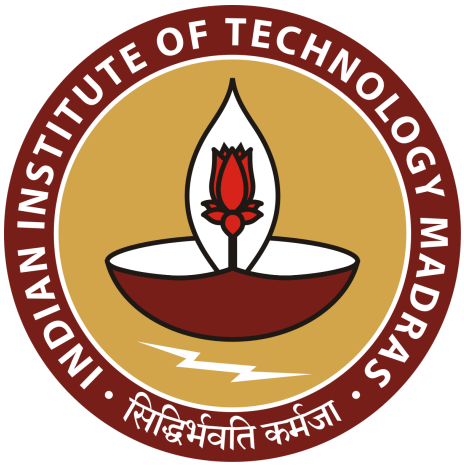


INDIAN INSTITUTE OF TECHNOLOGY MADRAS

CHENNAI – 600 036

Unveiling Challenges:

Addressing and Analyzing Operational Issues in a Small Business

Name: Kannan S

Student Email: 21f3000990@ds.study.iitm.ac.in

Roll No.: 21f3000990

Final Report for the Project on Business Data Management

May 2024

Contents

Executive Summary

This project focuses on the business establishment, XYZ, and the problems faced by it, as established in the proposal phase. Owned and run by Mr. R, it is a Salem-based multi-channel business that delivers onions to various customers. The problem it faces mainly concerns a lack of customer retention and challenges in expanding the establishment.

In the mid-term report, basic data collection, and analysis were conducted. This process highlighted the need to focus on customer-centric strategies and geographical diversification. Additionally, the importance of targeted approaches towards restaurants was also recognized.

The final report delves into a thorough analysis of Customer Retention Dynamics, Feedback Sentiments, Churn Rates, Customer Lifetime Value (CLV), and Geospatial Insights. It is achieved by employing advanced analytical methodologies such as Exploratory Data Analysis(EDA), Sentiment Analysis, Topic Modeling, and Geospatial Analysis, to provide actionable insights for strategic decision-making.

Key findings explore the effectiveness of targeted marketing towards restaurants in enhancing customer engagement and retention, identification of prevalent themes of negative feedback for targeted improvements in service quality, and gradual decline in churn rates over time indicating effective retention strategies. Moreover, insights from customer segmentation and CLV analysis highlight the significance of restaurants in driving business growth and underscore opportunities for personalized marketing initiatives. The geospatial analysis reveals concentrated customer bases in specific localities, informing potential strategic expansion of the business.

Detailed Explanation of Analysis Process/Method

Overview of the analysis process used:

The Midterm report utilized a range of analytical techniques to investigate customer retention dynamics across different dimensions, including geographical location, business type, and revenue. Through

the application of heatmaps, we identified strategic opportunities for business expansion. Specifically, our analysis suggests directing marketing efforts towards restaurants, leveraging methods such as localized promotional activities like posters and word of mouth, and soliciting feedback from clientele to better cater to their needs. Such initiatives could yield significant improvements in customer retention rates, leading to enhanced revenue streams, profitability, and sustainability.

Currently, the project has transitioned its focus towards the implementation of advanced methodologies to enable a more thorough analysis. The main goal of these analyses is to acquire a holistic comprehension of customer needs through the examination of negative feedback and retention metrics, as well as the identification of high-value customers. Furthermore, the objective is to derive critical insights that will guide strategic expansion efforts into new markets, ultimately bolstering overall profitability and sustainability.

Detailed Explanation of Methodologies Employed:

1. Exploratory Data Analysis

EDA was performed over the daily sales and customer dataset to illustrate and compare the results after applying methodologies discussed in the midterm report, python libraries Matplotlib, Seaborn, and pandas used for aggregation and visualization.

NOTE: The datasets used can be accessed here

Data Preprocessing:

- **Data Cleaning:** Checked and handled any missing or inconsistent data in daily sales and customer datasets.

Customer Data Analysis:

- **Comparing Unique Customers by Business Type and Locality Each Month**
    - The number of unique customers by business type and locality each month will be aggregated and visually observed to compare and interpret before and after applying targeted marketing toward restaurants.
- **Comparing Overall Customers Share by Business Type and Locality**

- Overall Customer share by each business type and locality will be aggregated and visually observed to compare and interpret before and after applying targeted marketing towards restaurants.

2. Customer Feedback Analysis

Customer feedback Analysis will be performed to identify negative feedback with higher frequency in a series of steps as follows

- Sentiment Analysis
- Topic Modeling
- Frequency Analysis

Data Preprocessing:

Before performing analysis on customer feedback data, a series of preprocessing steps have been performed as follows

- **Data Cleaning:** Checked and handled any missing or inconsistent data in the customer feedback dataset.

- **Text Preprocessing (For Topic modeling)**: Feedback text data were preprocessed by removing punctuation, converting text to lowercase, removing stopwords, and stemming or lemmatizing words to their base form.

- **Tokenization (For Topic modeling)**: Tokenize the preprocessed text into individual words or tokens.

Sentiment Analysis (refer)

Sentiment analysis is a natural language processing (NLP) technique used to determine the sentiment expressed in a text. The goal of sentiment analysis is to automatically extract subjective information from text and classify it as positive, negative, or neutral

- Python library TextBlob is used to analyze the sentiment of the feedback.
- Assigned a sentiment score to each feedback ranging from -1 to +1, where -1 means negative feedback and +1 means positive feedback and 0 means neutral feedback.

Negative Feedback Filter:

- After performing sentiment analysis, The feedback with negative sentiment scores was filtered out to perform topic modeling.

Topic Modeling

Latent Dirichlet Allocation (LDA) Modeling (refer):

Latent Dirichlet Allocation (LDA) is a generative probabilistic model used for topic modeling, a technique for discovering abstract topics or themes within a collection of documents. The Sklearn Python library will be used to perform LDA.

- **Hyper Parameter Tuning for number of topics (n_components) using coherence score (refer):**
    - To find the optimal number of topics in Latent Dirichlet Allocation (LDA) we have trained multiple models with n_components values ranging from (2 to 40)
    - We calculated coherence scores for different values of n_components models and selected the n_component parameter that maximizes coherence which is 3.
- **LDA Implementation:**
    - Latent Dirichlet Allocation (LDA) was performed with optimal parameters to label the similar feedback in the feedback data.
    - After LDA modeling, the frequency of each label was counted to identify the feedback that is most common among customers across business Types and locality.

Frequency Analysis

- After Performing LDA, The topics/labels with higher frequency will be observed.

3. Customer Churn Analysis (refer)

Customer churn analysis performed on daily onion sales data to identify churn rate given different time periods across different business types and locality. Customer churn analysis performed over daily sales data to understand and mitigate the risk to retain valuable customers.

Data Preprocessing:

- **Data Cleaning:** Checked and handled any missing or inconsistent data in daily sales dataset.
- **Data Merge:** Merged customer data personal data (Customer ID, Locality, Business Type) with each delivery made by the customer in sales dataset.

Risk Period / Customer Attrition Period Calculation :

- Churn rate was calculated between every 5 days period, Starting from day 0 to the 90th day to identify the Risk Period.
- The threshold for identifying the Risk Period will be visually determined by observing sudden drops in the bar chart and sustained trends in the churn rate over consecutive periods.

4. Customer Segmentation and Lifetime Value Analysis: (refer)

CLV (Customer Lifetime Value) Analysis is performed by segmenting customers by FM (Frequency and Monetary) values to identify the most valuable customer in terms of segmentation for each business.

Data Preprocessing:

- **Data Cleaning:** Checked and handled any missing or inconsistent data in daily sales dataset.
- **Data merge:** Merged customer data personal data like (Customer ID, Locality, Business Type) with each delivery made by the customer in sales dataset.

Customer Segmentation :

- **Frequency**: Counted the number of onion purchases made by each customer. This involves aggregating the data by customer ID and counting the number of unique purchase dates.
- **Monetary**: Calculated the total monetary value of purchases made by each customer. This requires multiplying the quantity of each purchase by its price and then summing these values for each customer.

Top Customers Visualization by segmented Lifetime value:

- The top 20 customers by segmentation Recency and frequency will be visualized in a bar chart and observed

5. Geospatial Analysis (refer):

Geospatial Analysis of customer data involves mapping individual customer locations to gain insights into business distribution and assess geographical coverage for strategic planning and targeted marketing efforts.

Data Preprocessing:

- **Data Cleaning:** Checked and handled any missing or inconsistent data like misspelled addresses in the customer dataset.
- **Data Transformation:** Using the Python library Geopy (refer), the Locality of customer location coordinates(longitude, latitude) will be identified and merged with customer data.

Plotting on the Map:

- The Python library folium(refer) is utilized to pinpoint the business in the map around the business locality with the coordinates identified using Geopy and observed.

Results and Findings

1. Customer Analysis:

- **Comparing Unique Customers by Business Type and Locality Each Month**

Note: The Yellow line separates the bar chart and compares the data before applying targeted marketing (Nov 2023 - Dec 2023) and After applying targeted marketing (Jan 2024 - Apr 2024) towards restaurants and localities.

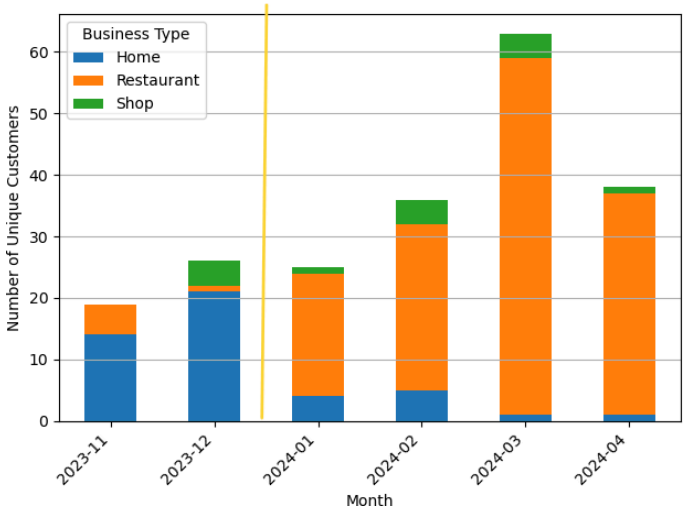


Figure 1.1

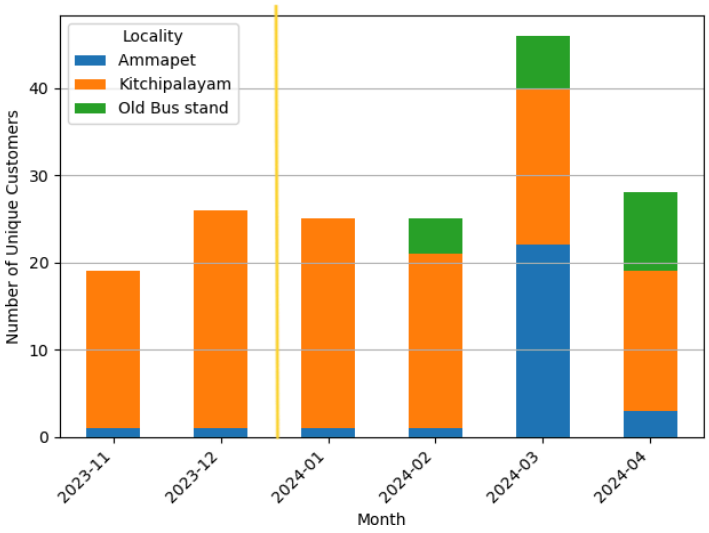


Figure 1.2

Figure 1.1, A stacked Bar chart illustrates the monthly distribution of unique customers across three distinct business types: Home-based, Restaurants, and Shops.

Before Applying Targeted Marketing Toward Restaurants:

- **November 2023**: Home-based establishments attracted the highest number of unique customers with 14, followed by restaurants with 5, while shops had no recorded customers.

- **December 2023**: Data saw a surge in overall customer numbers, with home-based establishments leading at 21, followed by shops with 4, and restaurants with 1 unique customer.

After Applying Targeted Marketing Toward Restaurants:

- **January 2024**: witnessed a shift in the trend, with restaurants dominating with 20 unique customers, followed by home-based establishments with 4, and shops with 1.

- **February 2024**: The previous month's pattern continued, with restaurants maintaining their lead at 27, followed by home-based establishments with 5, and shops with 4 unique customers

- **March 2024**: The restaurants experienced a remarkable spike in unique customers at 58, while home-based establishments and shops also observed slight increases.

- **April 2024**: restaurants sustained their lead with 36 unique customers, while home-based establishments and shops recorded 1 each.

.**Figure 1.2**, Stacked Bar chart illustrates the monthly distribution of unique customers across three distinct Localities.

Before Targeted Marketing Towards Specific Localities:

- **November 2023**: Kitchipalayam: Led with 18 unique customers, while Ammapet had only 1. Old Bus stand had no recorded customers.

- **December 2023**: Kitchipalayam: Maintained its lead with 25 unique customers. Ammapet had 1 customer, and the Old Bus stand remained without recorded customers.

After Targeted Marketing Towards Specific Localities:

- **January 2024**: Kitchipalayam: Maintained its lead with 24 unique customers. Ammapet had 1 customer, and the Old Bus stand remained without recorded customers.

- **February 2024**: Kitchipalayam: Continued to lead with 20 unique customers. Ammapet maintained 1 customer, while Old Bus stand recorded 4 unique customers.

- **March 2024**: Kitchipalayam: Sustained its lead with 18 unique customers. Ammapet observed a significant increase to 22 customers. Old Bus stand continued its growth with 6 unique customers.

- **April 2024**: Kitchipalayam: Led with 16 unique customers. Ammapet saw a decrease to 3 customers. The Old Bus stand maintained its growth, recording 9 unique customers.

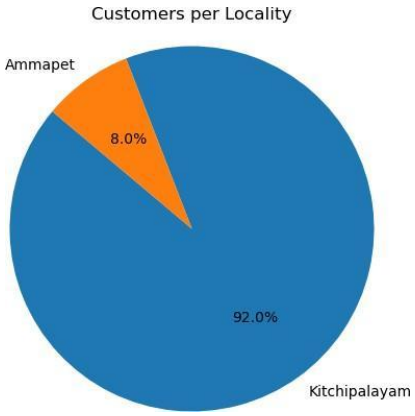
Conclusions:

Figure 1.1: Before applying targeted marketing towards restaurants, home-based establishments attracted the highest number of unique customers in November 2023, followed by restaurants and shops. However, after implementing targeted marketing towards restaurants, there was a notable shift in the trend, with restaurants dominating the customer count in January 2024 and subsequent months. This indicates the effectiveness of targeted marketing strategies in increasing customer engagement and retention, particularly in the restaurant sector.

Figure 1.2: Before targeted marketing towards specific localities, Kitchipalayam consistently led in customer count, followed by Ammapet and Old Bus stand. Following targeted marketing efforts, Kitchipalayam maintained its lead, while Ammapet experienced fluctuations in customer count and Old Bus stand saw significant growth. These findings suggest that targeted marketing initiatives can have varying impacts on customer engagement across different localities, highlighting the importance of localized marketing strategies for maximizing customer acquisition and retention

- **Comparing Overall Customer Share by Business Type and Locality**

Nov 2023 - Dec 2023



Jan 2024 - Apr 2024

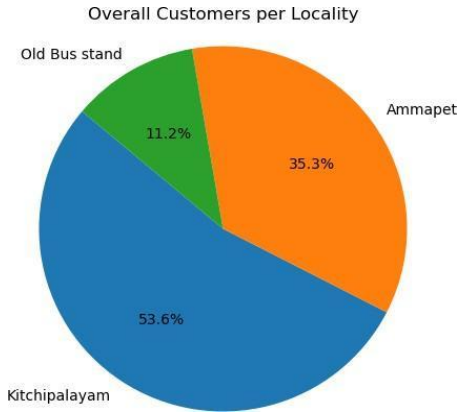
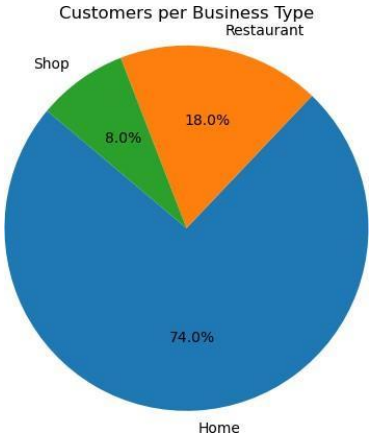


Figure 1.3

Nov 2023 - Dec 2023



Jan 2024 - Apr 2024

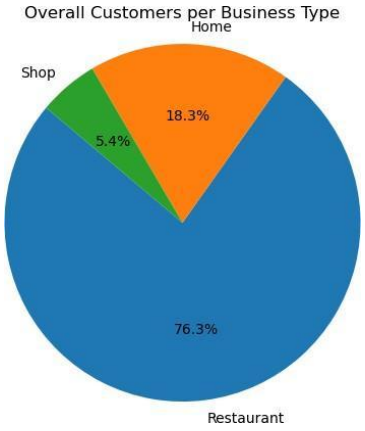


Figure 1.4

Customers Share by Locality:

Figure 1.3 The pie chart compares the overall customer shares by locality between the timeline (November 2023 - December 2023) and (January 2024 - April 2024)

November 2023 - December 2023 :

- In the initial analysis, 8% of customers are from Ammapet,92% of customers are from Kitchipalayam. This data underscores the notable concentration of customers within the Kitchipalayam area, suggesting its significance in terms of customer base

- The data indicated Kitchipalayam as a primary market hub. With only 8% of customers coming from Ammapet, it suggests a potential opportunity for business expansion or customer outreach to that locality.

January 2024 - April 2024 :

- The strategic focus on targeted customer outreach within specific localities has yielded significant business growth metrics.

- Notably, Ammapet experienced a notable surge in market share, escalating from 8% to 35.3%. Furthermore, the expansion efforts extended to include a previously untapped locality, Old Bus Stand, capturing an initial customer share of 11.2%.

- Despite these gains, Kitchipalayam retains its prominence as a pivotal contributor to overall customer share with 53.6%, demonstrating the enduring significance of established markets amidst expansion initiatives.

Customers Share by Business Type:

Figure 1.4 The pie chart compares the overall customer shares by Business Type between the timeline (November 2023 - December 2023) and (January 2024 - April 2024)

November 2023 - December 2023 :

- In the initial analysis, The restaurants comprised 18%, shops 8%, and homes 74% of the total customer base. This breakdown highlights a substantial portion of customers being residents (home-based), followed by restaurants and shops.

- Indicating a significant market segment that likely values convenience and delivery services. Additionally, while restaurants and shops constitute smaller portions of the customer base,
  there's still potential for growth and customer retention.

January 2024 - April 2024 :

- The implementation of targeted marketing initiatives directed at restaurants led to a substantial shift in customer distribution dynamics.

- Notably, there was a remarkable surge in the proportion of customers patronizing restaurants, soaring from 18% to 76.3%.

- Concurrently, there was a pronounced decline in the percentage of customers affiliated with home-based establishments, plummeting from 74% to 18.3%.

- Conversely, customers associated with shop categories witnessed a marginal reduction, declining from 8% to 5.4%.

- This analysis underscores the profound impact of targeted marketing strategies on reshaping customer preferences and market dynamics within the restaurant sector.

2. Customer Feedback Analysis:

Sentiment Analysis

The sentiment analysis conducted on the customer feedback dataset aimed to discern patterns of dissatisfaction and areas requiring improvement. Through the process, negative feedback was isolated, providing actionable insights to pinpoint specific aspects of the customer experience that warrant attention and enhancement.

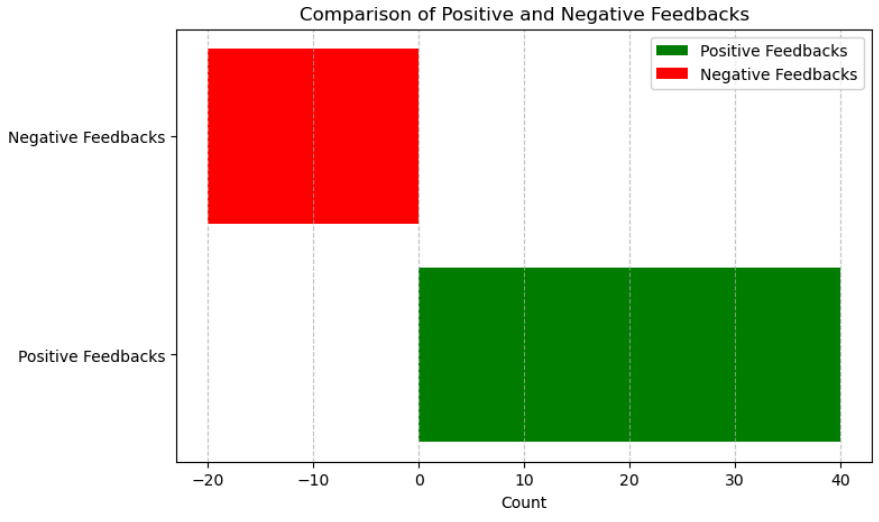


Figure 2.1

Figure 2.1, The diverging bar chart effectively contrasts the volume of positive and negative feedback within the onion delivery business, revealing a ratio of 2:1, with 40 instances of positive feedback and 20 instances of negative feedback. This visualization underscores the overall sentiment balance and highlights areas of strength and potential improvement in customer satisfaction.

Topic Modeling (LDA):

- Latent Dirichlet Allocation (LDA) is used to uncover underlying themes or labels within the corpus of negative feedback data, derived from sentiment analysis. By applying LDA, we aim
  to discern prevalent topics or issues raised by customers, facilitating a deeper understanding of their concerns and enabling targeted remedial actions for enhanced customer satisfaction.

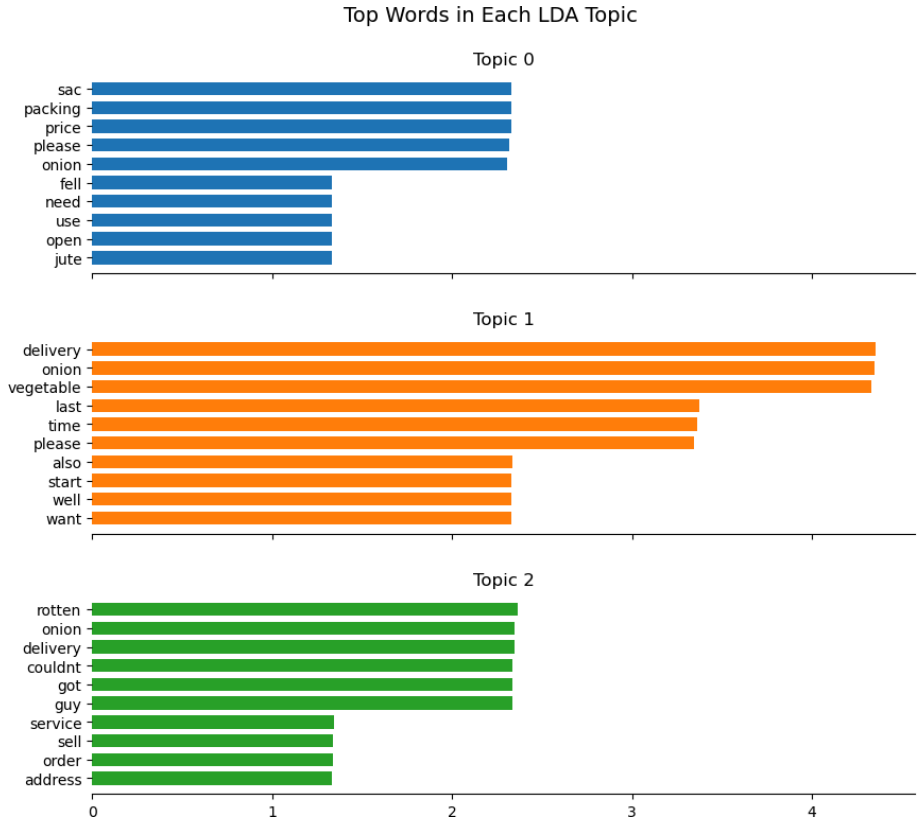


Figure 2.2

Figure 2.2, The bar plots illustrate each plot represents a topic generated by the LDA model. The bars in each plot correspond to the top words associated with that topic, with their respective weights.

LDA Interpretation :

- **Topic 0**: This topic appears to focus on issues related to packaging and pricing. Words like "sac," "packing," "price," "jute," and "open" indicate discussions about the packaging material used, pricing concerns, and suggestions for improvement in packaging methods.

- **Topic 1**: This topic seems to be related to general aspects of onion delivery and service. It includes words like "delivery," "onion," "vegetable," "please," and "well," which suggest discussions about the delivery process, the quality of onions, and customer requests or feedback about other vegetables.

- **Topic 2**: This topic revolves around complaints or feedback related to the quality of onions received. Words such as "rotten", "couldn't", "service", "sell" and "order" suggest instances where customers received onions of poor quality or faced issues with the delivery or ordering process.

These interpretations provide insights into the main themes or topics discussed in the feedback data. It appears that customers are discussing various aspects of the onion delivery service, including delivery experience, packaging quality, pricing concerns, and the quality of onions received.

Frequency Analysis:

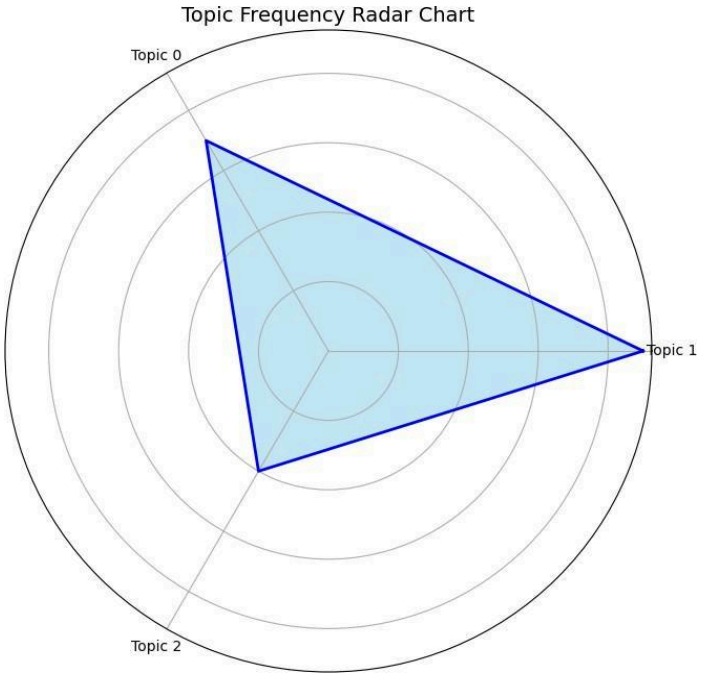


Figure 2.3

Figure 2.3, The radar chart illustrates that Topic 1 has the highest frequency of occurrences as 9, followed by Topic 0 as 7 and then Topic 2 as 4. The frequency distribution of topics helps prioritize areas for improvement or focus in addressing customer concerns. In this case, efforts may be directed towards addressing the issues discussed under Topic 1 more comprehensively due to its higher frequency of occurrence, followed by Topics 0 and 2.

Analysis Accuracy:

The accuracy of our topic modeling and frequency analysis can be compared with raw feedback data by manually grouping similar negative reviews and labeling them with customer concerns as follows:

- 10 similar reviews - Faster delivery (Topic 1)
- 7 similar reviews - Fresher quality of onions (Topic 1 and Topic 2)
- 6 similar reviews - Better packaging (Topic 0)

Feedback Analysis Conclusion :

- The analysis of customer feedback reveals distinct themes within the dataset.
- Topic 1 emerges as the most prominent, characterized by discussions surrounding delivery experiences, and onion quality assessments.
- Conversely, Topics 0 and 2 exhibit a lower prevalence, focusing on concerns such as packaging, pricing, and additional inquiries regarding onion quality.
- Prioritizing resolutions based on the frequency of customer needs facilitates efficient resource allocation, enabling the organization to address significant issues promptly, thereby enhancing service quality.

3. Customer Churn Analysis:

Customer churn analysis performed on daily onion sales data to identify churn rates given different time periods across different business types and localities. Customer churn analysis performed over daily sales data to understand and mitigate the risk of retaining valuable customers

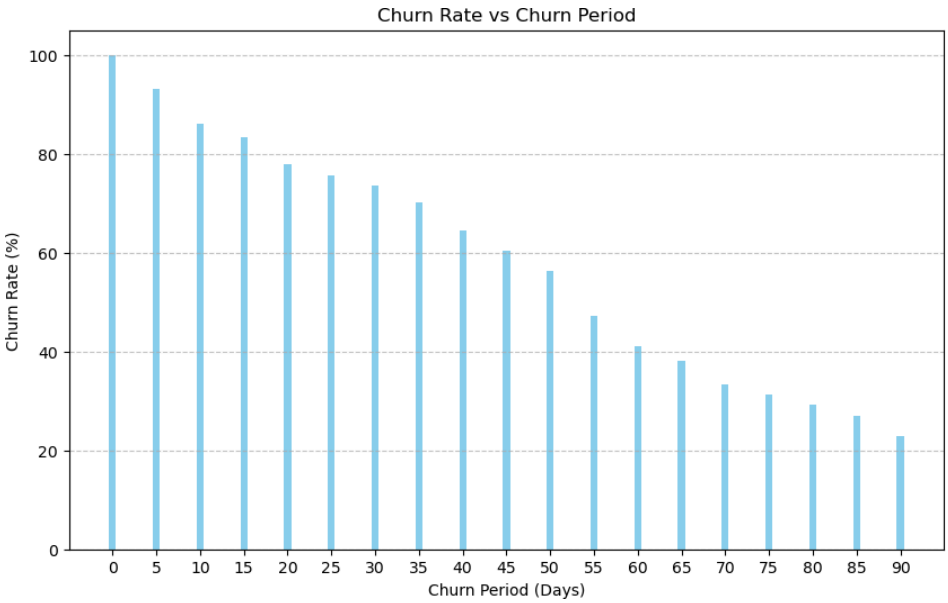


Figure 3.1

Figure 3.1, Bar chart illustrates the customer churn rates over a time period of 5 days between the 0th day to the 90th day.

Initial Observation:

- The churn rates start at 100% on day 0, As days progress, the churn rate decreases linearly.

Gradual Decline in Churn Rates:

- From day 0 to day 90, there is a consistent downward trend in churn rates, indicating an improvement in customer retention over time.

- This gradual decline in churn rates may suggest that the business is implementing effective retention strategies or improving the quality of its services, leading to increased customer satisfaction and loyalty.

Significant Drops in Churn Rates:

- Two notable drops in churn rates are observed:
    - Between day 11 and day 12: The churn rate decreases from approximately 56.25% to 47.22%.
    - Between day 13 and day 14: The churn rate decreases from approximately 47.22% to 40.97%.

- These significant drops may indicate periods of successful retention efforts or the introduction of new features or incentives that encourage customers to stay.

Risk Period Identification:

- The risk period, where the churn rates are relatively high, appears to be in the initial days of the observation period (day 0 to day 10).

- After day 10, the churn rates stabilized and gradually declined, indicating a reduction in customer attrition and a more stable customer base.

Insights:

- Overall, the data suggests a positive trend in customer retention, with churn rates decreasing over time.
- The significant drops in churn rates highlight specific time periods where the business may have implemented effective retention strategies or addressed underlying issues leading to customer churn.

4. Customer Lifetime Value Analysis:

- **CLV by Frequency:**

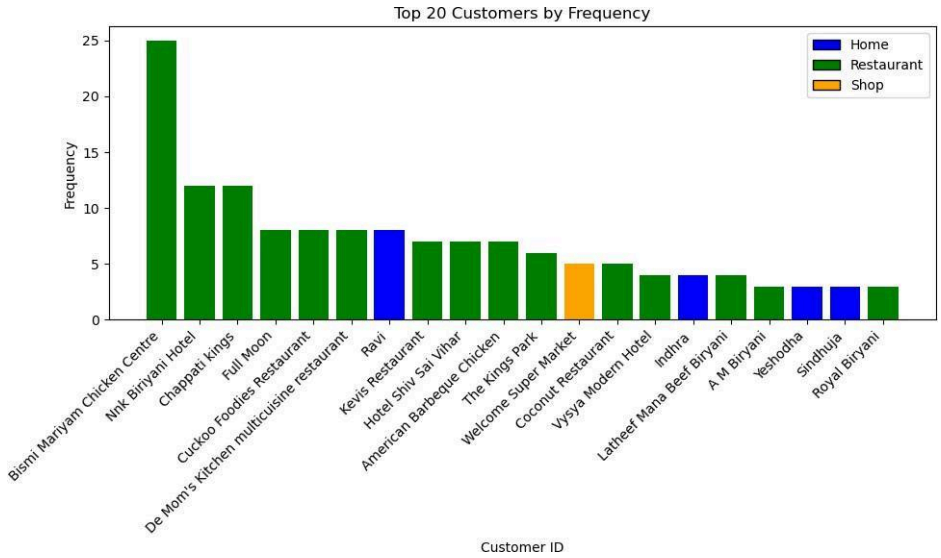


Figure 4.1

- **CLV by Monetary (Total Revenue as Monetary units):**

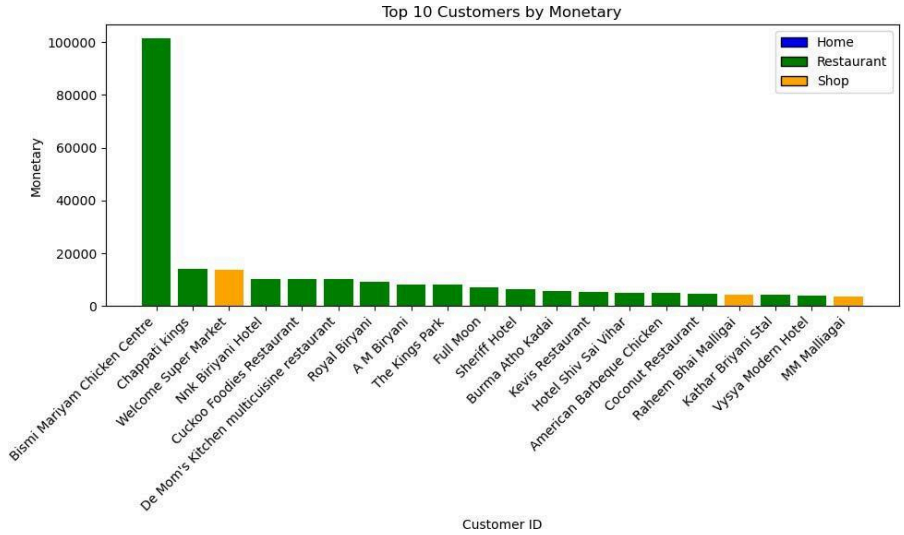


Figure 4.2

CLV By Frequency :

Figure 4.1 presents a bar chart depicting the top 20 customers based on their frequency of engagement with the business.

Upon analysis of the chart, it becomes evident that restaurants exhibit a notably higher frequency compared to other business types. Notable examples include "Bismi Mariyam Chicken Center" with a frequency count of 25, "NKN Biriyani Hotel" with a frequency count of 12, and "Chapati Kings" also with a frequency count of 12. This dominance underscores the significance of restaurants within the customer engagement landscape, highlighting their recurrent interactions and potentially strong brand loyalty.

CLV By Monetary:

Figure 4.2 illustrates a bar chart showcasing the top 20 customers ranked by their monetary contributions to the business.

Upon examination of the chart, it becomes apparent that once again, businesses in the restaurant sector exhibit dominance over other business types. Notably, "Bismi Mariyam Chicken Center" emerges as the top contributor in terms of monetary value, amounting to 101470 units. Following closely are "Chapati Kings" with a monetary value of 14166 units and "Welcome SuperMarket," representing the shop business type, with a monetary value of 13600 units. This trend underscores the substantial financial impact of restaurant-based establishments within the business ecosystem, suggesting robust patronage and potentially higher average transaction values compared to other sectors.

Conclusion:

Restaurants dominate in both frequency and monetary contributions, as evidenced by CLV analysis. "Bismi Mariyam Chicken Center" leads in monetary value, reaffirming the significant financial impact of restaurant-based businesses. This underscores their paramount importance for sustained growth and profitability within the business ecosystem.

5. Geospatial Analysis :

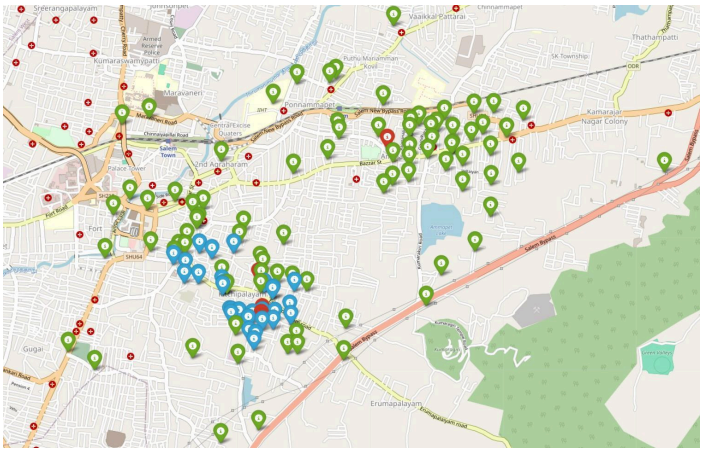


Figure 5.1

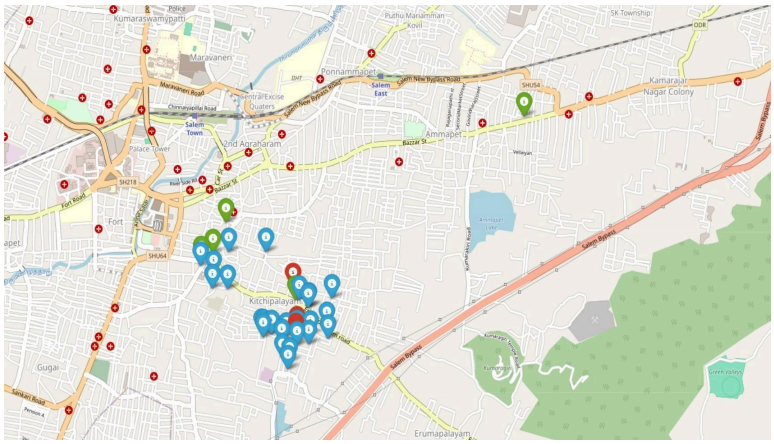


Figure 5.2

Figure 5.2 - Before Targeted Marketing (November 2023 - December 2023):

This visualization, generated using the folium library in Python, portrays the distribution of businesses across geographical regions in Salem. Blue pins denote residential areas, green pins represent restaurants, and red pins signify retail outlets. During the period of November to December 2023, there were a total of 34 customers analyzed.

Insights:

- A concentrated customer base is observed in the Kitchipalayam vicinity.
- There's a prevalent distribution of home-based customers in the area during this period.

Figure 5.1 - After Targeted Marketing (January 2024 - April 2024):

Similar to the previous visualization, this figure showcases the geographical distribution of businesses in Salem but for the period of January to April 2024. The color-coded pins remain consistent with blue for residential areas, green for restaurants, and red for retail outlets. The total number of customers analyzed in this timeframe is 148.

Insights:

- Following targeted marketing efforts directed towards restaurants in the Ammapet locality, there's a significant enhancement in the overall customer base.
- This suggests a successful impact of the marketing campaign in attracting more customers to the restaurant businesses in the specified locality.

Interpretation of Results and Recommendation

Results Interpretation:

1. **Customer Analysis and Targeted Marketing**: The implementation of targeted marketing towards restaurants resulted in a significant shift in customer distribution, with restaurants dominating the customer count in subsequent months. This indicates the effectiveness of
   targeted marketing strategies in increasing customer engagement and retention, particularly in the restaurant sector.

2. **Geospatial Analysis and Expansion Strategy**: Geospatial analysis revealed concentrated customer bases in specific localities like Kitchipalayam and Ammapet. This suggests
   opportunities for strategic expansion in areas with growing customer bases, such as Ammapet, while maintaining a strong presence in established markets like Kitchipalayam.

3. **Customer Feedback Analysis**: Negative feedback topics related to faster delivery, onion quality, and packaging were identified through sentiment analysis and topic modeling. Addressing these concerns is crucial for improving customer satisfaction and loyalty, thereby enhancing overall business performance.

4. **Customer Churn Analysis**: Gradual decline in churn rates over time suggests effective retention strategies and service improvements. Analyzing factors contributing to churn during the initial risk period can help identify areas for improvement and implement proactive measures to mitigate customer attrition.

5. **Customer Lifetime Value (CLV) Analysis**: Restaurants emerge as key contributors to both frequency and monetary value, highlighting their significance in driving business growth. Strengthening partnerships with high-value restaurant customers and tailoring offerings to meet their needs can further maximize CLV and foster long-term customer relationships.

Recommendations:

Targeted Marketing Strategy: Continue targeted marketing efforts towards restaurants, leveraging personalized promotional activities to enhance brand visibility and attract more customers. Explore opportunities for localized marketing campaigns in specific localities to further increase customer engagement and retention.

Expansion Strategy: Utilize insights from geospatial analysis to inform expansion strategies, focusing on areas with growing customer bases while maintaining a strong presence in established markets. Prioritize strategic partnerships with high-value customers, particularly in the restaurant sector, to drive sustainable revenue growth.

Customer Satisfaction Enhancement: Address customer concerns identified through feedback analysis by improving delivery processes, ensuring product quality, and enhancing packaging standards.
Implement continuous feedback mechanisms to gather insights into customer preferences and expectations, enabling proactive service improvements.

Retention Strategy: Implement data-driven retention strategies to minimize churn during the initial risk period, such as offering loyalty programs, personalized discounts, and proactive customer support.
Analyze churn patterns and customer feedback to identify root causes of attrition and implement targeted interventions to improve customer retention rates.

CLV Maximization: Foster strong partnerships with high-value customers, offering customized solutions and incentives to enhance their loyalty and increase their lifetime value to the business. Explore opportunities for upselling and cross-selling to maximize revenue potential from existing customer relationships and drive long-term profitability.

By implementing these recommendations, XYZ can overcome operational challenges, enhance customer satisfaction, and drive sustainable business growth in the competitive market landscape.