***Jean-Simon Mpeck Solving Business Problems Using Predictive analytics***

***September 30, 2024***

***Dataset*** : Nike Shoes Sales

***URL Link*** : <https://www.kaggle.com/datasets/ulrikthygepedersen/nike-shoes-sales>

***Problem statement* :**

Nike aims to optimize its product offerings and pricing strategy to enhance customer satisfaction and maximize revenue. Given the dataset, which includes product attributes such as listing pricing, sales price, discounts, ratings, and reviews, the goal is to identify patterns that influence consumer purchasing behavior. This analysis will help in understanding which product characteristics are associated with higher sales and customer satisfaction, enabling Nike to make data-driven decisions.

***Business problem* :**

Nike faces the challenge of optimizing its product offerings and pricing strategies to enhance customer satisfaction and drive potential sales. Despite having a dataset with various product attributes such as listing prices, sale prices, discounts, ratings, and reviews, the company lacks direct sales data to gauge consumer interest accurately. This gap necessitates an in-depth analysis of the existing data to identify key factors influencing perceived value and customer preferences. By leveraging clustering techniques, Nike aims to uncover patterns within the dataset that can inform strategic decisions regarding product positioning and marketing efforts, ultimately ensuring that its offerings resonate with target consumers and maximize market potential.

***Analysis :***

To address the business problem, I will employ a cluster analysis as the primary method.

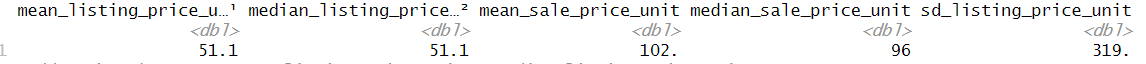
That cluster analysis will help segment products based on their characteristics (price, reviews, rating, etc.), revealing distinct groups that may attract different types of consumers. By analyzing these segments, Nike can gain insights into which attributes are most appealing to consumers, even in the absence of direct sales data, according to the available dataset. I divided the values in the listing\_price and sale\_price cells by 100 in the Excel file to facilitate their proper import into the R environment.

***Justification of the appropriate method :***

The choice of cluster analysis in this study is justified by its ability to uncover inherent groupings within the Nike product dataset, facilitating a nuanced understanding of consumer behavior and product characteristics. By segmenting products based on attributes such as listing price, sale price, rating, and reviews, cluster analysis allows for the identification of distinct market segments that share similar features. This method is particularly effective when dealing with multidimensional data, as it provides insights into patterns and relationships that might not be immediately apparent through univariate analyses. Additionally, cluster analysis supports targeted marketing strategies by enabling Nike to tailor its offerings to specific consumer groups, ultimately enhancing customer satisfaction and optimizing inventory management. The method's flexibility and ability to handle varied data types make it a robust choice for this analysis, ensuring that the findings are both actionable and relevant to the company's strategic objectives.

***Results :***

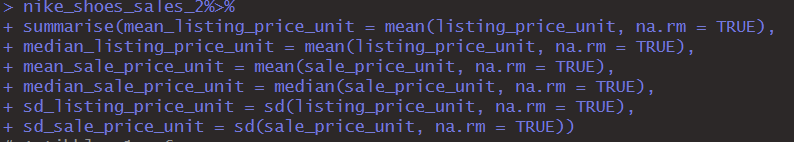
*Descriptive analysis*

****

The descriptive analysis of the dataset reveals insightful statistics regarding the pricing of Nike shoes. The mean listing price is approximately $51.10, indicating that, on average, shoes are priced affordably within a competitive market segment. The median listing price is also $51.10, suggesting that the distribution of prices is symmetrical and that half of the products are priced below this threshold, reinforcing the idea of accessibility for consumers.

In terms of sale prices, the average is significantly higher at $102.00, while the median sale price is slightly lower at $96.00. This discrepancy between the mean and median sale prices indicates that a few high-priced items might be influencing the average, suggesting a diverse product range with some premium offerings. The substantial standard deviation of $319.00 in the listing price further highlights this variability, reflecting the presence of both budget-friendly and high-end Nike products in the inventory. Overall, these statistics suggest that Nike effectively caters to a wide range of consumers, balancing affordability with higher-end options, which may enhance its market appeal and competitiveness.

**R code**

******

*Cluster Centers Analysis*

**Cluster 1 :**

**Listing Price Unit:** 0.86 (high relative to the mean)

**Sale Price Unit :** 0.61 (also high)

**Rating :** 0.68 (above average)

**Reviews *:*** 3.68 (moderate volume of reviews*)*

Cluster 1 represents high-end Nike products, likely with premium pricing. These items have received relatively positive ratings and a moderate number of reviews, indicating a strong market presence among consumers willing to pay more.

**Cluster 2 :**

**Listing Price Unit** : -0.02 (around average)

**Sale Price Unit** : 0.04 (slightly above average)

**Rating** : 0.75 (good ratings)

**Reviews** : 0.03 (very few reviews)

Cluster 2 appears to contain products that are fairly priced, with good ratings but very few reviews. This suggests they may be newer or less popular models that haven't garnered significant consumer attention yet, but they are well-received by those who have purchased them.

**Cluster 3 :**

**Listing Price Unit:** -0.07 (slightly below average)

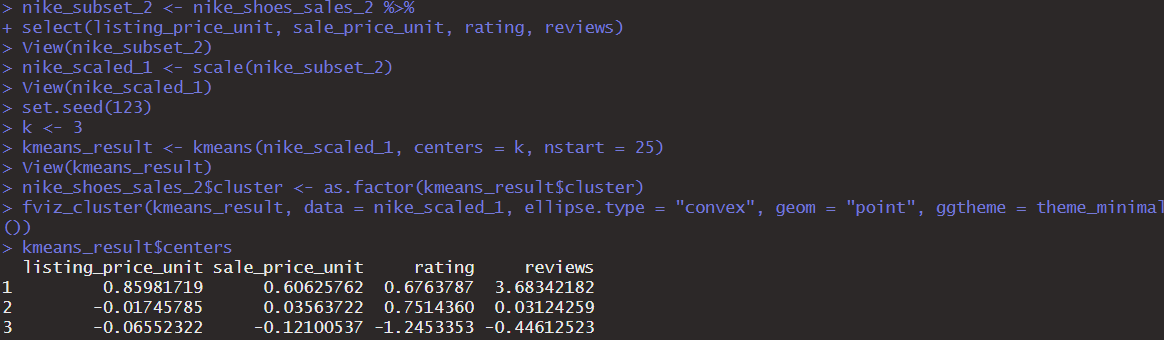
**Sale Price Unit:** -0.12 (low)

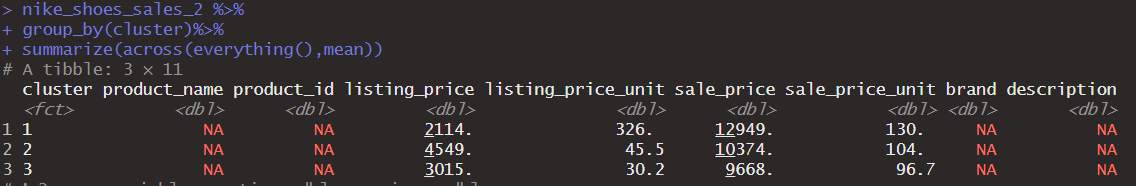
**Rating:** -1.25 (low ratings)

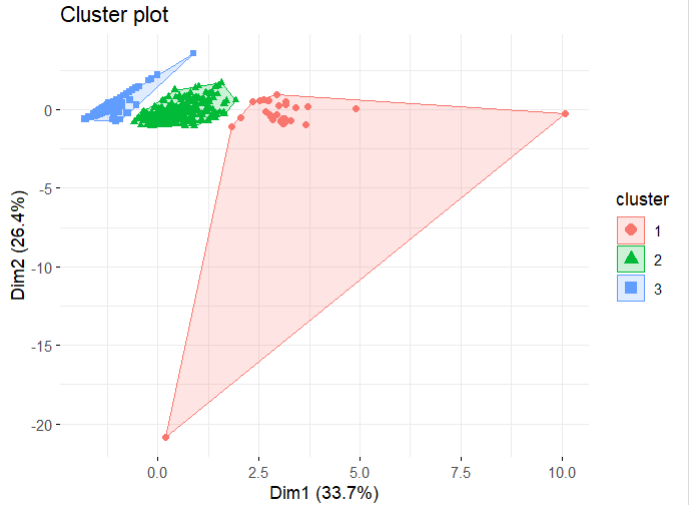
**Reviews:** -0.45 (below average volume)

Cluster 3 seems to represent lower-end products or those facing challenges in the market. These items have lower pricing and poor ratings, indicating that they may not meet consumer expectations. The low volume of reviews further supports the idea that they are either unpopular or underperforming.

**R code**





******

***Suggestions :***

Moving forward, it is essential to implement a few strategic recommendations based on the analysis outcomes. First, focus on enhancing the product offerings in Cluster 1, where the listing prices and sale prices indicate high potential profitability, while also addressing the relatively moderate review scores. This could involve targeted marketing campaigns or product improvements to boost customer satisfaction and drive higher ratings.

Next, consider exploring the characteristics of Cluster 2, which has lower listing and sale prices but better ratings. This cluster may represent a price-sensitive segment with strong demand for affordable, high-quality products. Tailoring product lines to meet the needs of this demographic could capture additional market share.

Revisit the normalization process for the reviews to ensure all features are comparable, which will enhance the accuracy of future analyses. Implementing continuous monitoring and analysis of consumer feedback can provide real-time insights to adapt strategies dynamically.

Last but not least, expanding the dataset to include sales data, customer demographics, and additional product attributes would facilitate deeper insights into consumer behavior, enhancing the overall effectiveness of clustering techniques and guiding more informed decision-making.

***Limitations :***

Several limitations should be considered when interpreting the results of this analysis. First, the dataset's size, with only 643 records, may limit the generalizability of the findings. A larger dataset would provide more robust insights and improve the reliability of the clusters identified.

Second, the normalization of certain features, such as reviews, may lead to misinterpretations if not accurately scaled or if outliers are present. Additionally, the absence of sales data restricts the understanding of actual market performance, making it challenging to correlate the clustering results with real sales trends.

While cluster analysis is useful for identifying patterns, it doesn’t establish causal relationships. External factors influencing customer behavior, such as seasonal trends or competitive actions, may not be captured in the dataset.

Finally, the choice of clustering method and the number of clusters can significantly affect the outcomes. If the initial assumptions about the data distribution or structure are incorrect, the resulting clusters may not accurately reflect meaningful segments of the market. Hence, ongoing evaluation and refinement of the analytical approach are crucial for improving future analyses.

***Conclusion :***

The analysis reveals distinct segments within Nike's product offerings. The high-end products (Cluster 1) are performing well in terms of consumer satisfaction, which aligns with the company's strategy to target premium segments. Conversely, Cluster 2 indicates potential opportunities for growth by leveraging positive ratings to boost visibility and sales, even with low review counts. Lastly, the underperforming products in Cluster 3 require immediate attention, either through product redesign, marketing strategies, or discontinuation, to improve overall brand perception and profitability.

This clustering analysis provides valuable insights that Nike can use to tailor its marketing strategies, enhance product offerings, and better align with consumer preferences. Addressing the challenges in Cluster 3 while capitalizing on the strengths of Clusters 1 and 2 could drive overall growth and strengthen Nike's dominant market position.