**MARKET BASKET INSIGHTS**

**Phase 4: Development Part-2**

* INTRODUCTION
* PROBLEM STATEMENT
* IMPORTING DATA
* DATA CLEANING
* DATA ANALYSING
* DATA VISUALISING
* APRIORI ALGORITHM
* COMPUTING SUPPORT AND CONFIDENCE
* COMPUTING FREQUENT ITEMS USING APRIORI ALGORITHM
* ASSOCIATION RULES ARE USED FOR FREQUENT ITEMS
* COMPUTING ZHANG’S RULE
* TRANSFORMING ANTECEDENT,CONSEQUENT AND SUPPORT
* GENERATING HEAT-MAP
* GENERATING SCATTER-PLOT
* GENERATING PARALLEL COORDINATES PLOT
* CONCLUSION

**INTRODUCTION:**

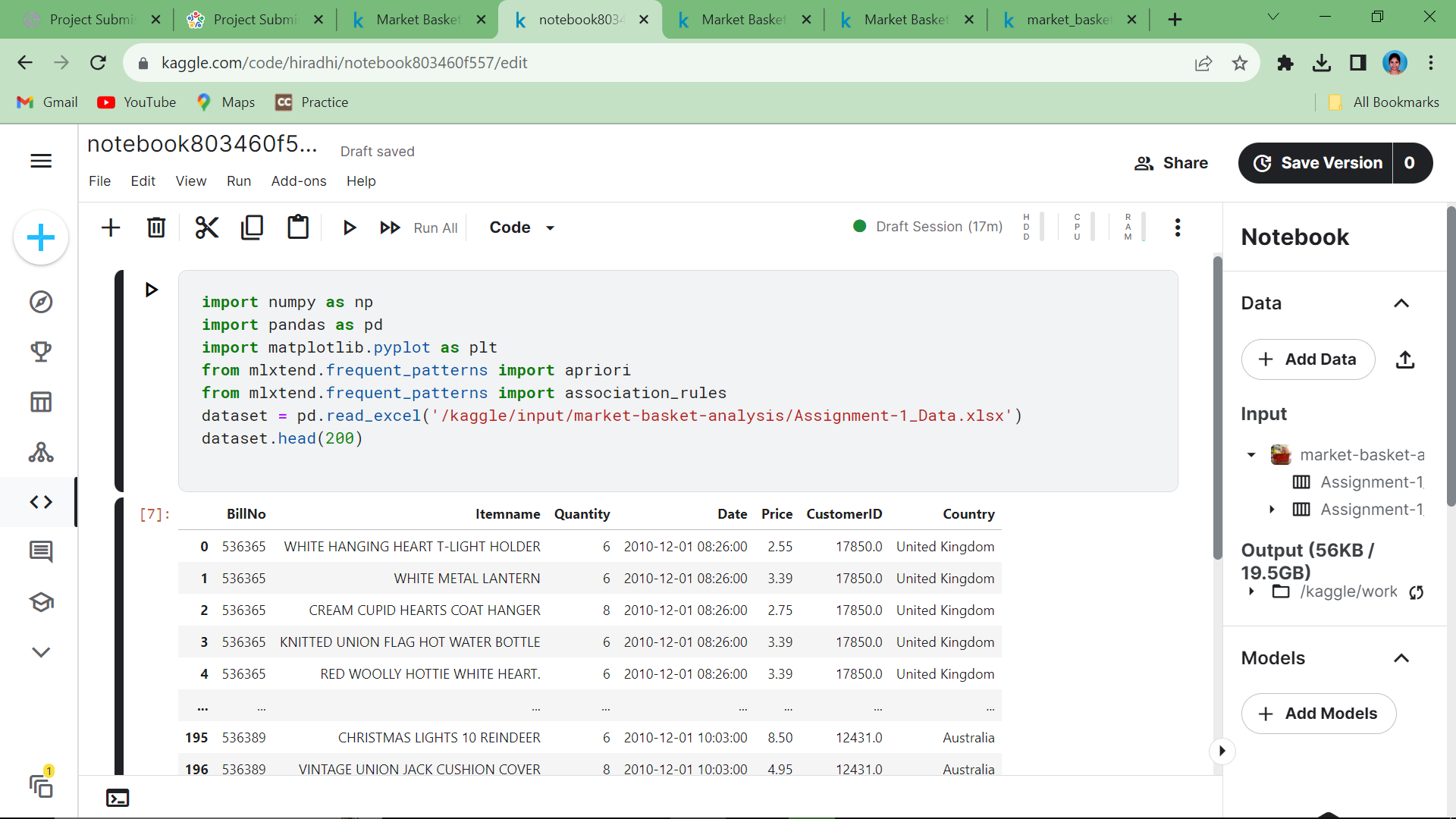
* **Market Basket Analysis**, known as Association rule analysis is a data analysis technique used in the field of data mining and business analytics.
* It aims to uncover relationships and associations between items that are frequently purchased or used together in transactions. This analysis helps businesses understand customer behavior, improve product recommendations, enhance marketing strategies, and optimize inventory management.
* The primary focus is on identifying patterns and rules that reveal how different products or services are related in customers' buying habits, which, in turn, can inform decision-making and business strategies.

**PROBLEM STATEMENT:**

Unveiling the customer behaviour through association analysis: Utilize the Market Basket Analysis on the provided dataset to uncover hidden patterns and associations between products, aiming to understand customer purchasing behaviour and identify potential cross selling opportunities for the retail business.

**IMPORTING DATA:**

Data importing is a crucial step as it forms the foundation for discovering associations and patterns among items, which can be used to make business decisions such as product placement, recommendation systems, and pricing strategies.

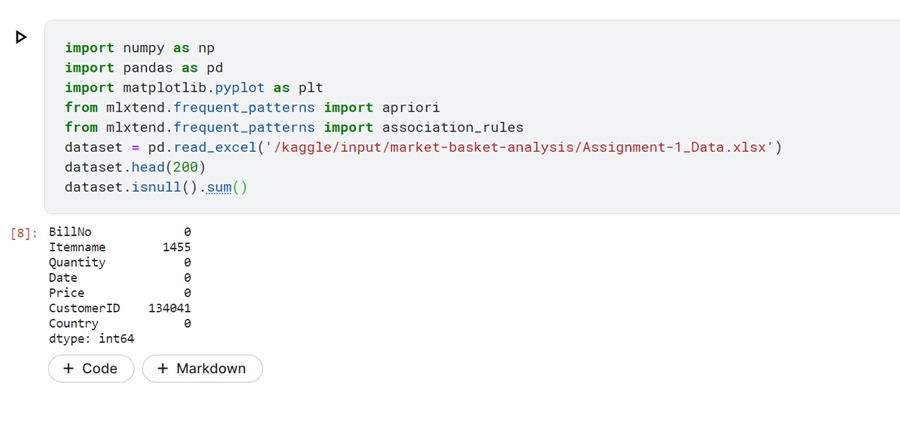


**Data cleaning:**

It is a crucial step in Market Basket Analysis (MBA) to ensure that your transactional data is accurate, reliable, and ready for analysis**.**

Data cleaning is achieved by

* + - * Removing duplicate values
      * Adding missing values



**Data Analyzing:**

It is the process of examining transactional data to identify patterns and relationships between products or items purchased together by customers.

The following steps are used for data analyzing:

* Frequent Itemset Generation
* Support Counting
* Increase k by 1 (k = k + 1)
* Generate Candidate k-Itemsets
* Remove candidate k-itemsets that have infrequent (k-1)-subsets.
* Generate Frequent k-Itemsets
* Form association rules based on frequent k-itemsets.
* Evaluate rule confidence.

**Data Visualization:**

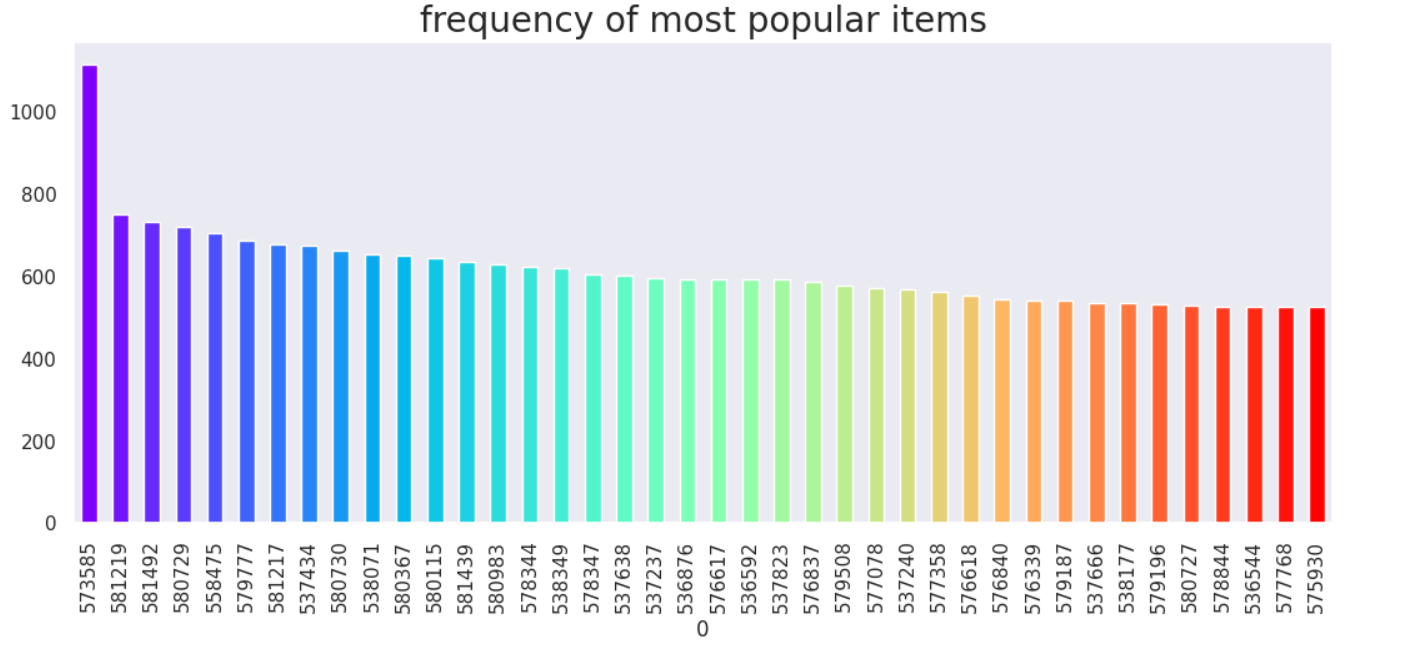
Some data mining tools and libraries that implement the Apriori algorithm for Market Basket Analysis come with built-in visualization features. These visualizations often include support-confidence lift plots, item set diagrams, and association rules.

**1.Frequent itemsets are shown using bargraph:**

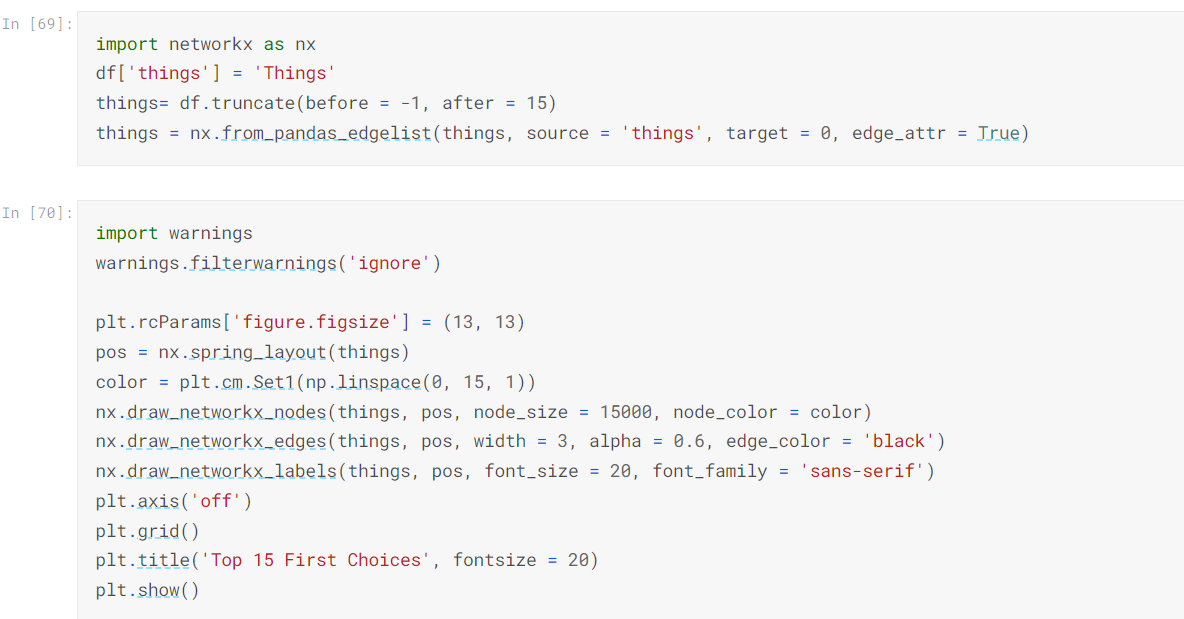
Discovering frequent itemsets is a crucial step in identifying patterns and associations between products in a retail or

e-commerce dataset.

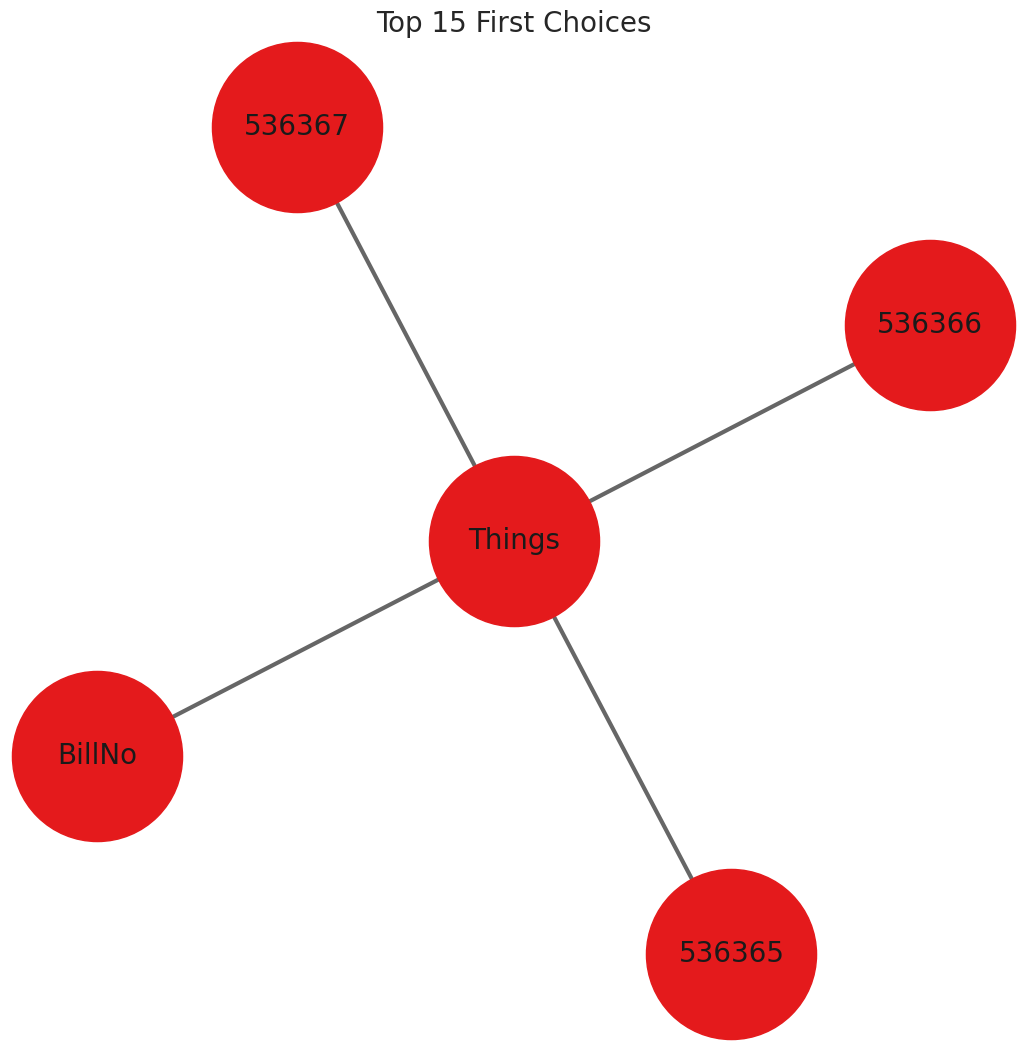
****



**2.Selecting priority from dataset:**



The given below diagram shows the top priority choices in the dataset.



**3.Association rules**:

Association rules are a fundamental concept in Market Basket Analysis , which is used to identify interesting relationships or associations between items in transaction datasets, particularly in retail and e-commerce.



**APRIORI ALGORITHM**

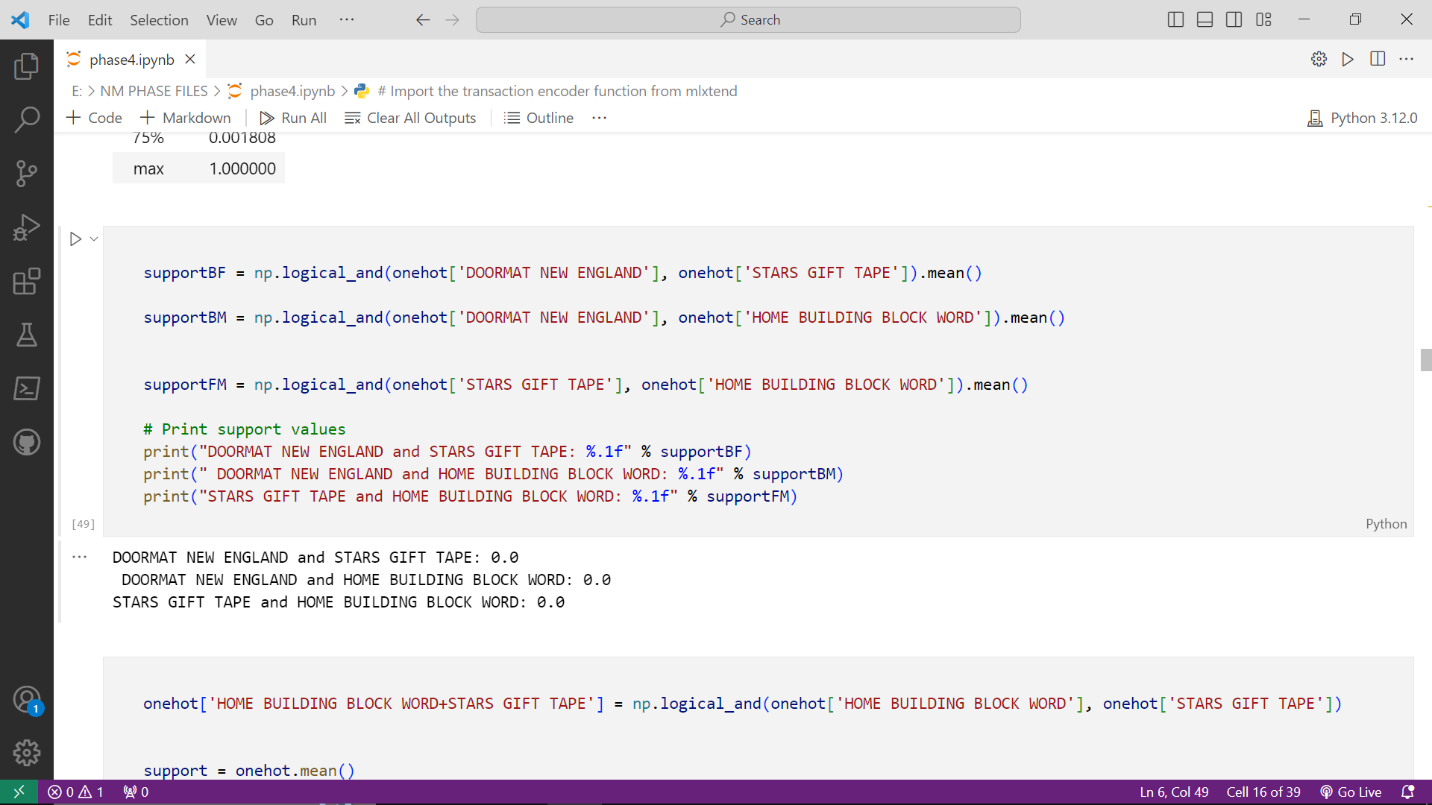
The Apriori algorithm is a data mining technique used for discovering frequent itemsets and association rules in large datasets. It works by:

* Identifying frequent individual items in the dataset based on a minimum support threshold.
* Generating candidate itemsets of increasing size by joining frequent itemsets and pruning those that do not satisfy the Apriori property.
* Counting the support of candidate itemsets in the dataset.
* Repeating steps 2 and 3 until no more frequent itemsets can be found.
* Once frequent itemsets are identified, association rules can be derived to reveal relationships between items in transactions. The Apriori algorithm is commonly used in market basket analysis and recommendations.

**1.COMPUTING SUPPORT AND CONFIDENCE:**

In the Apriori algorithm for association rule mining, "support" and "confidence" are two important metrics used to assess the significance and strength of association rules between items in a transaction database.

1. **Support** quantifies how often an itemset or association rule appears in the dataset.
2. **Confidence** measures the reliability or strength of an association rule.



**2.COMPUTING FREQUENT ITEMS USING APRIORI ALGORITHM:**

* Increase k by 1 (k = k + 1)
* Generate Candidate k-Itemsets
* Remove candidate k-itemsets that have infrequent (k-1)-subsets.
* Generate Frequent k-Itemsets
* Form association rules based on frequent k-itemsets.
* Evaluate rule confidence.

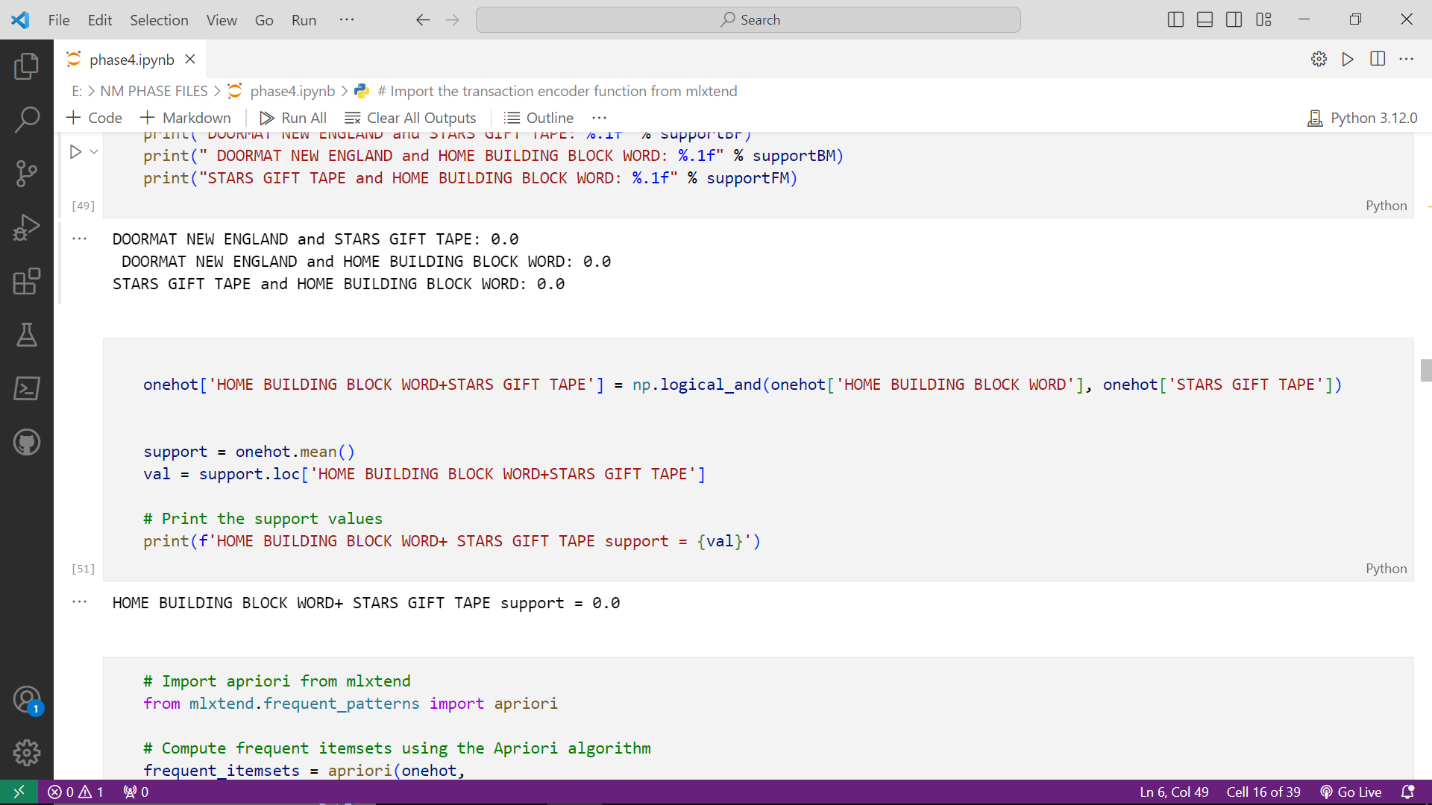
**3.ASSOCIATION RULES ARE USED FOR FREQUENT ITEMS:**

Association rules in market basket analysis are statistical patterns or relationships that reveal how frequently items are purchased together in a dataset of customer transactions.

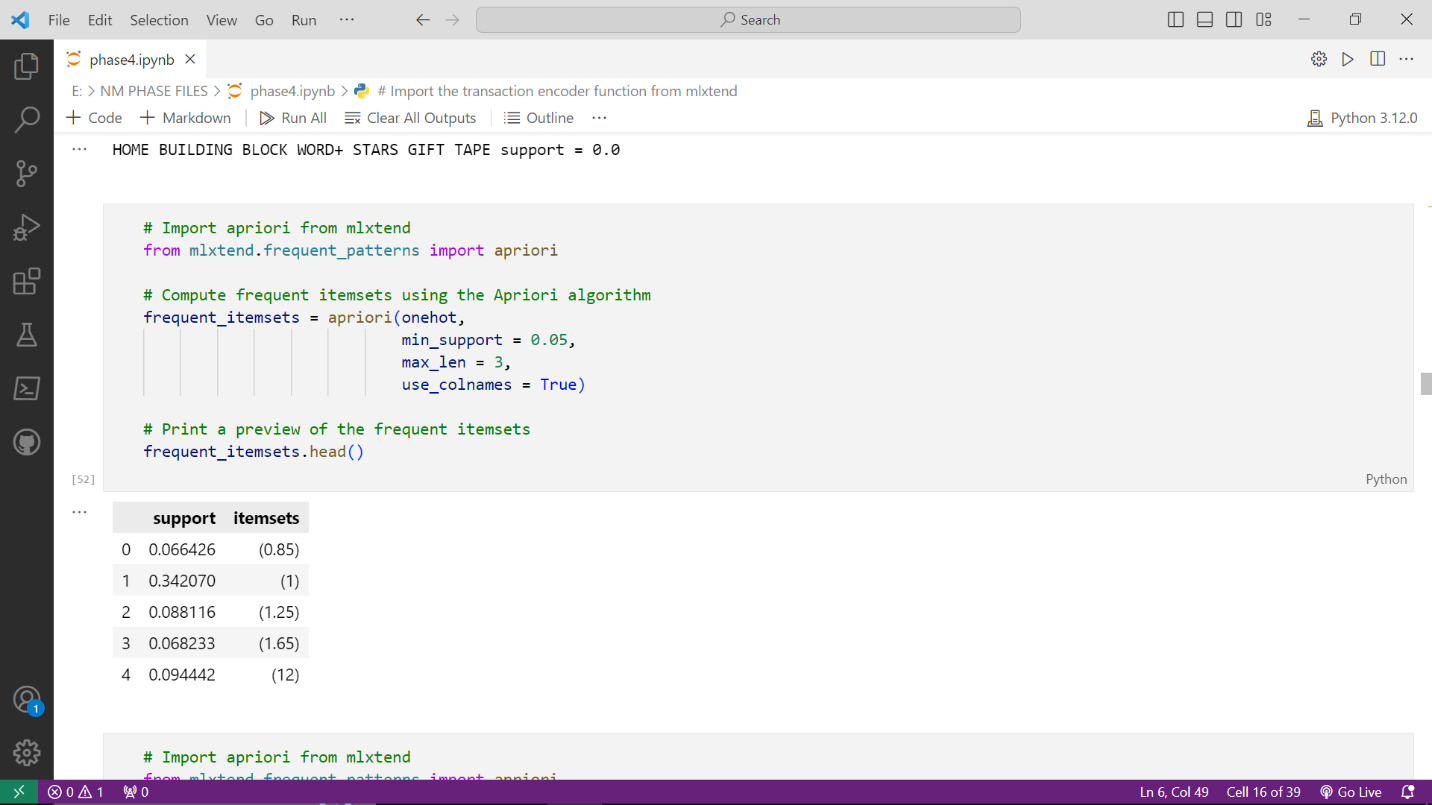
These rules help businesses understand customer purchasing behavior and can be used for various purposes, such as optimizing product placement, cross-selling, and making targeted product recommendations.

**Antecedent (if-part)**: This is the item or set of items that appear in the customer's basket.

**Consequent (then-part):** This is the item or set of items that tend to appear in the same basket along with the antecedent.



**4.COMPUTING ZHANG’S RULE:**



Using Zhang’s Rule in Apriori algorithm, both ‘DOORMAT NEW ENGLAND’,’STARS GIFT TAPE’ is calculated for the support.

**5.TRANSFORMING ANTECEDENT, CONSEQUENT AND SUPPORT:**

"Antecedent" and "Consequent" refer to two sets of items in association rules, which are used to discover patterns in transactional datasets.

1. **Antecedent:**

The antecedent (also known as the "left-hand side" or "LHS") is the set of items that appear in the condition part of an association rule.

It represents the items that are used as a condition or premise in the rule.

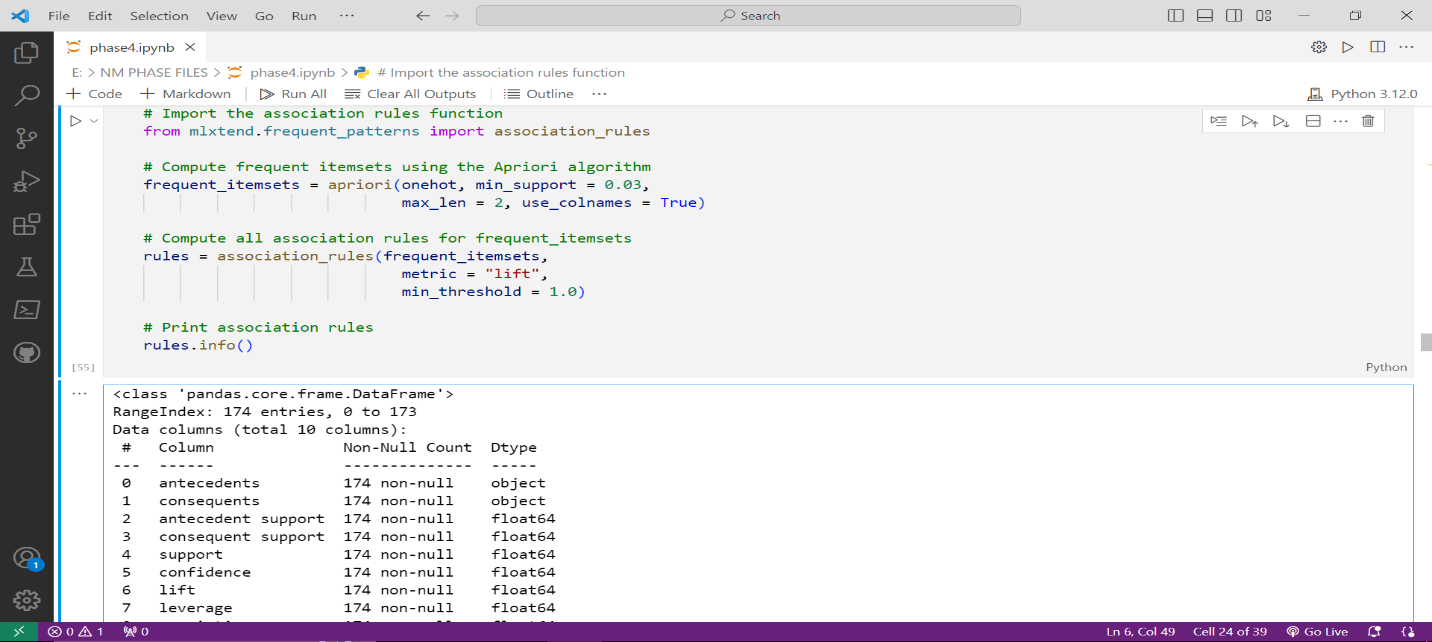
For example, in the association rule "{item A, item B} => {item C}," the antecedent is {item A, item B}.

**2.Consequent:**

The consequent (also known as the "right-hand side" or "RHS") is the set of items that appear in the result part of an association rule.

It represents the items that are predicted or implied by the antecedent.

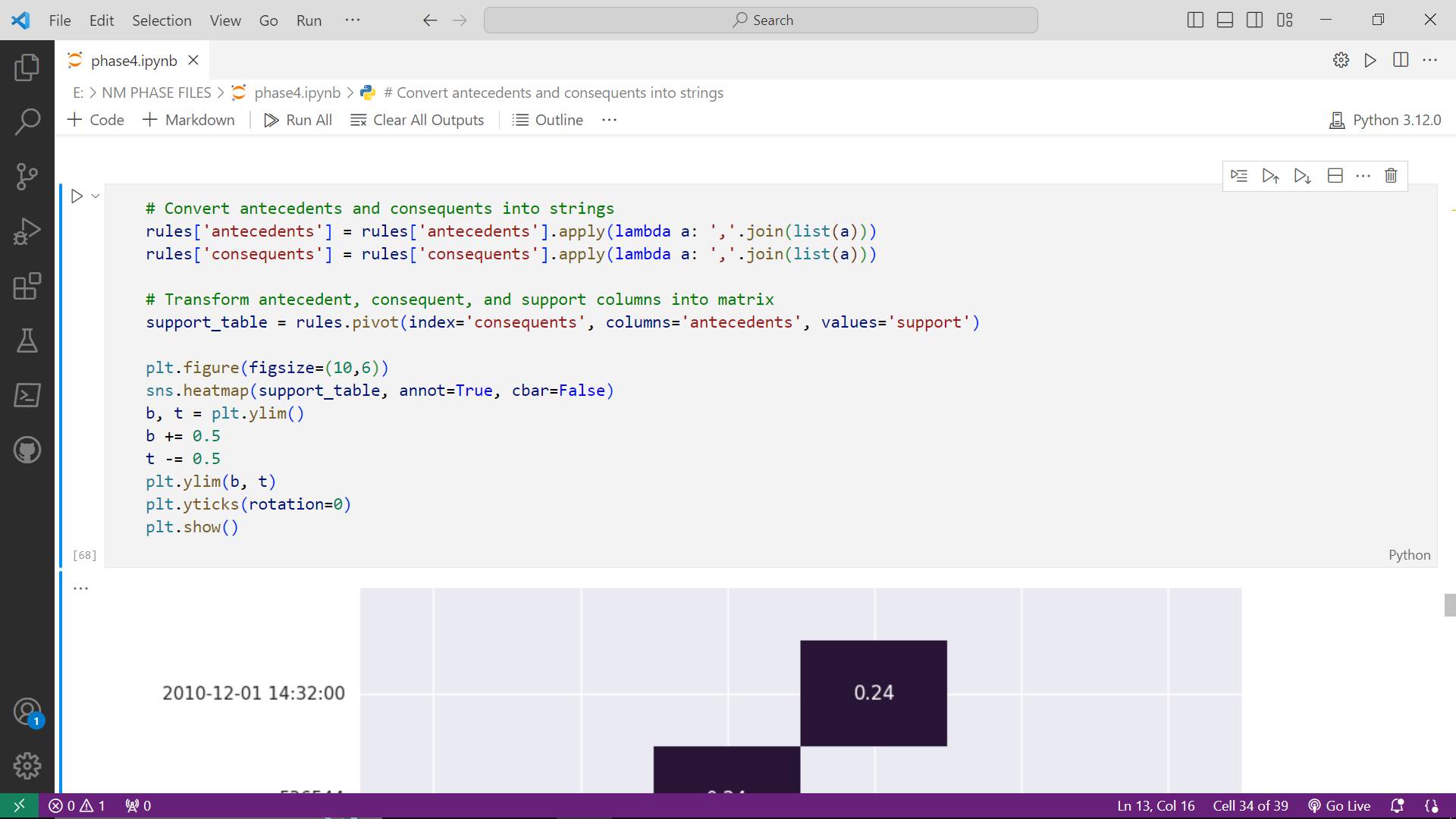
In the same example rule "{item A, item B} => {item C}," the consequent is {item C}.

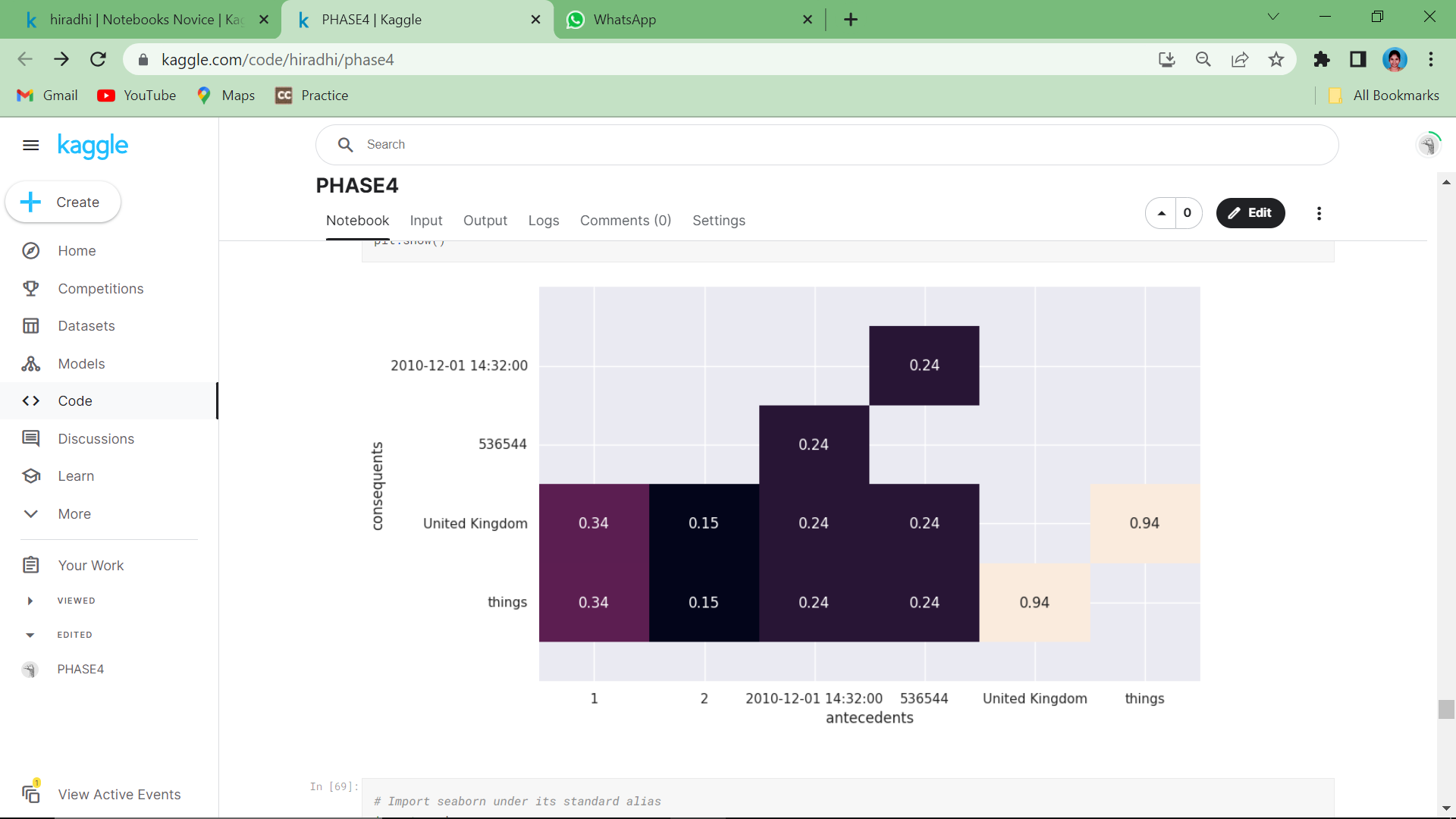


**6.GENERATING HEAT-MAP:**

A heatmap in market basket analysis is a visual representation that displays the strength of associations or correlations between items purchased together in a transactional dataset.

Heatmaps are a common way to illustrate the relationships between items in terms of their co-occurrence or support.

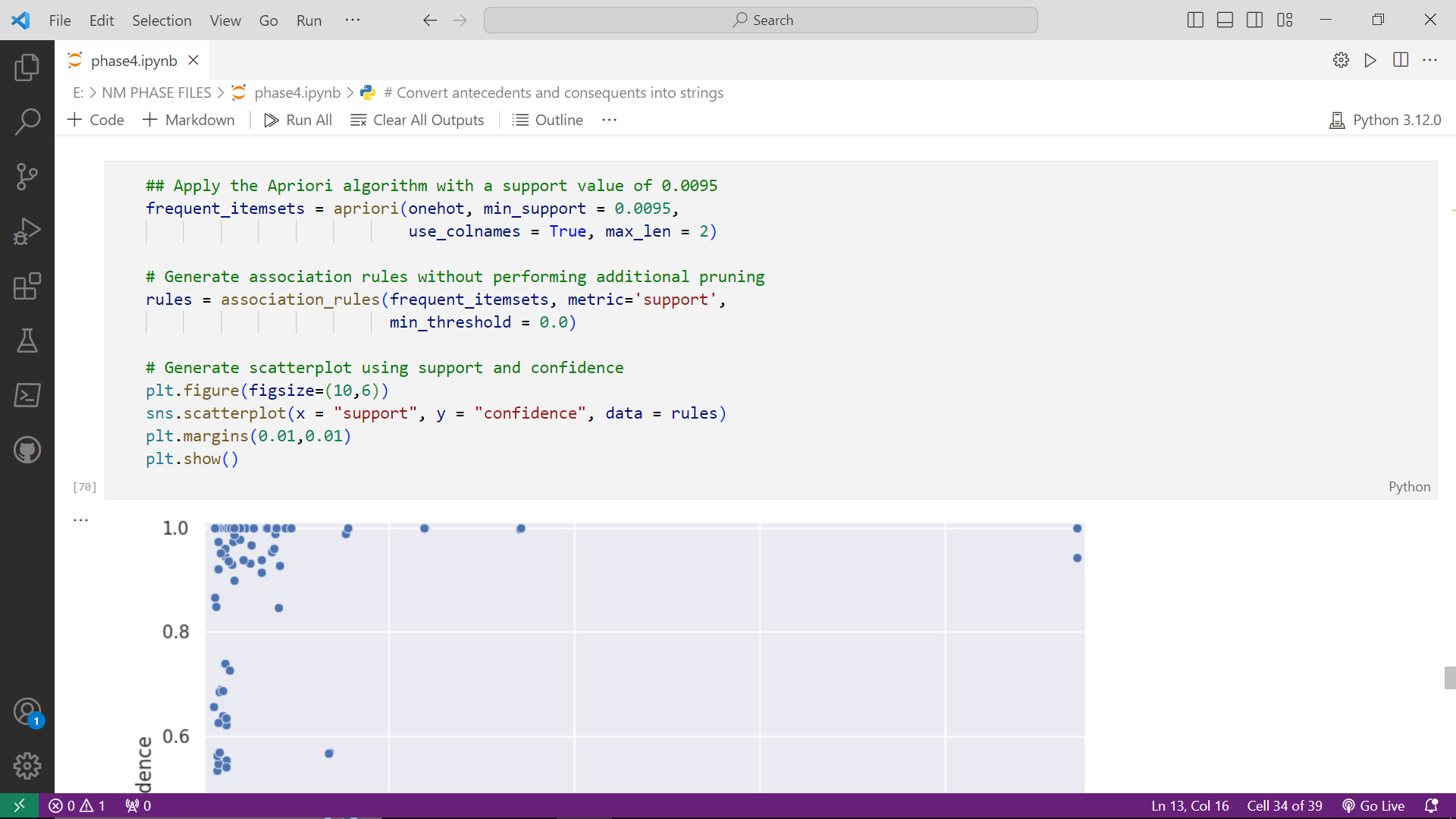


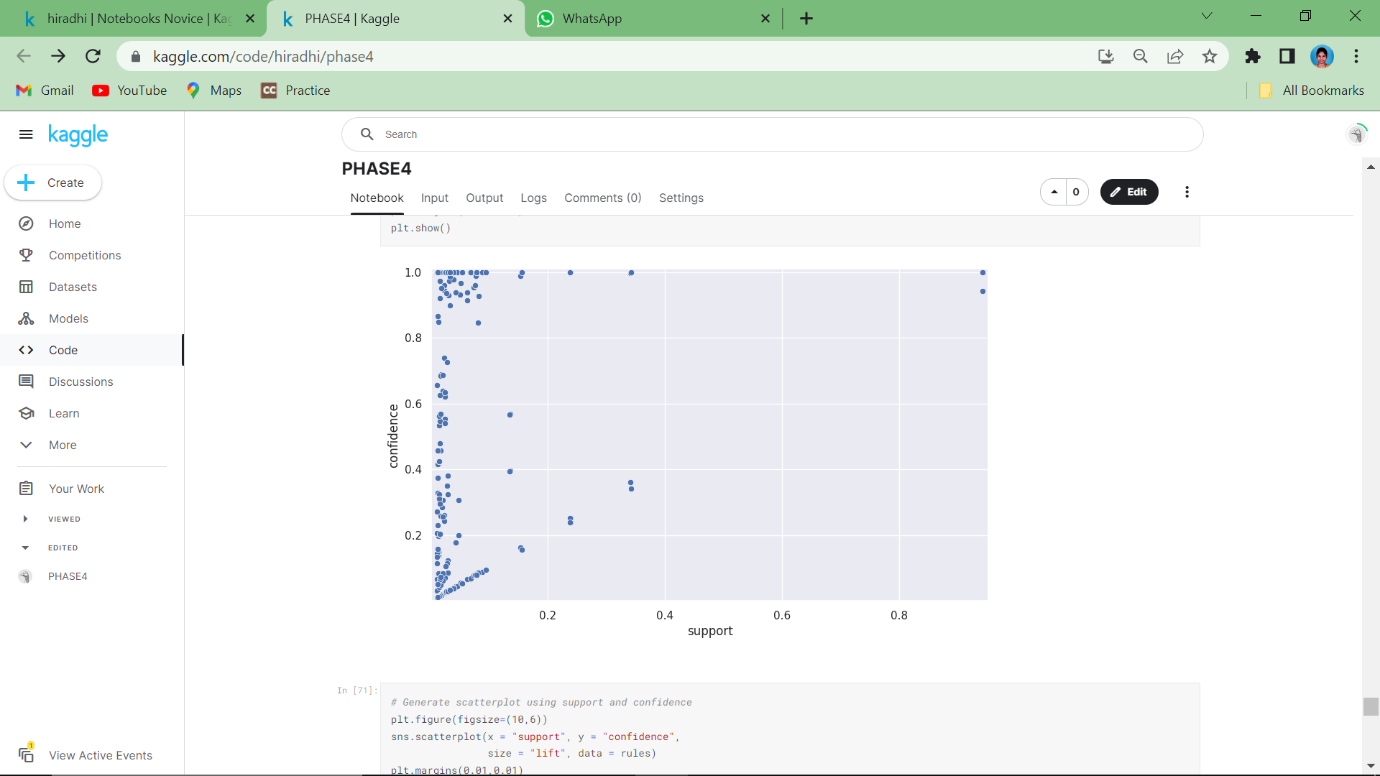


**7.GENERATING SCATTER-PLOT:**

Scatterplots are more commonly associated with the visualization of continuous numerical data to identify patterns and relationships between two continuous variables.

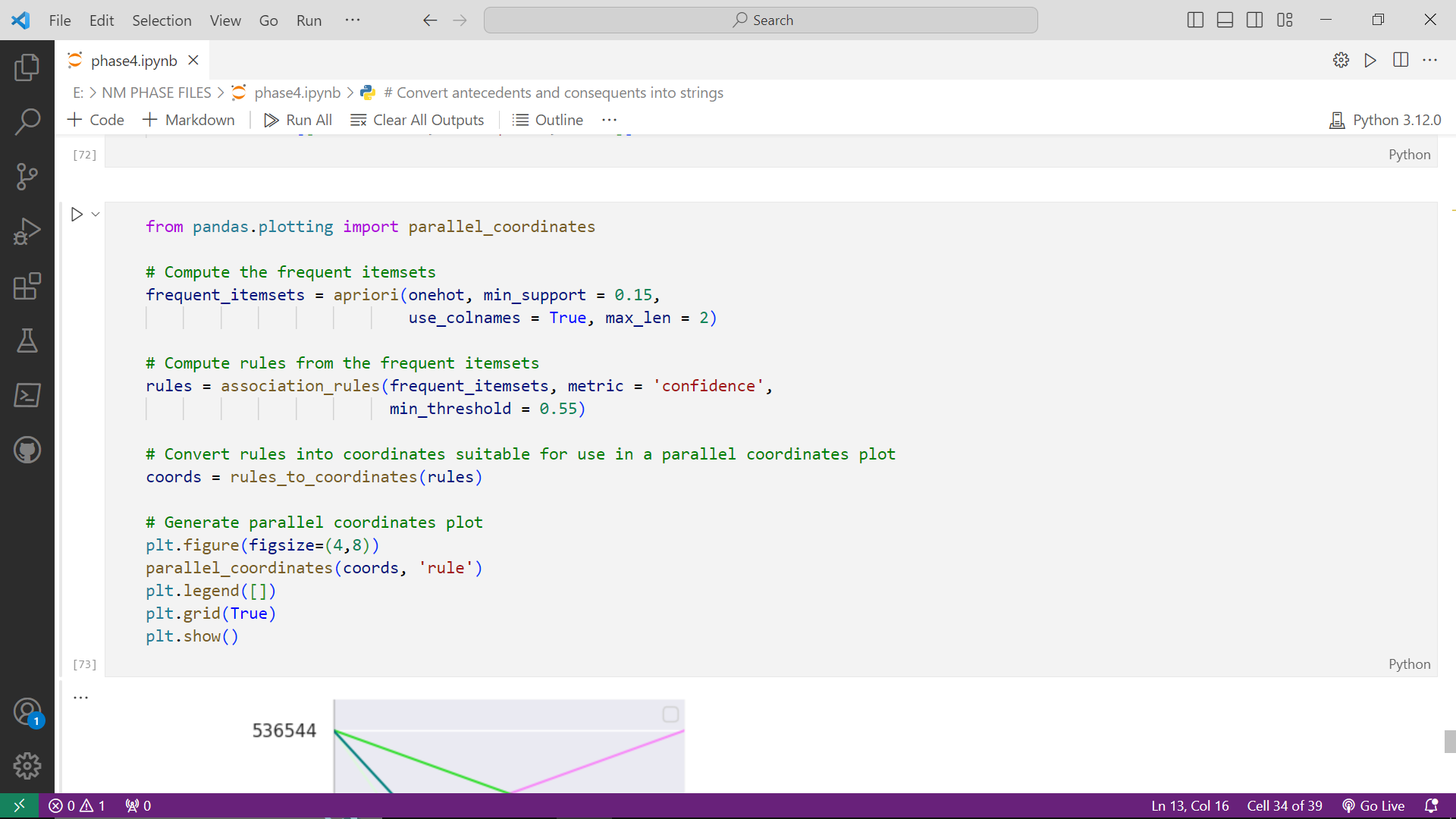
Here, both support and confidence are used for generating scatter-plot.

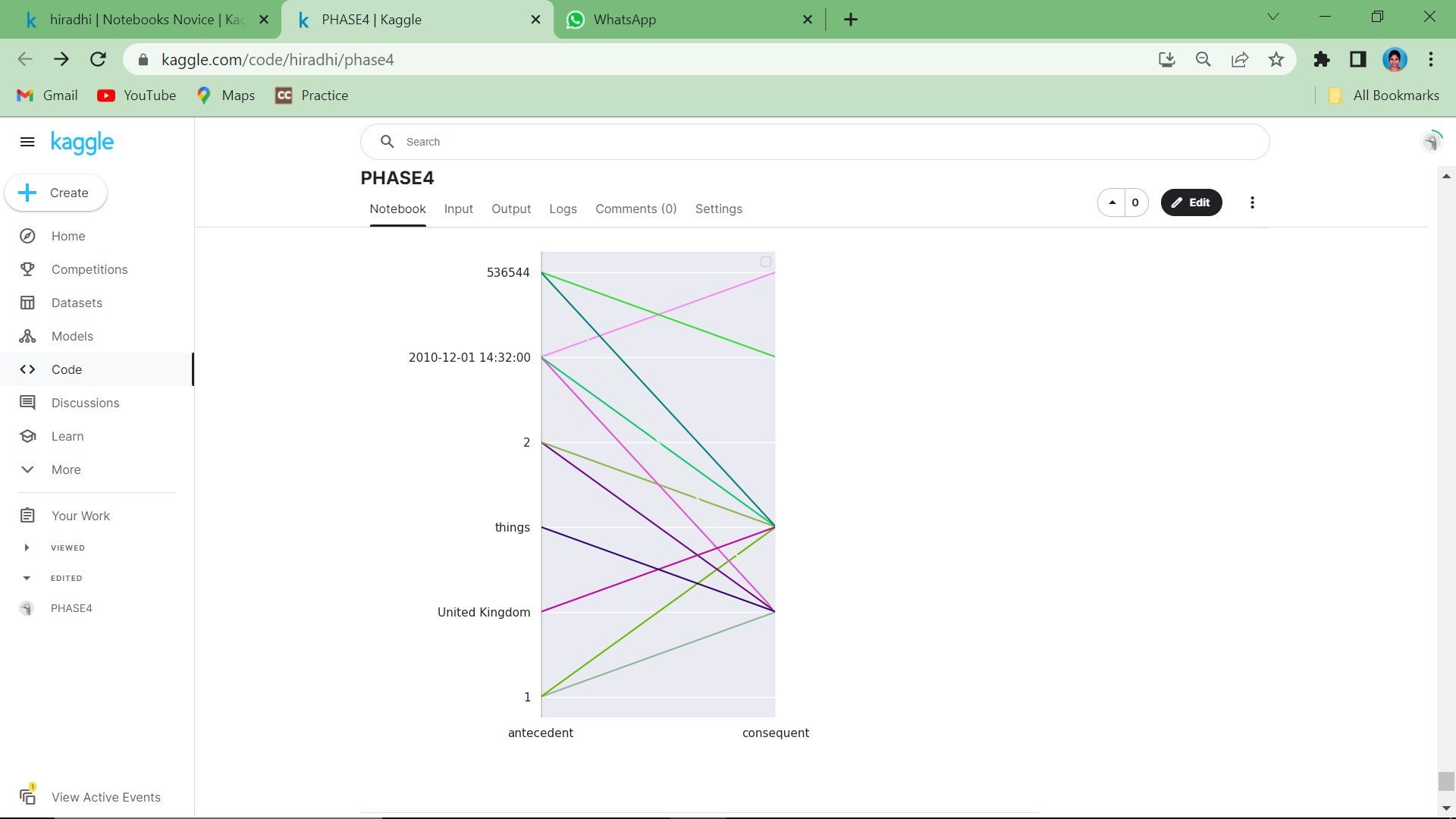




**8.GENERATING PARALLEL COORDINATES PLOT:**

The parallel coordinates plot will allow us to visualize whether a relationship exist between an antecedent and consequent. We can think of it as a directed network diagram. The plot shows connections between 22 objects that are related and indicates the direction of the relationship.





**CONCLUSION:**

Market Basket Analysis is a valuable tool that helps businesses maximize revenue, optimize operations, and provide better customer experiences. It is a versatile technique that can be applied in various industries, from retail to e-commerce, and from hospitality to healthcare, to drive growth and improve decision-making.

It is necessary to revisit the site and changes are made according to current trends.