CAPSTONE PROJECT

FINAL REPORT

**Submitted by,**

**Group-2**

**1. Pandey, Utkarsh (N01578892)**

**2. Unnikrishna Pillai, Ramkrishna Panicker (N01598588)**

**3. Goyal, Nikhil (N01530197)**

**4. Jacob, Kiran Chacko (N01597568)**

**5. Kurian, Richard (N01598201)**

**6. Manoharan Sudha, Gokul (N01535489)**

**7. Irem Midillic (N01562987)**

**1. Executive summary**

This project aimed to enhance the efficiency of a restaurant’s front-of-the-house operations, addressing issues such as inaccurate staffing, inefficient table management, and missed marketing opportunities caused by intuition-based decision-making. These inefficiencies resulted in increased operational costs and poor customer experiences, impacting the restaurant’s profitability.

To address these challenges, the project employed a combination of machine learning models and interactive dashboards. The **customer segmentation model**, based on K-Means clustering, allowed the restaurant to group customers based on their behaviors, enabling tailored promotions and optimized staffing levels aimed at these specific clusters of customers. The **combo recommendation model**, powered by the Apriori algorithm, identified frequently paired menu items, facilitating targeted combo offers that enhanced customer satisfaction.

In addition to these models, **interactive dashboards** were developed to provide real-time insights into key operational areas such as staffing, table management, and customer traffic. These dashboards allowed restaurant managers to make data-driven decisions, improving staff allocation, predicting customer demand, and managing inventory effectively. The dashboards also supported strategic marketing efforts by providing data for personalized promotions.

The outcome of this project was a more data-driven approach to front-of-the-house operations, leading to optimized staffing, improved forecasting, better inventory management, and more effective marketing. By utilizing machine learning models and real-time dashboards, the restaurant was able to improve operational efficiency, customer experience, and profitability.

**2. Introduction**

This report will focus on addressing inefficiencies faced by a local restaurant named Johny’s Pizza in the front-of-the-house operations of a restaurant, with a particular focus on staffing, dining area management, and marketing. The primary issue under consideration is the reliance on intuition-based decision-making in these areas, which often leads to inaccurate staffing levels, poor table management, and missed opportunities for targeted marketing. These inefficiencies not only affect the customer experience but also lead to increased operational costs and lost revenue potential.

Section 1 will introduce the business problem in detail, providing a comprehensive understanding of the challenges faced by the restaurant industry in managing front-of-the-house operations effectively. Section 2 will address the key analytical questions that need to be attended by the project. Section 3 will contain the scope of the project.

Section 4 will include the data sources from where datasets containing the data about key operations about the restaurant can be accessed and downloaded. This section will also discuss about the key entities present in the datasets that pertains to the project. Section 5 will describe about different data manipulation, data preprocessing, data cleaning techniques and data output.

Section 6 will present the methodology used to develop a data-driven solution to address these challenges. This will include the customer segmentation model, implemented using K-Means clustering, and the combo recommendation model, based on the Apriori algorithm.

In Section 7, the business value of the proposed system will be discussed. This section will explain how the system can optimize staffing, improve forecasting for future demand, better manage inventory, and enhance marketing efforts by segmenting customers and recommending personalized offers.

Section 8 will provide a comparative analysis of the K-Means and Apriori algorithms, contrasting them with other available techniques, and explaining why they are the best fit for the restaurant’s needs.

Finally, Section 9 will detail the results of hyperparameter tuning for both models, demonstrating how adjustments in key parameters led to more accurate and effective predictions.

**3. Business problem overview**

The restaurant industry faces significant challenges in front-of-house operations due to reliance on intuition-based decision-making. Inefficiencies in staffing levels, table management, and marketing strategies result in increased operational costs, suboptimal customer experiences, and missed revenue opportunities. For example, inaccurate staffing can lead to either excessive labor costs or poor service quality, while ineffective table utilization hinders revenue maximization. Similarly, limited use of data-driven marketing leads to generic promotions that fail to retain customers or attract new ones.

These challenges directly impact profitability and customer satisfaction, posing risks to the restaurant’s competitiveness in a highly dynamic market. To address these issues, a shift toward data-driven decision-making is essential. By leveraging machine learning models and analytical dashboards, the restaurant can optimize staffing, enhance customer engagement, streamline operations, and improve overall business performance, ensuring sustainable growth and a superior customer experience.

**4. Analytics questions**

In this section, we outline the critical analytical questions that guide our project, focusing on leveraging advanced data analysis techniques to uncover actionable insights. These questions are designed to address Johny's operational challenges and growth opportunities by exploring key aspects of customer behavior, sales trends, and product performance. By framing these questions, we aim to establish a structured approach to data exploration, ensuring that the analyses align with the business objectives and provide meaningful outcomes. The subsequent questions serve as the foundation for implementing data-driven solutions that enhance decision-making and operational efficiency.

* How to analyze the organization wide performance?
* How to plan the services of the restaurant to cater and support the visiting customers
* How to allot priorities and manage the inventory?
* Are employees becoming more productive YoY in terms of sales, tips generated and paid hours?
* How the employees need to be planned in each section of the restaurant
* Which area is generating more revenue and tips and needs more development with respect to space?
* Does the restaurant have programs to retain or increase customer visits?
* How to make processes like combo creation more data driven?

**5. Scope statement**

This project focuses on leveraging advanced data analytics to optimize the front-of-house operations of a restaurant, addressing inefficiencies caused by intuition-based decision-making in staffing, table management, marketing, and promotions. By implementing machine learning models and interactive dashboards, the project aims to provide actionable insights that enhance operational efficiency, customer experience, and profitability.

Key deliverables include a customer segmentation model using K-Means clustering to classify customers based on Recency, Frequency, and Monetary (RFM) analysis and an Apriori-based combo recommendation model for identifying and promoting frequently ordered menu combinations. These tools will enable the restaurant to create tailored marketing campaigns, optimize menu offerings, and improve customer retention.

Additionally, the project includes dashboards that provide insights into staffing levels, table utilization, and revenue patterns, facilitating better resource planning. Predictive analytics will support inventory management by forecasting customer demand, reducing waste, and avoiding stockouts. The scope also covers analyzing employee productivity trends to improve labor allocation and identifying high-revenue areas requiring development.

The project emphasizes scalability and long-term applicability, enabling the restaurant to adapt to changing customer behaviors, optimize operations, and maintain a competitive edge while delivering a superior dining experience.

**6. Data sources/key data entities and flows**

For this business case, a variety of sources are used to collect data, with each source contributing valuable insights into the operations. The main data sources two specialized systems: SQUARE POS for item-wise sales data and RESY OS for customer visit details.

1. **SQUARE POS (Point of Sale System)**

Square is a widely used POS system in which item-level sales transaction data is captured in real-time. Detailed information about each sale, including the item sold, quantity, and transaction time, is provided. Integration with other systems is also supported by Square to offer a complete view of business operations. The final report generated from Square contains transaction-level details of the orders placed each day, along with information about payments, the assigned employee, customer IDs, and more. The data is maintained in a relational database, and for this analysis, the input used is the flat file in CSV format, which consists of a Fact table associated with several dimensional tables.

Files used from SQUARE

Fact Table

1. summary\_items\_detailed in .csv format

Dimension Tables

1. dim\_date in .csv format
2. dim\_employee\_master in .csv format
3. dim\_Itemmaster in .csv format
4. summary\_section in .csv format
5. **Resy OS (Restaurant Management Software)**

Resy OS is a restaurant management platform in which customer visit details, reservation management, and table bookings are tracked. Operations are streamlined and customer experiences are managed more effectively through the capture of customer details and their visits.

1. **Key Business entities**
2. **Customer**

**Attributes**: customer\_ID, customer\_name,email\_id,phone\_No

**Importance**: The number of customers visiting the store is relevant for all operational planning of the restaurant including staff, space and inventory

**Systems of Record (SOR):** RESY OS

**ii**. **Employee**

**Attributes**: EmployeeName,Area

**Importance**: Manpower planning is crucial for restaurants for ensuring better service, generating more sales and manage the space efficiently.

**Systems of Record (SOR)** : SQUARE POS

**iii**. **Item**

**Attributes:** ItemName, Category, CategoryRollup, Price

**Importance**: Item master gives the details of the items with the Categories and the latest price for making the transaction.

**Systems of Record (SOR):** SQUARE POS

A screenshot of a computer

Description automatically generated

**7. Brief overview of data manipulation process and data output**

**7. 1 Data Cleaning**

Raw data extracted from Square POS App and Resy OS Application over 2 years; cleaned, transformed, and merged into 5 final files. The 5 files are namely, ***Sales\_summary\_report.csv,*** ***manpower\_sales.csv,*** ***Customer\_activity.csv,*** ***Item\_Transaction.csv*** and ***summary\_section.csv.*** The data cleaning process executed to get these final finals will be explained below.

* Converted ‘**Date’** and ‘**Time’** to datetime objects; converted ‘**Gross\_Sales’** and ‘**Net\_Sales’** from strings to floats by removing dollar signs, commas, and handling parentheses for negatives.
* Corrected anomalies in ‘**Qty’** and ‘**Party\_Size’** (e.g., large negative values outside store hours; invalid product **'CUSTOM AMOUNT'** replaced with 0).
* Applied one-hot encoding to ‘**Area’** column, creating ‘**Area\_BACK’** and ‘**Area\_FRONT’** for later analysis.
* Dropped rows with **NULL** or invalid strings in ‘**CUSTOMER\_ID’** column.
* One-hot encoded **ITEMS** column in ‘summary\_items\_detailed.csv’ for Apriori model.
* Removed irrelevant variables from the original files for model development.

**7. 2 Data Output**

After the preliminary cleaning of the datasets, they have been consolidated to form 5 different csv file which is then used as the training data for our models which are **regression model, associative rule mining and clustering.** The 5 files are namely, ***Sales\_summary\_report.csv,*** ***manpower\_sales.csv,*** ***Customer\_activity.csv,*** ***Item\_Transaction.csv*** and ***summary\_section.csv.*** The files have been structured in a manner that they are model ready with data preprocessing methods like null value handling, standardization and so on have been applied.

Figure 2


**Figure 1**

Figure 1 shows the structure of **Sales\_summary\_report.csv**, created by consolidating three files: **summary\_items\_detailed.csv**, **dim\_date.csv**, and **Reservations.csv**. A groupby operation on ‘Date’ aggregated ‘Qty’ and ‘Net\_Sales’ in **summary\_items\_detailed.csv**, which was then merged with **dim\_date.csv**. **Reservations.csv** was also grouped by ‘Date’, with ‘Party\_Size’ aggregated to ‘No\_of\_customers’, and merged with the combined dataframe

A screenshot of a computer screen

Description automatically generated

Figure 2

Figure 2 shows **manpower\_sales.csv**, created by merging **dim\_employee\_master.csv** and **summary\_attendance.csv** on the ‘Name’ column. The ‘Area’ column was one-hot encoded, and the data was grouped by ‘Date’ to aggregate employee counts for the Back and Front of house. This dataframe was then merged with the **Sales\_summary\_report** dataframe from the previous step.

A screenshot of a computer screen

Description automatically generated

Figure 3

Figure 3 shows **customer\_activity.csv**, created by processing **summary\_items\_detailed.csv**. The relevant columns, ‘Customer\_ID’, ‘Net\_Sales’, and ‘Date’, were extracted. The dataframe was first grouped by ‘Customer\_ID’ and ‘Date’ to aggregate ‘Net\_Sales’. It was then grouped by ‘Customer\_ID’ again, calculating **No\_of\_Visits** (distinct ‘Date’ values), **Avg\_Monetary\_Value** (average ‘Net\_Sales’), and **Last\_visit\_date** (maximum ‘Date’ for each customer).A screenshot of a computer screen

Description automatically generated

Figure 4

Figure 4 shows **Item\_Transaction.csv**, created by processing **summary\_items\_detailed.csv**. Only the ‘TransactionID’ and ‘Item’ columns were retained. The ‘Item’ column was one-hot encoded, and the data was grouped by ‘Transaction\_ID’, with each transaction represented by a row and item presence indicated by 0 or 1 (or the number of occurrences for multiple items). The **summary\_section.csv** was used as is, with no additional preprocessing required.

**8. New solution design and it’s fit into the existing IT architecture**

Today, companies collect a huge amount of data through tools like Square, CRM, and other business technology apps. However, much of this data goes unused, missing the chance to uncover valuable insights. As Jeff Bezos once said, *“If you don’t understand the details of your business, you are going to fail”*, without fully using this information, businesses may overlook ways to understand their customers better, improve their operations, and stay ahead in the market. Companies can make smarter, more effective choices by focusing on data-driven decisions and using tools that turn data into clear insights,

Johny's Pizza, in the process of scaling up, decided to leverage years of data that had been sitting idle to support its decision-making. As data professionals, we have been tasked with building models and dashboards that utilize the data stored through two services—Square and ResyOS—that the restaurant has used for its daily operations. The restaurant is transitioning from making decisions solely based on intuition to making decisions driven by solid data insights.

**8.1 The Front-Of-The-House problem**

In the context of our study, the restaurant business can be divided into two that drive day-to-day activities they are:

1. Back-of-the-house: inventory management, kitchen operations, food quality control, logistics
2. Front-of-the-house: Customer service, Dining area management, staff management, customer feedback, marketing, and promotion.

Intuition-based decision-making in the front of the house for staffing, dining area management, and marketing and promotion can cause many problems like inaccurate staffing levels, inefficient table management, and very limited marketing opportunities. This leads us to the front-of-the-house problem

“Intuition-based decision-making in the front-of-house operations—specifically in staffing, dining area management, and marketing and promotion—leads to inefficiencies such as inaccurate staffing levels, suboptimal table management, and missed marketing opportunities. These challenges result in increased operational costs, poor customer service, and lost revenue potential, ultimately hindering the restaurant’s ability to deliver a consistent and exceptional customer experience while maximizing profitability."

Our proposed system addresses this problem with the help of models and dashboards, they are:

1. **Customer segmentation model**
2. **Combo recommender model**

Our proposed system provides business value in the form of:

1. **Optimize staffing levels**: Maintaining an optimum number of staff based on the volume of customers who might come into the restaurant during the hours of working will help reduce labor costs whilst improving service quality.
2. **Improved forecasting and long-term planning**: By gathering historical data on customer traffic and using it in the prediction model, the restaurant can improve its long-term planning. This could involve planning for future hiring, scaling operations, optimizing the layout, or introducing new services based on customer behavior trends
3. **Marketing and customer retention**: The Customer segmentation model and Optimum combo creation model will help us classify customers into different categories so that each category can have its tailored promotions which can help in building a loyal customer base, furthermore the proposed system can also help in creating baskets of items in the menu which are highly associated with each other which would help the restaurant to create attractive combo offers
4. **Better inventory management**: With the predicted volume of customers the system gives the restaurant an option to anticipate food demand more accurately, reducing waste and avoiding stockouts.
5. **Optimized table management and seating strategy**: With the insights from our dashboards the restaurant can optimize table assignments, ensuring that high-revenue tables or areas are utilized efficiently.

**8.2 Methodology**

1. **Customer segmentation model**
   1. **Model Overview**

Algorithm: Kmeans clustering

Module: sklearn.cluster

K-Means works well for RFM (Recency, Frequency, Monetary) analysis because it quickly groups customers with similar behaviors, like "high-value" or "inactive" customers. It's fast and can handle large datasets efficiently. However, K-Means assumes that the customer groups are round-shaped, so if the data has more complex patterns, other methods like DBSCAN or GMM might be a better fit.

* 1. **Model training and validation**

The model-ready dataset is made into clusters with 4 initial clusters. By using

the elbow method we find the optimal number of initial clusters. Then make our conclusions based on the result.

1. **Combo recommendation model**
2. **Model Overview**

Algorithm: Apriori

Module: mlxtend.frequent\_patterns

Apriori is useful for a restaurant's menu item association because it efficiently identifies frequent item combinations that customers often order together. This can help the restaurant understand customer preferences, optimize menu placement, and create targeted promotions or combo meals.

1. **Model training and validation**

The model-ready dataset with the transaction of each item in the restaurant is used to make rules of association. From this, we make a basket of goods based on the confidence level and lift value.

**8.3 Comparative Analysis**

1. **K-means clustering**

K-Means clustering is popular because it's simple, fast, and works well with large datasets. It groups data points into clusters by minimizing the distance between them and their cluster center, making it great for identifying natural groupings in data. This makes it easy to implement and understand, which is why it's often used for tasks like customer segmentation, where each group shares similar traits.

Compared to hierarchical clustering, which can be slow and resource-heavy with large data, K-Means is much more efficient and scalable. Also, unlike DBSCAN, which struggles with clusters of different densities and needs careful tuning, K-Means is simpler to use and requires fewer adjustments, making it a go-to for many clustering tasks.

1. **Apriori Algorithm**

Apriori is a great association algorithm because it’s easy to understand and works well for discovering patterns in transactional data. It finds frequent itemsets—groups of items that are often bought together—and then creates rules like "customers who buy X are likely to buy Y." This makes it especially useful for things like market basket analysis and upselling products.

While Apriori is simple to implement and interpret, it can be slower with very large datasets. However, it's still a solid choice for many businesses because it helps uncover valuable insights quickly. Other algorithms like FP-Growth may be faster for bigger datasets, but Apriori’s simplicity and effectiveness make it a go-to for many tasks.

**The fit of the new solution into the existing IT architecture**

The outcome of the project is a dashboard showing reports, metrics and KPIs, based on the data that is downloaded from both SQUARE POS endpoint and RESY OS endpoint.

The solution will be hosted in a cloud server as a different system which will be used by the restaurant. The integration with the existing architecture of the restaurant happens only at the data ingestion layer of this system where datasets will be downloaded from both the SQUARE POS endpoint and RESY OS endpoint into the cloud-based software.

Since the cloud-based solution is a stand-alone solution no additional integration is required beyond this data ingestion process, reducing additional overheads and complexity. This software will work without any dependencies or disruptions to the existing IT architecture of the restaurant.

**9. New solution implementation and outcome testing**

**9.1 Solution Implementation**

The solution implementation phase prioritizes enhancing the predictive model, conducting rigorous testing, and performing fine-tuning. Integration with the client’s systems will follow after foundational model development, once more resources and understanding are secured.

This phase addresses challenges faced by Johny Pizza Store, including managing expanded locations and a growing customer base. The reliance on manual documentation has led to inefficiencies and delays, impacting customer satisfaction. To counter these challenges, the proposed solution leverages Power BI for creating interactive dashboards and clustering techniques for customer segmentation, aimed at improving operational efficiency and customer engagement.

**Transforming Data Utilization**

The implementation phase transitions Johny's to a data-driven approach by replacing manual planning methods with actionable insights derived from:

**1.Power BI Dashboards**: Focused on visualizing key business metrics.

**2. Customer Segmentation**: Using clustering techniques to classify customers based on behavior.

This approach enhances decision-making and resource allocation, enabling management to identify underperforming areas and redirect efforts effectively. Personalization strategies based on customer segmentation foster engagement, retention, and profitability.

**Key Components**

1. Power BI for Business Intelligence:

* **Data Visualization**: Dynamic dashboards simplify the interpretation of complex data, facilitating strategic decision-making.
* **Integration with Diverse Data Sources**: Seamlessly connects with relational databases, cloud services, and files, unifying customer behavior, sales, and inventory data.
* **KPI Tracking**: Real-time monitoring of metrics ensures timely insights into operational and financial performance.
* **User-Friendly and Scalable**: Intuitive for non-technical users and adaptable for growing data volumes.

2. Customer Segmentation with K-Means Clustering:

* **RFM Analysis**: Segments customers based on Recency, Frequency, and Monetary value.
* **Interpretability**: Clear cluster centroids simplify identifying customer groups like "high spenders" or "lapsed customers."
* **Scalability**: Efficiently handles large datasets, ideal for dynamic segmentation.

**Supporting Models**

**1. Combo Creation Model:** Using the **Apriori algorithm** for market basket analysis, this model generates association rules to identify popular product combinations, enabling data-backed combo offers.

* **Key Metrics**: Rules with high confidence and lift provide actionable insights.
* **Customizable Thresholds**: Tailored to Johny’s sales data.

**2. Data Processing Pipeline:**

* **Data Connectivity and Cleaning**: Power BI and Python streamline data integration and preprocessing, ensuring consistency and readiness.
* **Feature Engineering**: Key features like Recency, Frequency, and Average Monetary Value are calculated and normalized for clustering.
* **Filtering and Validation**: Retains relevant customer data for actionable analysis.

**Training and Validation**

* **Interactive Dashboards**: Iteratively refined to ensure clarity, usability, and stakeholder alignment.
* **Clustering Model Tuning**: Optimal cluster counts determined via **Elbow Method** and **Silhouette Score**.
* **Apriori Algorithm**: Association rules validated with metrics like support, confidence, and lift.

**Expected Outcomes**

1. **Improved Operational Efficiency**: Visual analytics and KPIs enable data-driven resource allocation.
2. **Enhanced Customer Engagement**: Personalized offers based on customer segmentation.
3. **Scalable Solutions**: Ensures the system grows with Johny's business.
4. **Real-Time Insights**: Supports timely, strategic decision-making.

This solution positions Johny Pizza Store for sustainable growth, operational excellence, and customer satisfaction, establishing a strong foundation for future expansions.

**9.2 Outcome Testing**

This section analyzes **Dynamic Interactivity**, **Vertical-Specific Dashboards**, and **Customizable Visualizations** in Power BI, prioritizing their impact on decision-making, user engagement, and scalability. These features were chosen for their operational efficiency, ease of use, and ability to provide actionable insights, aligning with Johny's growing data requirements.

**Methodology**

The dataset, comprising item sales, customer interactions, and attendance, was preprocessed to ensure consistency. Dashboards were designed in Power BI to enhance interactivity and align with business goals.

* **Dynamic Interactivity**: Enabled deeper trend analysis through slicers, filters, and drill-down options.
* **Vertical-Specific Dashboards**: Customized for different business verticals, targeting relevant KPIs.
* **Customizable Visualizations**: Tailored visual elements effectively displayed KPIs like revenue growth and customer retention.

**Testing and Review**

1. **Usability**: User feedback highlighted accessibility and actionable insights.
2. **Scalability**: Power BI handled growing data seamlessly, accommodating new store locations.
3. **Performance Analysis**: Dashboards reliably reflected real-time data updates.

**Customer Segmentation with K-Means**

Using recency and visit frequency, K-Means segmented customers into four clusters, identifying behavioral patterns to enable targeted marketing and loyalty strategies.

**Combo Creation with Apriori Algorithm**

Frequent itemsets and association rules were generated with **min\_support=0.05** and confidence metrics. Of 25 rules, the top 5 (ranked by lift) revealed impactful product combinations, aiding in marketing, bundling, and inventory decisions.

These solutions streamline Johny’s operations, enabling data-driven decisions and sustainable growth.

**10. Potential solution optimization**

**Optimization of K-Means Clustering**

To improve the performance and accuracy of our K-Means clustering model, we focused on

optimizing key hyperparameters. Optimization is an essential step in ensuring that the

model produces meaningful and well-defined clusters that align with the inherent

structure of the data. By fine-tuning parameters such as the number of clusters (k) and

the initialization of centroids, we aimed to enhance the segmentation results and reduce

potential issues like poor convergence or overfitting.

**Hyperparameter Tuning**:

To determine the optimal number of clusters (k), we used two evaluation methods: the

Elbow Method and the Silhouette Score.

1. **Elbow Method**: The Elbow Method helps us determine the ideal number of clusters by

plotting the inertia (within-cluster sum of squares) against different values of k. From the

plot, we can observe a sharp drop in inertia for values of k between 2 and 4. After this

point, the decrease becomes more gradual, indicating that additional clusters do not

provide significant improvements in cluster formation. This suggests that k=4 is the

optimal number of clusters based on the Elbow Method.

2. **Silhouette Score**: The Silhouette Score measures the quality of the clusters by evaluating

how similar a data point is to its own cluster compared to other clusters. Higher

silhouette score indicates better-defined clusters. In our case, the Silhouette Score

shows the highest value at k=4, suggesting that the clusters are most distinct and

well-separated at this number. This reinforces the choice of k=4 as the optimal number

of clusters for segmentation.

**Defining Initial Centroids**:

In addition to optimizing k, we also focused on the initialization of cluster centroids, which

are the starting points for K-Means clustering. We used specific strategies for centroid

initialization to enhance the quality of the segmentation. By defining better initial

centroids, we aimed to reduce the likelihood of poor convergence, which could lead to

suboptimal clusters. This step is crucial as it helps the algorithm converge faster and with

a more accurate clustering outcome.

**Outcome**:

By fine-tuning the hyperparameters, specifically the number of clusters (k=4) and the

initialization of centroids, we achieved better segmentation, resulting in well-separated

and meaningful clusters that align with the dataset's underlying structure. The final

cluster configuration is expected to improve the segmentation and classification process,

ensuring more relevant and actionable insights from the data.

**Optimization of Apriori Algorithm for Association Rule Mining**

To optimize our Apriori algorithm for generating association rules, we focused on fine-tuning

two key hyperparameters: minimum support and minimum confidence. By adjusting

these parameters, we sought to balance the number of rules generated with their

quality, as measured by metr ics like lift.

**Hyperparameter Tuning**:

In our optimization process, we evaluated different combinations of minimum support and

minimum confidence. The goal was to identify the support and confidence levels that

maximize the number of high-quality rules—particularly those with a lift greater than 1.

1. **Minimum Support**: This hyperparameter controls the frequency threshold of itemsets

appearing in the dataset. A higher support value limits the number of itemsets

considered frequent, whereas a lower support value leads to more itemsets being

mined. We tested values of minimum support ranging from 0.01 to 0.10.

2. **Minimum Confidence**: Confidence is a measure of the likelihood that a rule holds true

given the antecedent. We varied the minimum confidence between 0.45 and 0.55. This

allows us to fine-tune the precision of the generated rules, focusing on rules with higher

reliability.

**Loop and Evaluation**:

We ran the Apriori algorithm across several combinations of these two parameters,

calculating the number of rules with lift greater than 1, the number of rules with lift less

than 1, and the maximum lift value for each combination.

**Outcome**:

The evaluation showed that with a minimum support of 0.01 and confidence between

0.45 and 0.53, there were many rules with lift greater than 1, reaching a maximum lift of

3.629357. However, despite the high number of such rules, a significant portion had lift

values below 1, suggesting less meaningful associations. After considering business

context, a minimum support of 0.05 and confidence of 0.45 were selected. This

combination provided a more manageable number of high-quality rules (25 rules with

lift >1), offering a better balance between relevance and frequency for actionable

insights.

**11. Appendix**

A screenshot of a computer screen

Description automatically generated

This report depicts the RFM model implementation. Here 4 distinct clusters can be seen well defined

A screenshot of a graph

Description automatically generated

The above report provides analytics about the sales data by providing KPIs like sales in Month-to-Date and sales of year-to-date and so on. Also, the report give insights about the sales details over category, Number of categories

A screenshot of a graph

Description automatically generated

The report provides insights about the employee strength for a specific time period and also provides insights about the effectiveness of the employees for the restaurant growth by providing important KPIs like Sales per hour, Average bill per employee and so on

A screenshot of a graph

Description automatically generated

This report focuses on the sales happening over specific areas in the restaurants. It visualizes the sales happening, and the tips generated at specific seating section in the restaurant.

A graph of sales comparison

Description automatically generated

This report an important visualization providing insights about the Year-over-year sales of specific items that are sold in the restaurant and provide insights on the importance of that item.

A screenshot of a computer

Description automatically generated

This report provides insights into customer interactions happening in the restaurant.