REPORT

ON

CUSTOMER ANALYTICS AND SEGMENTATION: MACHINE LEARNING AND DEEP LEARNING INSIGHTS FOR CRM

BY

NEELESH NAGPAL 2020A7PS0139U CS

COURSE NUMBER: CS F376

Prepared in Fulfillment of the
Design Project Course

AT

BITS Pilani, Dubai Campus
Dubai International Academic City (DIAC)
Dubai, U.A.E

First Semester
(SEPTEMBER 2022 - DECEMBER 2022)

BITS Pilani, Dubai Campus
Dubai International Academic City (DIAC)
Dubai, U.A.E

Course Name: Design Project **Course No:** CS F376

Duration: 1-Sep-2022 to 30-Dec-2022 **Date of Start:** 1-Sep-2022

Title of the Project: Customer Analytics and Segmentation: Machine Learning and Deep Learning insights for CRM

Student Name: Neelesh Nagpal **Student ID:** 2020A7PS0139U

Discipline of Student: Computer Science Engineering

Name of Project Supervisor: Dr. Tamizharasan P S

Key Words: Customer Analytics, Descriptive and Predictive analysis, STP Framework, Machine Learning model, Deep Learning model, Customer retention and conversion.

Project Areas:

- Segmentation, Targeting and Positioning
- K-means clustering and PCA
- Business Intelligence
- Logistic and Linear Regression
- Artificial Neural Networks
- Elasticity modelling
- IDA (Intelligent Data Analysis)

Abstract:

The importance of gathering vast amount of data by companies every day is growing exponentially as they steer themselves towards a data-driven future. Customer analytics deals with analyzing data using machine learning techniques to harness its true power in future prediction that enhances productivity and profitability. This paper provides machine learning and deep learning insights in customer analytics. The gathered data is analyzed based on socio-demographic factors to gain initial insights. Further, K-Means algorithm is deployed to segment the dataset. However, K-Means showed better performance with PCA. This stage helps us form hypothesis which is needed for subsequent modelling. The exploratory analysis of the dataset provides us with insights about customer behavior about the brand choice and revenue each segment brings to the company. Linear

and Logistic regression is used for predictive analysis to estimate how would consumer behavior be affected with changes in prices of products. This helps us model brand choice elasticities and purchase probabilities. Customer analytics is pointless if companies cannot predict customer behavior in the future. Deep Learning provides the power to make such predictions. A black box model is created using TensorFlow framework that can make 90%+ accurate predictions.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Srinivasan Madapusi, Director of BPDC for providing the students with an opportunity to gain firsthand knowledge from working on projects under the guidance of the experienced faculty.

Next, I would like to extend my gratitude to Dr. Tamizharasan P S, my Project supervisor for his constant guidance, support and motivation that helped me understand the insights of CRM data analysis and segmentation process using ML and Deep learning techniques. His experience and ability helped me grasp how research happens at a professional level.

I would also like to thank my family for their constant encouragement and positivity that helped me in the successful completion of the project.

Student Name: **NEELESH NAGPAL**

ID Number: **2020A7PS0139U**

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1. LITERATURE REVIEW

J.R. Saura et al. (2021) developed a methodology to amalgamate B2B digital marketing with modern CRM AI-based techniques. They adopted Multiple Correspondence Analysis (MCA) framework. This paper gave a future direction of developing different CRM techniques in a traditional B2B setup using AI and Digital marketing.

N. Singh et al. (2020) developed a model using modern data mining and machine learning techniques to analyze customer dataset. Segmented data was classified using MLP, J48 and Naïve Bayes algorithm with MLP showing the most accurate and efficient results with an accuracy of 98.33% and run time of 0.26 sec.

Ahmad et al. (2019) developed a customer churn prediction model to filter out valuable customers. This model used machine learning techniques using big data. They used AUC (Area Under Curve) and SNA (Social Network Analysis) as their standard performance measures. They implemented four machine learning algorithms to a telecom company's dataset: Decision Tree, XGBOOST (best results), GBM and Random Forest.

Fares N. et al. (2019) worked to develop a machine learning approach for customer profiling. Their main objective was to increase customer engagement and increase basket size for maximum profits in fast fashion industry. They adopted Seasonality modelling along with human profile recognition as the standard measure.

M. Makhtar et al (2018) prepared a model for churn-based classification and prediction based on rough set theory. The results bypassed other commonly used machine learning algorithms like Bayes and even MLPNN in terms of accuracy. If companies can know beforehand which customers have a higher possibility of churning, they can focus their marketing efforts on that segment to avoid churning.

Sabbeh, S. F. (2018) implemented various machine learning based techniques for churn prediction on a telecom dataset. Predictive analysis helps filter out which customers have a higher probability of being retained by the company. This paper's aim was to compare the performance and benchmark eight different models when fed the same data. Models used were: Logistic Regression, CART, Ada Boost, MLP, Bayes Algorithm, Linear Discriminant Analysis, SVM and K-Nearest. Ada Boost and random forest gave the best results with a tied accuracy of 96%.

Serhat Peker et al. (2017) presented a modified RFM model with two extra parameters in form of (L): Length, of engagement of the customer with the brand and (P): Periodicity, which determines the frequency of visits to the store. Due to two additional parameters, segmentation results were significantly better.

Bahari and Sudheep Elayidom (2015) gave a framework that combined sophisticated data mining techniques with CRM to predict the behavior of customer based on their demographics and purchase history. They used Bayes algorithm to classify the customers and fed the same data to MLPNN which gave better results.

Cormac and Rozaki (2017) focused on integrating machine learning techniques to generate churning levels of customers. The report also focused on minimizing the CAC (Cost of Acquisition). C.5 Algorithm was analyzed within Bayes modelling to segment customers according to their social demographics.

He and Li (2016) developed a 3-variable solution for customer segmentation in calculating CLTV (Customer Lifetime Value) which can be visualized in three dimensions. RFM Model was chosen as the basis for segmentation, and they concluded that targeted marketing is more efficient than one size fits all strategy. Their model considered the Customer lifetime value, customer activity and their satisfaction.

1.2.CUSTOMER ANALYTICS:AN OVERVIEW

Customer analytics is the direct application of data science and analysis techniques by organizations to leverage the data gathered by them to increase their sales, optimize revenue, and improve inventory management. In today's competitive economy, two major drivers that aid businesses in creating value and dominating market spaces are data science and marketing. With the rapid growth and development of technology, companies are showing an increased interest in business intelligence using machine learning techniques. Customer analytics using state of the art technology, newer and faster machine learning algorithms is becoming the core area of interest in CRM (Customer Relationship Management). With customers connected to the internet now more than ever, where everything that interests them is just a click away, it has become essential for companies to understand their target customers to undercut their competition. Customer data has become the backbone for business intelligence. With the application of appropriate ML techniques and neural networks, this data can help provide companies with hidden information about their sales which would rather be invisible using traditional statistical modelling.

Once analyzed, the business strategists can produce suitable marketing strategies aimed at the high value customer segments. This has helped the companies to become data driven and customer centric.

It is a known fact for the companies that it is much more financially draining to acquire new customers than retaining the existing ones and this is where customer analytics steps in. Demographic data analysis of the customers can help companies figure out their highest value segment and the one which is costing them the most but not giving equitable returns. This can help strategists maneuver the marketing mix appropriately and allocate resources to retain and acquire customers from the high value segment.

1.3. STP FRAMEWORK: SEGMENTATION, TARGETING AND POSITIONING

Competition in today's market has become extremely cut-throat. Therefore, to survive in the challenging times, one size fits all approach cannot be used anymore provides the solution to this problem by providing a pathway for tailored marketing strategies. STP framework stands for Segmentation, Targeting and Positioning. This model forms the backbone of modern marketing. First, the market is segmented based on certain demographics. Each segment contains people with certain similarities such that they can be grouped. Once segmented, the best segment of customers is targeted. The companies then position their offerings in a certain manner such that they become more tempted to purchase them. The model can also identify niche markets and new customer or market prospects, thereby improving the effectiveness and cost-efficiency of marketing initiatives framework works best in B2C models as there is more data to work with as compared to B2B.

1.3.1. SEGMENTATION

Market segmentation is the first step for customer analytics using the STP Framework. The population of customers is divided into segments such that each segment has people with similar purchase behavior and similar demographics. This process helps companies produce marketing solutions targeting the segments instead of each customer individually since people in the same segment would respond similarly to the marketing strategies. This segmentation differs on a range of factors depending on the industry trends and requirements:

1.3.1.1. Geographic Segmentation

It segments the customers based on where they live and their environment. It helps companies filter out the regions where their product will be more successful. Their settlement sizes are taken in account while segmenting datasets. For e.g. It is a waste of company resources to market expensive products in a low-income neighborhood.

1.3.1.2. Demographic Segmentation

This technique segments the customers based on variables like age, gender, spending score, income, family, etc. This helps companies figure out which segment is more profitable than the other one.

1.3.1.3. Behavioral Segmentation

This segmentation considers the behavior of the customer rather than their age, gender, etc. This segment the customers based on their purchasing behavior and previous interactions with the brand. This helps companies figure out the brand choice probability and loyalty of the customers.

1.3.1.4. Psychographic Segmentation

This segmentation answers the question WHY? Instead of how and what. More focus is on the market instead of customer traits. This segmentation helps to figure out the actual need of people by analyzing factors like people's lifestyles and hobbies.

1.3.2. TARGETING

The next step after getting the segments is targeting. Business analysts look at the most profitable segment of customers and the ones with high potential value of becoming loyal customers to the company. They filter out the customers that are worth pursuing. While targeting, companies must consider a variety of factors:

1.3.2.1. Profit Maximization

Companies must keep in account the CAC (Customer Acquisition Cost) while marketing. They should be able to generate more revenue from the newly acquired customers than the CAC to stay profitable.

1.3.2.2. Reachability

Companies must target those segments to whom their marketing strategies are reachable and are unhindered by factors like lack of technological access.

1.3.2.3. Size and Difference

The segments targeted must be large enough to spend a company's resources on. If the segments are too small, the returns generated by acquiring them may not justify the CAC, let alone generate profit. Moreover, any two segments should differ by a noticeable amount, lack of it results in needless effort duplication.

1.3.3. POSITIONING

Once the segmentation and targeting are complete, companies have in hand the segment they want to pursue but how to do it is achieved through positioning. They must be able to position their goods or services in the market in such a way that it appeals to the target segment. Positioning the product in a proper manner is vital as this step decides how customers will view the offered product and decide if are they better off buying this product over its competitors or not.

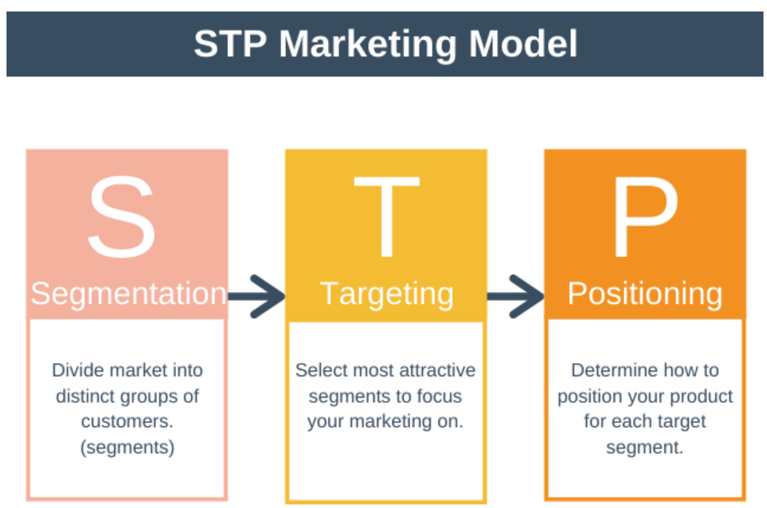

FIG.1

STP MARKETING MODEL

CHAPTER 2

HIERARCHICAL CLUSTERING

2. HIERARCHICAL CLUSTERING

2.1.SEGMENTATION DATASET: OVERVIEW

The segmentation dataset used for modelling purpose in this project consists of two thousand entries from different individuals who made a purchase from an FMCG store. All the categorical variables have been assigned to numerical values and dataset has been standardized using the StandardScaler () method. Outliers and missing values have been removed to improve accuracy of the segmentation result. The dataset has eight columns: ID, Sex, Marital Status, Age, Education, Income, Occupation and Settlement size. These demographics help provide a proper picture of the individual's background. The dataset can be reviewed through visual data representation techniques like the Scatterplot and heatmaps. Scatterplot can be plotted between any two of the demographics for all entries to see their correlation.

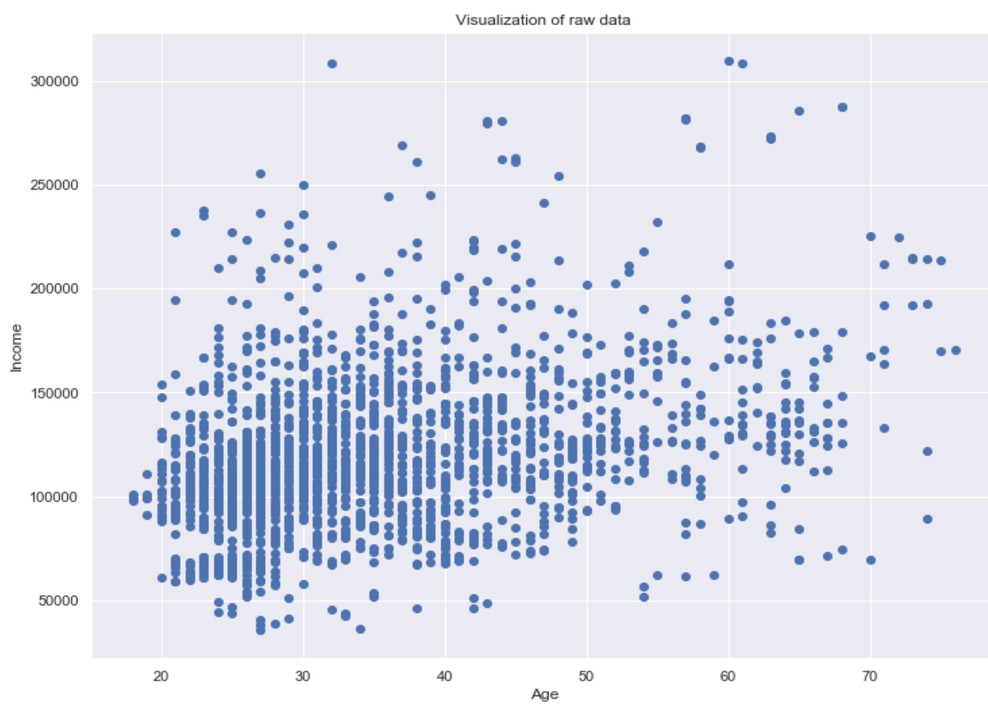

FIG.2

SCATTERPLOT: AGE VS INCOME

2.2.HIERARCHICAL CLUSTERING IMPLEMENTATION

Hierarchical clustering is an unsupervised machine learning technique that groups together data points that are like each other in so called clusters. It works on the principle that each data point in a cluster is as similar as possible to each other while data points in different clusters are as different as possible. Since the data unlabeled and there is no need of a target variable, this model does not have to be trained. There are different techniques available to visualize hierarchical cluster results of which dendrograms and linkage matrix is the most used. Dendrograms come in handy to find the number of clusters which is useful when feeding the input to flat clustering solutions like K-Means.

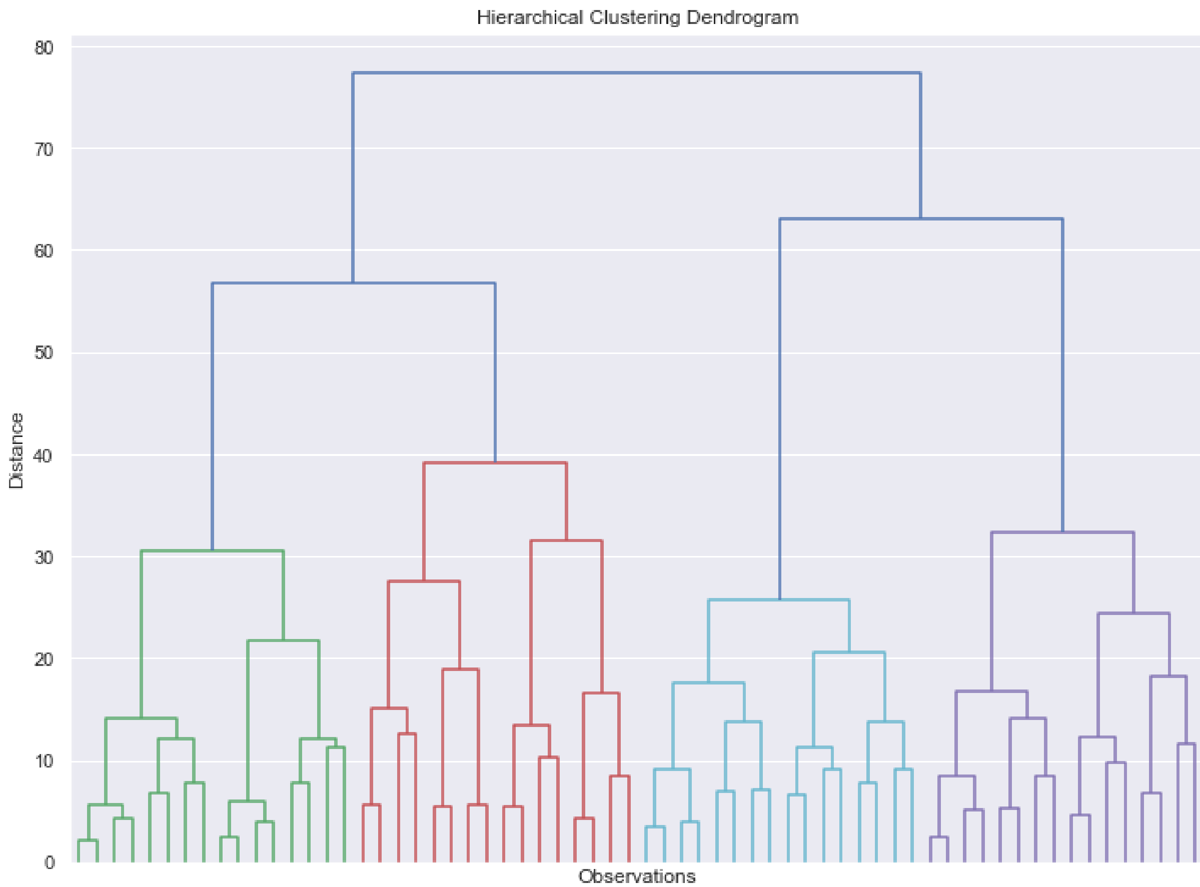

FIG.3
DENDROGRAM RESULT

The results of the hierarchical clustering are as shown above. The linkage matrix returns the solution as four clusters (highlighted in distinct colors) to be the ideal number of cluster solution for the given dataset. Truncate method has been used to simplify the dendrogram.

CHAPTER 3

K-MEANS CLUSTERING

3. K-MEANS CLUSTERING

3.1.K-MEANS OVERVIEW

K-Means clustering is a very commonly used method for segmenting the dataset for further analysis. It is an iterative algorithm that reassigns all datapoints to the nearest cluster whose centroid updates after every iteration. It runs till no more update to the centroid's position is possible. K-Means too is an unsupervised learning algorithm and comes under flat clustering techniques.

K-Means requires to feed the number of clusters as an input before implementing it. The ideal number of clusters is found using the Elbow Method. K-Means is run for a range of different clusters then the WCSS (Within Cluster Sum of Squares) v/s Number of clusters is plotted and compared to see where the steepest drop in WCSS is.

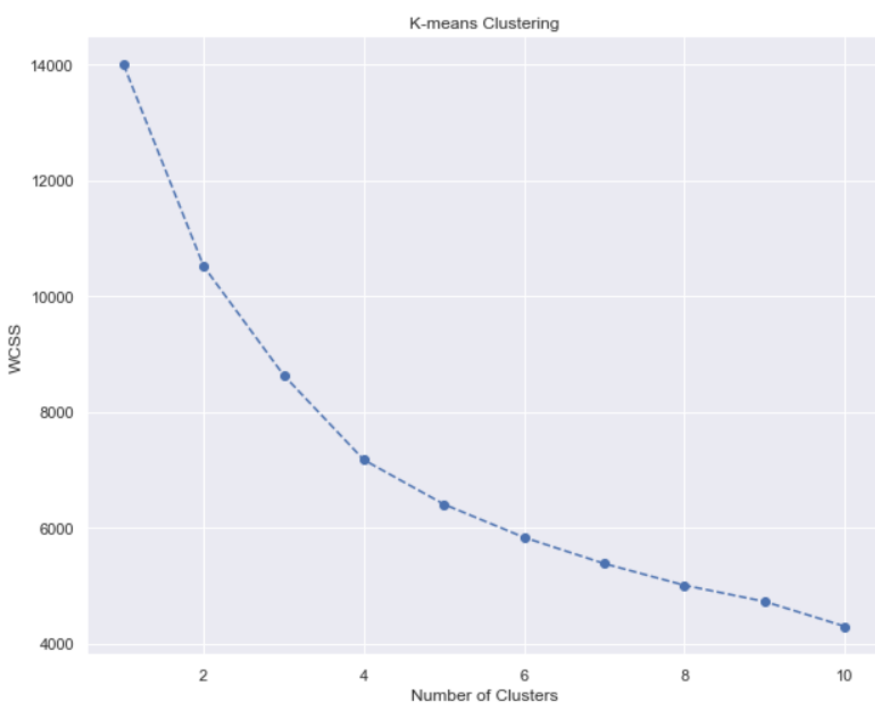

FIG.4

WCSS vs CLUSTERS

The above plot shows the elbow of the graph is at n=4 clusters. It can be verified by the results obtained from the hierarchical clustering solution which gave four as the number of clusters as well, hence the method checks out. K-Means is now run by inputting four as the number of clusters using the fit () method after creating an instance of the K-Means.

3.2.K-MEANS RESULT

Once implemented, the K-Means clustering results are interpreted. For each cluster, the datapoints are interpreted by using groupby (). mean (). This returns the mean values of all demographics in each cluster. This step gives an idea of what kind of people are in each cluster. In K-Means, since there is no target variable, it is up to us to analyze each segment separately and give them labels accordingly.

The first segment consists of almost equal number of men and women of which mostly are older people with good educational background and higher incomes; hence it is appropriate to rename this segment as 'Well-Off.' The second segment consists of people with poor educational background, lesser incomes, and smaller settlement sizes. This segment can be renamed as 'Fewer Opportunities.' The third segment is the segment of all average values so it can be renamed as 'Standard.' The last segment consists of mostly unmarried males with a high income and decent job. This segment can be renamed as 'Career Focused.'

[N Obs] and [Prop Obs] are two added columns to the data frame which shows what proportion of the total values does each segment consist of.

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

FIG.5

SEGMENT SUMMARY

To better visualize the data, the original dataset (unstandardized) is plotted again between the Age and Income variables, but this time each datapoint will belong to its respective cluster. The colors of each label are shown along with the graph. The well-off segment datapoints are well distinguished, located towards the right of the graph unlike the rest of the data. The other three segments are not so well separated. This problem can be taken care of by deploying PCA (Principal Component Analysis), which is done in the next section.

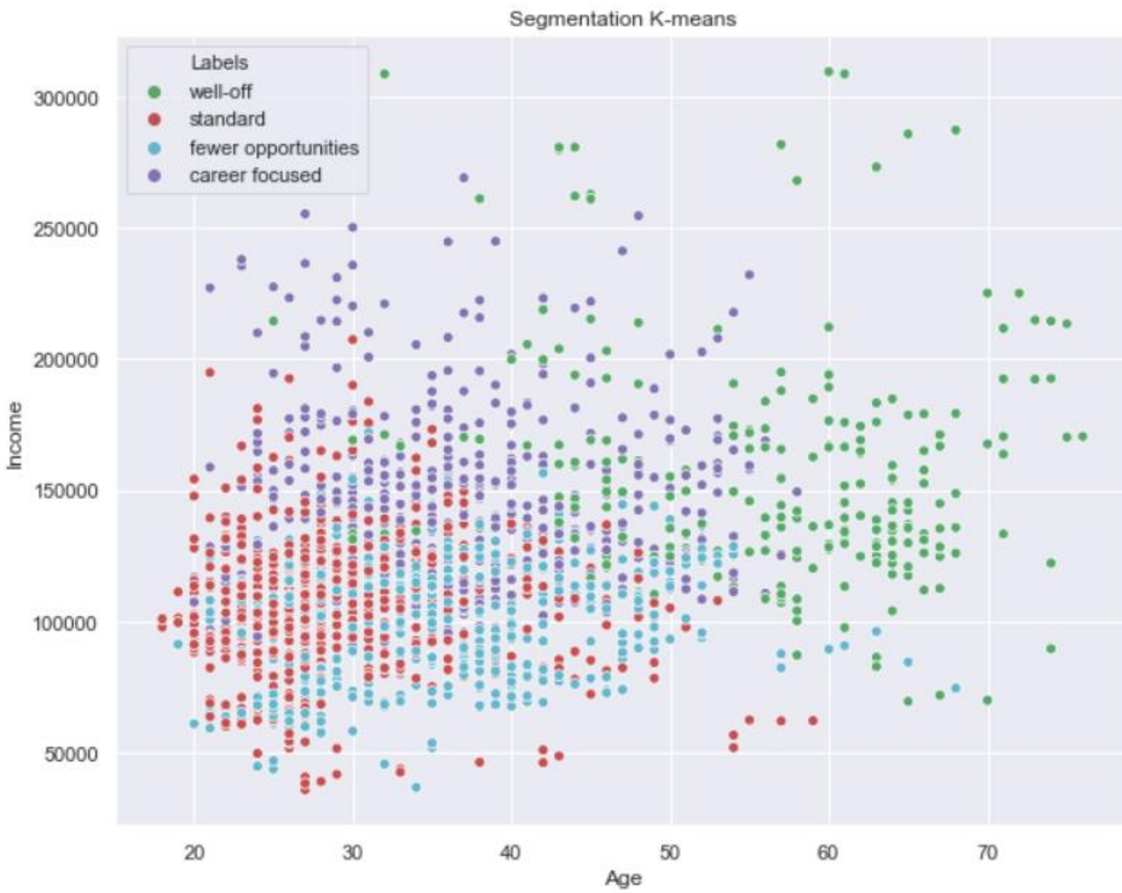

FIG.6

K-MEANS SEGMENTATION VISUALISED USING SCATTERPLOT

3.3.PRINCIPAL COMPONENT ANALYSIS AND K-MEANS

3.3.1. PCA INTRODUCTION AND APPLICATION

PCA stands for Principal Component Analysis. It is a method which is deployed to reduce the dimensions of the dataset. PCA enhances the simplicity and visualization of machine learning algorithms however it comes at the expense of losing some information which impacts the accuracy. So, the tradeoff must be such that, minimum information is lost, and simplicity is maximized by reducing the variables in the dataset. It must be noted that original variables are lost, and new variables are created when PCA is applied.

PCA creates as many components as there are features in the dataset and they are arranged in decreasing order of importance. Importance of a component is judged by how much of the variance in the dataset is explained by that component.

In our segmentation dataset, there are seven components. Thus, after applying PCA we can see how much variance is explained by each of them. It can be visualized by plotting cumulative sum of variance ratios of each component. It is a general practice to opt for a minimum 80% explained variance and the corresponding number of components is chosen as the basis for subsequent modelling.

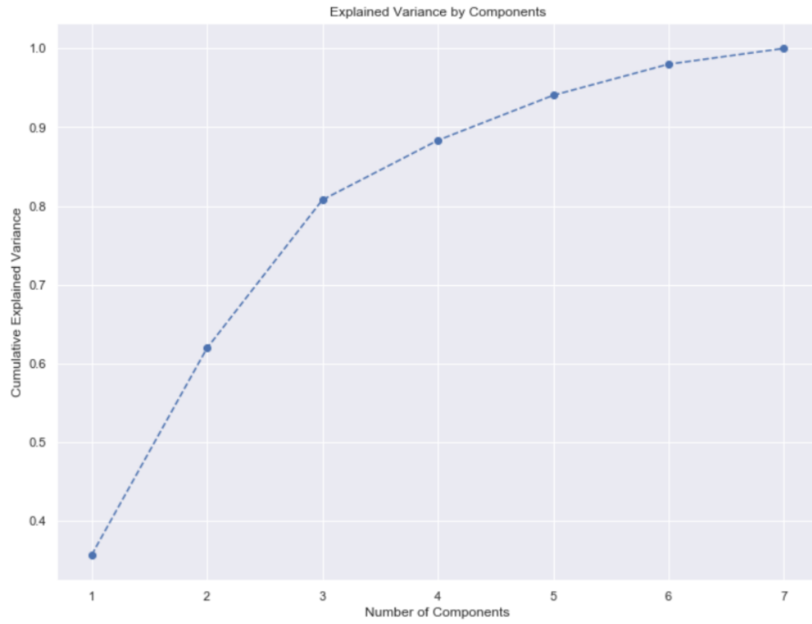

FIG.7
CUMULATIVE EXPLAINED VARIANCES V/S PCA COMPONENTS

It can be seen that the first three components are able to account for about ~80% of the data's variance. This way a 7 feature problem has been reduced to just 3 variable problem. The correlations of these components with each of the original 7 features of the dataset can be visualised using a heatmap. It must be noted that all the datapoints from the segmentation dataset are seven dimensional so they must be transformed to three dimensional such that each observation is described by these three PCA components.

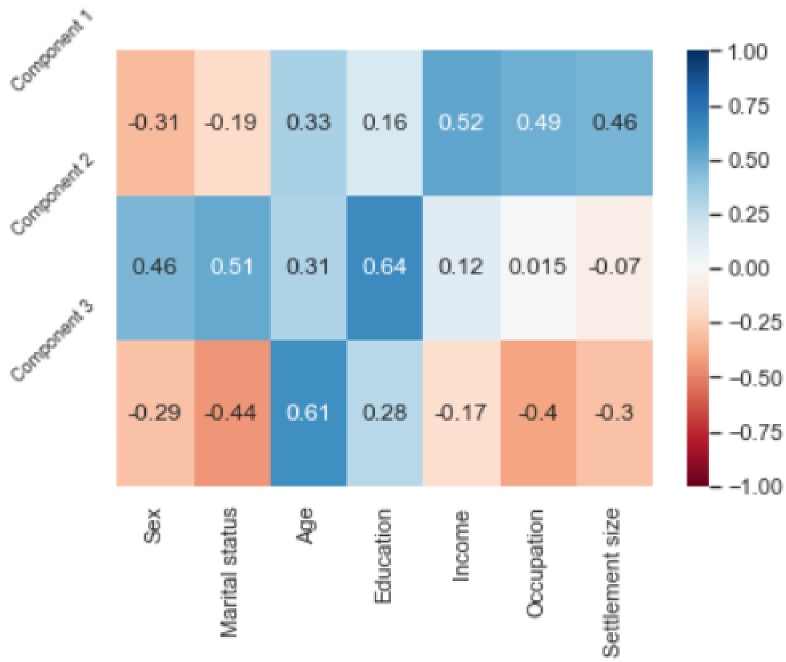

FIG.8

HEATMAP OF CORRELATIONS

HEATMAP INFERENCES:

[1] Component 1 has positive correlation with age,education,income,occupation and settlement size.It can be inferred that Component 1 majorly explains about the career focus of an individual.

[2] Component 2 has sex,marital status and education as its most prominent determinants of variance.This component doesn't explain the career of an individual,rather explains their education and lifestyle.

[3] Component 3 has marital status,age and occupation as major contributors to correlations. However, marital status and occupation are negatively correlated to component nevertheless they are important. This component explains the experience of an individual.

3.3.2. K-MEANS APPLICATION WITH PCA

The data once adjusted as per the new component values is fed into K-Means algorithm. Like before, WCSS v/s No. of clusters is plotted to find the optimal clusters for the algorithm. The elbow of the graph is seen at N=4.

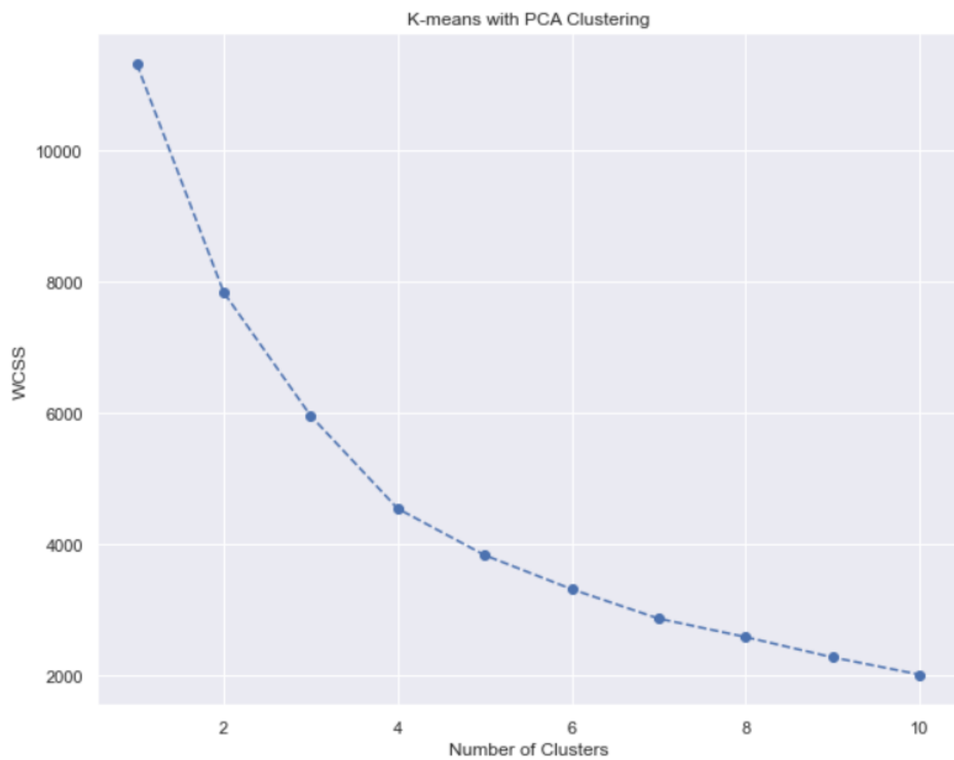

FIG.9
WCSS vs CLUSTERS(PCA)

A new data frame is created with few added attributes. It holds each component's average score for every segment, respectively. Since, segments are like the last clustering result, labels can be retained.

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

FIG.10
PCA SEGMENT SUMMARY

K-Means clustering solution can be visualized by plotting a scatterplot between any of the three components. The differences in all four segments(clusters) are much more pronounced as compared to the traditional K-Means clustering solution. The scatterplot obtained is shown below:

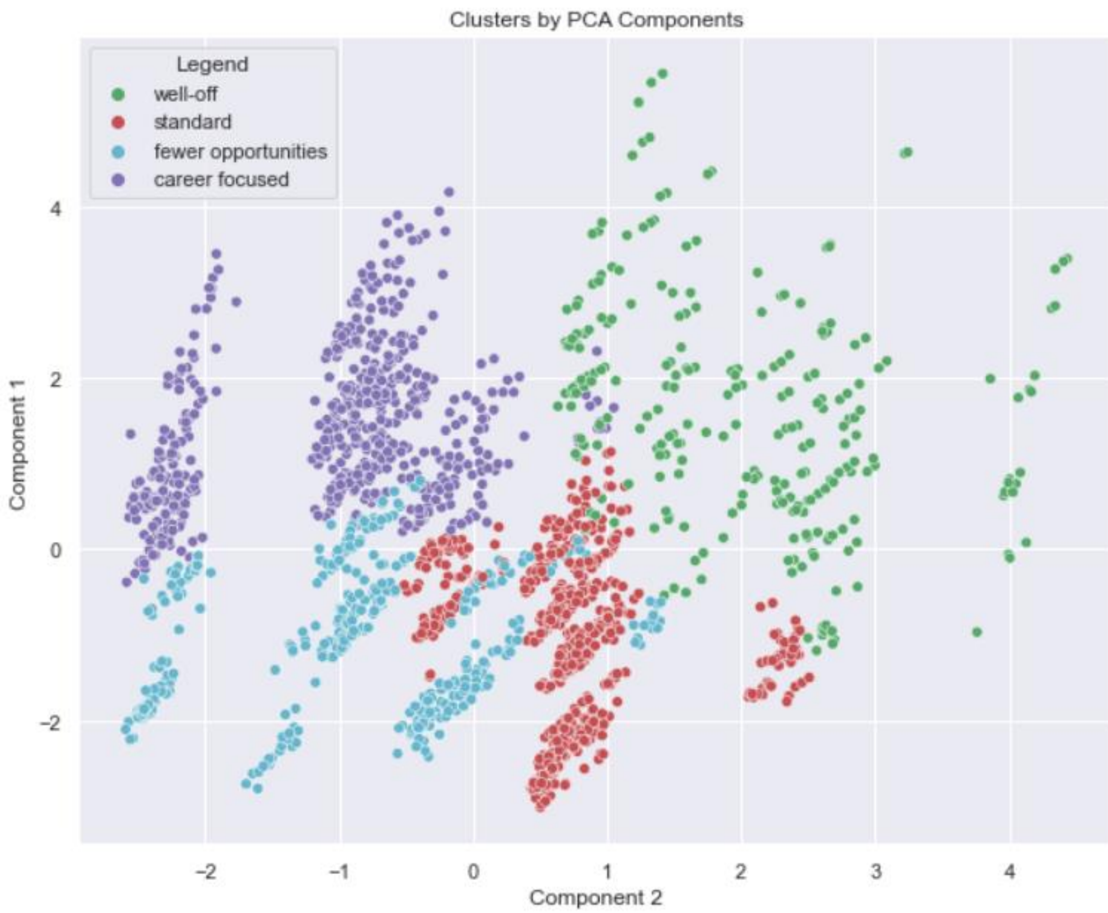

FIG.11
K-MEANS AFTER PCA

CHAPTER 4

PURCHASE ANALYTICS

4. PURCHASE ANALYTICS

4.1. INTRODUCTION

Once segmented, next step of STP framework is Targeting. The marketing mix must be adjusted such that companies can target the most profitable customers. There are three factors that the store would consider calculating:

[1] Purchase Probability

[2] Brand Choice Probability

[3] Purchase Quantity

These incidences can be modelled using machine learning algorithms. Linear and Logistic Regression are used because of their improved interpretability.

4.2. PURCHASE DATASET

The dataset holds transaction records of 500 different individuals for 2 years from the same store where the segmentation data was collected. Several different observations can relate to the same customer as they may have visited the store multiple times. The item considered is chocolate candy bars. The dataset has 24 columns of which 7 are similar to the segmentation dataset. This would help in segmenting the customers in the purchase dataset.

4.3. APPLYING K-MEANS

The new customers are segmented using the already built K-Means with PCA model to group the customers in the four originally obtained clusters. The segmentation model is loaded using pickle package. The dataset must be standardized before loading to the model.

Once applied, analysis of the segmented dataset is carried out by grouping the data based on individuals then based on their segments to gain insights.

4.4. DESCRIPTIVE ANALYSIS BY SEGMENTS

Descriptive analysis of the segmented purchase dataset deals with analyzing the customer shopping habits. Since, it is impossible to analyze each customer separately, they must be grouped into segments as per the K-Means model obtained earlier. For improved interpretation, each transaction is grouped to everyone. For everyone, average purchase incidences are calculated by dividing their total purchase incidences by the times each of them visited the store. A master data frame is developed having the store visits, purchase incidences and predicted segments for all customers.

4.4.1. SEGMENT PROPORTIONS

This measure depicts the proportion of each segment in the purchase dataset. The individuals are grouped to their respective segments for subsequent modelling. Once grouped, segment proportions are displayed for this dataset.

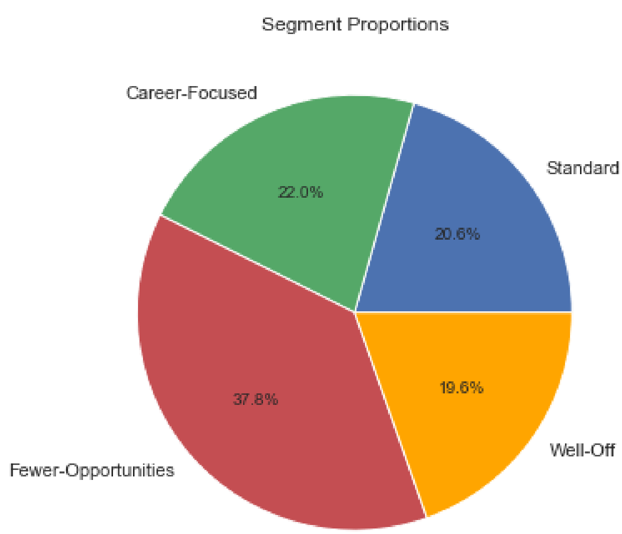

FIG.12

SEGMENT PROPORTIONS

Fewer opportunities form most customers in the dataset, followed by career focused. The number of people in the standard and well-off segment are almost equal, accounting to ~20% of the customers.

4.4.2.PURCHASE OCCASION AND INCIDENCES BY SEGMENTS

The average number of store visits is an effective measure to analyze the engagement of the market with the people irrespective of whether they make a purchase or not. The average number of visits are calculated for each segment.

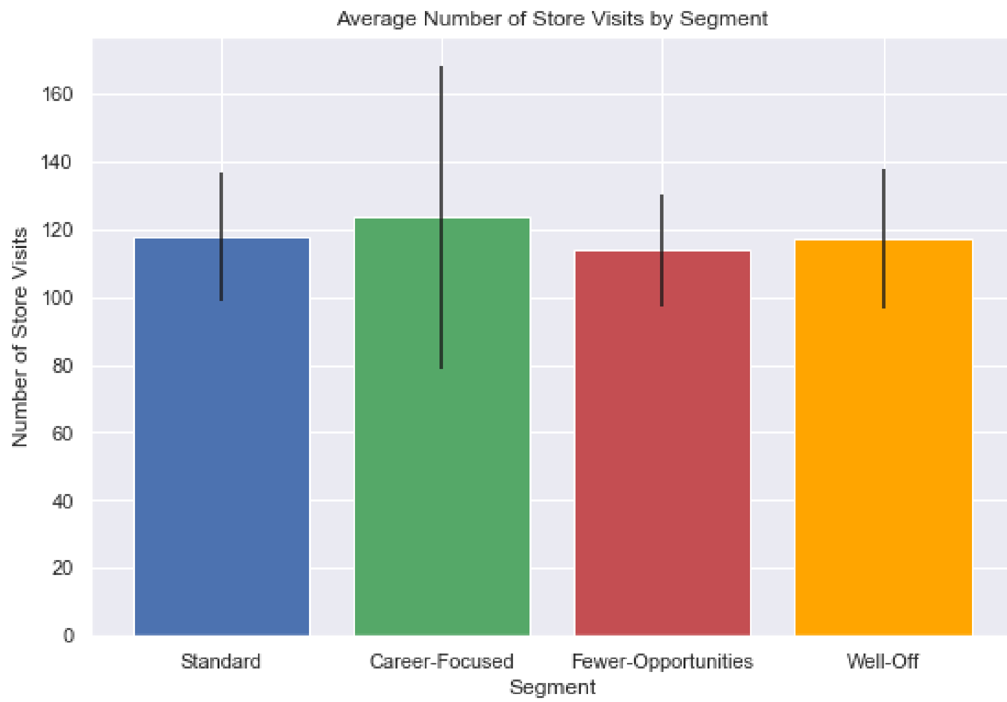

FIG.13

AVERAGE STORE VISITS BY SEGMENT

The results are visualized using bar charts wherein each bar stands for each segment. The height of each bar gives the average number of visits to the store. The vertical line stands for the standard deviation of segment. This depicts homogeneity and the dispersion of the datapoints within their respective segments. The customers from the Fewer-Opportunities segment visited the store least number of times and those from the Career-Focused paid the greatest number of average visits. However, the standard deviation in the Career-Focused segment is quite high implying poor homogeneity when it comes to visiting the store.

Standard, Fewer-opportunities, and Well-off segment are quite similar in terms of their average store visits.

The average number of purchases by each segment depicts how many visits by each segment were fruitful to the store. It holds the information about how profitable each segment was. This visualization can be plotted in a similar fashion.

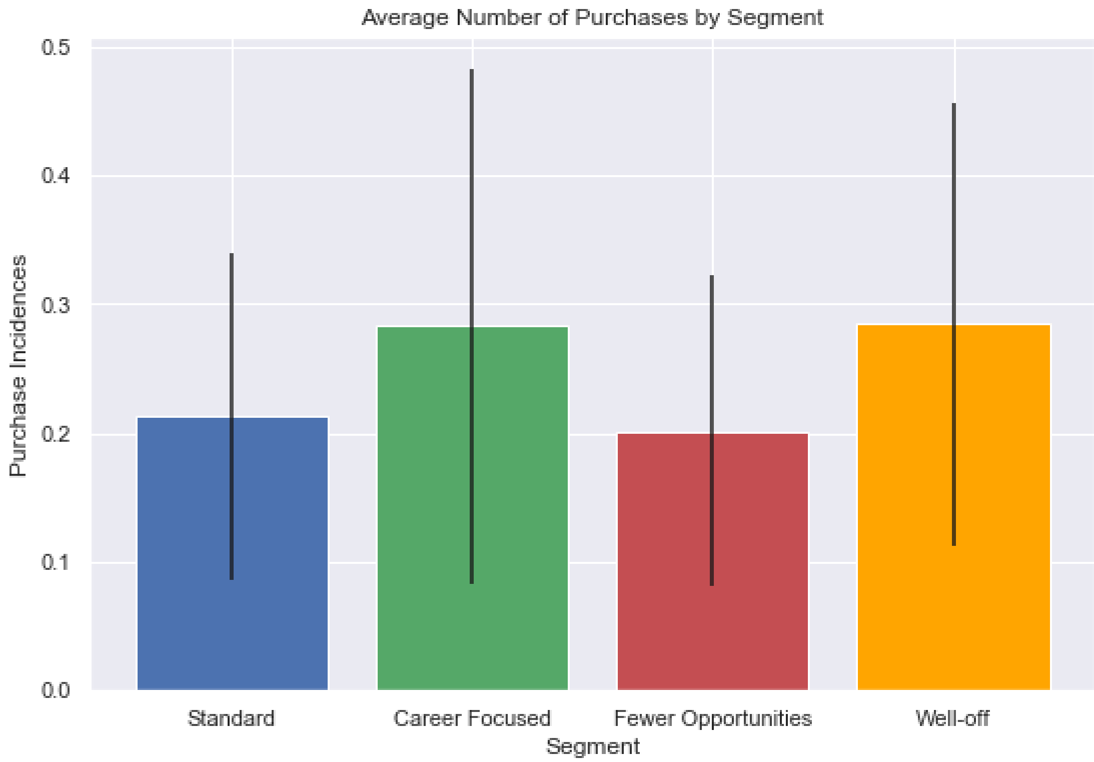

FIG.14
AVERAGE PURCHASE INCIDENCES BY SEGMENT

Career-focused segment and Well-Off segment buys products most often however, there is a large uncertainty in Career-focused segment as implied by the maximum standard deviation. This means that customer in this segment despite having similar incomes, prefer to spend their money differently. Fewer opportunities segment visits the store least however has the most homogenous behavior of customers as implied by the lowest standard deviation.

4.4.3. BRAND CHOICE

This measure aids in analyzing which brand are most popular among the customers. There were 5 brands of chocolate candy bars which were available for purchase. Descriptive analyses of brand choice return the most preferred brand for each segment. This can be visualized by heatmap.

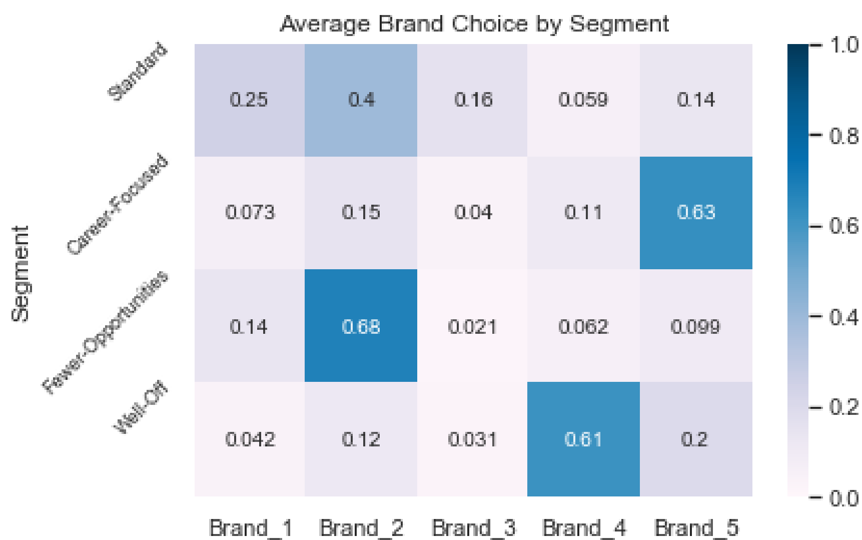

FIG.15
AVERAGE BRAND CHOICE BY SEGMENT

HEATMAP INFERENCES

[1] The brands in the heatmap have been arranged in increasing order of their selling price. Brand_1 is the cheapest while Brand_5 is the most expensive.

[2] Fewer opportunities segment shows a strong preference for Brand_2 with ~70% customers opting for this brand. Since, Brand_2 is not the cheapest one, it indicates that price of the product isn't the only factor customers take into account.

[3] 63% of Career-focused individuals opted for Brand_5, the most expensive one. It can be inferred that customers of this segment are looking for luxury status. This indicates that there is an opportunity to increase the price of Brand_5 chocolates even more.

[4] 61% of Well-off segment customers opt for Brand_4, which is not the most expensive one.

[5] The Standard segment shows heterogeneous brand preference with Brand_2 being the most preferred followed by Brand_1 and Brand_3. This segment is clearly not interested in buying expensive candy bars.

4.4.4. REVENUE DISSECTING

The next performance measure in the exploratory analysis is revenue dissecting. In this, brand revenues generated by each segmented are analyzed one at a time to see which segment brings in the most revenue and for which brand.

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

FIG.16
BRAND DISSECTING

DATAFRAME INFERENCES:

[1] The Career-focused segment brings in the most total revenue as compared to all the other segments. Standard segment is the least profitable in terms of total revenue.

[2] Despite being the second-largest segment, Career-focused brings the highest revenue. This finding is in line with the previous analysis that this segment buys the most expensive brand as well (Brand 5).

[3] The standard segment is almost same in size to the Career-focused but brings in just a little more than half total revenue as compared to the latter.

[4] The well-off and Fewer opportunities segments, spend almost same despite the latter being twice the size of former.

CHAPTER 5

PREDICTIVE ANALYTICS

5. PREDICTIVE ANALYTICS

5.1. PURCHASE INCIDENCE MODELLING

Modelling purchase incidence means calculating the probability of a customer making a purchase on his/her visit to the market irrespective of brand choice. Different statistical models can be used to calculate this probability, here Logistic regression has been used. In training the model to fit the chosen dataset, dependent and independent variables must be declared properly. Here, the dependent variable is incidence of purchase, and the predictor is set as the mean price of product. The result obtained in the coef_ method is array ([[-2.3558334]]). It is negative, signaling that with increase in price, the purchase probability decreases.

5.1.1. PRICE ELASTICITY OF PURCHASE PROBABILITY

Price elasticity of purchase probability means the observed change in purchase probability with a 1% change in price of any commodity. The price elasticity can be calculated by making use of the following formula:

$$E = (1 - Y) * \beta_1 * P$$

Where E: Elasticity, Y: Purchase Probability, β1: Logistic regression coefficient of price,
P: Price.

The price elasticities are calculated here for a range of 0.5 to 3.5 with a step of 0.01. The behavior of purchase probability is analyzed with a change in price for each step. To better visualize this result, price range and price elasticity are plotted.

FIG.17

PRICE V/S ELASTICITY

PLOT INFERENCE:

[1] The price elasticity falls as price increases, indicating that less purchases are made as price of the product increases.

[2] The rate of descent is slow in price range 0.5-1.1 but becomes steeper after the 1.1 mark.

[3] All price elasticities are negative because there is inverse relationship between price and purchase probability.

[4] When Price elasticity< |1| it implies inelastic and otherwise elastic.

MARKETING RECOMMENDATION:

[1] Prices at which elasticities are inelastic can be increased gradually since the effect on purchase probability is not that significant.

[2] If elasticities are greater than 1, it implies a change of more than 100% fall in purchase probability on increasing the price. It is recommended to reduce the prices at these values.

[3] Since the elasticities show a gradual increasing trend in the graph, it can be observed that price elasticity becomes elastic from inelastic at ~1.25 price point where the elasticity is closest to 1.

[4] So, for prices less than 1.25, the pricing point can be increased without compromising too much on purchase probability. This would bring in more revenue per item sold.

[5] For prices more than 1.25, it is recommended to reduce the prices to improve purchase probability.

5.1.2. PRICE ELASTICITY BY SEGMENT

Once the data for an average customer has been plotted for analyzing price elasticity trends, next step is to extend the research to segments. The price elasticities are plotted for all segments on the same graph for better comparison. The average customer is represented with grey, well-off segment in yellow, career-focused in green, standard in blue and fewer opportunities in red.

FIG.18

PRICE ELASTICITIES BY SEGMENT

INFERENCES:

[1] The career-focused segment is displayed in green. This segment is less elastic as compared to the average price elasticity, displayed in grey.

[2] The Fewer-opportunities segment is displayed in red. This segment is more price sensitive as compared to all the others. The slope of this segment falls the steepest of all implying that with an increase in price the elasticities become increasingly elastic and much faster as compared to other segments.

[3] The price elasticities for the Standard segment displayed in blue seem to differ across price range. This may be since the standard segment is least homogenous, which was discovered during the descriptive analysis.

5.2. PURCHASE QUANTITY MODELLING

Calculating purchase probability is not much useful as it doesn’t give an idea about the quantity of product the customer will purchase. Modelling Purchase Quantity gives retail stores an idea which brand is most preferred by their customers, and this can help them manage promotions and inventory accordingly.

In this case, to model and predict the purchase quantity, the Product price and Promotion incidence have been assumed as independent variables. On fitting the data to the Linear Regression model, the coefficients obtained are both negative. This implies that quantity of purchase falls with increase in price as well as if there is a promotion.

5.2.1. PRICE ELASTICITY ESTIMATION

Price elasticity estimation measures the change in quantity purchased with a unit change in price of a commodity assuming all other factors remain unchanged. Purchase quantity elasticity in Linear Regression model is calculated as:

$$E = beta * \frac{Price}{Quantity(purchase)}$$

Where, Beta: Coeffecient obtained in Linear Regression, Price is the price of preferred brand on promotion and Quantity is the predicted quantity at a particular price point.

The result has been plotted for both cases, once assuming there is a promotion on brands and once when there is no promotion.

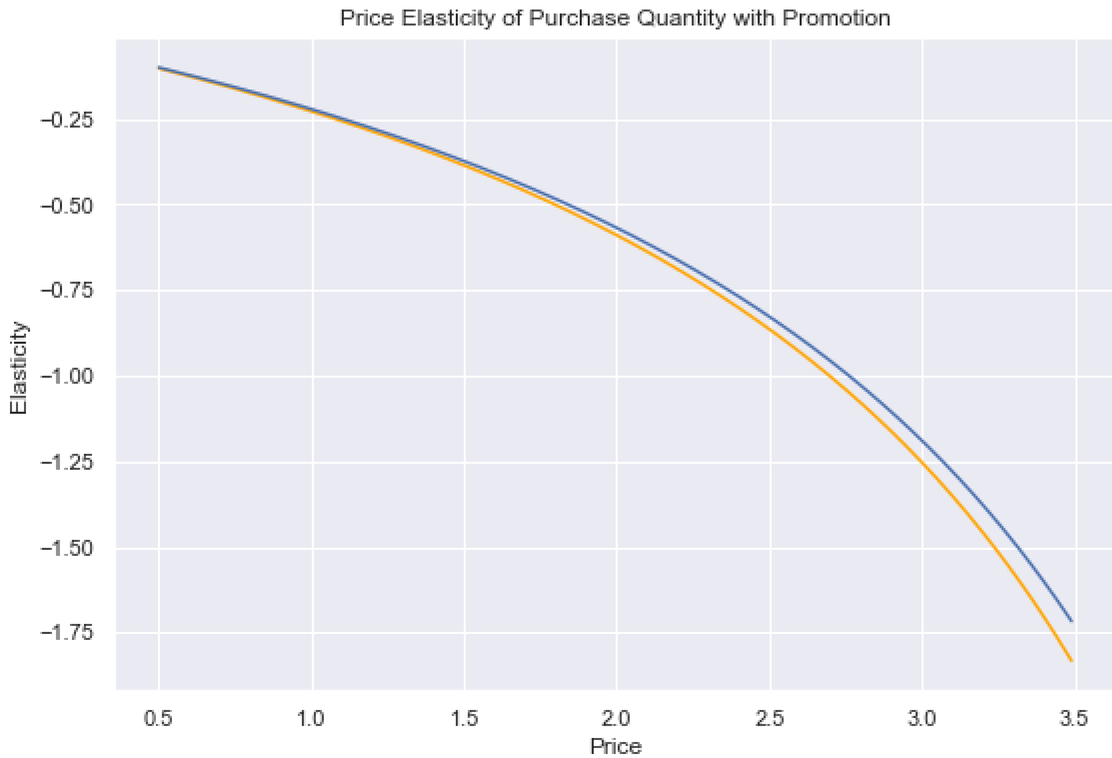

FIG.19

PRICE ELASTICITIES OF PURCHASE QUANTITY

The line in orange is the case when there is a promotion and blue is when there is no promotion. It can be observed that customers are a bit more elastic when there is a promotion.Since these 2 lines overlap pretty much, it can be concluded that price and promotion don’t hold a high predictive power in modelling purchase quantity.

CHAPTER 6

DEEP LEARNING IN CUSTOMER ANALYTICS

6. DEEP LEARNING IN CUSTOMER ANALYTICS

6.1. DEEP LEARNING FOR CONVERSION PREDICTION

Deep Learning is at the frontier of Data science in modern times. Deep learning gives the capability to develop black box models where we are only interested in the input and output and need not worry about the complexity of neural networks that helped reach a particular prediction. These black box models using Artificial Neural Networks (ANNs) can provide a framework for predicting if a customer will convert to a recurring visiting one or not.

6.2. BUSINESS CASE STUDY

The dataset consists of ~15000 records of purchase history of an audiobook company covering 2 years of engagement on the platform. Each entry corresponds to a customer transaction. It provides information about various metrics about the customer behavior while purchasing audiobooks. The average book length and overall book length are self-explanatory. The price indicates the cost customer paid. Review is a Boolean feature; it is 1 if customer left a review and 0 if they didn't. It is being assumed that customers leaving a review are likely to convert again. Review values on a scale of 10 are recorded as well. To make the dataset consistent, missing values here have been replaced with the average customer review. Minutes listened and completion are features indicating engagement on the application. Support requests is an integer indicating the number of times a customer faced an issue on the application. There is a recency variable as well to mark the difference between last visit to the app and purchase date. Since, this is a supervised technique, target variable is needed.

The target variable is a Boolean indicating to 1 if a customer converted and 0 if they didn't convert. By converting, it is implied that the customer visited the store again within the next 6 months of the time considered here which is 2 years.

6.3. TACKLING BUSINESS CASE

The real-world data is unclean and must be preprocessed before feeding into an algorithm. It starts by balancing the dataset, dividing the dataset into test, train and validation and finally save the data in a tensor friendly format to be fed into the machine learning algorithm.

6.3.1.PREPROCESSING DATA

While preprocessing the data, balancing of the dataset is integral to avoid the possibility of the machine learning from false patterns in data. To balance the dataset, equal priors are required which is achieved by making the number of 1s and 0s equal in the target variable and removing the entries causing an imbalance from the dataset. The dataset has been divided into 2 parts, one containing all features that hold a predictive power which form the input of the model and the other one contains the target value column which forms the output of the model. The inputs are standardized using the StandarScaler () module. The data entries must be shuffled as well before batching to improve the accuracy of the model. The data has been divided into 3 sets with the conventional proportions of 80-10-10 forming the train, validation and test data respectively. Each of them has ~50% priors as were intended in balancing the dataset.

6.3.2.DEEP LEARNING MODEL

The Deep learning model developed here consists of 10 units in the input layer that correspond to the 10 features in the dataset. The output layer consists of 2 units, each indicating 0 or 1 as the output of the target variable. The number of hidden layers has been chosen as 2 and number of units in each layer have been set to 50.

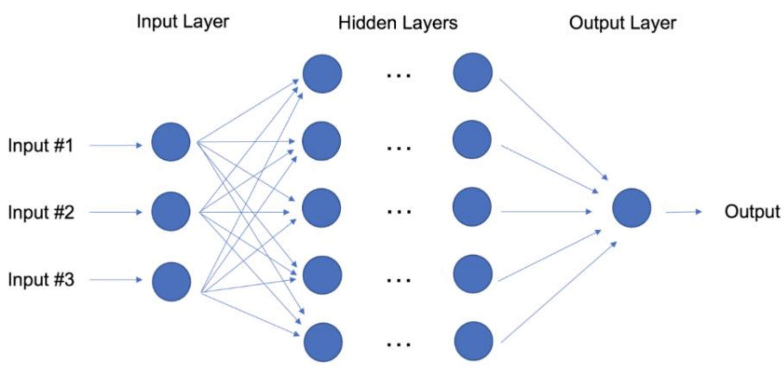

FIG.20

NEURAL NETWORK

6.3.3. MODEL TRAINING, VALIDATION AND TESTING

The Model is trained and validated using the previously trisected data. The validation accuracy achieved is ~91.28%. The model is then deployed to make predictions on unseen data of test inputs. The predicted values are then compared with the actual target variables. The test accuracy achieved is ~90.85%. This accuracy is slightly less than validation as the data may have overfit the training data. This result implies that the model can correctly predict about customer conversion in ~9 out of 10 cases by analyzing the same input features. The values for these predictions depict the probability of customer conversion between 0 and 1.

The Model can be saved as an HDF (Hierarchical Data Format) file which can be deployed to predict on new data without any target variables since the result of the future is not known.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7. CONCLUSION AND FUTURE WORK

7.1.CONCLUSION

The report concludes that application of sophisticated machine learning and deep learning techniques improve a company's understanding of their customer behavior. Understanding customer behavior is key, as it allows companies to position themselves in a way that improves CRM and generate more revenue. Using unsupervised machine learning techniques like K-Means algorithms with PCA provides segments of customers from which company may target only those which seem retainable. Traditional data representation in form of pie charts, bar charts and heatmaps form the foundation of descriptive analysis which analyzed revenue, brand choice, store visits and purchase incidences by segments. Logistic and linear regression are fundamental machine learning techniques that have been used in this report for the exploratory analysis as part of the predictive analytics section. These algorithms provided an insight as to how the purchase probability and purchase quantity changes with successive increase or decrease in prices of goods and services by modelling price elasticities. Marketing recommendations have been stated in the report where it has been found that prices can be increased without much trade-off in loosing purchase probability. Brand promotions, loyalty programs and other techniques can be used to rope in customers on the brink of churning. Feedforward Neural Networks (FNNs) have been deployed as part of the customer conversion prediction section of the paper. These black box models are extremely powerful in making accurate predictions about customer conversion. If companies can accurately predict if a particular customer will convert or not, they can save on massive amounts of time and money by targeting the ones with higher probability of converting rather than trying to retain an already dissatisfied customer.

7.2.SCOPE FOR FUTURE WORK

The report provides a foundation for further research in customer analytics which has an endless scope in trying to extract as much information as possible from the data collected by companies. The algorithms used here can be optimized for better results or using different machine learning techniques like random forest, Naïve Bayes classifier to make model elasticities. The deep learning model makes use of FNNs here, they can be fine-tuned by manipulating the number of hidden layers, using different optimizers, using more input as features. Other deep learning models like CNNs can be deployed to make conversion predictions if the available data is in form of images.

REFERENCES

[1] Singh, N., Singh, P., Singh, K.K. and Singh, A. (2020) 'Machine learning based classification and segmentation techniques for CRM: a customer analytics', Int. J. Business Forecasting and Marketing Intelligence, Vol. 6, No. 2, pp.99–117.

[2] Adebiyi, S.O., Oyatoye, E.O. ad Amole, B.B. (2016) 'Improved customer churn and retention decision management using operations research approach', EMAJ: Emerging Markets Journal, Vol. 6, No. 2, pp.12–21.

[3] Mitra, P.K. and Banga, C. (2019) 'Predicting Indian basket crude prices through machine learning models-a comparative approach', International Journal of Business Forecasting and Marketing Intelligence, Vol. 5, No. 3, pp.249–266.

[4] Claudio Marcus, (1998),"A practical yet meaningful approach to customer segmentation", Journal of Consumer Marketing, Vol. 15 Iss 5 pp. 494 – 504

[5] Ahmad, A.K., Jafar, A. and Aljoumaa, K. (2019) 'Customer churn prediction in telecom using machine learning in big data platform', Journal of Big Data, Vol. 6, No. 1, p.28.

[6] Fares, N., Lebbar, M. and Sbihi, N. (2019) 'A customer profiling' machine learning approach', for in-store sales in fast fashion', in Ezziyyani, M. (Eds.): Advanced Intelligent Systems for Sustainable Development (AI2SD'2018). AI2SD 2018. Advances in Intelligent Systems and Computing, Vol. 915, pp.586–591, Springer.

[7] Chubing Zhang, Xinchun Wang, Annie Peng Cui, Shenghao Han, Linking big data analytical intelligence to customer relationship management performance, Industrial Marketing Management, Volume 91,2020, Pages 483-494, ISSN 0019-8501.

[8] Jose Ramon Saura, Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics, Journal of Innovation & Knowledge, Volume 6, Issue 2,2021, Pages 92-102, ISSN 2444-569X.

[9] Nikolay Archak, Anindya Ghose, Panagiotis G. Ipeirotis, (2011) Deriving the Pricing Power of Product Features by Mining Consumer Reviews. Management Science 57(8):1485-1509.

[10] Customer segmentation in e-commerce: Applications to the cashback business model, Journal of Business Research, 88 (2018), pp. 407-414.

[11] Shaokun Fan, Raymond Y.K. Lau, J. Leon Zhao, Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix, Big Data Research, Volume 2, Issue 1,2015, Pages 28-32, ISSN 2214-5796.

[12] Kelleher et al., 2015, J.D. Kelleher, B. Mac Name, A. D'Arcy Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies, MIT Press, Cambridge, Massachusetts (2015)

[13] Peker, Serhat, Altan Kocyigit, and P. Erhan Eren. "LRFMP model for customer segmentation in the grocery retail industry: a case study." *Marketing Intelligence & Planning* (2017).

[14] T. Femina Bahari and M. Sudheep Elayidom / Procedia Computer Science 46 (2015) 725 – 731, An Efficient CRM-Data Mining Framework for the Prediction of Customer Behavior

[15] Dullaghan, Cormac, and Eleni Rozaki. "Integration of machine learning techniques to evaluate dynamic customer segmentation analysis for mobile customers."

[16] Wedel, M. and Kannan, P.K. (2016) 'Marketing analytics for data-rich environment', Journal of Marketing, Vol. 80, No. 6, pp.97–121.

[17] Liu, C.C., Chu, S.W., Chan, Y.K., Yu, S.S., 2014. A Modified K-Means Algorithm – TwoLayer K-Means Algorithm. In: 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, pp. 447–450. doi: 10.1109/IIH-MSP.2014.118.

[18] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: a literature relationship and classification," Expert Systems with Applications, vol. 36, pp. 2529–2602, 2009.