ANALYSIS OF RETAIL CUSTOMER RETENTION POTENTIAL

Avinash Chowdary Kongara
Computer Science
University of Missouri Kansas city
Kansas city, Missouri
ak28d@umsystem.edu

Harish Ravilla
Computer Science
University of Missouri Kansas city
Kansas city, Missouri
hry5g@umsystem.edu

Sai Purna Praneeth
Computer Science
University of Missouri Kansas city
Kansas city, Missouri
pcpb6@umsystem.edu

Sai Kiran
Computer Science
University of Missouri Kansas city
Kansas city, Missouri
smxxt@umsystem.edu

Ravi Chandra
Computer Science
University of Missouri Kansas city
Kansas city, Missouri
rtty5@umsystem.edu

Varun
Computer Science
University of Missouri Kansas city
Kansas city ,Missouri
vn6pm@umsystem.edu

ABSTRACT

"Companies have concentrated on customer retention and loyalty programs since loyal customers are a company's most valuable asset. In today's extremely competitive marketplaces, customer loyalty is vital for a company's survival. For a retail organization to avoid attrition (customer churn) or dropouts in the middle, it must engage customers in engaging continuous interactions. Profitability, retention, production volumes, potential revenue loss, putting additional expenses on assets, and, most significantly, the CRM component should be the primary focus. Customer retention tactics are designed to keep relationships with value-adding customers by minimizing the rate of defection."

"Retailers build the measure to identify the services they require without losing them by analyzing past data, and once they have become obsolete from the plan, we must attempt to reclaim them. Our technique begins by examining each attribute of the customer's past data and developing relationships between the attributes. We must concentrate on medium-level clients by meeting their wants because they are responsible, proactive, and reactive in their demands. If they're happy, they can renew their plans and stay with the company."

This will help the industries to analyze their positives and negatives to maintain and improve the customers base around the specific location, to provide discounts for attracting the customers.

After implementing the strategies, this project ends by categorizing the customer retention potential as high/medium/low.

Design– This is carried out using a quantitative dataset of with 15 attributes containing all the information related to customers, advertise email and transactions history.

Motivation – The major contribution of this project is the clear link that it establishes between customer retention performance.

KEYWORDS

Customer retention, Retailers

1 INTRODUCTION

Relationship marketing has become a vital strategic tool for organizations in today's dynamic market, where client desires and tastes change regularly. The importance of relationships and the need to join relationship networks is highlighted by the rapid developments in nearly all business types. As a result, organizations are increasingly focusing on establishing long-term profitable connections in business, internal, and consumer markets in order to give greater value to their customers."

"To avoid customer churn or dropouts in the middle, a retail organization grabs customers' attention through fascinating continuous interactions." Retailers build the measure to identify the services they require without losing them by analyzing past data, and once they have become obsolete from the plan, we must attempt to reclaim them. We must concentrate on medium-level clients by meeting their wants because they are responsible, proactive, and reactive in their demands.
""If they are happy, they can renew their plans and stay with the company. So let's take a deeper look at the patterns and emphasize the theme even further."

"Businesses utilize retention marketing to try to increase customer loyalty. This will assist industries in analyzing their advantages and disadvantages in order to preserve and improve their consumer base in a certain place, as well as providing discounts to entice customers. Business CRM also aspires to create a system that can process data in real time and engage with clients more quickly."

The following figures show and support the economics of retention marketing programs:

• According to Marketing Metrics, selling to a current customer has a 60–70% chance of success, whereas selling to a new prospect has a 5–20% chance of success (Charlton, 2012).

• According to Marketing Metrics, selling to a current customer has a success percentage of 60–70%, but selling to a new prospect has a success rate of 30–40%.
• According to Marketing Metrics, selling to an existing customer has a 60–70% success rate, whereas selling to a new prospect has a 5–20% success rate (Charlton, 2012).

Companies that establish and manage client retention strategies efficiently can gain a competitive advantage, better profitability, and an amazing reputation, as these data shows.

2 PROJECT DESCRIPTION

Our strategy starts with analyzing each attribute of the previous data of the customer and forming the relations between each of the attributes like
 • Total orders vs preferred delivery day.
 • Total orders vs customer feedback or response using mail.
 • Order frequency vs Total orders.
 • Total orders vs Delivery type.
 • Orders vs Automatic subscription.
 • Number of days vs Profits.
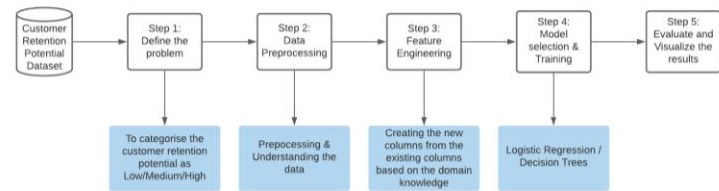 • Total orders vs City

2.1   PROJECT WORKFLOW

Fig.2.1

2.1.1   DATA PRE-PROCESSING
 • Imported all the required packages such as pandas, numpy, matplotlib.
 • Refined all the null values, special, foreign characters and merged all the datasets to single csv file based on the customer ID.

2.1.2   PROJECT IMPLENTATION
 • We have used python and spark cluster for the implementation of the project.
 • In this implementation, we have performed various operations like group by, aggregation, merge, and sort for better understanding of the dataset. These operations are performed by using pandas library.
 • We designed the code to compare various attributes and to visualize the relations between them, used matplot libraries to perform this.
 • Created new columns like making business, number of days with business with the help of existing data to make a clear visualization.

natural appearance and makes it simpler to interpret data, patterns, and outliers within the dataset.
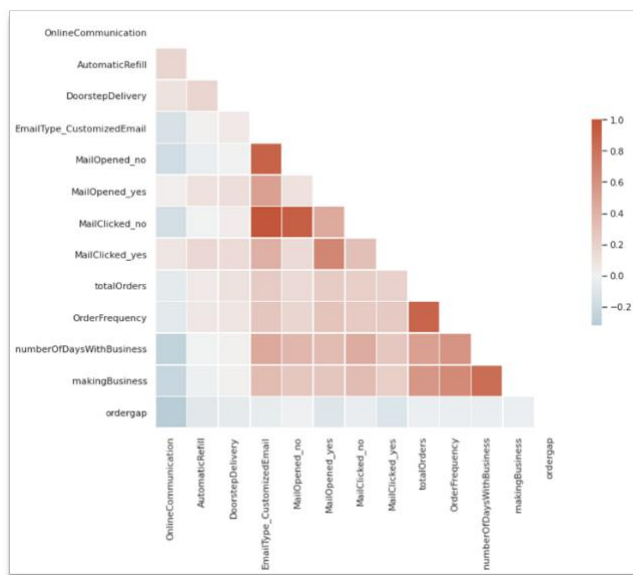
2.1.3   DATA VISUALIZATION

Data visualization provides clear idea of the customer retention potential dataset by giving it visual context through graphs. This gives the data a more

Some Benefits of the data visualization are

- Correlations in Relationships
- Trends Over Time
- Frequency

The various visualizations are:

Fig 1

Fig 2

- If many business days are more, we can have more business by performing efforts to get profits.

Observations:

- Email_type_customized:
    - Mails opened and clicked are slightly correlated.
    - Mails not opened and not clicked are highly correlated.
    - When the Emails are Customized, the mails have not been opened and not clicked.
- Total_Orders:
    - Order_Frequency and Making business are Maximally correlated.
    - If order Frequency is more, we can make business, corelated in terms of profitability.
- No of Days with business:
   - Making business is highly correlated.

Fig 2

Observations:

- Annual wise: Total orders were witnessing more in 2013 and again falling in the year 2014.
- Monthly wise: For the months from August to January, the sales are high when compares to other months.
- Week wise: The total orders are relatively same in for all the days in week except Thursday and Friday.
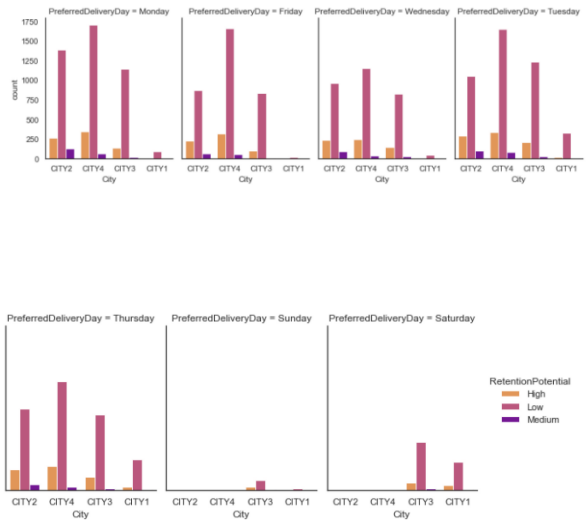- Day wise: The orders are comparatively more in the middle days of the month than start and end days.

Fig 3

Total OrderFrequency by Day
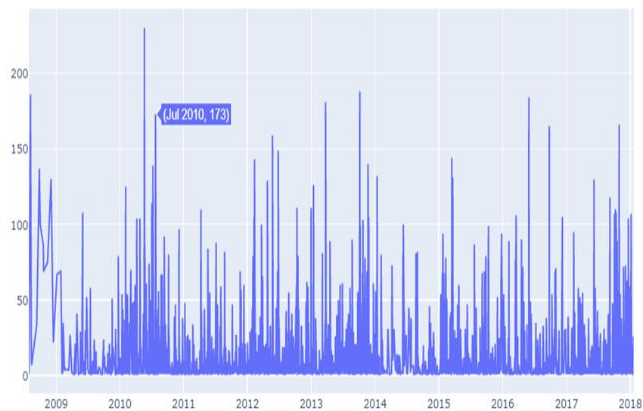
Observations:

- The order frequency witnessed more between the years 2010 and 2011.

Comparision Between the City , PreferredDeliveryDay , Customer RetentionPotential

Some other comparision

➢ This plot shows the details how door step delivery is distributed across the cities.
➢ 0 indicates that the customer is not intrested in doorstep delivery and 1 indicated that he opted for doorstep delivery
➢ And also it is clearly shown how high/medium/low is scattered across all cities

➢ The above visualizations shows how cities are scattered withrespect ro count of the orders.
➢ And we also included the concept of retention potential in this plots so that it would get more clear.
➢ Additionally we have spllitted the total thing into preferred delivery date so that it make the visualization more detail.
➢ From the above two graphs we can see that weekends is quite low when compared with the other days
➢ There might be various reasons for that.So that the company can make some strategical moves and develop their company
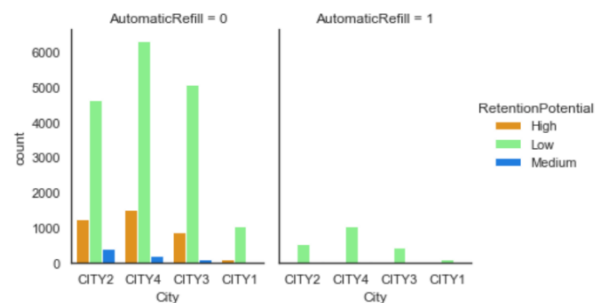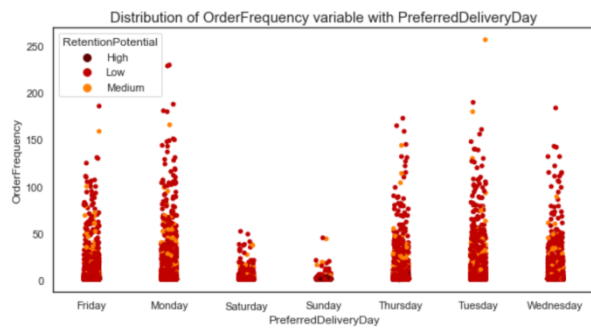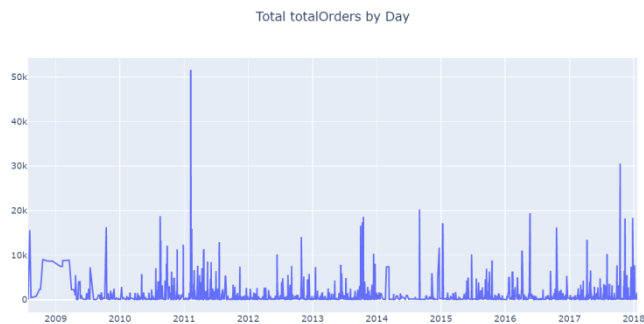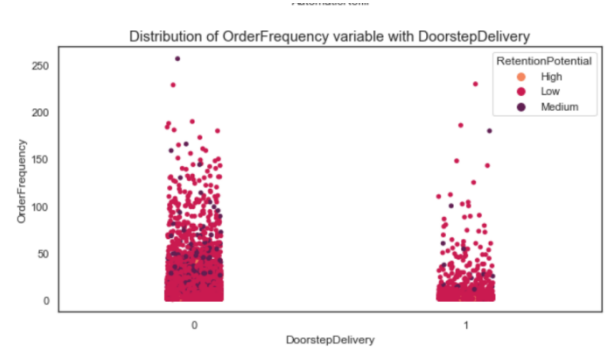➢ It is clearly shown how the retention potential is distributed all across the week

➢ This plot shows about the distribution of count w.r.t cities and included the retention potential with inclusion of automatic refill
➢ There will be less number of customers using this feature, but the number will be constant.
➢ So, the company can make some useful decision by this plot

Distribution of OrderFrequency variable with PreferredDeliveryDay

Distribution of OrderFrequency variable with DoorstepDelivery

➢ We focused on orderfrequency and preferred deliver day
➢ How retention potential is distributed across all days of week
➢ Since many customers have low retention potential, they are more dominant
➢ Asusual, weekends the plot is low as we discussed above.

➢ The above plot shows about the type of delivary the people prefers.
➢ According to the above data it is observed that most of the people prefered collecting from pick up stores rather than delivering at their door step.

Total totalOrders by Day

Distribution of TotalOrders variable with PreferredDeliveryDay

➢ The above observation shows that about the number of orders ordered over 10 years period.
➢ From the above observation the year 2011 has large number of orders placed and years 2014 and 2015 recorded very low number of orders.
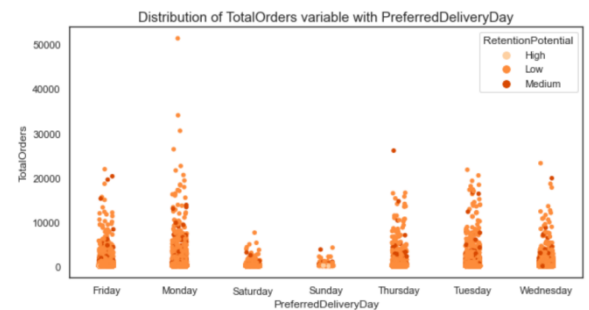
➢ The above plot shows about the day of delivery people mostly prefers.
➢ From the above graph it people mostly prefer working days to deliver their order rather than in weekends.
➢ It is because due to lack delivery services to isolated areas.

Orders Using AutoRefill

OrderFrequency vs TotalOrders
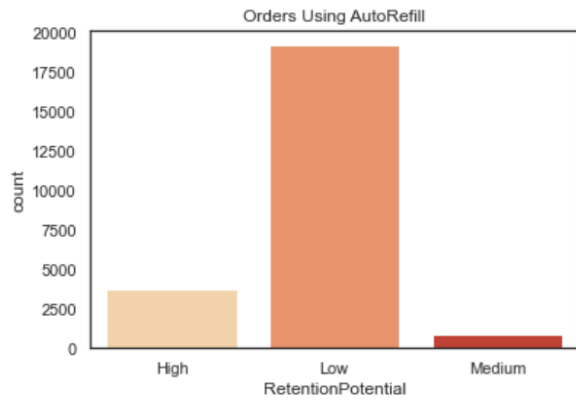
> ➢ From the above observation there are more people people belong to low retention potential category where company should concentrate more.
> ➢ In addition company should maintain the people who belongs to high and medium category without loosing them.
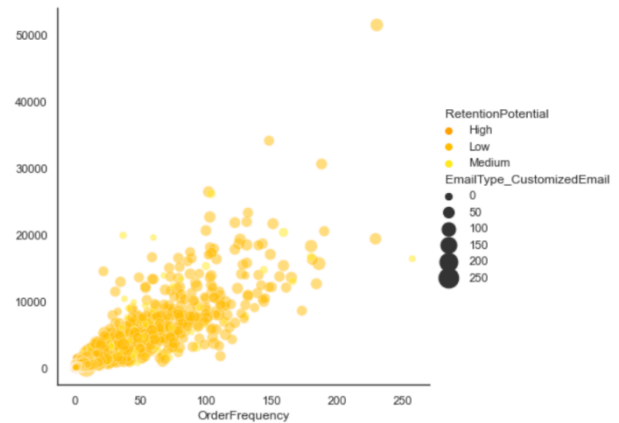
> ➢ The above scattered plot shows plot of retential potential data.

5. Desicion Tree:-

PreferredDeliveryDay vs totalOrders
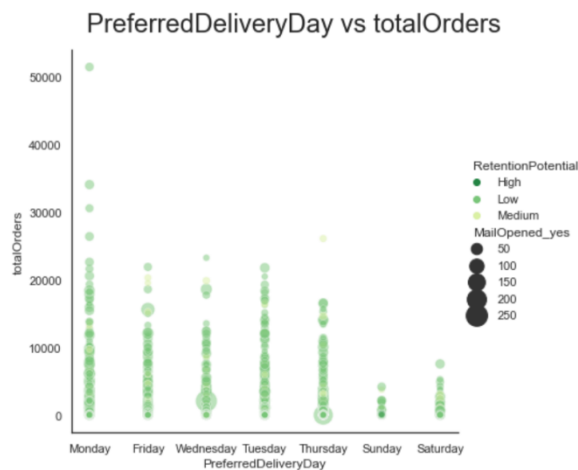
> ➢ The above plot shows with the preferred delivery day in addition with the mail opened_Yes or No indicated with the black dot.
> ➢ It is shown that more number of people preferred on the weak days rather than weekends.
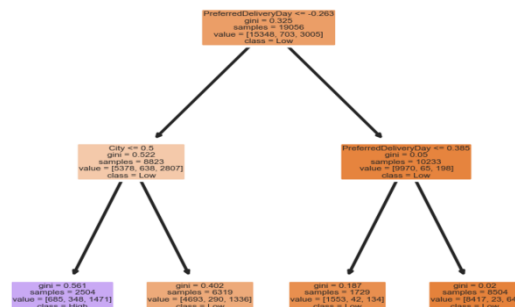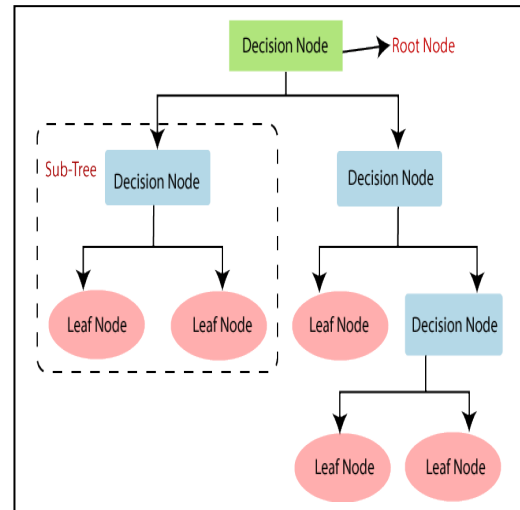
Used normalization as a scaling technique in which values are shifted and rescaled so that they end up ranging between -1 and 1.

Gini Index: It calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly

The supervised decision tree Algorithm techniques include the decision tree Algorithm. It can be used to solve a classification challenge or a regression problem.

The purpose of this technique is to develop a model that predicts the value of a target variable, and the decision tree solves the problem by using the tree

Representation where the leaf node corresponds to a class label and characteristics are represented on the internal node of tree.

It's a versatile tool that can be used in a variety of situations. Both classification and regression problems can be solved with decision trees. name implies that it use a tree-like flowchart to display the predictions that result from a sequence of feature-based splits. It begins with a root node and finishes with a leaf decision. The decision to make strategic splits has a significant impact on tree's accuracy. The decision criteria for classification and regression trees are different.

To decide whether to break a node into two or more sub-nodes, decision trees employ a variety of techniques. The homogeneity of the generated sub-nodes improves with the generation of sub-nodes. To put it another way, the purity of the node improves as the target variable grows. The decision tree divides the nodes into sub-nodes based on all available variables, then chooses the split that produces the most homogeneous sub-node.

In addition the below plot shows the gradual increase over the years 2012 and sudden increase in the year 2013 and after decrease in the sales.

From the year 2013 there is gradual decrease in the orders over the years. This might be due to some offers offered by the company in 2013 which made people to order products.

CONTRIBUTIONS

NAME	RELATED WORK
Avinash Chowdary	Analyzing the dataset and processing it for making it into a single csv file by grouping different datasets. Assisted in visualizations and report preparation.
Harish Ravilla	Research on different datasets and analyzing. Code Execution and assisting in plotting outputs. Assisted in PPT's, presentations and report.
Praneeth	Literature reviews on different concepts involved in the project.
Sai Kiran	Detailed research on different datasets and pre-processing the dataset.
Ravi Chandra	research in suitable algorithms and on the expected outcomes of the project
Varun	Assisted in project execution and in processing the dataset.

Conclusion: -

The customer's classification as LOW, MEDIUM, HIGH Retention Potential applies to all industries. To meet the profitability expectations without facing the condition of attrition. Business should concentrate more on medium retention potential customers in which vigorous efforts will lead them to the high potentiality category. Industries should fulfil the needs of customers having high potential, without any difficulties. Another focus must be on the people who tend to have low retention potential by improving the approaching analysis with captivating & fascinating attempts from industries to pull customers towards their services.

REFERENCES

Literature Reference:

[1] Ang, Lawrence, and Francis Buttle. "Customer retention management processes: A quantitative study." *European journal of marketing* (2006).

[2]Nasır, S. (2017). Customer Retention Strategies and Customer Loyalty. In Advertising and Branding: Concepts, Methodologies, Tools, and Applications (pp. 1178-1201). IGI Glob