Bibek Kumar Sah **180010**

Big Data & Data Analytics – II

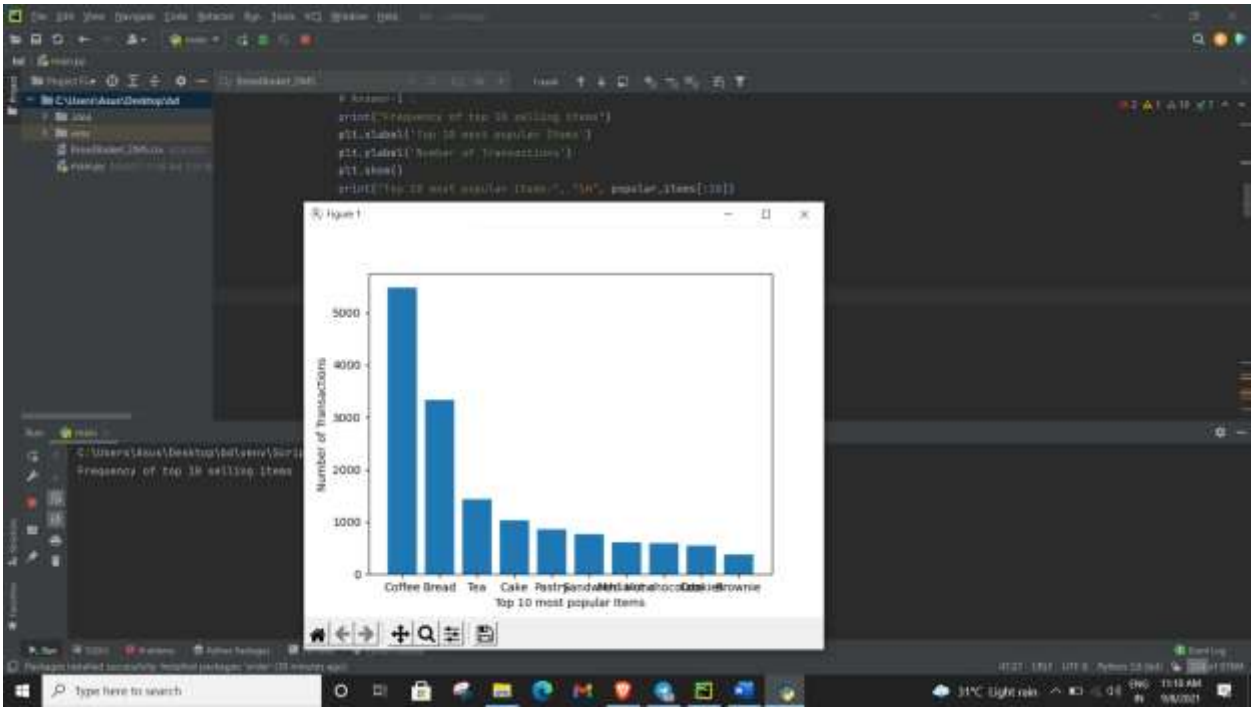
W5 – Project Activity-1

Market Basket Analysis Using Apriori Algorithms

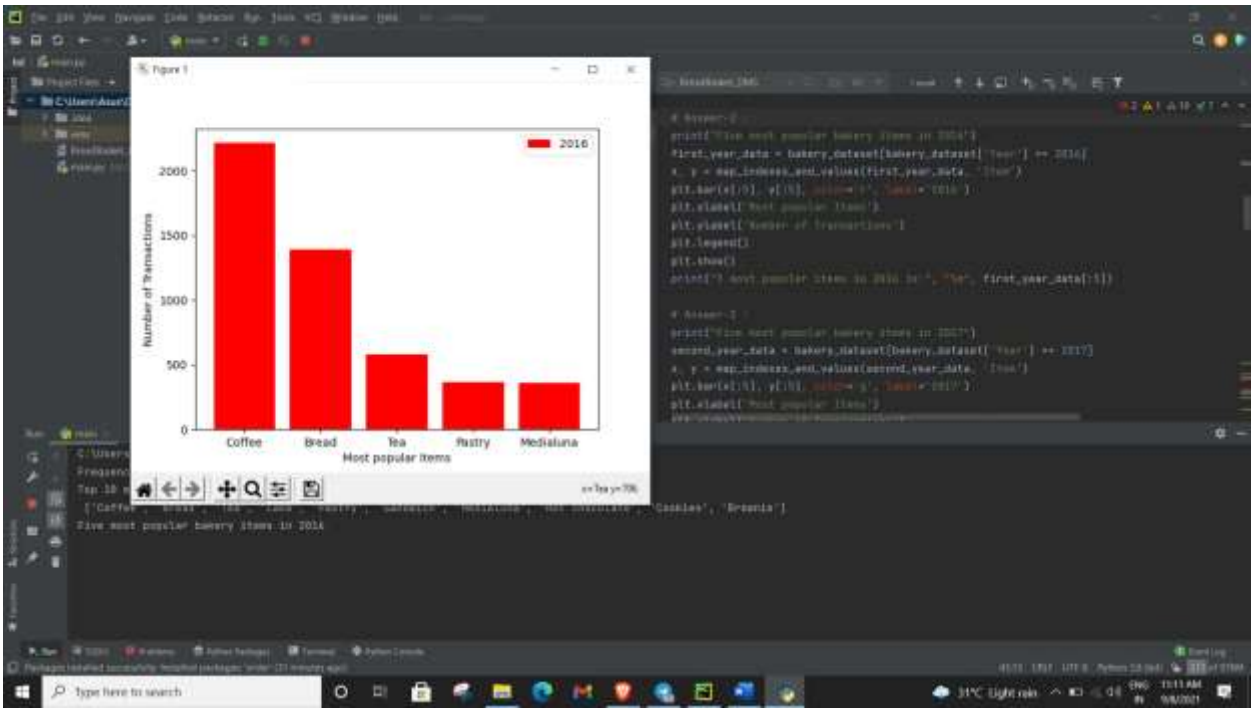
https://github.com/bbksa/Apriori-Algorithm.git

Section: A (Data Analysis)

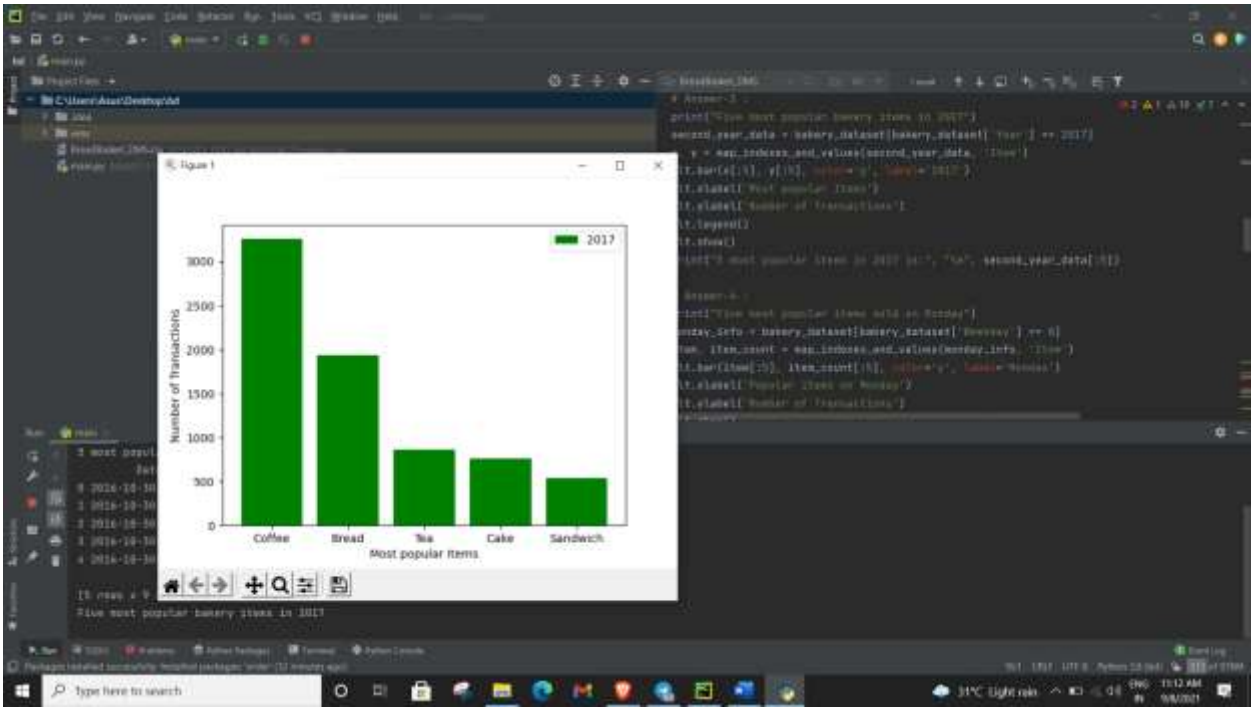
1. **Find out the frequency of top 10 selling items.**



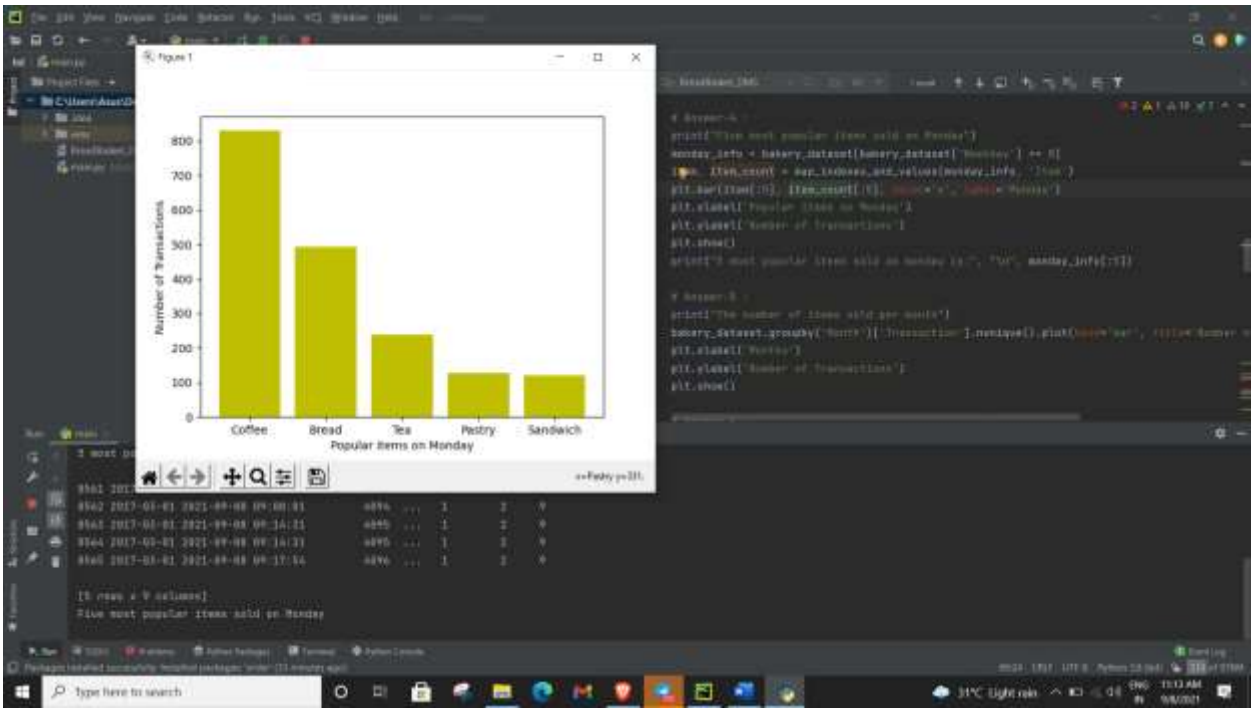
2. **Find out the five most popular bakery items in 2016.**



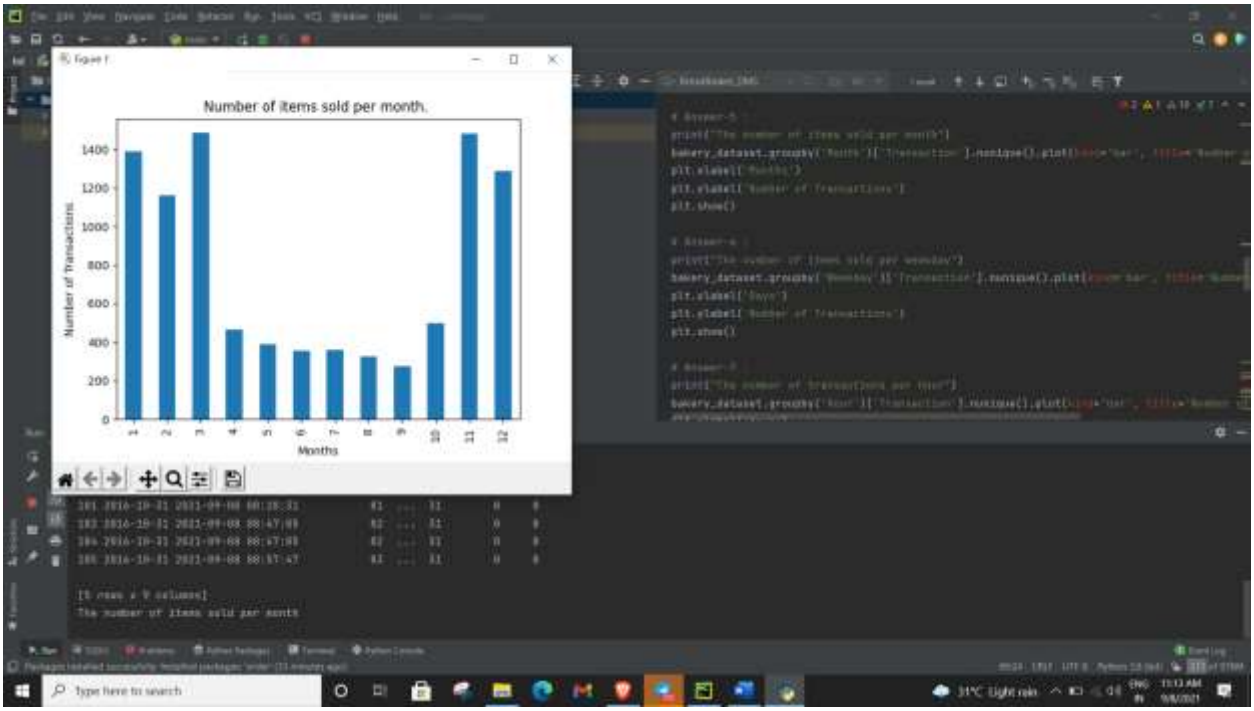
3. **Find out the five most popular bakery items in 2017.**



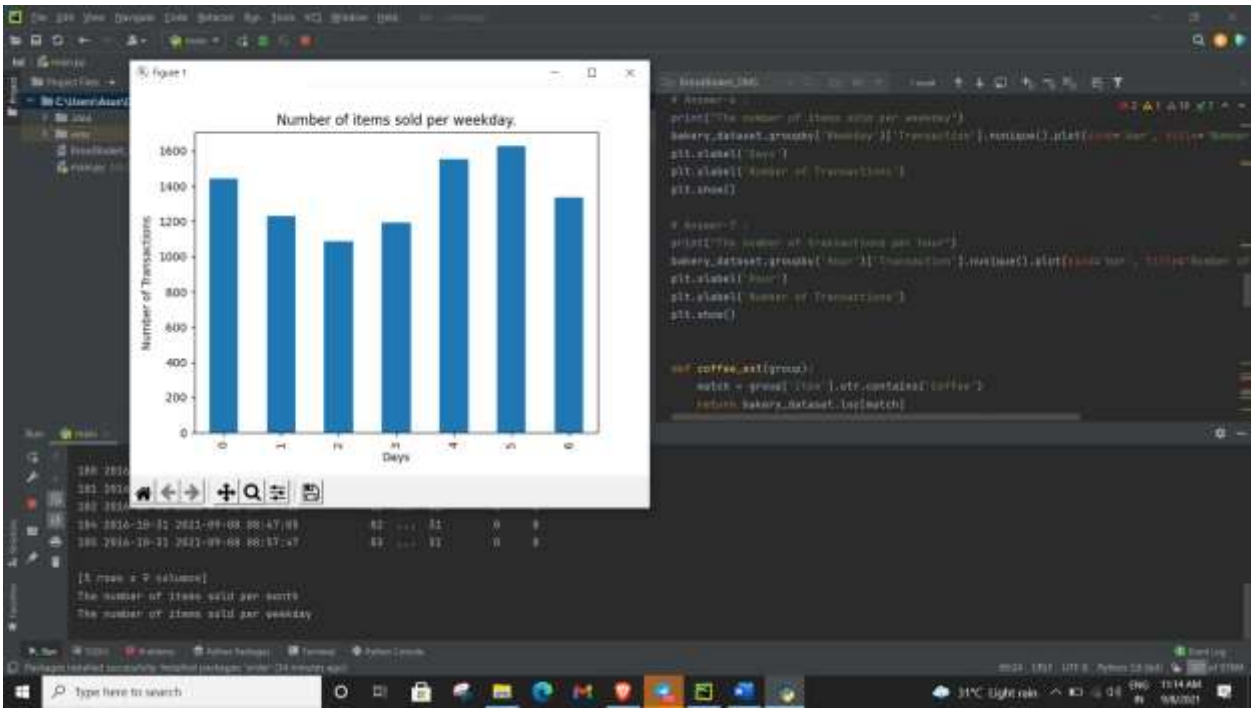
4. **Find out the five most popular items sold on Monday.**



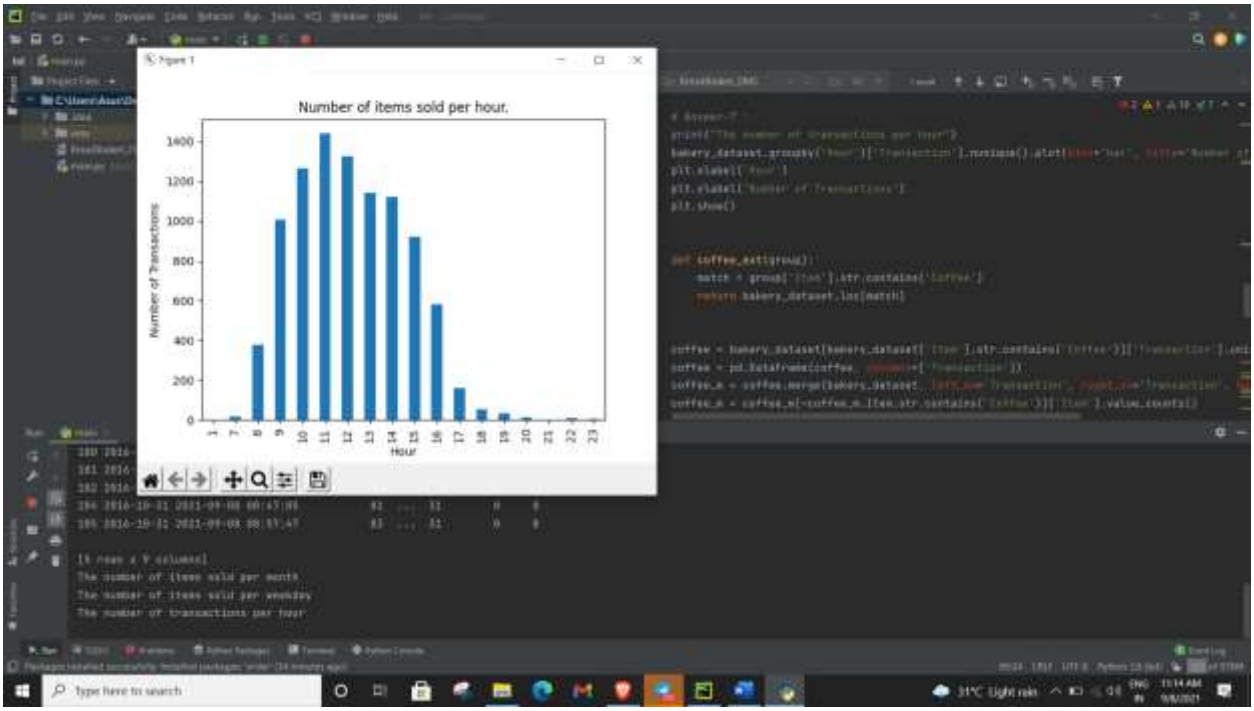
5. **Find out the number of items sold per month.**



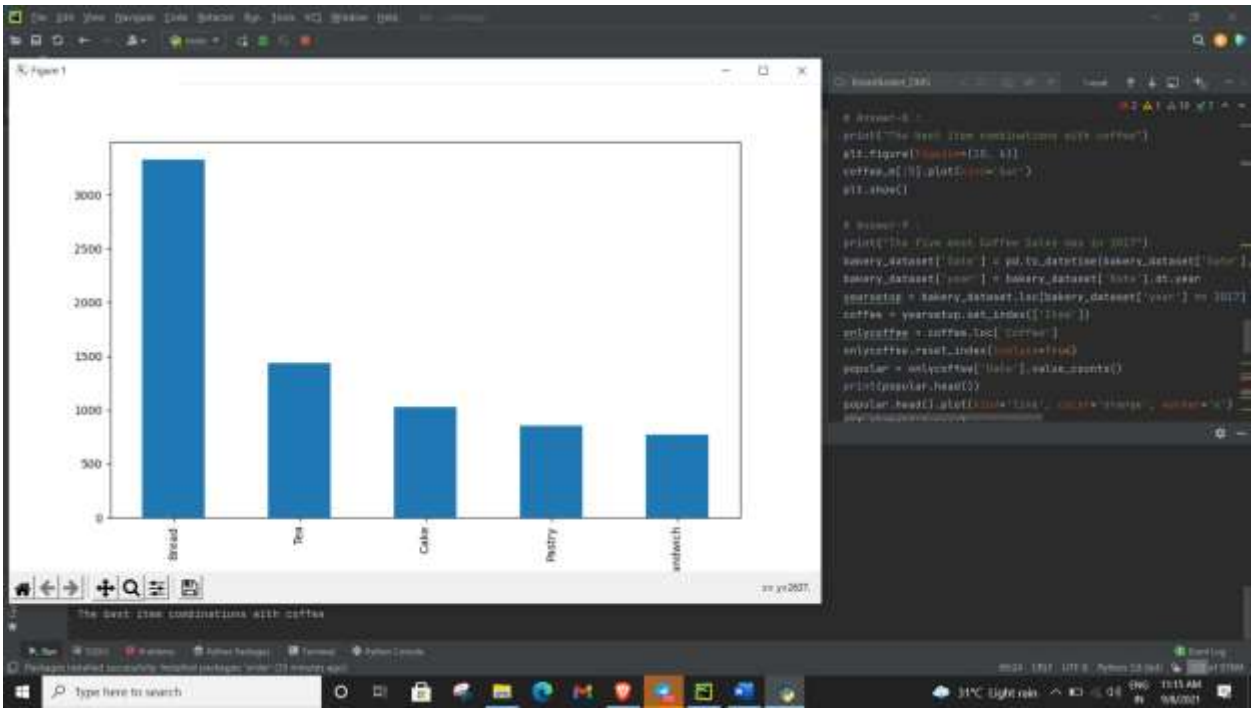
6. **Find out the number of items sold per weekday.**



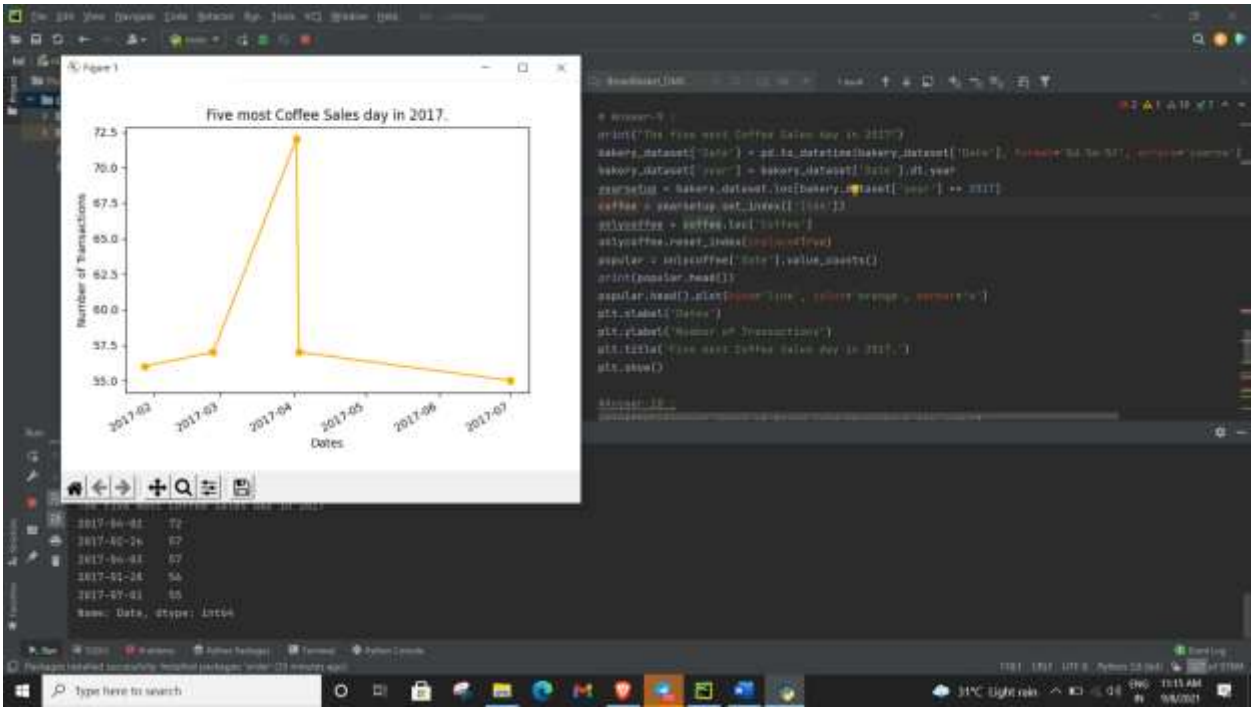
7. **Find out the number of transactions per hour.**



8. **Find out the best item combinations with coffee.**



9. **Find out the five most Coffee Sales day in 2017.**



10. **Find out historical chart of Bread sold throughout the week.**

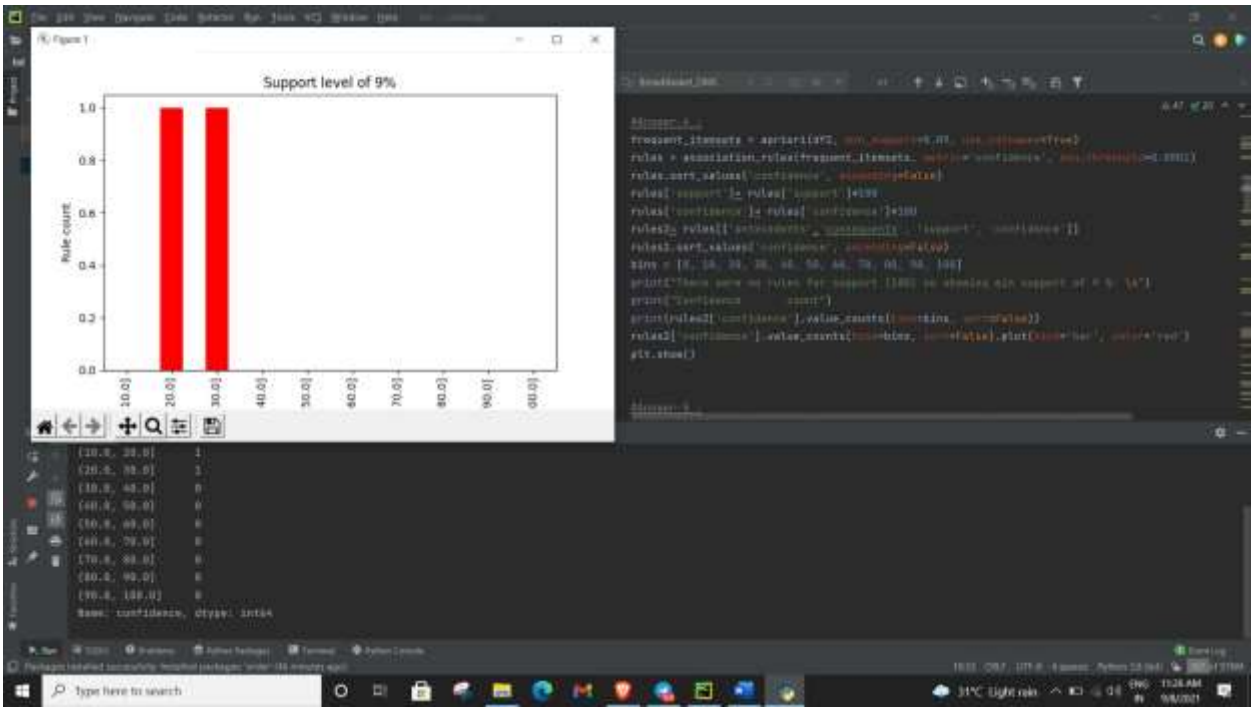
Section-B (Market Basket Analysis

1. **Give an example where you can apply the Apriori algorithm.**
   Apriori algorithm is a classical algorithm in data mining. It is used for mining frequent itemsets and relevant association rules. It is devised to operate on a database containing a lot of transactions, for instance, items brought by customers in a store.
   It is very important for effective Market Basket Analysis, and it helps the customers in purchasing their items with more ease which increases the sales of the markets. It has also been used in the field of healthcare for the detection of adverse drug reactions. It produces association rules that indicates what all combinations of medications and patient characteristics lead to ADRs.
   One basic example is when we go grocery shopping then which items, we frequently purchase together has been analyzed by the shop owner, is using apriori algorithm. So that the shopkeeper then arrange that frequently bought together items in same shelf so that it will be easy to buy by the customer.

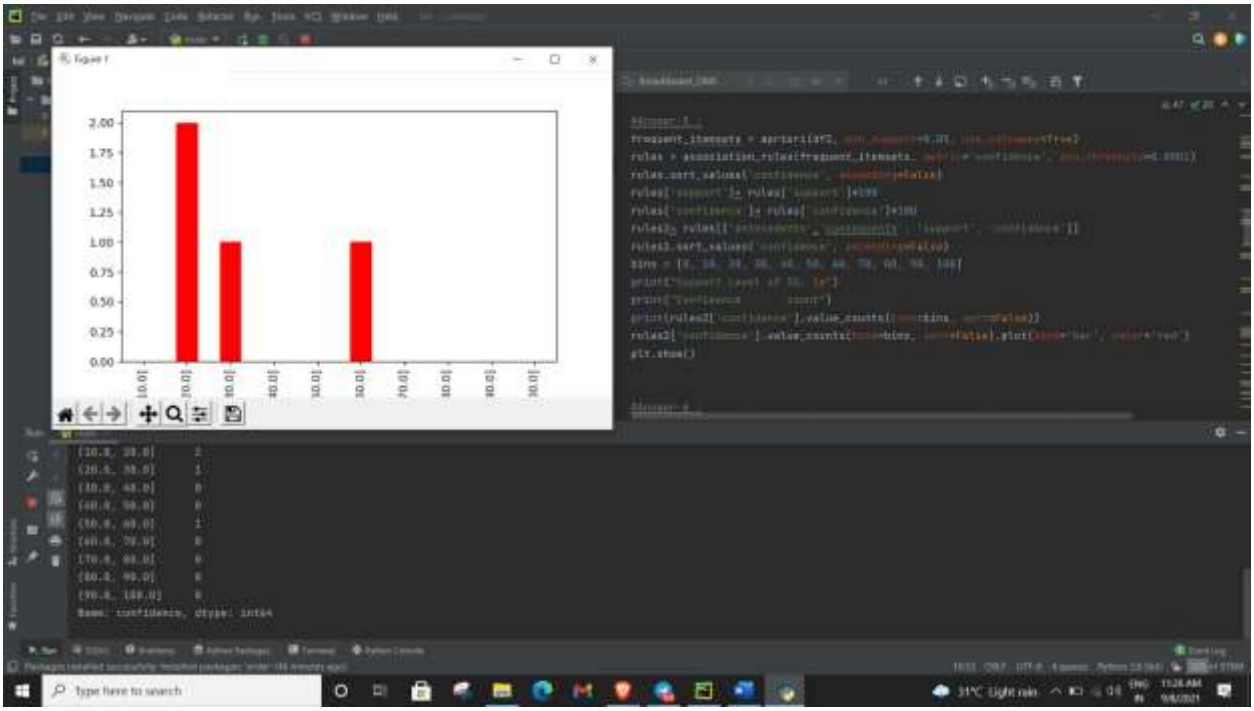
2. **What happens when we decrease the support level? Why?**
   Support represents the popularity of that product of all the product transactions. Support of the product is calculated as the ratio of the number of transactions includes that product and the total number of transactions. Support of the product = (Number of transactions includes that product)/ (Total number of transactions) When we decrease the support level, we decrease the amount of popularity we need.

3. **What happens when we increase the confidence level? Why?**
   This explains how likely Y is purchased when X is purchased. This defines association between two items. For example, when a person buys milk is more likely to buy bread as well or vice versa. This is measured by the proportion of transactions with item X, in which item Y also appears. Expressed as {X -> Y}. Calculated by the proportion of number of transactions in which both (X & Y) occurs to support of the item X. When we increase the confidence level, we tend to increase the probability of that item to be transacted with the compared item.

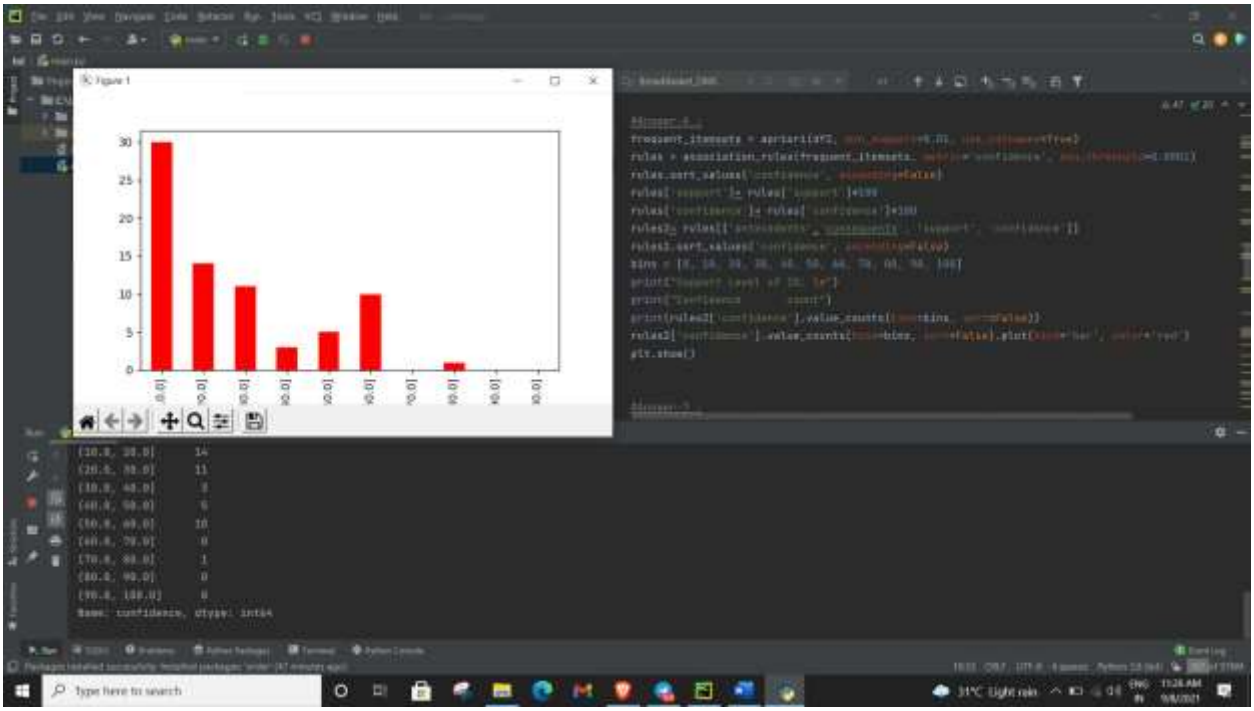
4. **How many rules are generated with a support level of 5% and a confidence level of 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%? Also visualized the generated rules.**



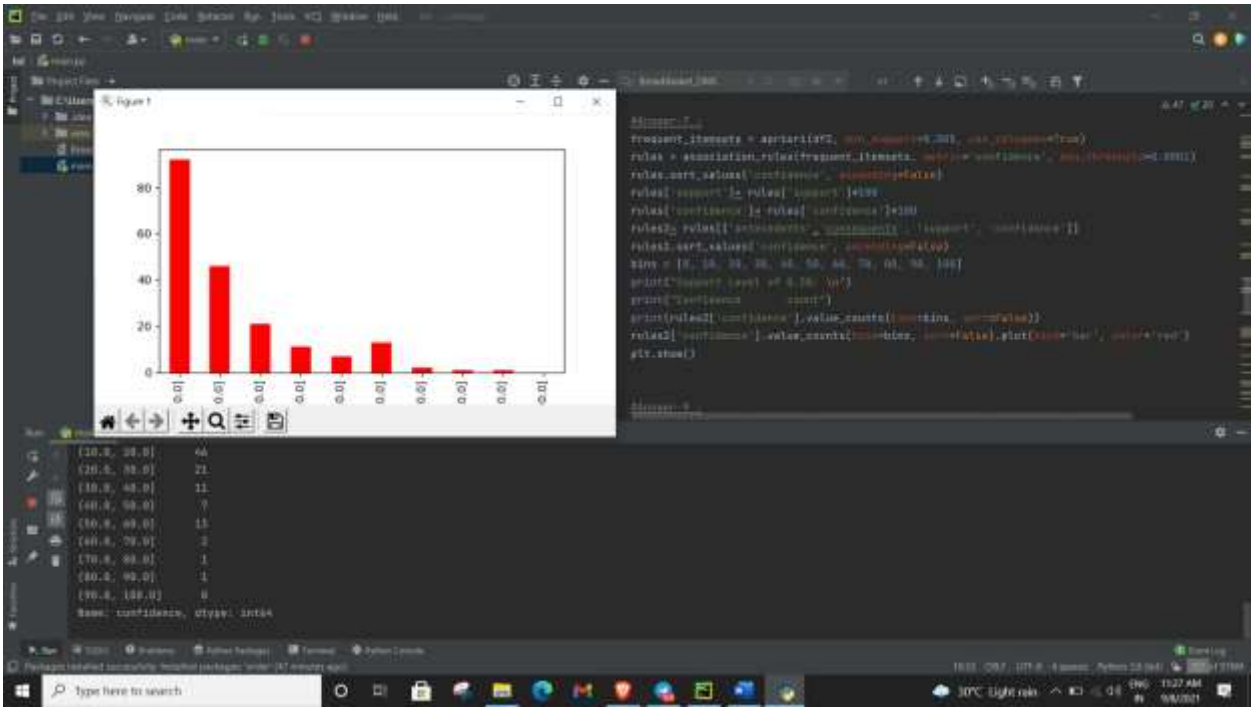
5. **How many rules are generated with a support level of 1% and a confidence level of 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%? Also visualized the generated rules.**



6. **How many rules are generated with a support level of 0.5% and a confidence level of 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%? Also visualized the generated rules.**



7. **Analyze the results of questions 4, 5, 6 and choose the optimal threshold value for support and confidence. What values do you choose for support and confidence? Why?**



8. **Execute the Apriori algorithm with the value of support and confidence obtained in the previous question-7 and generate the association rules. Also interpret the result.**

9. **What recommendations would you give to the owner of the bakery?**