

# Turtle Games

## Predicting Future Outcomes

Cristina Alonso Robles

July 2023





# Methodology

1. Background/context of the business
2. Analytical Approach, using Python and R
3. Visualisations and Insights, Patterns, Predictions and Recommendations



# 1. Background/ context of the business



# Turtle Games' Business Problem

Even though Turtle Games collects data from sales as well as customer reviews, they do not have a clear understanding of their overall sales performance nor markets or segments within their customer base.

## Why should Turtle Games try to solve this problem?

By properly gathering and analysing customers and sales data, Turtle Games can:

1. Improve their overall sales performance.
2. Understand their customers buying behaviours and reward their most loyal customers.
3. Invest more efficiently in new games and make predictions on what games would sell the most in the future.
4. Plan and use their marketing budget more efficiently by targeting specific markets and customers segments based on their customers buying behaviours within their current customer base.

# Turtle Games' Business Objectives

Improving overall sales performance by utilising customer trends



Turtle Games wants to understand:

- How customers accumulate loyalty points.
- How useful are remuneration and spending scores data.
- How groups within the customer base can be used to target specific market segments.
- How social data (e.g. customer reviews) can be used to inform marketing campaigns.
- The impact that each product has on sales.
- How reliable the data is (e.g. normal distribution, skewness, or kurtosis).
- What the relationship(s) is/are (if any) between North American, European, and global sales.

A photograph of a young woman with long red hair and glasses, wearing a teal cardigan over a white top. She is sitting at a desk in an office, looking down at a silver laptop. The office has multiple desks, bookshelves, and plants in the background. A large black desk lamp is positioned above her. In the foreground, there's a blurred view of another person's desk with papers and a mug.

## 2. Analytical Approach, using Python and R

- A. Making predictions with Linear regression using Python
- B. Making predictions with clustering using Python
- C. Analysing customer sentiments with reviews
- D. Visualising sales data to gather insights in R
- E. Cleaning, manipulating and visualising sales data in R
- F. Making recommendations to the business

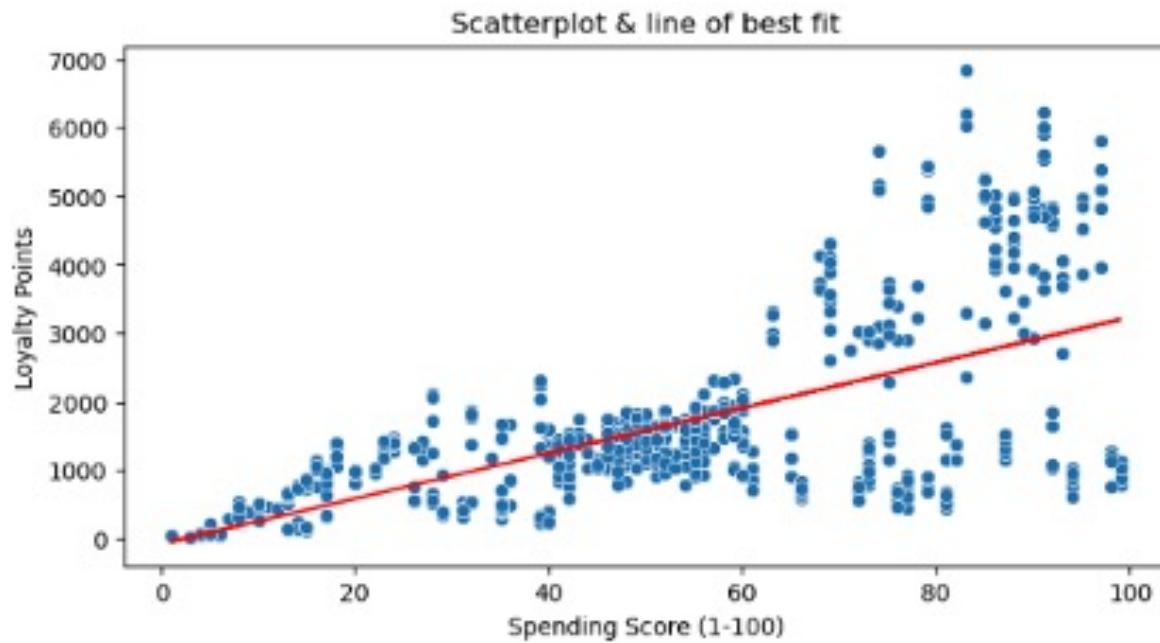
A complex network graph visualization on a dark purple background. It features numerous small white dots representing nodes, connected by a dense web of thin white lines representing edges. A prominent, larger cluster of nodes is located in the upper left, while a more linear sequence of nodes extends from the bottom left towards the center. Some nodes in the sequence are highlighted with larger circles.

### 3. Visualisations, and Insights

Patterns,  
Predictions and  
Recommendations

## Question 1: How customers accumulate loyalty points?

Correlation between loyalty points and spending score: 67%



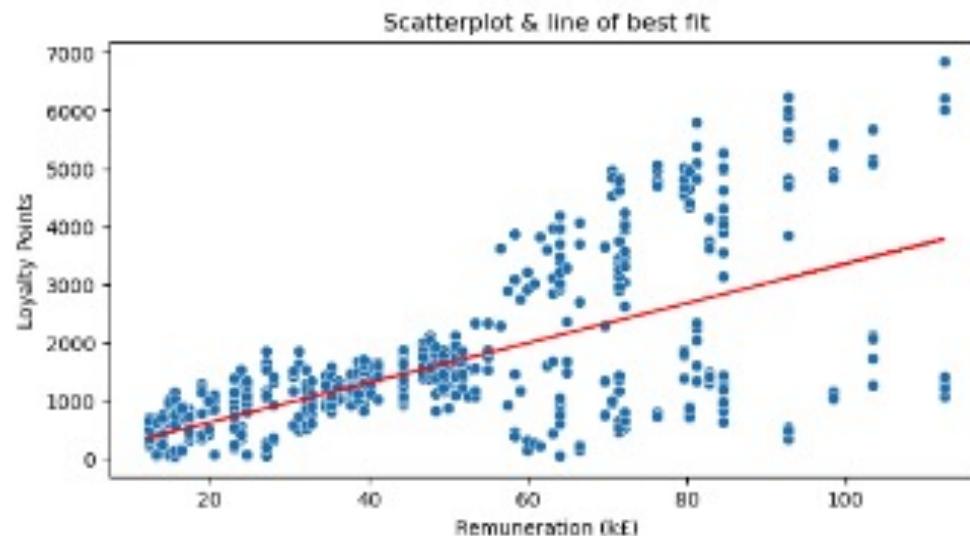
This indicates a **positive linear relationship** between the two variables.

**Customers with higher spending scores tend to accumulate more loyalty points, however this does not imply causation between these two variables**, since there are customers with high spending scores that do not accumulate many loyalty points, (all those blue dots below the red trend line).

Overall, **spending score is not a very significant estimate of the loyalty points**, and do not have a strong effect on the loyalty points in this linear regression model.

## Question 1: How customers accumulate loyalty points?

Correlation between loyalty points and remuneration: 62%.

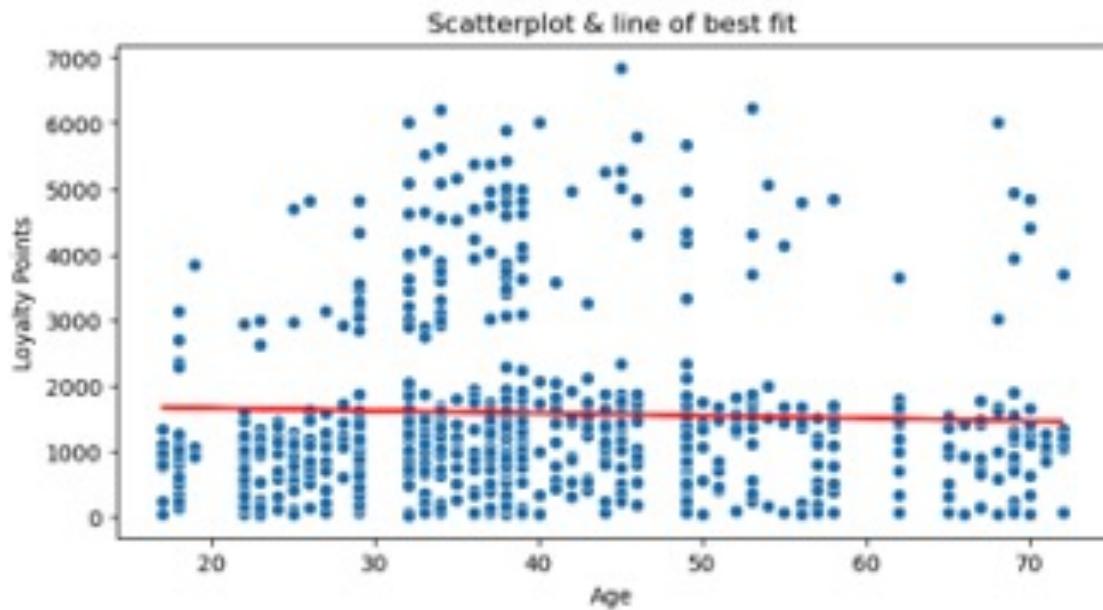


This indicates a **positive and moderately strong linear relationship** between the two variables, meaning that **customers with higher remunerations tend to accumulate more loyalty points**, however this does not indicate causation between these two variables, since there are customers with high remunerations who do not accumulate many loyalty points.

**Remuneration is not a very significant estimate of the loyalty points** and do not have a strong effect on the loyalty points in this linear regression model.

## Question 1: How customers accumulate loyalty points?

Correlation between loyalty points and age: -4.2%.



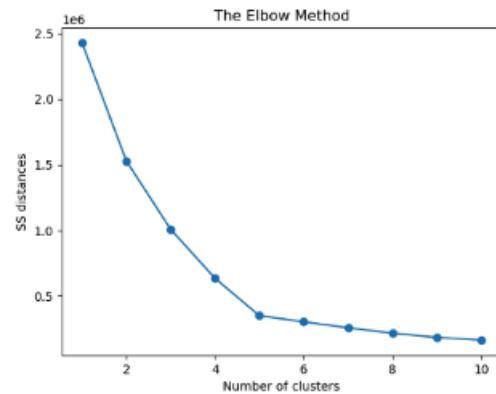
This indicates that there is a **weak negative linear relationship between age and loyalty points**. As age increases, loyalty points tend to slightly decrease, and vice versa. However, since the correlation is negative **this linear relationship is very weak**, and this is why we see the points scattered around without following a clear pattern.

It's important to note that a **weak correlation does not imply causation**. It does not imply that age is the cause of changes in loyalty points or vice versa.

The weak correlation may indicate that **age alone is not a strong predictor of loyalty points**. There may be other factors or interactions with age that better explain variations in loyalty points, hence we can say that age is not a significant estimate of the loyalty points.

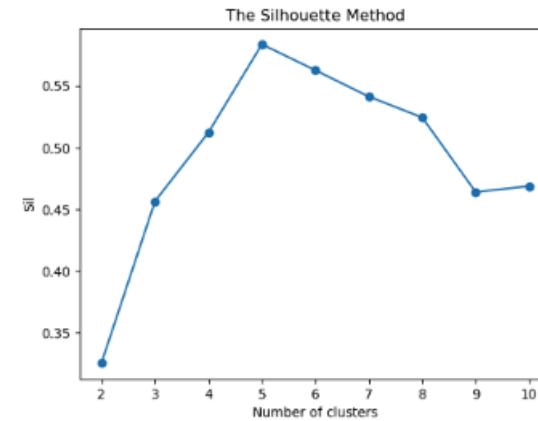
## Question 2: How useful are remuneration and spending scores data in providing data for analysis and in finding groups within the customer base that can be used to target specific market segments?

Considering that **Turtle Games** is looking to **identify groups within the customer base that can be used to target specific market segments for their marketing campaigns**, our job is to identify the **optimal number of clusters** for this purpose

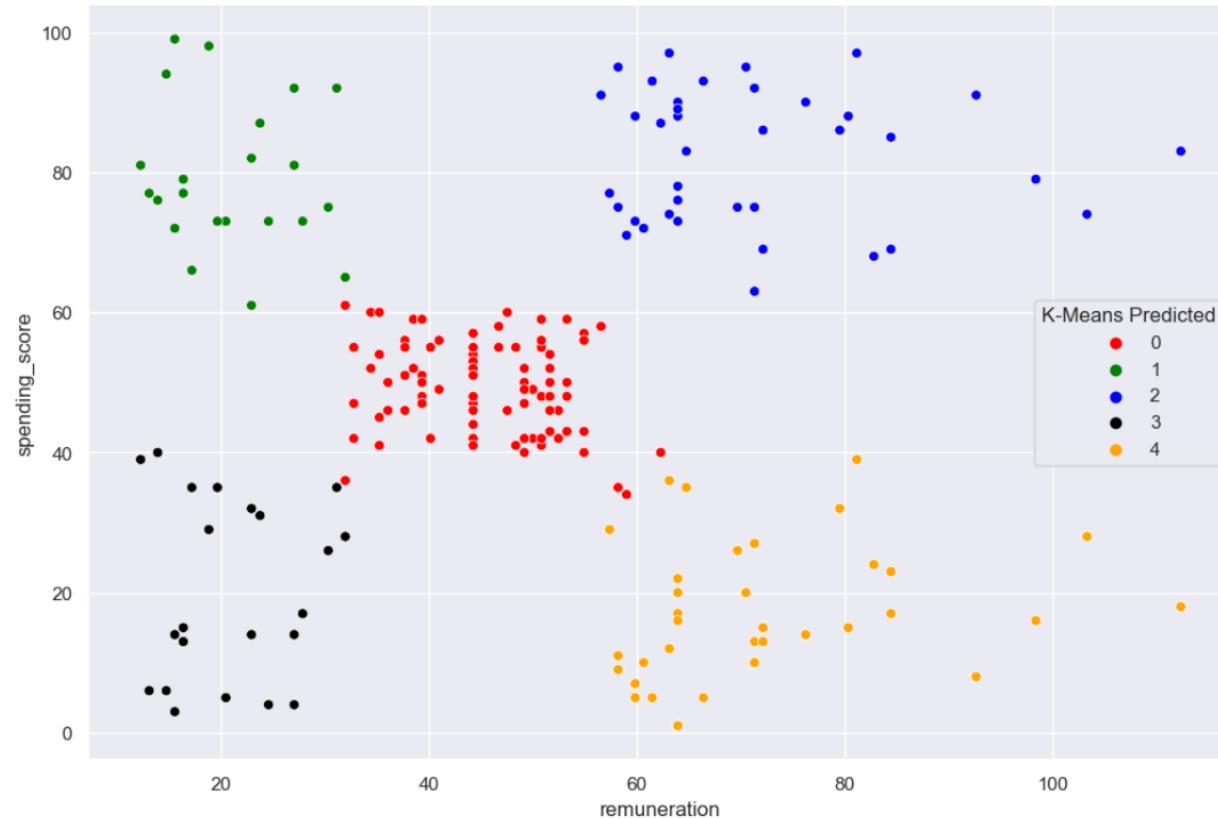


The number of predicted values per class seems to indicate a better distribution for  $k=5$  than  $k=7$  and  $k=3$ , hence I selected five clusters as the optimal number of clusters based on all the analysis and insights mentioned above.

Based on the Elbow and Silhouette methods, **five main clusters have been identified**, meaning that the five cluster solution has the highest average silhouette coefficient.



There is a distinct demarcation in terms of remuneration and spending scores, however here's not a causative relationship between these two variables.



**Clusters two and four** are the groups with the highest remuneration however this is not an indication of a high spending score.

And **clusters three and one** are the customers with the lowest remuneration, but again this is not an indication of a lower spending score.

The fifth group is cluster zero, which might be the easier segment to target for Turtle Games, since this group of customers seems to behave more homogenously and isolated in terms of all of the observations within our data set grouping customers with a remuneration between (£35K - (£55K) approx. and a spending score between 35 - 60 points), with no other clusters showing this combination of remuneration and spending score, unlike the rest of the clusters that share some similarities of behavior in terms of sharing either the same remuneration level with another cluster or sharing the same spending score with another cluster.

## **Segments for Turtle Games to target in their marketing campaigns**

**Segment 1:** customers with a remuneration between (£35K - (£55K) and a spending score between 35 - 60 points)

**Segment 2:** customers with a remuneration between (£15K - (£35K) and a spending score between 60 - 100 points)

**Segment 3:** customers with a remuneration between (£55K - (£100K+) and a spending score between 60 - 100 points)

**Segment 4:** customers with a remuneration between (£15K - (£35K) and a spending score between 5 - 40 points)

**Segment 5:** customers with a remuneration between (£55K - (£100K+) and a spending score between 5 - 40 points)



By grouping Turtle Games within the above segments, we could say that **the spending score is an indication of the number of purchases or the amount of money spent by customers during a specific range of time**, indicating that **regardless of the customer remuneration, there are customers with a low remuneration that tend to have a high spending score** (we can assume that this is because they spend more), these customers belong to segment 2, and then to give another example of how customers behave, we can see that for instance, **customers within segment 5, with a high remuneration above 60K, tend to have a lower spending score, below 40 points** (we can assume that this is because they spend less).

In future analysis, it might be worth looking at the behavior of these segments by considering a specific **period of time**, let's say one year, one month, or even a quarter, so we can then compare the customers behavior across different **years, months or quarters and send them surveys to gather more insights from their buying behavior**.

It might also be interesting for future analysis, to introduce a third variable into our clustering analysis, such as the customer **age, gender or education**.

### **Question 3: How customer reviews can be used to inform marketing campaigns?**

## A. Review Word Cloud



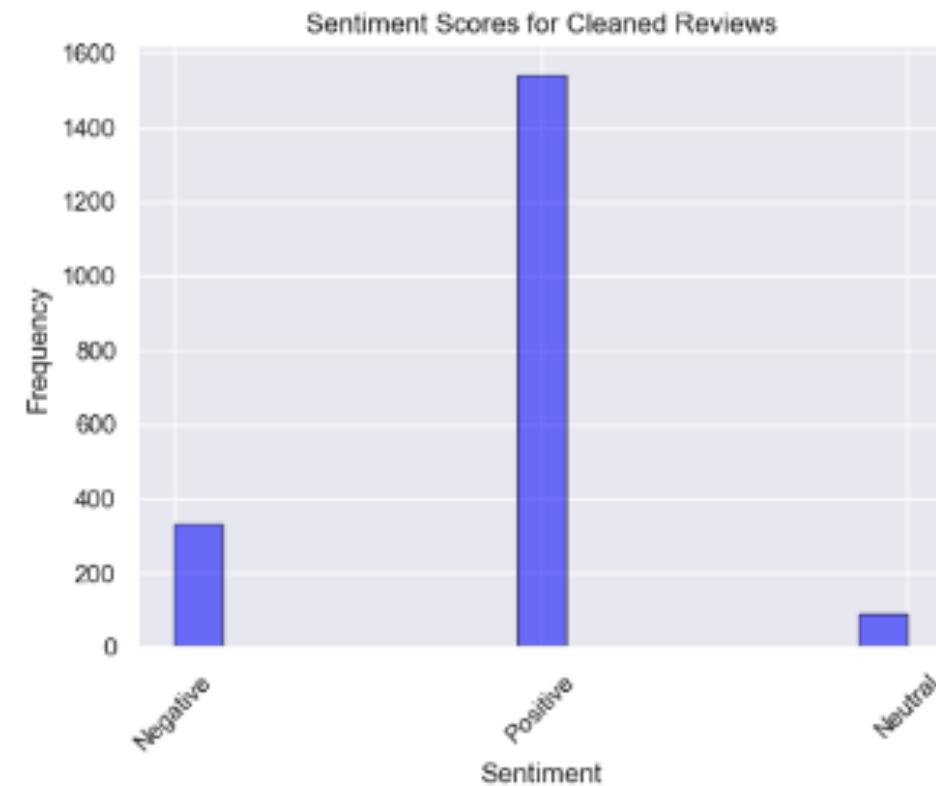
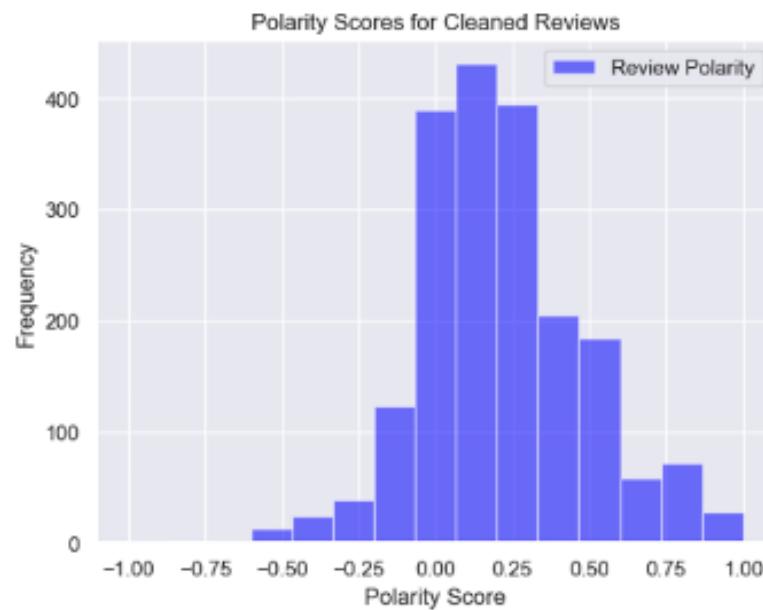
Review Words	Frequency
game	1671
great	580
fun	552
one	530
play	502
like	414
love	323
really	319
get	319
cards	301
tiles	297
time	291
good	289
would	280
book	273

## B. Summary Word Cloud

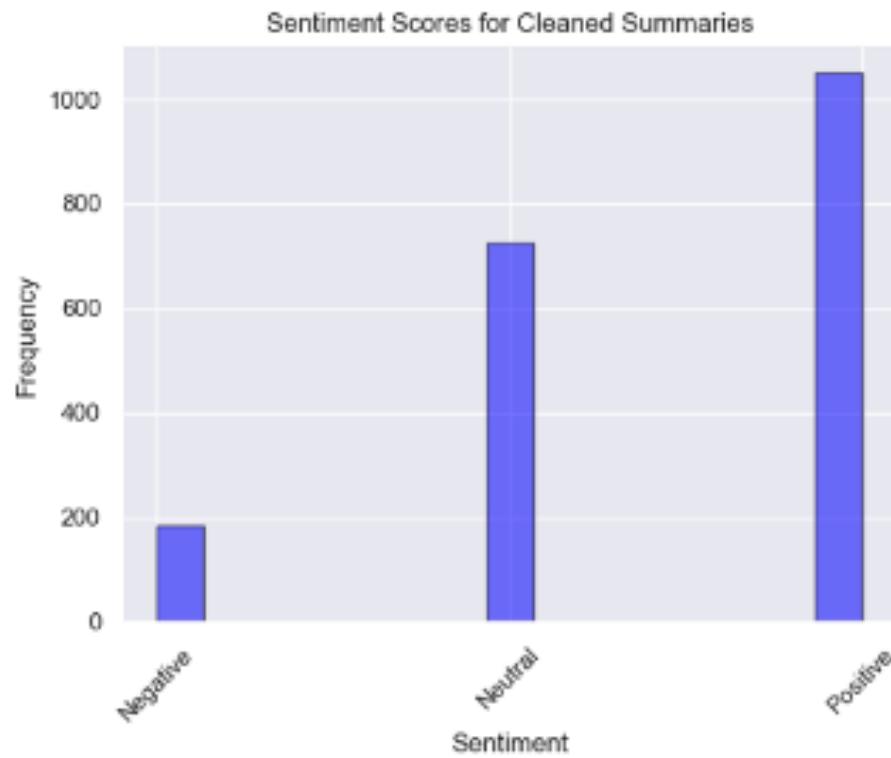
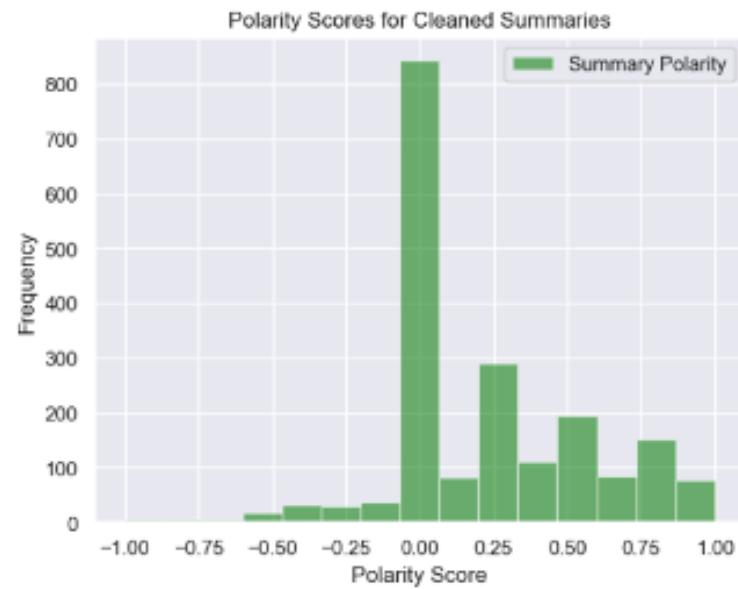


Summary_Words		Frequency
stars	427	
five	342	
game	319	
great	295	
fun	218	
love	93	
good	92	
four	58	
like	54	
expansion	52	
kids	50	
cute	45	
book	43	
one	38	
awesome	36	

## Polarity Score and Sentiment Scores: Reviews



## Polarity Score and Sentiment Scores: Summaries





## Top 20 negative reviews

```
[ 'im sorry find product boring frank juvenile',
  'comes dms screen space screen absolute premium fact space wasted art terribly informative needed art well makes completely useless reason gave stars technically speaking least still stand block notes dice rolls drops ball completely',
  'really cheaply produced cardboard playing pieces cost',
  'high hopes game big fan fourth edition dd though like e enjoy im going tell one better ill tell one prefer also enjoy games modular tiles really went game biased towards enjoying sadly couldnt rescue game exceedingly repetitive play turns youll get entire experience game offer nothing new exciting game hopelessly shallow based much chance strategy nonexistent despite modular tiles game offers virtually customization pieces seem wellmade fortunately might supply use actual dd since store bought doesnt returns however quality pieces matched game really wanted enjoy game played whole games uncounted number aborted attempts accepting fact simply doesnt work extremely disappointing',
  'wish id watched gameplay videos ashardalon buying others mentioned rpg board game board game particularly bad board game punishingly difficult little theme begs house rules clean terrible encounter card system tried playing aggressively quickly overwhelmed critters kill tried playing cautiously party wrecked encounter cards doesnt feel like proper dungeon crawl youve killed five monsters lost half characters hp youve left sight staircase like combat game set wacky arena bad stuff happens time seen another way like puzzle game balance risk generally much reward generally little hopefully survive long enough reach objective way feel like dungeon crawl importantly poor introduction rpg elements clearly minority wellreviewed game simply wish offer dissenting opinion ashardalon fun play im selling copy im going try mice mystics instead looks like hoping ashardalon would balance teenfriendly story adventure combat need dm im also eager try arcadia quest grant two stars review instead one bits box fantastic excellent miniatures board pieces to pnotch presentation looks fantastic take box',
  'book pages size x note card much fun',
  'great game kids love play',
  'game tiles board tile stands made paper using times sustain paper board tiles move board making game messy inconvenient manage shame done brilliant game',
  'guess look closely information game impression similar acquire used play quality one use looks cheap would paid get better quality',
```

## Top 20 negative summaries

```
[ 'disappointing',
  'fact space wasted art terribly informative needed art',
  'disappointing',
  'disappointing',
  'disappointing',
  'small boring',
  'mad dragon',
  'bad quality all made paper',
  'bad expecting',
  'promotes anger instead teaching calming methods',
  'disappointed',
  'disappointed',
  'disappointed',
  'disappointed',
  'another worthless dungeon masters screen galeforce',
  'hated running rpg campaign dealing towns',
  'boring',
  'horrible nothing say would give zero stars',
  'boring unless craft person',
  'worst value ive ever seen']
```



## Top 20 positive reviews

```
['make game come alive battles go much quickly smoothly nicely designed great way give detail depth battles',
'excellent expansion lords waterdeep importantly adds sixth player option game',
'pigeon books elementary school library dont pigeon missing students love pigeon',
'disclaimer one villain cards came factory defect machine burned rarity far rest best miniatures ive ever cave bears
wifes favorite love otyugh insane number orcs duergar extra kobolds devils frickin awesome seriously well balanced ev
en new game features stand even better previous castle ravenloft game',
'bought doll year old boy class preschool teacher perfect children love animals separately also book trio last year
s always favorite hands story preschoolers used one previous school belonged school time bought set',
'impressed quality puzzle easy fun put together',
'great resource bhis care coordinators works well kids teens says',
'opinion best dungeon crawler setup played tote hours makes fun ride fun kids mine ',
'wonderful ball manipulate consists sided pieces fit together sided ball approximately inches diameter assemblies m
ade',
'love expand current game new additions completely change half expansion two different expansions',
'grandson could put wants play',
'daughter loves little books theyre perfect size keep car diaper bag purse keep hand times stuck waiting doctors off
ice anywhere else',
'far favorite dd board games fun easy learn even got family play',
'quick fun easy learn wide age range fast play great family game',
'great gift wedding party dr themed wedding everyone loved gift reasonably priced much fun',
'yes quick wonderful accurate',
'love game playing years best played people',
'loads fun youve played dd boardgames set wrath ashardalon opinion wellrounded among includes gold tokens spend buyi
ng treasures adventures balanced set heroes memorable unique abilities without overpowered make game easy relatable s
etting havent played boardgames dont one standalone game rules board pieces cards characters interchangeable game dun
geon crawl turn explore new room dungeon deal things find monster trap encounter anyall fully cooperative need one pl
ayer dungeon master plays different rules others acts villain even play game solo really wanted ai rules monsters enc
ounter game determined card dont worry making people feel like theyre singled ie game wont break friendships unlike r
isk monopoly approachable dont ever played game dungeons dragons understand even appreciate hand could actually great
way introduce people dd youre inclined rules distilled simplified versions actual tabletop roleplaying game typical r
ound play including setup pickup could last anywhere minutes hours extreme end',
'daughter loves stickers awesome seller thank',
'somuchfun seriously addictive small card set family skeptical first prying away table almost midnight one wants qui
t playing best stocking stuffer gift ever good sense buy']
```

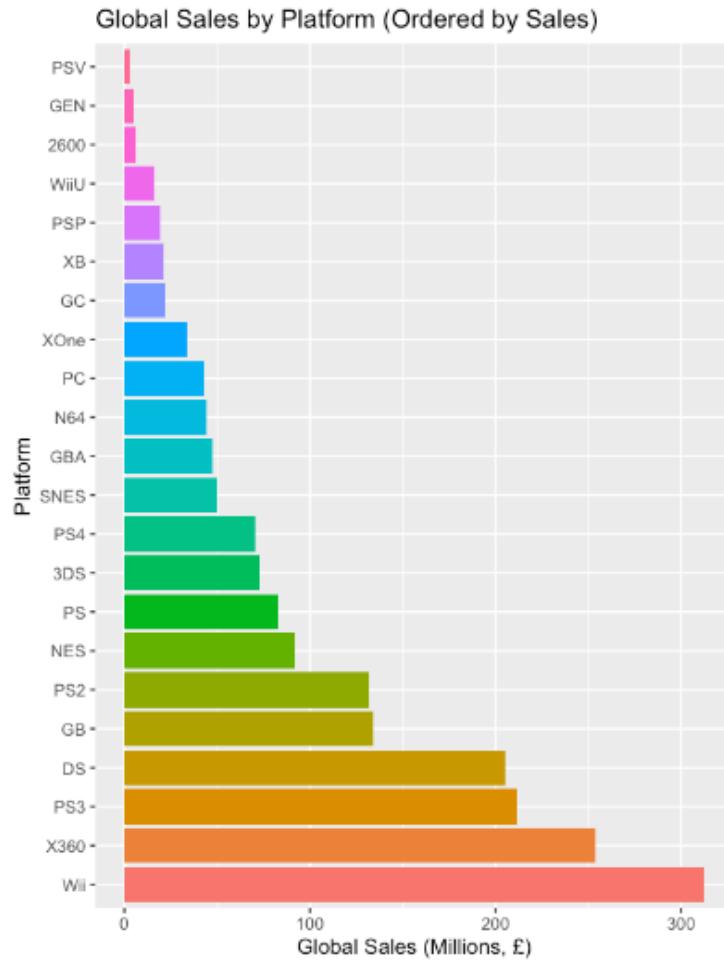
## Top 20 positive summaries

```
['awesome',
'adds six player option excellent expansion',
'pigeon perfect addition school library',
'best boardgame ever',
'perfect preschoolers',
'excellent puzzle',
'perfect',
'best dungeon crawler',
>wonderful ball manipulate',
'awesome expansion',
'awesome',
'theyre perfect size keep car diaper',
'best one series',
'excellent',
'perfect gift',
>wonderful',
'one best games ever',
'best among dd boardgames',
'awesome seller thank',
'one best games ever']
```

## Question 5: What is the impact that each product has on sales?

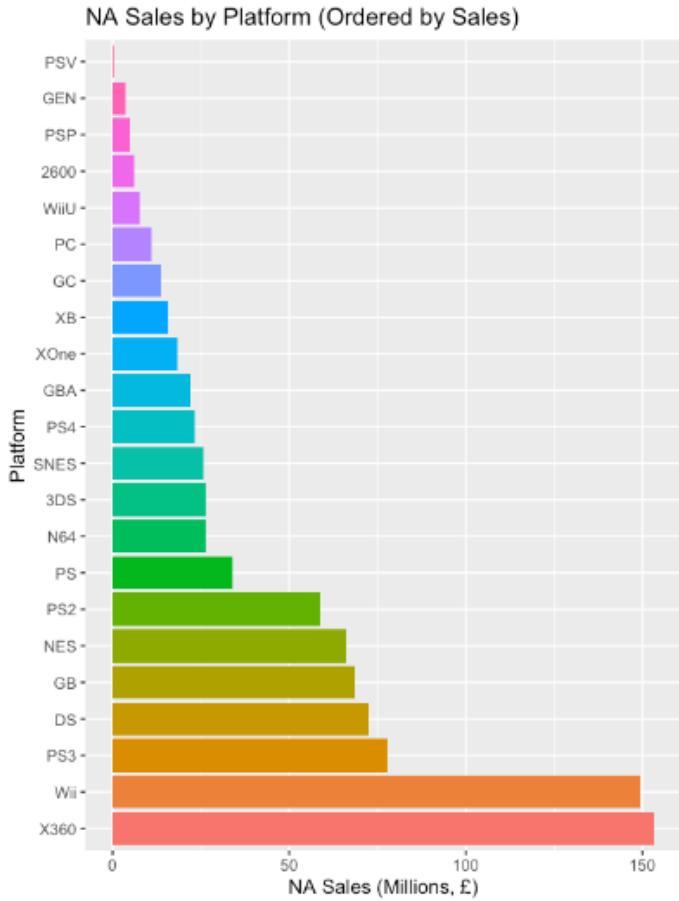
Product-ID	Global Sales	Product-ID	NA Sales	Product-ID	EU Sales
107	67.8	107	34.0	107	23.8
515	45.9	123	26.6	515	18.9
123	37.2	326	22.1	195	10.6
254	29.4	254	21.5	3967	10.2
195	29.4	515	19.2	2371	9.26
231	27.1	948	14.4	876	9.25
249	25.7	535	13.1	3645	9.14
948	25.4	195	13	979	9.07
876	25.3	231	12.9	231	9.03
263	24.6	876	12.8	399	9.02

## Question 5: What is the impact that each platform has on sales?



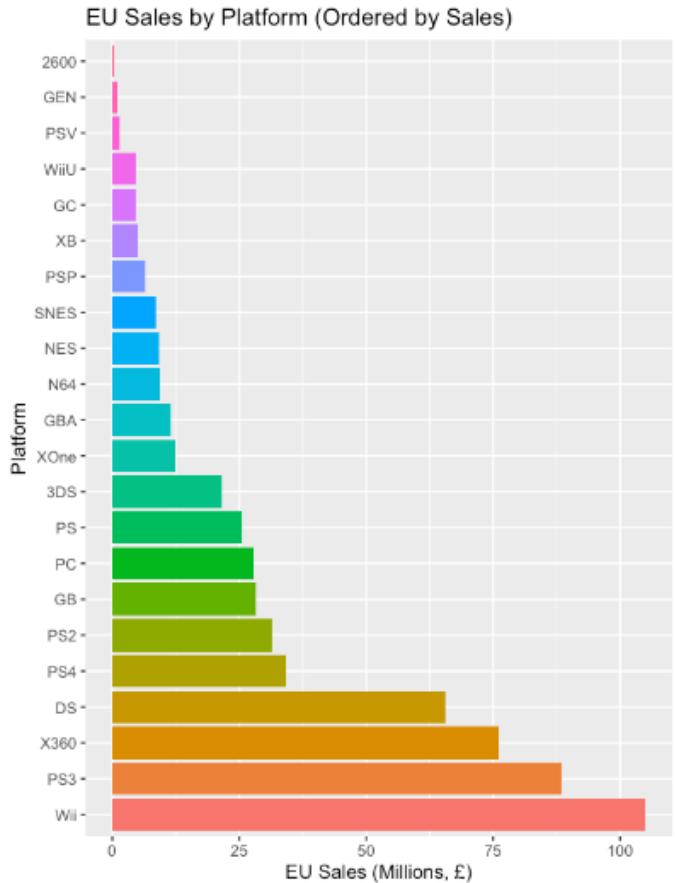
Platform	Global Sales
1 Wii	313
2 X360	254.
3 PS3	212.
4 DS	205.
5 GB	134.
6 PS2	132.
7 NES	91.4
8 PS	82.9
9 3DS	73.2
10 PS4	70.5

## Question 5: What is the impact that each platform has on sales?



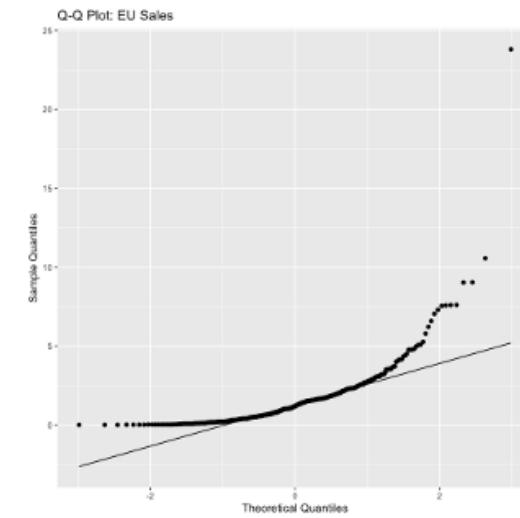
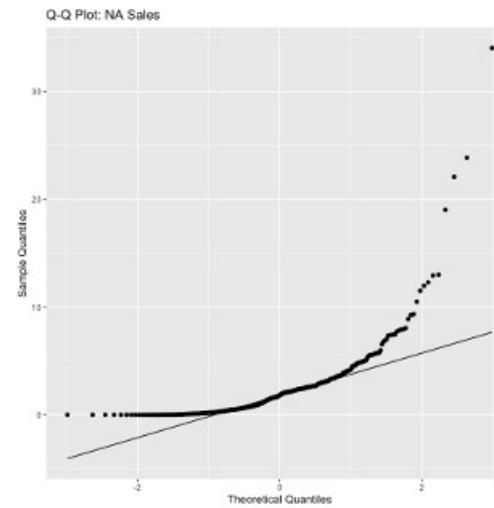
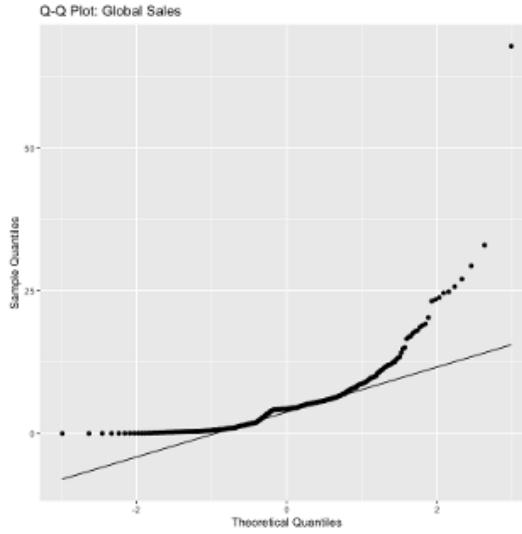
Platform	NA Sales
1 X360	153.
2 Wii	150.
3 PS3	77.8
4 DS	72.6
5 GB	68.7
6 NES	66.0
7 PS2	58.7
8 PS	34.0
9 N64	26.4
10 3DS	26.4

## Question 5: What is the impact that each platform has on sales?



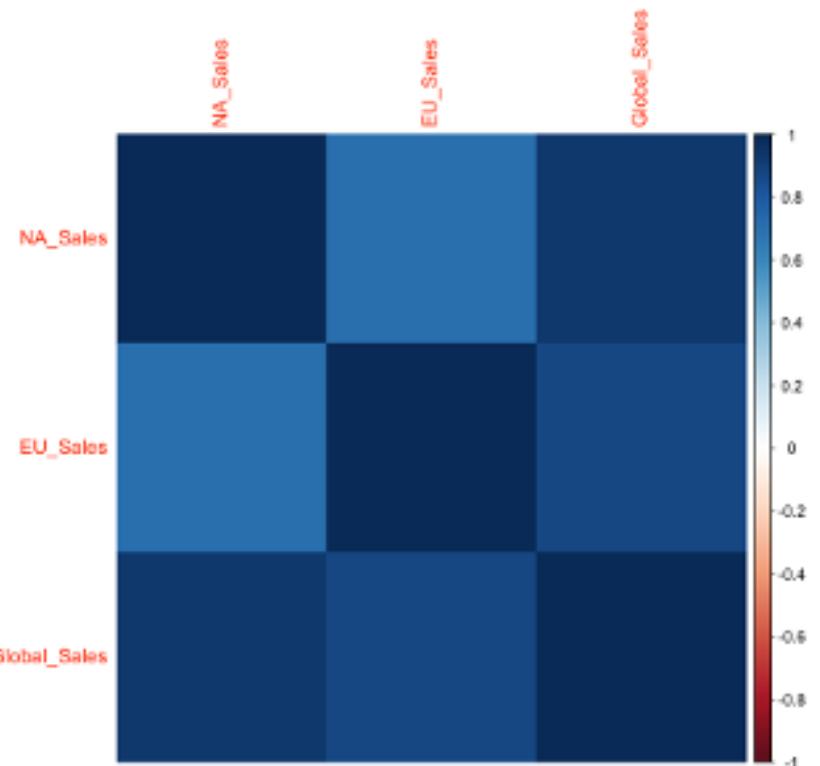
Platform	EU Sales
1 Wii	105.
2 PS3	88.5
3 X360	76.0
4 DS	65.6
5 PS4	34.1
6 PS2	31.5
7 GB	28.2
8 PC	27.9
9 PS	25.6
10 3DS	21.6

## Question 6: How reliable the data is (e.g. normal distribution, skewness, or kurtosis)?



As we can see, the points for the three Sales columns deviate from the line in an S-shape, this indicates that the **data might have heavy tails, indicating non-normality**

## Question 7: What the relationship(s) is/are (if any) between North American, European, and global sales?

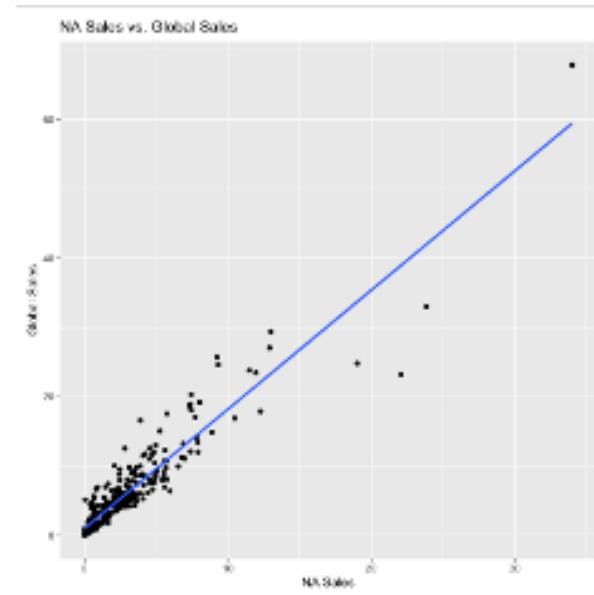
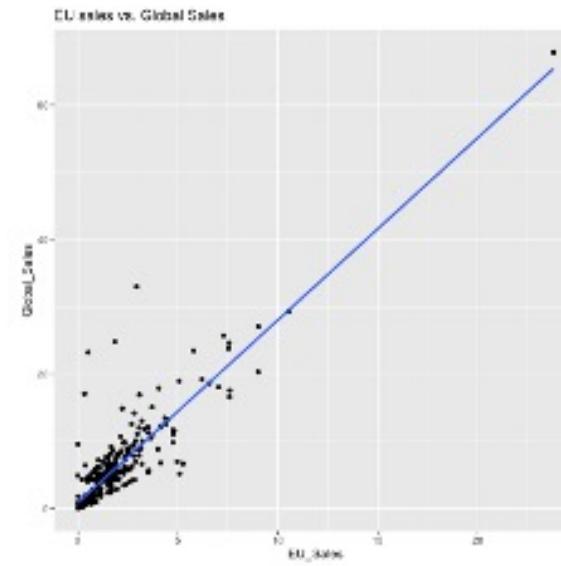
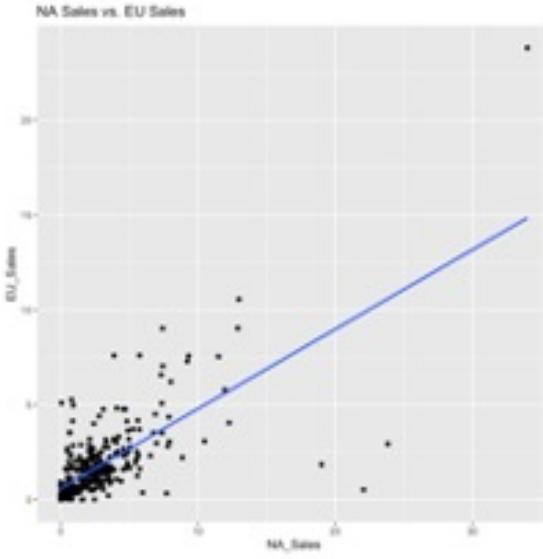


As expected, there is a **high positive correlation between the global sales and the NA sales and EU sales**, since global sales in the world is the sum of EU sales, NA sales and other sales.

For future analysis, it might worth **digging into the 'other sales'** category and calculate the sales under this uncategorised group.

**Turtle Games should properly collect and label the 'other sales' category to identify the market / markets that belong to this category.** This would be very useful to inform future marketing campaigns too since we would be able to see socio-economic and demographic profile of the customers under this category.

## Question 7: What the relationship(s) is/are (if any) between North American, European, and global sales?



- NA Sales are greater than EU sales.
- Overall, we can see that the global sales are much greater than the EU and NA sales.
- We can also see how there are some global sales values that deviate more from the normal distribution, meaning that these outliers are the greatest sales values observed within the global sales column, such as the black dot shown way over the 60M mark for Global Sales.

# Q & A



# Thank You

---

Cristina Alonso Robles

Data Analyst

[aroblescristina@gmail.com](mailto:aroblescristina@gmail.com)

