ABSTRACT

Identifying valuable customers as well as retaining them has become a key component for any business to succeed in this competitive market. Customer retention is one among the foremost important issues for companies. Companies always seek to scale back customer churn so as to extend the customer lifetime value and reduce the value of acquisition of latest customers. By specializing in customer churn prediction and identification, companies can predict beforehand which customers are getting to churn and thus decrease customers churn rate through related personalized actions. The key issue here is the way to predict customer churn at an early stage. This project identifies related issues in customer relationship management and provides new definitions and strategies.

This project establishes a framework which consists of five main stages- Customer Segmentation, Customer Lifetime Value Prediction, Churn Prediction, Predicting Next Purchase Day, Predicting Sales. During these stages exploratory data analysis will be performed and various machine learning models will be applied.

These outcomes can be used for customized or personalized product and service developments, to improve customer service efficiency and related decision-making more effectively and more particularly enabling strategic promotion campaigns to customers.

CONTENTS

ABSTRACT i

LIST OF FIGURES iv

ABBREVIATION vi

1. INTRODUCTION 10

2. LITERATURE SURVEY 11

2.1 Survey Existing system 11

2.2. Limitation Existing system or research gap 13

2.3 Problem Statement and Objective 14

2.4 Scope 14

3. PROPOSED SYSTEM 15

3.1 Analysis/Framework/ Algorithm 15

3.2 Technologies Used 20

3.3 Design Details 21

4. IMPLEMENTATION 22

4.1 Motivation and Objective 22

4.2 Dataset Used 22

4.3 Analyze Data and Understand various metrics 23

4.4 Customer Segmentation 27

4.5 Customer Lifetime Value Prediction 28

4.6 Churn Prediction 30

4.7 Predicting Next Purchase Day 30

5. RESULTS 31

6. PROJECT TIMELINE 33

7. CONCLUSION 35

8. FUTURE SCOPE 36

REFERENCES 37

Publication Paper 38

ACKNOWLEDGEMENT 39

LIST OF FIGURES

Figure No.	Figure Title	Page No.
3.1	Flowchart	21
4.1	Dataset Used	23
4.2	Dataset Overview	23
4.3	Monthly Revenue	24
4.4	Monthly Growth Rate	24
4.5	Monthly Active Customers	25
4.6	Monthly Total orders	25
4.7	Average Revenue per Order	25
4.8	New Customer Ratio	26
4.9	Monthly Rate of Retention	26
4.10	Segments Distributed on Scatter Plot	27
4.11	Lifetime Value Vs Overall RFM Score	29
4.12	Actual Vs Predicted Sales	30
5.1	Monthly Revenue Result	31
5.2	Segments Distributed on Scatter Plot	31
5.3	Lifetime Value Vs Overall RFM Score	32
5.4	Actual Vs Predicted Sales	32
6.1	Timeline Table for Implementation Plan	33

ABBREVIATION

Sr. No	Abbreviation	Full form
1	Dlib	Library for Machine Learning
2	USB	Universal Serial Bus
3	SQL	Structured Query Language
4	OpenCV	Open-Source Computer Vision Library
5	API	Application programming interface
6	SDK	Software Development Kit
7	DFD	Data Flow Diagram
8	DB	Database
9	LTV	Lifetime Value
10	RFM	Recency, Frequency and Monetary Value
11	B2C	Business to Consumer

1. INTRODUCTION

In today's day and age control over the customers and having the customers by your side in a B2C company is very important. These small proportions of customers in the company often imply the massive amount of revenue that is generated throughout the year. The revenue that is generated is disproportionately large within the overall revenue. It is many times the understanding of the customer's behavior that often plays a key role in the businesses being able to deliver their goods to the customers in the most efficient way and in a most organized way. More often than not using this knowledge of customers' behavior and in general how customers may be dealing with the prices, products of a particular company can become an advantage and maybe even a very important metric in determining how the sales of the company might be or look like in the near future of such companies.

Also, there is so much vast availability of the information that can be gathered by the normal metrics that the business or even the normal retail stores keep to keep the track of the customers, their orders, number of orders, etc can also become a very key factor in helping the business grow in a particular way that they may be a profitable business at the same time they can also deliver good and services to the costumes in a way that in even all the customers may like.

In this project, we define it as an interdisciplinary field that brings together techniques from data analysis, machine learning, pattern recognition, statistics, data visualization and neural networks.

2. LITERATURE SURVEY

2.1 Survey Existing system

In this part, we discuss a little of the related literature which is appropriate, relevant, and applicable to this work. Consumer acquisition and consumer retention and its prediction were studied over a previous couple of years in nearly all enterprise domains like banking, retail, telecommunications, and so on. With fast modifications within the marketplace, and then strategies are made primarily based on the change in customer behavior keeping in mind the troubles prompted from the past has been a difficulty. a unique statistical strategy and machine learning strategy of various sorts are being used to deal with this problem [4]. Initial work related to customers' acquisition depended on Probit Models and Logit Models [7,8]. Afterward, Giuffrida et al. 2000[9] found out that the multivariate decision tree induction algorithm was proved to have better performance than the logit model in searching for the simplest and most useful customers as targets. In today's time, clustering and other segmentation techniques are used widely to create clusters and segments of the same type of customers using the historical, transactional and personal data. Breto et al 2015 [4] found out data mining techniques such as clustering and few other techniques to discover sub groups to find patterns in the preferences of customers who are related to the fashion industry. Hsu et al 2012 [3] used a clustering method which is hierarchical in order to segment customers based on transactional dataset. It will also consider the similarities between different types of items in the transaction. But, in this project we have adopted a mixed approach of clustering for identification of segments and a different approach for impact of customers.

Also the focus on churning of customers is to determine the customers who might be leaving and checking whether we should retain them or not and is it worth it [5]. Building an accurate customer churn prediction model while keeping it efficient is a crucial problem for both students and industry experts in recent years. Profiling helps a company to take actions for keeping customers who might leave, which helps in churn reduction, because it is observed that keeping old customers is less expensive than acquiring new customers [6]. Reichheld and Sasser [10] work has increased the interest of people in customer’s retention because they reported that a 5% increase in customer’s retention can help company profitability to rise by anywhere between 25% - 85%. Also, it was reported that depending on business domains, retaining already existing customers is around 6 times more cost effective than acquiring new

customers.

Customer’s churn is an important parameter for the corporation because acquiring a replacement customers can cost almost seven times that of retaining an already existing customers. Customer’s churn can be a huge blocker for a rapidly growing company and proper retention strategy should be used so as to avoid the customer’s churn rate [2].

2.2. Limitation Existing system or research gap

It can be grueling to address the issue of client accession and retention because numerous enterprises haven't yet begun to develop client databases that retain the position of complication (e.g., data storages) that's necessary to dissect directly the numerous factors that may impact the client life cycle [8].

Also, during early ages logit and probit models were used [7,8]. The logit model assumes a logistic distribution of crimes, and the probit model assumes normal distributed crimes. These models, still, aren't practical for cases when there are two or more cases, and the probit model isn't easy to estimate (mathematically) for further than 4 to 5 choices.

In the work that was done previously the system to calculate and measure the position of the threat to customers and the customer churn was not that effective. previous research [4], Also raised a few challenges. It had validated that during this sphere, like also in various other fields like advertising, finance, and production. It requires a trial and other strategies which is also time-consuming for fine-tuning and preparing the information of the styles and DM approaches. It has also been proved that dictionary that the close involvement is critical for the field specialist for the achievement of the design.

After the difficulties encountered during the deposition and taking them into account and also the limitations this assist on it some of the additional explorations is also demanded in the resulting areas:

● Finding the highest quality wide variety of clusters and developing approximation tactics for it.

● Defining a distance degree in line with the assessment standards and it truly is specific to the business trouble .

● Testing and applying other group discovery algorithms.

● Remodeling the information to permit the use of different algorithms, evaluating the one of a kind results and figuring out the most efficient bone[ ]

2.3 Problem Statement and Objective

In today's competitive environment, a successful company must provide better customized services that are not only acceptable to customers but satisfy their needs as well, in order to survive and succeed in gaining an advantage against competition. It has been proven by many studies that it is more costly to acquire new customers than to retain old ones. Consequently, evaluating current customers in order to enhance their lifetime value becomes a critical factor to decide the success or failure of a business.

Our objectives are:

1. Customer Segmentation

2. Customer Lifetime Value Prediction

3. Churn Prediction

4. Predicting Next Purchase Day

5. Predicting Sales

2.4 Scope

Customers lifetime value prediction and churn prediction can help to save companies from wasting their money, time and resources.Predicted sales can be utilized for planning. We can plan our demand and supply actions by looking at the forecasts.

It helps to see where to invest more. It is an excellent guide for planning budgets and targets.Consumer's satisfaction can be studied through telephone interviews or questionnaires on segmented customers and can be an important reference for maintaining those customers.

3. PROPOSED SYSTEM

3.1 Analysis/Framework/ Algorithm

3.1.1 Import Data:

This Online Retail II data set contains all the transactions occurring for a UK-based and registered, non-store online retail between 01/12/2009 and 09/12/2011.The company mainly sells unique all-occasion gift-ware. Many customers of the company are wholesalers.

Dataset contains 1,067,371 rows x 8 columns.

Dataset Link:

https://www.kaggle.com/mashlyn/online-retail-ii-uci

3.1.2 Analyze data and understand various metrics:

In this step we will get to know about our data well. We will understand different metrics from this dataset and have a brief understanding of the dataset.

Main aim of this project is growth. We want more customers, more orders, more revenue, more signups, more efficiency. To achieve this understanding our metrics is important.

Firstly, we will find out the North Star Metric. **The North Star Metric** is the single metric that best captures the core value that your product delivers to customers. This metric depends on your company's product, position, targets & more. Airbnb's North Star Metric is nights booked whereas for Facebook, it is daily active users. In our example, we will be using an online retail dataset. For an online retail, we can select our North Star Metric as **Monthly Revenue.**

Equation for it is:

Revenue = Active Customer Count * Order Count * Average Revenue per Order

Other metrics we will be considering are:

- Monthly Revenue Growth Rate
- Monthly Active Customers

- Monthly Order Count
- Average Revenue per Order
- New Customer Ratio
- Monthly Retention Rate
- Cohort based Retention Rate

3.1.3 RFM Clustering:

RFM stands for Recency - Frequency - Monetary Value. As the methodology, we need to calculate Recency, Frequency and Monetary Value (we will call it Revenue from now on) and apply unsupervised machine learning to identify different groups (clusters) for each.

3.1.4 Customer Segmentation:

We have analyzed the major metrics for our online retail business. Now we know what and how to track. Also, you can't treat every customer the same way with the same content, same channel, same importance. They will find another option which understands them better. In this step we will focus on customers and segment them.

Theoretically we will have segments like below:

- Low Value: Customers who are less active than others, not very frequent buyer/visitor and generate very low - zero - maybe negative revenue.
- Mid Value: In the middle of everything. Often using our platform (but not as much as our High Values), fairly frequent and generates moderate revenue.
- High Value: The group we don't want to lose. High Revenue, Frequency and Low Inactivity.

3.1.5 Customer Lifetime Value Prediction:

We segmented our customers and found out who are the best ones. In this step we will measure one of the most important metrics we should closely track: **Customer Lifetime Value.** We invest in customers (acquisition costs, offline ads, promotions, discounts & etc.) to generate revenue and be profitable. These actions make some customers super valuable in terms of lifetime value but there are always some customers who pull down the profitability. We need to identify these behavior patterns, segment customers and act accordingly.

The equation for calculating lifetime value is:

Lifetime Value = Total Gross Revenue - Total Cost

So the path for **Customer Lifetime Value Prediction** is:

- Define an appropriate time frame for Customer Lifetime Value calculation
- Identify the features we are going to use to predict future and create them

- Calculate lifetime value (LTV) for training the machine learning model
- Build and run the machine learning model

3.1.6 Customer Retention Rate:

We know our best customers by segmentation and lifetime value prediction, we should also work on retaining them. That's what makes Retention Rate one of the most critical metrics.

Product-market fit describes a scenario in which a company's target customers are buying, using, and telling others about the company's product in numbers large enough to sustain that product's growth and profitability. Retention Rate is an indication of how good your product market fit (PMF) is.

To calculate customer retention rate (CRR) we can use the following formula involving the customers we have at the start (S), at the end (E) and customers acquired during the period we are measuring (N).

CRR = ((E-N)/S) x 100.

3.1.7 Churn Prediction:

One of the powerful tools to improve Retention Rate and hence the PMF is Churn Prediction. By using this technique, we can easily find out who is likely to churn in the given period. We will follow the following steps to develop a Churn Prediction model:

- Exploratory data analysis
- Feature engineering
- Investigating how the features affect Retention by using Logistic Regression
- Building a classification model with XGBoost

3.1.8 Predicting Next Purchase Day:

Predicting the next purchase day of the customer can provide us with various opportunities. We can build our strategy on top of that and come up with lots of tactical actions

Here we will follow the steps below:

- Data Wrangling (creating previous/next datasets and calculate purchase day differences)
- Feature Engineering
- Selecting a Machine Learning Model

3.1.9 Predicting Sales:

Before this step, almost all our prediction models were on customer level (e.g. churn prediction, next purchase day, etc.). It is useful to zoom out and look at the broader picture as well. This step helps us justify our efforts on the customer side and how it affects the sales. It can be utilized for planning. We can plan our demand and supply actions by looking at the forecasts. It helps to see where to invest more. We will build a deep learning model for predicting sales.

The implementation of our model will have 3 steps:

- Data Wrangling
- Data Transformation to make it stationary and supervised
- Building the LSTM model & evaluation

3.2 Technologies Used

Python

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small- and large-scale projects. As flask is written in python language by default python is the main programming language for our website and also for our machine learning model, eye, mouth and head tracking is also programmed using python as machine learning is simplified in python and easier to integrate with the website.

Libraries Used: Pandas, Numpy, Seaborn, Plotly, Matplotlib, Sklearn, XGBoost, Streamlit

Machine Learning

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

Data Analysis

Data analysis is the process of cleaning, changing, and processing raw data, and extracting actionable, relevant information that helps businesses make informed decisions. The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.

3.3 Design Details

Architecture Diagram:

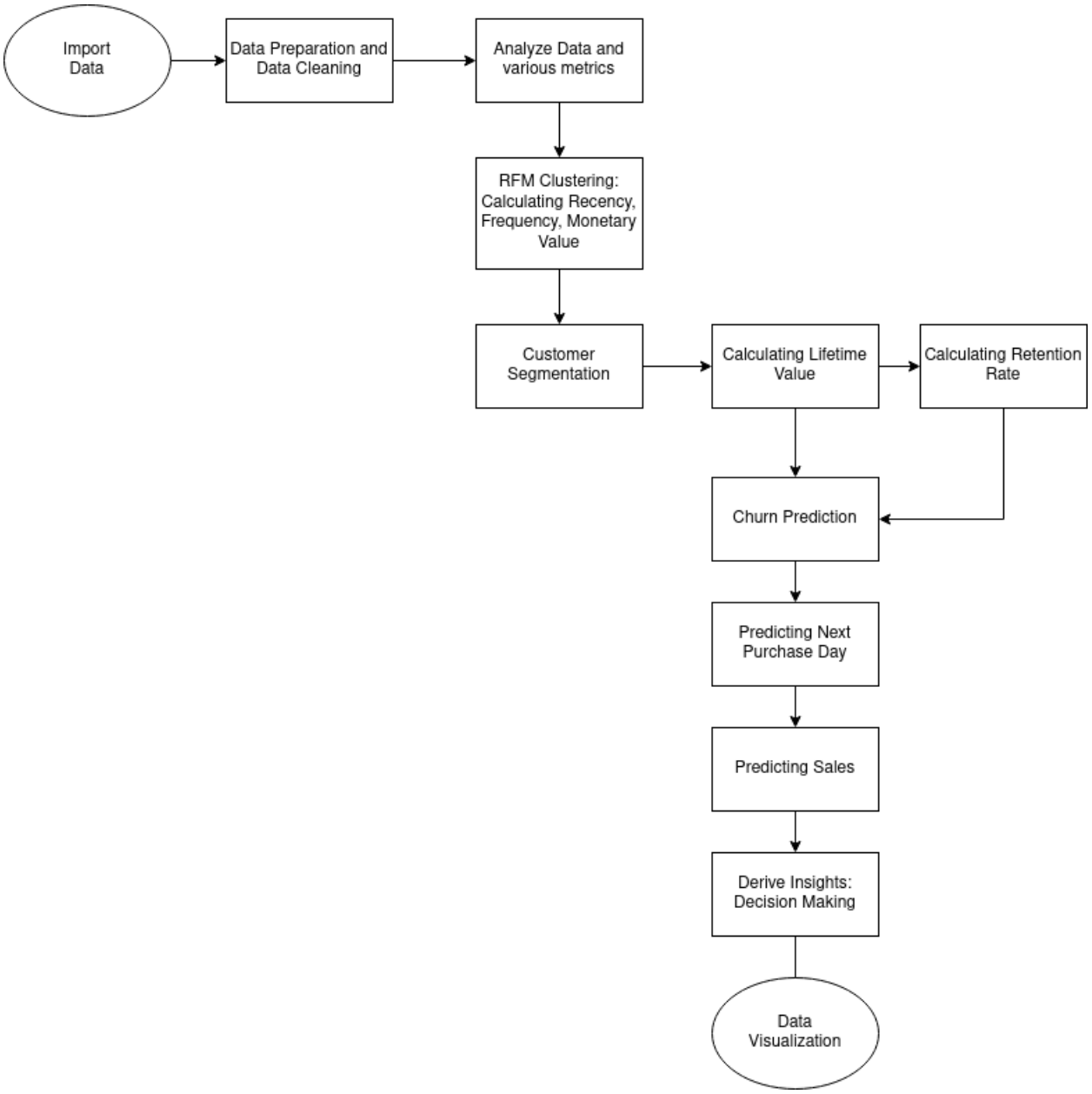


Fig. 3.1 Flowchart

4. IMPLEMENTATION

4.1 Motivation and Objective

To be able to be successful in today's environment, for a business, it is important for them to not only deliver goods and services in a way that is not profitable to them but is also meeting the requirements of the customers as well. Also, many studies have proven that acquiring new customers or customer acquisition is far more costly than that of retaining the already loyal and existing customers.

So as to evaluate the customers to increase their Lifetime value as it decides the success or failure for any company.

Our objectives are :

1. Segmentation of Customer
2. Prediction of Customer Lifetime Value (LTV)
3. Prediction of Churn
4. Prediction of Next Purchase Day
5. Prediction of Sales

4.2 Dataset Used

The dataset that is used for this project is an Online Retail II dataset. The transactions that it contains are from 01/12/2009 to 09/12/2011.

This Dataset contains 1,067,371 rows x 8 columns.

Fig. 4.1 Dataset Used

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
5	489434	22064	PINK DOUGHNUT TRINKET POT	24	2009-12-01 07:45:00	1.65	13085.0	United Kingdom
6	489434	21871	SAVE THE PLANET MUG	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
7	489434	21523	FANCY FONT HOME SWEET HOME DOORMAT	10	2009-12-01 07:45:00	5.95	13085.0	United Kingdom
8	489435	22350	CAT BOWL	12	2009-12-01 07:46:00	2.55	13085.0	United Kingdom
9	489435	22349	DOG BOWL , CHASING BALL DESIGN	12	2009-12-01 07:46:00	3.75	13085.0	United Kingdom

Fig. 4.2 Dataset Overview

4.3 Analyze Data and Understand various metrics

In this step, our main goal or the main objective is to understand the data that we have well. We have various metrics that are available in the dataset, and by trying to understand these metrics, we would be able to get a clear understanding of the data that we are dealing with, what are the various metrics and we could also get a quick understanding of the dataset. The main objective of the project is growth. Every business would like to have these metrics and would like to have more customers, orders, revenue, growth .

We will be using something called as north star metric. A north star metric is a metric used which is the most significant metric for a company to decide its overall growth. A north star metric is one metric of measurement which can be helpful and which can be most predictable in deterministic a company's growth. For a metric to qualify as a north star metric, it has to do three things cause revenue , Reflect the customer value and measure the progress of the business. For the web retail store we have selected or not star metric as the monthly revenue .

We have all the necessary information we need in the dataset, they are: ID of Customer, Quantity, ,Unit Price, Invoice Date.

With all these features, we can get our North Star Metric formula as:

Revenue = (Active Customer Count) * (Average Revenue per Order) * (The Order Count)

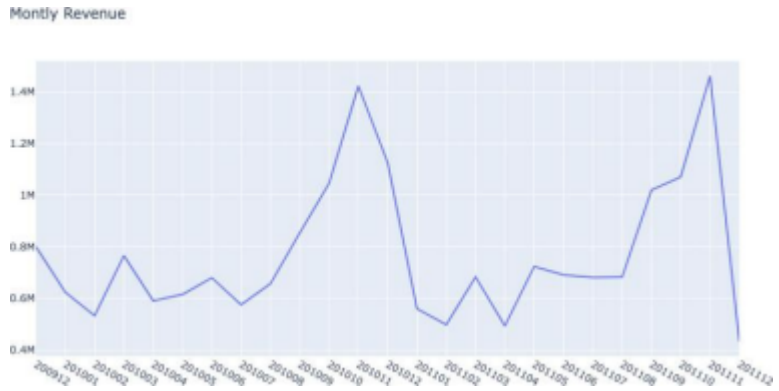


Fig. 4.3 Monthly Revenue

Here we see 36.5% growth in November (December is excluded in the code since it hasn't been completed yet).

This graphic is showing us that from August onwards the revenue is growing. To figure out the reasons for this we consider a few other metrics.

Other metrics to be considered are:

1. Monthly Growth Rate

We have computed the Monthly Revenue Growth Rate to understand the growth rate.

Fig. 4.4 Monthly Growth Rate

2. Monthly Active Customers

We can also find the monthly active customers by counting the unique customer ID that are available in the data set. For example, if we consider that in April the monthly active customer numbers were drawn by around 10% here we are able to see a similar kid of trend for the number order numbers as well.

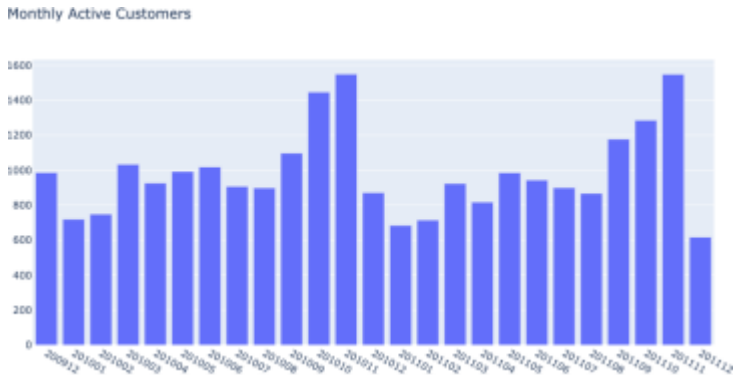


Fig. 4.5 Monthly Active Customers

3. Monthly Order Count

Here we have seen that the number of others have gone down. Now as we can see that the active customer count directly is affecting the order count decrease. hence we can check or average revenue for an order as well.

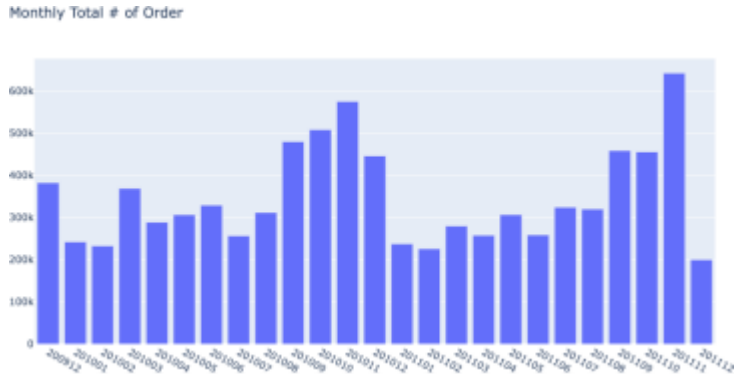


Fig. 4.6 Monthly Total Orders

	InvoiceYearMonth	Revenue
0	200912	16.978696
1	201001	16.794060
2	201002	16.360383
3	201003	17.444639
4	201004	15.831928
5	201005	16.081746
6	201006	15.671030
7	201007	16.014851
8	201008	18.245852
9	201009	19.209146
10	201010	16.582635
11	201011	17.107149
12	201012	16.604712
13	201101	13.614680
14	201102	16.093027
15	201103	16.716166
16	201104	15.773380
17	201105	17.713823
18	201106	16.714748
19	201107	15.723497
20	201108	17.315899
21	201109	18.931723
22	201110	16.093582
23	201111	16.312383
24	201112	16.247406

Fig. 4.7 Average Revenue per Order

4. New Customer Ratio

New Customer ratio is a good way to know if we are losing our already available loyal customers we are not able to gain new customers .

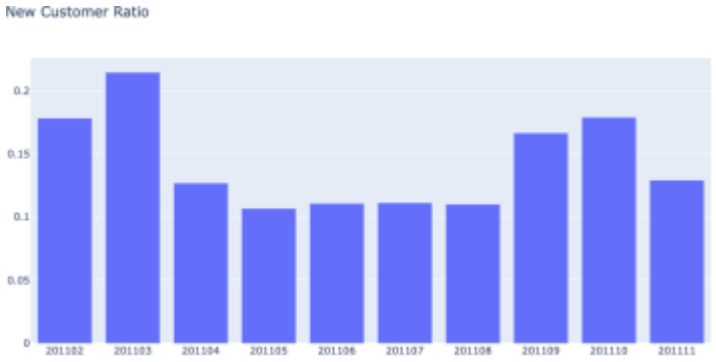


Fig. 4.8 New Customer Ratio

5. Monthly Retention Rate

Counting monthly retention rates in order to check how well is our product market fit and how sticky our product is.

Monthly Retention Rate **= (Retained Customers From Previous Month)/(Active Customers Total)**

We can observe that the Monthly Rate of Retention increased from October to December and went to levels previous levels afterwards.

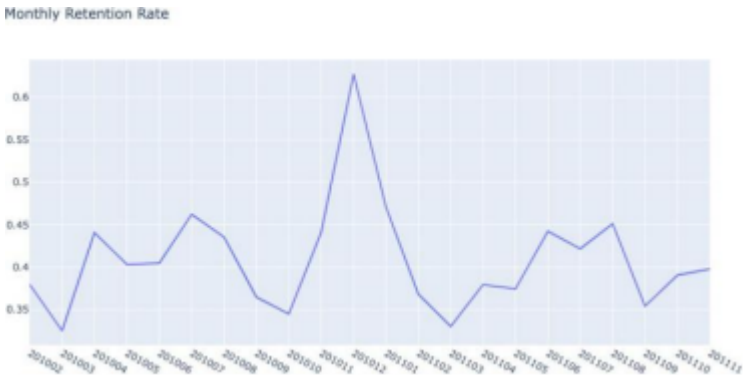


Fig. 4.9 Monthly Rate of Retention

4.4 Customer Segmentation

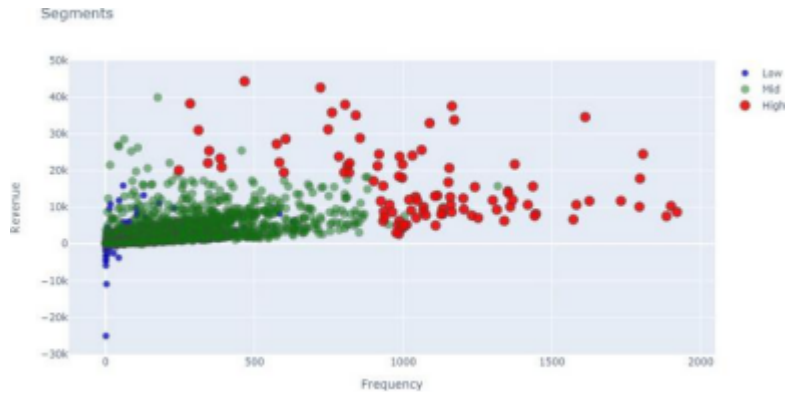
Now we have our main metrics to calculate the customer segmentation to stop as all the customers cannot be treated equally since the customer's used differently for goods and services or products, some use frequently and some not so frequently. so we have tried to segment the customers into three types according to RFM segmentation / clustering. RFM is a framework which helps us to analyze the customer behavior using the factors which we think are : (Recency) ,(Frequency) and (Monetary Value) .

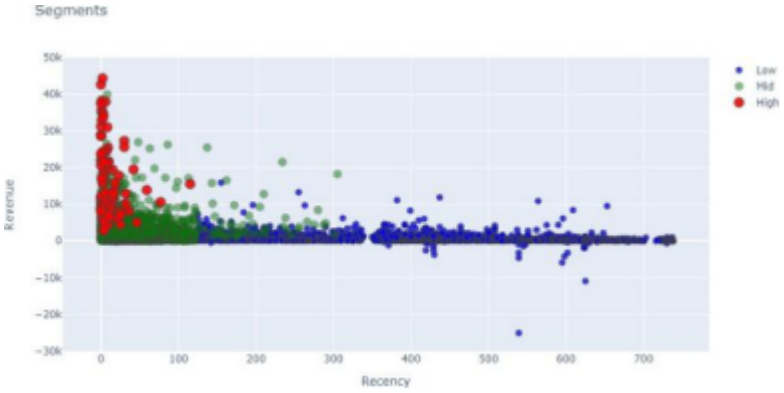
First we have calculated recency and monetary value and we are applying K-means clustering method and we have used elbow method to find out how many clusters should be used, frequency we are calculating using the total number of orders by each customer .

So we have calculated the Recency, Frequency and Monetary value (which is also revenue) and have applied an unsupervised machine learning algorithm RFM clustering to identify different groups (clusters) for each of them .

Customers are segmented as:

- Low Value: These are the customers who are less active than other customers, also not very frequent buyers or visitors and generate very low - zero - maybe negative revenue.
- Mid Value: They generate moderate value. They are fairly frequent and use our service or product often.
- High Value: These customers generate High RFM value. They are frequent buyers.





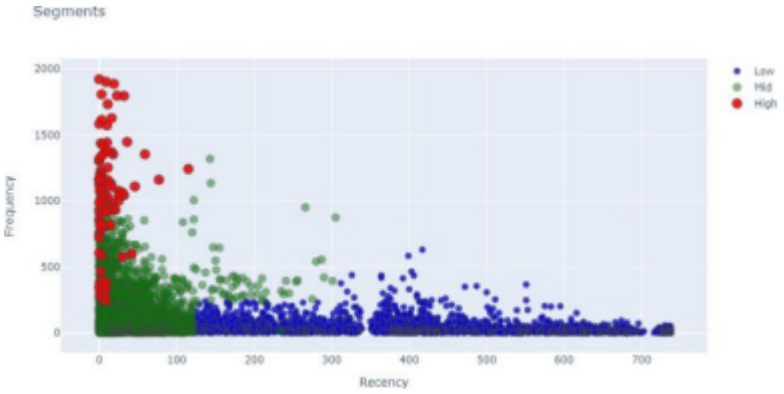


Fig. 4.10 Segments distributed on scatter plot

4.5 Customer Lifetime Value Prediction

As we have already segmented our customers and then found the high-value customers, this time we're going to measure customer lifetime value prediction which is one of the most important metrics we need to consider for the growth of the business. As a business, we invest a lot of money in the customers by giving promotions and discounts, etc to generate revenue and be profitable. These investments make some of our customers more valuable in terms of their lifetime value and these customers help the business keep going and keep the business on a stable scale and be profitable. At the same time value these customers there are always some customers who can bring down the overall profitability. For this, we need to spot these behavior patterns, do segmentation of the customers and act accordingly.

The lifetime value of the customers is also very important for any business as a business usually develops a graph of 5 to 10 years for its growth trajectory, so by knowing how loyal our customers are, what is the customer lifetime value of the current customers, by what rate are the new customers coming in and for how long are the existing customers using the goods and the products or the services that the company is providing is very beneficial to the company. As companies sell a variety of goods and services to a variety of customers, there is so much data that is collected by the company. This data needs to be processed properly and can be looked at for the proper growth of the company. When we realize that certain customers' lifetime value is more or way more than the other customers, we can then segregate a certain type of campaign for these loyal customers. This type of data can help the management team of the company to make new strategic decisions on how to approach in providing the services and goods to the customers.

The formula for calculating of the lifetime value is:

Lifetime Value **= (Total Gross Revenue) - (Total Cost)**

We have taken 3 months of data and calculated the recency frequency and monetary value and used it for predicting the subsequent next six months. Then we calculated 6 months of Life Time Value for every customer that we are going to use for training our model. There is no cost distinct inside the data set. that is the reason revenue becomes our Life Time Value directly.

Then we merged our three months and six months of data frames to see the correlations between the lifetime value and the characteristics that we have set.

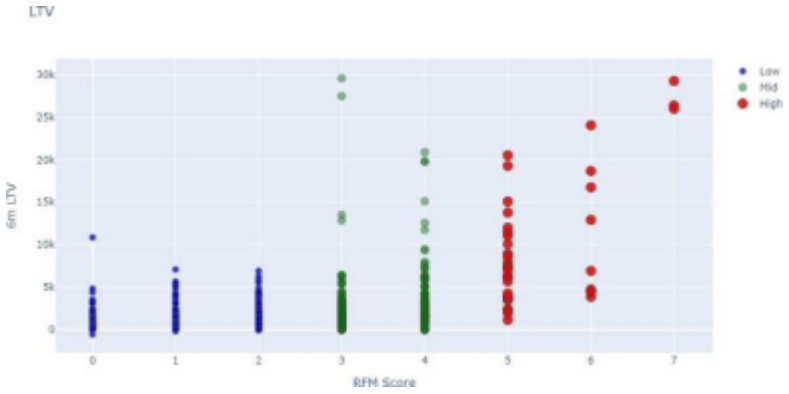


Fig. 4.11 Lifetime value vs overall RFM score

For calculating lifetime value:

1. We defined a time frame of 3 months.

2. Next, we identify the necessary features and calculated them Using the RFM analysis, these features are :

a) Recency
b) Recency cluster
c) Frequency
d) Frequency cluster
e) Revenue
f) Revenue cluster

3. Next, we performed feature engineering and split ourfeature set and label (LifeTime Value) as X and y. We use X to predict Y.

4. Created dataset for training and testing. Mechine Learning model will be built using Training set . Our model will be applied to test the set and see its real performance.

5. Achieved 84% accuracy on test set.

4.6 Churn Prediction

To develop a churn model, it is necessary to first analyze the data. Next, we can apply feature engineering to transform the raw data and extract more features from it.

In feature engineering we apply the following:

1. The numerical columns would be grouped using clustering techniques.
2. Label encoder is applied to categorical features which are binary.
3. get_dummies() is applied to features which are categorical which have multiple values.

Then we use the XGBoost Model for binary classification to calculate the churn rate.

4.7 Predicting Next Purchase Day

The knowledge of which customers are possibly buying again and at what time is also very important for any business or company. as the delay is that a pause between the ordering of goods and delivery of the goods and services can cause a lot of customers to check which in turn could produce less revenue for that particular period. For predicting of the next purchase day the first step is data wrangling. Candidates for feature engineering are selected as below :

- RFM clusters and scores.
- The days which are between the last three purchases which would have been done
- The Mean and the standard deviation of the difference between the purchases which would have been done in days.

Then, we have used the XGBoost Machine Learning model and also hyperparameter tuning is applied to further improve the accuracy of the model.

Fig. 4.12 Actual vs Predicted sales

5. RESULTS

1. Analyze Data and Understand various metrics

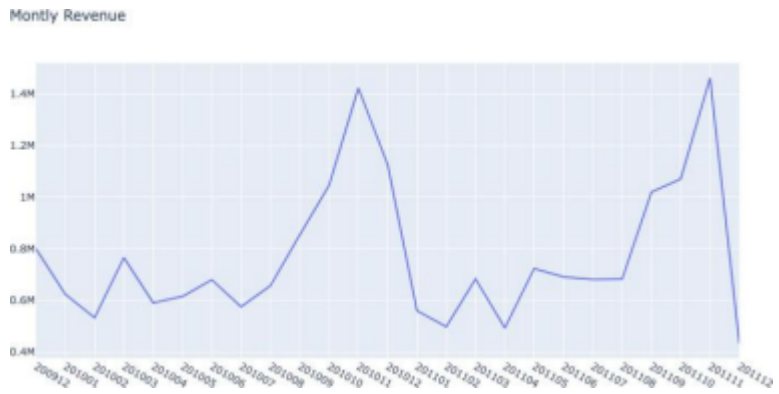
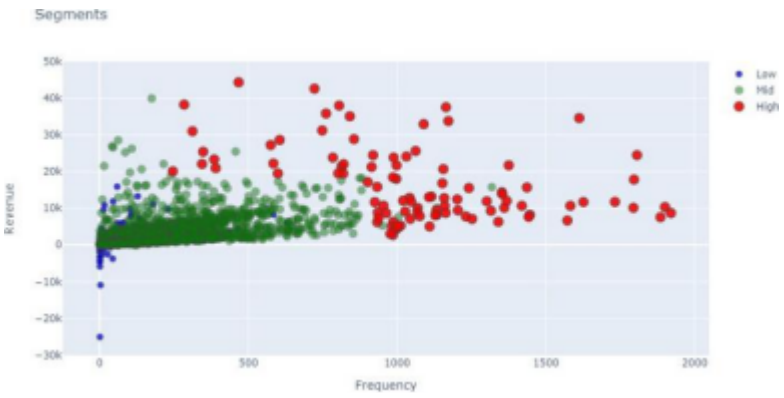
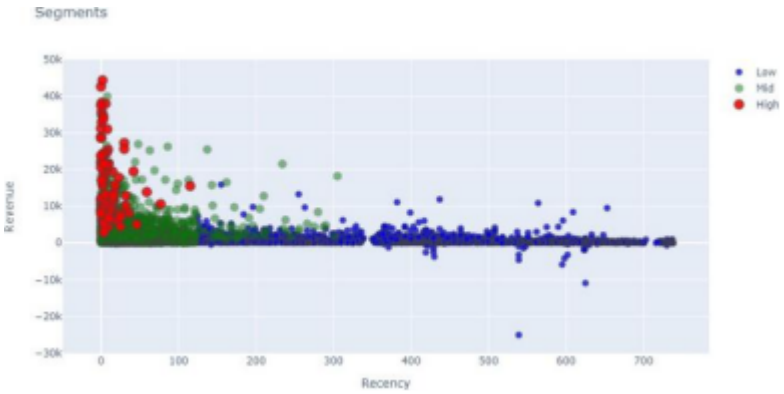


Fig. 5.1 Monthly Revenue

2. Customer Segmentation





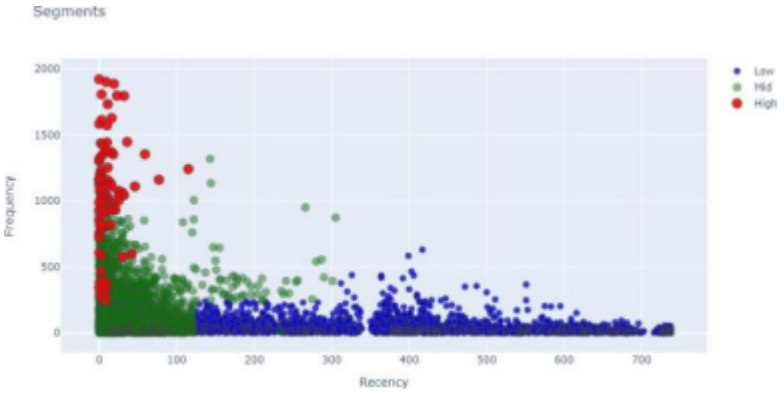


Fig. 5.2 Segments distributed on scatter plot

3. Customer Lifetime Value Prediction

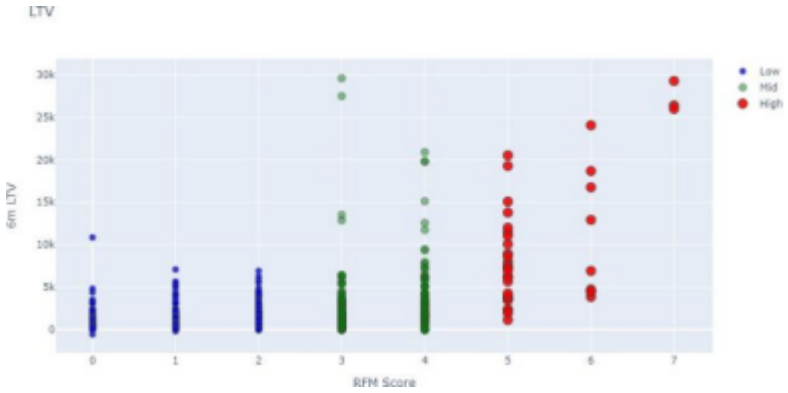


Fig. 5.3 Lifetime value vs overall RFM score

4. Predicting Next Purchase Day

Fig. 5.4 Actual vs Predicted sales

7. CONCLUSION

The Project has focused on modules like Customer Segmentation, Customer Retention, Customer Churn Prediction, Predicting next purchase day, and Predictive Sales modules.

It is quite evident that big organizations and companies are concerned about their customer churn and customer retention. as customer acquisition costs have grown during these times these companies would like to retain their existing oil customers and also reduce the churn rate for their customers. Also, Predicting the next purchase day of the customers and predicting the sales can be really helpful for these organizations and for their businesses. In this paper, we have used these five modules in order to create data analysis by using the data set.

The insights and the graphs that these modules have generated can be very helpful for these organizations to understand the metrics of their seals, their customer interaction, and their customers.