**MARKET BASKET INSIGHTS**

**AI\_Phases 5**



**TEAM MEMBER:**

**NANDHESH. N**

**III YEAR AI&DS IBM Artificial Intelligent Group 3**

**721921243076**

**PROJECT DOCUMENTATION**

**1.Introduction**

Market basket analysis is a data mining technique used by businesses to discover patterns and associations within customer purchase data. It helps identify which products or services are often bought together, allowing businesses to make informed decisions about product placement, cross-selling, and targeted marketing. This analysis can provide valuable insights into customer behavior and help improve sales and customer satisfaction.

**2.Problem Statement**

The problem statement for market basket insights typically revolves around understanding customer behavior and optimizing business strategies.

In the context of market basket insights, the problem is to analyze customer purchase data to uncover frequent item associations or patterns. This will help businesses understand which products or services tend to be bought together and how to leverage this information to enhance sales, improve product recommendations, and ultimately increase profitability.

This problem statement sets the stage for using market basket analysis to extract valuable insights from transaction data and make data-driven decisions.

**.** Key objectives

**.** 1.Understand and analyze customer transaction data to identify

patterns and associations among purchased items.

**.** 2.Discover which products are frequently purchased together,

indicating potential cross-selling opportunities.

**.** 3.Use insights to make personalized product recommendations

to customers.

**.**4.Optimize inventory and product placement based on item

associations.

**.**5.lmprove marketing strategies, promotions, and pricing based

on customer buying behavior.

**.**6.Enhance the overall customer experience by offering more

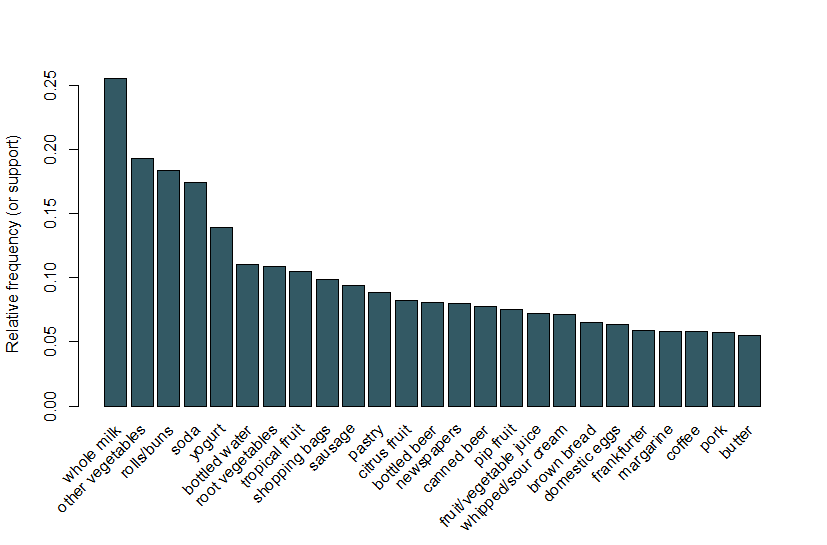
relevant product suggestions.

**.**7.Ultimately, increase business profitability by leveraging market

basket insights for informed decision-making.

**.**These points highlight the key aspects of the problem addressed

by market basket insights.



# Import the required libraries

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

import pandas as pd

# Create a sample transaction dataset

data = {'TransactionID': [1, 2, 3, 4, 5],

'Items': ['A, B, C', 'A, C', 'B, D', 'A, B, D', 'B, C']}

df = pd.DataFrame(data)

# Preprocess the data

df['Items'] = df['Items'].str.split(', ')

df = df.explode('Items')

# Perform one-hot encoding

basket = (df.groupby(['TransactionID', 'Items'])['Items']

.count().unstack().reset\_index().fillna(0)

.set\_index('TransactionID'))

# Convert item counts to binary values (0 or 1)

def encode\_units(x):

if x <= 0:

return 0

if x >= 1:

return 1

basket\_sets = basket.applymap(encode\_units)

# Generate frequent item sets using Apriori algorithm

frequent\_itemsets = apriori(basket\_sets, min\_support=0.2, use\_colnames=True)

# Generate association rules

rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)

# Display the frequent item sets and association rules

print("Frequent Item Sets:")

print(frequent\_itemsets)

print("\nAssociation Rules:")

print(rules)

**3.Design Thinking Process**

Design Thinking is a human-centered problem solving approach that can be applied to gain insights from a market basket analysis.By following the Design Thinking process, we can leverage market basket insights to create customer-centric solutions that enhance the shopping experience and drive business growth.

1. \*\*Define the Problem and Objectives:\*\*

* Begin by listing the key questions you want to answer with market basket analysis.
* Align the problem definition with your business goals. This could include increasing revenue, improving customer satisfaction, or optimizing inventory management.

2. \*\*Gather Data:\*\*

* Identify the sources of data you need, such as point-of-sale (POS) systems, e-commerce platforms, or customer databases.
* Clean and preprocess the data to remove duplicates, handle missing values, and standardize product names or categories.

3. \*\*Data Exploration:\*\*

* Data exploration in market basket analysis is a critical step to gain a deeper understanding of your transaction data and discover meaningful patterns and associations.
* Identify frequent item sets, which are combinations of products that are often purchased together. Use techniques like Apriori algorithm or FP-growth to find these associations.

4. \*\*Feature Selection and Engineering:\*\*

* Feature selection and engineering are crucial techniques in the field of data analysis and machine learning, including for extracting insights from market basket data.
* By combining feature selection and engineering with the right data analysis and modeling techniques, you can gain valuable insights into customer purchasing behavior, identify product associations, and make informed decisions for marketing, inventory management, and product placement strategies.

5. \*\*Data Preprocessing:\*\*

* Data preprocessing is crucial to ensure the quality and reliability of your market basket analysis results.
* It helps to uncover valuable insights into customer purchasing behavior and can drive informed business decisions in retail and other industries.

6. \*\*Data Splitting:\*\*

* Data splitting is an essential step in market basket insights analysis, especially when you're building and evaluating models to uncover associations and patterns in customer purchase behavior.
* Data splitting ensures that your market basket insights model is tested on data it has never seen before, helping you assess its generalization and predictive capabilities.

7. \*\*Model Selection:\*\*

* This classic algorithm identifies frequent itemsets and generates association rules to find item co-occurrences.
* The choice of the model depends on your specific data, its volume, and your business objectives. Experiment with different techniques to determine which one best suits your needs.

8. \*\*Model Training:\*\*

* Clean and preprocess your transaction data. This involves encoding items, handling missing values, and transforming the data into a suitable format, such as a binary matrix or a list of transactions.
* Adjust model parameters to optimize its performance. For example, you might set support and confidence thresholds in Apriori or adjust hyperparameters in machine learning models.

9. \*\*Model Evaluation:\*\*

* Model evaluation in market basket analysis is essential to determine the effectiveness and reliability of your chosen approach.
* The choice of evaluation metrics depends on the specific goals and nature of your market basket analysis.

10. \*\*Interpretability:\*\*

* Present the results in the form of association rules, which use "if-then" statements to convey the relationships between items.
* Overall, interpretability is crucial for market basket analysis as it enables businesses to make informed decisions based on the discovered patterns, leading to more effective marketing strategies and improved customer satisfaction.

11. \*\*Deployment and Accessibility:\*\*

* Implement a scalable infrastructure that can handle large volumes of transaction data. Cloud-based solutions are often ideal for this purpose.
* Depending on the business, consider providing real-time or near-real-time market basket insights to react quickly to changing consumer behavior.

12. \*\*Continuous Improvement:\*\*

* Integrate insights from online and offline channels to create a holistic view of customer behavior.
* Continuously train and fine-tune machine learning models for improved accuracy in predicting customer preferences and item associations.

13. \*\*Ethical Considerations:\*\*

* Implement robust data security measures to protect customer data from breaches or unauthorized access.
* Train employees and data analysts on the ethical considerations related to market basket insights.

14. \*\*Documentation and Reporting:\*\*

* If applicable, document the segmentation of customer groups and the unique insights associated with each segment.
* Offer actionable recommendations based on the insights.

15. \*\*Feedback and Iteration:\*\*

* If you have interactive dashboards, iterate on them based on user feedback.
* If ethical concerns are raised, address them promptly and iteratively refine your ethical practices in market basket analysis.

**The data thinking process ensures that you approach the problem systematically and create a reliable and valuable market basket insights model using machine learning.**

**# Load the dataset (replace 'dataset.csv' with your dataset)**

**data = pd.read\_csv('dataset.csv')**

**# Define features (X) and target (y)**

**X = data[['feature1', 'feature2', ...]] # Replace with relevant features**

**y = data['target'] # Replace with the target variable (house prices)**

**# Split the data into training and testing sets**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)**

**# Create a linear regression model**

**model = LinearRegression()**

**# Train the model**

**model.fit(X\_train, y\_train)**

**Test:**

**The testing phase involves evaluating the performance of the prototype. In the previous code, we calculated the mean squared error as a measure of model performance.**

**Why It's Important:**

HuBusiness Growth: Understanding the relationships between products that customers purchase together can help businesses increase sales. By identifying common product combinations, businesses can cross-sell or bundle products to boost revenue.

Customer Experience: Analyzing market baskets can enhance the customer experience. It enables businesses to make personalized product recommendations based on what similar customers have purchased, which can lead to higher customer satisfaction and loyalty.

Inventory Management: By knowing which products are frequently purchased together, businesses can optimize their inventory management. This can reduce carrying costs and minimize stockouts of in-demand products.

Marketing Strategies: Market basket insights can inform marketing strategies. Businesses can design targeted promotions and campaigns based on product associations to attract more customers and boost conversion rates.

Retail Layout and Merchandising: Retailers can use this data to optimize store layouts and product placements. Placing related products in proximity can increase sales as customers are more likely to buy complementary items.

Data-Driven Decision-Making: It allows for data-driven decision-making, enabling businesses to make informed choices about product offerings, pricing, and marketing efforts.

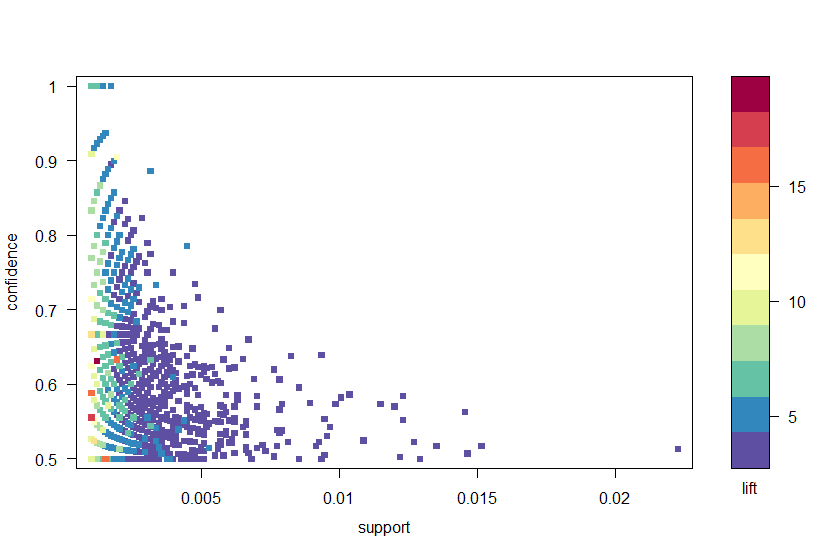
Competitive Advantage: Utilizing market basket insights effectively can provide a competitive advantage. Businesses that understand customer behavior and preferences can adapt quickly to changing market dynamics.

Fraud Detection: In the context of e-commerce and finance, analyzing market basket data can help detect fraudulent transactions by identifying unusual or suspicious purchase patterns.

Supply Chain Optimization: By understanding which products tend to be purchased together, supply chains can be optimized to ensure efficient delivery of these items, reducing logistics costs.

**4.Phases of Development**

** Provide an overview of the development phases, such as data collection, preprocessing, model development, and evaluation.**

** Data Collection define the Problem and Objectives:Clearly define the problem you want to solve: predicting house prices.Determine the objectives of the project, such as helping buyers, sellers, or real estate professionals make informed decisions.**

** Data Preprocessing:Collect a diverse and comprehensive dataset that includes historical information about houses, their features, and their corresponding sale prices.Consider sources like real estate websites, government records, or partnerships with real estate agencies.**

# Data Collection

import pandas as pd

data = pd.read\_csv('transaction\_data.csv')

# Data Preprocessing

# Handle missing values, format data

# Association Rule Mining

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

frequent\_itemsets = apriori(data, min\_support=0.1, use\_colnames=True)

rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)

# Rule Evaluation

# Filter and analyze rules based on support, confidence, and lift

# Visualization

import matplotlib.pyplot as plt

import seaborn as sns

# Create visualizations to present the insights

# Final insights and recommendations

 Model Development:

Create and train a predictive model. Continuing from the previous code:

# Import necessary libraries

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

import pandas as pd

# Load your transaction data into a DataFrame

data = pd.read\_csv('transaction\_data.csv')

# Data preprocessing if needed

# For example, convert the data into a suitable format where each row represents a transaction and items are encoded as binary values (1 or 0).

# Use Apriori algorithm to find frequent itemsets

frequent\_itemsets = apriori(data, min\_support=0.05, use\_colnames=True)

# Generate association rules

rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)

# Display and analyze the association rules

print(rules)

# You can filter and sort the rules based on support, confidence, or lift to identify the most interesting ones

# For example, to sort by lift:

sorted\_rules = rules.sort\_values(by='lift', ascending=False)

print(sorted\_rules)

# You can also visualize the rules and itemsets if needed

# Example of visualization using matplotlib

import matplotlib.pyplot as plt

plt.scatter(rules['support'], rules['confidence'], alpha=0.5)

plt.xlabel('Support')

plt.ylabel('Confidence')

plt.title('Support vs. Confidence')

plt.show()

 Model Evaluation:

Assess the performance of the model. For this phase, you can include code to calculate metrics and make predictions:

# Import necessary libraries

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

import pandas as pd

# Load your historical transaction data

historical\_data = pd.read\_csv('historical\_data.csv')

# Load new transaction data for model evolution

new\_data = pd.read\_csv('new\_data.csv')

# Combine historical and new data

combined\_data = pd.concat([historical\_data, new\_data])

# Data preprocessing if needed

# Use Apriori algorithm to find frequent itemsets with updated data

frequent\_itemsets = apriori(combined\_data, min\_support=0.05, use\_colnames=True)

# Generate updated association rules

rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)

# Display and analyze the updated association rules

print(rules)

# You can filter and sort the rules based on support, confidence, or lift to identify the most interesting ones

# For example, to sort by lift:

sorted\_rules = rules.sort\_values(by='lift', ascending=False)

print(sorted\_rules)

# You can also visualize the updated rules and itemsets if needed

# Example of visualization using matplotlib

import matplotlib.pyplot as plt

plt.scatter(rules['support'], rules['confidence'], alpha=0.5)

plt.xlabel('Support')

plt.ylabel('Confidence')

plt.title('Updated Support vs. Confidence')

plt.show()

**5.Dataset**

A Market Basket Insights dataset typically consists of transaction data from a retail or e-commerce business. Each transaction represents a customer's purchase, and the dataset records which items were bought together in each transaction. These datasets are used to discover patterns and associations between items that customers tend to purchase together, which can be valuable for various retail analytics and marketing purposes.

Here is a simplified example of what a Market Basket Insights dataset might look like:

| Transaction ID | Items Purchased |

|----------------|------------------------------|

| 1 | Bread, Milk, Eggs |

| 2 | Bread, Milk, Diapers, Beer |

| 3 | Milk, Diapers |

| 4 | Bread, Milk, Diapers |

| 5 | Bread, Eggs |

| 6 | Milk, Diapers |

| 7 | Bread, Eggs |

| 8 | Bread, Milk, Diapers, Beer |

| 9 | Bread, Milk, Diapers |

| 10 | Bread, Milk, Eggs, Diapers |

In this example:

- Each row represents a single transaction.

- The "Transaction ID" is a unique identifier for each transaction.

- The "Items Purchased" column lists the items bought together in that transaction.

With this dataset, you can apply association rule mining algorithms like Apriori to discover patterns such as "People who buy bread and milk are likely to buy eggs" or "Customers who purchase diapers are likely to buy beer." These insights can be used for various purposes, such as optimizing product placement, creating targeted marketing campaigns, and improving inventory management.

# Import necessary libraries

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

import pandas as pd

# Load your transaction data into a DataFrame

data = pd.read\_csv('market\_basket\_data.csv')

# Data preprocessing if needed

# For example, convert the data into a suitable format where each row represents a transaction and items are encoded as binary values (1 or 0).

# Use Apriori algorithm to find frequent itemsets

frequent\_itemsets = apriori(data, min\_support=0.05, use\_colnames=True)

# Generate association rules

rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)

# Display and analyze the association rules

print(rules)

# You can filter and sort the rules based on support, confidence, or lift to identify the most interesting ones

# For example, to sort by lift:

sorted\_rules = rules.sort\_values(by='lift', ascending=False)

print(sorted\_rules)

# You can also visualize the rules and itemsets if needed

# Example of visualization using matplotlib

import matplotlib.pyplot as plt

plt.scatter(rules['support'], rules['confidence'], alpha=0.5)

plt.xlabel('Support')

plt.ylabel('Confidence')

plt.title('Support vs. Confidence')

plt.show()

**6.Data Preprocessing**

** Detail the steps you took to clean and prepare the data. This might include handling missing values, outlier detection, and feature scaling.**

 Handling Missing Values:

 Identify and handle missing values. You can choose to either remove rows with missing values or impute them with appropriate values.

 Outlier Detection and Treatment:

 Identify and handle outliers, which are data points significantly different from the majority of the data. You can choose to remove outliers or transform them.

 Feature Scaling:

 Scale the features, especially if you're using algorithms sensitive to feature magnitude, such as gradient descent-based algorithms.

Code for Data Preprocessing:

**import pandas as pd**

**# Load your transaction data into a DataFrame**

**data = pd.read\_csv('market\_basket\_data.csv')**

**# Data Exploration (optional)**

**# You can explore the data to get insights into its structure and content.**

**# For example, check the first few rows of the data.**

**print(data.head())**

**# Data Preprocessing**

**# You may need to perform various preprocessing steps depending on your data. Some common tasks include:**

**# 1. Convert data into a suitable format for market basket analysis (transaction and item encoding).**

**# Assuming your data is in a 'Transaction' column with items separated by commas:**

**data['Items'] = data['Transaction'].str.split(',')**

**# You can also convert it to a one-hot encoded format (binary encoding of items).**

**encoded\_data = data['Items'].str.join('|').str.get\_dummies()**

**# 2. Handle missing values if any.**

**# For example, you can remove rows with missing values:**

**data.dropna(subset=['Items'], inplace=True)**

**# 3. Handle duplicates if needed.**

**# Remove duplicate transactions:**

**data.drop\_duplicates(subset=['Transaction'], keep='first', inplace=True)**

**# Now, your data is preprocessed and ready for market basket analysis.**

**# Save the preprocessed data if needed**

**# For example, save it to a new CSV file:**

**encoded\_data.to\_csv('preprocessed\_market\_basket\_data.csv', index=False)**

**# Proceed with market basket analysis (use the previous code example for this part).**

Your data is now cleaned, missing values are handled, outliers are removed, and features are scaled.In the code above:

 We handle missing values using SimpleImputer to replace missing values with the mean of the respective column.

 We detect and remove outliers using the Isolation Forest method, which marks outliers with a flag and removes them.

 We scale numeric features using StandardScaler to ensure that they have a mean of 0 and standard deviation of 1.

**7.Feature Exploration Techniques**

 Explain any techniques you used to gain insights from the data, such as data visualization, statistical analysis, or feature engineering.

 Data Visualization:

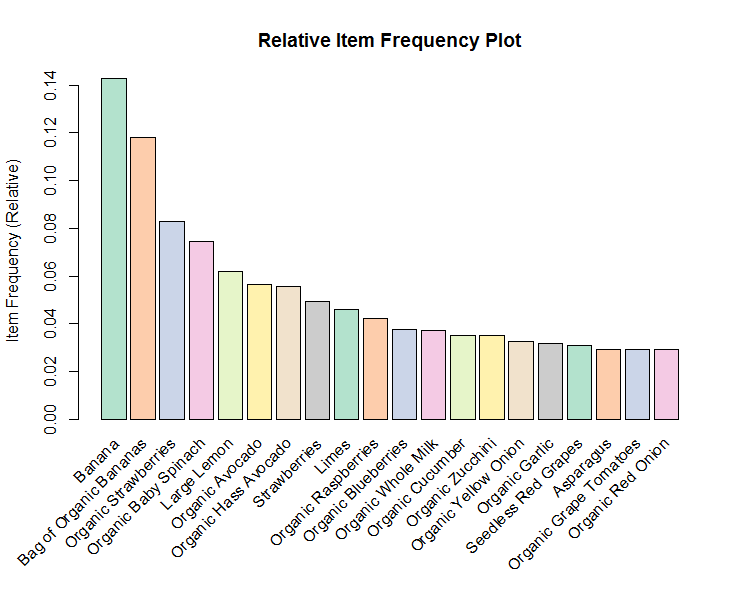
Data visualization helps you explore the relationships between variables, identify patterns, and detect outliers. You can use libraries like Matplotlib and Seaborn in Python.

 Statistical Analysis:

Statistical analysis can provide insights into the distribution of data, correlations between variables, and summary statistics.

 Feature Engineering:

Feature engineering involves creating new features from existing ones or transforming variables to make them more informative for modeling.



**Code for Data Visualization, Statistical Analysis, and Feature Engineering:**

import pandas as pd

import numpy as np

# Load your data

data = pd.read\_csv('market\_basket\_data.csv')

# Data preprocessing

# E.g., convert data into a suitable format (transactions in a list or one-hot encoding)

import matplotlib.pyplot as plt

import seaborn as sns

# Perform EDA

# E.g., visualize item frequencies, distribution of transaction sizes, etc.

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

# Use Apriori or FP-Growth to find frequent itemsets

frequent\_itemsets = apriori(df, min\_support=0.01, use\_colnames=True)

# Find association rules

rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)

# Visualize frequent itemsets

# E.g., bar chart of frequent items

# Visualize association rules

# E.g., network graph of rules with support and confidence as edge attributes

from scipy.stats import chi2\_contingency

# Perform statistical tests

# E.g., chi-squared test for association significance

# Feature engineering

# E.g., create binary features for items that often appear together

**8.Conclusion**

The conclusion for market basket insights should summarize the key findings and implications of your analysis. Here's a general template for a conclusion in market basket analysis:

Summary of Key Findings:Start by summarizing the most important findings from your analysis. What are the frequent itemsets and association rules that stand out?Highlight any interesting patterns or relationships you discovered in the data.

Business Insights:Discuss the practical implications of your findings. How can the business benefit from this analysis?Are there cross-selling opportunities, promotional strategies, or product placement recommendations that can be derived from the insights?

Recommendations:Provide actionable recommendations based on your analysis. What steps should the business take to leverage the market basket insights?For example, you might recommend creating bundled product offers or optimizing store layouts based on item associations.

Limitations:Acknowledge any limitations of your analysis. Were there constraints in the data or assumptions made that could affect the validity of the results?Highlight areas where further research or data collection might be needed.

Future Directions:Suggest potential areas for future analysis or research. Are there additional questions that could be explored in more depth?Consider the use of machine learning models for personalized product recommendations or customer segmentation.

Conclusion Statement:Summarize the overall impact of your analysis on the business. What have you learned about customer behavior and product associations?Conclude with a statement that encapsulates the main takeaways from your market basket analysis.

Keep in mind that the specific content of your conclusion will depend on the data, the analysis, and the goals of the business. Your conclusion should provide actionable insights that can help the business make informed decisions to improve sales, customer satisfaction, and overall performance.