Customer Analytics Report

Customer Analytics:

The importance of understanding customers cannot be overstated in running a successful business, with KYC (Know Your Customer) being crucial for many companies. KYC facilitates tailoring offerings to customer needs, enhancing communication and delivery. Customer analytics, especially segmentation, is vital for effective marketing, encompassing various customer characteristics and behaviours. The STP (Segmentation, Targeting, Positioning) framework is fundamental for exploring and understanding potential customers.

The dataset used here pertains to a B2C (Business to Consumer) model, offering ample data points for analysis. FMCG (Fast Moving Consumer Goods) companies, like supermarkets, provide abundant data, making them ideal for customer analytics courses. Segmentation involves dividing customers into groups with similar characteristics, aiding in predicting purchase behaviours and marketing responses. Marketers utilize demographic, geographic, and psychographic data for segmentation when consumer behavior data is limited or available. Targeting evaluates segment profitability, determining which segments to focus on and how to promote products. Positioning entails aligning product characteristics with customer needs, considering presentation and distribution channels. The Marketing Mix framework further guides positioning by addressing product presentation and distribution strategies.

The concept of the marketing mix is to develop the best product or service and offer it at the right price through the right channels. In this case study, we'll perform customer analytics that answers three fundamental questions about positioning and the marketing mix.

1.Will the customer buy a product from a particular product category when they enter the shop?

2.Which brand is the customer going to choose?

3.How many units is the customer going to purchase?

These three questions outline three main components of the purchase process and examine how each of them is influenced by the marketing mix tools. These components are

1.The customer visits a store and decides whether or not to buy a product.

2.Provided that the customer had decided to buy a product from the product category of interest,

the customer chooses which of the available brands to buy.

3.Lastly, the customer buys a particular quantity or number of items that the product from the selected brand.

The marketing mix comprises four groups of variables: product, price, promotion, and place, often referred to as the four P's of Marketing. Product encompasses core attributes such as features, design, branding, and packaging. Price involves decisions regarding product cost, long-term pricing, discounts, and payment terms. Promotion extends beyond mere discounts to include communication and advertising strategies, encompassing TV commercials, flyers, and unique events like sending a Tesla into space. Sales promotions, known as merchandising, include price reductions, display promotions, and feature promotions. Display promotions involve placing products in prominent locations, while feature promotions utilize various marketing materials like printed ads.

Place refers to distribution strategies, categorized into intensive distribution, selective distribution, and exclusive distribution. Intensive distribution involves widespread availability across many stores, while selective distribution limits availability to select stores. Exclusive distribution restricts products to specific outlets, often associated with luxury items like Tesla or Rolex. The marketing mix tools aid in making decisions regarding product characteristics, pricing, promotion strategies, and distribution channels.

Segmentation of Data:

The segmentation dataset comprises information on 2000 individuals from a specific area, treated as representative of the entire county or district. This data is derived from the purchasing behaviour of these individuals in an FMCG store, collected through loyalty cards used at checkout. Pre-processing steps include numerical encoding of variables and handling missing values, while ensuring data volume restriction and customer privacy protection. The dataset consists of seven demographic and geodemographic variables alongside an individual ID column. The columns represent individual attributes as follows: biological sex (coded as 0 for male and 1 for female), marital status (coded as 0 for single and 1 for non-single), age in years, education level (coded from 0 to 3 for other/unknown education, high school, university, and graduate school, respectively), annual income in dollars, occupation (coded as 0 for unemployed/unskilled, 1 for skilled employee/official, and 2 for management/self-employed/highly qualified employee/officer), and settlement size (coded as 0 for small city, 1 for mid-sized city, and 2 for big city).

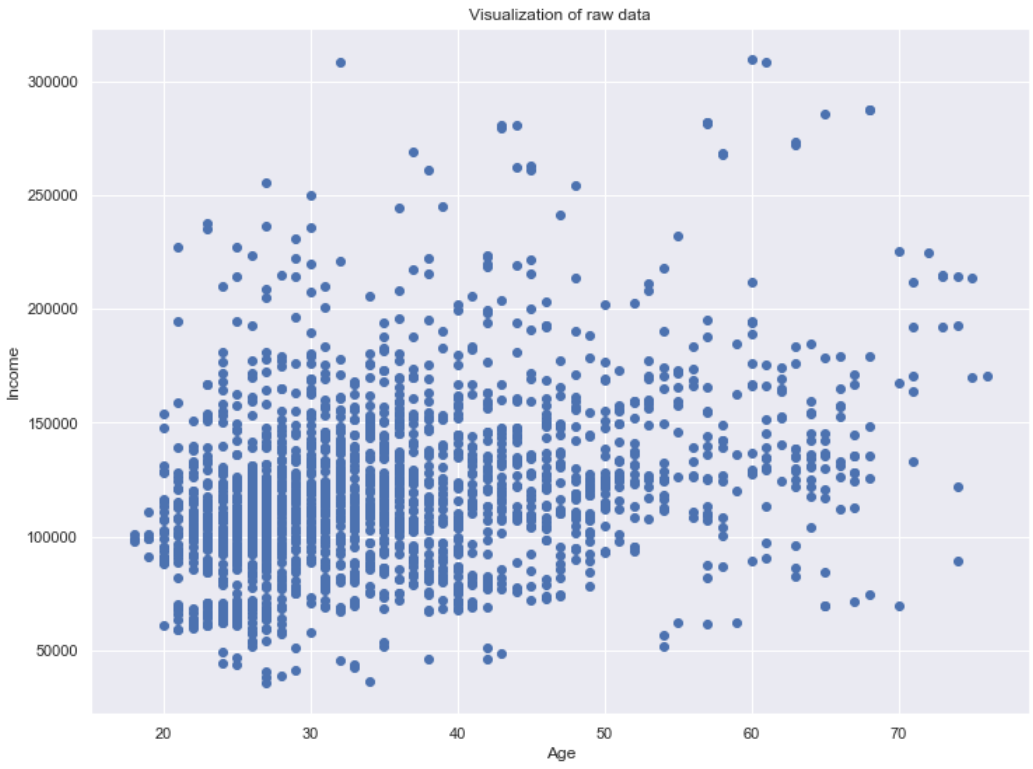
We have done an initial exploration of the dataset. We inferred that proportion of women in dataset=45.7% and the average value of the age is 35.9 years and average income is 120954.41.

Further analysis involved calculating Pearson correlations between variables to explore relationships. Strong positive correlations were observed between age and education ($r = 0.65$), as well as between occupation and income ($r = 0.68$). These correlations indicated that older individuals tended to have higher levels of education, and higher income was associated with certain occupations. Additionally, positive collinearities were observed, such as between occupation and settlement

size, suggesting potential relationships between employment status and urbanization.

	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
Sex	1.000000	0.566511	-0.182885	0.244838	-0.195146	-0.202491	-0.300803
Marital status	0.566511	1.000000	-0.213178	0.374017	-0.073528	-0.029490	-0.097041
Age	-0.182885	-0.213178	1.000000	0.654605	0.340610	0.108388	0.119751
Education	0.244838	0.374017	0.654605	1.000000	0.233459	0.064524	0.034732
Income	-0.195146	-0.073528	0.340610	0.233459	1.000000	0.680357	0.490881
Occupation	-0.202491	-0.029490	0.108388	0.064524	0.680357	1.000000	0.571795
Settlement size	-0.300803	-0.097041	0.119751	0.034732	0.490881	0.571795	1.000000

To enhance visualization of correlations, a heatmap was created using Matplotlib and Seaborn libraries, displaying positive correlations in blue and negative correlations in red. The heatmap revealed strong positive correlations between age and education, occupation and income, as well as other positive collinearities such as occupation and settlement size. These insights into feature relationships are crucial for segmentation purposes, enabling the identification of similar consumer groups.



Next preprocessing step we need to perform is standardization. Standardization is crucial for our segmentation models as it ensures that all features are treated equally and that the differences between their values are comparable. Without standardization, the algorithms may assign disproportionate weights to certain features based solely on their numerical values, leading to biased results. By transforming the features to fall within the same numerical range, standardization

enables fair comparison and accurate modelling. After standardization, we will now start building segmentation models.

Hierarchical clustering:

The primary objective of clustering is to group individual observations in such a way that observations within the same group are highly similar to each other while being distinctly different from observations in other groups. There are two main types of clustering: hierarchical and flat.

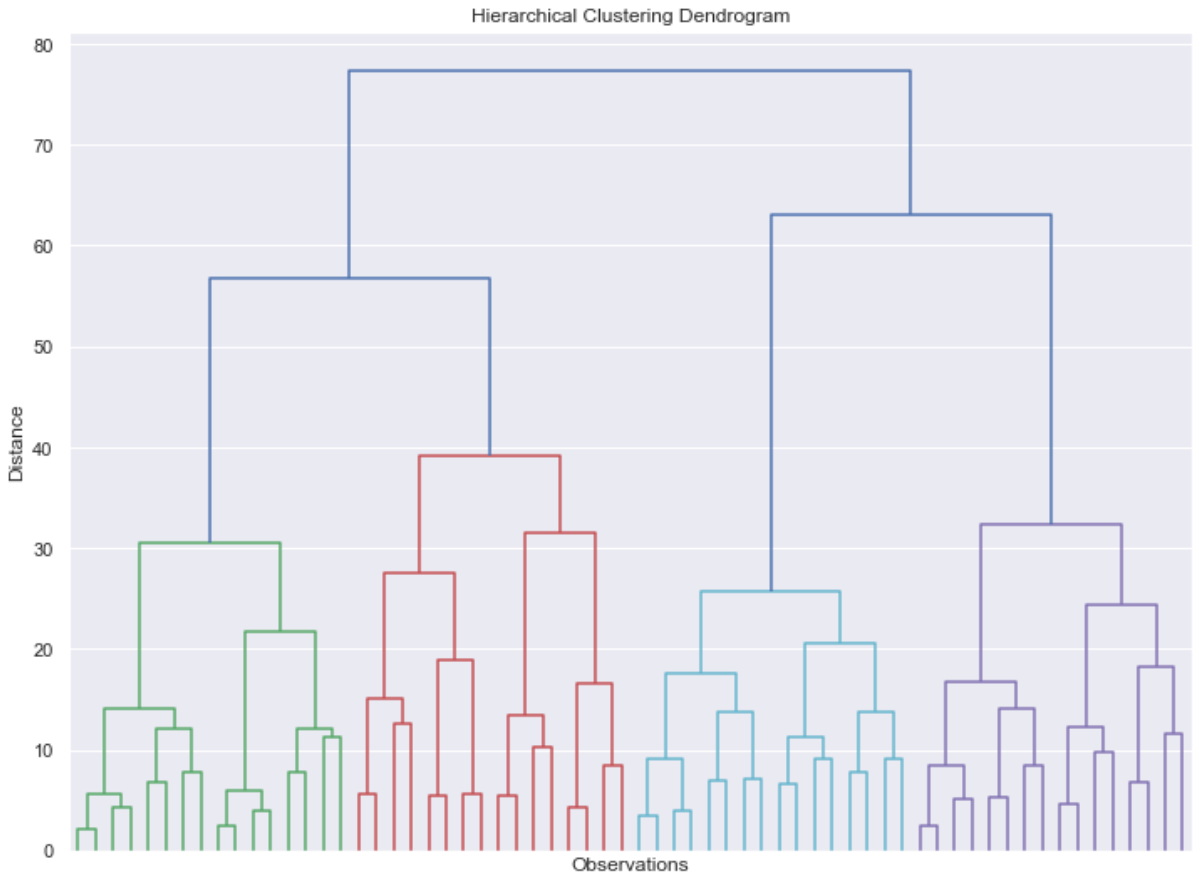
Hierarchical clustering involves organizing observations into a hierarchy of clusters, akin to the taxonomy of the animal kingdom, where subgroups are successively divided into more specific categories. Divisive clustering, or top-down, begins with all observations in a single cluster and progressively splits them into smaller clusters. Agglomerative clustering, or bottom-up, starts with individual observations as separate clusters and merges them into larger clusters iteratively. While both approaches should yield similar results, agglomerative clustering is preferred for its computational efficiency.

The distance between observations, crucial for determining cluster similarity, can be measured using various metrics such as Euclidean distance, squared Euclidean distance, Manhattan distance, or maximum distance. Additionally, distances between clusters as a whole must be considered, with the Ward method being one of the most commonly used approaches. The Ward method calculates the average of the squares of the distances between clusters, facilitating effective segmentation in practical implementations.

The dendrogram, a tree-like representation of hierarchical clustering results, provides a visual assessment of clustering solutions. It showcases observations (2000 individual points) grouped based on their distances, with vertical lines representing distances between points and horizontal lines indicating merges between clusters. To enhance clarity, dendrograms can be truncated to display only a specified number of merged cluster levels.

Determining clusters involves identifying a horizontal line on the dendrogram to cut, with the longest uninterrupted vertical line often indicating the optimal clustering solution. This line delineates clusters, with each vertical line below representing the beginning of a distinct cluster.

Moreover, the linkage method used in hierarchical clustering automatically assigns colours to clusters, aiding in visual differentiation. While hierarchical clustering may be slower with larger datasets, its advantage lies in determining the number of clusters, making it a valuable precursor to flat clustering techniques, which are typically faster and more commonly employed for segmentation. But if we have no prior knowledge about the number of clusters, we can start by performing hierarchical clustering to determine them. Then we can employ flat clustering for the segmentation itself.



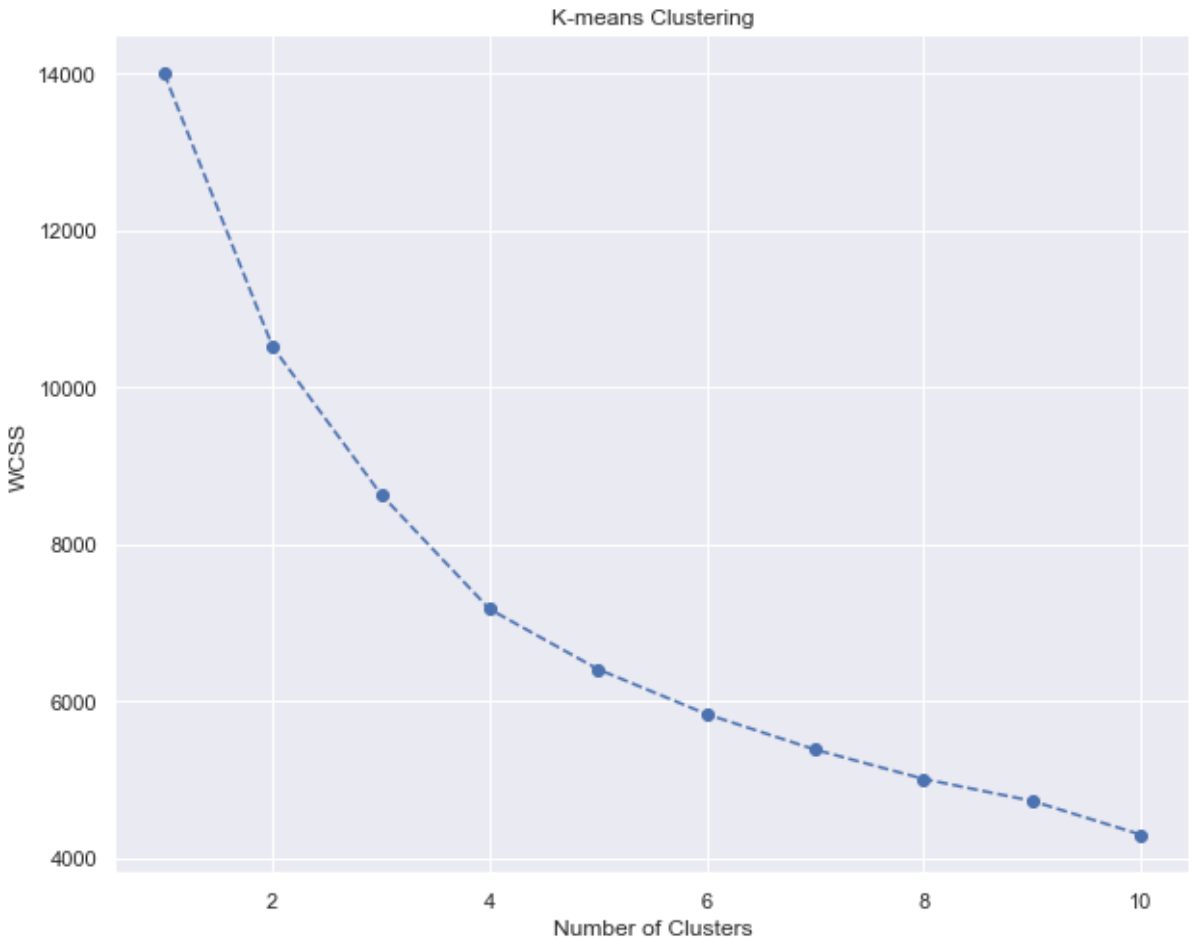
K means clustering:

K-means clustering is widely used due to its simplicity, consisting of several key steps. Firstly, the number of clusters, denoted as K, must be determined. Next, cluster seeds, or starting centroids, are chosen either randomly or based on prior knowledge. These seeds represent the initial clusters. Points in the dataset are then assigned to the nearest seed based on Euclidean squared distance. Subsequently, centroids are recalculated as the geometric centers of their respective clusters. This process iterates until a stable clustering solution is achieved.

However, K-means is susceptible to certain drawbacks. The squared Euclidean distance used in proximity calculations can be sensitive to outliers, leading to the creation of singleton clusters. This issue can be mitigated through techniques like K-median clustering, albeit at higher computational costs. Additionally, specifying the number of clusters beforehand can result in suboptimal solutions if chosen improperly. Furthermore, K-means tends to enforce spherical or blob-shaped clusters, which may not accurately represent elongated or irregularly shaped clusters in the data.

Despite these limitations, K-means often produces satisfactory results for segmentation tasks, particularly when dealing with population segments that exhibit spherical or similar shapes. Hence, it remains a valuable tool in clustering analysis for customer segmentation and other applications.

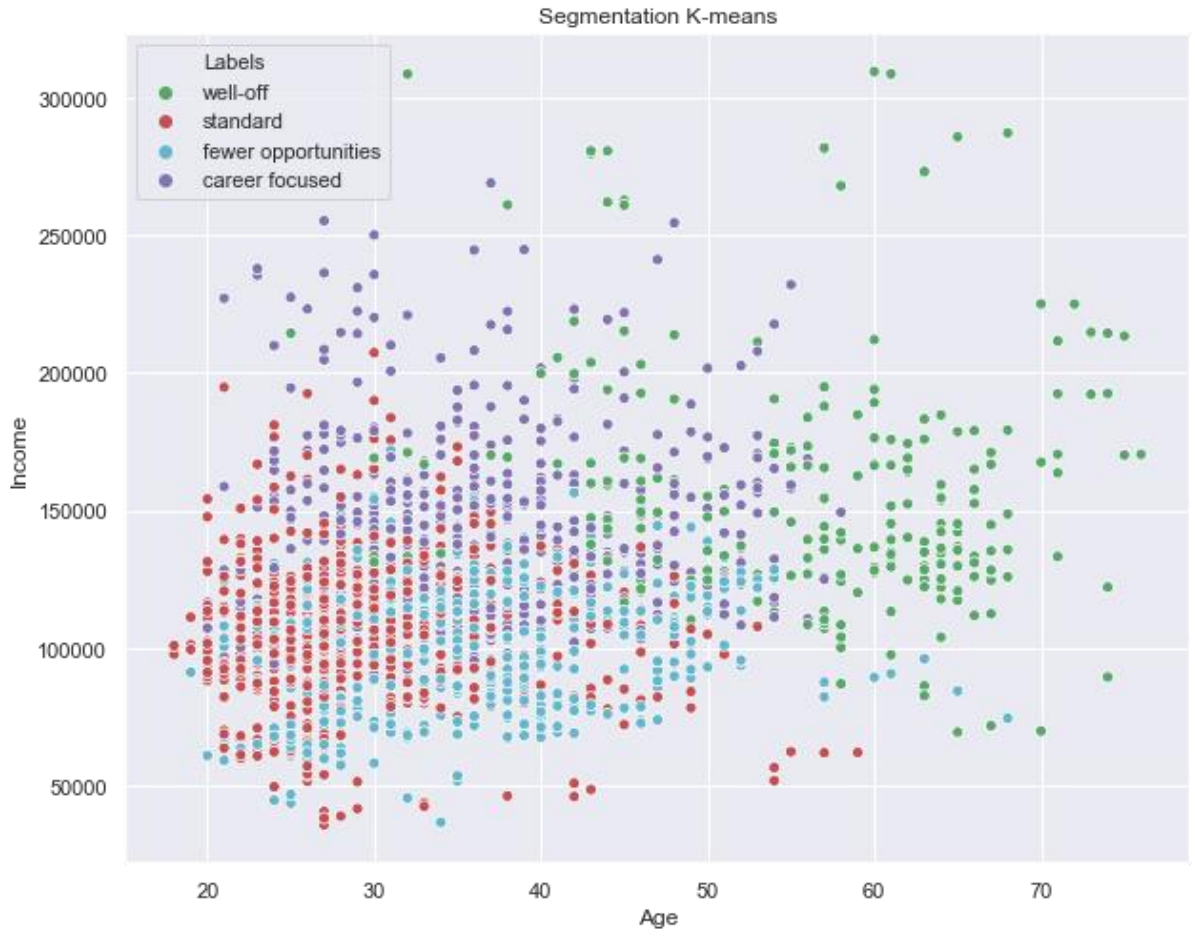
To determine the optimal number of clusters, we employed the within-cluster sum of squares (WCSS) metric. We initialized an empty list to store the WCSS values and then ran a loop to iteratively calculate the WCSS for different numbers of clusters, ranging from 1 to 10. For each iteration, we applied the K-means clustering algorithm, specifying the number of clusters and utilizing the K-means++ algorithm to initialize centroids. We then fitted the K-means model to the standardized data and computed the WCSS value, storing it in the inertia attribute. Subsequently, we plotted the WCSS values against the number of clusters to visualize the elbow method. This method involves identifying the "elbow" point on the graph, where the rate of decrease in WCSS diminishes. We observed that the graph exhibited a clear elbow at the four-cluster mark, indicating that four clusters would be appropriate for our segmentation task. Finally, we performed K-means clustering with four clusters, initialized with K-means++, and fitted the data to obtain the clustering results. In the subsequent section, we will analyse and interpret the outcomes of the algorithm.



We successfully segmented our data using K-means clustering and analyzed the results to gain insights into the characteristics of each cluster. We began by adding a new column named "Segment K-means" to our dataframe, containing the predicted

clusters for each observation. Then, we calculated the mean values of each feature for each cluster using the groupby method. By interpreting these mean values, we assigned descriptive labels to each segment, such as "well-off," "fewer opportunities," "standard," and "career-focused." Additionally, we determined the size and proportions of each cluster relative to the entire dataset, providing valuable information about the distribution of segments. After naming the segments, we visualized the segmented data by creating a scatter plot of income against age, with each individual point color-coded according to its cluster. The plot revealed distinct separation for the "well-off" segment, while the other segments were less clearly delineated. Despite some limitations, such as overlapping clusters, the K-means algorithm performed reasonably well in segmenting the data. In the subsequent section, we aimed to improve the clustering results by combining K-means with principal component analysis (PCA).

Segment K-means	Sex	Marital status	Age	Education	Income	Occupation	Settlement size	N Obs	Prop Obs
well-off	0.501901	0.692015	55.703422	2.129278	158338.422053	1.129278	1.110266	263	0.1315
fewer-opportunities	0.352814	0.019481	35.577922	0.746753	97859.852814	0.329004	0.043290	462	0.2310
standard	0.853901	0.997163	28.963121	1.068085	105759.119149	0.634043	0.422695	705	0.3525
career focused	0.029825	0.173684	35.635088	0.733333	141218.249123	1.271930	1.522807	570	0.2850



K means clustering based on PCA:

PCA provides a powerful tool for dimensionality reduction while preserving the essential information in the dataset, facilitating further analysis and interpretation. PCA provides a powerful tool for dimensionality reduction while preserving the essential information in the dataset, facilitating further analysis and interpretation. We performed PCA to reduce the dimensionality and selecting the optimal number of components to retain based on explained variance. PCA generated seven components, each representing a different proportion of the variance in the data. These components were ordered by importance, with the first few explaining a significant portion of the variability. The explained variance ratio attribute was used to quantify the variance explained by each component, totalling 100% across all components.

A critical step in PCA analysis is selecting the number of components to retain while preserving as much information as possible. A line chart depicting the cumulative explained variance against the number of components chosen was plotted to aid in this decision-making process. The graph illustrated how the cumulative explained variance increased with the number of components selected.

The decision on the number of components to retain depends on the desired level of information retention. Generally, it's suggested to retain enough components to explain 70-80% of the variance. Here choosing three components retained approximately 80% of the variance while significantly reducing the dimensionality of the data.

Further insights into these components were gained using the **components** attribute of PCA, resulting in a three by seven array indicating loadings, which represent correlations between original variables and components.

Component one exhibited positive correlations with age, income, occupation, and settlement size, indicating a focus on career-related aspects. Conversely, component two was primarily correlated with sex, marital status, and education, suggesting associations with lifestyle and education. Component three highlighted correlations with age, marital status, and occupation, emphasizing aspects of experience across various domains.

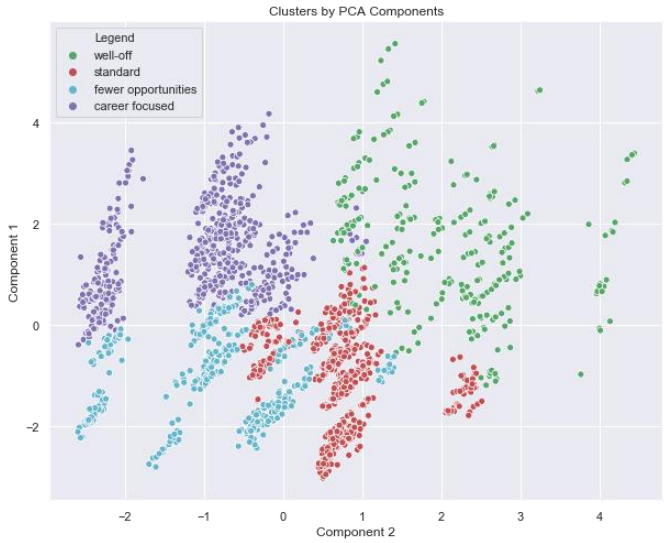
To transform the original seven-dimensional data into the three-dimensional space represented by the principal components, the **transform** method of the PCA class was employed. This generated PCA scores, encapsulating each observation's representation in the three principal components. These scores were saved for subsequent analysis, including K-means clustering based on the PCA scores.

Next our data segmentation was accomplished using K-means clustering with principal components as features. Visualizing the within-cluster sum of squares against the number of clusters revealed a prominent kink at the four-cluster mark, consistent with previous analyses, thus prompting the selection of four clusters.A K-means PCA model with four clusters was then instantiated and fitted with the

principal component scores using the same initializer and random state as in previous iterations. This resulted in the successful generation of a K-means clustering solution with four clusters. The segmentation of data using K-means clustering with principal components as features was performed.

Interpreting the clusters based on the three principal components—career, education/lifestyle, and experience—identified segments analogous to those from previous analyses: well-off, fewer opportunities, career-focused, and standard. The distribution of individuals across clusters was analysed, with the standard segment being the largest and the well-off segment the smallest. Additionally, visualization of the clusters on a 2D plane was performed, leveraging the first two principal components to clearly distinguish between the clusters and observe their separation, highlighting the efficacy of PCA in reducing dimensionality while preserving meaningful information.

Segment K-means PCA	Sex	Marital status	Age	Education	Income	Occupation	Settlement size	Component 1	Component 2	Component 3	N Obs	Prop Obs
standard	0.900289	0.965318	28.878613	1.060694	107551.500000	0.677746	0.440751	-1.107019	0.703776	-0.781410	692	0.3460
career focused	0.027444	0.168096	35.737564	0.734134	141525.826758	1.267581	1.480274	1.372663	-1.046172	-0.248046	583	0.2915
fewer opportunities	0.306522	0.095652	35.313043	0.760870	93692.567391	0.252174	0.039130	-1.046406	-0.902963	1.003644	460	0.2300
well-off	0.505660	0.690566	55.679245	2.128302	158019.101887	1.120755	1.101887	1.687328	2.031200	0.844039	265	0.1325



Purchase Analytics:

So already completed the Segmentation of STP framework. Next we will address positioning by leveraging machine learning algorithms, specifically focusing on linear regression and logistic regression. These approaches were selected for their interpretability and ease of implementation, allowing us to provide insights into three

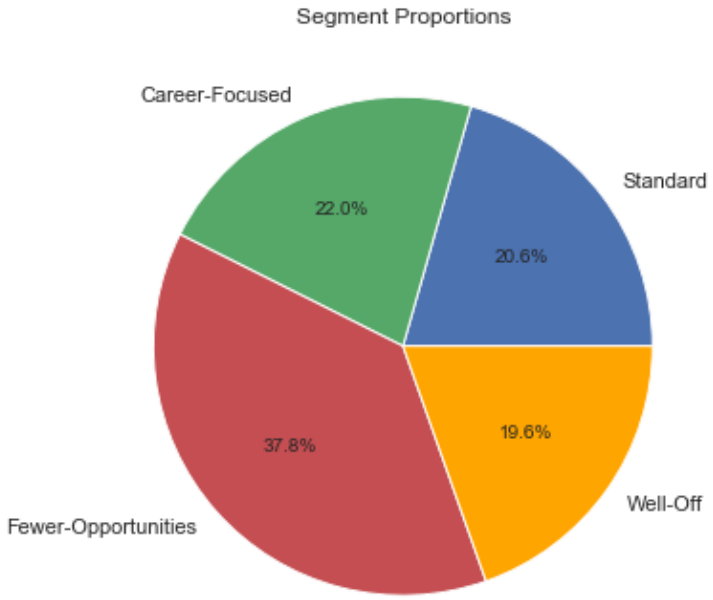
key questions: whether a customer will buy a product from a specific category, which brand they are likely to choose, and how many units they are expected to purchase.

The initial data exploration involved visually examining the dataset, comparing it to the segmentation dataset, and utilizing the segmentation model to segment new customers. Unlike the segmentation data, which contained information about individual customers, the purchase dataset consisted of transactional data from a grocery store, where each observation represented a transaction rather than a customer, resulting in multiple observations for the same individual. The dataset focused on purchases of chocolate candy bars, providing information on purchase incidents, chosen brands, purchase quantities, prices, promotions, and geodemographic characteristics, serving as a foundation for answering key questions related to purchase behavior.

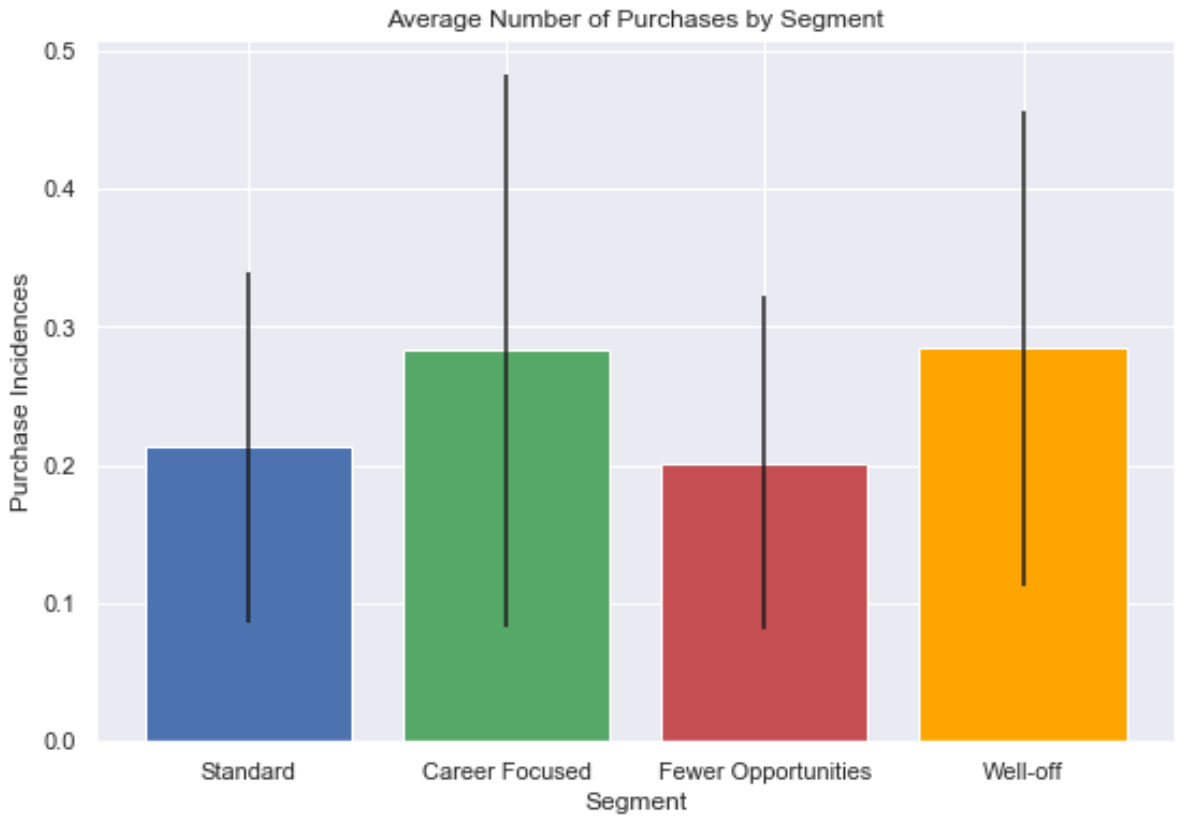
Descriptive statistics are not applicable to the dataset due to variations in the number of records per customer and per day, rendering them neither useful nor appropriate. Instead, the presence of missing values was checked using the **isnull()** command, revealing that there are no missing values in the carefully preprocessed dataset. With no missing data, the next step involves applying the segmentation model to group transactions by segments.

To assign new customers into segments, the data preprocessing steps from the segmentation model were replicated, including standardizing the variables and transforming them into principal components using PCA. The K-means clustering model previously trained on the segmentation data was then used to predict the segments for the new customers based on their PCA scores, resulting in the placement of new customers into the original four segments identified in the segmentation part.

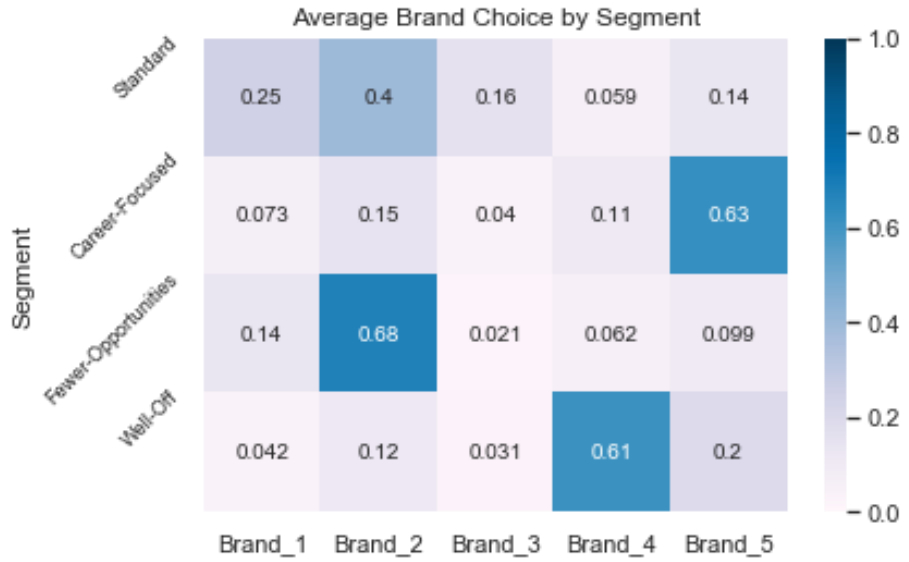
The purchase analytics is divided into 2 parts-descriptive analysis and modelling. First we will focus on descriptive analysis of the purchase data. First, we grouped the data by individuals and then by segments to gain insight into customers' shopping habits, including frequency of visits, amount spent, and products purchased. The dataset contains rows reflecting each purchase occasion, with varying numbers of rows per individual corresponding to store visits. We created a new dataframe organized by individual to analyze purchasing behavior on an individual level, including the number of purchase occasions and purchases per individual. Descriptive statistics such as average purchase frequency per individual and segment proportions were calculated to provide insight into shopping behavior and segment distribution. Finally, a pie chart was plotted to visualize the distribution of store visitors across segments, revealing that the "fewer opportunities" segment is the largest, followed by "career focused," with the "well-off" and "standard" segments nearly equally distributed.



We conducted a quantitative analysis of purchase behaviors by segments, focusing on how often customers from different segments visit the store and purchase chocolate candy bars relative to their store visits. The mean and standard deviation were calculated for each segment to show the average behavior and the homogeneity within each segment. Bar charts were used to visually represent the results, with each bar representing one of the four segments. The analysis revealed that the "career focused" segment visits the store most frequently and buys products more often, while the "fewer opportunities" segment exhibits the lowest frequency of store visits but shows greater homogeneity in purchase behavior. Finally, we emphasized the importance of relative values for comparison rather than absolute values.



So we completed the descriptive analysis of the purchase occasions of our segments. We conducted an analysis focusing on brand choice, specifically on observations where customers purchased at least one chocolate candy bar. Dummy variables were created for each of the five brands, and the data was aggregated by segment and customer ID to avoid biasing the results. The resulting average brand choice by segment was visualized using a heatmap, with brand proportions ranging from 0 to 1. Key insights included the strong preference for brand two among the "Fewer Opportunities" segment, the preference for brand five among the "Career Focused" segment, and the heterogeneous brand preferences within the "Standard" segment. This analysis provided high-level insights into brand preferences but did not explore the impact on revenue, which we will be addressing next.



We have completed the calculation of revenues for each segment based on the purchase of chocolate candy bars. This involved calculating the revenue for each brand separately, aggregating the revenues by segment, and calculating the total revenue for each segment. The analysis revealed that the "Career Focused" segment contributes the highest revenue, despite not being the largest segment in terms of size. Conversely, the "Standard" segment, while similar in size to the "Well Off" segment, contributes the least revenue. Additionally, we explored the revenue table from a brand perspective, identifying potential marketing strategies for each brand based on customer segments. These insights will inform future predictive analytics, where machine learning models will be utilized to estimate price elasticities and predict revenue outcomes.

Segment	Revenue Brand 1	Revenue Brand 2	Revenue Brand 3	Revenue Brand 4	Revenue Brand 5	Total Revenue	Segment Proportions
Standard	2611.19	4768.52	3909.17	861.38	2439.75	14590.01	0.206
Career-Focused	736.09	1746.42	664.75	2363.84	19441.06	24952.16	0.220
Fewer-Opportunities	2258.90	13955.14	716.25	1629.31	2230.50	20790.10	0.378
Well-Off	699.47	1298.23	731.35	14185.57	5509.69	22424.31	0.196

Modelling purchase incidence:

Our one of the fundamental question is Will a customer purchase a product from a specific product category upon entering the shop? In this segment, we aim to address this question by employing a statistical model designed to estimate the probability of purchase for each customer during each shopping trip. Subsequently, we will analyze the price elasticity of purchase probability under varying conditions.

A purchase occasion occurs when a customer visits the store, during which they may or may not make a purchase from the product category of interest. The dataset we utilize contains observations indicating when a customer visited the shop, with the

incidence variable denoting whether a purchase was made. Incidence is binary, with '1' indicating a purchase and '0' indicating no purchase.

Among the plethora of statistical models available, we opt for logistic regression due to its simplicity, interpretability, and widespread acceptance. Logistic regression, a classification method, generates probabilities ranging from 0 to 1. These probabilities can be interpreted in two key ways:

1. **Probability Estimate**: The output represents a real number between 0 and 1. For instance, an output of 0.77 implies a 77% chance of purchase, while an output of 0.21 indicates a 21% chance of purchase.
2. **Classifier**: If the output is below 0.5, it is classified as '0' (no purchase), and if it is above 0.5, it is classified as '1' (purchase).

Transitioning from theory to practice, we embark on applying the logistic regression model in a new Jupyter notebook to maintain a clear demarcation between descriptive analytics and predictive analytics. By doing so, we ensure streamlined and efficient workflow management. We begin by loading our data and importing three pickled objects: Scaler, PCA, and K-means with PCA. Preprocessing steps mirror those conducted in the descriptive statistics notebook, aiming to obtain a consistent dataframe for predictive analysis.

In this new notebook, we consolidate the aforementioned steps into succinct code, culminating in the creation of a new dataframe named 'Dfpa' (Data Frame for Predictive Analysis). This naming convention aids in distinguishing between the descriptive and predictive analytics sections within our project.

With our preparations complete, we are primed to delve into the application of the logistic regression model, marking the onset of predictive purchase analytics.

We will be predicting the probability of purchase so our dependent variable is whether a purchase taken place or not- incidence. It is widely acknowledged that a customer's decision to purchase a product is strongly influenced by its price. It's virtually certain that in any purchase analysis model you create, price will be the most prominent feature.

Hence, for this initial model, we adopt a simplistic approach supported by existing research on this subject. We seek to determine whether a purchase has occurred based on the average price of a product. While there are currently various brands of candy bars available, our focus is solely on discerning whether a purchase will occur at all. With this in mind, let's create a variable that represents the price, irrespective of the brand. We have several options here. We could take the minimum or maximum price, both of which are good indicators of a product's general expense. Other useful metrics include average price and median price. These are the two most common ways to represent an average for this chocolate bar model. We opt for the mean, representing the average price of chocolate bars for each row in our dataframe.

To create the variable, we'll calculate the average price by dividing the sum of prices by five (since there are five brands).. We created a variable 'purchase model' as

logistic regression. We fit the model with X and Y. We specify the solver argument as SAG to optimize for large datasets.we observed the price coefficient is -2.34, indicating a price decrease increases purchase likelihood. This model quantifies the price-purchase relationship precisely. Next, we'll examine both direction and magnitude of this effect.

We estimated our logistic regression model and now we can utilize the results to calculate the price elasticity of purchase probability. Price elasticity of purchase probability is defined as the percentage change in purchase probability in response to a 1% change in the respective aggregate price for the product category. To initiate this analysis, we first consider the range of prices available in our dataset by using the **describe** method on the price column from our dataframe. We observed that the minimum price is $1.1 and the maximum price is $2.8.Expanding this range somewhat, we opt for a range between 0.5 and 3.5 to gain a better understanding of how purchase probabilities and respective elasticities change. This range allows for comprehensive analysis while retaining valuable information.

We define our price range using NumPy, ranging from 0.5 to 3.5 with a step of 0.01 to indicate an increase of $0.01. This generates 300 values in total.Storing this information in a dataframe called 'Price Range', we proceed to predict the probability of purchase for each price point, storing the result in a variable named 'purchase probability'.The purchase probability represents the probabilities for both classes (0 and 1), where 0 implies no purchase and 1 indicates purchase. We extract the probabilities of purchase by taking the second column of the resulting array.

Next, we calculated the price elasticities The price elasticity (P) is calculated as the product of the price coefficient from the model, the price itself, and one minus the purchase probability.Creating a new dataframe called 'Price Elasticities', we add the price range array as an ID column. For each price point, we compute the corresponding price elasticity. The column containing the price elasticities is labeled 'Mean Price Elasticity'.This dataframe serves as a master record of all price elasticities calculated, providing valuable insights into average customer purchase behavior.

	Price_Point	Mean_PE
0	0.50	-0.094836
1	0.51	-0.098849
2	0.52	-0.102988
3	0.53	-0.107256
4	0.54	-0.111656
5	0.55	-0.116192
6	0.56	-0.120866
7	0.57	-0.125682
8	0.58	-0.130644
9	0.59	-0.135755
10	0.60	-0.141019
11	0.61	-0.146439

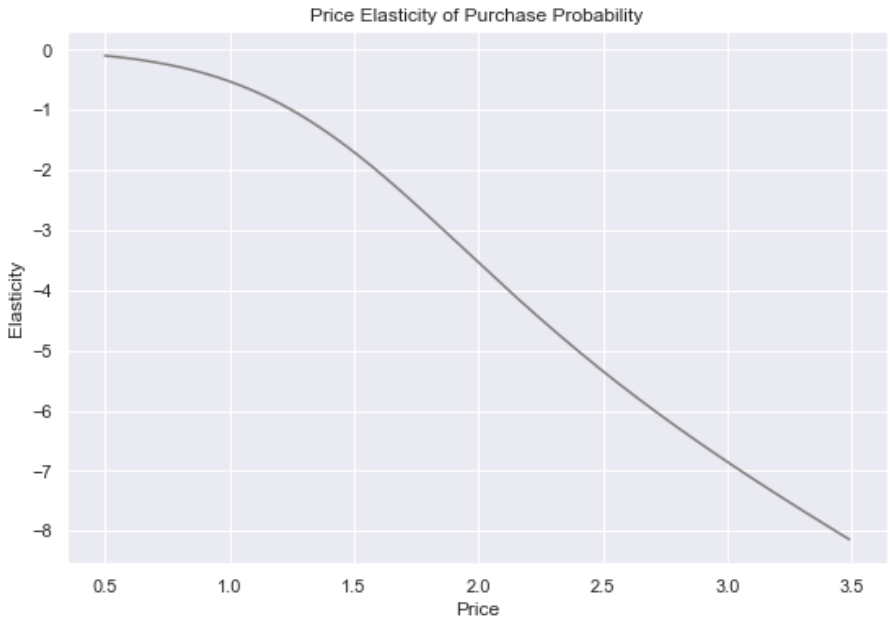
Utilizing a 9x6 figure, we plotted the price range on the x-axis and price elasticities on the y-axis. We opted for a gray color to maintain neutrality for subsequent additions to the plot. The x-axis was labeled as 'Price' and the y-axis as 'Elasticity', with the title set as 'Price Elasticity of Purchase Probability'.

The resulting graph illustrates the inverse relationship between price and purchase probability, with elasticity decreasing as price increases. Notably, all price elasticities are negative due to the inverse proportionality between price and purchase probability.

Elasticity measures the percent change in an output variable, purchase probability, given a percent change in an input variable, price. If the elasticity is smaller than one in absolute terms, it is deemed inelastic; if greater than one, it is elastic.

For instance, at a price of 1.10, the elasticity of -0.69 indicates inelasticity, as a 1% increase in price results in less than a 1% decrease in purchase probability. Conversely, at a price of 1.50, an elasticity of -1.7 suggests elasticity, as a 1% increase in price leads to almost a 2% decrease in purchase probability.

The transition from inelastic to elastic occurs around the 1.25 mark, indicating a critical price threshold. Below 1.25, price increases can be made without significant loss in purchase probability, while above 1.25, lowering prices could yield greater benefits.



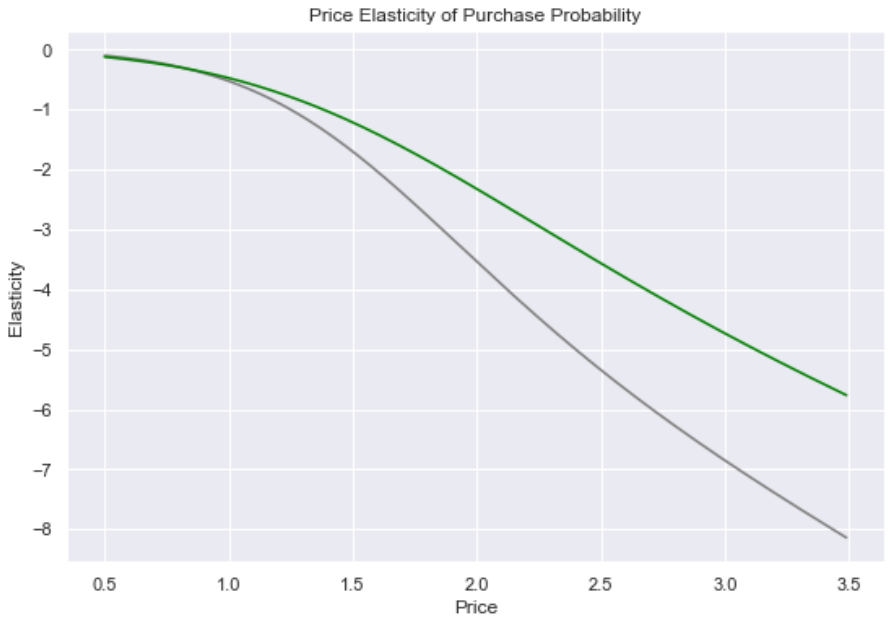
This analysis of the elasticity curve provides valuable insights into average customer purchase behavior. However, for more targeted marketing strategies, segmentation of data is necessary. Next, we will explore how to obtain purchase elasticities for different customer segments.

Purchase probability by segments:

Segment1- Carrer focused:

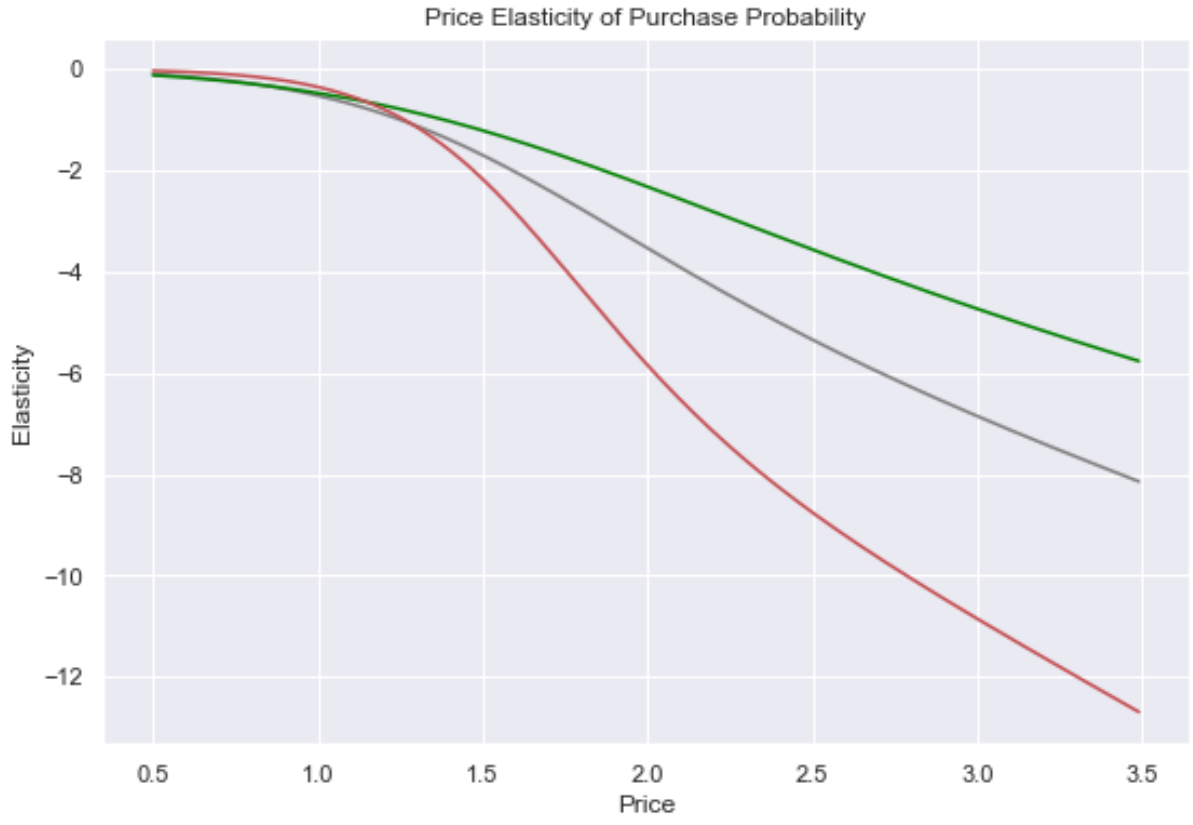
We refined our analysis by modeling purchase elasticities for specific segments of customers. Beginning with segment one, representing career-focused individuals, we created a new dataframe named 'DF Segment One' by selecting rows where the segment series equals one from the 'Purchase Analytics' dataframe. For this segment, we modeled the purchase probability using logistic regression, with the incidence column as Y and the average price across brands as X. The resulting model, named 'Model Incident Segment One', displayed a price coefficient of -1.7, indicating lower impact compared to the average customer.

Next, we calculated the price elasticities for segment one using the same methodology as before, with the results added to our 'Price Elasticities' dataframe in a new column titled 'Price Elasticity Segment One'. Visualizing these elasticities on a graph alongside the average customer's elasticities revealed that the purchase probabilities of the career-focused segment are less elastic, with the turning point to inelasticity occurring at $1.39.



Segment2: Fewer Opportunities

Moving on to segment two, representing the 'Fewer Opportunities' group, we repeated the same steps with the necessary adjustments. The resulting elasticities were added to the 'Price Elasticities' dataframe, and a red line was plotted on the graph to represent this segment., We observed that the 'Fewer Opportunities' segment is more price-sensitive and reaches its tipping point between elasticity and inelasticity at $1.07, lower than the average tipping point.



We do the same approach for other segements too. We display all elasticities of purchase probability on the same plot. We observe that the Career-focused segment are the least elastic when compared to the rest. So, their purchase probability elasticity is not as affected by price. The price elasticities for the Standard segment seem to differ across price range. This may be due to the fact that the standard segment is least homogenous, which we discovered during our descriptive analysis. It may be that the customers in this segment have different shopping habbits, which is why their customers start with being more elastic than average but then shift to being more inelastic than the average customer and indeed the Career-focused segment.

These findings provide precise insights into how purchase probabilities change with price variations for different customer segments, demonstrating the power of predictive data science in quantifying and predicting customer behavior.

Purchase probability with promotion feature:

We have completed the analysis by segments based on price, which revealed valuable insights into how different customer groups respond to price changes. However, there are additional factors beyond price that can influence a customer's purchasing decision, such as promotions. In this next phase of our analysis, we will incorporate a promotion feature to examine its effect on purchase probability elasticity.

For this model, we will utilize the entire dataset to assess the importance of promotions for all customers. Alternatively, one can refine the model to reflect specific customer segments, as demonstrated in our previous analyses.

As before, our target variable (Y) will be the incidence column from the purchase analytics dataframe, representing whether a purchase was made. In our feature set (X), we will include both the price and promotion features. For the promotion feature, we will calculate the mean promotion, reflecting the average promotional activities across all brands.

Once our feature set is prepared, we will estimate a logistic regression model using Y and X. Upon fitting the model, we will examine the resulting coefficients. For the price feature, we expect a negative coefficient, indicating an inverse relationship between price and purchase probability. Conversely, for the promotion feature, we anticipate a positive coefficient, suggesting that an increase in promotion leads to a higher purchase probability.

This model will quantify the precise relationship between price, promotion, and the probability of purchase, providing us with valuable insights for calculating purchase probability elasticity in the subsequent section.

	Mean_Price	Mean_Promotion
0	2.044	0.2
1	2.028	0.0
2	2.028	0.0
3	2.028	0.0
4	2.030	0.0

We will proceed with the price and promotion elasticity model using logistic regression. Firstly, we'll store the price range in a new dataframe named "price elasticity promotion", with the first column labeled "price range". To predict our model, we need to define the values for the promotion feature as well. Given the current aggregate nature of our model, where mean price and mean promotion are used without segmentation, we'll explore two cases: one where there are promotional activities for all brands at all price points, and another where there are no promotions at all. First, we'll focus on the first case..

We'll add a new column named "promotion" to our dataframe and set it equal to one, indicating the presence of promotions. We'll then proceed to predict the probabilities with this updated dataframe.To calculate the elasticities, we'll follow the same steps as in our previous models, reusing our code with updated names. Finally, we'll add the elasticities to our master dataframe and display the results.

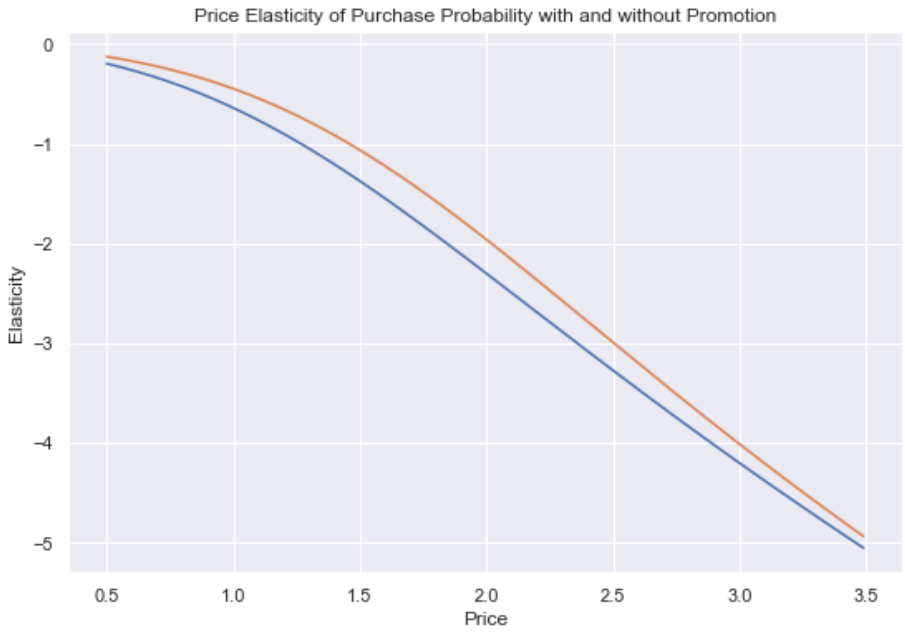
	Price_Point	Mean_PE	PE_Segment_0	PE_Segment_1	PE_Segment_2	PE_Segment_3	Elasticity_Promotion_1
0	0.50	-0.094836	-0.210803	-0.118646	-0.030918	-0.010656	-0.125711
1	0.51	-0.098849	-0.217367	-0.122815	-0.032685	-0.011265	-0.129826
2	0.52	-0.102988	-0.224040	-0.127078	-0.034539	-0.011904	-0.134020
3	0.53	-0.107256	-0.230822	-0.131435	-0.036484	-0.012574	-0.138295
4	0.54	-0.111656	-0.237713	-0.135889	-0.038523	-0.013277	-0.142650
5	0.55	-0.116192	-0.244714	-0.140440	-0.040662	-0.014014	-0.147087
6	0.56	-0.120866	-0.251826	-0.145090	-0.042904	-0.014787	-0.151607
7	0.57	-0.125682	-0.259050	-0.149840	-0.045254	-0.015597	-0.156211
8	0.58	-0.130644	-0.266386	-0.154693	-0.047717	-0.016445	-0.160899
9	0.59	-0.135755	-0.273835	-0.159649	-0.050297	-0.017335	-0.165674
10	0.60	-0.141019	-0.281399	-0.164710	-0.053000	-0.018266	-0.170535
11	0.61	-0.146439	-0.289076	-0.169878	-0.055831	-0.019242	-0.175485

Calculated the price elasticity of purchase probability with and with out promotions and plotted both on a graph. We utilized our elasticity promotion model to quantify two different scenarios: one where promotional activities were active at each price point, and the other where we assumed there were no promotions.

Now, we'll display both elasticity curves using a plot to compare and analyze them. Firstly, we'll plot the price range against the price elasticity with no promotion. Then, we'll plot the price range against the price elasticity promotion, showing both elasticities across the price range.

The X and Y labels on the graph will represent pricing elasticity, and we'll name the graph "price elasticity of purchase probability with and without promotion."

.



Observing the graph, we notice that the elasticity curve with promotion consistently sits above its counterpart without promotion across the entire price range. Consulting our master dataframe reveals that the inelasticity for no promotion ends at 1.27, whereas for promotion, it ends at 1.46, indicating a difference of almost $0.20.

This implies that even if a product has a regular price of $1.30, the purchase probability remains elastic when there's a promotional price reduction to $1.30 from the regular price of $1.50. This insight underscores the significance of promotional activities in influencing customer purchase behavior. Customers tend to be less price sensitive to similar price changes when promotional activities are present. Consequently, offering discounts, whether through large discount signs or psychological perceptions of bargain, can significantly impact customer purchasing decisions. Incorporating these findings into our model suggests that it's more beneficial to maintain a higher original price with constant promotions rather than having a lower original price.These models provide insights into purchase probability, but they can be further refined to include additional features or model more finely-grained data. The fundamental framework remains the same, setting the stage for exploring other intriguing dependencies in subsequent sections.

Up next is the model for the second significant purchase behavior outcome: brand choice

Modelling Brand choice:

Next, we are going to answer the second question about purchase behavior.

Which brand is the customer going to choose?

We're going to build a statistical model that estimates the brand choice probability for each brand. Then we will calculate price elasticity of brand choice probability for the average customer and then for each segment. Now, we will proceed with the brand

choice model, utilizing multinomial logistic regression to analyze customer behavior and predict the probability of selecting a particular brand. This approach is essential for marketers to enhance sales and increase customer satisfaction, benefiting both the company and the customer. We will leverage sklearn's capabilities to facilitate learning in a multi-class scenario, allowing us to discern brand preferences effectively.

In the process of constructing our regression model to ascertain customer brand choice, we've filtered our data to include only instances of purchases, storing this information in a new DataFrame named "brand choice." With a focus on predicting the brand, we've designated Y as the "brand" column from our "Brand choice" DataFrame. For the independent features in X, we've selected price variables corresponding to each brand, ensuring a comprehensive analysis. These features are listed and incorporated into our model, labeled as "model brand choice," which employs logistic regression with the SAG solver and multinomial scheme to accommodate the multi-class scenario. Following the model fitting process, our next step will involve interpreting the results.

We have extracted the coefficients from the model and organized them into a DataFrame labeled "brand choice coefficients," enhancing interpretability. By transposing the coefficients array, we established a conventional representation with features as rows and output variables as columns. Each coefficient is appropriately labeled according to the brand it represents, facilitating analysis. For instance, when examining brand one, we observe a negative coefficient for its own price, indicating a decrease in the probability of purchase as the price of the own brand increases. Conversely, positive coefficients for competitor brands suggest a higher probability of customers opting for our own brand as competitor prices rise. These insights highlight the interrelated nature of choice probabilities among brands, emphasizing the importance of understanding both own brand effects and cross-brand effects in marketing strategies. Moving forward, we are poised to calculate the own price elasticities for a chosen brand, further deepening our understanding of market dynamics.

	Coef_Brand_1	Coef_Brand_2	Coef_Brand_3	Coef_Brand_4	Coef_Brand_5
Price_1	-3.92	1.27	1.62	0.57	0.44
Price_2	0.66	-1.88	0.56	0.40	0.26
Price_3	2.42	-0.21	0.50	-1.40	-1.31
Price_4	0.70	-0.21	1.04	-1.25	-0.29
Price_5	-0.20	0.59	0.45	0.25	-1.09

Performing brand choice analysis is extremely useful from a brand perspective. Therefore, we'll concentrate on a specific brand to gain insight into developing a strategy to target customers. we focused on analyzing the own purchase probability elasticity for Brand five, the most expensive brand in our dataset. We constructed a

DataFrame named "on Brand five" to predict purchase probabilities based on changes in Brand five's price while keeping competitor prices fixed. Utilizing the multinomial logistic regression model previously built, we calculated the predicted probabilities for choosing Brand five and stored them in a variable called "own probability Brand five." Next, we derived the own purchase probability elasticity for Brand five using the elasticity formula, considering the appropriate coefficient from the brand choice coefficients table. This elasticity was then added to our master DataFrame under the column "own P Brand five." Finally, we visualized the elasticities of Brand five across different price points through a plotted graph, providing insights into how changes in Brand five's price impact its purchase probability. However, our curiosity extends beyond own brand effects, as we aim to investigate the influence of competitor pricing on Brand five's purchase probability, which we will delve into next.

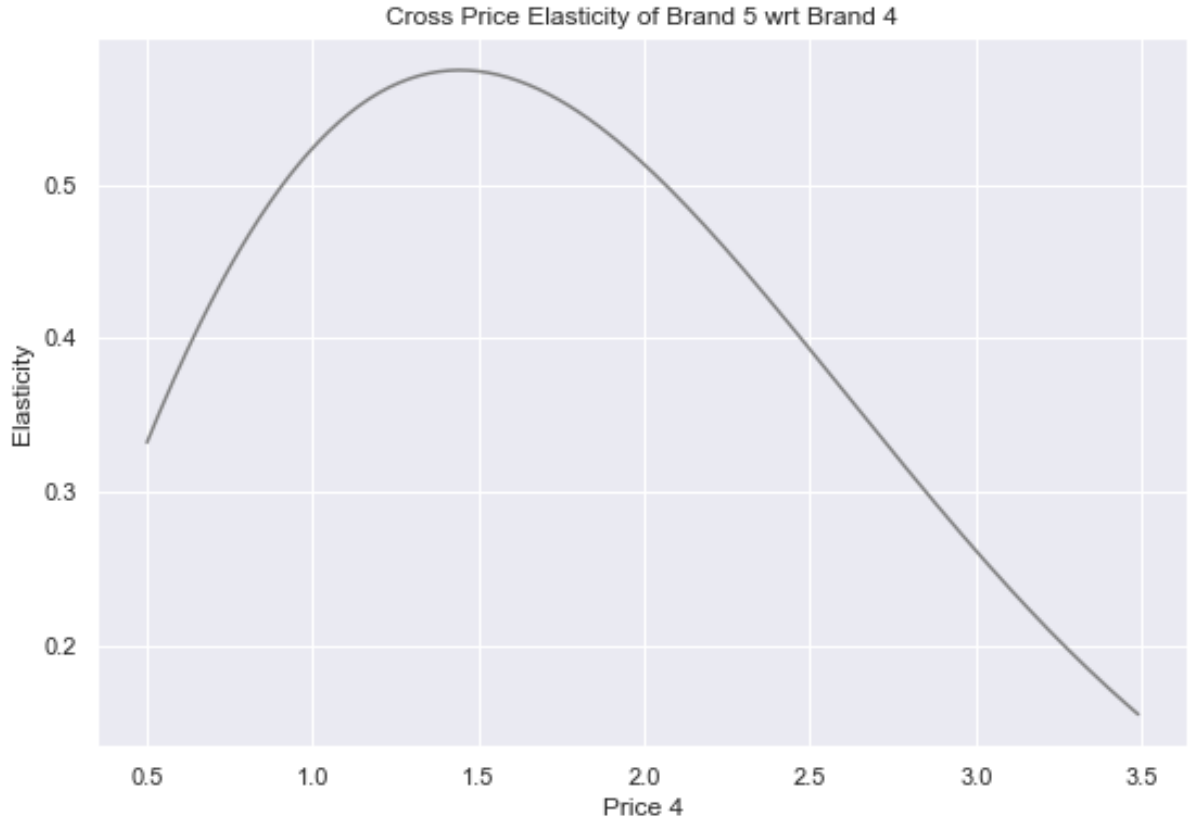
We will now examine the relationship between our own Brand five and a competitor.We could choose any of the remaining brands.However, we know that brand five is the most expensive. Price alone means nothing, but we can assume it is one of the highest quality. Based on that, it seems that the brand which comes closest is the fourth brand.Therefore, it would make the most sense to compare these two. So we will determine the cross price elasticities of brand five with respect to brand four.

we focused on computing cross-price elasticities for Brand five with respect to Brand four, a competitor brand. We began by creating a DataFrame named "Brand five Cross Brand four," containing price columns for both brands, with Brand five's price fixed at the mean value while Brand four's price varied across the price range.

Utilizing the multinomial logistic regression model, we predicted probabilities for choosing Brand four and extracted these probabilities for computation. The cross-price elasticities were then calculated using a derived formula, considering the appropriate coefficient for Brand five's price. These elasticities were incorporated into our price elasticities DataFrame under the column "Brand five Cross Brand four." Subsequently, we visualized the cross-price elasticities across Brand four's price range, revealing positive values indicative of Brand four being a substitute for Brand five.

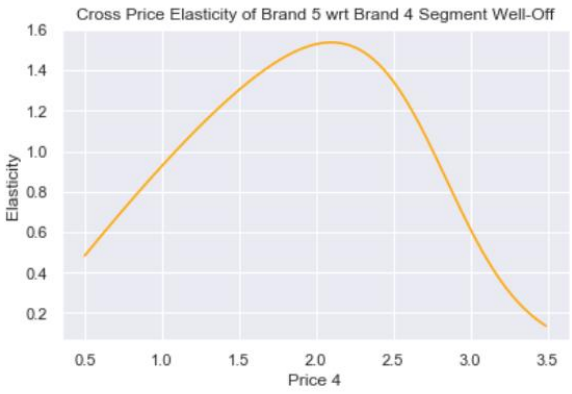
We observed that as Brand four's price increased, the purchase probability for Brand five also increased, albeit at a slower rate in the observed price range. This suggests that while Brand four serves as a weak substitute for Brand five on average, it could become a stronger competitor if priced substantially lower.

Understanding these dynamics, Brand five can devise targeted marketing strategies to attract customers who prefer Brand four. However, acknowledging the challenges of catering to the average customer, we recognize the importance of segmenting customers for more refined analysis. In the subsequent step, we will delve into segment-specific own and cross-price elasticities to further optimize marketing strategies.



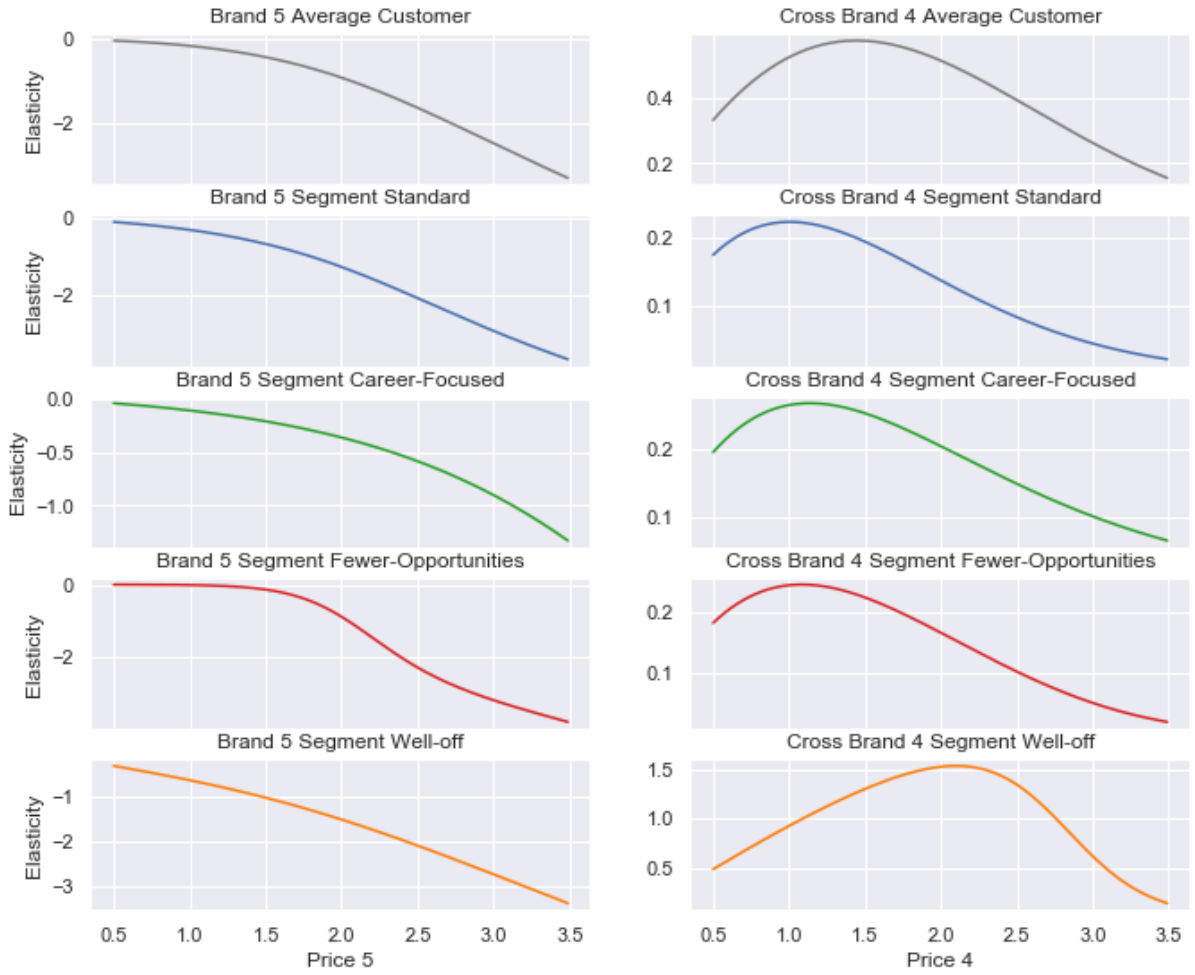
We undertook the task of modeling purchase probabilities for brand choice by segment, focusing initially on the well-off segment due to their strong preference for

Brand four. We filtered our data to isolate purchase instances for segment three, creating a DataFrame titled "Brand Choice Segment three." Using the same methodology as for the average customer, we computed own and cross price elasticities specific to the well-off segment, highlighting insights into their purchase behavior. Subsequently, we plotted the own and cross price elasticities side by side using subplots, providing a visual representation of the segment's response to price changes in both Brand five and Brand four. Analyzing the plots, we observed that the well-off segment exhibits elastic behavior towards Brand five, indicating their preference for Brand four. Furthermore, the positive cross price elasticities signify Brand four as a substitute for Brand five among the well-off segment. Leveraging these insights, we explored a strategic approach where adjustments in our own pricing could counteract competitor price changes, maintaining or even increasing market share.



We conducted a comprehensive analysis of purchase probabilities for Brand five across four segments, utilizing own and cross price elasticities derived from previous homework assignments. We plotted these elasticities side by side, enabling a comparison between the average customer and each segment. Notably, the standard segment exhibited greater elasticity compared to the average customer, suggesting a potential marketing strategy of lowering prices within certain price ranges to increase purchase probability. Conversely, the career-focused segment displayed low elasticity, indicating loyalty to Brand five and resistance to competitor brand pricing changes. The Fewer Opportunities segment demonstrated varying elasticity, being highly elastic at lower price points and less so at higher prices, suggesting a nuanced approach to marketing. However, due to limited purchase data for this segment, further insights were challenging to obtain, highlighting the importance of sufficient data for accurate modeling. Ultimately, we recognized the significance of segment-specific insights in informing targeted marketing strategies, particularly in addressing the needs of the well-off segment while retaining the loyalty of the career-focused segment. By leveraging segmentation information and understanding elasticity dynamics, brands can devise effective strategies to optimize market share and profitability. This analysis provides valuable insights into consumer

behavior and informs strategic decision-making in the competitive landscape. Our next task will be modeling purchase probability based on purchase quantity.



Modelling purchase quantity:

Next, we will address the final question concerning purchase behavior, focusing on determining the quantity or number of units of the product category of interest that customers will buy.

Our goal is to construct a statistical model capable of estimating the purchase quantity for each customer on each shopping trip. This involves analysing the data to calculate the price elasticity of purchase quantity under various conditions. At this stage of the customer journey, they have already visited the store, decided to purchase a product from the product category of interest, and selected the brand they intend to buy. The remaining decision for them is how many units they will

purchase. For example, while the decision to buy a car may result in either 0 or 1 unit purchased, when it comes to chocolate candy bars, the quantity may vary from one to several units, typically ranging from one to ten. Regardless, the outcome will be numerical.

In our dataset, there is a variable indicating the number of units a customer has purchased, referred to as the "quantity" variable. Its values are integers ranging from 0 to 15. Despite the discrete nature of our response variable, we are not dealing with a categorical scenario. Instead, we have a quantitative response that implies an ordering. Thus, we require a regression model to address this problem. We will employ linear regression for this purpose, as it offers the advantage of being easily interpretable and is widely applicable in the field of data science

After the customer has visited the store and decided to purchase a product from the relevant product category, it's crucial to note that our focus will solely be on instances where a sale is certain to happen, indicating a quantity greater than zero.To capture this outcome, we've created a new variable named "purchase quantity," which contains data from the dataframe where the "incidents" variable equals one, signifying a purchase occasion.

We've also generated dummy variables for each brand, considering that the dataset initially had a single variable for “brand”. Exploring the dependent variable, "quantity," through descriptive statistics reveals that the mean purchase quantity is 2.77, with a standard deviation of 1.8, indicating that on many occasions, more than one unit of chocolate candy bars was purchased.

Regarding independent variables, we've evaluated each column to identify potential predictors of purchase quantity. Variables such as "ID" and "day" were deemed irrelevant for predictive purposes. While the "last incidence brand" variable didn't show a clear relationship with the current purchase quantity, we've decided not to utilize the "last incremented quantity" due to its binary nature in this dataset. The most significant predictors identified were "price incidence" and "promotion incidence," representing the price of the chosen brand and whether it was on promotion, respectively. Dummy variables for brand and segments were already accounted for in these variables.

After selecting our independent variables, comprising price and promotion incidents due to data scarcity, we proceeded to create a linear regression model. Interpreting the coefficients, we found that for every dollar increase in price, there was an expected decrease of about 0.82 units in the purchase quantity. Additionally, if a promotion was present, there was an anticipated decrease of about 0.11 units in the purchase quantity. These findings suggest that higher prices deter purchases, while promotions might influence customers to buy slightly fewer units, possibly due to a simplified model or the influence of customer segments.

Moving forward, we plan to calculate the price elasticities based on these hypotheses. Next, we are going to calculate price elasticity of purchase quantity. The price elasticity of purchase quantity is the percentage change in purchase quantity in response to a 1% change in the unit price of the chosen brand.

We are preparing to calculate the price elasticity of purchase quantity across a range of prices, focusing initially on scenarios where the chosen brand is on promotion.

Firstly, we create a new dataframe named "price elasticity of quantity" and assign the price incidence column to the existing price range. To indicate the presence of a promotion, we add a new series named "Promotion Incidence" set to one for all price points. With the data assembled, we proceed to calculate the price elasticity using the formula: the regression coefficient of price multiplied by the price itself, divided by the predicted quantity.

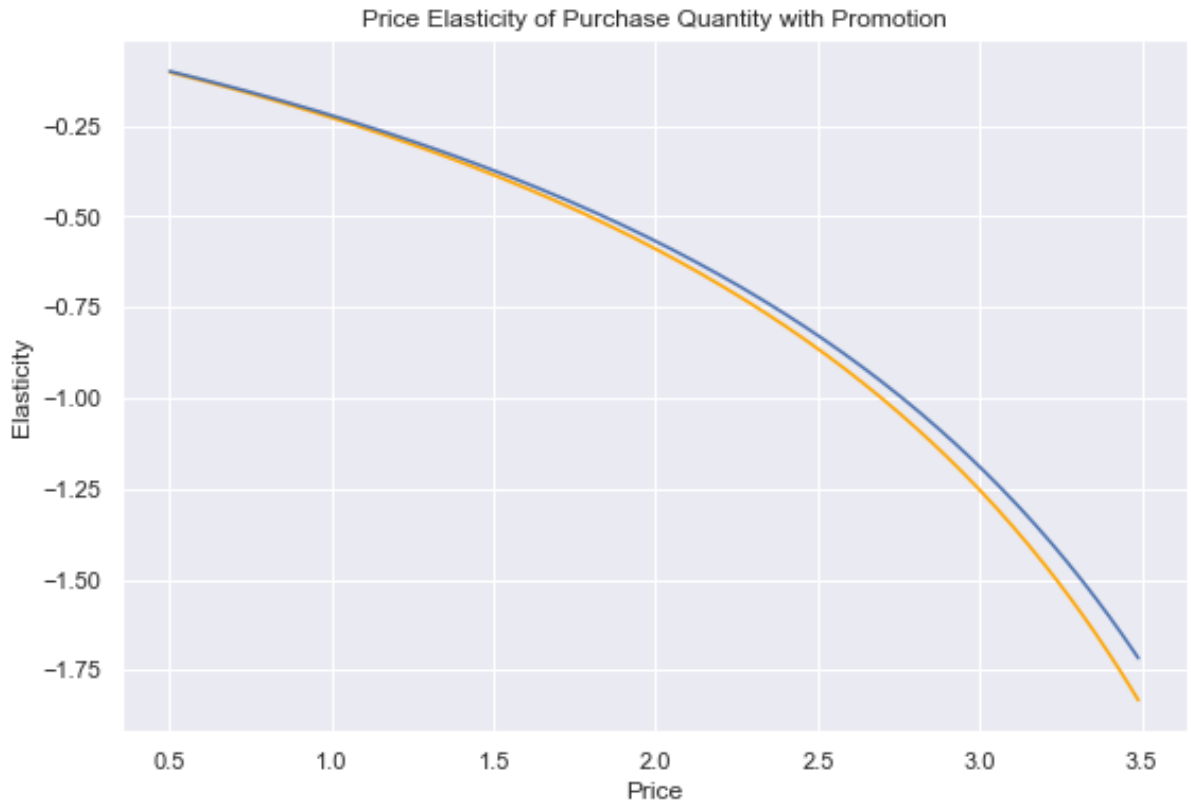
The regression coefficient is retrieved from the previously estimated model and stored as "beta quantity." We predict the quantity using the linear regression model, obtaining a series named "predict quantity." Next, we compute the price elasticities of purchased quantity with promotion and store the results in a variable named "price elasticity Quantity Promotion."

These results are then added to the price elasticity dataframe under the code name "P quantity promotion one." After calculating the elasticities, we plot the price range against the price elasticities to visualize the relationship.

We utilized our purchase incidence model to compute the price elasticity of purchase quantity under two conditions: when there is a promotion for the chosen brand and when there isn't. To visualize the results, we plotted the price range against the price elasticities, distinguishing between the two scenarios by assigning the color orange to elasticities with promotion.

From the plot, we observe that customers exhibit slightly more elasticity when a promotion is present. However, overall, customers display inelastic behaviour towards purchase quantity across prices ranging from $0.50 to approximately $2.70. Considering that our most expensive brand reaches a maximum price of $2.80, it appears that neither price nor promotion significantly influences customer decisions. This lack of distinction is evident as the two lines representing purchase quantity with and without promotion overlap at numerous price points.

Upon reflection, it becomes apparent that the variables included in our model may lack predictive value. Consequently, it may be impractical to focus extensively on purchase quantity analysis. An alternative explanation could be attributed to the imperfections in our methodology. By estimating a model based on average customer behavior, we overlook the unique preferences and behaviors of the four distinct customer segments. To address these limitations and refine our model, one potential approach is to calculate the price elasticity of demand for each brand individually. This entails filtering transactions for each brand and exploring the corresponding price elasticity of purchase quantity.


Price Elasticity of Purchase Quantity with Promotion
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
Elasticity
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Price