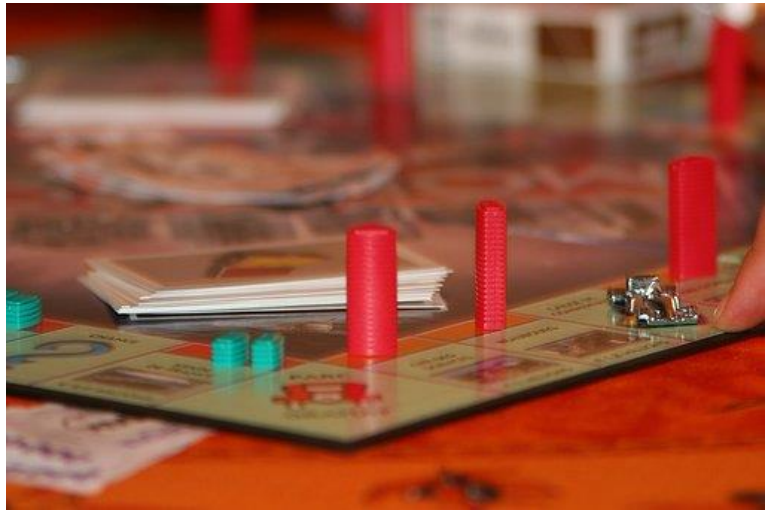
Capstone Project 2

Prediction of board game rates based on their reviews

Diane Deroualle – October 2019

Problem Statement



- Board games have regained popularity in recent years

- **Purpose: build a model with machine learning and natural language processing to predict the rates of board games considering the reviews of players, the number of players, the average time of a game, the number of rates… ….**

Dataset

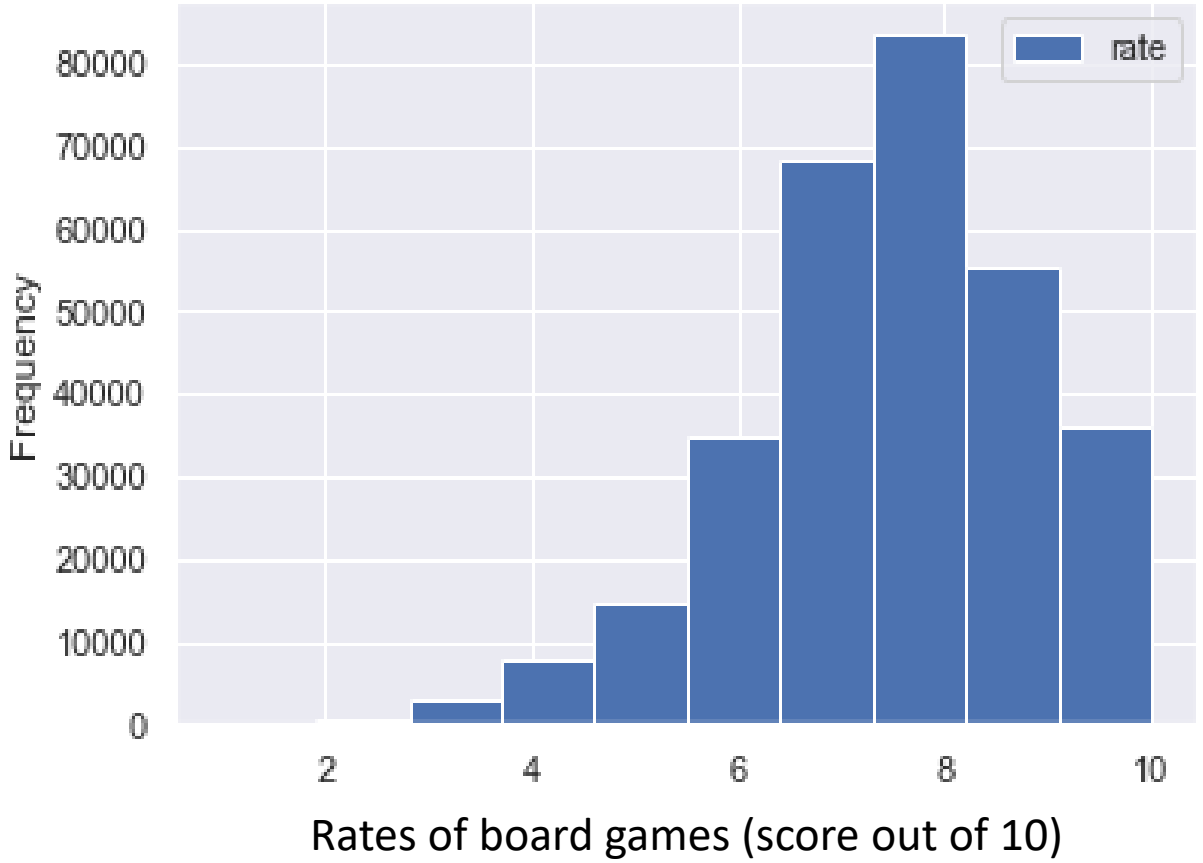
- Scraping and APIs on the website: https://boardgamegeek.com

- For each of the 50 most rated boardgames:

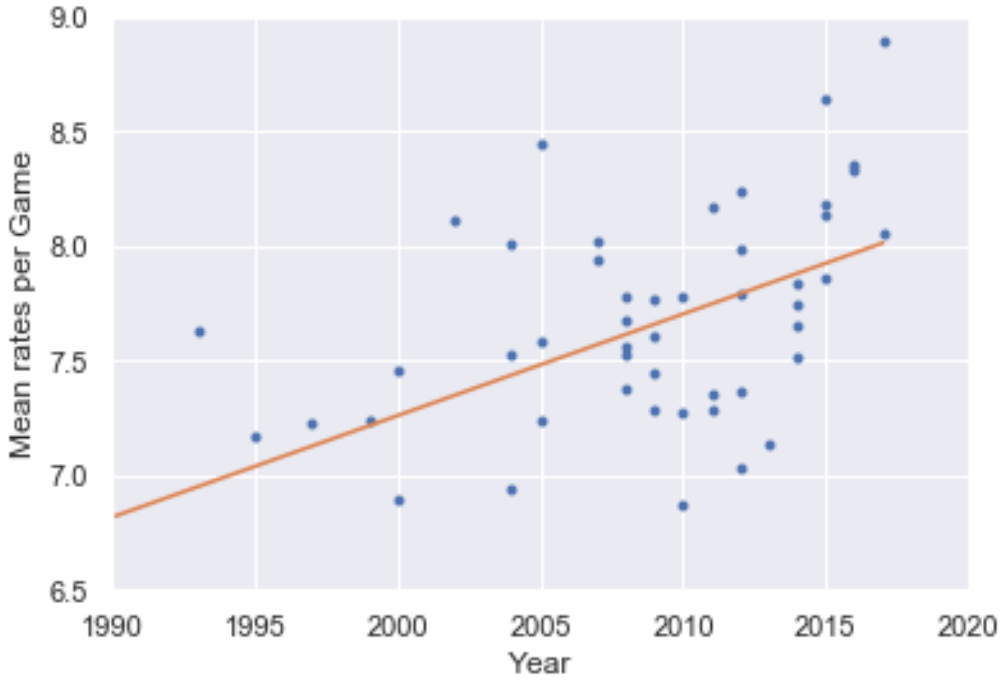
  - ID
  - Name
  - Year of design
  - Minimum and maximum number of players required
  - Minimum and maximum number of minutes required to complete the game
  - Minimum age required
  - Category
  - Number of rates
  - Username of players
  - Reviews
  - Rates (score out of 10)

- Dataset with 304864 raws

Distribution of all the rates

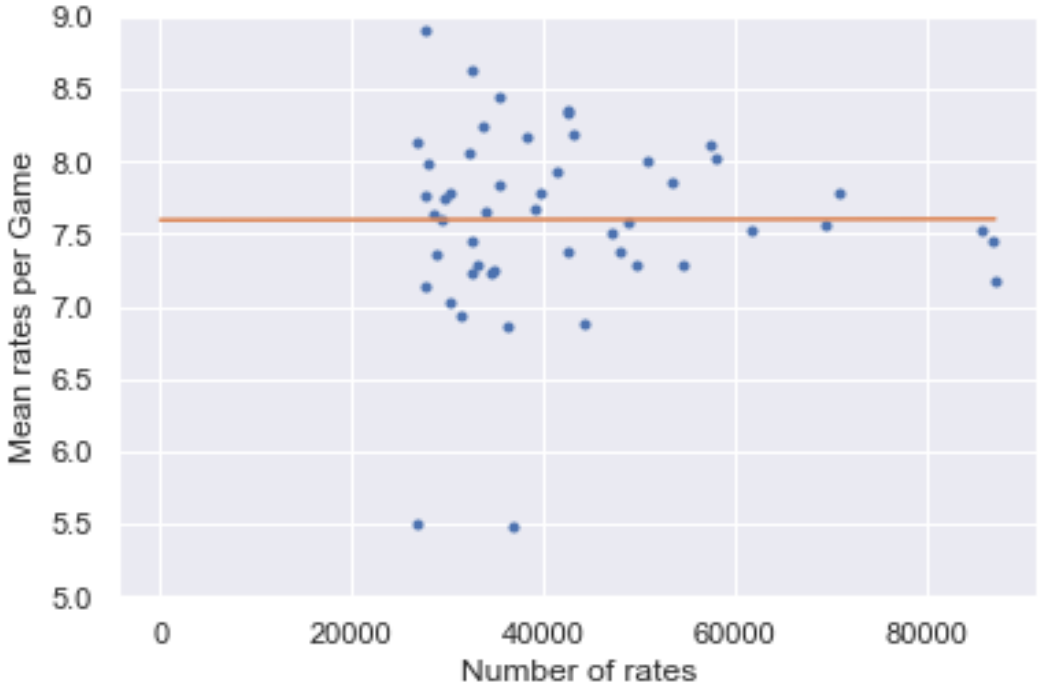


Relation between year of design and mean rates per game



- **Positive correlation between the *year of the design of a game* and the *mean of rates of this game*. More a boardgame is recent and more he seems to have higher mean rates.**

Relation between the mean rates per boardgame and the number of rates per games



- No correlation between the *number of rates of a game* and the *mean of rates of this game*.
- No bias in the rates related to the fact we took the 50 most rated games from boardgamegeek.com.