

## **TURTLE GAMES REPORT**

Turtle Games is a gaming company with a global customer base. They wish to determine how customers accumulate loyalty points, how groups within the customer base can be used to target specific market segments, how social data can be used to inform marketing campaigns.

### **USING PYTHON**

First, I imported the data turtles.csv and then used the Seaborn library to conduct the analysis of the data and used the nltk library to generate word clouds.

First, three simple regression models were built to help Turtle Games understand the relationship between loyalty points ( a dependable variable) and age, remuneration, and spending scores (independent variables). There is a correlation between spending and remuneration with loyalty (R Squared measure 0.45 (spending) and 0.38 (remuneration)) but almost no correlation with age (R2 0.002)).

Using the pair plots we were able to decipher the customer clusters that should be targeted is 5. I then used k-means modelling and using the elbow and shoulder methods I tested the suitability of 5 and other numbers 6/7. the optimal cluster number is k=5. As it provides a good separation between groups. Therefore the marketing team should be advised to focus on the following groups: low income and low spending, low income and higher spending, average income and average spending, higher income and low spending, and higher income and higher spending groups. Turtle Games could conduct further market research into these groups to find out more about what drives their spending.

### **USING R for sales data analysis**

I imported the raw csv file into R studio as a data frame using the tidyverse library and then explored the data. I checked the shape and grouped by product (dataframe). Upon first look I could see that the X360 is the most frequently used platform in the data set, then PS3 and then PC. The relationship between Global, EU and NA sales is roughly linear, as expected. EU, NA and Global sales were all skewed towards the right before removal of outliers. After removing the outliers, the data points do not seem as tightly clustered around the smoothing curve (ex: Global sales versus NA sales). The relationship between the variables is still linear.