Customer Analysis for Consumer Goods Company

- **Venkata Sai Bala Krishna Batchu**
- **Chandratej Kurella**
- **Rajashekar Reddy Ayluri**
- **Sai Rohit Reddy Nagella**
- **Reena Namani**

Project Overview

The goal of this project is to assist a retail or FMCG (fast-moving consumer goods) organization in developing pricing and marketing plans that will increase sales of each brand of candy bar. **Finding the 'sweet spot' for price to maximize three customer behaviors—buy likelihood, brand choice probability, and purchase quantity**—will help a business realize its full revenueboosting potential.

Regression models were trained using data from consumer purchase histories to forecast these three customer behaviors within a predetermined price range. In order to evaluate how changing prices, affect each of the behaviors, the results were then transformed into price elasticities. As a result, we will be able to identify the best pricing and promotion tactics.

First, in order to assist our study of customer behavior and better position our products, we will segment our client base. This will enable us to create marketing plans that are specific to the needs of customers from various socioeconomic backgrounds.

1.Segmentation

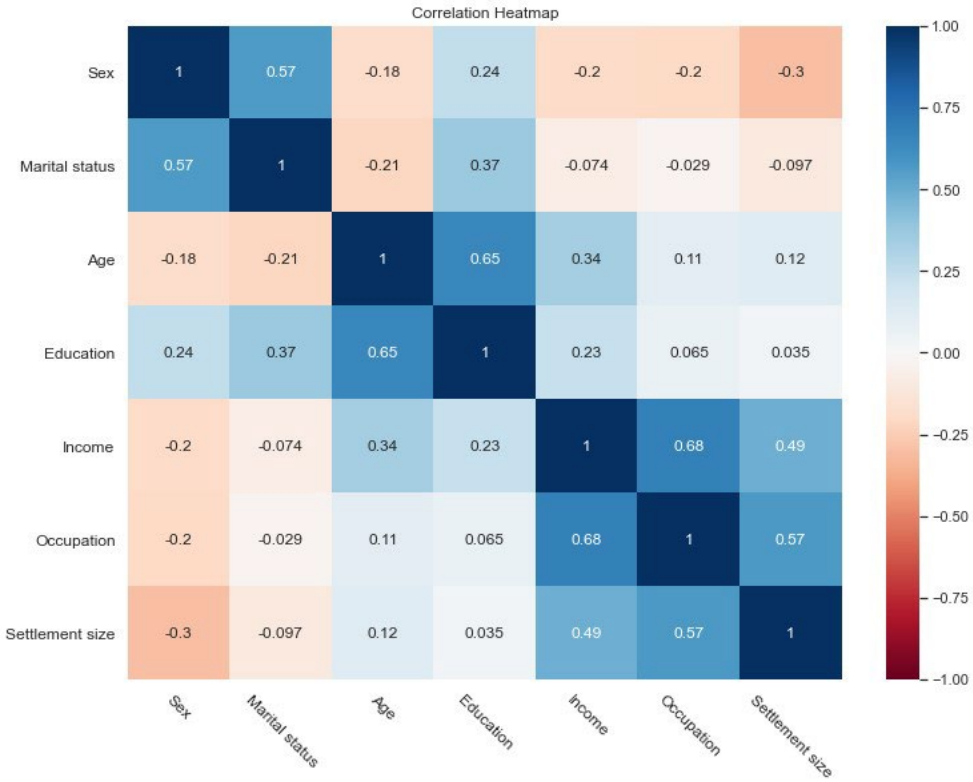
In this section, we'll segment our clientele by classifying them into several groups based on seven distinct characteristics. We will be able to examine purchase information by group and develop unique marketing plans for every one of them.

Dataset overview

	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
ID							
100000001	0	0	67	2	124670	1	2
100000002	1	1	22	1	150773	1	2
100000003	0	0	49	1	89210	0	0
100000004	0	0	45	1	171565	1	1
100000005	0	0	53	1	149031	1	1

- Sex: 0 - male, 1 - female
- Marital status: 0 - single, 1-non-single
- Education: 0 - other/unknown, 1 - high school, 2 - university, 3 - graduate school
- Occupation: 0 - unemployed, 1 - skilled, 2 - highly qualified
- Settlement size: 0 - small, 1 - midsized, 2 – big

Correlation estimate



A connection can be seen between some pairs of variables, such as **income and occupation**, **education and age**, and **settlement size and occupation**. It suggests that we can scale back how we represent our consumers without losing too much information, enabling us to classify our clientele more precisely.

1.2 Clustering

Standardization

We standardize our data first to ensure that all features are given the same weight.

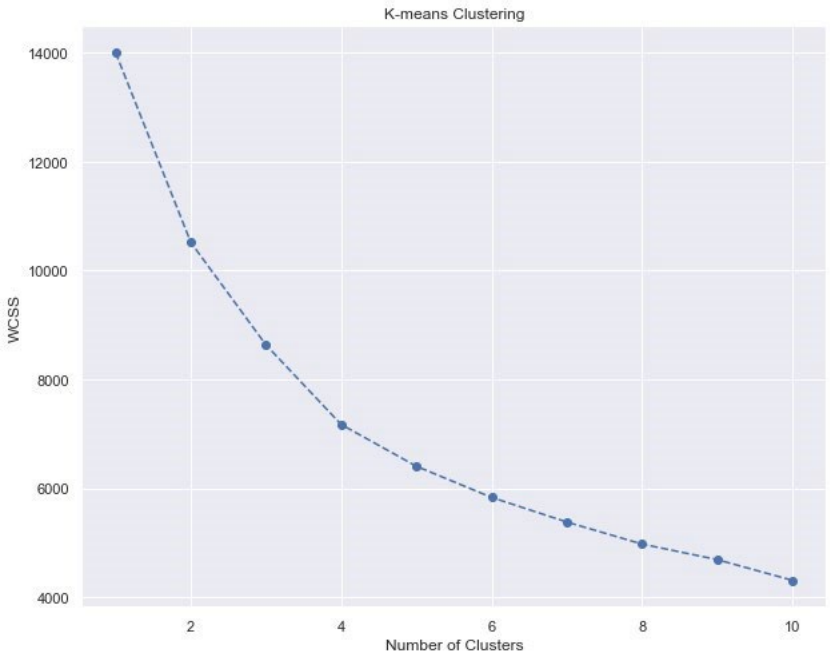
Standardizing data, so that all features have equal weight scaler = StandardScaler()

#Create an instance segmentation_std = scaler.fit_transform(df_segmentation) #Apply

the fit transformation

K-Means Clustering

First, using 1 to 10 clusters, we cluster using K-means, and we display the Within Cluster Sum of Square (WCSS).



We chose 4 clusters to segment our clients using the "Elbow method," and we obtained the following traits for each group.

We have 4 distinct customer categories.

Well-off: elderly, highly educated, and earning a large income

Fewer prospects for those who are unmarried, middle-aged, have poor incomes or low-level jobs, or have tiny households.

Young, educated, and single with an emphasis on their careers

Typical: others

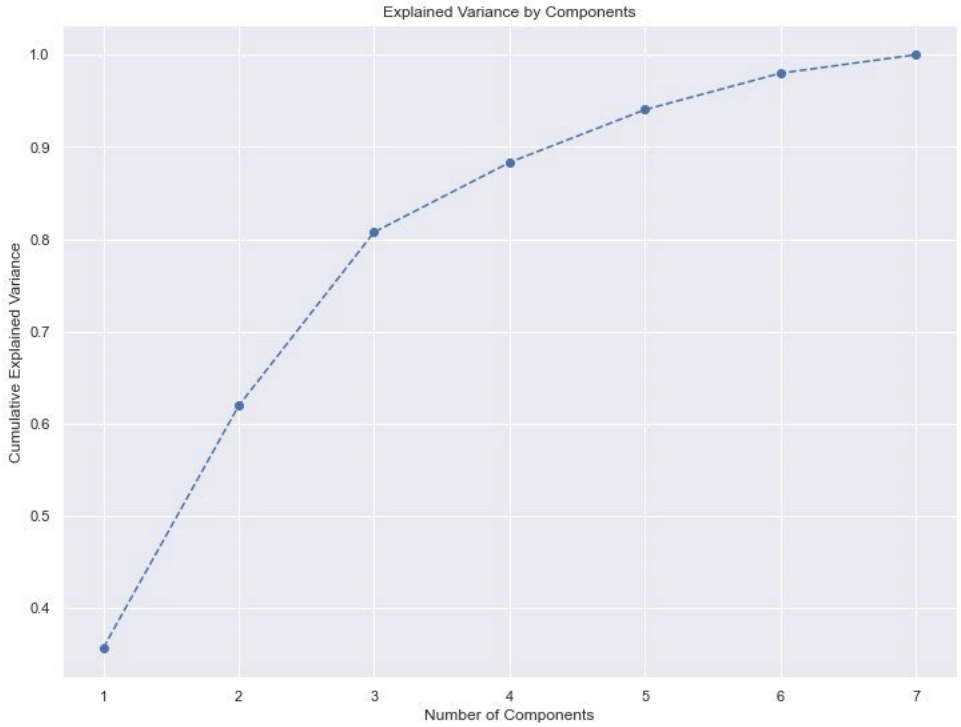
However, it's challenging to distinguish between the groups if we only use 2 dimensions to depict the segmentation.

However, it's challenging to distinguish between the groups if we only use 2 dimensions to depict the segmentation.

Therefore, we need to perform the clustering with PCA

PCA

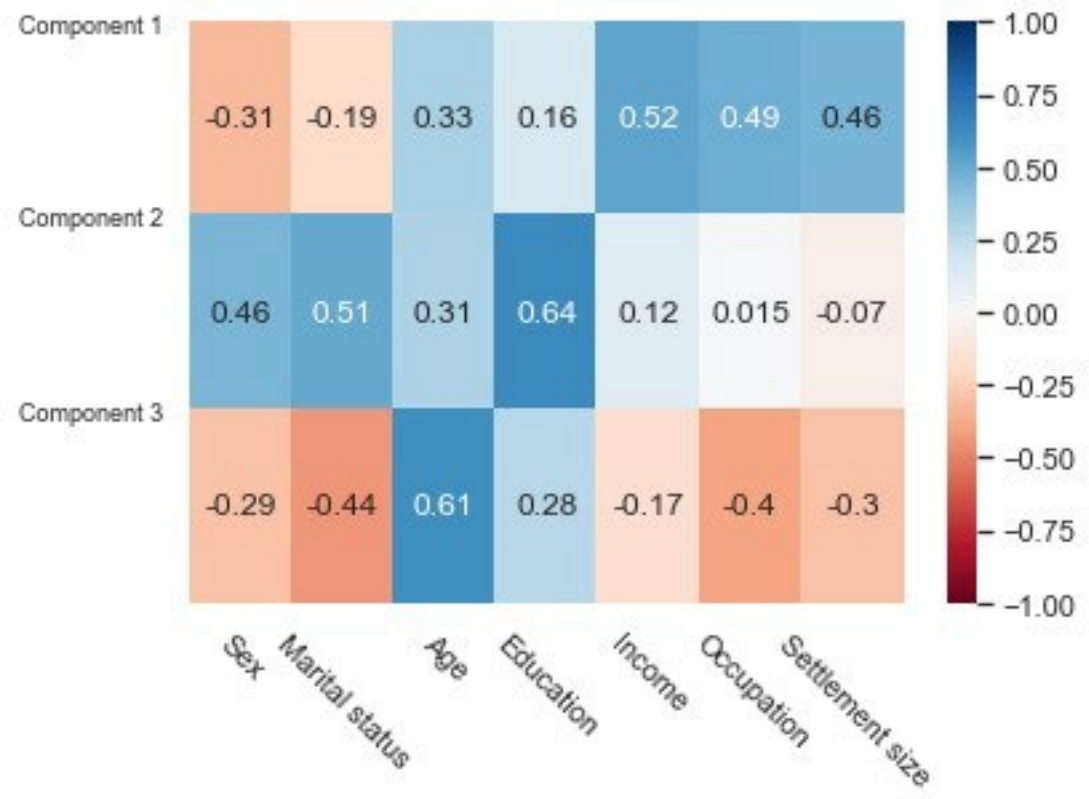
After fitting the PCA with our standardized data, we visualize the explained variance



Our data are represented by 3 components that account for more than 80% of the variance.

We obtain the loadings (i.e. correlations) of each component on each of the seven original features after fitting our data with the chosen number of components.

Visualize the loadings by heatmap



A dimension of each component's individual features is shown.

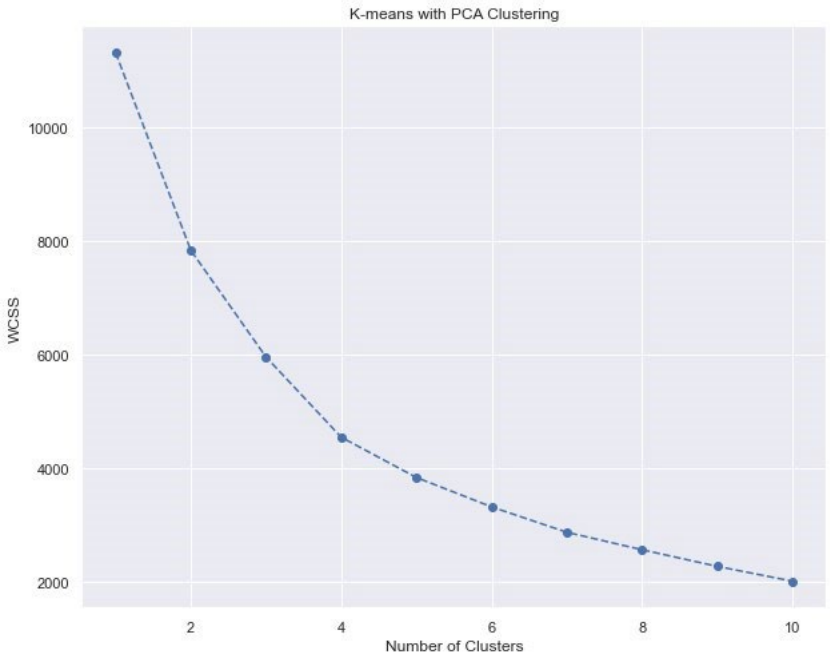
Component 1: relates income, occupation, and settlement size to the career focus.

Component 2: relates gender, marital status, and education to indicate the individual's lifestyle and education.

Component 3: relates marital status, age, and occupation to the amount of experience (work and life).

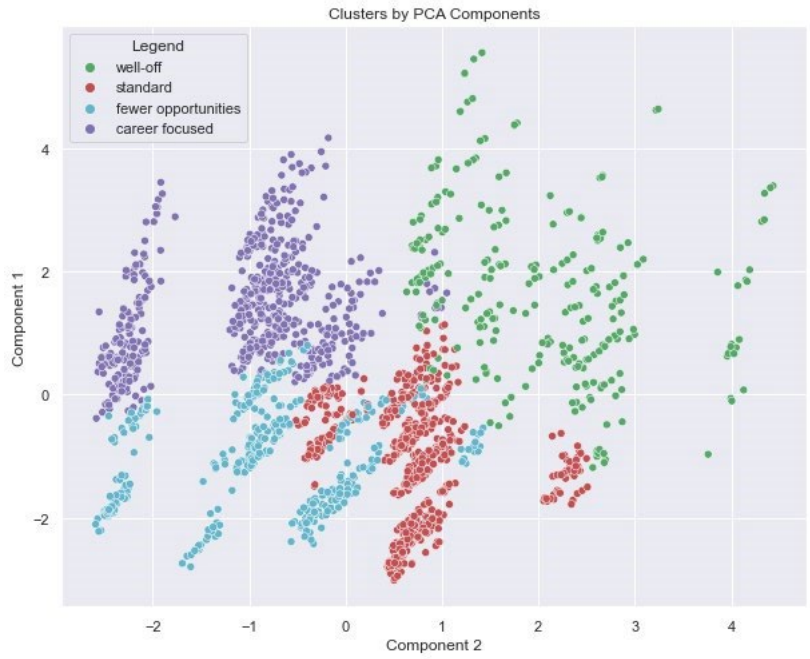
PCA and K-Means Clustering

We obtain the WCSS below by fitting K means to the converted data from the PCA.



Again, we choose 4 clusters to fit our data, and get the below results

We plot data by 2 PCA components: Y axis - component 1, X axis - component 2

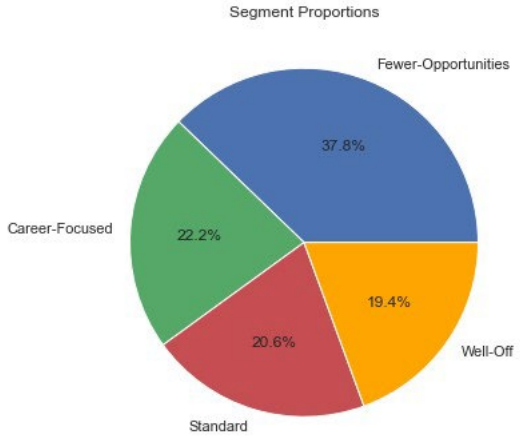


2. Purchase Descriptive Analytics

In this section, we want to learn more about our customer's past purchasing habits, like how frequently they shopped and purchased candy bars, which brand they favored most frequently, and how much they spent. The outcomes can be utilized to verify our third part's predictions.

2.1 Data segmentation

To segment our clients in the purchase dataset, we use the standardization, PCA, and Kmeans clustering models from the preceding section. What we have is as follows **We visualize the proportions of total number of purchases as per segments**

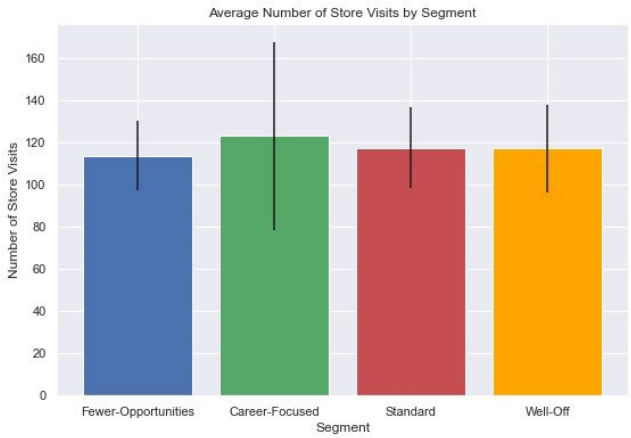


insights: In our store, **we'll typically observe candy bars being purchased in smaller groups with fewer opportunities**. There are several potential

- They represent the largest clientele groups (add more observations)
- **They go to the store more frequently than the other customers (more visits result in more purchases).**
- whenever they go shopping, people are more likely to buy candy bars. We shall look into this more below.

2.2 Purchase Incidence and Purchase Occasion

Use **a bar chart to show the average number of store visits for each of the four segments, and a straight line to show the standard deviation.**



insight:

The **standard deviation for "career orientation" is fairly high.** This means that customers in this segment are at least homogeneous. That is, **they are at least similar in terms of how often they visit the grocery store.**

Standard, low-opportunity, and wealthy clusters are very similar in terms of average in-store purchases. This is welcome information and will make comparisons easier in future analyses.

View average purchases by segment to help understand how often each group buys candy bars **Number of purchases by segment**

Insight:

The standard deviation is highest for carrier orientation. **This could mean that one part of the segment buys the product very frequently and another part less.** Consumers in this segment earn about the same, but they may spend their money differently.

The **most uniform segment seems to be the segment with the least opportunity**. This is indicated by the segment with the lowest standard deviation or the segment with the shortest vertical line. The

standard segment is also consistent, with an average purchase count of about 25 and a standard deviation of 30.

2.3 Brand Choice

First, we select only rows where the incidence is one. Then we make dummies for each of the 5 brands.

	Brand_1	Brand_2	Brand_3	Brand_4	Brand_5	Segment	ID
6	0	1	0	0	0	0	200000001
11	0	0	0	0	1	0	200000001
19	1	0	0	0	0	0	200000001
24	0	0	0	1	0	0	200000001
29	0	1	0	0	0	0	200000001
...	...	...	...	...	...	...	...
58621	0	1	0	0	0	0	200000500
58648	1	0	0	0	0	0	200000500
58674	0	1	0	0	0	0	200000500
58687	0	1	0	0	0	0	200000500
58691	0	1	0	0	0	0	200000500

Visualize the brand choice by segments (on average, how often each customer buy each brand in each segment)

Insights: Each segment has preference on 1 or 2 brands

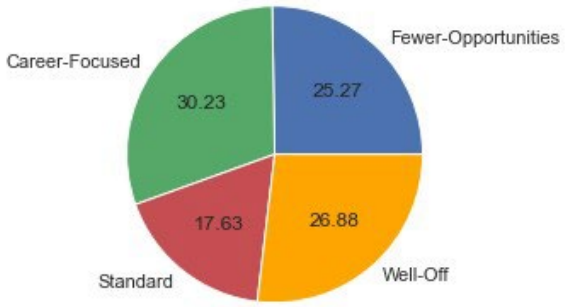
Well-off and Career-focused prefer pricy brands

Fewer-opportunities and standard prefer low price products

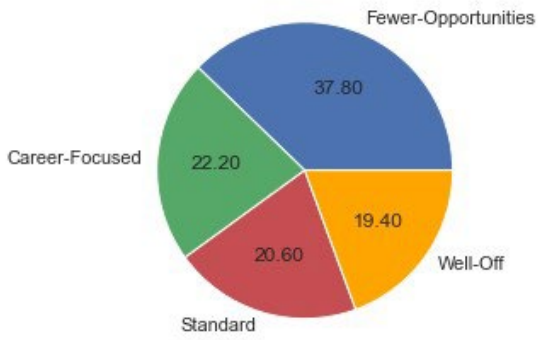
2.4 Revenue

Compute the total revenue for each of the segments.

	Revenue Brand 1	Revenue Brand 2	Revenue Brand 3	Revenue Brand 4	Revenue Brand 5	Total Revenue	Segment Proportions
Segment							
Fewer-Opportunities	2258.90	13909.78	722.06	1805.59	2214.82	20911.15	0.378
Career-Focused	736.09	1791.78	664.75	2363.84	19456.74	25013.20	0.222
Standard	2611.19	4768.52	3909.17	861.38	2439.75	14590.01	0.206
Well-Off	699.47	1298.23	725.54	14009.29	5509.69	22242.22	0.194



Percentage of Total Revenues by Segments



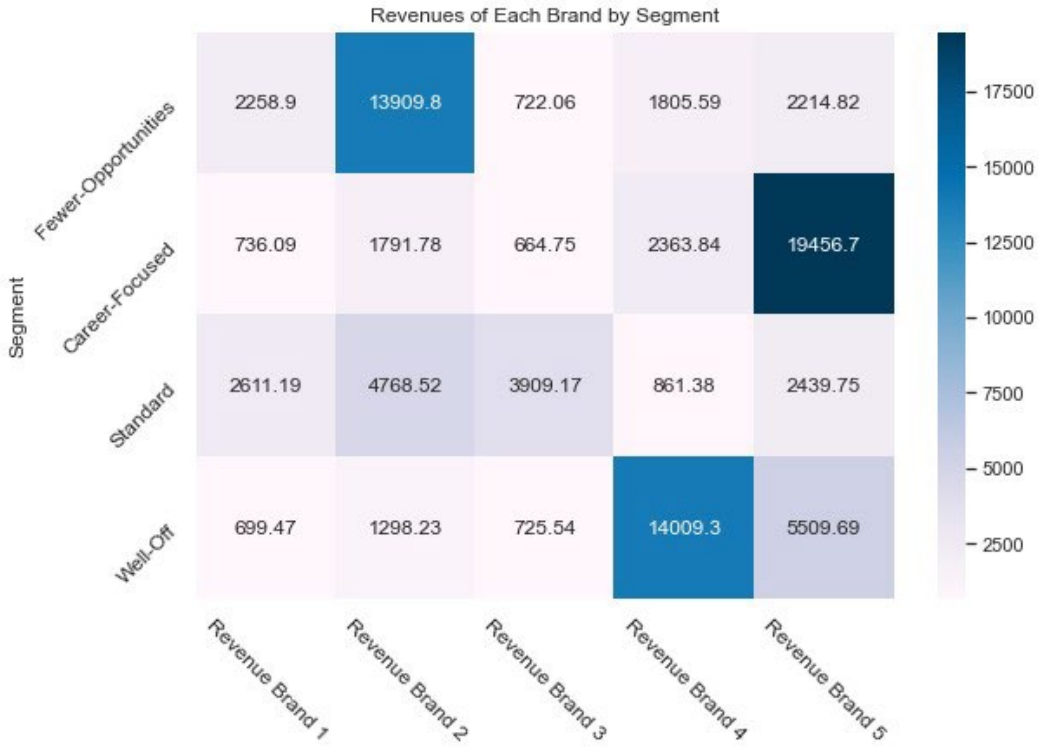
Segment Proportions

insight:

Career focus brings the highest revenue, but is nowhere near the largest standard segment by total purchases

Well-offs have the second highest turnover despite being the smallest segment

Standards are not the smallest segment, but they contribute the least as **they tend to buy lower priced products.**



Insights:

Brand 3 does not have any segment as its loyal customers. **If brand 3 reduces its price, the standard segment could pivot towards it since they seem to be struggling between brand 3 and brand 2**.

Well-off segments mostly prefer brand 4, followed by brand 5. They seem to be **not affected by price**. Therefore, **brand 4 could cautiously try to increase its price.** (hypothesis here: will retain most of the customers and increase the revenue per sale)

For career-focused, Brand 5 could increase its price.

3. Purchase Predictive Analytics

3.1 Purchase Probability

We implement the standardization, PCA, and K-means clustering models from part 1, to segment our customers in purchase dataset.

Price Elasticity of Purchase Probability

	Price_1	Price_2	Price_3	Price_4	Price_5
count	58693	58693	58693	58693	58693
mean	1.392074	1.780999	2.006789	2.159945	2.654798
std	0.091139	0.170868	0.046867	0.089825	0.098272
min	1.1	1.26	1.87	1.76	2.11
25%	1.34	1.58	1.97	2.12	2.63
50%	1.39	1.88	2.01	2.17	2.67
75%	1.47	1.89	2.06	2.24	2.7
max	1.59	1.9	2.14	2.26	2.8

Then we fit our 'test price range' in our model to get the corresponding Purchase Probability for each price point.
Next, we apply below formula to derive the price elasticity at each price point

insight:

Overall price should be lowered to increase overall purchase probability

If the price is below the 1.25, we can raise the product price without losing too much potential to buy. Above 1.25 you can get more profit by lowering the price.

This is not good news as the average price for all brands is over 1.25. Each segment needs further research.

Purchase Probability by Segments

insight:

Wealthy segments are the least resilient compared to the rest of the segments. Therefore, the **elasticity of purchase probability is not affected by price. Fewer opportunities are more price sensitive than other groups**

The price elasticity of the low-opportunity segment appears to vary by price range (low for low prices, high for high prices).

Here's why: more accurate because it has more

observations

This segment enjoys candy bars so much that higher prices on the low end won't hurt them. When it gets expensive it doesn't make financial sense for them to invest in it.

Purchase Probability with and without Promotion Feature

insight:

At the same time, by applying promotions, we can raise our prices a bit without fear that they will be less likely to buy our products.

The elasticity of the customer's purchase probability becomes less elastic due to the promotion

This is an important insight for marketers. Because, **according to our model, people are more likely to buy an item when there is a promotion than to buy it at the same price when it's not on sale**.

3.2 Brand Choice Probability

We get the following coefficients:

	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

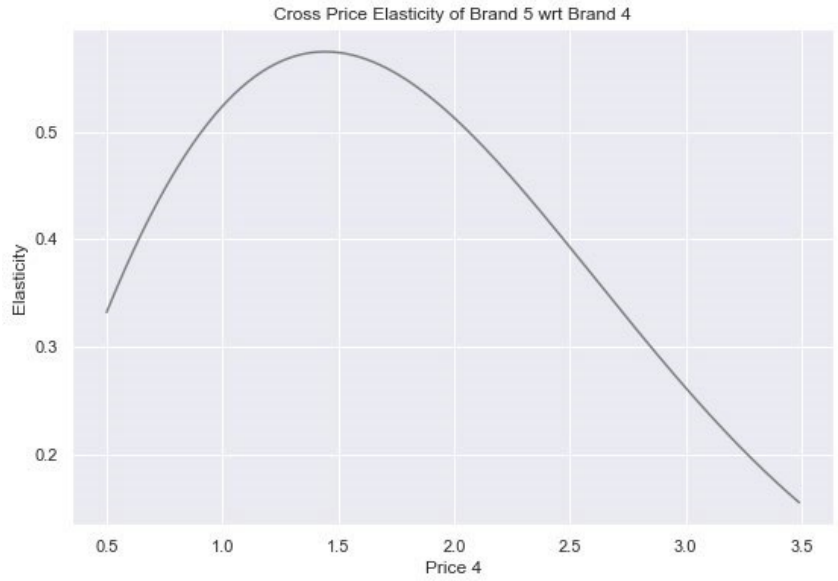
interpretation:

Each coefficient shows how price changes affect the likelihood of choosing the corresponding brand. In general**, the lower your own price and the higher the price of other brands, the more likely you are to choose a brand**

Own Price Elasticity Brand 5

Interpretation: It shows us how it would affect brand 5 if they change their own price.

We visualize the cross-price elasticity of purchase probability for brand 5 vs brand 4



Interpretation: **It shows us how it would affect brand 5 if brand 4 change their price.**

insight:

Brand 4 is a strong alternative to Brand 5 at all prices up to $1.65.

Note:

The **observed price range for Brand 4 is between $1.76 and $2.6 in the region**

These prices are outside the natural range of Brand 4, so if Brand 4 were significantly lower in price, it would be a very strong competitor to Brand 5.

The elasticity drops from the 1.45 level, but is still positive, suggesting a slowdown in the increase in purchase probability at level 5.

Brand 4 is no substitute for Brand 5 when it comes to the average customer.

Brand 5 can create marketing strategies that target customers who choose Brand 4 and encourage them to purchase Brand 5. well-off and retaining the career-focused segment, the most frequent buyers of brand 5

- For Career-focused segment, Brand 5 could increase its price, without fear of significant loss of customers from this segment

The **Career-focused segment is the most inelastic and they do not seem to be that affected by price**

The cross price elasticity also has extremely low values, meaning they are unlikely to switch to brand 4

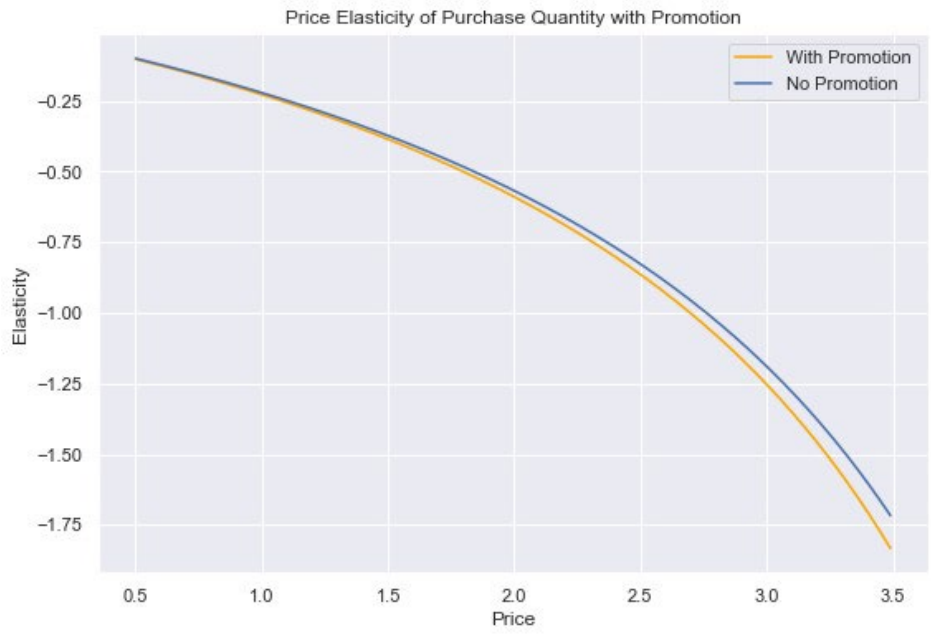
- **For the Well-off segment, we'd better decrease brand 5 price to gain market share from this segment** ○ For this segment, own elasticity is much higher than 'career-focused' ○ **Welloff also purchase the competitor brand 4 most often by having highest cross brand elasticity, meaning a tiny increase in price will lose customers**

3.3 Purchase Quantity:

To determine price elasticity of purchase quantity, also known as price elasticity of demand, we're interested in purchase occasion, where the purchased quantity is different from 0.

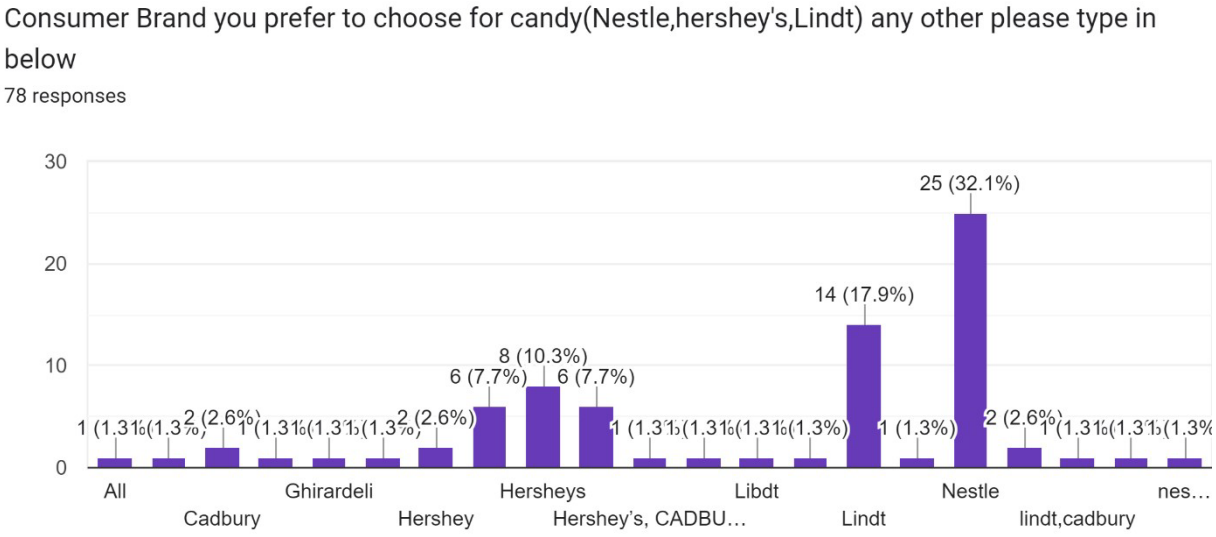
Interpretation: It appears that promotion reflects negatively on the purchase quantity of the average client, which is unexpected.

Plot the two elasticities (with and without promotion) side by side:

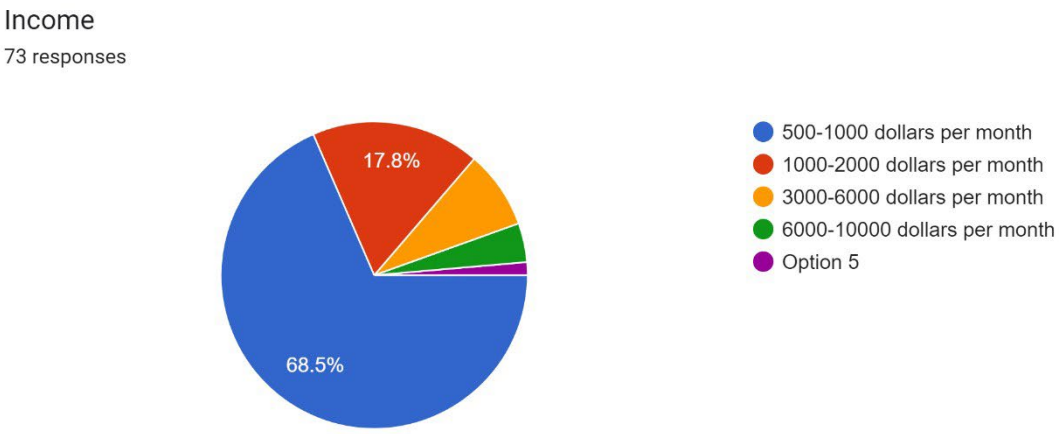


Insight: It appears that **promotion does not appear to be a significant factor in the customers' decision what quantity of chocolate candy bars to purchase.**

Our survey Data to analyze the best brand and price range



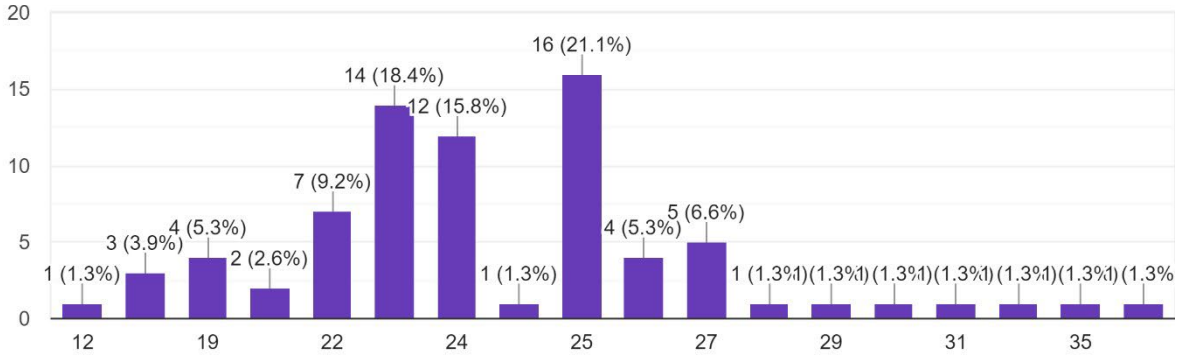
Nestle Seems to be the top candy bar brand preferred by consumers.



As the majority people are students for the survey response the income range lies around 1000 dollars per month.

Age of consumer

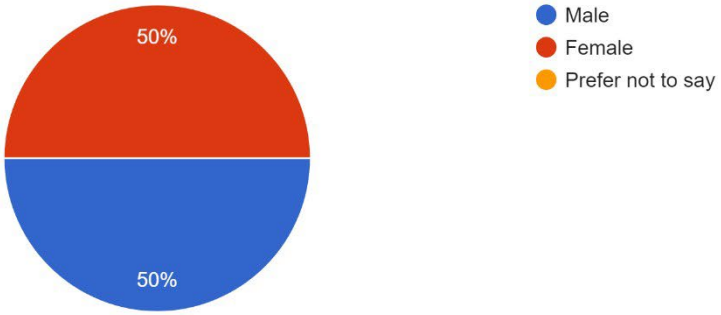
76 responses



Average age of consumer as per survey seems to be 24.

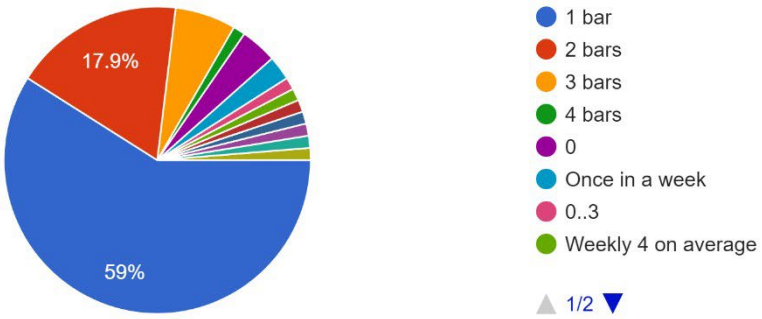
Sex

78 responses

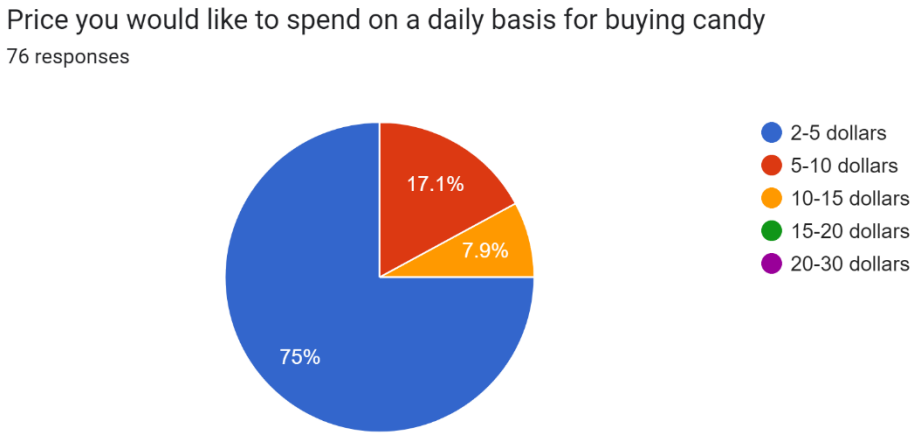


Quantity of candy consumed in a day on an average

78 responses



Quantity of candy bar consumed in a day on an average is around 1 bar to the max.



Price of candy bar maximum of the respondents is preferred under 5 dollars.


Thank You!