cct | College Dublin
Computing • IT • Business

BUSINESS REPORT

Customer Personality Analysis

Marketing Proposal

For Year 2022-23

PREPARED BY:

Leandro Marigo
David Massey
Bobbi McDermott
Oluwatoyin Fawole

CONTENTS

INTRODUCTION

We have used data collected from a period of 2012 to 2014, with Customer information attaining their purchases. We aim to use this data to predict behaviour using analytical techniques to establish a clear marketing plan to increase sales. We are following the CRISP-DM process to help keep a clear structure. There are six phases to it, each of which has additional standardized tasks. The general tasks outline what must be completed in a phase before proceeding to the following one. The explicit definition of the Business Understanding phase in CRISP-DM is one of its advantages over competing models. Following the process towards shared business objectives thanks to the Business Understanding phase and the simple description

The dataset for this project is provided by Dr. Omar Romero-Hernandez and the License is CC0: Public Domain

Customer Personality Analysis | Kaggle

Version Control and Collaboration Platform

GitHub Repository

Google Drive

SECTIONS

Section 1	Business Description
Section 2	Technologies used for Data Mining and Analysis
Section 3	What has been accomplished so far
Section 4	Obstacles and Strategies
Section 5	Results and Analysis, Next Steps
Section 6	Conclusion

SECTION 1: BUSINESS DESCRIPTION

Grocery store with a base of 2240 customers operating both physically and online

1. Hypothesis

Data collected over 2 years includes the profiling of the customer, products they bought, how much they spent, when their last visit, and what portal they used to purchase. The goal of this customer personality analysis is to choose the most efficient method for analyzing this data to identify spending patterns and improve marketing campaigns.

To increase customer engagement, our goal is to determine key performance indicators (KPIs) and create focused marketing campaigns. KPIs in terms of grocery retailers include.

- Revenue per visitor
- Average order volume and basket size
- E-commerce contribution to total orders and revenue
- Pick rate and time to fulfillment
- Fulfillment cost per order
- Profit
- Inventory

(Team, 2022)

Our data is limited in its capacity to address all these indicators, so we aim to use the information to address the KPIs that we can. From there, devise a strategy to collect additional data on inventory, fulfillment, and cost metrics using tracking and reporting through various campaigns. This will enable the client to improve marketing strategies and lucratively engage with their target market (Team, 2019)

2. General Goal

Our goal is to realize patterns in the data to see where marketing strategies, product placement, promotions, and customer integration can boost sales and expand customer reach. Our priorities are to determine key indicators and questions.

We will do this with two strategies in mind:

(i) **Retention and Upsell:** Get more money out of existing customers
(ii) **Acquisition**: Obtain new customers and/or a new customer reach

With this in mind, we will use the data to assist the company in changing its product to better suit its target customers across various customer categories. For instance, a company can assess which customer segment is most likely to purchase the product and then market the product exclusively to that specific segment rather than investing money to market a new product to every consumer in the company's database.

Our approach will be to utilize the data accumulated over this period to increase sales, customer frequency, and customer expansion and realize and expand the target audience.

The data analytics will help the shop to determine the appropriate marketing approach to achieve Retention Upsell and Acquisition.

3. Objectives and Strategy

Data is essential in driving customer engagement through various channels; online, in-store, and via an app (Colvin, n.d.). It's about building a relationship with the customer based on their profiling obtained through data.

Objectives

Objective: Determine the "Big Fish" (high-spend customers matched with high frequency) with two subcategories:

(i) **Strategy 1**: High spenders with regular visits. Analyze their basket for what they aren't buying. The benefit of this will be to get them to expand their product categories by sending them vouchers for what they don't purchase

(ii) **Strategy 2:** High spenders with infrequent or low visits: The idea would be to look at what they are buying and look at running a promotion around that and complementary products

(iii) **Strategy 3:** Low spenders with regular or infrequent visits. The objective here would again be to analyze the basket and establish buying patterns around these products to get them to shop exclusively at the store

Potential Campaigns After Profiling

From this platform, the data can help us profile customers to expand on this strategy, for instance:

(i) **Age:** Look at age within the 'Big Fish' two categories. This can be used in different ways, for example.

e.g., If the dominant age group is in their 20s, an ad campaign or social campaign targeting that profile, or creating brand ambassadors would be appropriate

(ii) **Average Spend:** The idea here would be getting this category to spend more

e.g., Buy one get one free offer

(iii) **Low income:** Again, the shop would ideally look at what they are buying and from this perspective adopt strategies such as

e.g., Create promotions around the products they are buying

e.g., Consider adding complimentary items to their inventory

Strategy: Retention and Upsell

By conducting these analyses, it allows the retailed to encourage spending, incentivize loyalty and introduce new buying options

Strategy: Acquisition

Profiling current customers can help identify trends and create campaigns around popular patterns.

Challenge

Our challenge with this dataset is that there isn't information about specific basket items themselves. However, we have valuable information on spending habits, and personal circumstances, such as family, income, education, expenditure, sex, and age. We also know the channels used to purchase. The same strategies can be used in their product category and personal circumstance while collecting fresh data to leverage marketing campaigns and increase revenue.

Limitations

We were unable to do a market basket analysis to determine customer purchase trends because the product inventory is categorized by section, and individual product categories are not provided. Implementing a new inventory management system that enables more precise product categorization is one possible solution. This would help establish what buying incentives would be appropriate, ie., loyalty programs, meal deals, brand teaming (Stodall, 2011) sampling, and seasonal promotions.

This would highlight:

- High Spend customers are buying and create promotions of what would go with that or would make them buy more
- If there is an opportunity to collaborate with a competitor with the business's own brands?
- Use in-store tactics based on what the business's target consumers are buying

4. Success Indicators / Criteria

Data Analytics can be used to measure a company's success (Sharma, 2016). We can achieve specific goals to outline a roadmap that would define the company's current state and desired goals (Sharma, 2016)

By aligning our objectives as analysts with that of the client, the analysis will be successful in its own right (Hicks and Peng, 2019). The indicators of this success are that, if the recommendations given, based on the examination, benefit the client, increase sales (Hussain, 2019), and help establish a brand identity (Kyamko, 2022). For example, Lidl and Aldi establish a positioning as bargain product companies aimed at low-cost items (BMarketingstrategy.com, 2021), whereas Supervalu, recognized their size limitations next to their international competitors and concentrated on a deep-rooted Irish own brand collaboration and emotional connection ("ADFX Awards | SuperValu: How a brave local brand defied the forces of globalization," n.d.)

SECTION 2: TECHNOLOGIES USED FOR DATA MINING AND ANALYSIS

1. Tools Employed

The approach for data mining was employed using Jupyter Notebooks in a Python 3.10.9 environment to execute the project.

2. Models Implemented and Considered

To mine the customer data and derive intuitive conclusions, we implemented several machine learning algorithms, being:

(i) K-means clustering
(ii) DBSCAN
(iii) Hierarchal clustering
(iv) Neural Networks
(v) Principle Component Analysis

We tried these based on their relative performance and in consideration of our business remit.

Artificial Neural Networks, recognize data based on their relationships, similar to that of a human brain, and make predictions on patterns. (IBM, 2023), therefore we trialed it for prediction. This algorithm wasn't suited to our analysis and didn't perform well. For this reason, we discounted it from the analysis.

Principal Component Analysis uses dimensionality reduction for high-dimensional data to preserve important information. Our data didn't' perform well under this so it was also discounted.

3. Libraries

Data Manipulation and Analysis Libraries

NumPy

Pandas

Dataprep.eda

DateTime

Data Visualization Libraries

Matplotlib.pyplot

Matplotlib.ticker

Seaborn

Data Preprocessing and Machine Learning Libraries

Sklearn.preprocessing.LabelEncoder

Sklearn.preprocessing

Sklearn.preprocessing.OrdinalEncoder

Sklearn.preprocessing.MinMaxScaler

Sklearn.cluster.KMeans

Sklearn.metrics.silhouette_score

Scipy.cluster.hierarchy

Scipy.cluster.hierarchy.linkage

Scipy.cluster.hierarchy.dendrogram

Scipy.cluster.hierarchy.cut_tree

Scipy.cluster.hierarchy.dendrogram

Scipy.cluster.hierarchy.linkage

Scipy.cluster.hierarchy.cophenet

Scipy.cluster.hierarchy.fcluster

Scipy.spatial.distance.pdist

Scipy.cluster.hierarchy

Deep Learning Libraries

Tensorflow

Keras

Keras.Models Sequential

Keras.Models Dense

Model Training Libraries

Sklearn.model_selection.train_test_split

Sklearn.preprocessing.StandardScaler

Other Libraries Utilized

os

math

4. Machine Learning Algorithms

(i) KMeans
(ii) Hierarchical
(iii) DBSCAN

We decided on these algorithms as they best identified profiling assumptions and ultimately led us to our marketing strategy appropriate to the client.

Comparing the models, K-Means is a centroid-based clustering that divides data into unique clusters. It is relatively simple to implement and works well with large datasets, and is suitable for numerical data. It assumes spherical clusters and requires a defined K value.

Hierarchical clustering is an agglomerative clustering method that determines the number of clusters based on data, however, it is computationally expensive and slow for large datasets. KMeans is more scalable and easier to interpret.

DBSCAN is a density-based spatial clustering algorithm that can group data points into clusters based on their density without requiring you to specify the number of clusters. (Yıldırım, 2020) Flexible cluster shape and size, robust against noise and outliers, and no need to specify the number of clusters.

SECTION 3: WHAT HAS BEEN ACCOMPLISHED SO FAR

1. Data Methodology

Approach

Using Data Analytics, we approached the data with a focus on

- Prescriptive
- Descriptive
- Predictive

Analytics can be vase but this methodology is complementary and valuable to business success and survival (University of New South Wales, 2020). We will utilize the tools available to create insights that will highlight growth strategies.

It's also important for the business to maintain a cognitive recognition of consumers' capabilities and needs in terms of their daily lives. Purchase behaviour directly reflects needs, desires, material and non-physical interests (Tao et al., 2022). External influences are attributed to these. These can then be beneficial in buying predictions of consumer psychology, using influences such as current trends, climate, and economic and political circumstances (Tao et al., 2022)

Exploratory Data Analysis for Data understanding

EDA immediately shows the distribution of all the Features to give an insight into how to approach cleaning and Feature Selection

The Dataset provided on Kaggle had a tabulated csv file that required us to render it into a data frame by specifying the separator as " \t"

The data frame contained 29 features each with 2240 observations except for the float type feature "Income "with 24 missing values. Other object-type includes integer and object types.

'Education' and 'Marital Status' have categorical values that we will encode to get a numerical dataset.

'Dt_Customer' is in a datetime format that will need to be converted.

There are zero valid data values in the data frame variables including values in the amounts spent on some products including Fish, zfruits, sweets, and catalog purchases

Marital_Status contains unrecognizable values so we will convert them to Null.

We drop the following columns as they do not hold information about our business plan.

- Campaign Marketing
- Complain
- Z_CostContract
- Z_Revenue
- Response

Data Cleaning and Preparation

The object type variables "Marital status ", and "Education" were analyzed, values falling outside the distinct class context within the Marital Status variable "YOLO'' and Absurd" were first replaced with "Nan " values and dropped since they constitute less than 5% of the data.

These variables were further engineered by encoding to convert them to numerical types that can then be computerized.

The date of purchases was also reformatted from object type to date-time format which represents the standard for time stamps and further converted to the integer type.

Variables considered to hold irrelevant information regarding the goal of this project were dropped. These include Campaign marketing, complaint, z_Costcontract, z_revenue, and response.

The distribution of the missing values in "Income was first observed to determine the most appropriate approach to resolve them". Very few outliers, skew >2, and normal distribution resolved to replace missing values with the median value of the data values.

Missing Values

24 missing values were identified in the "Income" feature and 4 in the "Marital Status" feature. The 4 missing values under Marital Status were dropped as they won't impact the overall data loss. The 24 missing values under Income were replaced by the median of the feature as data is significantly skewed to the left.

"Another technique is median imputation in which the missing values are replaced with the median value of the entire feature column. When the data is skewed, it is good to consider using the median value for replacing the missing values. Note that imputing missing data with median value can only be done with numerical data." (Kumar, 2020)

Feature Engineering

Some features were modified and added to the dataset to aid in data understanding and visualization.

Profiling

This included transforming 'Dt_Customer' to the same format as 'Year_Birth' (DateTime) and based on that we created a new column named 'Consumer_Duration' containing the number of years that the individual is a customer.

An 'Age' column was created from the 'Year_Birth' column for general distinction.

The ages of the customers were calculated by simply subtracting their birth year from the current year and a new column was appended and named "age". The same was done to calculate the time interval between the customer's last visit and the current date. These new columns were created for

better understanding and viewing of the data. The old variables/ columns were dropped since they are no longer needed.

Based on the feature "Age" customers were mapped into three different categories.

- Gen Z = between 0 and 30 years old
- Gen Y = between 31 and 55 years old
- Gen X = between 56 and 83 years old (excluding outliers)

Another column was created for total expenditure by adding the amounts spent on each product for each of the customers.

A new column named 'Total_Children', concatenated from the which is the sum of the two original columns 'Kidhome' and 'Teenhome'. That will ultimately reduce the number of observations to be plotted and analyzed.

From there we visualized 'Total Spend' for 'Items Bought' and the number of customers by spend profile. These profiles were divided into low, average, and high spending.

Correlations

A correlation matrix was done to analyze the relationship between the variables in the dataset. For better viewing, diverging pallets were used and the values of the correlation were annotated.

It was observed that the highest positive correlation occurs between total spending and the number of wines with (0.89) followed by meat (0.84) meaning more is spent on these products than the total amount spent. The amount spent also correlated strongly with income having a (0.79) correlation value. In these cases, age was not a determining factor.

On the other hand, Income correlated negatively with the number of web visits showing that people with lower income checked the site more often than people with higher income. The number of children also had a slightly negative correlation with total spending. Education has a very weak correlated value of 0.09 with total spending and cannot be considered as an index for measuring total spending.

Some features were modified and added to the dataset to aid in data understanding and visualization.

Preprocessing: Outlier Treatment and encoding

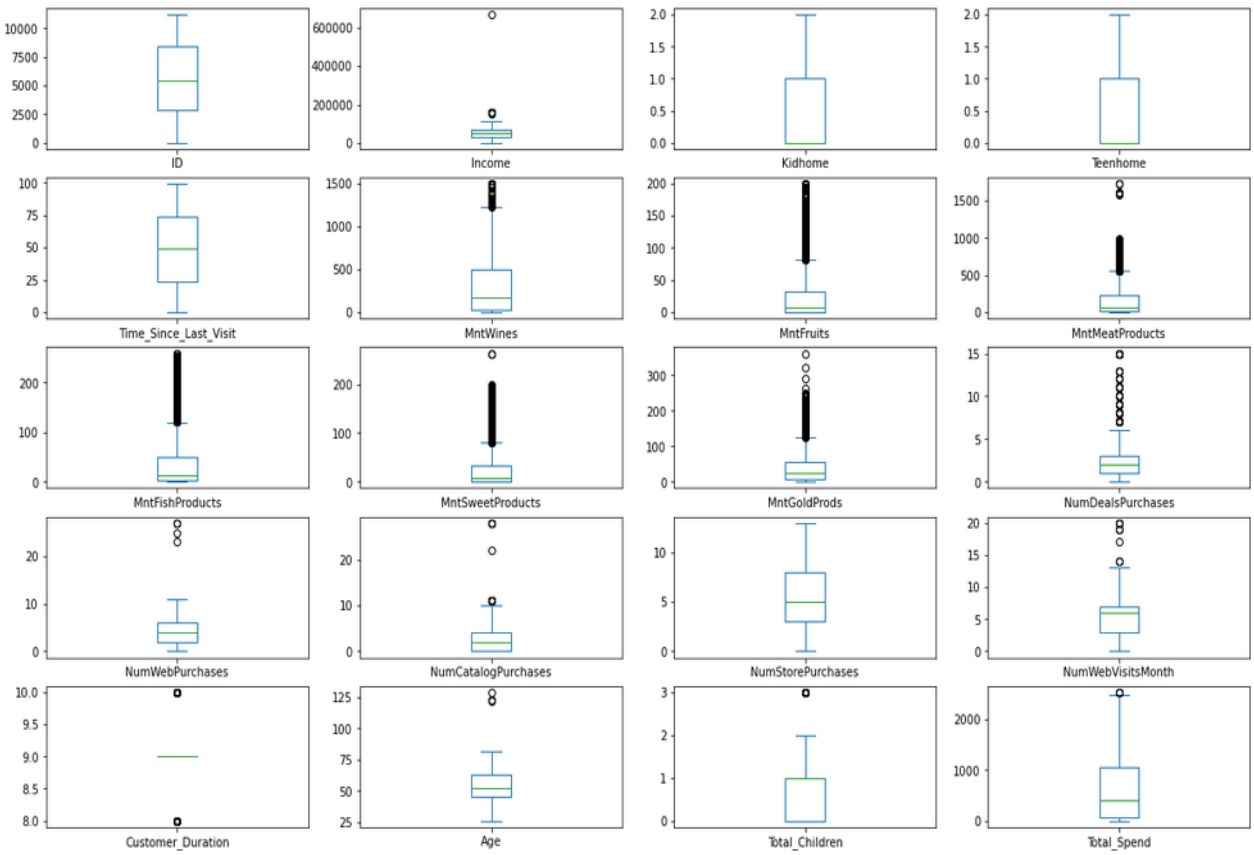
Outliers

Considering the features are skewed the Interquartile range proximity rule was applied to identify outliers in the numerical features.

Outliers were identified under features "Income", "Age" and "Total Spend".

Looking at visualization (box plot) and the outlier's detection via the Interquartile range proximity rule we have decided to drop outliers that were much higher than the other outlier points identified. Income higher than 600k was dropped and age higher than 100 was also dropped, leaving a few real outliers that are closer to the mean/median and that will not affect the models we are planning to implement.

It's noted features containing outliers that were handled during data preparation/outlier detection were realized using the quartile percentile.



Encoding

- Education = Ordinal encoder (as it is an ordinal variable)
- Age_Profile = Ordinal encoder (as it is an ordinal variable)
- Marital_Status = Dummy Encoder
- Cust_Profile = Label encoder (as it is our label for customer spend profile)

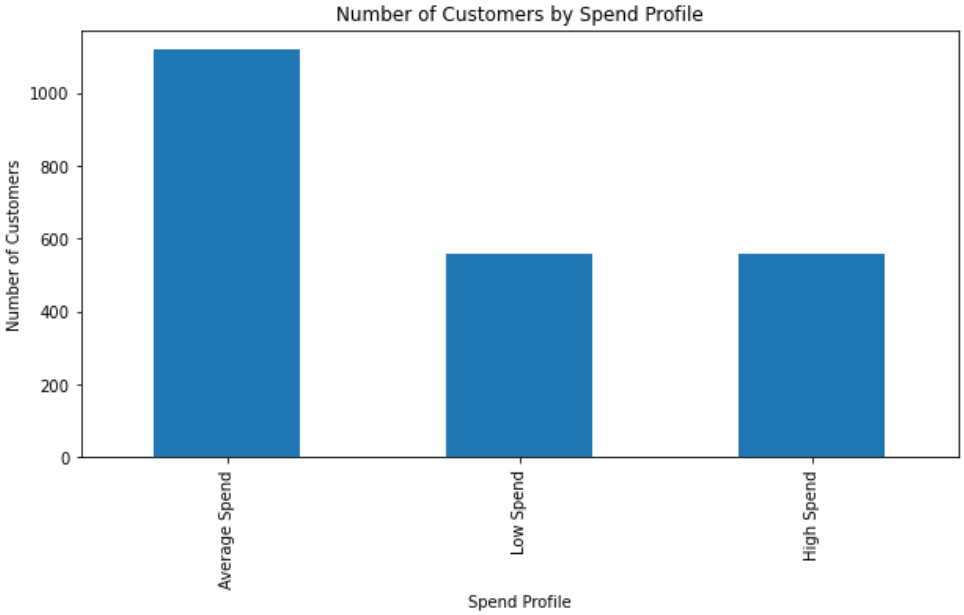
2. Data Visualisations

In our early EDA stage, we used pair plots and heatmaps to identify the correlation between the features and observed that features with high correlation are:

- Total_Spend / All products spend features
- Total_Spend / Income
- Total_Spend / NumCatalogPurchases
- Total_Spend / NumStorePurchases
- Total_Spend / NumWebPurchases
- NumCatalogPurchases / MntMeatProducts
- NumCatalogPurchases and NumStorePurchases / MntWines
- All products spend features / Income
- All products spend features among themselves (eg. MntWines / MntMeatProducts)
- NumCatalogPurchases and NumStorePurchases / Income (while NumWebVisitsMonth is inversed correlated to Income)

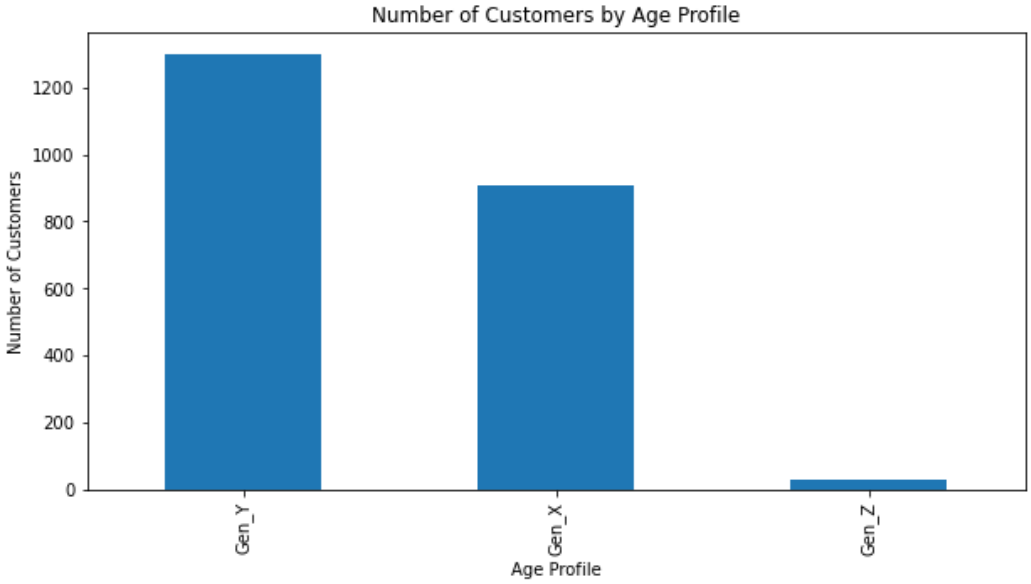
The Total_Spend feature was used to create three different categories of customers (Figure 1)

- High Spend = between 1046 and 3000 $USD
- Average Spend = between 69 and 1045.5 $USD
- Low Spend = between 0 and 68.75 $USD

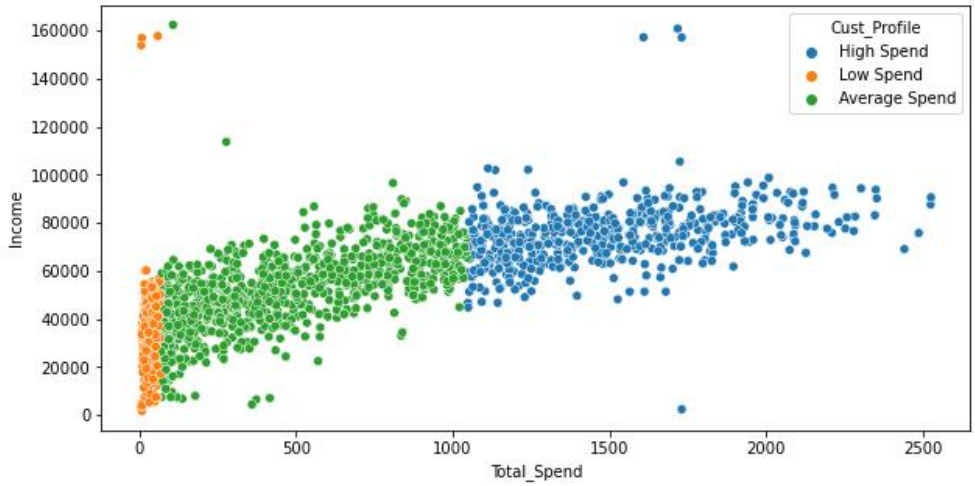


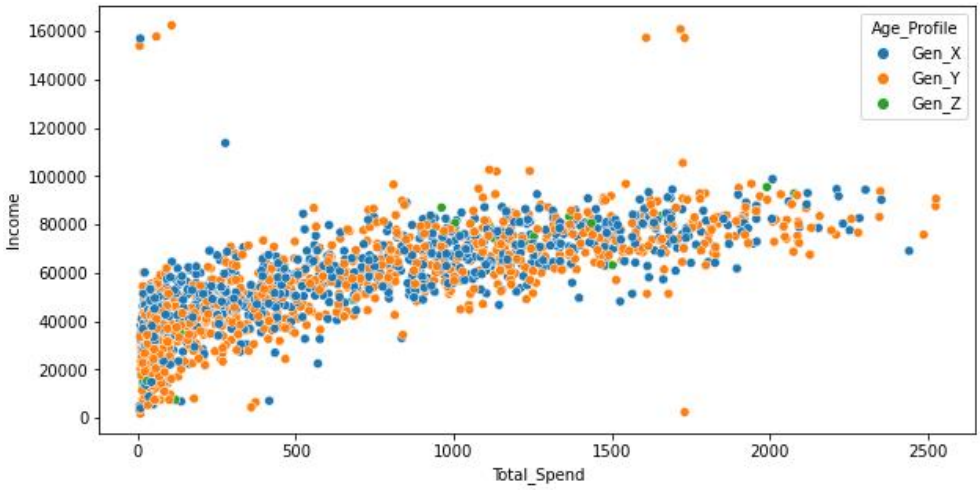
Age was also split into three different categories (Figure 2)

- Gen Z = between 0 and 30 years old
- Gen Y = between 31 and 55 years old
- Gen X = between 56 and 83 years old (excluding outliers)



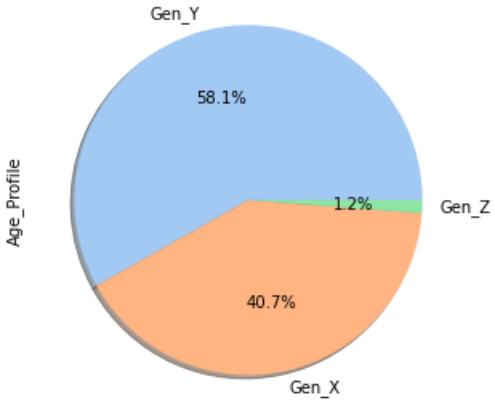
Using the categories created for Spend and Age, we can see that total spending is highly correlated to the customer's income and education level but not so much to the customer's age group. We aim to find a better correlation across the different customer profiles and what they buy to establish the best strategy approach after applying the clustering and market basket analysis models.

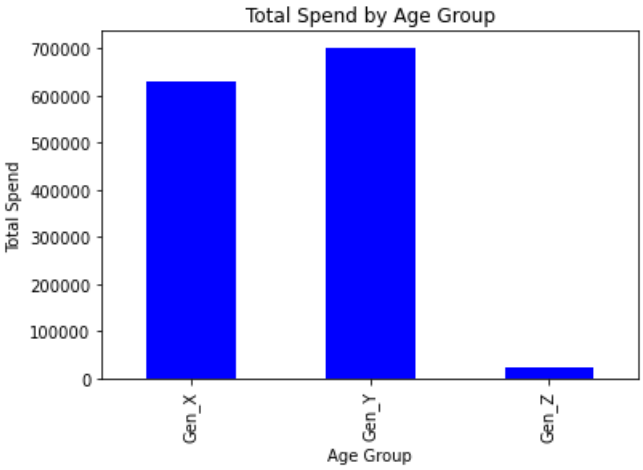


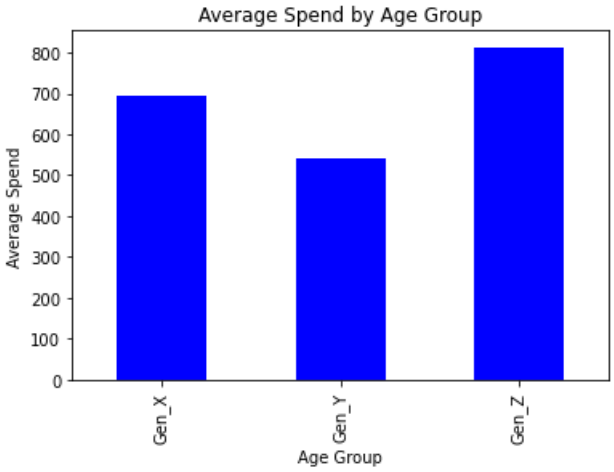


Age Group Insights

Most of our customers (58.1%) are classified under Gen Y = Age between 31 and 55. On average highest spenders are classified under Gen Z = Age between 56 and 82 - (max age excluding outliers)

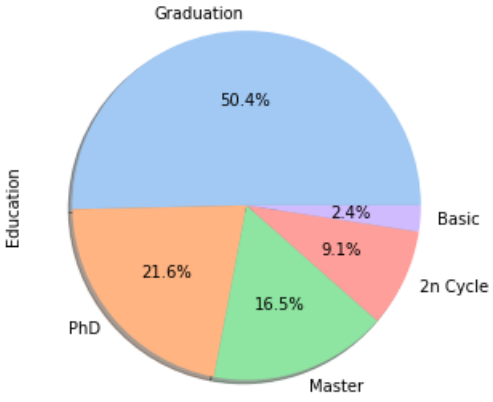


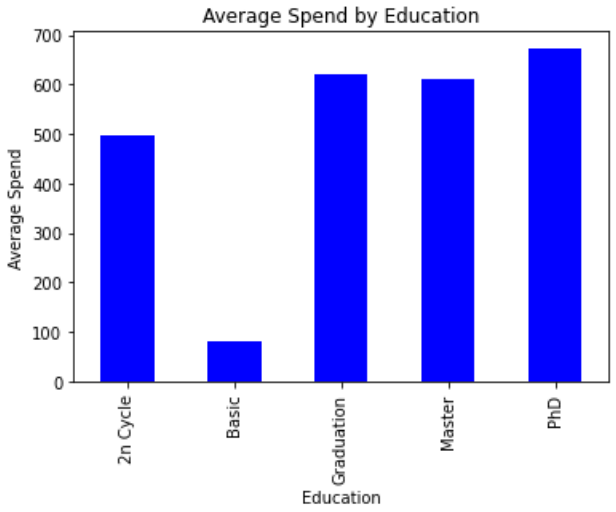




Education Group Insights

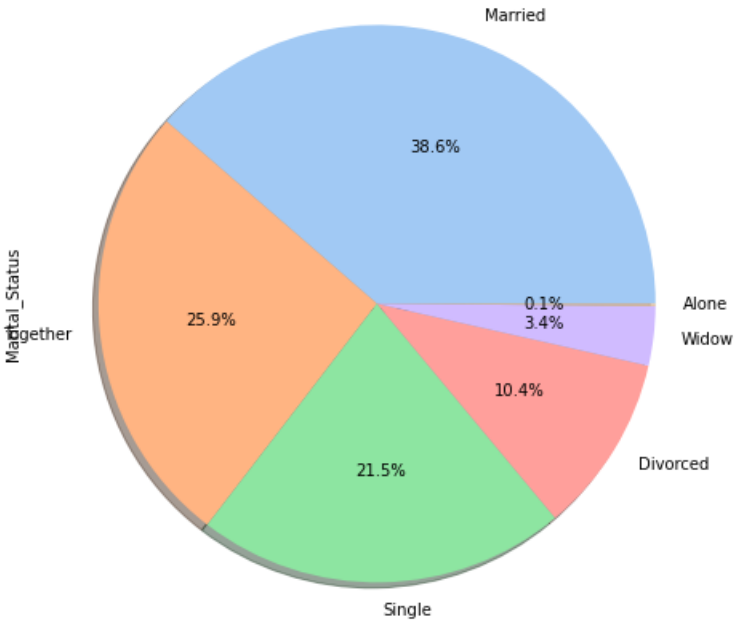
Around 50% and most customers have only graduated. We can see in the graphs below the total spend by education level as well as the average spend by education level, which shows that the highest spenders on average are customers with a Ph.D.

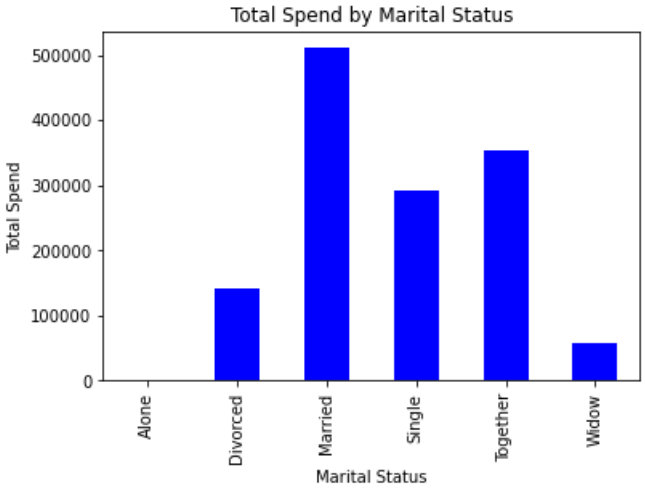


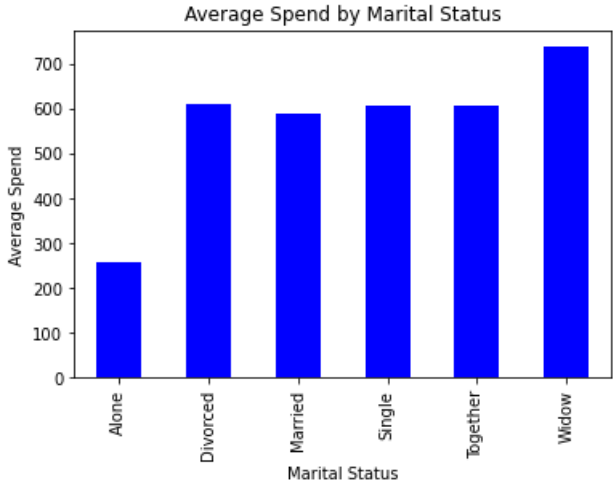


Marital Status Group Insights

Most of our customers 38.6%) are Married. On average highest spenders are widow customers




Total Spend by Marital Status
Total Spend
500000
400000
300000
200000
100000
0
Alone
Divorced
Married
Single
Together
Widow
Marital Status


Average Spend by Marital Status
Average Spend
700
600
500
400
300
200
100
0
Alone
Divorced
Married
Single
Together
Widow
Marital Status

3. Feature Selection and scaling

A seaborn heatmap was used to establish feature importance which establishes that the columns ID and Time_Since_Last_Visit have a low correlation with the other features, and the columns "Kidhome" and "Teenhome" are duplicated with the column "Total_Children" so before passing through the model I will drop these columns.

We have scaled the data to allow us to process the data without distorting differences in their ranges This was done through a min/max scaler to normalize the data and improve the performance of the algorithm.

SECTION 4: OBSTACLES AND STRATEGIES

1. Obstacles

The Dataset did not provide much information regarding the specific products themselves, therefore we could not employ market basket analysis and demographic analysis was a better approach. The benefits are engaging with your target customer to communicate and gather information for marketing purposes to expand your data reach (Kelly, 2023)

Working as a team remotely was challenging in terms of time management and coordination.

There was also a lack of information as to the store's budget, technological and social reach, and marketing platform.

2. Strategies

We concentrated on customer profiling to compensate for the lack of product and campaign data. Demographic information assists in recognizing and understanding your target customer. Demographic trends are essential to recognize as current Economical, Geographical, Social, and Political environments influence buyers (Wallace, 2022)

Remote collaboration required us to be diligent and organized in our communication. We initially set up a shared online drive to maintain version control but migrated to GitHub as it's a good portfolio basis along with version control, open-source contributions, and document maintenance

Tasks were divided evenly. Each member was assigned an algorithm and it was a collaborative effort to establish the best results.

The lack of information motivated a strategy that would elevate the client's position for better data collection and sales processes which would lead to the next stage of marketing, in turn, Retention Upsell, Acquisition, and possibly expansion.

The analysis will realize Strengths, Weaknesses, Opportunities, and Threats (SWOT) that can be leveraged for effective business decisions (Schooley, 2023)

SECTION 5: RESULTS AND ANALYSIS, NEXT STEPS

After evaluating several models, the Machine Learning Algorithm that provided the most useful and relevant insights was Kmeans. As it had the highest level of accuracy and clear definition of clusters, it was the best model to identify purchasing behaviour for each customer group. This accuracy in

the analysis is essential to deliver informed insights for beneficial business and marking decisions that will achieve increased sales. The following characteristics determined the following:

1. Results, Engagement, and Next Steps

Results

We determined the algorithms best suited to this analysis were, DBSCAN, Kmeans, and Hierarchical clustering.

Insert dbscan and hierarchical performance

Ultimately, Kmeans would provide us with the analysis we needed for an efficient and effective predictive system to provide the client with a forward strategy

The clusters demonstrated the following profiling:

Spend

- Cluster 0 is low /average spend;
- Cluster 1 is low spending;
- Cluster 2 is average/high;
- Cluster 3 is High spending;

Clusters could be classified as follows

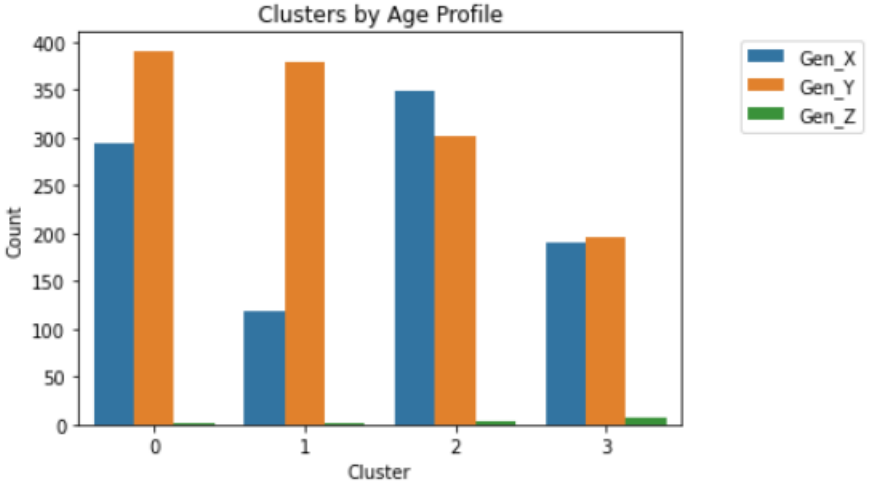
- Cluster 0 and 1 are customers that we want to retain and upsell;
- Cluster 2 are customers we want to focus on upselling through promotions and bringing new customers;
- Cluster 3 we will look into what products they are not buying and create promotions around that and also use them to bring more customers;

Income

- Cluster 0 is low /average Income;
- Cluster 1 is low Income;
- Cluster 2 is average/high;
- Cluster 3 is High Income;

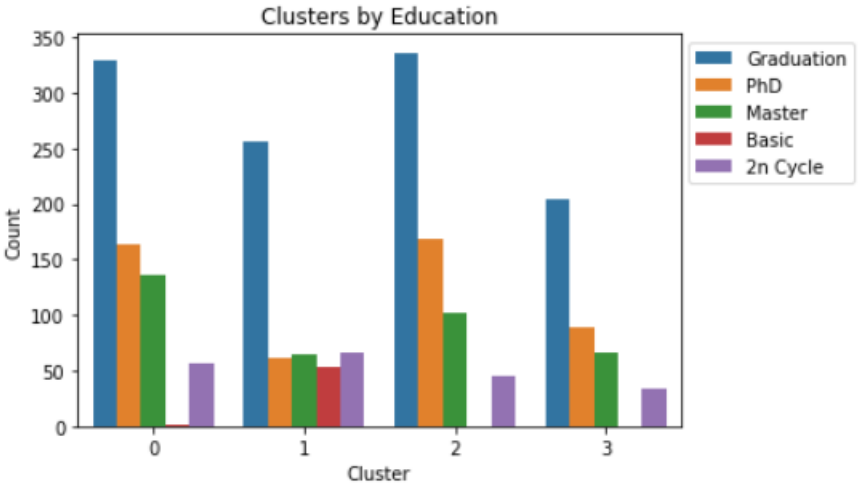
Age

- Cluster 0 average age between 50 and 55 years old;
- Cluster 1 are younger customers with an average of 49 years old;
- Cluster 2 are older customers with an average between 55 and 60 years old;
- Cluster 3 are average age customers between 50 and 55 years old;



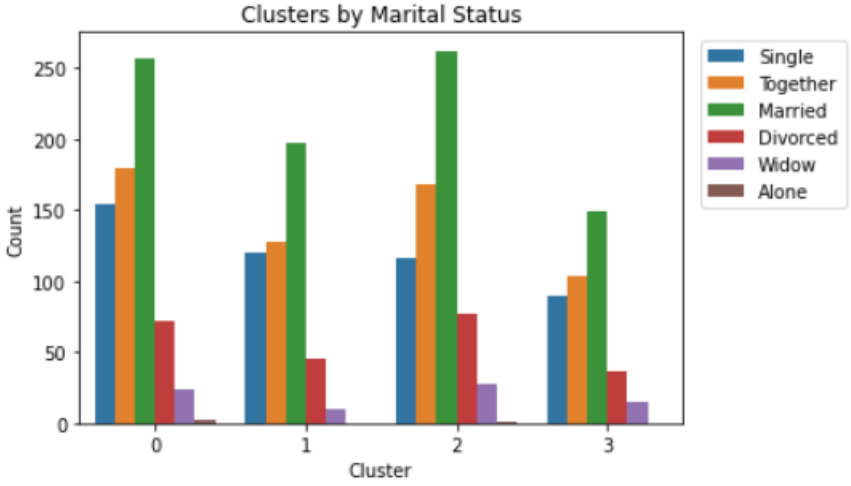
Education

Education is similarly distributed across all clusters with higher education in clusters 0 and 2, however, we have to take into account that these two clusters are the biggest clusters so we can assume education is pretty flat across the different clusters with no differentiation.



Marital Status

This was not a differentiator as it did not differ much per cluster



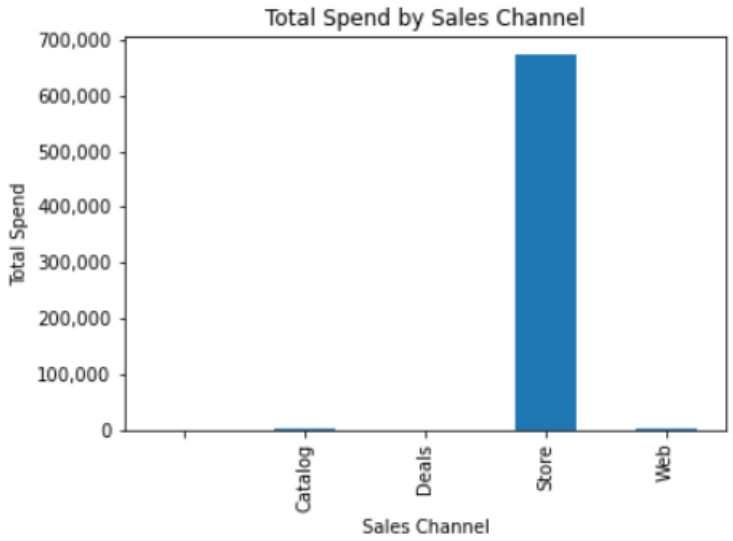
Number of Children

Cluster Number of children characteristics:

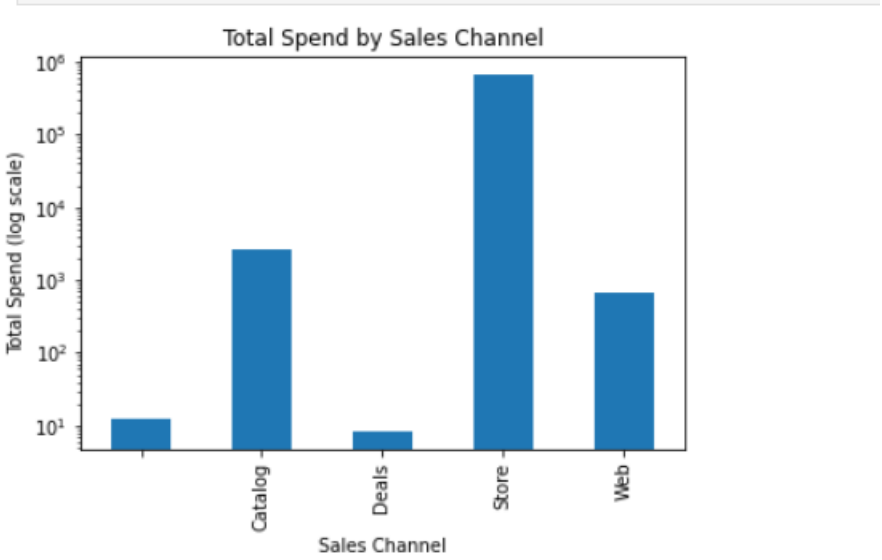
- Cluster 0 between 1 and 2 children;
- Cluster 1 customers has in average 1 child;
- Cluster 2 between 0 and 1 child;
- Cluster 3 majority of customers do not have children;

Engagement

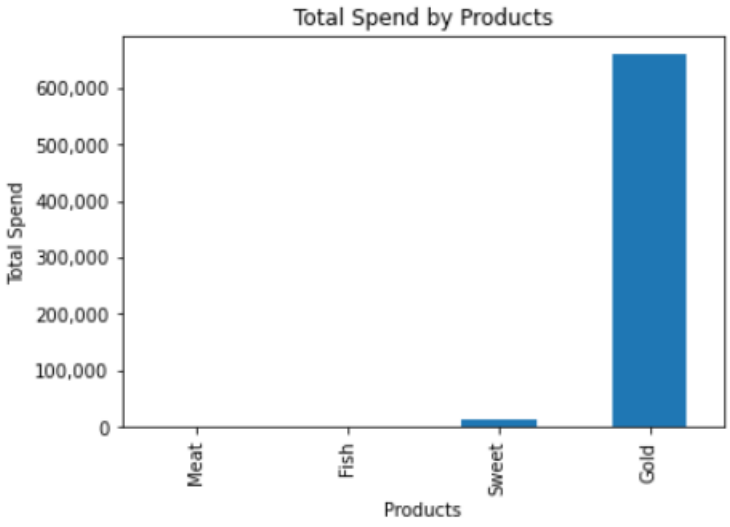
The best performer was 'Store Bought' over the web, catalogue, or deal purchases. This would indicate that two strategies would need to be employed when considering Retention and Upsell / Acquisition



The main sales channel is in store



The gold category outperformed other categories



2. Next Steps

Customer Profiling for Marketing and Sales

Budgeting Families

- Focus on campaigns that stress the value and affordability of items; Offer family-friendly products and services, as the majority of this cluster has 1-2 Children
- Upsell through online advertisement, promotions, and discounts to drive additional spending in-store

GenZ Savers

- Provide specials and deals that are pocket-friendly because people in this group have low incomes and frugal purchasing patterns
- Create marketing efforts that promote the convenience and enjoyment of shopping in-store and are geared toward younger consumers
- Examine providing goods and services in the areas of technology, fashion, and entertainment that are tailored to the wants and needs of young adults.

Secure Older Adults

- As this demographic has average to high income and spending habits, create marketing campaigns that emphasize the quality and dependability of products and are geared toward older consumers
- Consider providing discounts and promotions that appeal to clients who prefer making purchases online
- Provide goods and services that are geared toward the wants and needs of senior citizens, such as travel, recreation, and health and wellbeing

High-Income Singles

- These are our top clients. To appeal to this cluster's high spending tendencies, emphasize the exclusivity and elegance of the products

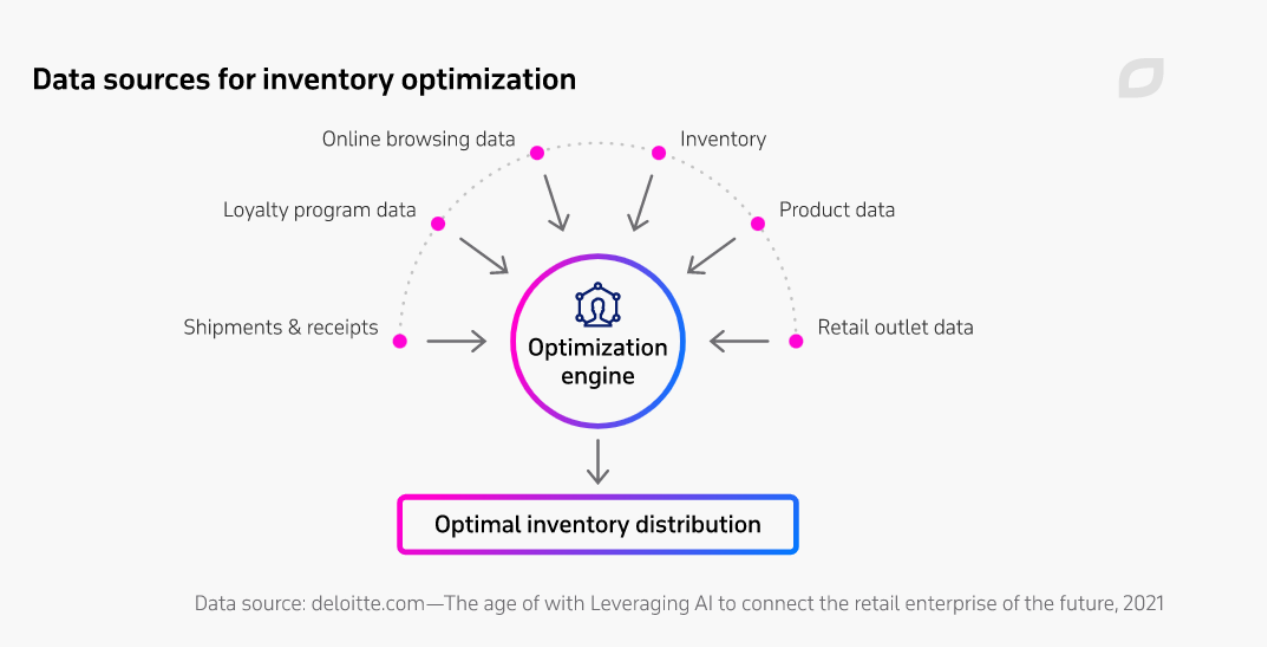
- Develop a unique customer experience, benefit-rich products and services, and a premium card with special offers

- Offer products and services that cater to the needs and interests of singles, such as organic produce, premium products, healthy products, meat, fish, wine, and beer

Data Collection

It is evident from our initial look that a better grasp of data collection, monitoring, and maintenance is essential for business continuity. It highlighted that there was a need for multiple channels, including 'virtual sources such as social media and e-commerce platforms, in combination with that of physical store collection (Di Stefano, 2023)

Another feature that was evident in that it needed attention was inventory and the lack of information or variety around that. It is prudent to monitor not only buying habits but also behaviours around product placement and frequently co-purchased items (Di Stefano, 2023)to forward this study to Market Basket Analysis



("The Age of With™," 2021)

Further Machine Learning Resources

A Proposition from the analysis would be to use the outcomes to leverage market position. An approach would be to interpret causal inference to study the cause-and-effect relationships of these campaigns. (Malik, 2022) Various Python libraries can be used in this capacity such as DoWhy or Causalnex

SECTION 6: CONCLUSION

The goal of the analysis is to bring about both profitable and reputational change in the organization. Reputation exists in both the character of a business but also in the audience it hopes to engage with (Eccles et al., 2007) and this reputation has a direct correlation with the type of marketing that is employed and the profitability of that organization (Comparably, 2021) Customer profiling and sales analytics, provide the resources needed to market for organizational change to increase profit margins and increase reach. These trends are the groundwork to develop a roadmap to achieve the same.

We researched through rival retailers the approaches that best enable the business to increase sales, profit, and engagement and improve on planning through inventory, data collection, marketing, and technological updates (McKinsey & Company, 2015)

The analysis process was one of trial and error. We determined that Neural Networks was not providing good results and accuracy was low. We also performed PCA which did not improve the performance of our models. Both these processes were dropped

Based on our identified target groups as Budgeting Families, GenZ Savers, Secure Older Adults, and High-Income Singles. Through data preparation, cleaning, feature processing, engineering, and machine learning we've been able to gain insight into the business potential for targeted marketing initiatives, future campaigns, and increased revenue and business expansion.

Our ultimate goal was to establish profiling to fuel the marketing strategy of 'Retention and upsell' and 'Acquisition'. We realized statistical trends in the data using Kmeans and this can be fed into the strategies outlined in Objectives and Strategies

BIBLIOGRAPHY

Comparably, 2021. Why Reputation is Undeniably Important in Business | Comparably [WWW Document]. URL https://www.comparably.com/news/why-reputation-is-undeniably-important-in-business/ (accessed 8.16.21).

Eccles, R.G., Newquist, S.C., Schatz, R., 2007. Reputation and Its Risks. Harvard Business Review.

Hicks, S., Peng, R., 2019. Evaluating the Success of a Data Analysis.

Hussain, D.M., 2019. What does success look like in Data Science? | LinkedIn [WWW Document]. URL https://www.linkedin.com/pulse/what-does-success-look-like-data-science-m-maruf-hossain-phd/ (accessed 12.16.22).

Sharma, S., 2016. The Key Success Criteria for Implementing a Successful Data Strategy | LinkedIn [WWW Document]. URL https://www.linkedin.com/pulse/why-your-business-intelligence-bi-project-fail-samir-sharma/ (accessed 12.16.22).

Stodall, H., 2011. Doritos unites with Pepsi Max to cross-sell crisps and drinks [WWW Document]. The Grocer. URL https://www.thegrocer.co.uk/doritos-unites-with-pepsi-max-to-cross-sell-crisps-and-drinks/221151.article (accessed 12.16.22).

Kumar, A. (2020). Python - Replace Missing Values with Mean, Median & Mode. [online] Data Analytics. Available at: https://vitalflux.com/pandas-impute-missing-values-mean-median-mode/.

ADFX Awards | SuperValu: How a brave local brand defied the forces of globalisation [WWW Document], n.d. URL https://adfx.ie/databank/supervalu-how-a-brave-local-brand-defied-the-forces-of-globalisation.html (accessed 5.3.23).

BMarketingstrategy.com, 2021. Marketing Strategy of LIDL - Business Marketing Strategy. URL https://bmarketingstrategy.com/marketing-strategy-of-lidl-lidl-marketing-strategy/ (accessed 5.2.23).

Colvin, J., n.d. Customer acquisition and retention: Which should you focus on? [WWW Document]. URL https://www.mparticle.com/blog/customer-acquisition-and-retention/ (accessed 5.5.23).

Di Stefano, A., 2023. Machine Learning in Retail: 10 Use Cases & Business Benefits [WWW Document]. URL https://www.itransition.com/machine-learning/retail (accessed 5.5.23).

IBM, 2023. What are Neural Networks? | IBM [WWW Document]. URL https://www.ibm.com/topics/neural-networks (accessed 5.6.23).

Kelly, K., 2023. Understanding Your Customers: How Demographics and Psychographics Can Help [WWW Document]. URL https://extension.psu.edu/understanding-your-customers-how-demographics-and-psychographics-can-help (accessed 5.4.23).

Kyamko, M., 2022. 7 Strategies to Help You Build a Strong Retail Brand. crowdspring Blog. URL https://www.crowdspring.com/blog/retail-branding-strategies/ (accessed 5.2.23).

Malik, A., 2022. The Beginner's Guide to Causal Inference for Making Effective Business Decisions [WWW Document]. Medium. URL https://towardsdatascience.com/the-beginners-guide-to-causal-inference-for-making-effective-business-decisions-a9c7ca64d9dd (accessed 5.5.23).

McKinsey & Company, 2015. Big-Data-Ebook.

Schooley, S., 2023. How SWOT Analysis Can Help Grow Your Business [WWW Document]. Business News Daily. URL https://www.businessnewsdaily.com/4245-swot-analysis.html (accessed 5.4.23).

Stodall, H., 2011. Doritos unites with Pepsi Max to cross-sell crisps and drinks [WWW Document]. The Grocer. URL https://www.thegrocer.co.uk/doritos-unites-with-pepsi-max-to-cross-sell-crisps-and-drinks/221151.article (accessed 12.16.22).

Tao, H., Sun, X., Liu, X., Tian, J., Zhang, D., 2022. The Impact of Consumer Purchase Behavior Changes on the Business Model Design of Consumer Services Companies Over the Course of COVID-19. Frontiers in Psychology 13.

Team, C., 2022. Online Grocery KPIs: Learn the 8 Metrics That Matter [WWW Document]. Constructor. URL https://constructor.io/blog/grocery-kpis/ (accessed 5.2.23).

Team, L., 2019. Marketing Effectiveness: What It Is and 4 Ways to Measure It. Leadspace. URL https://www.leadspace.com/blog/marketing-effectiveness/ (accessed 5.2.23).

The Age of With™: Leveraging AI to connect the retail enterprise of the future [WWW Document], 2021. Deloitte Canada. URL https://www2.deloitte.com/ca/en/pages/consumer-industrial-products/articles/age-of-with-ai.html (accessed 5.5.23).

University of New South Wales, 2020. Descriptive vs. Prescriptive vs. Predictive Analytics Explained [WWW Document]. Business Analytics. URL https://www.techtarget.com/searchbusinessanalytics/tip/Descriptive-vs-prescriptive-vs-predictive-analytics-explained (accessed 5.3.23).

Wallace, F., 2022. 6 major demographic macro trends shaping retail [WWW Document]. Farnaz Global. URL https://www.farnazglobal.com/perspectives/6-major-demographic-macro-trends-shaping-retail (accessed 5.4.23).

Yıldırım, S., 2020. DBSCAN Clustering — Explained [WWW Document]. Medium. URL https://towardsdatascience.com/dbscan-clustering-explained-97556a2ad556 (accessed 5.6.23).