Ai\_phase-4

**PROJECT:**

**MARKET BASKET INSIGHTS**

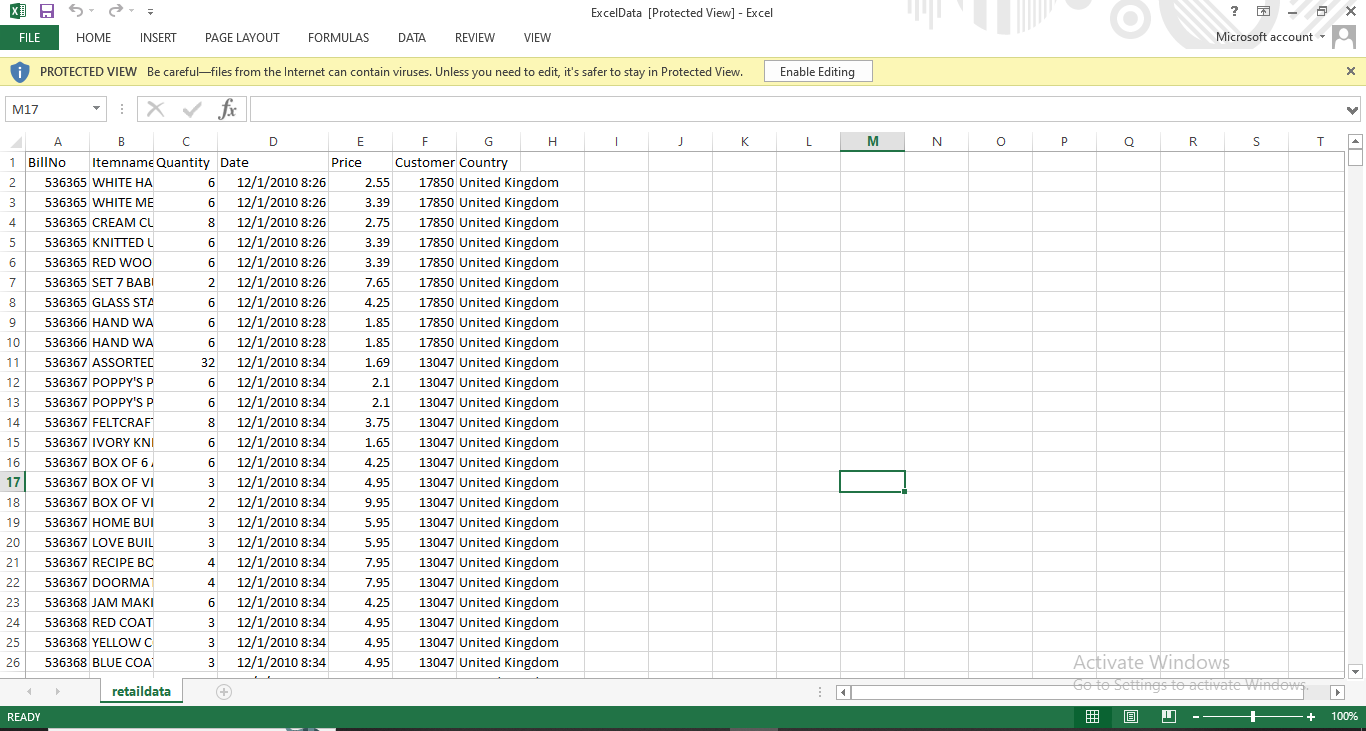
Introduction

The "Market Basket Insights" 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 4 steps:

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**

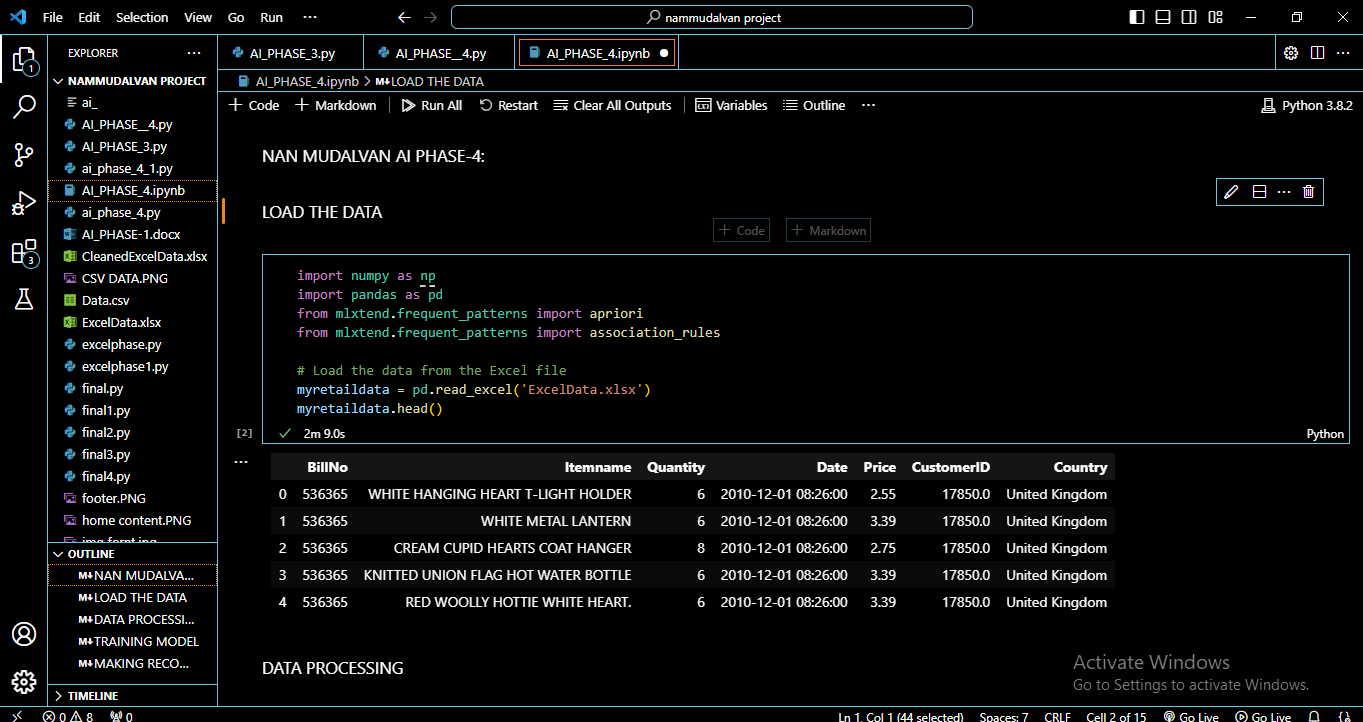
**pip install mlxtend**

**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.

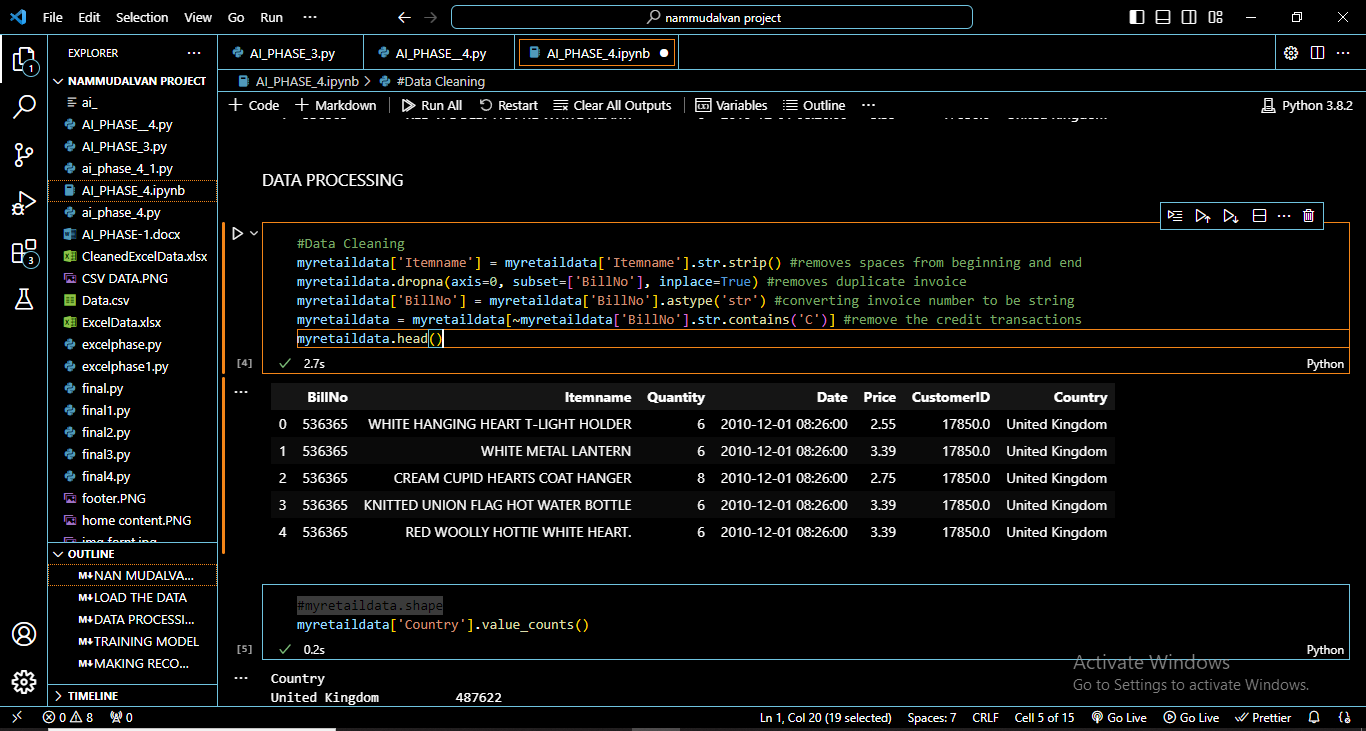
This data contains information about customer transactions, and it forms the basis for our market basket analysis.



* 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:

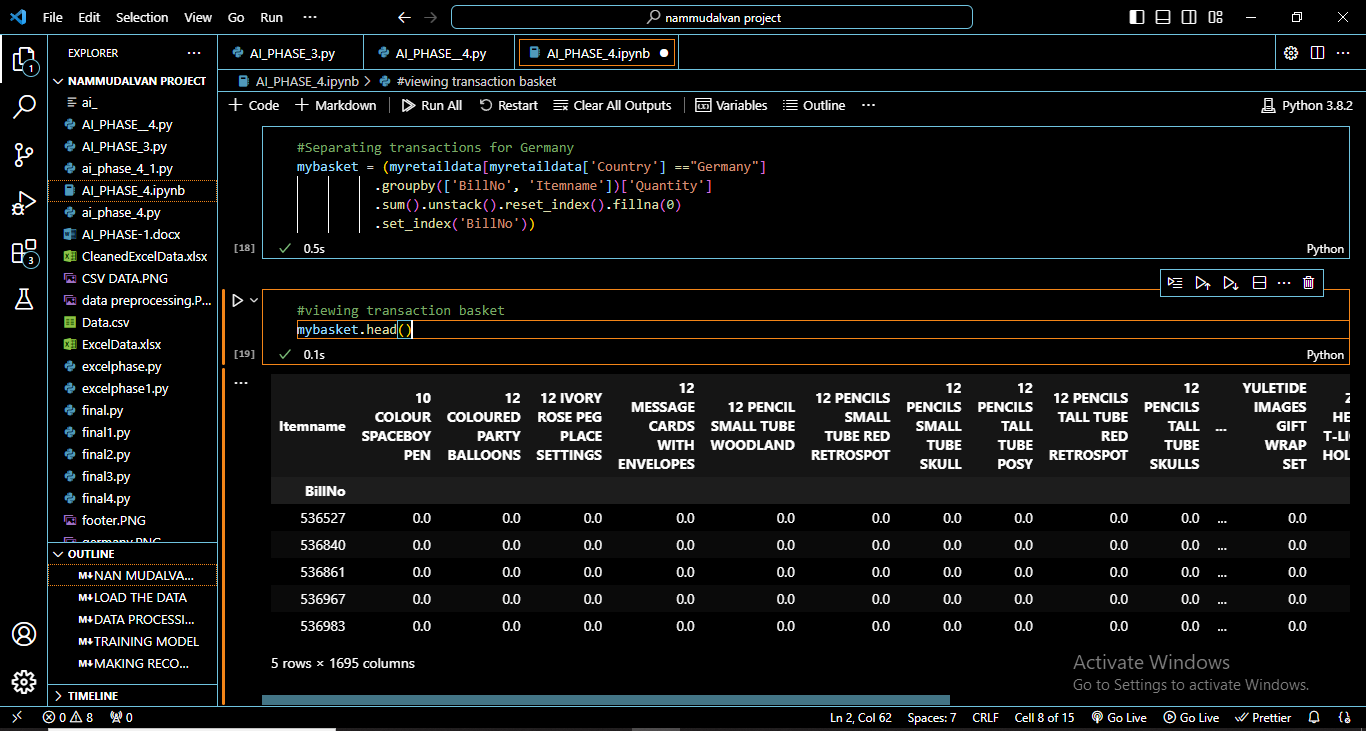
Value count function is used to display all the country



Step 7: Separating Transactions for Germany

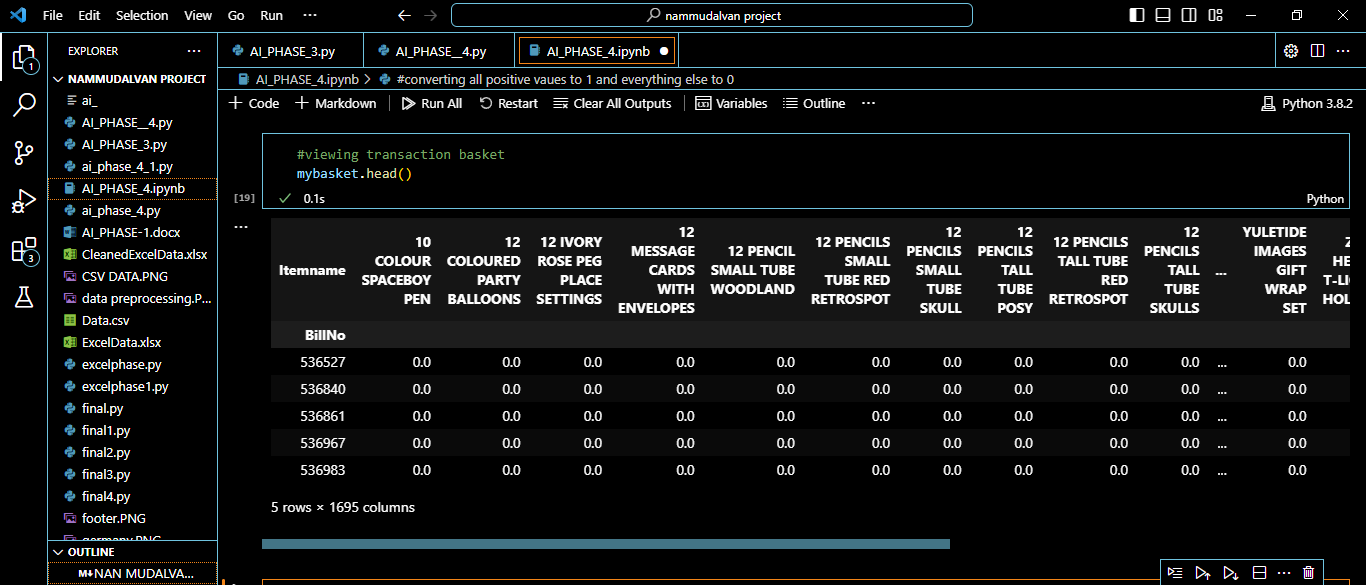
To focus on a specific market segment, we filter and reshape the dataset to separate transactions for a particular country, such as Germany.

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



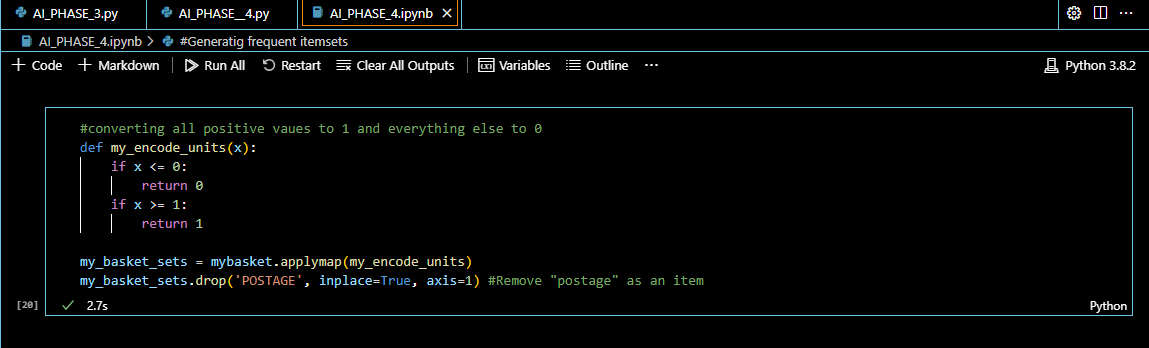
Step 8: Viewing Transaction Basket

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.



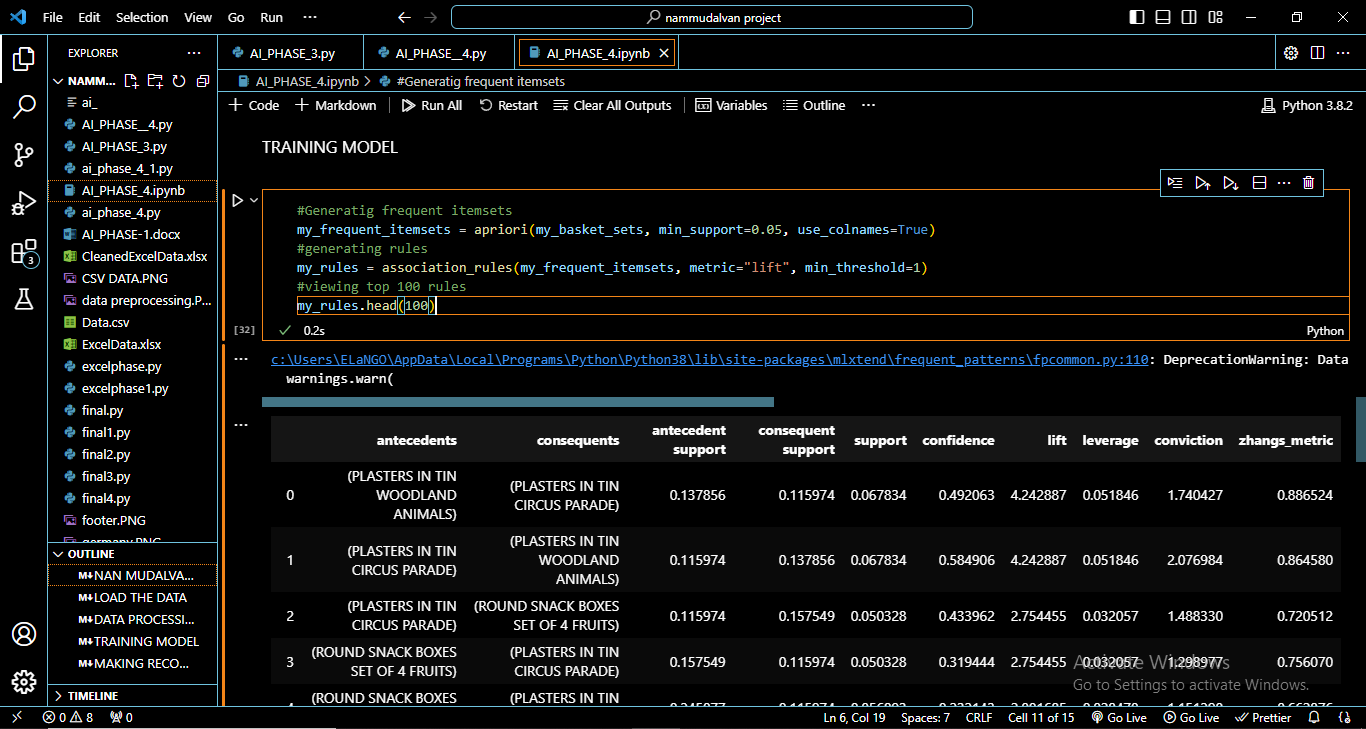
Step 10: Training the Model and Generating Frequent Itemsets

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

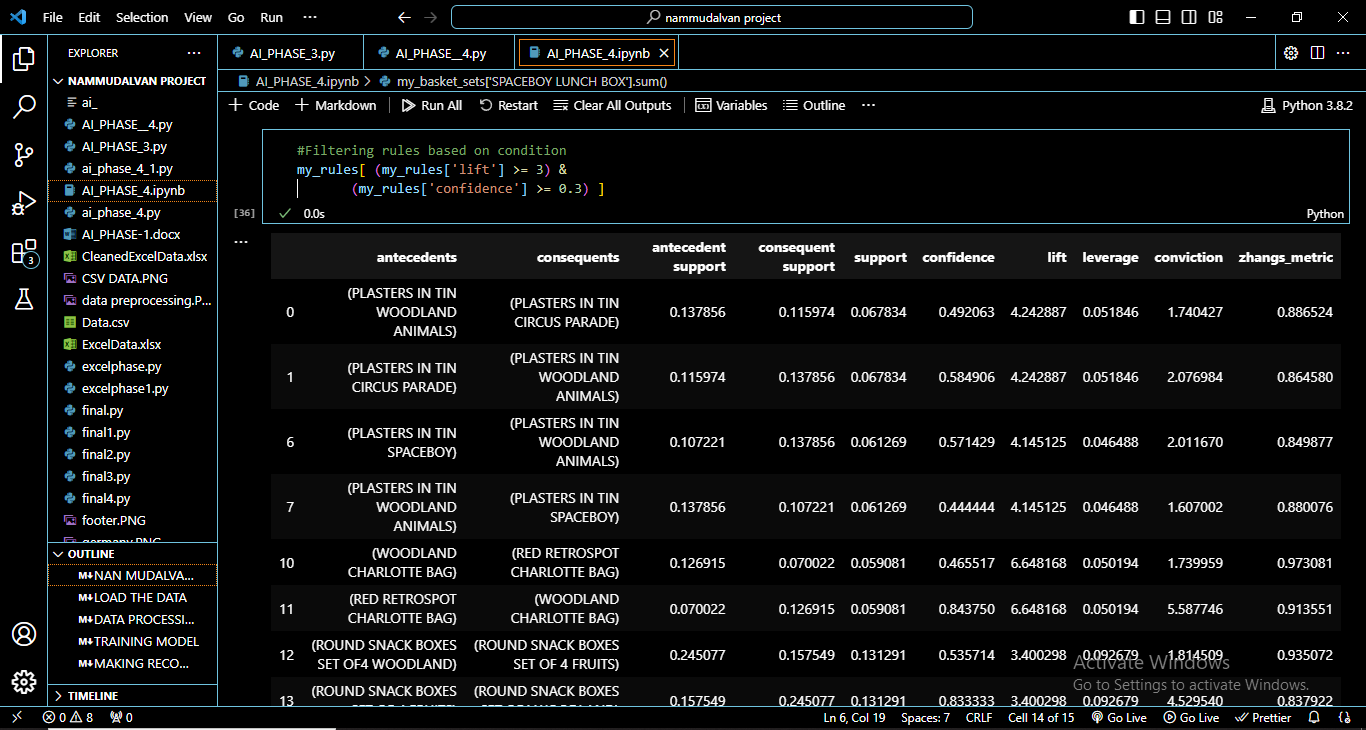
my\_basket\_sets['SPACEBOY LUNCH BOX'].sum()

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 "Market Basket Insights" 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.