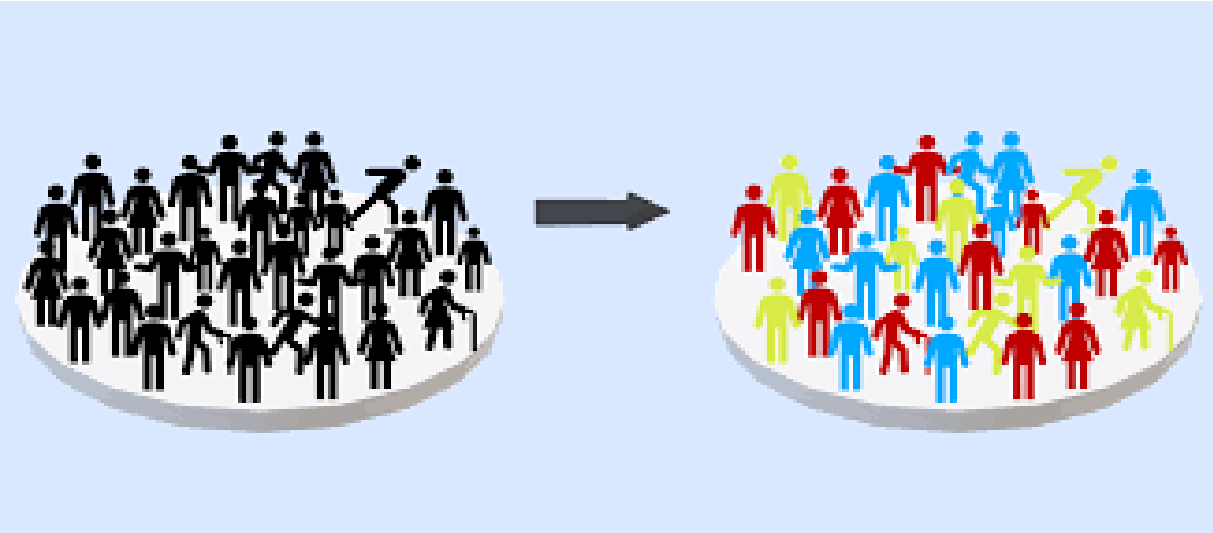
Analytics Specialization and Applications (BUSI4370 UNUK) (SPR1 23-24)

Coursework 1- Customer Analytics Study



Student ID: 20542740

1.Executive Summary:

Task Description:
The objective of this analysis is to conduct a comprehensive examination of customer data obtained from a national convenience store chain. The dataset comprises of four files containing transactional information for 3000 customers over a span of six months. By segmenting customers based on their spending behavior across various product categories, visit frequencies, and quantities purchased, businesses can derive valuable insights into customer preferences. These insights can then inform tailored marketing strategies aimed at enhancing customer engagement and satisfaction.

Technical Approach:
The technical approach involves loading the datasets and conducting exploratory data analysis (EDA) to understand its structure and characteristics. Data cleaning techniques are then applied to address missing values and outliers. Furthermore, data types are adjusted as needed to ensure consistency and accuracy throughout the analysis process, enhancing the reliability of the findings. Feature engineering is performed to create new variables capturing important aspects of customer behavior. Feature selection helps identify the most relevant features, reducing dimensionality and computational complexity. Combining relevant tables consolidates information on spending patterns, visit frequencies, and demographic characteristics into a single dataset. Feature generation or selection enhances the dataset's quality and interpretability, facilitating meaningful analysis. Clustering algorithms, such as k-means, segment the dataset into distinct groups based on similarities in feature space. Cluster evaluation metrics validate the effectiveness of the segmentation, ensuring meaningful patterns are captured. The technical approach combines data aggregation, feature scaling, principal component analysis (PCA), and k-means clustering to derive actionable insights from the data.

Data Summary:
In addition to the comprehensive examination of customer data, the dataset comprises multiple tables providing detailed insights into customer spending. These tables are categorized by product categories and include demographic details such as total spend, total quantity purchased, and average spend etc. The preprocessing steps applied to the dataset involve handling missing values, converting currency values, and creating new features that capture relevant aspects of customer behavior. These below datasets collectively contribute to a rich dataset that enables a holistic understanding of customer behavior, allowing for robust analysis and strategic decision-making.

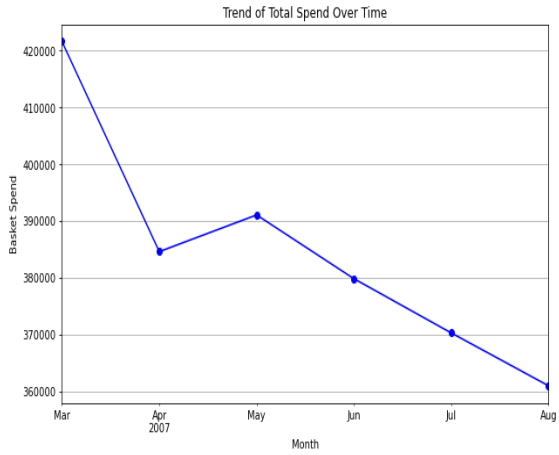
Datasets	Line items	Customers	Baskets	Category spend
No. of Datapoints	1,461,315	3,000	195,547	3000
No. of Features	6	5	5	20

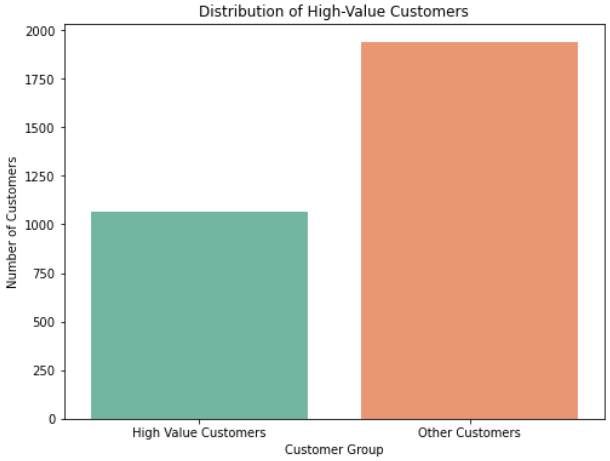
Results Summary:
Following the clustering analysis, six distinct customer segments are identified based on their spending patterns and demographic characteristics. Each segment exhibits unique behaviors and preferences, enabling businesses to tailor their marketing efforts and service offerings to better meet customer needs. The clusters are evaluated using metrics such as total spend, total quantity purchased, and average spending per unit quantity for various given categories to assess their effectiveness in capturing meaningful differences among customer groups.

Insights Summary:

The clustering analysis yields actionable insights into customer behavior and preferences, enabling businesses to make informed decisions regarding marketing strategies, product offerings, and customer engagement initiatives. Insights include the identification of high-value customer segments, understanding of spending preferences across product categories, and opportunities for targeted promotions and personalized marketing campaigns.

The diagrams provide key insights into customer spending behavior above threshold value assuming to be 800 and trend of total spend over time, aiding in strategic decision-making and targeted marketing efforts.

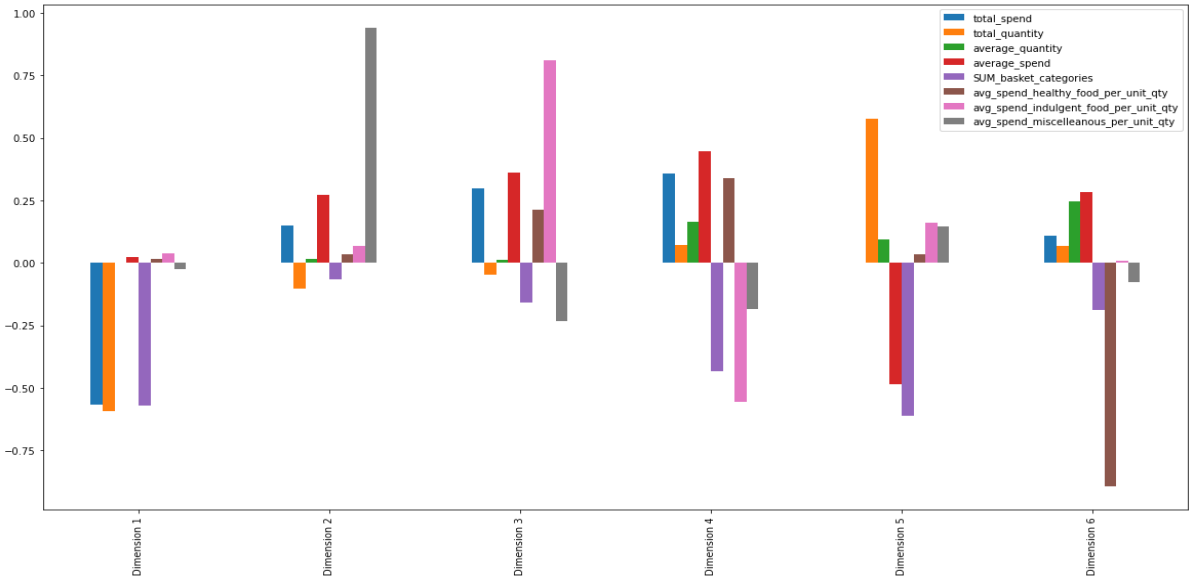




2.Feature Description:

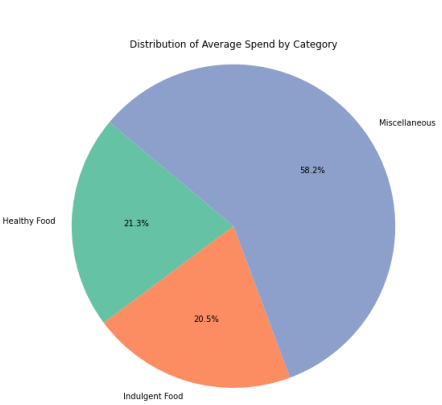
In the realm of data analysis, identifying the most relevant features is crucial. Key indicators such as "Total Spend," "Total Quantity," "Average Quantity," "Average Spend," and "Sum of Basket Categories" provide essential insights into customer behavior, encompassing aspects like purchasing power, frequency, and spending habits. These metrics serve as the foundation for robust analysis and strategic decision-making. In addition to leveraging existing features obtained through data aggregation at the customer level, the creation of novel metrics enhances our understanding of customer behavior. Transactions grouped by customer numbers, along with aggregated quantities and spends across distinct product categories, reveal nuanced preferences and purchasing patterns. Categories like "Healthy Food," "Indulgent Food," and "Miscellaneous Items" mirror diverse consumer choices, catering to health-conscious, indulgent, and convenience-driven preferences, respectively.

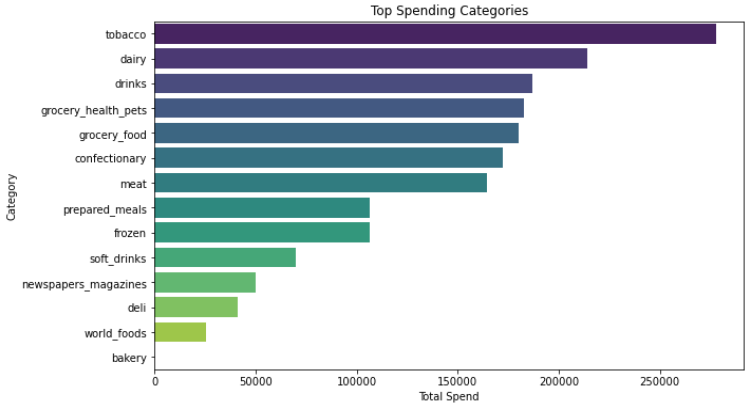
The introduction of the "Average Spend Per Unit Quantity" metric further enhances insights into spending efficiency within each category, offering a standardized measure across product types. Furthermore, employing principal component analysis (PCA) aids in identifying underlying data patterns and relationships. By reducing dimensionality, PCA clarifies interrelationships between features, facilitating visualization and interpretation. Application of PCA as a feature engineering technique to reduce dimensionality and eliminate correlations among the original features. The analysis revealed new dimensions, or components, each representing a combination of the original categories present in the data. By interpreting these components, we gained insights into the underlying structure of customer spending behavior. PCA results highlight significant proportions of variance explained by the first and second principal components (84.4%), as well as the first four components (96.37%). This underscores the suitability of these dimensions for visualization and analysis, enabling informed decision-making processes and final feature selection for clustering. The diagram below depicts the application of PCA. First principal component is associated with large increases in "total spend", "total quantity" and "sum basket categories" features.



3. Customer Base Summary:

The exploratory analysis of the customer base reveals a diverse range of spending patterns and preferences among customers. Aggregating spending and quantity data across product categories provides detailed insights into customer spending within each product category. Total spend and Total quantity varies widely, indicating diverse purchasing power. Categorizing transactions into Healthy Food, Indulgent Food, and Miscellaneous Items highlights consumer preferences. The dominance of Miscellaneous Items underscores the need to cater to diverse needs, with Healthy Food and Indulgent Food reflecting evolving trends in health-consciousness and indulgent consumption. These insights enable businesses to prioritize marketing efforts and allocate resources effectively, tailoring product assortments and initiatives to resonate with target audiences. Below diagrams highlight the distribution of average spend by category and top spending categories.

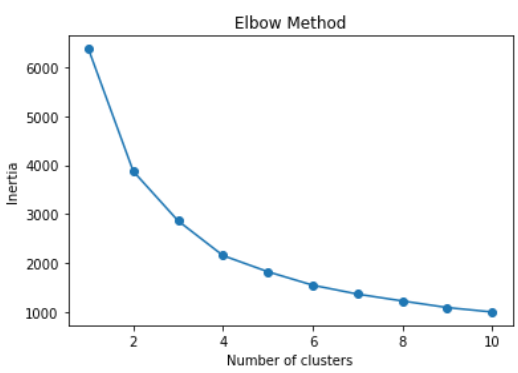


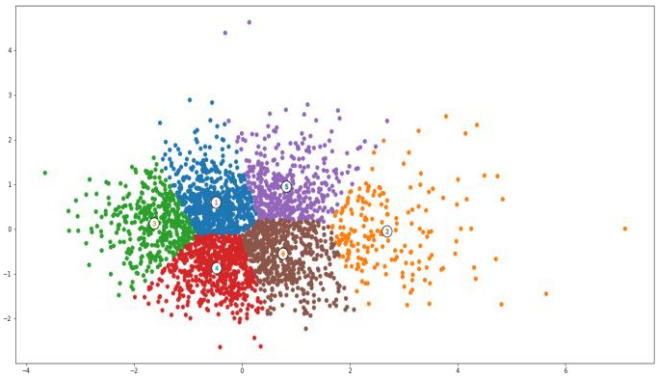


4.Segmentation Methodology:

The segmentation methodology utilizes the K-means clustering algorithm to group customers based on spending behavior, quantity purchased, and product category preferences. Determining the optimal number of clusters (k) is crucial, balancing model complexity with interpretability. The K-means algorithm iteratively to partition datasets into k clusters, aiming to optimize cluster cohesion and separation, thereby uncovering distinct customer segments. The approach integrates the Elbow Method and Silhouette Score to ascertain the optimal k. The Elbow Method identifies the inflection point in cluster cohesion, guiding the selection of an appropriate number of clusters. The diagrams below shows that K=6 is the optimal value with cluster visualization. Additionally, in practical terms,

having six segments allows for a nuanced understanding of customer behavior and preferences, enabling tailored marketing strategies and personalized customer experiences.





Concurrently, the Silhouette Score quantifies cluster quality, ensuring a balance between intra-cluster cohesion and inter-cluster separation. In subsequent steps, we systematically explore a range of k values, spanning from 2 to 10 clusters. For each k value, we meticulously evaluate the Silhouette Score, quantifying the effectiveness of clustering in capturing meaningful customer segments. The k value associated with the highest Silhouette Score selected as the optimal number of clusters, ensuring the robustness and effectiveness of the segmentation approach. K values are in the range of 0.35 to 0.32.

5.Results:

Upon conducting K-means clustering analysis with 6 clusters, we have segmented our customer base to gain insights into distinct purchasing behaviors. Following delineates each segment both statistically and descriptively:

Segment 1: Balanced Shoppers

Statistically: Segment 1 comprises 694 customers with a mean total spend of £910 and a mean total quantity of 600 items. These customers exhibit a balanced approach to spending, reflected in their moderate spending habits across different product categories. On average, they allocate £1.11 per unit quantity on healthy foods, £1.07 on indulgent foods, and £4.40 on miscellaneous items. ***Descriptively***: Individuals in this segment are pragmatic shoppers who prioritize value for money. They tend to make informed choices, opting for a mix of healthy and indulgent items while also exploring miscellaneous options to meet their diverse needs.

Segment 2: Budget-Conscious Buyers

Statistically: In Segment 2, consisting of 194 customers, individuals exhibit frugal spending habits.The average total spend is £148, and the mean total quantity is 108.55 items. Despite their limited spending, these customers allocate a relatively higher amount towards indulgent foods (£1.246 per unit quantity) compared to healthy foods (£1.198) and miscellaneous items (£2.618).

Descriptively: This segment represents budget-conscious shoppers who prioritize affordability over luxury. They are likely to seek value in their purchases and prioritize cost-effectiveness in their shopping behaviors.

Segment 3: Gourmet Enthusiasts

Statistically: Segment 3 consists of 425 customers who indulge in high-quality foods. Mean total spend is significantly higher at £1705.7, with a mean total quantity of 1278 items. Despite their focus on gourmet items, these customers allocate a moderate amount towards healthy foods (£1.11 per unit quantity) and miscellaneous items (£2.803).

Descriptively: This segment represents gourmet enthusiasts who appreciate premium products. They are willing to invest in high-quality items to elevate their culinary experience while maintaining a balanced approach to their spending habits.

Segment 4: Pragmatic Shoppers

Statistically: With 649 customers, Segment 4 comprises pragmatic shoppers who prioritize practicality in their purchases, the mean total spend is £745, and the mean total quantity is 701 items. These individuals exhibit moderate spending habits and allocate a relatively lower amount towards indulgent foods (£0.96) compared to healthier options (£1.049) and miscellaneous items (£1.036).

Descriptively: This segment represents shoppers who value efficiency and seek to optimize their spending without compromising on quality. They make pragmatic choices based on their needs and preferences.

Segment 5: Health-Conscious Consumers

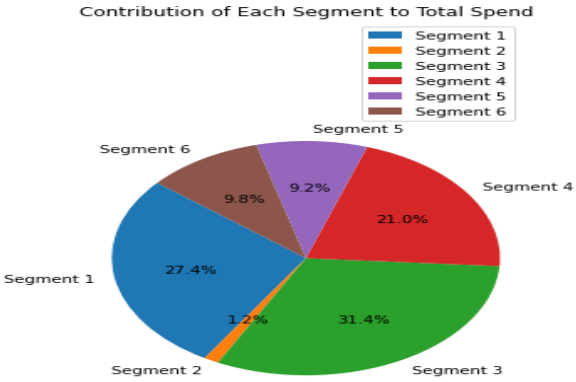
Statistically: Segment 5, consisting of 448 customers, represents health-conscious consumers who prioritize nutritious options in their purchases, the mean total spend is £475, with a mean total quantity of 271 items. The average spend per unit for healthy food is £1.15, reflecting their commitment to maintaining a balanced diet. Meanwhile, they also allocate a significant portion of their spending to indulgent treats, with an average spend per unit of £1.20. For miscellaneous items, they spend an average of £6.35 per unit.

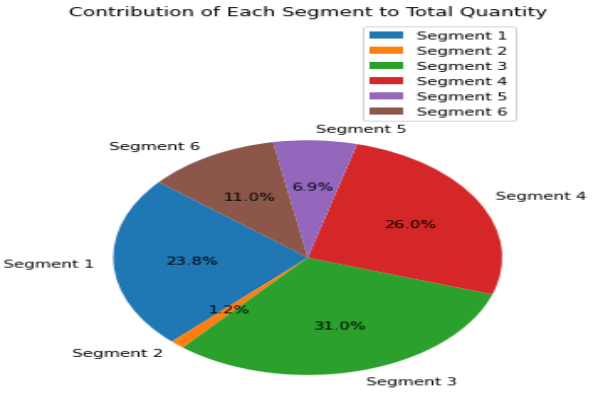
Descriptively: This segment reflects shoppers who prioritize their well-being and actively seek out nutritious options to support their healthy lifestyle choices.

Segment 6: Convenience Seekers

Statistically: In Segment 6, comprising 590 customers, individuals prioritize convenience and efficiency in their purchases, the mean total spend is £382.04, and the mean total quantity is 327.58 items. These customers exhibit moderate spending habits and allocate a balanced amount towards healthy foods (£1.12 per unit quantity), indulgent foods (£1.059), and miscellaneous items (£1.399).

Descriptively: This segment represents shoppers who prioritize ease and accessibility in their shopping experience. They often opt for quick and convenient meal solutions, seeking to streamline their shopping process while still making informed choices.





6.Summary:

The clustering analysis unveiled distinct customer segments reflecting varied spending patterns and product preferences. This segmentation facilitates targeted marketing efforts, personalized product recommendations, and optimized inventory management. By discerning the unique needs of each segment, businesses can enhance customer satisfaction, drive revenue growth, and foster lasting customer relationships. Leveraging this data-driven approach enables companies to allocate resources effectively, enhance customer engagement, and gain a competitive advantage in the market, ultimately maximizing profitability.

Business Case for Clustering Solution: Segmenting customers allows us to tailor our marketing efforts and product offerings to cater to the specific needs and preferences of each segment. Gaining

a deeper understanding of the distinct behaviors exhibited by different consumer groups, we can allocate resources more efficiently, improve customer targeting, and ultimately drive revenue growth.

Recommendation for Key Segments: Based on the analysis, two segments that are of particular importance for the company to focus attention on are Segment 3, "Gourmet Enthusiasts" and Segment 5, "The Health-Conscious Consumers."

1.Segment 3: Gourmet Enthusiasts: This segment represents customers who prioritize premium and gourmet food items. Given their willingness to invest in high-quality products, targeting this segment with premium offerings and exclusive promotions can yield significant returns. By curating a selection of gourmet products and enhancing the shopping experience for these customers, the company can capitalize on their discerning tastes and increase both revenue and brand loyalty.

2.Segment 5: Health-Conscious Consumers: With a focus on health and wellness, this segment presents opportunities for the company to expand its offerings of nutritious and health-conscious products. By catering to the unique needs and preferences of wellness enthusiasts, such as offering organic, gluten-free, or plant-based options, the company can position itself as a trusted provider of healthy choices. Investing in marketing campaigns and partnerships that resonate with this segment's values can help attract and retain these customers, driving long-term growth and market differentiation.

Key Insights and Recommendations:

1.Customized Marketing Strategies: Tailoring marketing messages and promotions to align with the unique preferences and behaviors of each segment is crucial for enhancing customer engagement and driving sales.

2.Product Assortment Optimization: Optimizing our product assortments to cater to the preferences of different segments can significantly improve customer satisfaction and loyalty.

3.Continuous Monitoring and Analysis: Regularly monitoring customer behaviors and preferences through data analysis enables us to make dynamic adjustments to our marketing strategies and product offerings, ensuring they remain aligned with evolving consumer demands.

Further Recommendations for Business Analysis: To further refine the marketing strategies and drive sustainable growth, the recommendation is to conduct in-depth analysis in the following areas:

1.Segment-specific Product Development: Exploring opportunities to develop new products or services tailored to the specific needs and preferences of key segments can help us capture additional market share and increase customer loyalty.

2.Customer Lifetime Value Analysis: Analyzing the lifetime value of customers within each segment will enable us to prioritize resource allocation and retention efforts effectively, ensuring maximum return on investment.

3.Market Basket Analysis: Conducting market basket analysis to understand cross-selling and upselling opportunities within each segment will allow us to optimize revenue generation and enhance overall customer experience.

4. Leveraging social media and digital marketing channels: to engage with customers across different segments and enhance brand visibility and awareness.

5.Sentiment analysis: Gain insights into customer sentiment and satisfaction through sentiment analysis of customer reviews and feedback, informing reputation management and customer experience improvements.

By leveraging the insights gained from our clustering analysis and implementing targeted strategies, we can strengthen our competitive position, foster deeper customer relationships, and drive sustainable business growth.