Background/context of the business:

*Turtle Games, a leading gaming company, is seeking to understand its customer behaviour, market trends, and the effectiveness of its products. With a diverse range of products from physical board games to digital platforms, understanding the nuances of their market is paramount. The business aims to utilise insights derived from their data to improve product quality, market more effectively, and strategize product releases to maximise customer loyalty and sales***.**

Analysis of Loyalty Point Accumulation

In Module 1, our primary focus revolved around preparing and analysing the data to gain insights into how customers accumulate loyalty points.

- Spending Score vs Loyalty Points: Our analysis indicated a noteworthy trend – as customers spend more, they accumulate more loyalty points. Statistical tests, particularly the p-value, confirmed that spending score significantly contributes to explaining the variation in loyalty points. The graphical representation also supported this, showing a clear linear increase in loyalty points as spending score rises.

- Remuneration vs Loyalty Points: Similar to spending score, an increase in remuneration correlated positively with loyalty points. Moreover, remuneration proved to be highly significant in predicting loyalty points, as evidenced by a high F-statistic and a low p-value. Just like spending score, this relationship displayed a linear increase in loyalty points with higher remuneration.

- Age vs Loyalty Points: The relationship between age and loyalty points showed a slight decrease in loyalty points with increasing age. Despite the visual representation of a declining trend, the statistical results, including the marginally significant p-value and low R-squared, suggested that this regression model had limited explanatory power. In essence, it didn't accurately capture the relationship between age and loyalty points.

Insights from Customer Segmentation

Module 2 delved into customer segmentation to gain a deeper understanding of the different groups of customers

Both clustering methods consistently pointed towards the ideal number of clusters being five. This conclusion was further corroborated by applying the k-means clustering algorithm, which demonstrated the highest similarity among the five clusters. Upon implementing the five-cluster model, we observed that the cluster with the most significant customer representation fell in the middle of the remuneration and spending score variables. This insight suggested that additional focus and efforts should be directed towards maximising sales and customer retention within this cluster. Simultaneously, the other clusters could benefit from tailored engagement strategies to maximise their potential.

Utilising Natural Language Processing for Review Insights

In Module 3, our goal was to identify the most common words and assess the sentiment of the reviews left by customers.

We performed several data manipulation methods and the results strongly suggested that the marketing team should focus on content related to the fun and game nature of the products. Additionally, the word "book" appeared prominently in both lists, signifying an opportunity for targeted marketing efforts which was confirmed by further exploring the rows that included this word.

Sentiment analysis of the columns revealed predominantly positive sentiment. Furthermore, customers of card games and D&D games appeared to be the most satisfied, as these types of games were common purchases. Surprisingly, even in negative reviews, some terms with strongly negative polarity were used in a less critical context. In some cases, customers highlighted how these products served as a solution to anger issues or as a therapeutic coping mechanism, further suggesting that marketing direction could be directed to maximising these areas.

Examining the summary column further, negative sentiment reviews indicated concerns related to the quality of cardboard in the products. This finding suggested that discussions with the production departments might be necessary to address these issues.

Initial exploration of markets and first insights

Product Insights: One revelation is the consistent dominance of the same top-selling product (Product 107) across all markets as well as Wii, PS3 and X360 being the highest selling platforms in the company. Platforms, such as XB and X360, feature popular titles that resonate globally. However, nuances emerge, highlighting the European Union's (EU) ascendancy in sales for platforms like psv, PC, and ps4.

Histograms for the markets reveal some commonalities in the distribution of sales, yet notable differences stand out. NA sales exhibit a higher concentration of smaller sales figures (under 1 million units), while the European market displays a more consistent distribution, with moderate but persistent sales figures across various products. Additionally, the EU presents a more consistent market sales-wise with North America demonstrating a significantly higher number of outliers on their sales.

Deep Dive into Product and Market Dynamics

NA boasts a considerably larger market size compared to the EU, with NA holding a significantly higher market share.

Product-wise, we limit our analysis to the top 25 products per market. Barplots emerge as a preferred choice to uncover valuable insights. They reveal the highest-selling products in each market, confirming that NA sales typically contribute to the biggest part of the global market share.

Normality tests indicate significant deviations from the normal distribution for all three markets. The Shapiro-Wilk normality tests for EU Sales, NA Sales, and Global Sales all yielded low p-values, indicating non-normality in their distributions. Additionally, high skewness and kurtosis values suggest the presence of heavy tails or outliers in these sales data as it was also demonstrated on the plots as well as linear relationships among them.

Making predictions for the business

Both North American (NA) Sales and European Union (EU) Sales exhibit substantial and positive impacts on Global Sales. Each unit increase in EU Sales corresponds to an estimated increase of 1.34197 units in Global Sales, while a unit increase in NA Sales translates to an estimated increase of 1.15543 units in Global Sales.

In our modelling process, we identified multicollinearity as a potential challenge. To address this, we employed Lasso Regression, a regularisation technique that introduces a penalty to the absolute values of the regression coefficients.

Due to the consistency of results that both models offer we would suggest that Leveraging the synergy between EU and NA markets by crafting cross-market strategies that account for the interplay of sales in these regions is paramount. Additionally, exploring expansion opportunities within the EU and NA regions, including in-depth market research and entry into emerging markets within these regions, holds significant potential as well as exploring the potential of region favourites across the other markets.