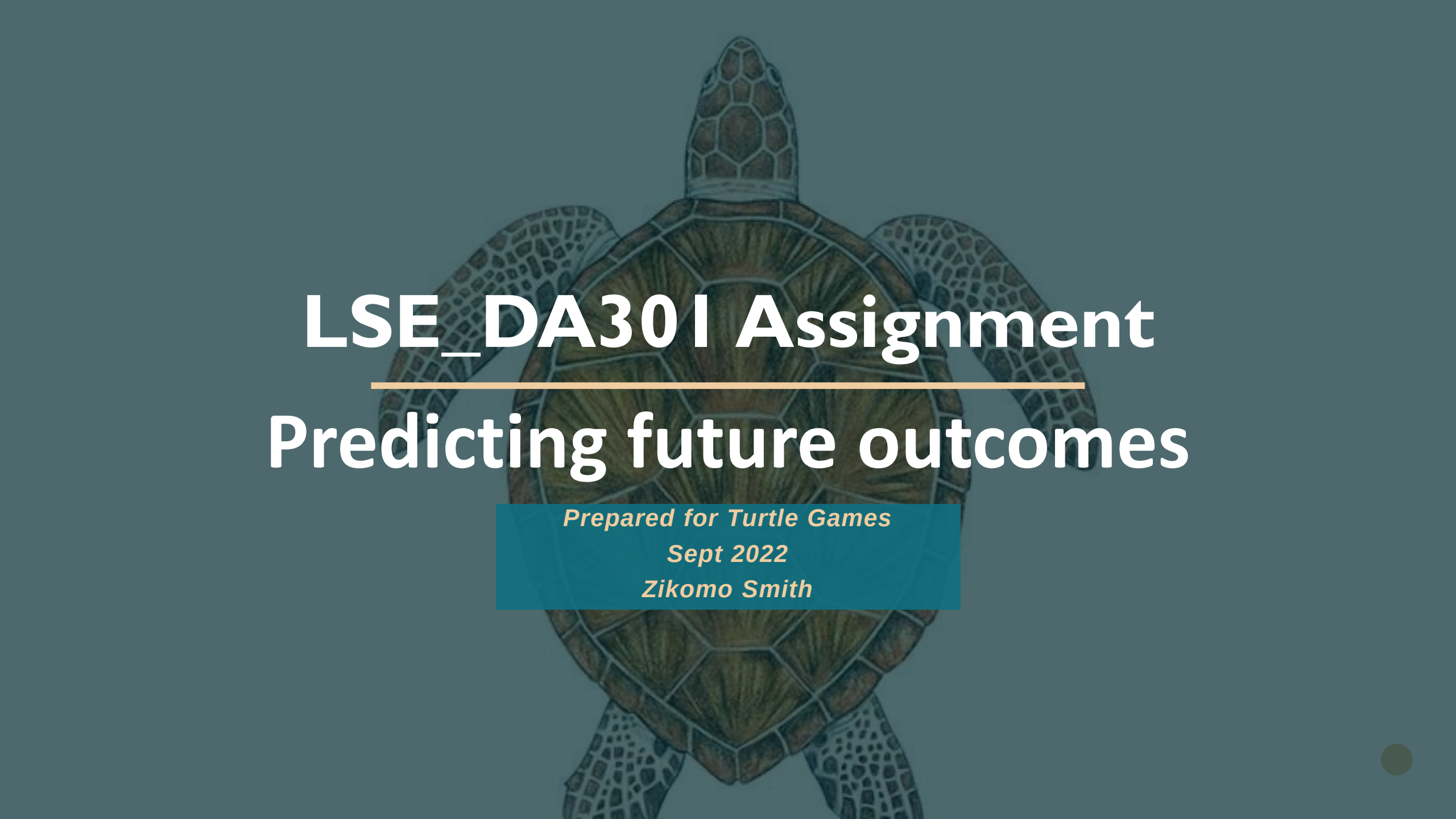


LSE_DA301 Assignment

Predicting future outcomes

Prepared for Turtle Games

Sept 2022

Zikomo Smith

Agenda

1. Background/ Context
2. Results & Recommendations
3. Key patterns and trends
4. The data analysis process
5. Process considerations moving forward

Background/ Context

Context of data analysis

Turtle Games wants to improve its overall sales performance by assessing customer trends to build a business strategy.

Turtle Games must understand the insights that current customer and sales data provide, whether the datasets are reliable, and how to proceed with the data at their disposal.

Recommendations and Predictive Modelling Results

Key recommendations

1. **Create 5 distinct audience groups derived from k-means clustering to identify and market to wealthier customers** who are currently not spending as much as expected with us.

2. **Include customer sentiment as a marketing KPI** to monitor whether marketing maintains the positive perception of Turtle Games.

3. **Reassess the engagement strategy of customers with relatively fewer loyalty point (0-2000)** as they tend to share the most comments.

4. **Increase the sales of product IDs 123, 254, & 948 in Europe to drive incremental revenue** – all three appear within the top 10 products in terms of sales but sell almost double the amount in North America compared to Europe

5. **Refine predictive models further to account for the long tail of products** that are more difficult to assess

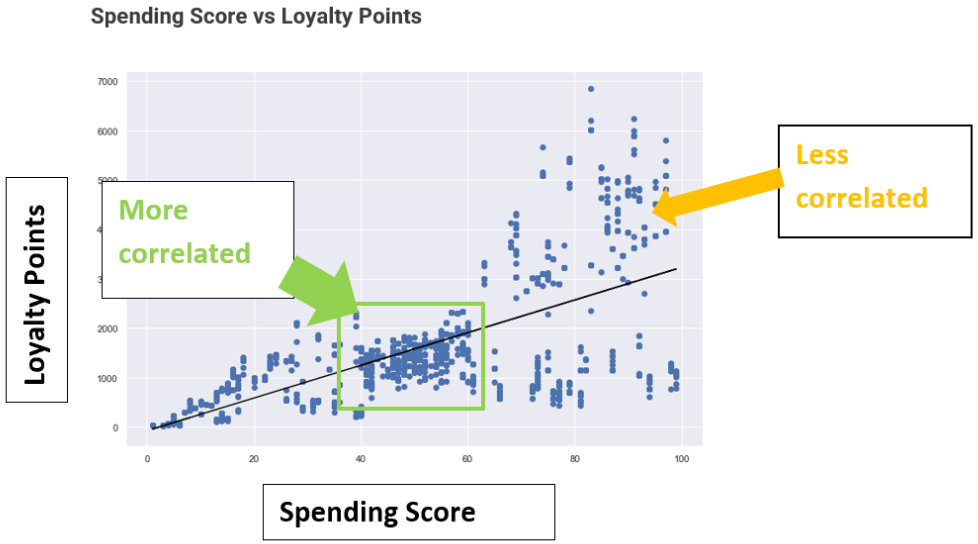
Results of scenario based predictive modelling to determine model robustness

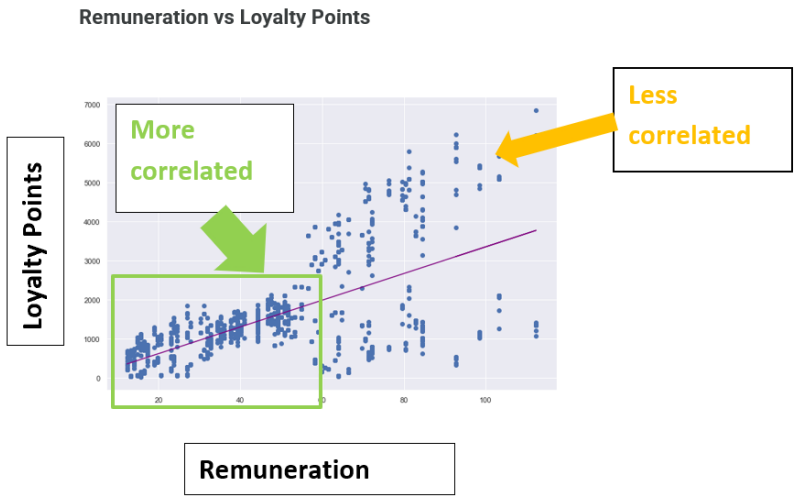
Scenario	EU Sales (M)	NA Sales (M)	Predicted Global Sales (M)	Actual Global Sales (M)
1(Product ID107)	23.80	34.02	66.3587	67.8
2(Product ID99)	3.93	1.56	8.645582	6.04
3(Product ID176)	2.73	0.65	6.282032	4.32
4(Product ID258)	2.26	0.97	6.078803	5.6
5(Product ID326)	22.08	0.52	28.58208	23.21


FIFA 22
EA SPORTS
FIFA
NOWY KOMENTARZ
TOMASZ SMOKOWSKI
JACEK ŁASKOWSKI
Electronic Arts
PC DVD ROM
3
www.pegi.info
BESTCENNIK

Insights

Higher spending customers and richer customers tend to accumulate more loyalty points

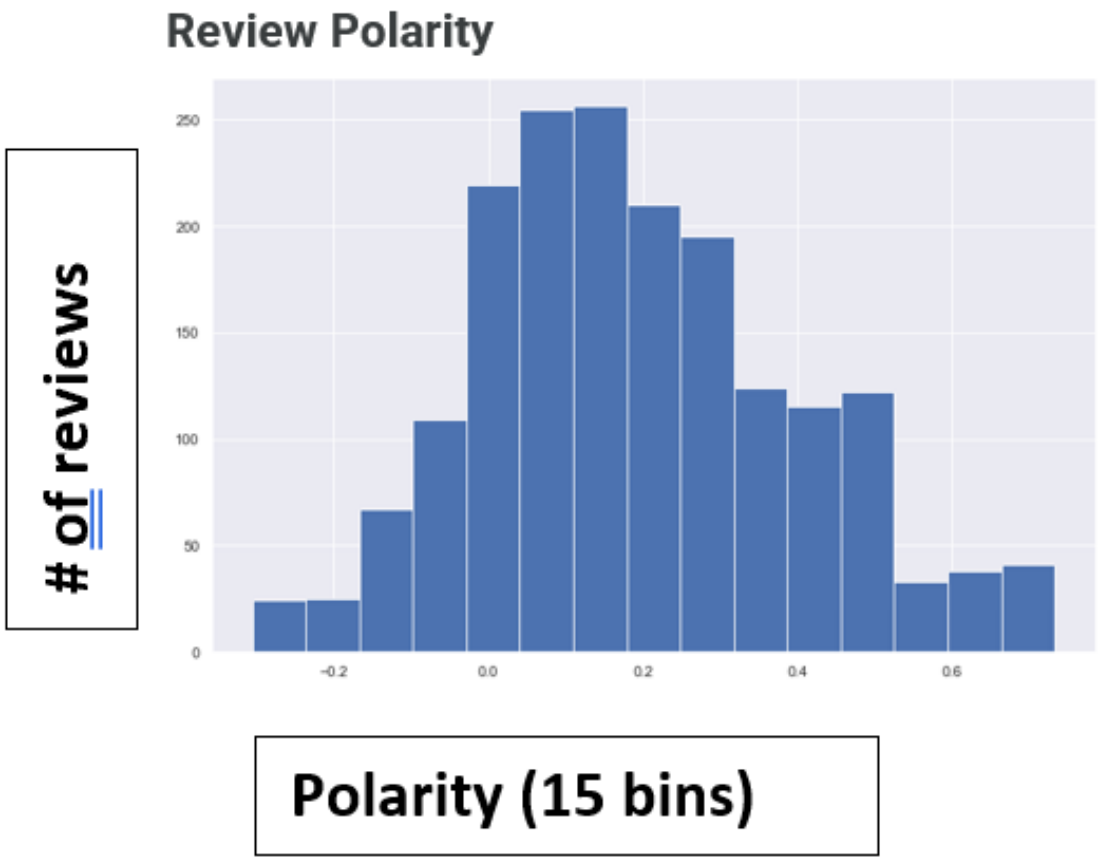




Cluster analysis based on remuneration and spending suggests 5 distinct groups to target

- Cluster analysis grouped by customer spending score relative to customer remuneration.

Current review sentiment is positive



Positive Reviews mentioned both price and quality – the word ‘great’ was particularly prominent.

	summary	polarity_summary
1507	sturdy bright colors awesome	0.850000
155	beautiful	0.850000
1764	perfect item i loved it	0.850000
642	beautifully made	0.850000
35	beautiful coloring book	0.850000
40	so beautiful	0.850000
309	she made four beautiful puppies from the kit a...	0.850000
703	great quality very cute and perfect for my tod...	0.816667
1190	great start for any wargamer looking for orcs ...	0.800000
508	great	0.800000
1262	great for the price	0.800000
199	great product darling puppies	0.800000
1783	great puzzle toy	0.800000
516	great therapist tool	0.800000
1327	great expansion	0.800000
1472	great expansion	0.800000
1805	great for your coffee table	0.800000
1280	a great tile set for any fantasy gaming group	0.800000
1175	another great dungeon command set	0.800000
1473	great expansion set	0.800000

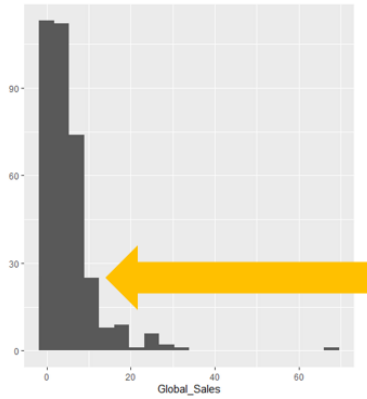
The long tail of product IDS account drive the most sales whilst top products sell unevenly across NA and the EU

Top Selling Products, NA & EU Comparison

Product	NA_Sales_sum	EU_Sales_sum	Global_Sales_sum
<int>	<dbl>	<dbl>	<dbl>
107	34.0	23.8	67.8
515	19.2	18.9	45.9
123	26.6	4.01	37.2
254	21.5	2.42	24.4
195	13	10.6	24.1
231	12.9	9.03	24.1
249	9.24	7.29	25.7
948	14.4	7.79	25.4
876	12.8	9.25	25.3
263	9.33	7.57	24.6

Large delta between NA & EU

Long tail of products

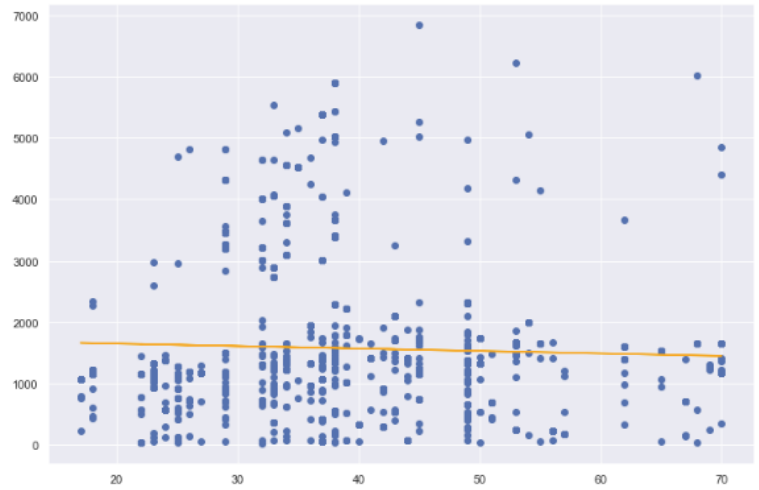


Long tail of products

The Data Analysis Process

Create linear regression models across multiple views

```
Predicted values:  [1659.01044907 1648.66145499 1621.06413743 ... 1579.6681611  1607.26547866
 1586.56749049]
Out[178]: [<matplotlib.lines.Line2D at 0x221bdce6190>]
```

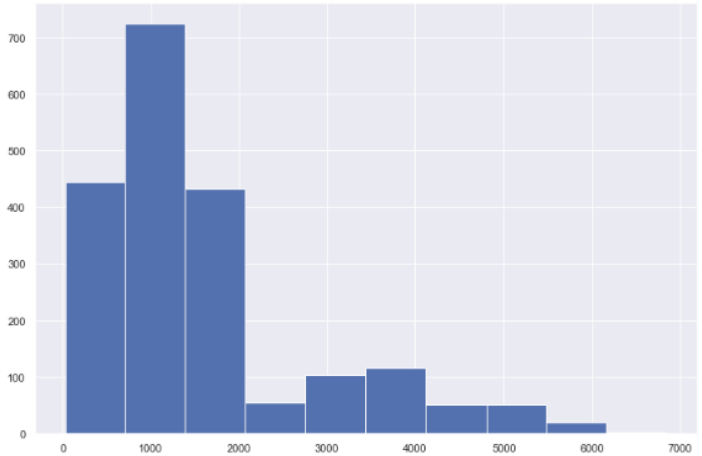



```
In [179]: # Solution - age vs loyalty
```

- Allows us to assess correlation of a variety of factors
- Able to assess whether there was any difference between audiences (no significant difference found).
- Quite cumbersome and time consuming with current code.

Use histograms to group data

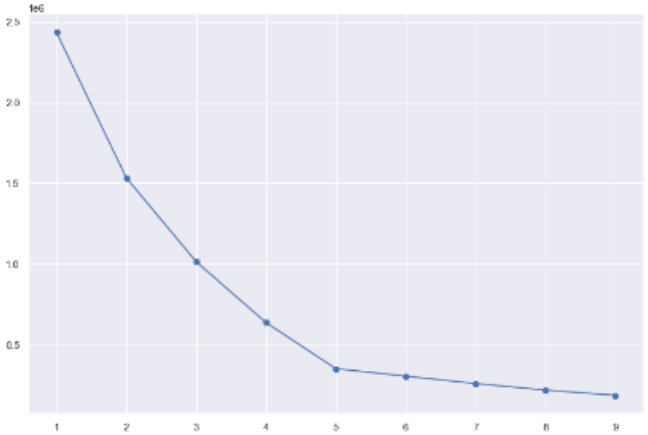
```
In [185]: #create histogram to understand the type of people sharing comments
          plt.hist(reviews['loyalty_points'])

Out[185]: (array([444., 724., 432.,  55., 103., 116.,  51.,  52.,  20.,   3.]),
           array([  25. ,  707.2, 1389.4, 2071.6, 2753.8, 3436. , 4118.2, 4800.4,
                  5482.6, 6164.8, 6847. ]),
           <BarContainer object of 10 artists>)
```



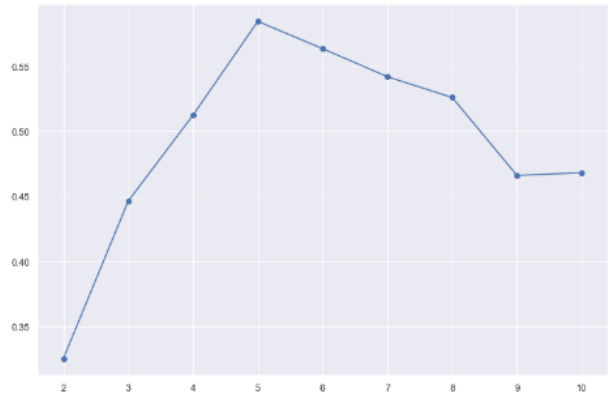
- Allows us to transform numeric loyalty data into a categorical feature of our customers.
- Gives us insight that can inform a loyalty marketing strategy.
- Limited in expressiveness.

Employ elbow and silhouette methods for clustering analysis

Elbow Method



Silhouette Method



- Use of multiple clustering methods producing similar results supports choice of 5 clusters.
- Requires more analysis to see whether we can effectively target clusters in reality.

Use multiple linear regression to make predictions

```
################################################################################

# 3. Create a multiple linear regression model

# Multiple linear regression model.
multi.fit = lm(Global_Sales_sum~EU_Sales_sum+NA_Sales_sum+Product, data=productsales)
summary(multi.fit)

################################################################################

# 4. Predictions based on given values
# Compare with observed values for a number of records.

# Compare with observed values for a number of records.
# Predict based on NA_Sales_sum of 34.02 and EU_Sales_sum of 23.80.

Question1 <- data.frame(EU_Sales_sum=c(23.80), NA_Sales_sum=c(34.02), Product=c(107))

predict(multi.fit, newdata=Question1)
```

- Allows us to predict sales based on region
- Allows for topline insight into relative importance/ potential sales targets per region
- Does not allow for more granular analysis into factors driving sales

Process Considerations Moving Forward

Future-proofing the analysis

1. **Continue the use of GitHub**
    a) Version control system reduces errors
    b) We can make data public to crowdsource solutions in the future

2. **Improve Predictive Modelling**
    a) Current models overshoot the delivered outcome across long tail product
    b) Share sales strategy and assumptions to build into the model and improve accuracy

3. **Incorporate media team analysis to drive strategic value of marketing segments**
    a) Current clustering techniques group customers, but do not let us know how targetable they are.
    b) Sharing extra data from media teams who understand how to activate against audiences will make the clustering analysis even more applicable.

THANK YOU!

Zikomo Smith

Zikomo.smith@mail.mcgill.ca