**Mini Project Report on**



**SALES DASHBOARD FOR CUSTOMER SEGMENTATION**



**Submitted in partial fulfilment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name: DISHA JAIN** **University Roll No.:2021878**

***Under the Mentorship of***

**Dr. Prakash Srivastava**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January-2025**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Sales Dashboard for Customer Segmentation”** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of Dr. Prakash Srivastava**,** Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name: Disha Jain University Roll no: 2021878

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | 4-5 |
| Chapter 2 | Literature Survey | 6-7 |
| Chapter 3 | Methodology | 8-10 |
| Chapter 4 | Result and Discussion | 11-12 |
| Chapter 5 | Conclusion and Future Work | 13 |
|  | References | 14 |

**Chapter 1**

**Introduction**

In the following sections, a brief introduction and the problem statement for the work has been included.

* 1. **Introduction**

Customers are the foremost requirement or the demand for a successful business. Entrepreneurship is the outstanding domain which is currently outgrowing in this market. Building new business or outstanding with the existing ones demands some rules that need to be seriously follow to introduce a brand in a market. Everything is made for customers; every effort is done to attract them. Different ways of attracting new ideal potential customers to avail different services and products or to retaining the existing ones and gaining their loyalty, so that they won’t churn from the services and the products they are using should be one of the major concerns of a business or a startup. It is very important to understand the customer segmentation to analyse what customer demands and how much loyal customers do that business own, so to prevent business losses. As it is very important to understand customers who are the foremost ask of any small or big business. Doing sales insight for customer segmentation a classifying customer on various basis helps to prevent churning which a major concern for any organization. Churn is a key driver for EBITDA (Earnings Before Interest, Taxes, Depreciation and Amortization). A churned customer gives less revenue or no revenue and increases market competition. Customer segmentation is a way of analysing customer characteristics and forming groups based on who they are, and their unique characteristics Customer segmentation enables businesses to use targeted messaging, rather than a one-size-fits-all approach to business results. For example, a streaming music company could segment its customers by customer listening, typical times of day, and music devices and then the company could use analytics to attract more music lovers who match those characteristics. It is a critical prediction for many businesses because acquiring new clients often costs more than retaining existing ones. If businesses, sales teams, and marketing teams segment their customers and continue to refine their distribution channels, they can: Create a beautiful and strong customer persona, Provide marketing messages to every customer segment, Message through marketing channels that can effectively reach customer segments, Make the most of the marketing budget, Overhauling marketing strategies throughout, Examine the reaction of parts to different price choices, Enhance products and customer experience based on segment feedback, Turn prospects into loyal customers. Various methods are used for customer segmentation like RFM (recency, frequency, monetary) analysis. The more competition there is, the higher rates companies can charge. Firms in low-competition markets may charge lower rates than highly competitive markets. The number of services associations can provide also affects the fees charged to the associations. The more services the company provides, the higher the company can charge. Customer segmentation is a method of organizing customer groups based on shared characteristics, behaviours, or preferences with the goal of providing a more relevant experience. So, by analysing and visualizing customer behaviour, decisions can be made respectively.

* 1. **Problem Statement**

The purpose of this project is to conduct a Customer Segmentation Analysis for an Automobile bike Company. Customer segmentation is performed by developing an RFM Model. RFM (Recency, Frequency, Monetary) analysis is a behavior-based approach grouping customers into segments. It groups the customers based on their previous purchase transactions. In this analysis the customer segment was divided into 11 groups. The analysis will help in determining which customers segments should be targeted to enhance sales revenue for the company. A Sales Dashboard for Customer Segmentation is developed using Tableau and the data quality assessment and analysis is done using Python.

**Chapter 2**

**Literature Survey**

In this chapter some of the major existing work in these areas has been reviewed.

Customer segmentation is not a completely unknown field. This is the most concerning field which already has the hands-on by some researchers. But since it is not that easy to make a complete precise model classify customers based on their behaviour of purchases, i.e., how much loyalty do they own to that company. So, the research is ongoing till now. With the constant repetition of artificial intelligence, new ideas in business world, new ongoing trend coming with entrepreneurship, the researchers are continuously developing the ideas to solve business problems or to help business using these machine learning tools as well as different ongoing tech. The most important thing that paper should include is precision, proper theoretical knowledge and proper outcome that can prove to helpful in real world. This paper includes information about the dataset of an automobile bike industry with detailed insight analysis using python and visualization which is the foremost part of this paper. There are different ways by which customers can be categorized on their loyalty but visualizing all these make it more prominent for a company to understand and analyse better. In recent years, different work has been recorded in this area of Customer segmentation. Different methods have been recorded by the researchers. Some of them are:

1- Omar Adwan et al. [1], new commercial competitors offer multi-layer perceptron neural networks (MLPNN): tor modelling and analysis, resulting in serious loss of profit. They have also used the data from the telecommunication industry but they use MLPNN technology to detect its precision.

There are different works in these areas using machine learning and different market analysis. The work in this field has also done using machine learning, i.e., unsupervised machine learning algorithms involving clustering to divide customers basis on their loyalty. These reviews lack some respect to the theoretical knowledge, new introduced methods and techniques which can be used in this field. Also, clustering can be misleading as human decisions are involved in that which cannot be appropriate every time. This time, we are using RFM (recency, frequency, monetary) analysis which is completely based on mathematical calculations which increases the chances of accuracy. Also, this paper includes visualization which makes everything very understandable. But achieving precision is still not so easy in case of predicting customer behaviour. So, the search for the precise one is still being continued. Based on the above analysis, in order to take the responsibility to solve these deficiencies by making our model more prone to be believed by people out there. There is a compressed overview of the research situation in this field of customer segment analysis. The model which is being represented by this review paper will be able to make people more prone to be more helpful to secure their customers by analysing their customer segmentation and may be helpful to the business out there. Some specific contributions are as follows:

* We have divided the model into two parts, one which takes the insight from the dataset and other visualize the analyses using dashboard of the visualized data based on the insights.
* We have use python to take proper insights from the data and performed RFM (recency, frequency, monetary) analysis to divide customers into major 11 categories based on their RFM score.
* After proper insightful analysis and calculating RFM score for each, we have used dashboard which we have designed using Tableau to visualize that analysis and make proper decisions in regards with customer as well as for that organization/company/business.

This paper focus on the techniques, different ways to find insights from the dataset, to choose the most affected areas, and then to visualize based on those categories. Precision increases when we use mathematics instead of assumptions thus making more surety over the decision. Customer segmentation is one of the most important aspects of this industry. Thus, there are different ways by which customer behaviour are being analysed, so to make perfect decisions and betterment of the company.

**Chapter 3**

**Methodology**

This section presents the methodology which is used in this model to build it from the very start to the end where it can be declared as a model to detect customer churn. It contains all the algorithms, methods, figures used to make this model. It contains detailed information about each step from data collecting to data visualization.

A systematic process is being followed to build a model which we will be discussing in this section. The very first step is data collection. For performing this customer segmentation, we have use data of a automobile bike industry. There are different files on this basis of which the categorization among customers has been done. Those datasets consist of Customers Demographics, Customer Address, New Customer and Transactions made by customer.

The first step towards generating useful insights from the data was the data preparation, quality assessment and data cleaning step. After the cleaning process exploratory data analysis on the dataset and identification customer purchasing behaviors to generate insights can be performed.

In the data cleaning step, the data quality of the following datasets was first assessed. Data cleaning steps have been performed on the following datasets CustomersDemographics.xlsx, NewCustomerList.xlsx, transaction\_data.xlsx and CustomerAddress.xlsx. The data cleaning steps involved dealing with numeric and non-numeric values, dealing with missing values and replacing all the missing values with null values and removing or reducing those columns which are of less use so to decrease data complexity.

After the data cleaning process, exploratory analysis on the dataset is performed and the following insights are obtained:

A graph of a number of people

Description automatically generatedA graph of a number of customers

Description automatically generated A graph of a person and person

Description automatically generated

**Fig. 3.1 New Vs Old Customers Age Distribution Fig. 3.2 Bike Purchases Over last 3 years by Gender**

**A graph of a bar chart

Description automatically generated with medium confidence A graph of a number of bars

Description automatically generated with medium confidence A graph of different colored bars

Description automatically generatedA graph of different colored bars

Description automatically generated**

**Fig. 3.3 New Vs Old Customers Job Industry Distribution Fig. 3.4 Wealth Segmentation by Age Category**

A graph of a number of customers who own a car

Description automatically generated

**Fig. 3.5 Car Owned by States**

After doing pre-processing and data exploration and comparison of data, we have done RFM (Recency, Frequency, Monetary) analysis. RFM analysis is used to categorize customers based on RFM score as the most loyal or least loyal customers.

In this stage of analysis, the customer segmentation was done by developing an RFM Model. The RFM (Recency, Frequency, Monetary) analysis is a behavior-based approach grouping customers into segments. It groups the customers based on their previous purchase transactions.

In this analysis the customer segment was divided into 11 groups. The groups being:

* Platinum Customers
* Very Loyal Customers
* Recent Customers
* Potential Customers
* Lost Customers
* Losing Customers
* Late Bloomer
* High Risk Customers
* Evasive Customers
* Becoming Loyal
* Almost lost Customers

RFM score is calculated based on these three bases. We have divided customers into four scores. Giving ranks according to their loyalty. Giving ranks (1,2,3,4) based on min, 25%, 50%,75% and max. Just giving rank 4 to top some %, giving rank 3 to next some % of customers, next some % of customers are given rank 2 and the bottom some % is given rank 1. This is done for all three, i.e. Recency, frequency and Monetary after arranging the data into descending order and then ranking them accordingly.

After assigning them ranks, RFM score is calculated considering weightage 100 for Recency, weightage 10 for Frequency and weightage 1 for Monetary respectively. Thus, formula generated for each customer is as follows:

RFM score= 100\*Recency(rank)+10\*frequency+1\*Monetary

Now, based on this RFM score, customers are categorized among these 11 categories ranging from Platinum being most loyal customers and almost lost customers being most unloyal customers.

Accordingly, comparison among recency vs monetary and frequency vs monetary is done and dashboard is being prepared as a result to visualize all these insights to do customer segmentation.

**Chapter 4**

**Result and Discussion**

This section contains the results that are produced after data cleaning, data exploring and then performing RFM analysis and calculating the RFM score and categorizing the customers among 11 categories. The comparison is done among recency vs monetary and recency vs frequency to get more clarity about the customer behaviour.

*Recency vs Monetary*:

The visualization shows that recent customers have purchased more products and generated relatively more revenue than the customers who visited a while ago.

A graph with blue dots

Description automatically generated

**Fig. 4.1 Recency Vs Monetary**

*Frequency vs Monetary*:

The visualization shows that customers belonging to Platinum/ Very Loyal/ Becoming Loyal Customer Segments have a greater frequency and generate greater monetary for the business.

A graph with blue lines

Description automatically generated

**Fig. 4.2 Frequency Vs Monetary**

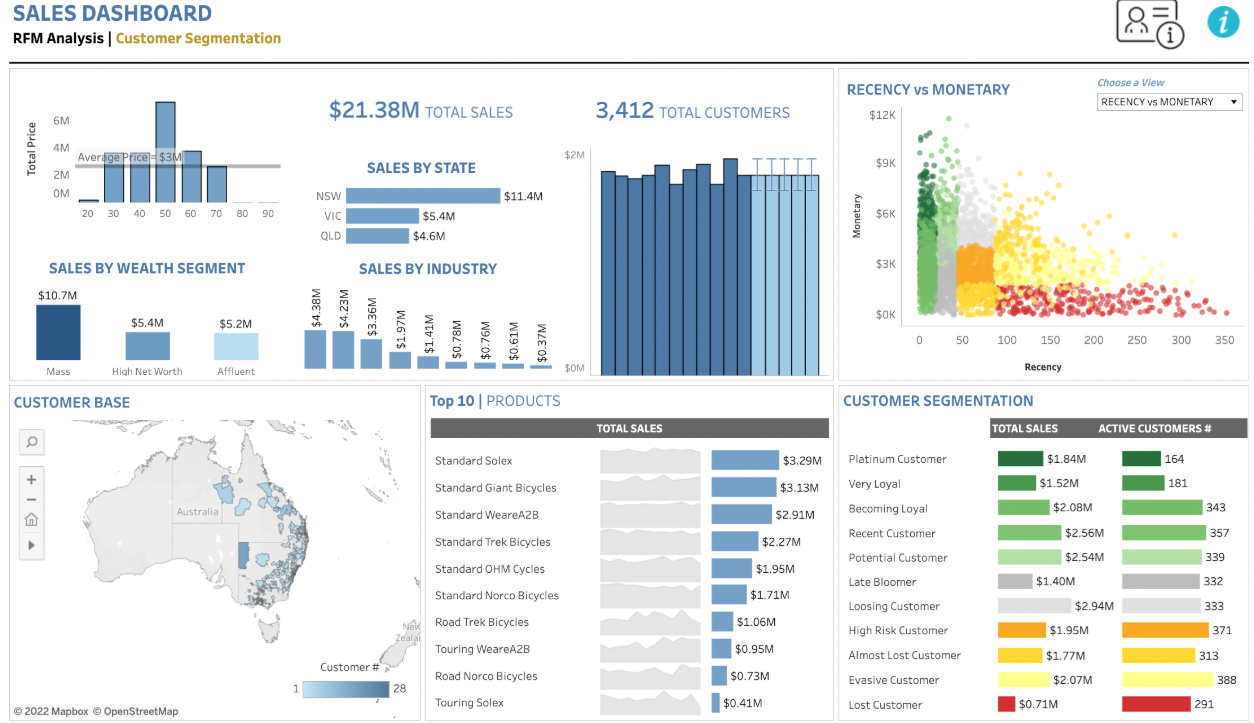
As a result of these comparisons and categorizing the customers among 11 levels of their loyalty, we have plotted the graph showing the number of customers for each category. This plot shows the number of customers present under a customer segment.

A graph of a number of customers

Description automatically generated

**Fig. 4.3 Number of customers by customer segment**

After all these categories, we have plotted a dashboard using Tableau which helps in visualizing every other number of customers that has been analyzed and depicted as loyal or not. This dashboard helps in many purposes and thus is needed in taking proper insights.



**Fig. 4.4 Sales Insight of Customer Segmentation**

And in this, all these analysis and insight of data resulted in this proper visualization using this dashboard, thus giving more precise results.

**Chapter 5**

**Conclusion and Future Work**

Business is not just a field; this world of start-ups demands more and more emerging ideas and new enlightened technologies. They ask for every way out so, that they make their successful and stable business. Retaining customers from churning to another service can be proved to be very helpful for them. These emerging technologies is a big push to the business world as well. Depicting sales insight for customer segmentation has proved to be a great use for new startups or the existing business. We have used RFM (recency, frequency, monetary) analysis through which we calculated the RFM score and categorize customers into classes. Recency meant for how recent a customer has purchased a product, frequency meant for the number of times a same customers purchase that product and monetary denotes the customers who buys for how much, does the most recent customers buys the costliest products or not. So, RFM score is based on these three aspects on which the customers are being analysed. This technique helps in precision of the result. We have involved mathematics but that includes some of our considerations, which cannot be 100% precise. Also, there has been various ways by which customer segmentation can be done, one is clustering which is unsupervised machine learning algorithm. The continuous research has been ongoing for betterment of both customer and the company, so increasing precision is very important. Predicting human behaviour through machine is a less precise thing. So, saying that every model containing human behaviour is accurate is incorrect to say. The only thing we fight for the most accurate result, the close one can get to the most optimal result. And that’s why, we this tech field try to work on closest correct prediction in every way possible. Also, our model, even after using a mathematical function and then visualization which somehow gives some percentage of precision, which on comparison with other methods may or may not be a better one. The only reason behind this is customer unpredictive behaviour which have no limits and criteria.

Future work can include, models being made by working on more re-fined dataset pre-processed, cleaned and having more decisive results. Applying more optimized techniques to increase the precision of the output is yet need to be examined. As this business is not an area of experiments, because they are the most affected fields, they work on accuracy only. In this field, a lot more is left to be analysed and be investigated.

**References**

[1] Omar Adwan, Hossam Faris, Khalid Zaradat, Osama Harfoushi, Najih Ghatasheh and others

Predicting subscriber churn in the telecom industry using multilayer preceptron neural networks: observations and analysis

[2] Ning Lu, Hua Lin, Ji Lu, Guangquan Zhang “A consumer Churn Prediction Model in Telecommunication Industry Using Boosting”, IEEE Transactions on Industrial Informatics, vol. 10, no. 2, May 2014 in the.

[3] Ammara Ahmed, D.Sc. Maheshwari Textiles Research and analysis of churn forecasting techniques for customer retention in telecommunication industries 2017 Fourth International Conference on Advanced Computing and Communications Systems (ICACCS), IEEE (2017)

[4]L. Ning, L. Hua, L. Ji, Z. Guangquan Predictive Modeling in Telecom Industry Using Boosting”, IEEE Definition. Ind. Please give suggestions. 10 (2014) 1659–1665.

[5] Coen W. de Bock, Dirk van den Poel, “Findings Analysis of variable-based group classification for the patient churn prediction”, Expert Systems Management 38 (2011); 12293-1 is available

[6] Kalkan IE, S¸ ahin C (2023) Evaluating cross-selling opportunities with recurrent neural networks in retail marketing. Neurocomputer Applications 35(8): 6247–6263