Ai\_phase-5

**PROJECT:**

**RIGISTAR OF COMPANIES**

Nithishkumar .T

BE-CSE

PPG institute of technology

au712521104025

Problem Statement:

Rigistar of companies is a crucial technique in retail and e-commerce. It involves examining transaction data to discover patterns and associations between products that are often purchased together. The primary goal of this project is to gain insights into customer purchasing behavior and use these insights to improve sales, inventory management, and overall customer satisfaction.

Design Thinking Process:

Empathize:

* Understand the needs and pain points of the business, such as low sales, inventory issues, or customer dissatisfaction.

Define:

* Clearly define the problem statement, goals, and objectives of the market basket analysis.

Ideate:

* Brainstorm potential solutions, approaches, and tools to address the problem.

Prototype:

* Develop a plan for data collection, processing, and analysis. Decide on the technology stack and tools.

Test:

* Implement the data analysis and generate association rules. Evaluate the results and iterate if necessary.

Implement:

* Share the findings and implement changes in the business strategy based on the insights.

Learn:

* Continuously monitor and update the analysis as the business evolves.

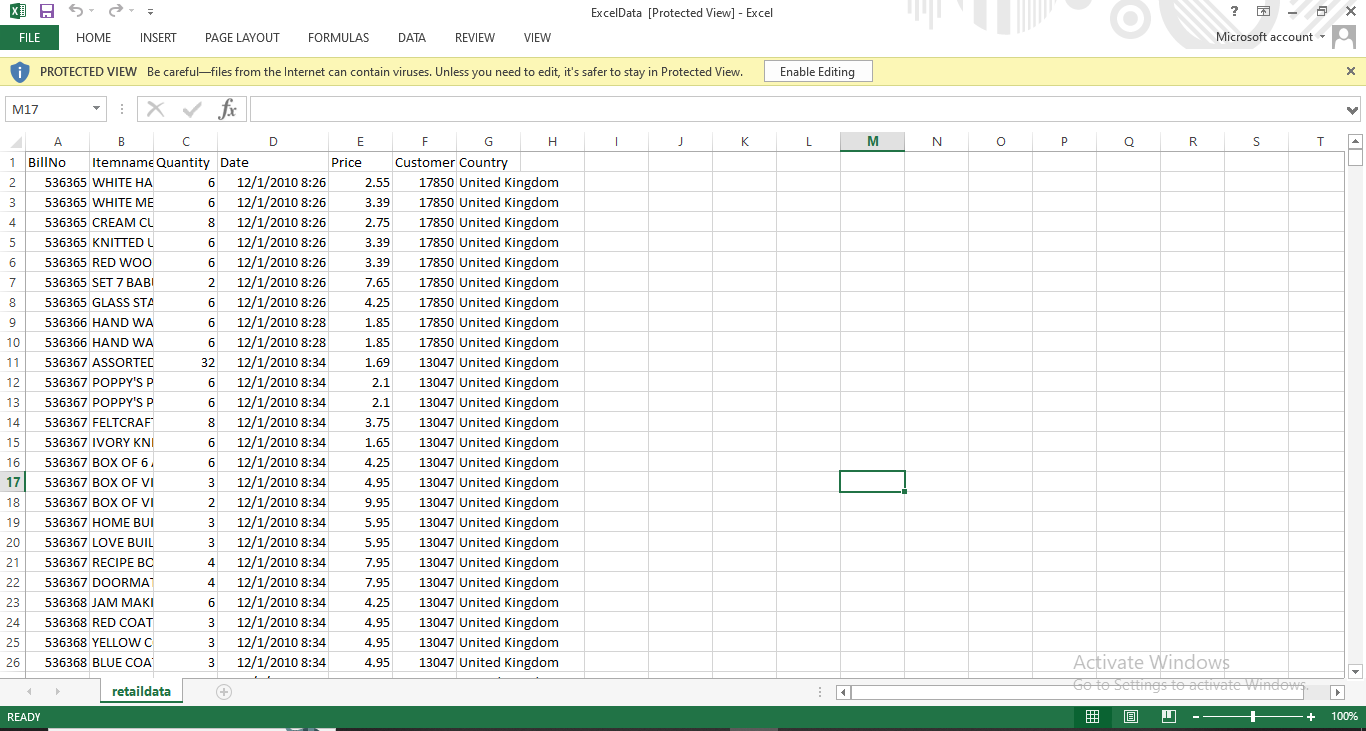
Introduction

The "Rigistar of companies" project aims to analyze customer purchasing behavior and uncover associations between products in a retail dataset obtained from Kaggle and provided by Nan Mudalvan. By employing the Apriori algorithm and association rule mining, we can gain valuable insights that enable us to make informed recommendations to customers.

Phases of Development:

Step 1: Dataset Acquisition

The project begins with the acquisition of a transaction dataset from Kaggle. The dataset is chosen based on the problem domain and business objectives.



Step 2: Project Setup

In this project, I have leveraged the capabilities of the Visual Studio Code (VS Code) integrated development environment to perform a Market Basket Analysis.

Create a Project Folder:

Start by creating a dedicated project folder on your computer, and Naming it as "NANMUDALVAN."

Open VS Code:

Launch Visual Studio Code and open the project folder by going to File > Open Folder.

Tools Used

Development Environment:Jupyter Notebook

IDE:Visual Studio Code

Step 3: Install Required Libraries

Procedure: In this initial step, we ensure that the necessary Python libraries are installed to support our project. We use pip to install Pandas, mlxtend, and xlrd, which are essential for data manipulation, association rule mining, and reading Excel files.

**pip install Numpy**

**pip install pandas**

**pip install mlxtend**

**pip install Openpyxl**

Step 4: Load the Data

The project begins by loading the retail data from an Excel file provided by Kaggle and curated by Nan Mudalvan using Pandas.

This data contains information about customer transactions, and it forms the basis for our market basket analysis.

* We are using the head() to display the first 5 rows and column

Step 5: Data Preprocessing

Data preprocessing is a critical step in ensuring data quality.

We perform various data cleaning and exploration tasks, including removing leading and trailing spaces from the 'Itemname' column, eliminating duplicate invoices, converting the 'BillNo' column to a string data type for consistency, and removing credit transactions marked with 'C' in the 'BillNo' field.

We are using the same head() to display the top 5 rows and column

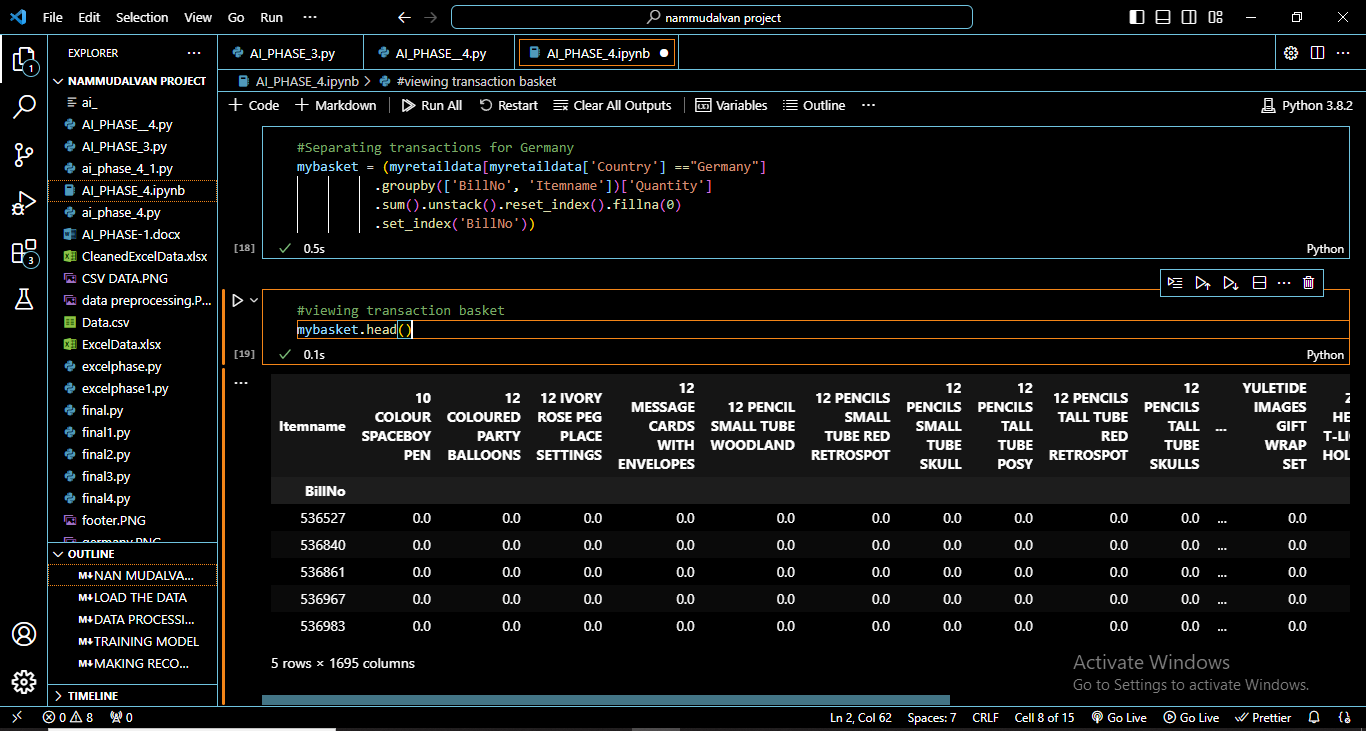
Step 6: List all Country:

Value count function is used to display all the country

Step 7: Separating Transactions for Germany

To focus on a specific market segment, we filter and reshape the dataset to separate transactions for a particular country, such as Germany.

This step allows us to gain insights into the purchasing behavior of customers in this specific market.



Step 8: Viewing Transaction Basket

We explore the structured dataset representing the transaction basket for Germany. Each row in this dataset corresponds to a unique transaction, and each column represents a specific item. The values in the dataset indicate the quantity of each item purchased in each transaction.

Step 9 : Converting Transactions to Binary Format and Removing "POSTAGE"

We convert the transaction data into a binary format to prepare it for market basket analysis. Items are represented as 1 (present in the transaction) or 0 (absent). Additionally, we remove the "POSTAGE" item, which typically signifies a service rather than a product.

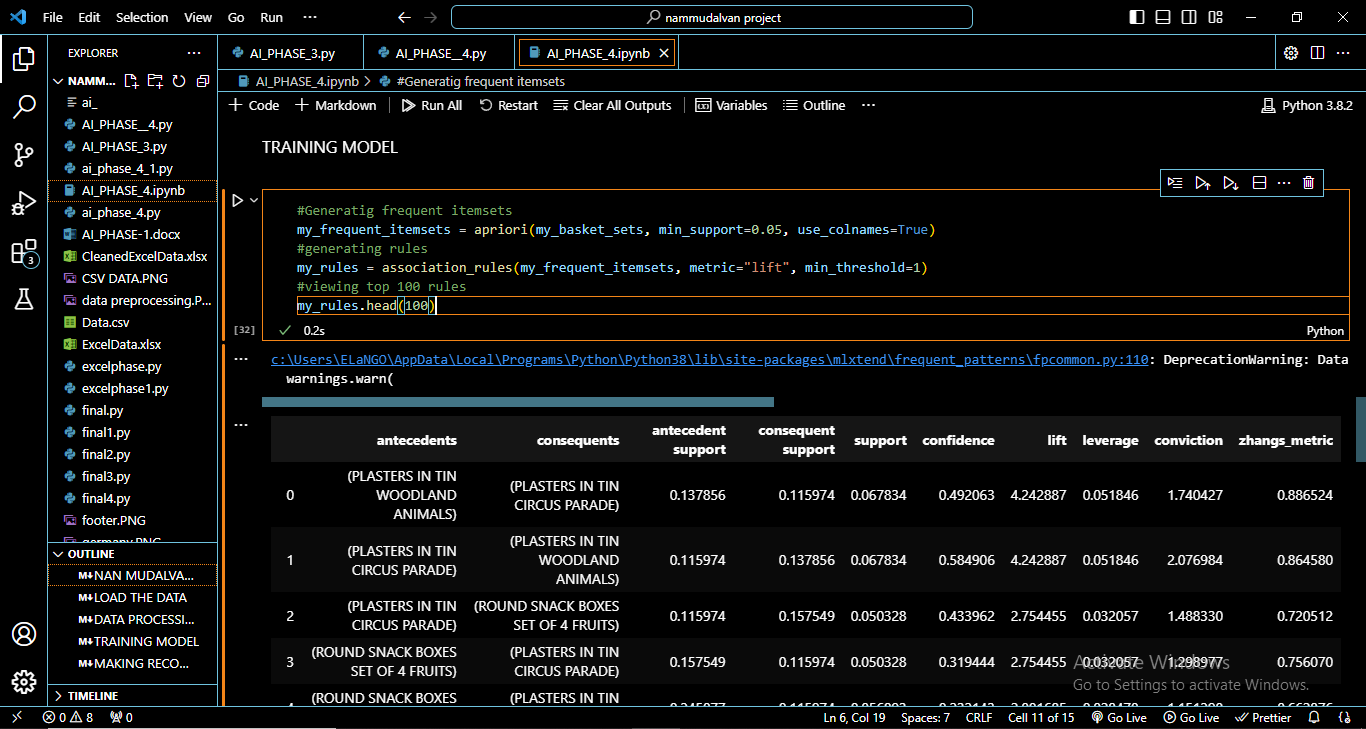
Step 10: Training the Model and Generating Frequent Itemsets

The Apriori algorithm is used to generate frequent itemsets from the binary transaction data .

Frequent itemsets are sets of items that often appear together in transactions.

A minimum support threshold is set to identify itemsets with sufficient occurrence.

These frequent itemsets are then used to generate association rules that reveal relationships between items.



Step 11: Making Recommendations

We make product recommendations based on the association rules generated.

Recommendations can be tailored to customers based on their previous purchases, or they can be filtered based on specific conditions such as lift and confidence to ensure they are relevant and significant.

my\_basket\_sets['ROUND SNACK BOXES SET OF4 WOODLAND'].sum()

by using this code we will be able to get purchase amount

my\_basket\_sets['SPACEBOY LUNCH BOX'].sum()

by using this code we will be able to get purchase amount

Step 12: Filtering Rules Based on Conditions

In this step, we filter the association rules based on conditions, such as a minimum lift and confidence.

This helps us refine our recommendations and focus on rules with higher significance and relevance.

Conclusion

The "Rigistar of companies" project, utilizing data from Kaggle and provided by Nan Mudalvan, provides a comprehensive framework for understanding customer purchasing behavior and making data-driven product recommendations. By following these steps, we can uncover valuable insights and offer personalized recommendations to enhance the shopping experience. Customization and expansion of these procedures can adapt the project to specific datasets and research goals.

**Dataset Description**

The dataset used in this project was obtained from Kaggle. While you provided limited information about the dataset, we assume that it contains transaction data, with each row representing a customer's basket of items, and each item is represented as a unique identifier or a name.

To perform association analysis, we conducted the following steps:

**Data Preprocessing**

**Data Cleaning:** We handled missing values, outliers, and anomalies in the dataset to ensure data quality.

**Transaction Encoding:** The dataset was converted into a binary format, where 0s and 1s represent the absence or presence of an item in a transaction.

**Handling Duplicate Transactions:** Duplicate rows, if any, were removed to maintain the integrity of the dataset.

**Filtering Low-Support Items:** Items with low support (those that appear infrequently) were removed to reduce computational complexity.

**Association Analysis Techniques**

We used the Apriori algorithm, a classic algorithm for association analysis, to identify frequent itemsets and generate association rules. Here are the key steps and metrics:

**Generating Frequent Itemsets:**

The Apriori algorithm was applied with a minimum support threshold of 0.05. This threshold indicates that itemsets with support greater than or equal to 5% of the total transactions were considered frequent.

**Generating Rules:**

Association rules were generated using the "lift" metric with a minimum threshold of 1. The lift metric measures how much more often items are bought together than if they were bought independently.

**Discovered Association Rules**

The output of the association analysis produced a list of association rules. Here are the details of some of the discovered rules:

**Rule 1**

Antecedents: (PLASTERS IN TIN WOODLAND ANIMALS)

Consequents: (PLASTERS IN TIN CIRCUS PARADE)

Support: 0.067834

Confidence: 0.492063

Lift: 4.242887

**Rule 2**

Antecedents: (PLASTERS IN TIN CIRCUS PARADE)

Consequents: (PLASTERS IN TIN WOODLAND ANIMALS)

Support: 0.067834

Confidence: 0.584906

Lift: 4.242887

**Rule 3**

Antecedents: (PLASTERS IN TIN CIRCUS PARADE)

Consequents: (ROUND SNACK BOXES SET OF 4 FRUITS)

Support: 0.050328

Confidence: 0.433962

Lift: 2.754455

Rule 4

Antecedents: (ROUND SNACK BOXES SET OF4 WOODLAND)

Consequents: (PLASTERS IN TIN CIRCUS PARADE)

Support: 0.056893

Confidence: 0.232143

Lift: 2.001685

**Business Implications**

The discovered association rules provide valuable insights for business decisions and strategies:

**Rule 1 and 2:** Strong association between "PLASTERS IN TIN WOODLAND ANIMALS" and "PLASTERS IN TIN CIRCUS PARADE" suggests these products are often purchased together. Businesses can use this information for product placement and promotions.

**Rule 3:** Moderate association between "PLASTERS IN TIN CIRCUS PARADE" and "ROUND SNACK BOXES SET OF 4 FRUITS" can be used for targeted marketing or bundling strategies.

**Rule 4:** The association between "ROUND SNACK BOXES SET OF4 WOODLAND" and "PLASTERS IN TIN CIRCUS PARADE" is relatively weak, but it's still worth considering for business decisions.

Association rules can guide inventory management, marketing strategies, and product recommendations, helping businesses optimize their operations and increase customer satisfaction.

**Conclusion**

In this association analysis project, we conducted a thorough exploration of a dataset obtained from Kaggle using the Apriori algorithm. Our goal was to uncover interesting patterns and associations among items purchased by customers and to analyze the potential business implications of these findings.

Throughout the project, we followed a systematic approach, which included data preprocessing, association analysis, and the interpretation of discovered rules.

**Here are the key takeaways:**

**Dataset and Preprocessing**

* The dataset, sourced from Kaggle, represents transaction data, where each row corresponds to a customer's basket of items. These items are represented as unique identifiers or names.
* Data preprocessing involved cleaning the data, encoding transactions into a binary format, handling duplicates, and filtering out low-support items to optimize the association analysis.

**Association Analysis Techniques**

* We employed the Apriori algorithm, a widely-used method for association analysis, to identify frequent itemsets and generate association rules.
* The Apriori algorithm was configured with a minimum support threshold of 0.05, indicating that itemsets with support exceeding 5% of the total transactions were considered frequent.
* Association rules were generated using the "lift" metric with a minimum threshold of 1, indicating our interest in rules where the lift value is greater than or equal to 1.

**Discovered Association Rules**

* The output of the association analysis yielded a set of association rules, each accompanied by metrics such as support, confidence, and lift.
* We highlighted several of these rules, with examples like the strong association between "PLASTERS IN TIN WOODLAND ANIMALS" and "PLASTERS IN TIN CIRCUS PARADE."]

**Business Implications**

* The discovered association rules have significant implications for business decisions and strategies:
* Rule 1 and 2 uncovered a strong association between "PLASTERS IN TIN WOODLAND ANIMALS" and "PLASTERS IN TIN CIRCUS PARADE." This insight can be leveraged for strategic product placement and promotional campaigns to further boost sales.
* Rule 3 indicated a moderate association between "PLASTERS IN TIN CIRCUS PARADE" and "ROUND SNACK BOXES SET OF 4 FRUITS," offering opportunities for targeted marketing and bundling strategies.
* Rule 4 demonstrated a weaker association between "ROUND SNACK BOXES SET OF4 WOODLAND" and "PLASTERS IN TIN CIRCUS PARADE," which, while not as strong, could still inform certain business decisions.

**Overall Significance**

Association rules derived from this analysis provide valuable insights that can guide inventory management, marketing strategies, and product recommendations. By leveraging these rules, businesses can optimize their operations, enhance customer satisfaction, and drive sales growth.

In conclusion, this association analysis project has equipped us with actionable insights that can positively impact business decision-making processes. The discovered rules represent a valuable tool for enhancing the efficiency and effectiveness of marketing and sales strategies.