Turtle Games
Customer Trends
Analysis Report

Filipe Bastos

April 2023

Contents

LSE Data Analytics
Career Accelerator   Filipe Bastos

1. Introduction

1.1. Background

As part of the Course 3, Advanced Analytics for Organisational Impact, final assignment we will help Turtle Games, a game manufacturer and retailer with a global customer base, to learn more about customer trends in order to achieve the business objective of improving overall sales performance.

To complete this analysis, we will be analysing two data sets, turtle_reviews.csv and turtle_sales.csv, to identify insights using Python and R.

1.2. Objectives

A fishbone diagram was created to establish a problem framework to make it easier to identify, plan and structure the analysis.
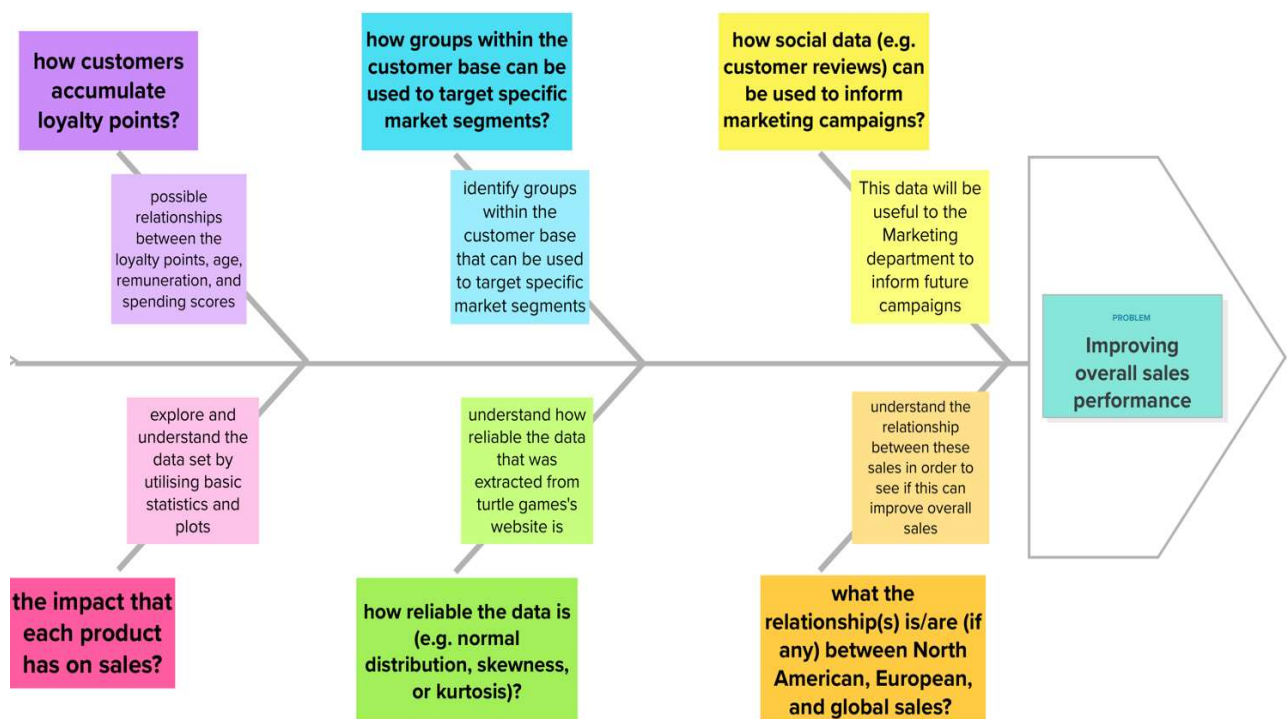

Figure 1 - Fishbone diagram

2. Analytical approach

2.1. Data Exploration

All the data exploration was completed using Python and R. In order to familiarise check if the data had to be cleaned we completed the following exploratory analysis.

1. **turtle_reviews.csv Data Set**

This data set has 2000 entries and 11 columns, below is a summary of the data types and the number of null values in each column.

Column	Non-Null Count	Dtype
gender	2000 non-null	object
Age	2000 non-null	int64
remuneration (k£)	2000 non-null	float64
spending_score (1-100)	2000 non-null	int64
loyalty_points	2000 non-null	int64
education	2000 non-null	object
language	2000 non-null	object
platform	2000 non-null	object
product	2000 non-null	int64
review	2000 non-null	object
summary	2000 non-null	object

Table 1 – turtle_reviews.csv info.

2. **turtle_sales.csv Data Set**

This data set has 352 entries and 5 columns, below is a summary of the data types and the number of null values in each column.

Column	Non-Null Count	Dtype
Product	352 non-null	int
Platform	352 non-null	chr
NA_Sales	352 non-null	num
EU_Sales	352 non-null	num
Global_Sales	352 non-null	num

Table 2 – turtle_sales.csv info.

2.2. Data Analysis

2.2.1. Introduction

In this chapter we will perform data wrangling manipulation to search for answers to general questions that will help us better understand the data and create visualisations to identify possible trends.

2.2.2. Loyalty Points

In this section we will investigate the possible relationships between the loyalty points, age, remuneration, and spending scores. We will evaluate possible linear relationships between loyalty points and age/remuneration/spending scores to determine whether these can be used to predict the loyalty points.
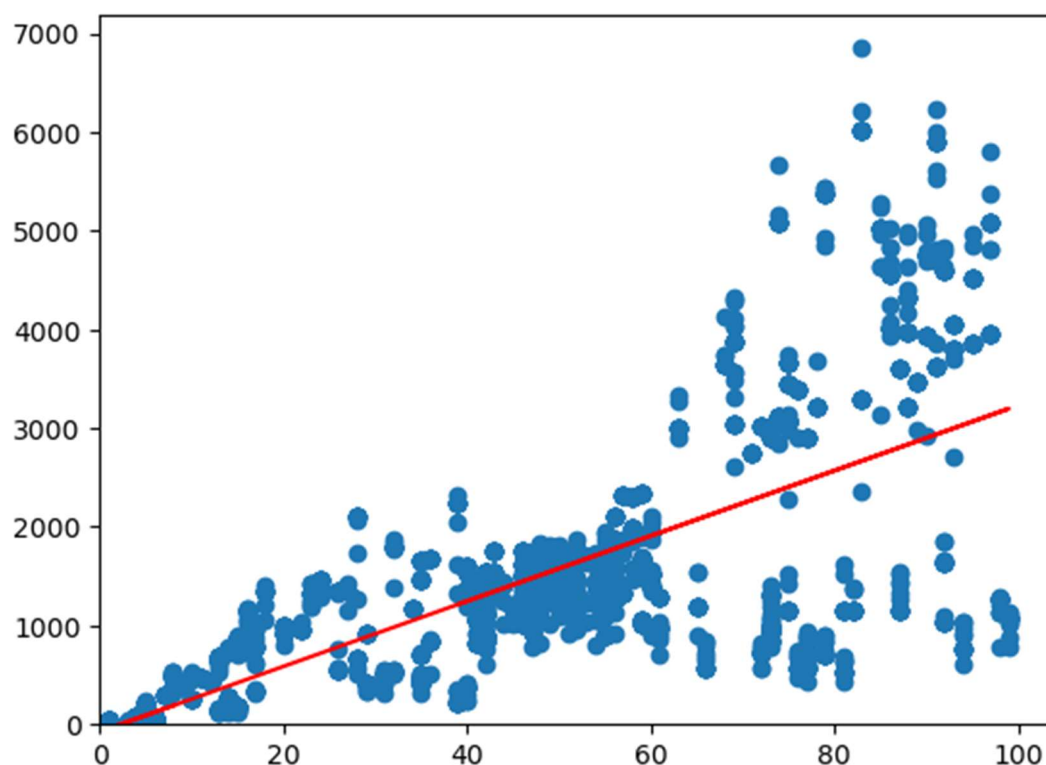

Figure 2 - Spending vs loyalty regression line

- R^2: 45% of the total variability of y (loyalty), is explained by the variability of X (spending).
- P – value is smaller than 0.05 so the set of variables of the regression model are significant.
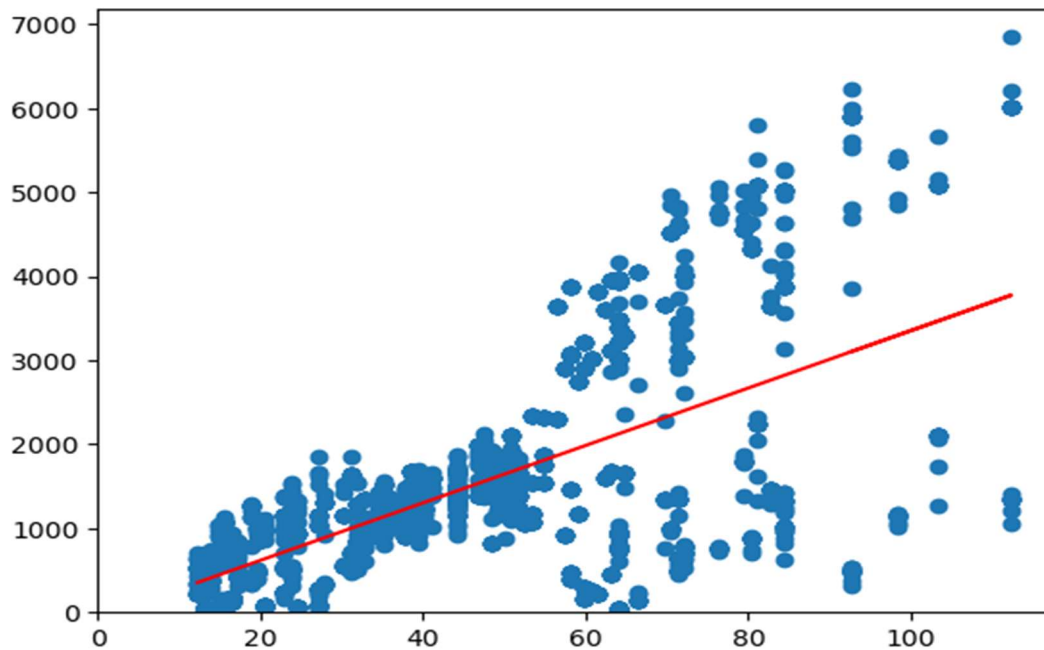
Figure 3 - Renumeration vs Loyalty regression line

- R^2: 38% of the total variability of y (loyalty), is explained by the variability of X (remuneration).
- P – value is smaller than 0.05 so the set of variables of the regression model are significant.
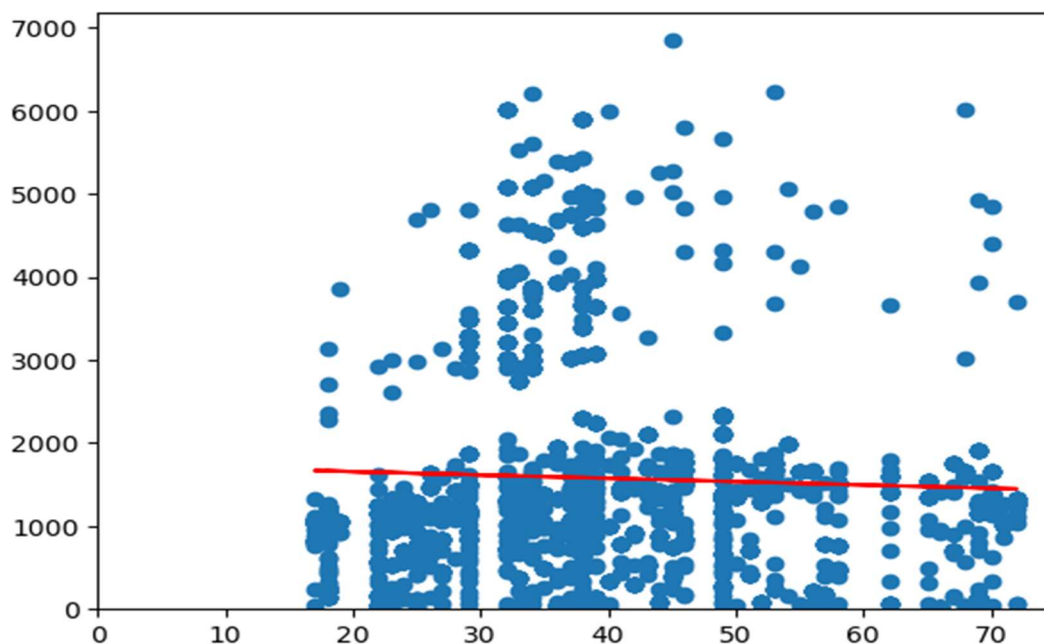

Figure 4 - Age vs Loyalty regression line

- R^2: 2% of the total variability of y (loyalty), is explained by the variability of X (spending).
- P – value is bigger than 0.05 so the model is not good.

Both the spending vs loyalty and the renumeration vs loyalty models are significant but the age vs loyalty model has a p value bigger than 0.05 so it cannot be used.

2.2.3. Grouping Customer Data Base

In this section we will identify groups, based on the remuneration and spending scores, within the customer base that can be used to target specific market segments.

In order to achieve this we used the K-means clustering method which will allow us to group costumers based on similar features. We used the Silhouette and Elbow methods to determine the optimal number of clusters.
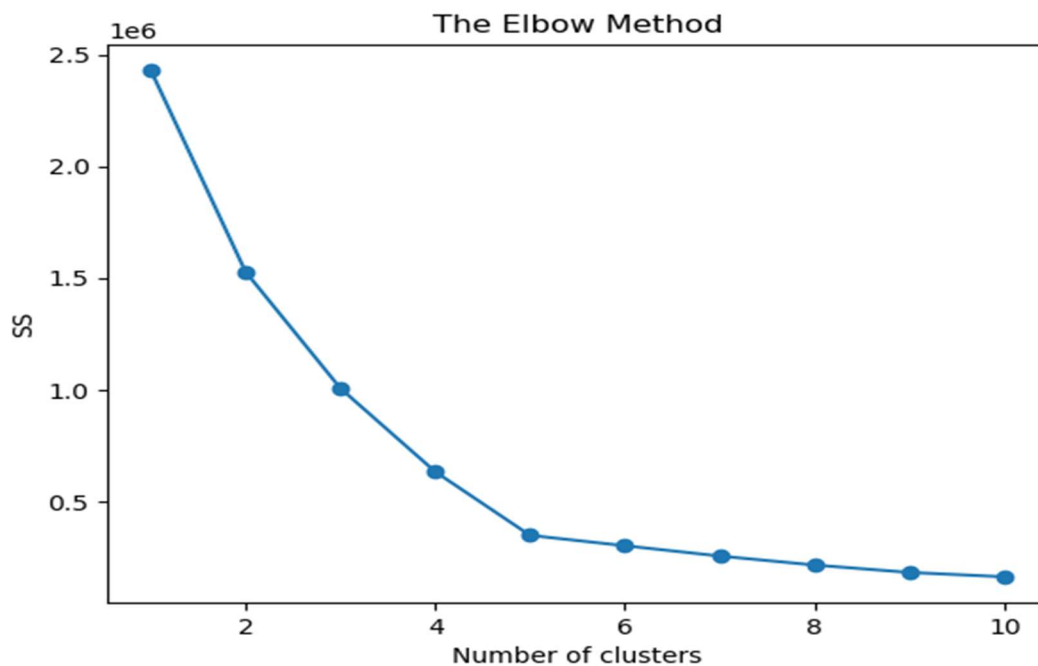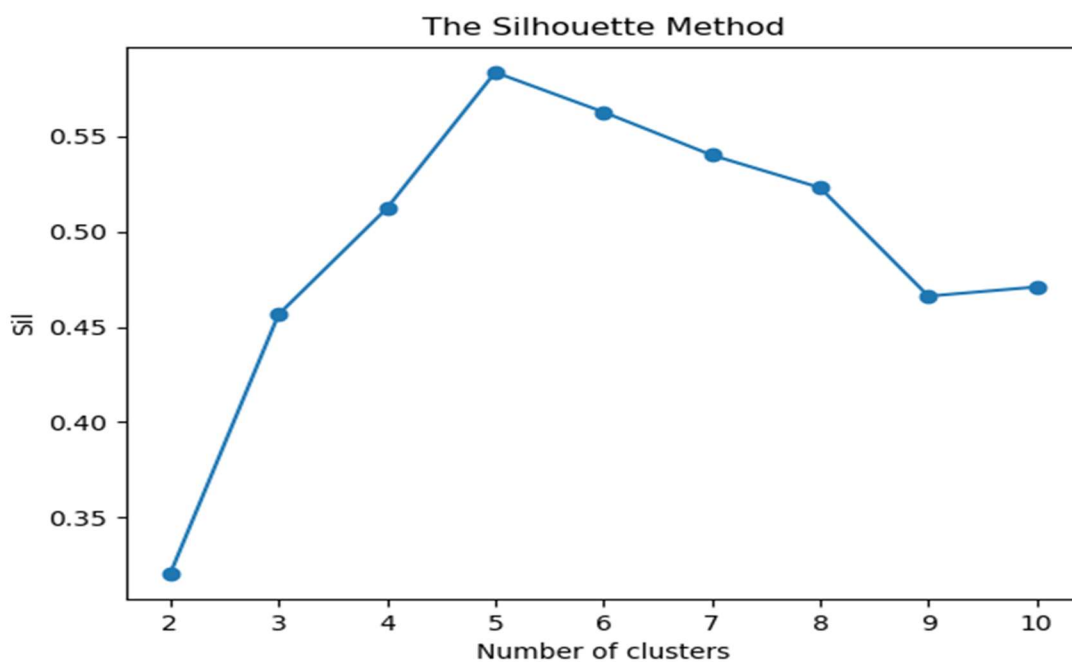

Figure 5 - Elbow Method

Figure 6 - Silhouette Method

LSE Data Analytics
Career Accelerator   Filipe Bastos

Figure 7 - Final model (number of clusters =5)

From our analysis it looks that k=5 (five clusters) give the best results (groups), and indicates a better distribution of the groups.

2.2.4. Analyse customer sentiments with reviews

In this section we will analyse the customer reviews. We will be using Natural language processing (NLP) methods, including sentiment analysis.

Below we can see the Wordcloud for both summary and review column with the most repeated words.

Figure 8 - Wordcloud

7

- Top 20 negative reviews and summaries

	review	review_polarity
208	booo unles you are patient know how to measure i didnt have the patience neither did my daughter boring unless you are a craft person which i am not	-1.000000
182	incomplete kit very disappointing	-0.780000
1804	im sorry i just find this product to be boring and to be frank juvenile	-0.583333
364	one of my staff will be using this game soon so i dont know how well it works as yet but after looking at the cards i believe it will be helpful in getting a conversation started regarding anger and what to do to control it	-0.550000
117	i bought this as a christmas gift for my grandson its a sticker book so how can i go wrong with this gift	-0.500000
227	this was a gift for my daughter i found it difficult to use	-0.500000
230	i found the directions difficult	-0.500000
290	instructions are complicated to follow	-0.500000
301	difficult	-0.500000
1524	expensive for what you get	-0.500000
174	i sent this product to my granddaughter the pompom maker comes in two parts and is supposed to snap together to create the pompoms however both parts were the same making it unusable if you cant make the pompoms the kit is useless since this was sent as a gift i do not have it to return very disappointed	-0.491667
347	my 8 yearold granddaughter and i were very frustrated and discouraged attempting this craft it is definitely not for a young child i too had difficulty understanding the directions we were very disappointed	-0.446250
538	i purchased this on the recommendation of two therapists working with my adopted children the children found it boring and put it down half way through	-0.440741
306	very hard complicated to make these	-0.439583
427	kids i work with like this game	-0.400000
437	this game although it appears to be like uno and have an easier play method it was still too time consuming and wordy for my children with learning disabilities	-0.400000
497	my son loves playing this game it was recommended by a counselor at school that works with him	-0.400000
803	this game is a blast	-0.400000
806	i bought this for my son he loves this game	-0.400000
824	was a gift for my son he loves the game	-0.400000

Table 3 - Top 20 negative reviews

	summary	summary_polarity
21	the worst value ive ever seen	-1.000000
208	boring unless you are a craft person which i am	-1.000000
829	boring	-1.000000
1166	before this i hated running any rpg campaign dealing with towns because it	-0.900000
1	another worthless dungeon masters screen from galeforce9	-0.800000
144	disappointed	-0.750000
631	disappointed	-0.750000
793	disappointed	-0.750000
1620	disappointed	-0.750000
363	promotes anger instead of teaching calming methods	-0.700000
885	too bad this is not what i was expecting	-0.700000
890	bad qualityall made of paper	-0.700000
178	at age 31 i found these very difficult to make	-0.650000
101	small and boring	-0.625000
518	mad dragon	-0.625000
805	disappointing	-0.600000
1015	disappointing	-0.600000
1115	disappointing	-0.600000
1804	disappointing	-0.600000
1003	then you will find this board game to be dumb and boring	-0.591667

Table 4 - Top 20 negative summaries

- Top 20 positive reviews and summaries

	review	review_polarity
7	came in perfect condition	1.000000
165	awesome book	1.000000
194	awesome gift	1.000000
496	excellent activity for teaching selfmanagement skills	1.000000
524	perfect just what i ordered	1.000000
591	wonderful product	1.000000
609	delightful product	1.000000
621	wonderful for my grandson to learn the resurrection story	1.000000
790	perfect	1.000000
933	awesome	1.000000
1037	awesome	1.000000
1135	awesome set	1.000000
1168	best set buy 2 if you have the means	1.000000
1177	awesome addition to my rpg gm system	1.000000
1301	its awesome	1.000000
1401	one of the best board games i played in along time	1.000000
1550	my daughter loves her stickers awesome seller thank you	1.000000
1609	this was perfect to go with the 7 bean bags i just wish they were not separate orders	1.000000
1715	awesome toy	1.000000
1720	it is the best thing to play with and also mind blowing in some ways	1.000000

Table 5 - Top 20 positive reviews

	summary	summary_polarity
6	best gm screen ever	1.000000
28	wonderful designs	1.000000
32	perfect	1.000000
80	theyre the perfect size to keep in the car or a diaper	1.000000
134	perfect for preschooler	1.000000
140	awesome sticker activity for the price	1.000000
161	awesome book	1.000000
163	he was very happy with his gift	1.000000
187	awesome	1.000000
210	awesome and welldesigned for 9 year olds	1.000000
418	perfect	1.000000
475	excellent	1.000000
543	excellent	1.000000
548	excellent therapy tool	1.000000
580	the pigeon is the perfect addition to a school library	1.000000
599	best easter teaching tool	1.000000
647	wonderful	1.000000
651	all f the mudpuppy toys are wonderful	1.000000
657	awesome puzzle	1.000000
662	not the best quality	1.000000

Table 6 - Top 20 positive summaries

2.2.5. Exploratory Analysis of Sales Data Set

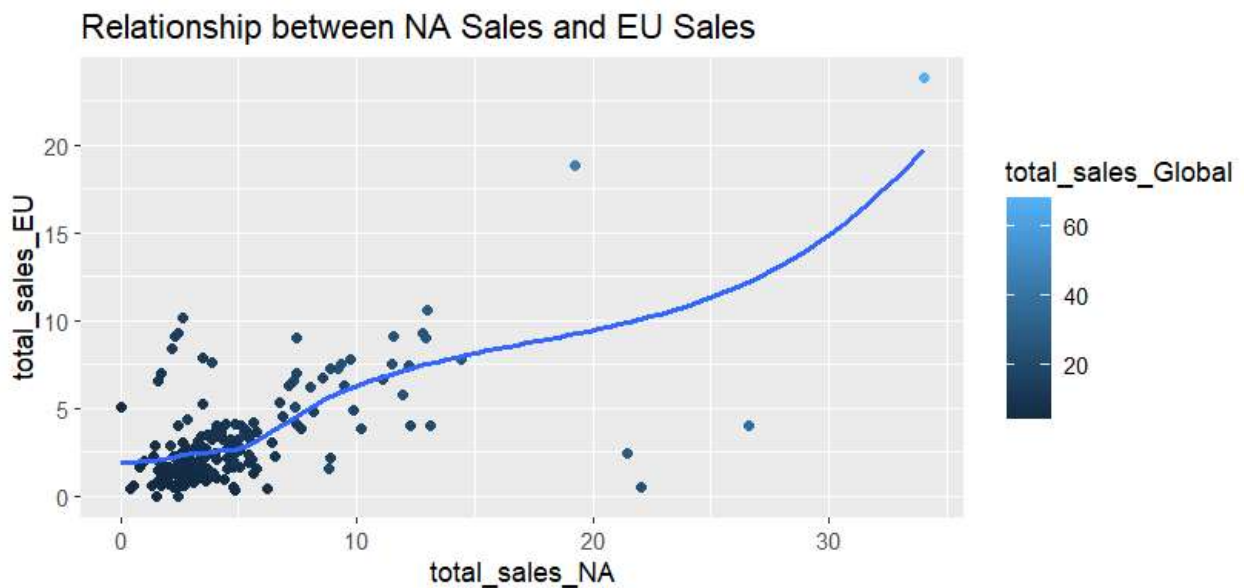In this section we will explore the sales data set.

Figure 9 - NA Sales vs EU Sales vs Total Sales

By looking at the scatterplot above there seem to be a positive correlation between NA sales, EU sales and Global Sales.
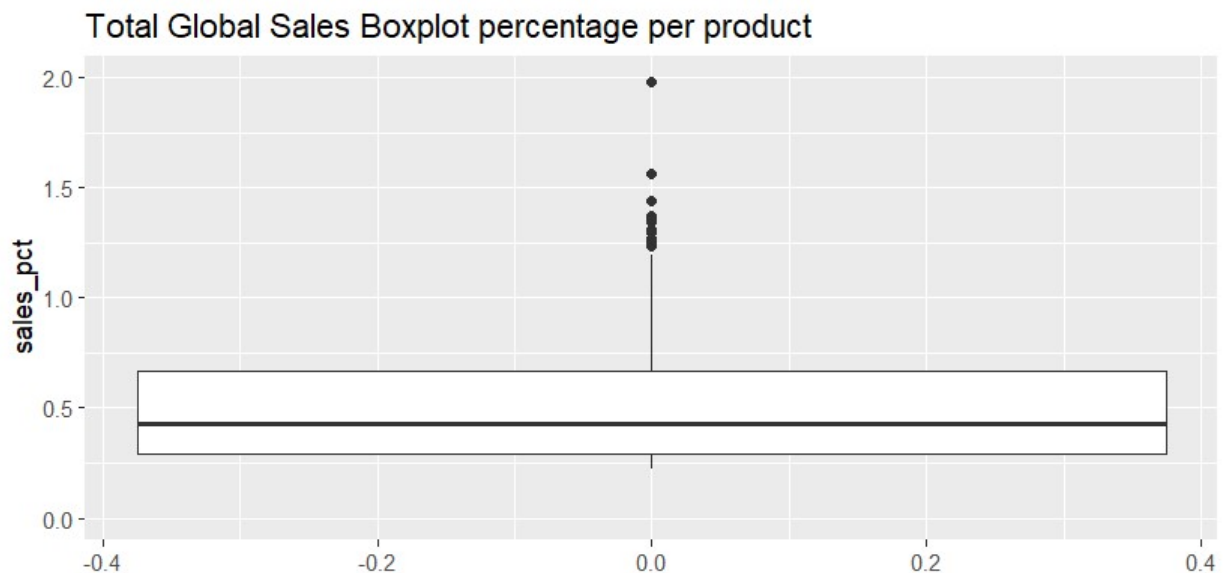

Figure 10 - Total Sales % per product Boxplot

The boxplot shows that circa 50% of the products have a weight 0.3% to 0.7% and the upper 25% have a weight between 0.6% to 1.25% of the Total Global Sales.

2.2.6. Data Statistical Analysis

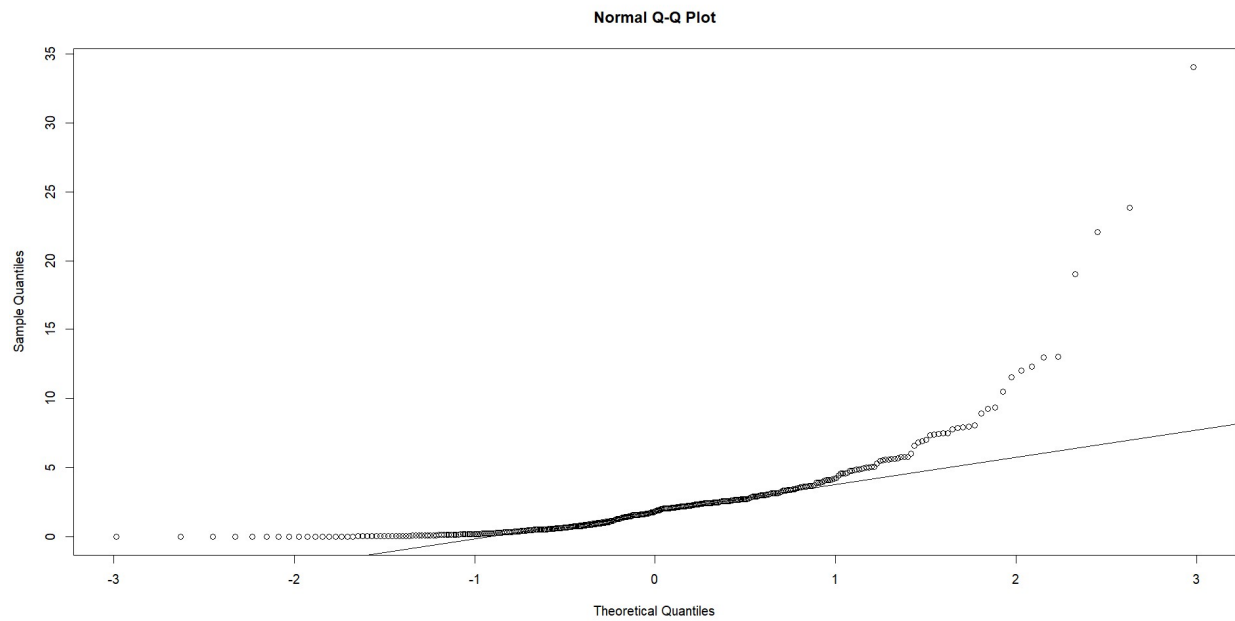In this section we will explore the sales data set.

LSE Data Analytics
Career Accelerator   Filipe Bastos

Normal Q-Q Plot

Figure 11 - Q-Q Plot NA Sales
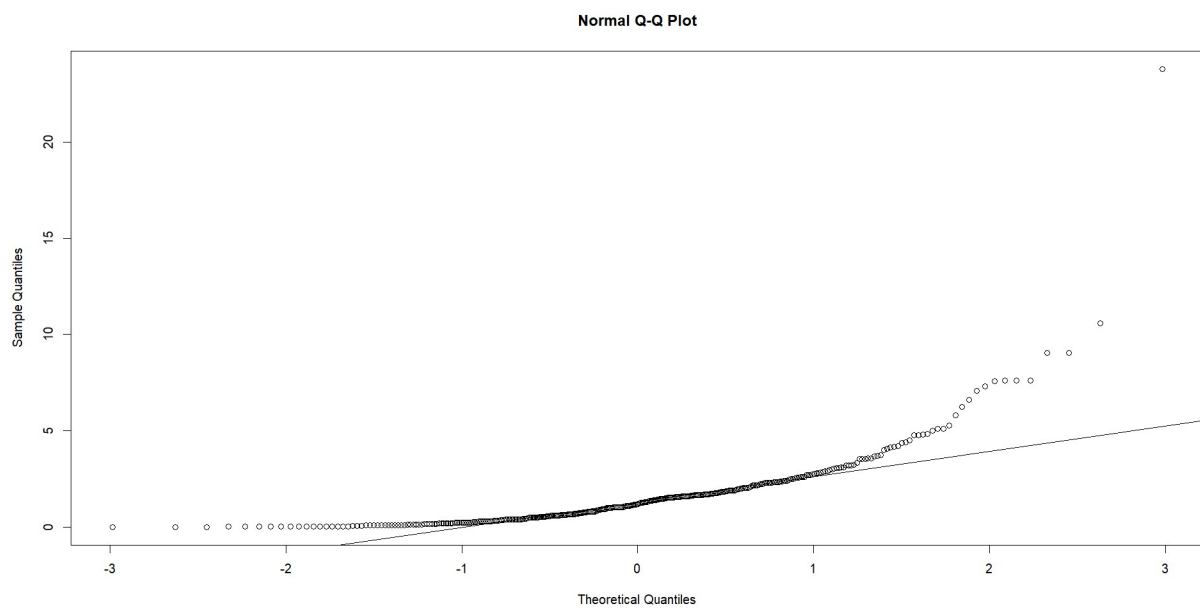
Normal Q-Q Plot

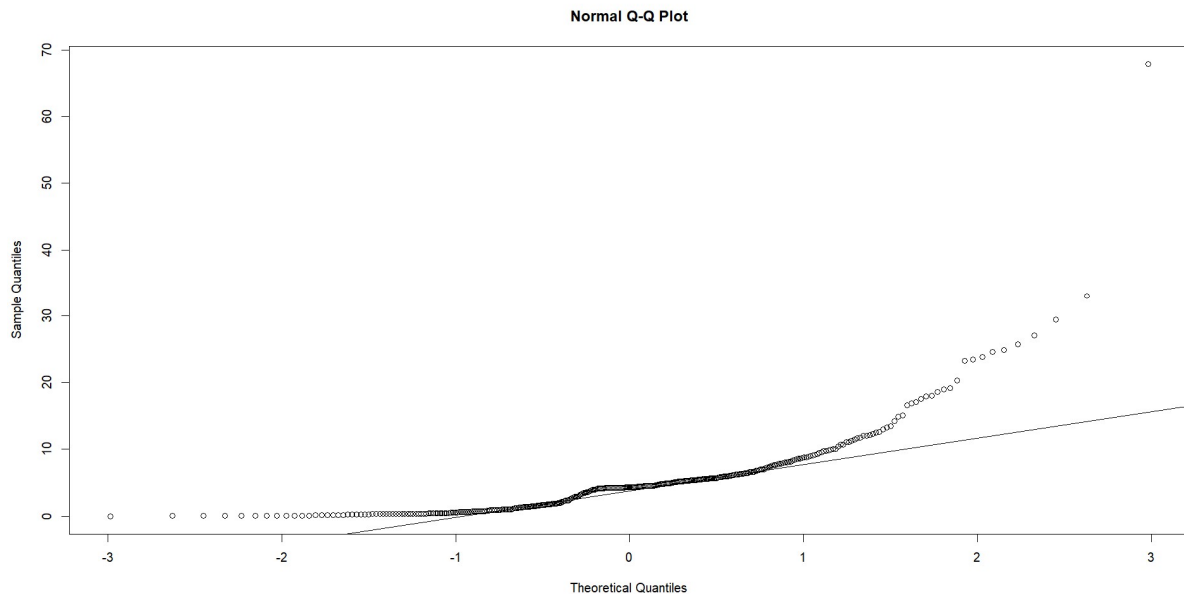*Figure 12 - - Q-Q Plot EU Sales*

11

Normal Q-Q Plot

Figure 13 - Q-Q Plot Global Sales

By looking at the Q-Q plots and at the Shapiro test results (circa 2.2e-16 for the three plots) we can conclude that the data isn't normally distributed.

2.2.7. Sales Data Set Correlations

In this section we will explore the determine the relationship between the Global Sales (dependent variable) and the EU and NA Sales (independent variables).

The correlation plot indicated that most of the variables tend to have a positive correlation – for example, NA Sales & EU Sales (0.71), NA Sales & Global Sales (0.93) and EU Sales & Global Sales (0.88).

We created a multiple linear regression model, predicted values with the model and concluded that it was very accurate.

Based on the correlation plot and accuracy of models in this case, it seems that we can use multiple linear regression to predict the Global sales based on EU and NA Sales

LSE Data Analytics
Career Accelerator   Filipe Bastos

3. Findings

3.1. Trends, Insights and Further Recommendations

Trends and Insights

- Loyalty points accumulation shows a moderate correlation with customer spending (45%) and remuneration (38%), but not with age.

- Based on customer similarities, it is recommended to segment them into five groups.

- Sentiment analysis reveals that negative comments often use words like "boring" and "disappointing", while positive comments tend to use words like "excellent", "perfect", "quality", and "awesome".

- Most products contribute between 0.3% to 0.6% of Total Global Sales, while the top 25% of products range from 0.6% to 1.25%.

- Sales data is positively skewed and has heavy tails, indicating that a few high-value sales have a significant impact on the overall distribution, and there may be many outliers to consider when analysing the data.

- There is a strong positive correlation between NA Sales and EU Sales (71%), NA Sales and Global Sales (93%), and EU Sales and Global Sales (88%).

Recommendations for further exploration

We would recommend further exploration on the following areas:

- Loyalty Points: Further analysis can be done to understand what factors affect customer spending and remuneration. This can help the company design strategies to increase customer loyalty and encourage them to spend more.

- Customer Segmentation: Once the customer groups have been identified, the company can tailor its marketing and sales strategies to each group. This can help in increasing customer engagement, retention and revenue.

- Sentiment Analysis: Further analysis can be done to identify the most common themes or topics that customers talk about. This can help the marketing team create targeted campaigns that address the customers' pain points and interests.

- Product Analysis: The company can identify which products contribute the most to sales and profitability. This information can help them focus their resources on developing and marketing these products.

- Sales Data: Further analysis can be done to identify the outliers and understand their impact on the overall distribution. This can help the company make informed decisions about how to manage their inventory and pricing strategy.

- Correlation Analysis: Further analysis can be done to understand the factors that contribute to the high correlation between NA Sales & EU Sales, NA Sales & Global Sales, and EU Sales & Global Sales. This can help the company design strategies to maximize their sales in these regions.