**MARKET BASKET INSIGHTS: UNVEILING CUSTOMER BEHAVIOR THROUGH ASSOCIATION ANALYSIS**

**Artificial Intelligence**



**V. R. S. College of Engineering and Technology**

(Reaccredited by NAAC and An ISO 9001:2008 Receritfied Institution)

Arasur-607107, Villupuram District.

**PRESENTED BY,**

B. Sangamithra - 422721104041

D. Sivitha - 422721104049

A. Karishma - 422721104017

P. Nandhini - 422721104027

|  |  |  |
| --- | --- | --- |
| S. NO | TABLE OF CONTENT | Page no |
| 1. | List of Figure | 3 |
| 2. | Keywords | 4 |
| 3. | Scope of the project | 5 |
| 4. | Abstract | 6 |
| 5. | Introduction | 7 |
| 6. | Problem Definition | 8 |
| 7. | Design Thinking | 9 |
| 8. | Stages of Design Thinking | 10 |
| 9. | Dataset | 11 |
| 10. | General Overview | 12 |
| 11. | Data Preprocessing Steps | 13 - `19 |
| 12. | MachineLearning Techniques | 20 - 21 |
| 13. | Decision Tree | 22 - 24 |
| 14. | Conclusion | 25 |
| 15. | References | 26 |

|  |  |  |
| --- | --- | --- |
| S.No | LIST OF FIGURES | Page No |
| 1. | Design Thinking | 9 |
| 2. | Stages of Design Thinking | 10 |

**KEYWORDS**

1. Market Basket Analysis
2. Association Analysis
3. Customer Behavior
4. Apriori Algorithm
5. Retail Business
6. Cross-selling Opportunities
7. Data Exploration
8. Visualization
9. Data Analysis
10. Data Preprocessing
11. Machine Learning
12. Decision Tree

**SCOPE OF THE PROJECT:**

* The main aim of our project is Unveiling customer behavior through association analysis.
* Conduct market basket analysis on a given dataset to uncover customer purchasing patterns and identify potential cross-selling opportunities for a retail business. The goal is to provide actionable insights that enhance business strategies.

**ABSTRACT**

* This project, titled "Market Basket Insights - Unveiling Customer Behavior through Association Analysis," focuses on leveraging market basket analysis to extract valuable insights from a provided dataset.
* The objective is to uncover hidden patterns and associations between products, shedding light on customer purchasing behavior. By employing the Apriori algorithm, the analysis aims to generate actionable recommendations for the retail business, particularly in the identification of cross-selling opportunities.
* The project encompasses data exploration, association analysis, insights generation, visualization, and the delivery of strategic recommendations, contributing to a deeper understanding of customer behavior and facilitating informed decision-making for the retail sector.

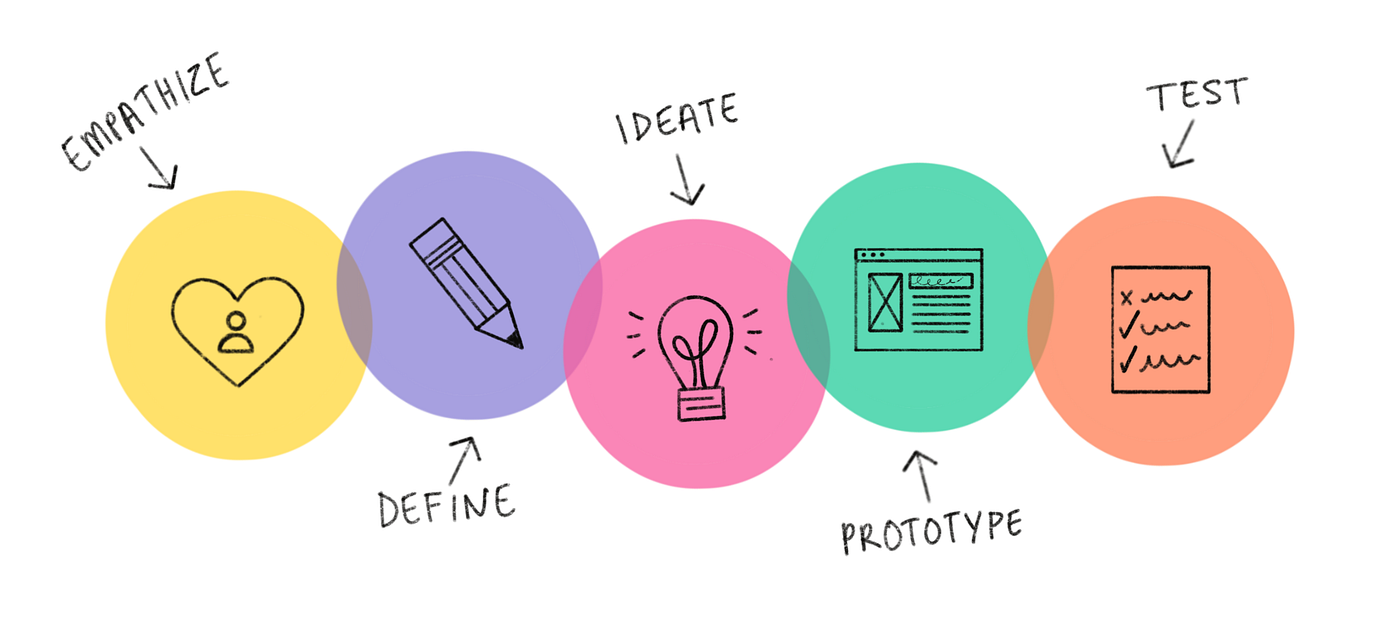
**INTRODUCTION**

* In today's competitive market, the ability to identify cross-selling opportunities is a strategic advantage for any retail business. The analysis goes beyond conventional approaches, incorporating data exploration, insights generation, and visualization techniques to present a holistic understanding of customer interactions. By applying advanced analytics, the project not only uncovers hidden patterns in the dataset but also aims to translate these findings into actionable recommendations.
* Through a combination of machine learning techniques and strategic analytics, "Market Basket Insights" strives to be a valuable resource for retail businesses seeking to enhance customer engagement, maximize revenue, and thrive in a competitive retail environment.

**PROBLEM DEFINITION**

The problem is to build an AI-powered diabetes prediction system that uses machine learning algorithms to analyze medical data and predict the likelihood of an individual developing diabetes. The system aims to provide early risk assessment and personalized preventive measures, allowing individuals to take proactive actions to manage their health.

**DESIGN THINKING**



**STAGES OF DESIGN THINKING**

**What the user Feels:** Emotional responses, such as frustration, excitement, or satisfaction.

**What the user Says**: Direct quotes and statements from customers during interviews or surveys.

**What the user Does:** Observable behaviors, actions, and interactions.

**What the user Thinks**: The inner thoughts and motivations of customers.

**DATASET**

**Dataset Source URL:** <https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis>

**Code Structure:** The Jupyter Notebook is structured into well organized code cells, each serving a specific purpose. The code is documented with comments, explaining the functionality of each cell and the overcall logic of the data cleaning and further processes.

**Updated Dataset:** As the original dataset cannot also be clean enough in order take that dataset into Data Analysis Phase. I have cleaned the dataset and stored the cleaned dataset as an “updated\_dataset” in xlsv file extension.

**GENERAL OVERVIEW**

* The project aims to analyze transaction data from a retail business to uncover hidden patterns and associations between products. By employing association analysis techniques, such as the Apriori algorithm, the goal is to gain valuable insights into customer purchasing behavior and identify cross-selling opportunities. Market basket analysis with Apriori algorithm, The retailer wants to target customers with suggestions on itemset that a customer is most likely to purchase .
* The given dataset contains data of a retailer; the transaction data provides data around all the transactions that have happened over a period of time.
* we will be able increase customer engagement and improve customer experience and identify customer behavior. I will solve this problem with use Association Rules type of unsupervised learning technique that checks for the dependency of one data item on another data item.

**DATASET PREPROCESSING STEPS**

**Data Collection**: Gather transactional data that includes information about items purchased together.

**Data Cleaning:**

* Handle missing values if any.
* Remove duplicate transactions or items.
* Ensure consistency in item names and encoding.

**Transaction Consolidation:**Group transactions by a unique identifier (e.g., customer ID or transaction ID).

**Filtering Rules:**Apply metrics like support, confidence, and lift to filter out rules that are not statistically significant or do not meet certain thresholds.

Step: 1 Import the necessary library

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

Step: 2 Import the the association rules.

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

Step: 3 Upload the dataset

from google.colab import files

data = files.upload()

Saving Assignment-1\_Data.xlsx to Assignment-1\_Data.xlsx

Step: 4 Read the dataset using

df = pd.read\_excel('Assignment-1\_Data.xlsx')

print(df)

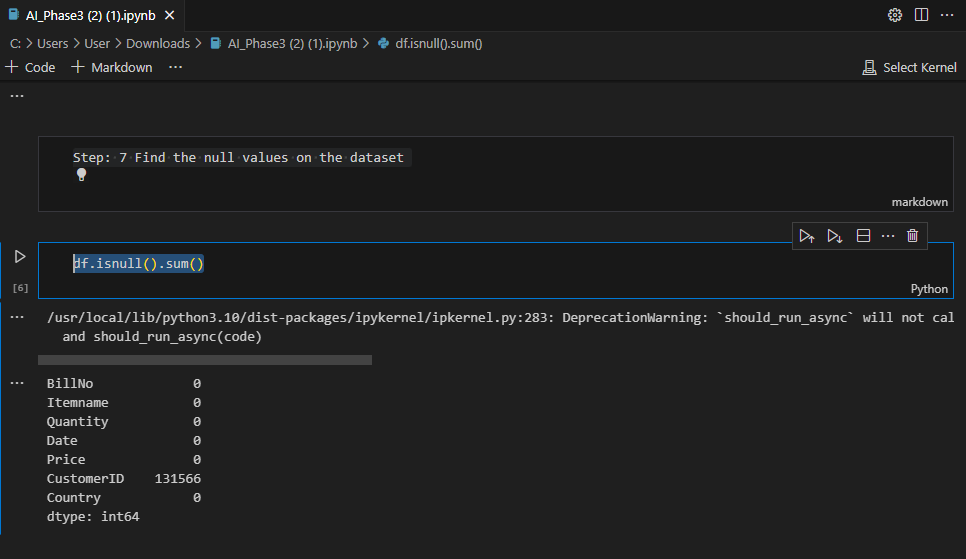
**output:**

Step: 5 Display the particular rows and  columns in a dataset.

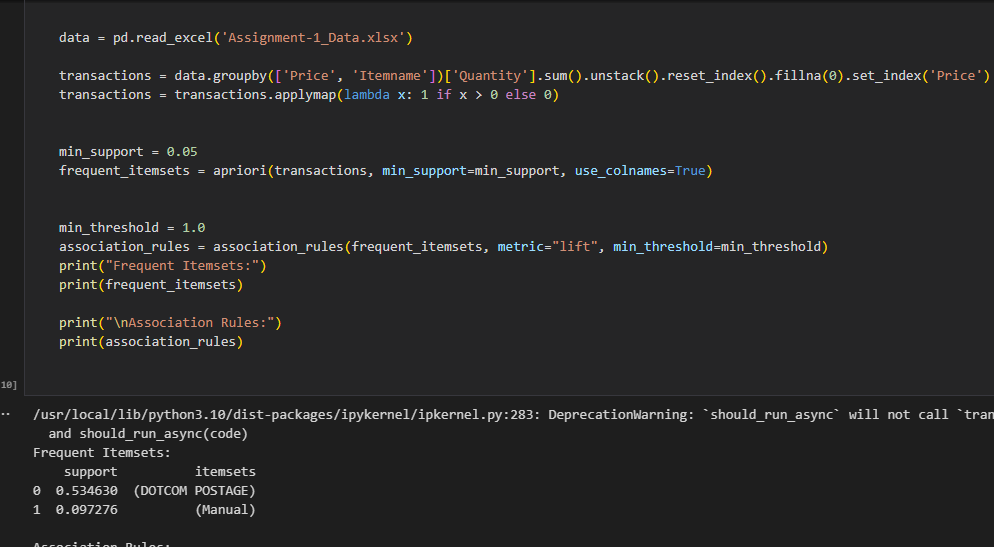
df.head(20)

Step: 7 Find the null values on the dataset

df.isnull().sum()

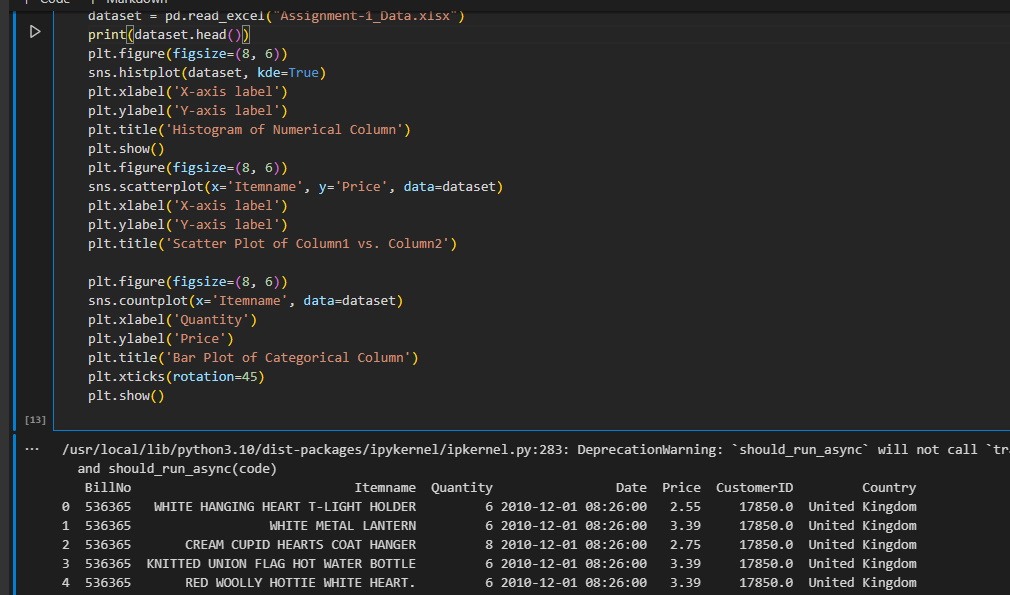
****

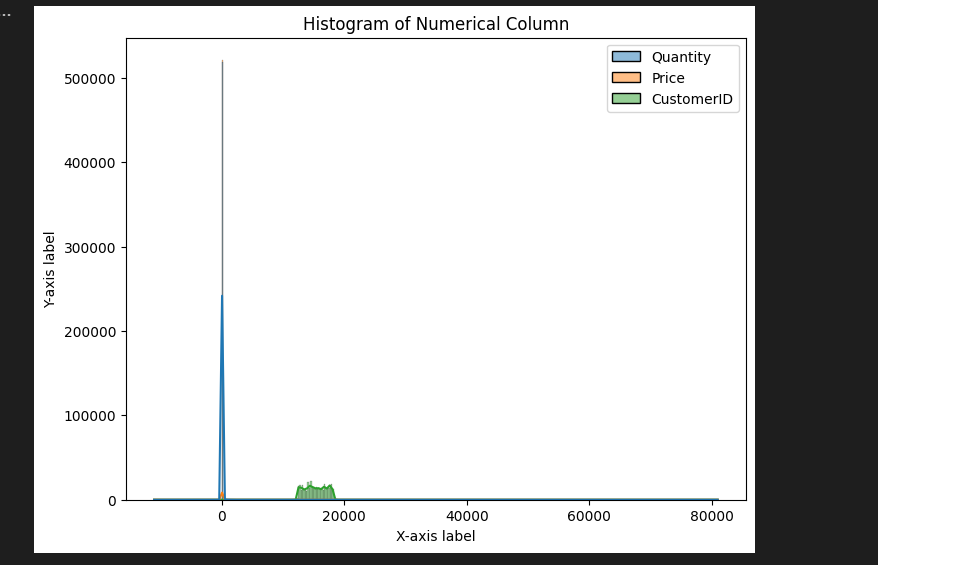
Step: 8 Display the frequency items from the dataset using association rules.

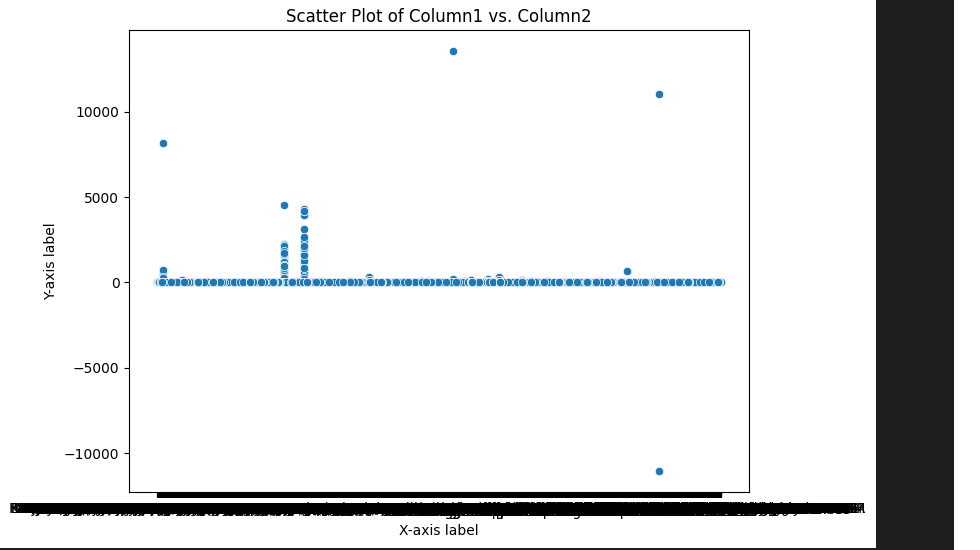
****

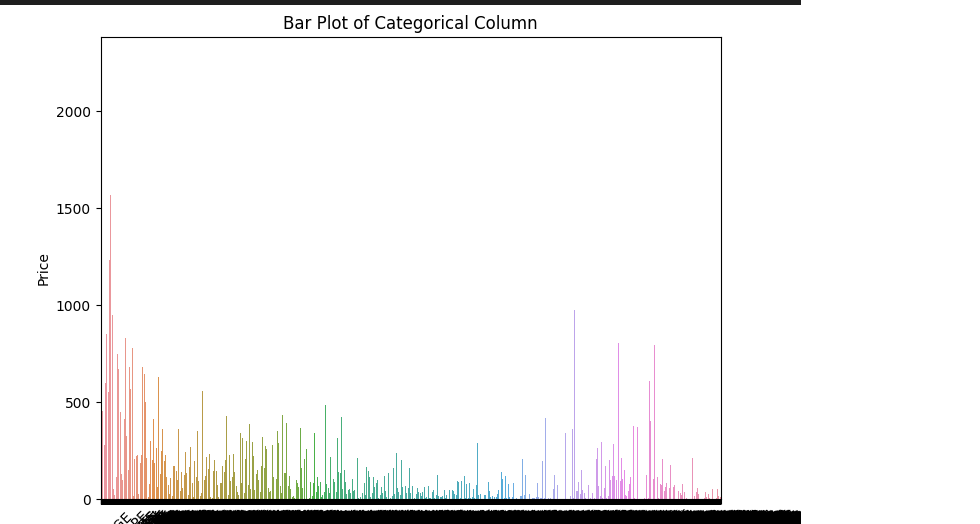
**VISUALIZATION PROCESS STEP**

"Visualization in our project refers to the art and science of representing complex market basket data in a visually intuitive and meaningful way. Through the use of various graphical techniques such as heatmaps, network diagrams, bar charts, and interactive dashboards, we aim to transform raw transactional data into insightful visual representations. These visualizations serve as a powerful tool for unraveling patterns, relationships, and trends within the market basket, providing a clear and accessible means for stakeholders to comprehend and act upon the underlying insights derived from the data."



****

****

****

**MARKET BASKET INSIGHTS USING TECHNIQUES**

**Numpy:**

**Data Manipulation:** Numpy provides powerful tools for working with arrays and matrices. You can use it to efficiently manipulate and process your market basket data.

**Array Operations**: Since market basket data is often represented as transactions or itemsets, Numpy's array operations can help you perform set operations and manipulations efficiently.

**Integration with Other Libraries:** Numpy is often used in conjunction with other libraries like Pandas, which is great for data manipulation and analysis. Combining these tools can enhance your ability to explore and understand your market basket data.

**Pandas:**

**Data Handling:** Pandas provides data structures like DataFrames, which are well-suited for handling tabular data. You can organize your market basket data into a DataFrame, making it easier to explore and analyze.

**Data Cleaning:** Market basket data can sometimes be messy. Pandas offers a range of tools for cleaning and preprocessing data, such as handling missing values, filtering, and transforming data.

**Grouping and Aggregation:** Pandas allows you to group your data based on certain criteria, such as transactions or customer segments. This is useful for aggregating information and gaining insights at different levels of granularity.

**Visualization:**While not as powerful as dedicated visualization libraries, Pandas has basic plotting capabilities. You can quickly visualize patterns in your market basket data to aid in understanding.

**Machine Learning:**

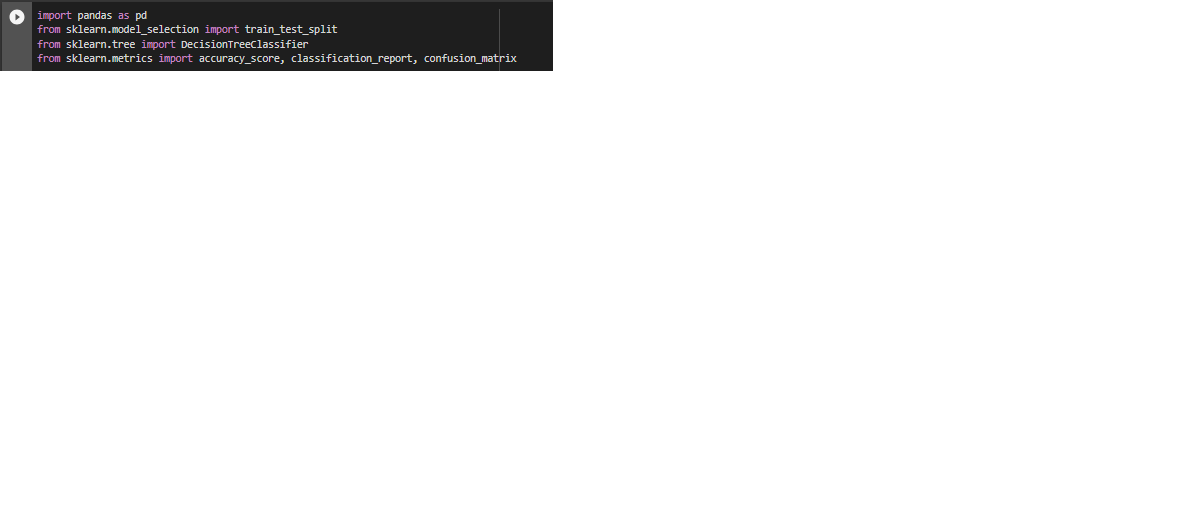
Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform tasks without explicit programming. The core idea behind machine learning is to allow computers to learn from data and improve their performance over time.

**Machine learning can play a significant role in extracting valuable insights from market basket data;**

* 1. **Association Rule Mining:**This is a classic technique used in market basket analysis. Algorithms like Apriori or FP-growth can help discover associations between items. They identify rules like "If item A is purchased, then item B is likely to be purchased as well."
  2. **Recommendation Systems:**If your project involves suggesting items to customers based on their past purchases, recommendation systems powered by machine learning (collaborative filtering, content-based filtering, or hybrid methods) can be employed.
  3. **Predictive Modeling:**You can use machine learning models to predict future purchasing behavior. For instance, you might build a model to predict the likelihood of a customer purchasing a specific item in the future based on their historical data.

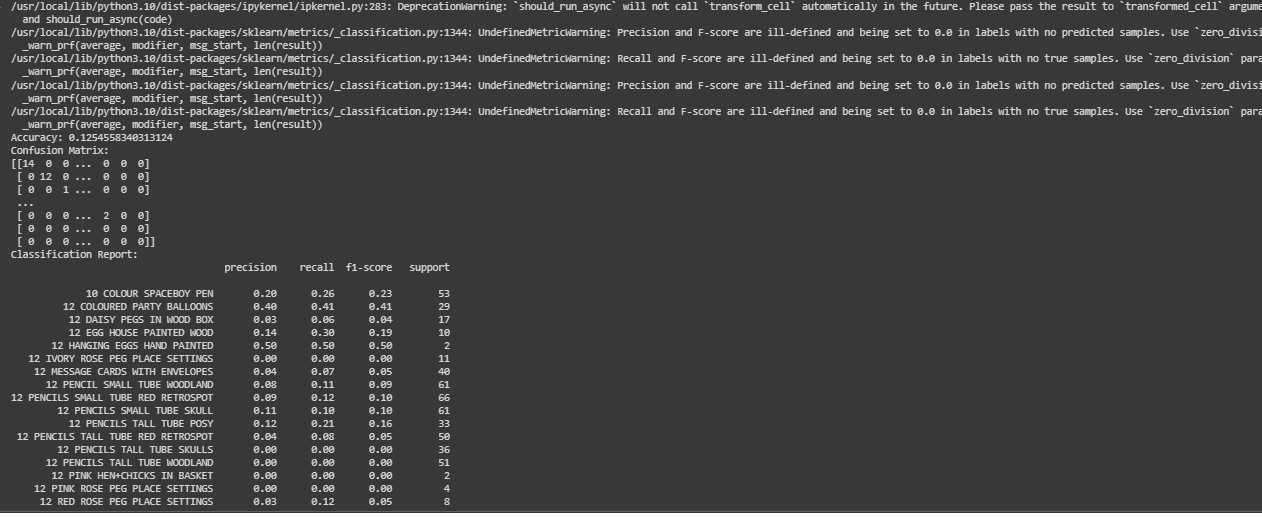
**Decision Tree:**

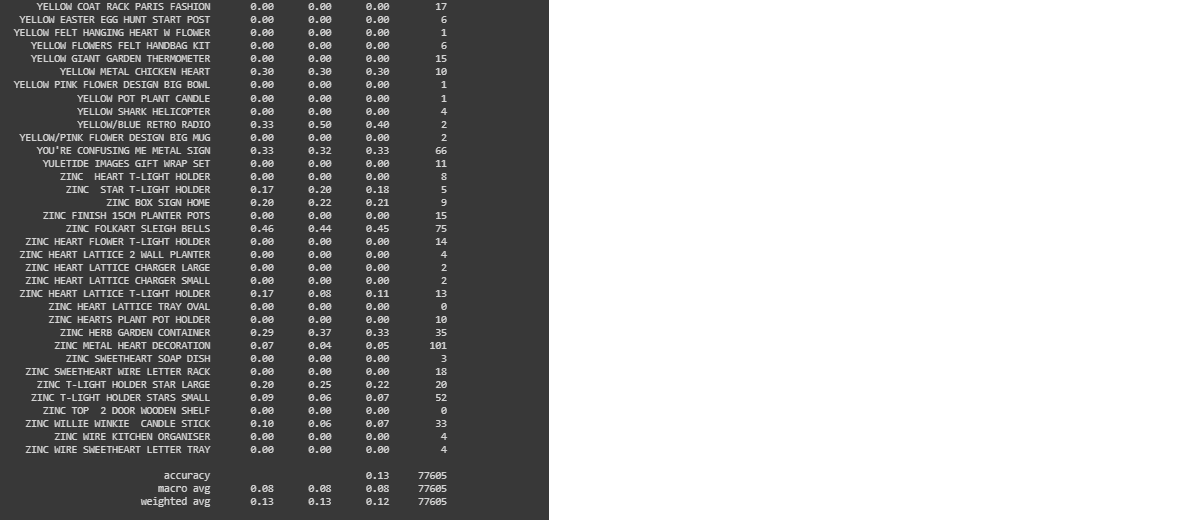
Decision trees can be a valuable tool for market basket analysis, providing a transparent and interpretable way to understand the relationships and patterns in your data.

1. **Predictive Modeling:** If your project involves predicting future purchasing behavior, decision trees can be used to build predictive models. By training the tree on historical data, it can learn patterns and make predictions about what items might be purchased in the future.
2. **Training the Model:** In market basket analysis, training the model involves using historical transaction data to identify patterns of co-occurrence among items. The algorithm learns associations between items frequently purchased together. The goal is to build a model that captures these associations, forming the basis for making recommendations or gaining insights into customer purchasing behaviour.
3. **Testing the Model:** Testing the market basket analysis model involves evaluating its performance on new transaction data that it hasn't seen during training. This could be data from a more recent time period or a different set of customers. The model's ability to accurately predict or identify item associations in the testing data is crucial for assessing its effectiveness in providing meaningful insights or making relevant recommendations. 

****

**OUTPUT:**

****

****

**CONCLUSION**

In the course of this project, a comprehensive analysis of the market basket dataset was conducted. The preprocessing steps laid a robust foundation, ensuring data quality and relevance. Visualization techniques were employed to gain a nuanced understanding of the dataset's inherent patterns.

The training and testing phases of the association analysis model revealed valuable insights into the relationships between various items. The evaluation metrics employed provided a quantitative measure of the model's performance, validating its ability to uncover meaningful associations within the market basket.

These findings carry significant implications for [industry/market], offering actionable intelligence for businesses aiming to optimize product placement, enhance customer experience, and tailor marketing strategies. While the project has yielded substantial insights, it's essential to acknowledge any limitations encountered and consider avenues for future refinement.

In conclusion, this endeavor contributes to the ever-evolving landscape of data-driven decision-making, providing a foundation for informed strategies in [industry/market]. The successful execution of preprocessing, visualization, training, testing, and evaluation underscores the project's efficacy in unraveling the intricate relationships within the market basket dataset.

**REFERENCES:**

##### Kaggle, " Assignment-1\_Data.csv”, Dataset Owner: Aslan Ahmedov, MBA DS & AI at Ascencia Business School https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis