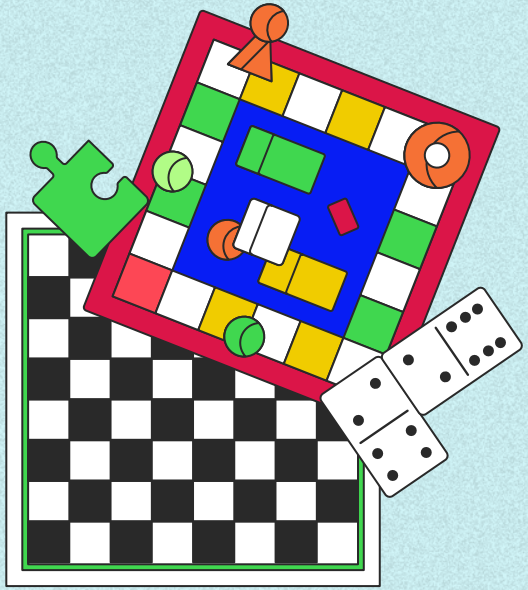
Board Games Analysis



Michael ELBAZ

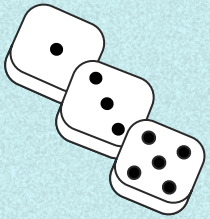
My goals

- Analyze the criteria relative to board-gaming industry
- Observe and determine what are the key factors influencing the rating of a board game
- Create a prediction model to estimate the chances of success of a board game

Project Management

On Notion :

Data gathering



- **Flat files**:
  - One big dataset (games)
  - Binary datasets (values =0/1)

[themes, artists, mechanics, user ratings, publishers, designers, subcategories]

- **API**:
  BoardGameGeek API

- **Web scrapping**:
  BoardGameGeek > boardgame browsing area

EDA & Data cleaning

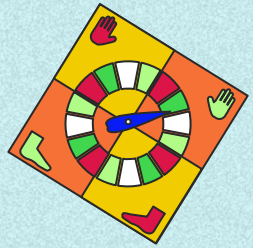
Initial shape
21925 rows / 48 columns

After cleaning
16832 rows / 21 columns

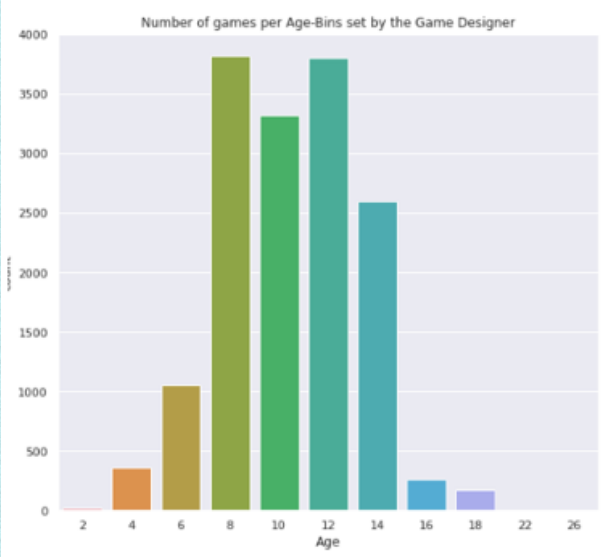
- ❑ Checking for data shapes and data types
- ❑ Handling missing values
- ❑ Checking for duplicates
- ❑ Removed useless columns
- ❑ Removed outliers on certain features (minimum age , Year published, Playtime > 10H)
- ❑ Renamed certain columns
- ❑ Normalization / formatting of all columns (lowercase, special characters,…)

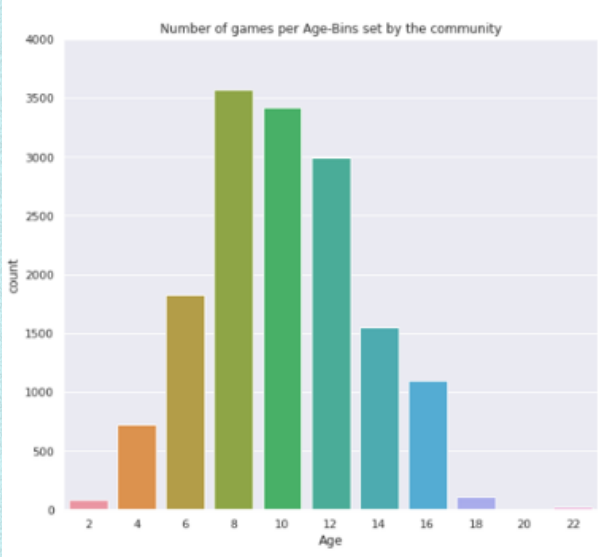
	bgg_id	name	description	year_published	game_difficulty	rating	min_players	max_players	co
1	2	Dragonmaster	dragonmaster tricktaking card game base old ga...	1981	1.9630	6.64537	3	4	
2	3	Samurai	samurai set medieval japan player compete gain...	1998	2.4859	7.45601	2	4	
3	4	Tal der Könige	triangular box luxurious large block tal der k...	1992	2.6667	6.60006	2	4	
7	8	Lords of Creation	interesting offering warfrog player god seek d...	1993	2.4000	6.10716	2	5	

Visualization

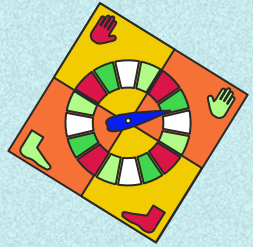


Recommended minimum age repartition



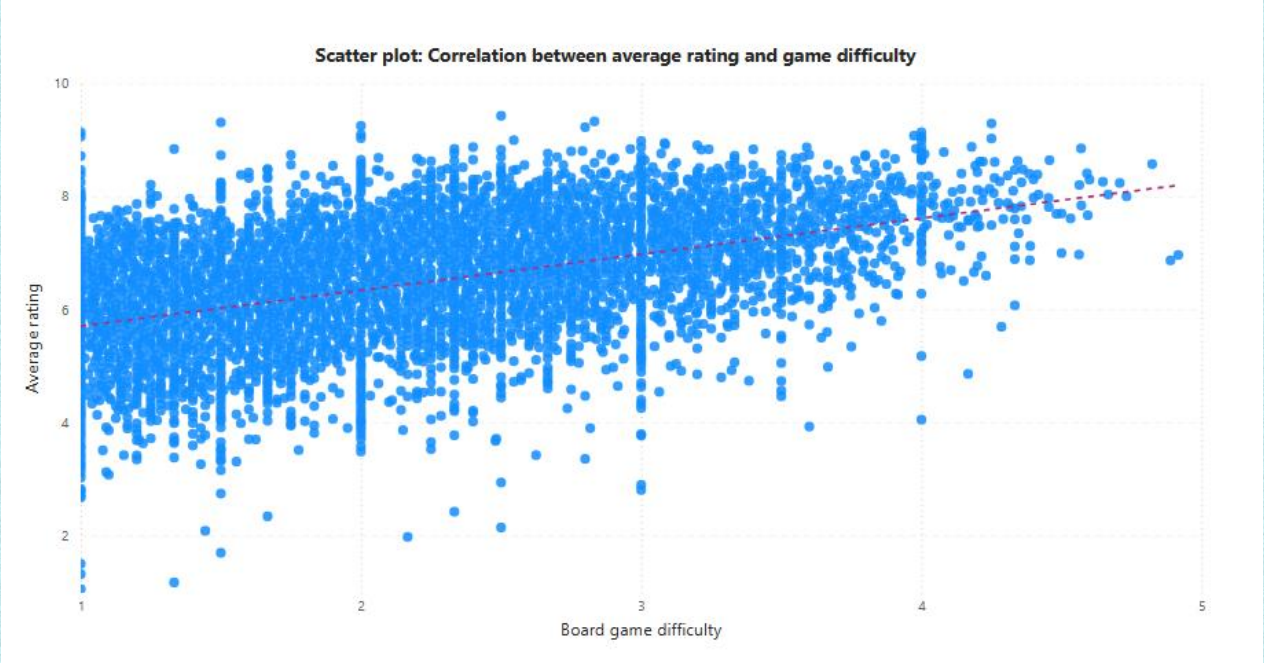


Visualization

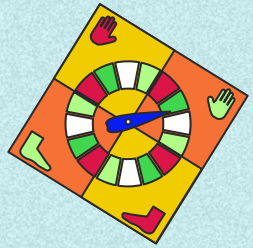


Game difficulty

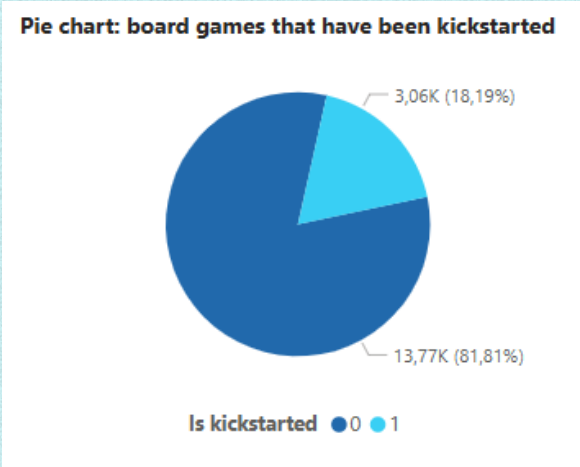
```
Correlation between GameDifficulty and Rating: 0.513108027942937
```

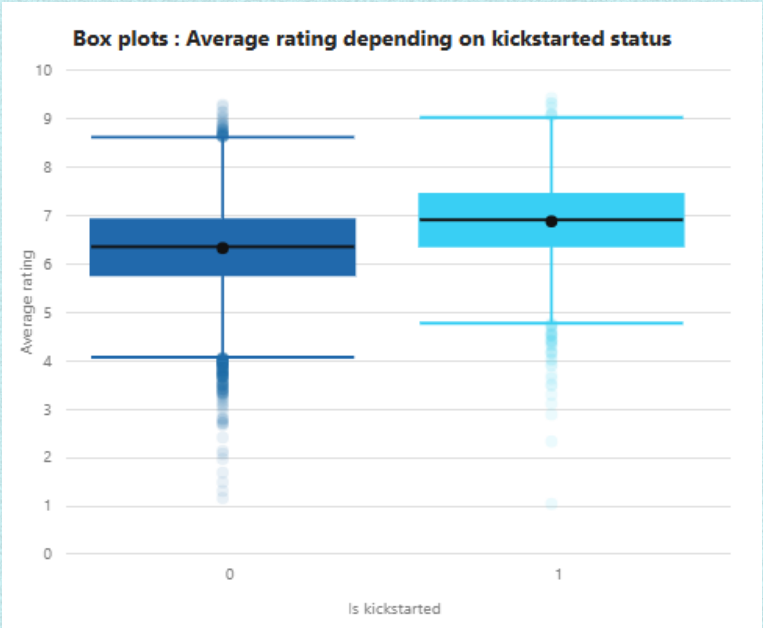


Visualization

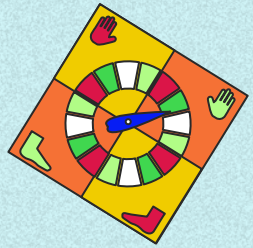


Kickstarter status (True or False)

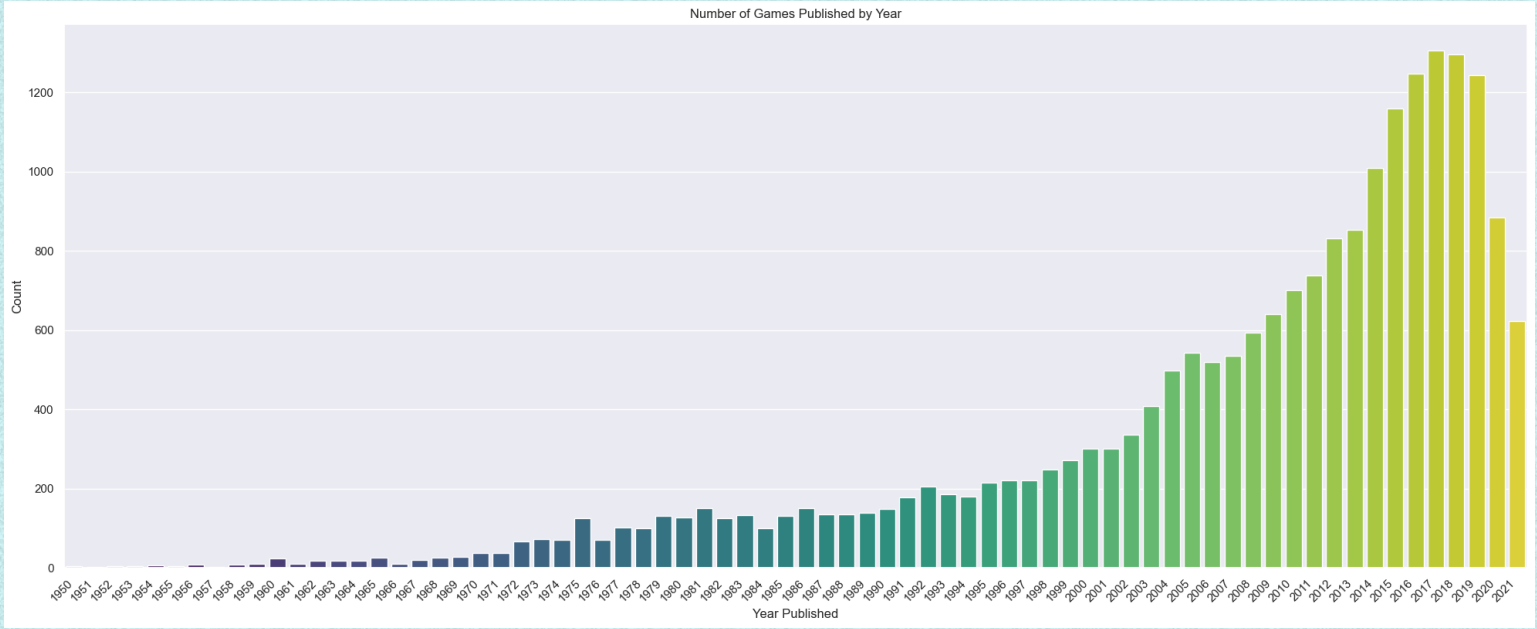




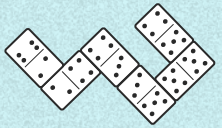
Visualization



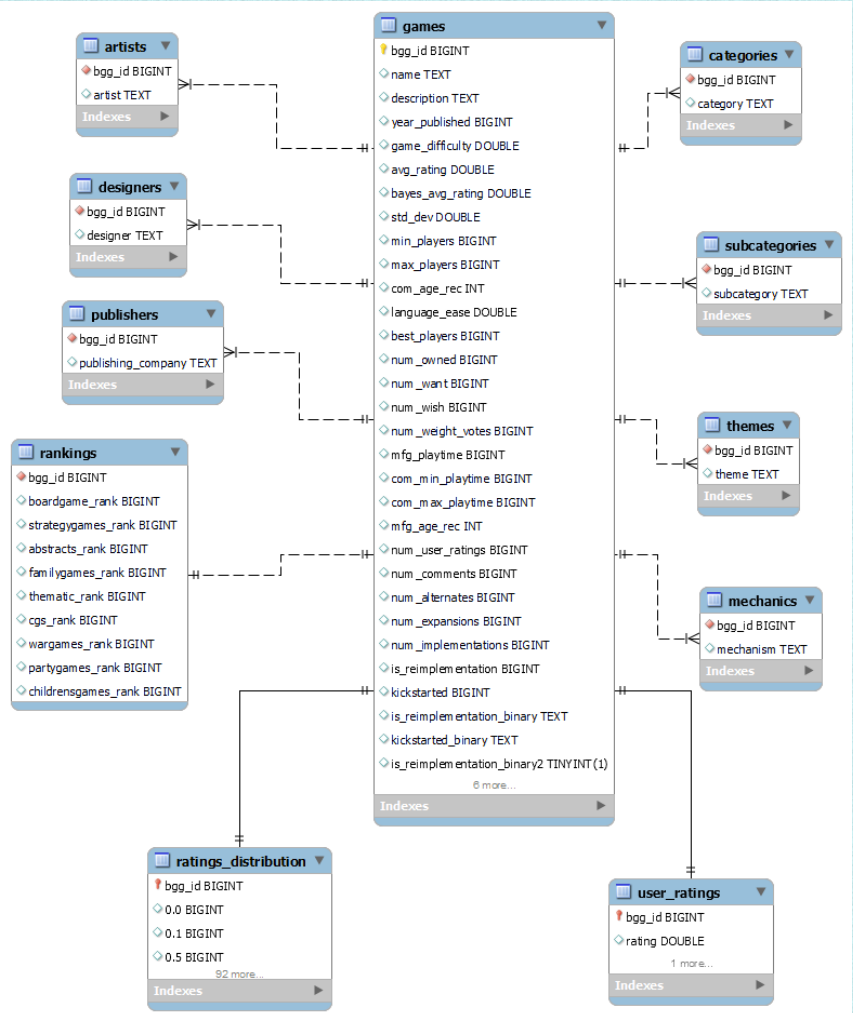
Year published



Database schema



- Connection from Python to MySQL
- Joins on bgg_id (common key to each dataset)
- ERD creation



Entity Relationship Diagram

- Star schema
- ‘games’ is the Fact table
- Other tables (binary columns) are the Dimension tables

Example of SQL Queries

Top board games in the 90's

```sql
-- top 10 best games in the 90's
SELECT
    g.bgg_id,
    g.name,
    rd.average_rating,
    rd.total_ratings,
    t.theme
FROM bgg_project.games AS g
LEFT JOIN bgg_project.ratings_distribution AS rd USING (bgg_id)
LEFT JOIN bgg_project.themes_improved AS t USING (bgg_id)
WHERE rd.total_ratings > 200
AND g.year_published >= 1990 AND g.year_published < 2001
ORDER BY rating desc
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

bgg_id	name	average_rating	total_ratings	theme
234	Hannibal: Rome vs. Carthage	7.8	4851	Ancient, Political
93	El Grande	7.8	24288	Renaissance, Medieval
4214	Stonewall in the Valley	7.74	318	American Civil War
490	Warangel	7.71	263	Fantasy, Mythology
42	Tigris & Euphrates	7.76	25960	Civilization, Ancient
215	Tichu	7.67	13710	NULL
939	Star Wars: The Queen's Gambit	7.56	2131	Fighting, Science Fiction, Movies / TV / Radio th...
463	Magic: The Gathering	7.63	33935	Fantasy, Fighting
552	Bus	7.56	2487	Transportation, Theme_Time Travel
555	The Princes of Florence	7.56	14732	Renaissance, City Building, Theme_Art

Top publishers between 1990 and 2021

```sql
-- Top 10 publishers between 190 and 2021
SELECT
    p.publisher,
    COUNT(g.bgg_id) AS total_games
FROM df_games_before_cleaning g
LEFT JOIN bgg_project.publishers_improved p ON g.bgg_id = p.bgg_id
WHERE p.publisher IS NOT NULL
AND g.year_published >= 1980 AND g.year_published < 2022
GROUP BY p.publisher
ORDER BY total_games DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

publisher	total_games
Low-Exp Publisher	540
SPI (Simulations Publications, Inc.)	72
GMT Games	72
Decision Games (I)	62
The Avalon Hill Game Co	57
Ravensburger	56
(Self-Published)	44
3W (World Wide Wargames)	40
Games Workshop Ltd.	37
Milton Bradley	35

Example of SQL Queries

Most present BG themes

```sql
-- What are the most popular themes
SELECT
    t.theme,
    COUNT(g.bgg_id) AS total_games
FROM bgg_project.games g
LEFT JOIN bgg_project.themes_improved t ON g.bgg_id = t.bgg_id
WHERE t.theme IS NOT NULL
GROUP BY t.theme
ORDER BY total_games DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch

theme	total_games
Fantasy	607
Animals	541
World War II	351
Science Fiction	316
Medieval	242
Humor	233
Economic	224
Trivia	218
Ancient	188
Fantasy, Fighting	184

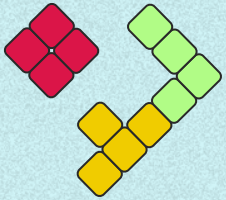
Number of games by year

```sql
-- Evolution of the  number of BG published over years
SELECT
    year_published,
    COUNT(*) AS total_games
FROM bgg_project.games AS g
WHERE year_published IS NOT NULL
GROUP BY year_published
ORDER BY year_published DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

year_published	total_games
2021	561
2020	796
2019	1124
2018	1164
2017	1158
2016	1104
2015	1014
2014	889

Example of SQL Queries



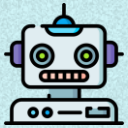
Some publishers rating

```sql
-- Some publishers average rating
SELECT
    p.publisher,
    ROUND(AVG(g.rating), 2) AS avg_rating,
    SUM(rd.total_ratings)
FROM bgg_project.games g
JOIN bgg_project.publishers_improved p ON g.bgg_id = p.bgg_id
JOIN bgg_project.ratings_distribution AS rd ON g.bgg_id = rd.bgg_id
WHERE p.publisher IN ('Low-Exp Publisher', '3M', 'Asmodee', 'Ravensburger', 'Hasbro', 'Avalon Hill Games, Inc.')
GROUP BY p.publisher;
```

publisher	avg_rating	SUM(rd.total_ratings)
Low-Exp Publisher	5.87	265325
Ravensburger	5.7	26337
3M	5.94	2023
Avalon Hill Games, Inc.	6.29	8603
Hasbro	6.01	22398
Asmodee	6.65	453

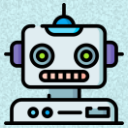
Flask API

flask --app hello_api run --port 8080



- Resources from MySQL queries
- Endpoints
  - http://127.0.0.1:8080
  - http://127.0.0.1:8080/boardgames
  - http://127.0.0.1:8080/bdgaoarmes/<int:bgg_id>
  - http://127.0.0.1:8080/boardgames/<int:bgg_id>/details
  - http://127.0.0.1:8080/boardgames/kickstarted

Flask API

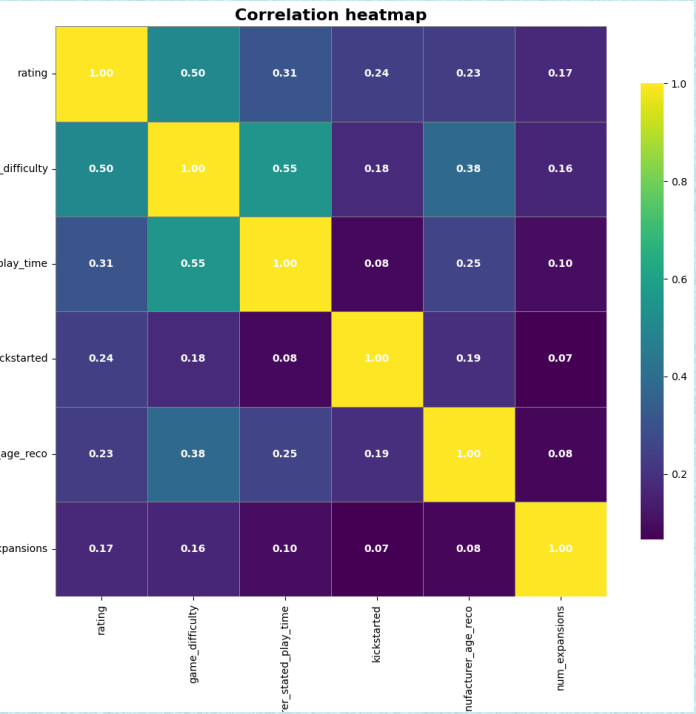



```
{
  "artists": null,
  "bgg_id": 260789,
  "community_age_reco": 12.0,
  "community_max_playtime": 90,
  "community_min_playtime": 30,
  "description": "assassin creed brotherhood venice cooperative tactical board game
see assassin campaign develop headquarters advance brotherhood agenda templarsplayer
publisher",
  "image_path": "https://cf.geekdo-images.com/WdfVXXWgbBDA9aaTr63M5Q__original/img/7
  "kickstarted": 1,
  "manufacturer_age_reco": 12,
  "manufacturer_playtime_reco": 90,
  "mechanic": "Action Points, Modular Board, Variable Player Powers, Scenario / Miss
  "name": "Assassin's Creed: Brotherhood of Venice",
  "publisher": null,
  "theme": "Renaissance, Video Game Theme"
}
```

```
{
  "bgg_id": 308119,
  "name": "Pax Renaissance: 2nd Edition",
  "num_expansions": 0,
  "num_implementations": 1,
  "year_published": 2021
},
{
  "bgg_id": 307862,
  "name": "Dollars to Donuts",
  "num_expansions": 1,
  "num_implementations": 0,
  "year_published": 2021
}
,
last_page": "/boardgames/kickstarted?page=33&page_size=100",
next_page": "/boardgames/kickstarted?page=2&page_size=100",
previous_page": "/boardgames/kickstarted?page=0&page_size=100",
```

Machine Learning

correlation heatmap, variables, p-values



Variable to predict (y) : Rating

Selected features (X) :

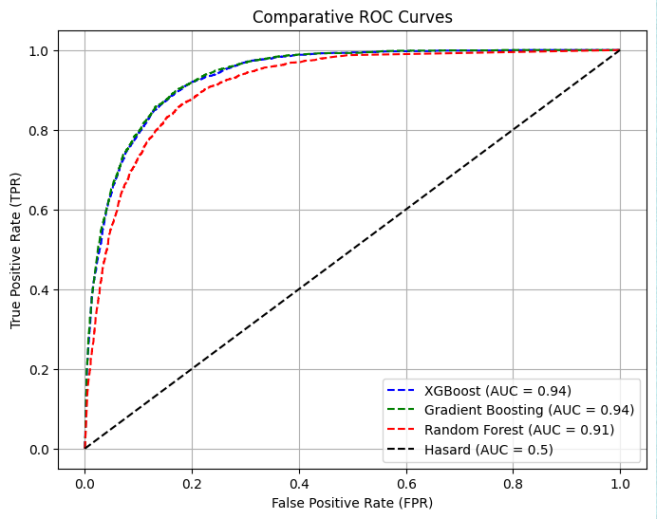
- Game difficulty
- Manufacturer stated play time
- Manufacturer age recommended
- Kickstarted status
- Number of expansions

p-values :
significant and have an impact on games rating

game_difficulty	0.000000e+00
manufacturer_stated_play_time	2.298157e-03
kickstarted	2.567740e-87
manufacturer_age_reco	1.048979e-02
num_expansions	6.235811e-29

ML : comparisons

	Random Forest	XG Boost	Gradient Boosting
Accuracy	93,17 %	79,47 %	79,66 %



Main results

- Random forest seems to be an effective predictive model to estimate the rating of a board game
- Game difficulty, playtime, age are “basic” information and have a real impact on the success of a board game
- Other additional parameter like the birth of a game on Kickstarter platform, the number of expansions also say something about the notoriety of a board game

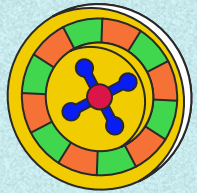
Next step and challenges

🔎 **To go further**:

- ❑ Hyperparameter tuning:
  Could allow improvement of the model
- ❑ Find additional data like the amount of purchase of board games to enrich my database and lead deeper analysis.
- ❑ Conduct a Clustering model to find the (nearest) existing game (present in base) when entering board game features.

🏋️ **Challenges**:

- ❑ Data collection :
  Especially for including information about publishing company size or designer notoriety to observe more clearly their impact on the average rating
- ❑ Time management
- ❑ Big challenge:
  Write the report and prepare the presentation 😉



Demo ?


Thank you
"The true joy of a game is found not in winning or losing, but in the shared experience of playing."
Unknown