What about board games?

Matteo Tarenzi

1 Introduction

Board games are becoming more and more popular not only among children but also among adults. They provide not only entertainment but also intellectual stimulation and collaboration/competition. While there is an ever-growing number of games available, many share common characteristics. These elements can define a game and help establish a general framework for board games. How these elements are used and combined plays a crucial role in a game's success from the players' perspective. Therefore, it could be valuable, to perform a study relative to some of these aspects, to comprehend how players react when they are introduced in different ways.

2 Research question

When reviewing or commenting a game, the players tend to focus on certain aspects. These features, can be defined as the fundamentals for every board game and can include gameplay elements and mechanics. The project's goal is to understand the use of some main games elements. In particular, the final objective is to obtain the player base sentiment associated with this defined set of features. For each aspect, the project will determine the polarization of that feature, and if possible, draw a general opinion that can be used to understand how the players experience them.

- **Luck**
- **Downtime**
- **Bookkeeping**
- **Interaction**
- **Complex and Complicated**
- **Bash the leader**

2.1 Project's approach

The workflow can be divided into four phases: the creation of the dataset, the mining of the aspects, the training of the model and finally the aspects insights. Each phase will be presented rapidly in this paragraph and deeply described in the *Methodology* section.
The first operation to perform has been obtaining the data regarding the board games. The popular Board Games Geek, an online community and resource regarding board games content provides access to his database. Taking advantage of the API then, it has been possible to download comments and other useful statistics relative to a large number of games. After that, to gather the possible aspects and sentiments for the study, it has been performed an Aspect Based Sentiment Analysis. The third step of the project has focused on implementing a classification model to use its explainability to obtain more information and comprehension of the aspects data. Finally, in the fourth phase, using the information and the results obtained before, it has been possible to aggregate aspects and draw some results for each of the aspect of interest. Every aspect has shared a similar pattern in the operations implemented within this phase. However, some of studies have included some extra tests that improved the comprehension of players opinions.

Within the goal set by the project, two key difficulties emerge. On top of everything, it is difficult to perform a sentiment analysis when the aspects of interest are so much related to the environment. The project will show, in fact, that popular packages struggles or fail when performing the sentiment analysis. Another significant issue lies in the construction of an opinion, which is inherently more complex and varied than mere sentiment. Understanding opinions may require a human touch, as it involves interpretation and context that automated tools may not adequately address.

3 Methodology

The following paragraphs will present in details the operations and the ideas behind the project. Each of them represent a step in the workflow, starting with the data collection and ending with the aspect opinion extraction.

3.1 Building the dataset

The construction of the dataset has been done using the resources shared by Board Games Geek. BGG gives access to their database through an API that can be used to collect the comments of users regarding various board games along with information and useful statistics. Collected data has been cleaned from undesired languages and finally filtered over some parameters that ensure that games were suitable for the project . A dictionary of board games has finally been created, it contains different statistics and metrics along with the comments and ratings.

3.2 Aspect mining and Sentiment extraction

Once that the textual data has been obtained it was time to perform the aspect mining. For this purpose, a syntactical approach has been choose. Starting by the individuation of nouns as possible aspects, the adjectives has been detected exploiting the syntax of each comment. Three function have been created to extract different types of connection between adjectives and nouns. Once that the nouns have been mined, the first function retrieves those adjectives that are directly connected to the noun. The second one is focused on finding those that are connected to the possible aspects through a verb. The last instead, operates over the 'of-and' construction. (ex. 'A good mix of luck and strategy). It has been necessary to face also other difficulties in the extraction like compound names and negations. Ended the mining, aspect-opinions pairs have been stored in a dataframe along with board games data (categories, review score and the sentence were the aspect were found). For each pair, the PMI has been computed and added to the dataframe. Next, the aspects have been filtered to eliminate punctuation among aspects and those pairs with a low PMI retaining only those nouns-opinions that had higher frequency respect to the expectations. The sentiment extraction has been computed at three level: opinion, sentence and whole comment. First, the sentiment has been extrapolated using *sentiwordnet* combining the sentiment metrics of the synsets coming from the word composing the adjectives. Using *VADER* then, it has been possible to compute the sentiment for the sentences associated to the pairs. This time, the polarity has been documented using three values that measure the intensity of that sentiment: *Sentence neg, Sentence neu, Sentence pos.* For the comment polarity, instead, it has been choose to use the review score given by the users even if this metric might involve some elements that were not mentioned by the user.
Parallel to the extraction, word embeddings have been created. The analyzed comments have been used to obtain a vector representation of the terms. These word embeddings will then be employed in the fourth phase to help with the aggregation.

3.3 Born Classifier

This part was focused on the creation of an algorithm for classifying the polarity of comments. More than the final predicting capabilities however, it was more interesting to observe how the model makes the decisions. In other words, for the purposes of this project, the ability to clearly explain and interpret the classification results holds greater importance than the final performance metrics. The focus is on ensuring that the decision-making process is transparent and understandable, even if it comes at the expense of optimizing for the highest possible accuracy or efficiency. Because of that, the Born classifier has been chosen precisely for his capabilities. Before training the model, however, it was necessary to build a dataset suitable for the classifier. A sparse matrix has been created from scratch. With comments as records and aspect-opinion pairs as the features; the values has been set using the *PMI*. If a feature(pair) was present in the comment, then a combination of *PMI* and *sentence polarity* was given, otherwise a default value was associated. The output variable has been created using the score of each comment and associating a categorical variable based on this schema:

Score less than 5.5	**Score between (5.5 , 7)**	**Score more than 7**
Negative	Neutral	Positive

Once that the model has been trained and tested, Born results have been observed to understand which pairs were the most important to the positive and negative classes. Moreover, the results have been used to obtain an initial idea of nouns-adjectives polarity.

3.4 Aspects aggregation and opinion crafting

Once that the previous results have been extrapolated, it was possible to move on to the fourth step of the project focusing the attention on the aspects of interest. To obtain the best results, it has been necessary to modify the approach to adapt it to each aspect. However, even if there are some differences, they all shares the same general pattern.
First, the embedding of the aspects has been exploited to derive a general meaning and some similar words that allows to create a vector of words. This vector of words has been used to lead the study instead of the only using the literal term. Born results associated with the pairs have been investigated to understand which of the adjectives were more important to determine a negative or a positive polarity. After that it has been constructed a summary dataframe that contains all the sentiment metrics that regard the pairs of interest. For each pair, the three metrics were aggregated using the mean to draw the general opinion towards that aspect. Finally, to create general polarity metrics, the statistics have been aggregated again. This time, instead of the simple mean it has been used a weighted mean with the frequency of the pairs as weights. Since most of the changes to this framework are very slight, they will not be presented in the paper. However, one of the aspects deserves a further investigation: *Complex vs Complicated*. The original dictionary that has been constructed from BGG database contains one important very metric: the game complexity. Along with other statistics, this can be exploited to better comprehend the aspect. For this purpose two different studies have been deployed. One had the goal of understanding whether or not there are differences between the complexity level of different game board categories. The second study instead, combines also the other statistics associated to the game to compute the correlation between them and the complexity level.

4 Results

This section will present the results obtained from the project. First, they will be presented the performance and the other information that can be derive from the Born classifier. In the second part of the section, the sentiment and the opinion of each aspect of interest will be investigated and commented using the comprehension derived during the fourth step of the project.

4.1 Born results

Model performances are not particularly satisfactory. The use of a sparse matrix of *PMI* and *sentence polarity* to predict sentiment does not demonstrate strong classification performance. Both precision and recall are quite bad. In particular, looking at **Table 1**, it is possible to observe that although the model can do slightly better for positive and negative classes, neutral comments are hardly poorly classified by the model, that does not surpass the random classification.
Despite these results, it is still possible to identify the most important features for both the positive and negative classes in the other table. Specifically, regarding the negative class, it is already possible to provide a preview of the subsequent findings.

Table 1

Class	Precision	Recall	F1-Score
Negative	0.42	0.51	0.46
Neutral	0.31	0.34	0.32
Positive	0.65	0.54	0.59
Accuracy		0.48	
Macro Avg	0.46	0.46	0.46
Weighted Avg	0.50	0.48	0.49

Feature	Positive	Neutral	Negative
(expansion, great)	0.024125	0.007345	0.006736
(classic, true)	0.021213	0.006398	0.001774
(player game, great)	0.019393	0.009452	0.002360
(wargame, favorite)	0.019366	0.000000	0.000000
(fun, great)	0.018311	0.012326	0.007687
(auction game, great)	0.015581	0.004629	0.000000
(racing game, best)	0.015346	0.005657	0.002504
Feature	**Positive**	**Neutral**	**Negative**
(disappointment, big)	0.001308	0.004811	0.019318
(gameplay, boring)	0.000000	0.006206	0.016243
(control, very little)	0.002797	0.004193	0.016015
(trivia game, bad)	0.000000	0.000000	0.015088
(execution, poor)	0.002069	0.000000	0.014388
(fun, not much)	0.003252	0.007729	0.014327
(experience, bad)	0.003927	0.004013	0.014313

4.2 Aspects discoveries

4.2.1 Luck

Based on the word embedding similarity it has been chosen to aggregate the term luck with the terms chaos and randomness. The tables below show the sentiment of noun-adjective aggregation . he first point to note is that synsets and sentence polarity are not effective at capturing the sentiment of aspects, particularly when it comes to negative emotions. For that reason it would be better to rely on the *Mean Stars* metric to investigate the opinions. Users tends to prefer games where chaos, luck or randomness has little impact. When looking at the worst reviews that includes any of these elements, in fact, results show that all of the comments mention an overwhelming amount. This, however, does not mean that the players do not enjoy a bit of luck when they play but they love to have it well balanced by some other factor or perhaps combined with something more skill-driven.

Chaos sentiment

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
chaos	enough	7.0	0.0833	0.1707	0.6135	0.2159	6.941
chaos	less	10.0	-0.1458	0.1961	0.6266	0.1770	6.823
chaos	sheer	6.0	-0.0625	0.1692	0.6786	0.1524	6.559
...	...	...	...	...	...	...	...
chaos	just much	9.0	0.2188	0.2877	0.6421	0.0702	5.000
chaos	total	23.0	0.0000	0.3753	0.4796	0.1450	4.976
chaos	way much	6.0	0.0673	0.3882	0.5634	0.0483	4.833

Luck sentiment

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
luck_factor	nice blend	6.0	0.2500	0.0000	0.4511	0.5489	8.286
luck_factor	nice balance	8.0	0.2396	0.0000	0.4865	0.5135	8.232
luck_factor	good combination	7.0	0.2824	0.0000	0.5048	0.4952	8.053
...	...	...	...	...	...	...	...
luck_factor	total	12.0	0.0000	0.0187	0.6401	0.3412	3.917
luck_factor	complete	12.0	0.3000	0.0190	0.6388	0.3423	3.847
luck_factor	uncontrollable	6.0	-0.4688	0.2163	0.5988	0.1849	3.786

Randomness sentiment

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
randomness	very little	5.0	0.0281	0.0238	0.8848	0.0915	7.688
randomness	just enough	22.0	0.1979	0.0271	0.8143	0.1586	7.599
randomness	little	32.0	-0.2250	0.0378	0.8023	0.1599	7.443
...	...	...	...	...	...	...	
randomness	even more	9.0	0.0134	0.0528	0.8918	0.0554	3.796
randomness	way much	6.0	0.0673	0.0407	0.8701	0.0892	3.810
randomness	pure	6.0	0.1607	0.0355	0.8243	0.1402	3.692

The table on the right shows the nouns that are more often associated with this aspect. This has been done finding those pairs associated with the adjective 'lucky' and sorting them by their frequency. In particular, it is possible to confirm that the mechanics that introduce more luck within a board games are the dice rolls and the card draws.

Object	**Adj**	**Count**
roll	lucky	56.0
draw	lucky	35.0
player	lucky	30.0
card	lucky	23.0
die rolls	lucky	16.0

4.2.2 Downtime

This aspect involves all those elements that have to do with the waiting and the wasting of time including the paralysis that players could face in light of too many choices or possible implications. Like the luck, games with less downtime are preferred to those that have lots of it. However, whether players do not mind or enjoy to have little alea inside their sessions, results show that downtime is associated with a pure negative opinion. This suggests that players might enjoy a more rapid play style rather than a slow pace. Further results about complexity however, describe how people are not scared by the length of the playtime but only by the unproductive time.

Most positive and negative opinions

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
downtime	low	8.0	-0.2500	0.2504	0.6680	0.0817	8.098
downtime	very little	30.0	0.0281	0.0206	0.8387	0.1407	8.035
downtime	little	47.0	-0.2250	0.0376	0.8127	0.1497	6.916
analysis_paralysis	little	13.0	-0.2250	0.0391	0.8418	0.1191	6.783
analysis_paralysis	major	5.0	0.0481	0.0749	0.8425	0.0825	6.727

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
downtime	massive	5.0	0.0000	0.1078	0.7611	0.1311	3.833
downtime	excessive	12.0	-0.3125	0.0648	0.8679	0.0673	5.215
downtime	huge	6.0	-0.1250	0.0177	0.7174	0.2648	5.392
downtime	big	9.0	0.0809	0.1257	0.7751	0.0992	5.440
analysis_paralysis	heavy	5.0	-0.0542	0.1675	0.7085	0.1238	5.462

The table on the right contains the categories of games where the downtime is mentioned more often and then, consequently, those that are more likely to experience it . On top there are wargames, economics and fantasies, all games that requires a lot of analysis and usually some preparation.

Game Type	Score
Wargame	83
Economic	75
Fantasy	55
Card game	48
Exploration	43

4.2.3 Bookkeeping

This aspect might be related to the previous one since writing down gameplay elements and consulting the manual could be easily seen as a sort of downtime . Because of that, bookkeeping shares the same polarity. Those games that require less effort are more likely to be appreciated while, on the other hand, repeatedly looking at the manual and regularly keeping note discourage players.

Bookkeeping sentiment

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
bookkeeping	tedious	6.0	-0.1250	0.0000	0.9336	0.0664	4.062
paperwork	too much	7.0	0.0312	0.0690	0.8520	0.0794	4.308
bookkeeping	way much	6.0	0.0673	0.0690	0.8195	0.1115	4.350
bookkeeping	too much	27.0	0.0312	0.0496	0.8303	0.1201	5.198
bookkeeping	much	5.0	0.1250	0.0795	0.8548	0.0658	5.875
paperwork	less	5.0	-0.1458	0.0000	0.7969	0.2031	6.125
bookkeeping	little	8.0	-0.2250	0.0767	0.7757	0.1477	6.865
bookkeeping	more	7.0	0.0000	0.0534	0.8265	0.1203	7.571

Again, the category that is mostly associated with bookkeeping is wargame. Due to the nature of these games, in fact, the player is required to look frequently at the statistics about his troops and take track of wounds and victory points.

Game Type	Score
Wargame	39
Science fiction	15
World War II	10
Economic	10
Negotiation	8

4.2.4 Interaction

A game is consider to be interactive when the gameplay or some element requires players to relate. This could be done by forcing communication and debate, increasing the cooperation or instead, heating the competition. Another way of being interactive, is including some aspect of the game that involves a different approach by the players like using their mobile phone or utilize some gadgets. Opposite to the previous aspects, all of these possibilities are largely appreciated by players. Results shows that players love to interact with others or directly affect the progression of a game. On the contrary, games where the play style is quite passive and their actions do not have enough engagement are not perceived well.

Most positive and negative opinions

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
interaction	human	5.0	0.0000	0.0846	0.7953	0.1201	8.038
influence	major	5.0	0.0481	0.0265	0.8997	0.0738	7.860
interaction	direct	35.0	0.0365	0.0509	0.8005	0.1486	7.356
interaction	high	21.0	0.0556	0.0628	0.7596	0.1776	7.096
interaction	complex	12.0	-0.0250	0.0148	0.7994	0.1859	7.089

Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
influence	little	9.0	-0.2250	0.0295	0.8960	0.0745	4.582
interaction	only	10.0	-0.0278	0.0757	0.8628	0.0615	4.839
influence	much	10.0	0.1250	0.0344	0.9072	0.0584	4.971
interaction	not enough	12.0	-0.2708	0.0510	0.8871	0.0618	5.136
influence	limited	42.0	0.0000	0.2201	0.3988	0.3814	5.267

As for the Luck, the table on the right shows those nouns that are more often associated with the interactivity. From the results it can be derive that interaction could be perceived at the general game level down to the individual elements. This suggests that there a lot of ways to increase the interactivity of a board game to make it more entertaining to the players.

Object	**Adj**	**Count**
game	interactive	59.0
fun	interactive	14.0
element	interactive	10.0
play	interactive	8.0
gameplay	interactive	6.0

4.2.5 Bash the leader

Bash the leader sentiment

category	noun	adj	Synset_pol	Sent_neg	Sent_neu	Sent_pos	stars
Age of Reason	bash_the_leader	bad	-0.595588	0.260	0.563	0.177	9.0
Card Game	bash_the_leader	overwhelming	-0.104167	0.000	1.000	0.000	6.5
Card Game	bash_the_leader	fun game	-0.051339	0.000	0.784	0.216	6.0
Card Game	bash_the_leader	strong	0.075000	0.000	0.696	0.304	6.0
Card Game	bash_the_leader	huge	-0.125000	0.218	0.647	0.135	6.0
Card Game	bash_the_leader	too much	0.031250	0.000	1.000	0.000	5.0
Card Game	bash_the_leader	too much	0.031250	0.000	1.000	0.000	5.0
Card Game	bash_the_leader	too much	0.031250	0.000	1.000	0.000	4.0
Card Game	bash_the_leader	severe	-0.270833	0.418	0.448	0.134	4.0
Pirates	bash_the_leader	mindless	-0.225000	0.326	0.674	0.000	4.0
Card Game	bash_the_leader	original	0.020833	0.000	0.685	0.315	3.0

Different from the other aspects, bashing is a particular mechanic that is not so frequent among the board games. From all the three hundred thousand comments, only eleven times the bashing has been mentioned, meaning that is something quite specific. The category of card games is more incline to experience this element. The results, in addition, shows something a bit unexpected about the sentiment. In fact, although the general polarity is negative, there are some comment that associate a medium or high score to strong bashing mechanics. A first theory is that the aspect could be not important enough to largely affects the final score of a comment, remembering also that this is a quite specific mechanic. Another explanation is that a sort of aggressive and chaotic competition, for absurd, could be entertaining for some types of players, especially within the category of card games.

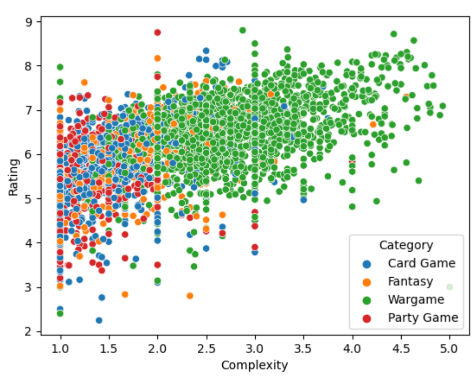
4.2.6 Complex vs Complicated

Complex vs Complicated sentiment

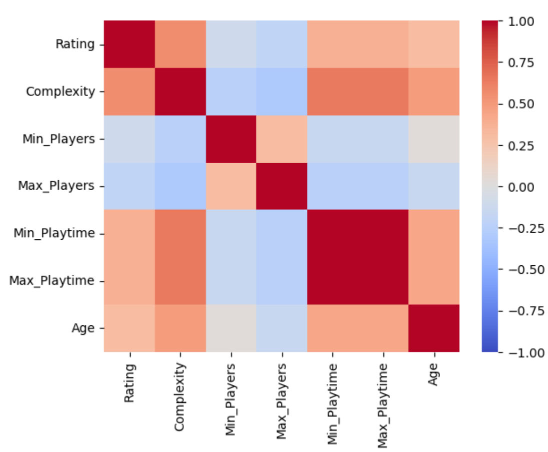
Object	Adj	Count	Synset_pol	Sentence_neg	Sentence_neu	Sentence_pos	Mean_Stars
complexity	right	13.0	0.1736	0.0352	0.8130	0.1519	8.491
complexity	just enough	13.0	0.1979	0.0102	0.8540	0.1358	7.917
complexity	right amount	10.0	0.0600	0.0000	0.8690	0.1310	7.880
complication	additional	6.0	0.0000	0.0000	0.8906	0.1094	7.667
complexity	right level	11.0	0.0868	0.0288	0.7972	0.1739	7.483
...	...	...	...	...	...	...	...
complexity	extra	14.0	0.0179	0.0513	0.8011	0.1476	6.029
complexity	needless	15.0	-0.0714	0.0432	0.8446	0.1121	5.638
complexity	too much	44.0	0.0312	0.0491	0.8198	0.1311	5.489
complexity	unnecessary	24.0	0.0000	0.0429	0.8711	0.0861	5.166
complication	needless	5.0	-0.0714	0.0856	0.7669	0.1475	4.632

Even if complexity and complicatedness are not the same thing, they share the same sentiment. The effort needed to master them is tolerable only if the difficulty is associated with some gain in terms of gameplay depth. For that reason, games with a unnecessary amount of complexity are not appreciated by the players. A well balanced amount, however, could be really pleasant, improving the experience and giving more challenges to those that are seeking them. The table on the right contains those therms that are associated with the complexity but not with the complicatedness delineating the differences between them.

Object	Count
strategy	29.0
side	12.0
interaction	12.0
decision	10.0
gameplay	10.0
wargame	10.0
situation	9.0
combat	7.0
problem	6.0
mechanism	6.0



The graph on the left shows the level of complexity and the score given to a game by each user for some different categories of games. War games are the ones with the maximum grade of complexity, followed by Fantasies, Card games and then Party games. Another thing that could be observe already from this graph is the relation between the complexity and the rating. It seems that there is a slight positive relation between them, meaning that as the complexity increases, also the ratings of the users rise.



Considering the results of the previous graph, it could be interesting to study the correlation between some of the metrics proposed by Board Games Geek. Focusing particularly on the complexity, it is possible to observe a slight positive correlation with the rating, the playtime and, as expected, the suggested age. On contrary, as the minimum and maximum number of players increases the complexity decreases. This can be explained thinking that most of the games that includes a lot of players are party games or games that need to reduce the complexity to avoid downtime.

4.3 Sentiment summary

As presented in the *Methodology*, the final step of the project has been to further aggregate the aspects to obtain some general metrics and final considerations. The table below shows the sentiment scores for each aspect of interest. The sentiment is presented again at three levels: at the adjectives levels, at the sentence level and at comment level. The first information that can be draw from the results is that no aspects of our study is always appreciated by the players, the reason why *Mean Stars* ranges between five and seven. Is not the aspect on his own that is positive or negative, but the context and the way a game makes use of it. After that, it can be noticed again that synsets and sentence polarity struggle when applied to our study. The reason can be detected by the neutrality score. Because aspects are not polarized on their own, the associated adjectives usually express a quantity rather than a direct sentiment. The final information that can be extrapolated from the table is that users tends to complain more about the use of bashing, bookkeeping and luck while instead, they praise interaction and complexity. The downtime, is experiencing an interesting phenomenon: although the aspect is not seen as positive among users, they are more incline to comment about it only when the game has little amount of it, praising the game for that reason.

ASPECT	**Synset_pol**	**Sent_neg**	**Sent_neu**	**Sent_pos**	**Mean_Stars**
LUCK	-0.02	0.08	0.65	0.27	5.90
DOWNTIME	-0.05	0.06	0.83	0.12	6.22
BOOKKEEPING	-0.02	0.05	0.83	0.12	5.48
INTERACTION	-0.05	0.07	0.77	0.16	6.22
BASH THE LEADER	-0.11	0.11	0.77	0.12	5.32
COMPLEX & COMPLICATED	-0.04	0.07	0.81	0.12	6.64

5 Conclusions and possible improvements

Results are satisfactory, but there are also areas of concern. On one hand, the project was able to build a solid comprehension of the aspects and define a general polarity for each of them. Every element has been studied in details, not only achieving the predetermined objective but also providing additional information. On the other hand, however, the failure of the sentiment metrics and the low performance of the classifiers undermine those results. There are several possible reasons. Firstly, since the framework of board games are quite specific, *VADER* and *sentiwordnet* were not able to capture effectively the polarity. Concerning the second issue, the data was likely insufficient for the intended purpose and the construction of the output variable using subjective thresholds has undoubtedly affected its effectiveness. With some modifications, the study could be possibly improved. The classifier could be changed to a more refined model that involves deep learning and could be fed with more data. With the right amount and typology of data moreover, a language model can be fine tuned over the board game topic providing more robust results for classification and sentiment extraction when combined with the ABSA.

Despite the difficulties, the information gathered through the project remains relevant. The knowledge of these aspects could prove valuable when analyzing the economic failures or successes of a board game and may also assist in profiling players or conducting research on specific categories of games.