**ONLINE RETAIL SALES**

**Group 09**

Darshan Reddy Chilakala Krishna Reddy

Sai Praneetha Jaladanki

Surya Chandra Phani Santhosh Dulam

1. Abstract and Introduction

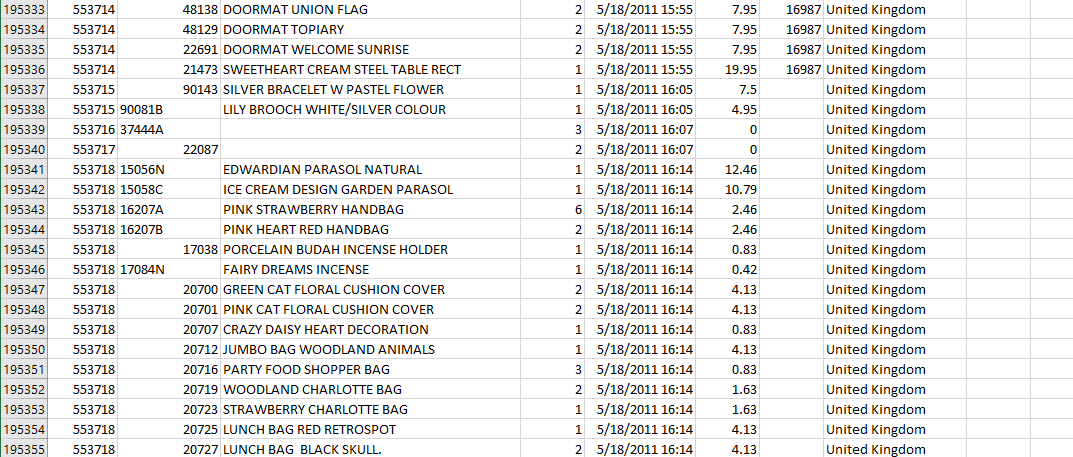
Online retail industry has been on the grow at an exponential rate for the past decade. Online retail gives the buyer a similar experience of an offline retail except that the buyer can have many more advantages. The main purpose of the retail domain analysis is to better understand the customers, purchase pattern of the customers and sales trends that helps in getting to know about the customers.

The data set features transactions occurring between 12/01/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The dataset in hand has been converted into various formats like .csv and .arff from .xlsx for various actions that need to be performed on the data.

The main goal of the analysis is to draw most favorable outcomes that help us to beter understand the sales trends, knowledge about the customer base and a brief idea of how the sales are going to be depending on the ongoing sales.

1. Method and Pseudo Code

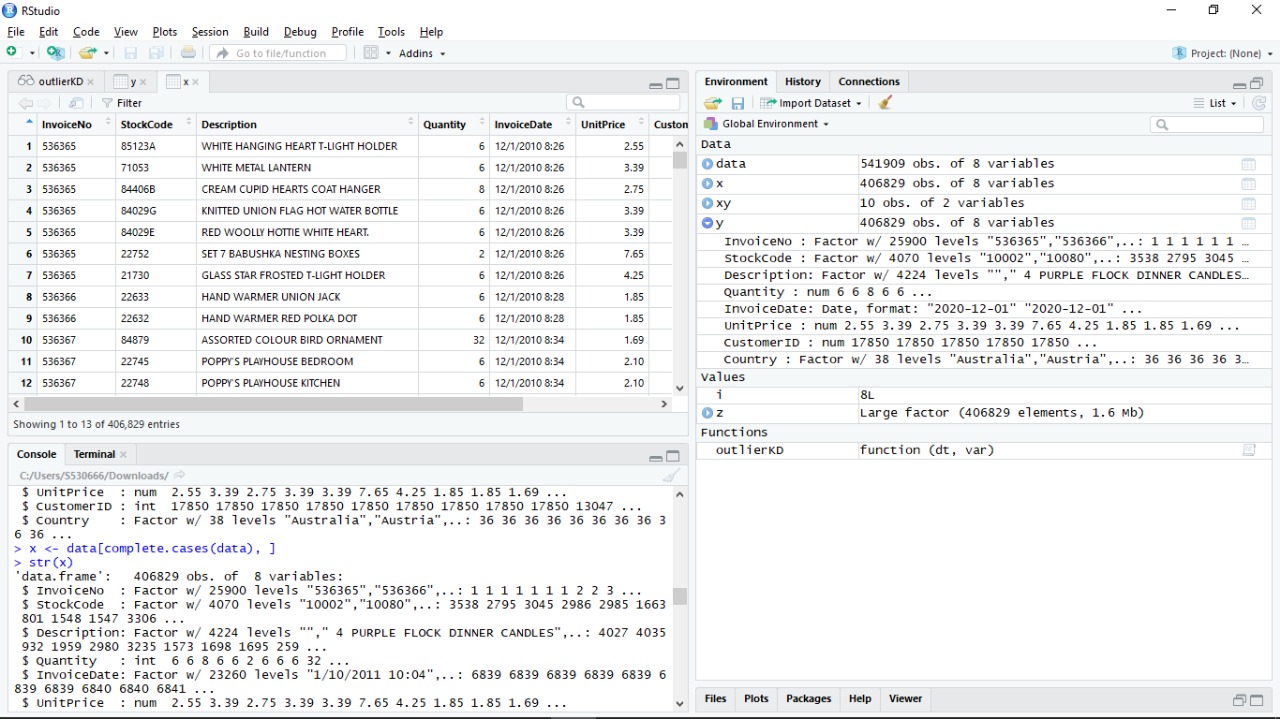
The raw data has 8 attributes and 5,81,587 rows. Invoice, description, quantity, unitPrice, Invoice date, CustomerID, Country are the attributes for our dataset. We have introduced another attribute called total price which is a product of unitPrice and quantity.

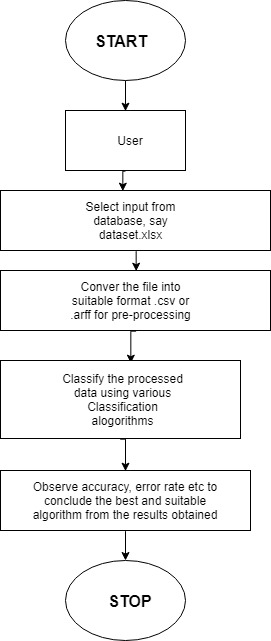


We used R to preprocess the data to clean data, remove noisy data and inconsistent data. We have used Weka for Prediction and Apriori, to predict the sales for next two months and to know what item have been bought most frequently, respectively. We have used Tableau and Microsoft Excel to visualize our outcomes.



|  |  |  |
| --- | --- | --- |
| **Dimension** | **Data Type** | **Description** |
| Invoice | Nominal | Invoice Number |
| Stockcode | Nominal | Code of the product |
| Description | Nominal | Describes the item |
| Quantity | Numeric | Count of the items purchased |
| unitPrice | Numeric | Cost of each unit |
| InvoiceDate | Numeric | Date of the invoice |
| CustomerID | Nominal | Unique ID assigned to the customer |
| Country | Nominal | Location of the item purchased |



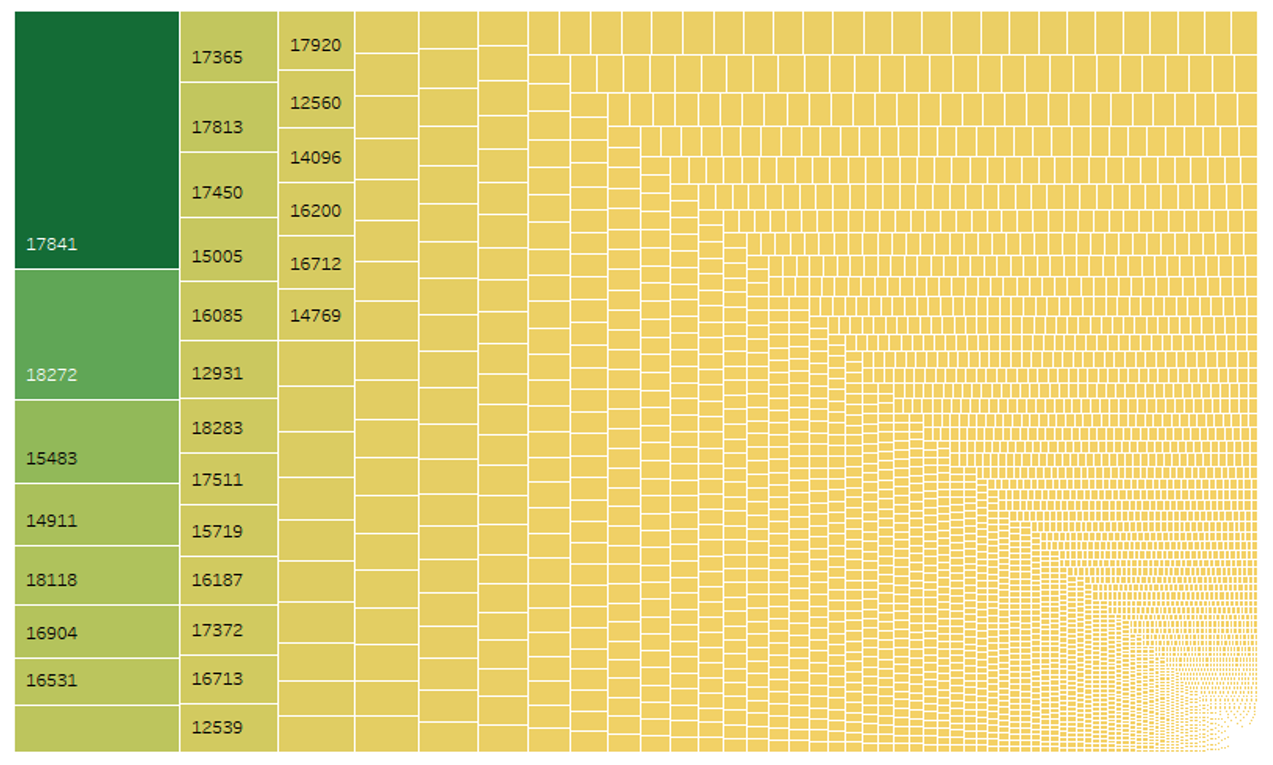


1. Outcomes and Results

Outcome – 1: The customer with most purchases

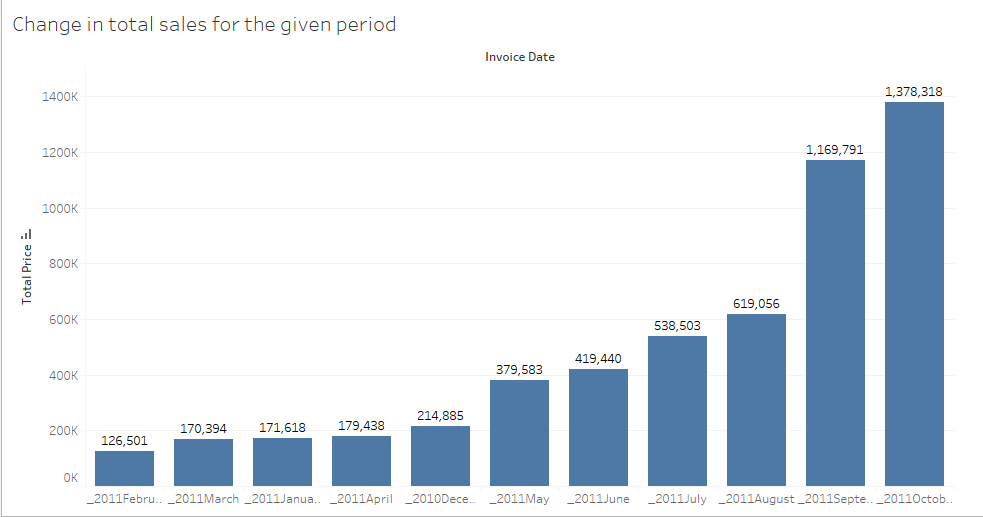
To analyze the most profitable customer to the sales, we have taken into the consideration the customer who spent the most money. The output has been visualized through Tableau. Each block represents an individual customer. The more the area a block occupies more the customer spends. The customers with the ID 17841 has the largest block and hence is the customer with the most number of purchases. The blocks at the rightmost corner represent the customers with least sales.

This analysis helps in finding out the most profitable to the sale



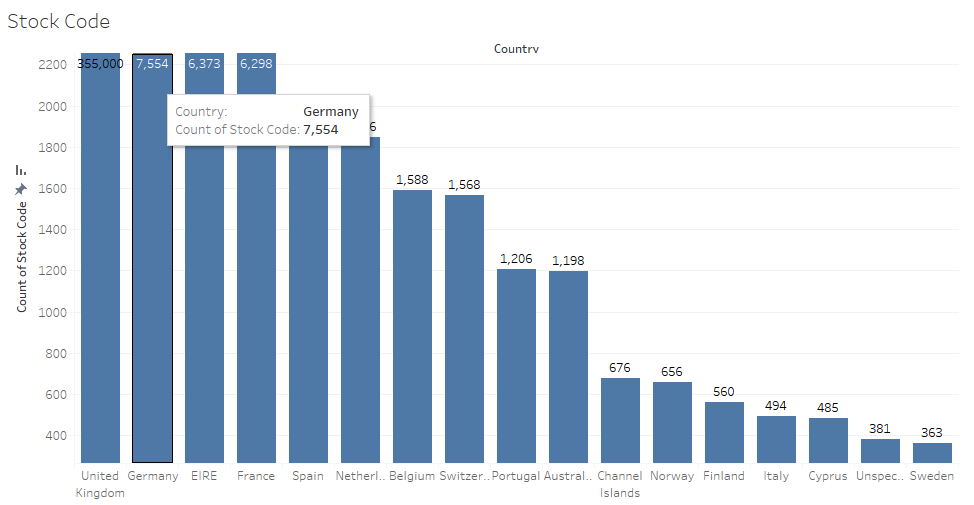
Outcome – 2: Percentage change in the total sale over a period of 11 months

To study the percentage in sales, we have taken the total amount spent every month. The output has been visualized through Tableau. It can be interpreted that there was some decline in sales for the first three months from Dec 2010 to Feb 2011.



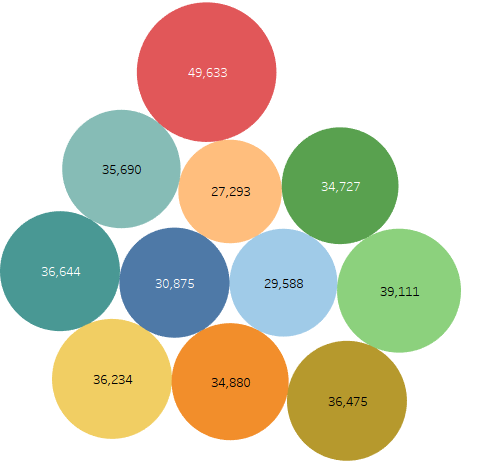
Outcome – 3: Which country has sold the highest stock

The output has been visualized through Tableau. It can be interpreted that the highest sales and stock is in the United Kingdom. The least number of sales can be taken as Sweden.



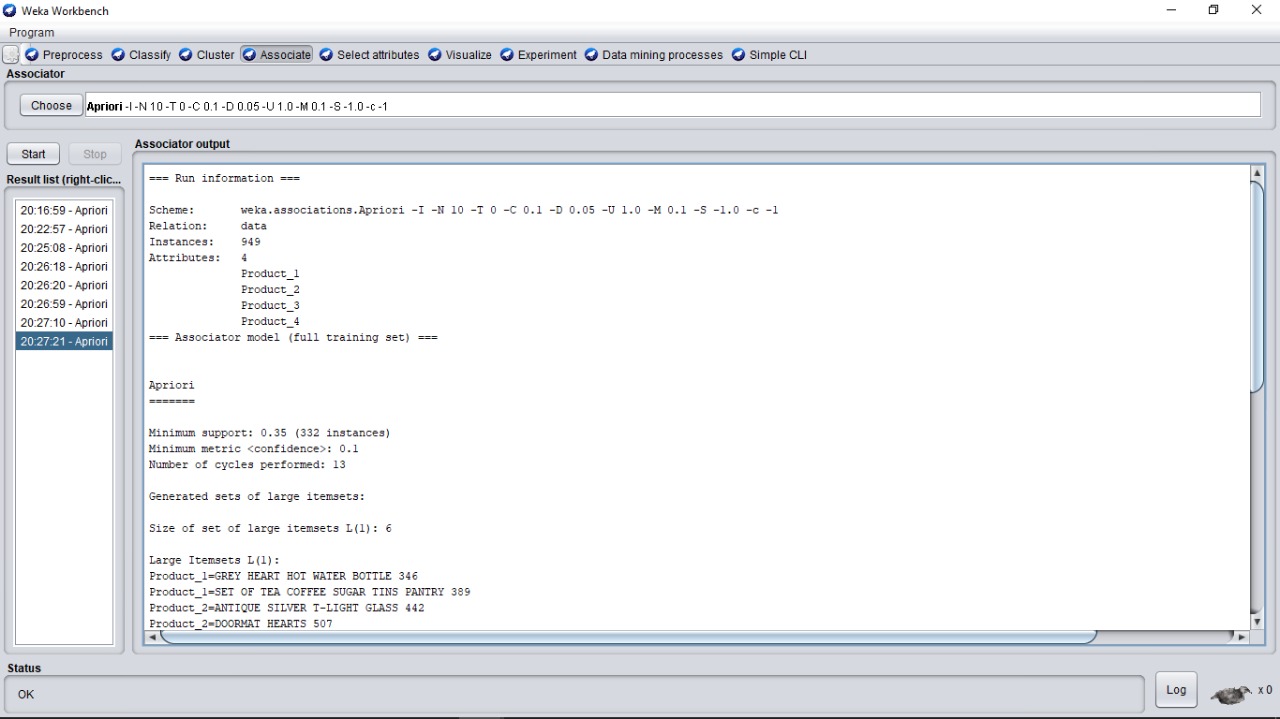
Outcome – 4: Highest number of units sold in which month

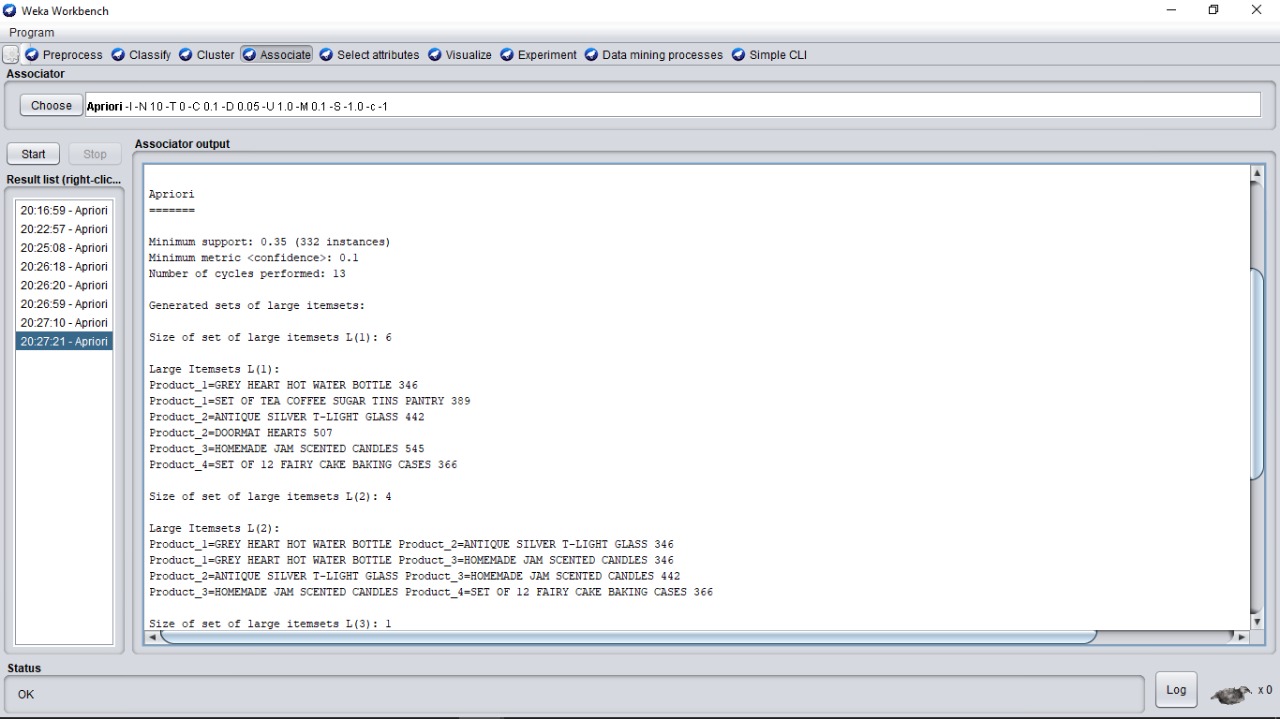
The output has been visualized through Tableau. It can be interpreted that the highest sales amongst the time period is observed in the month of September 2011. The least stock can be visualized in the month of February for the year 2011.

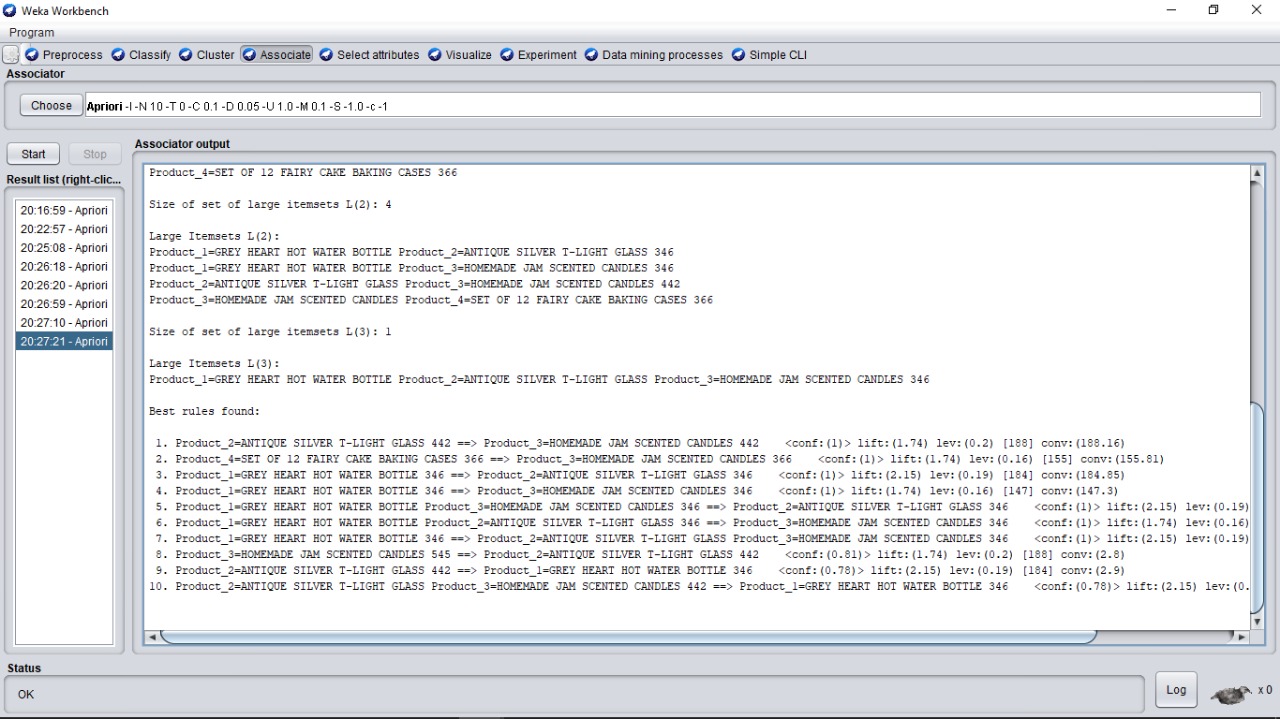


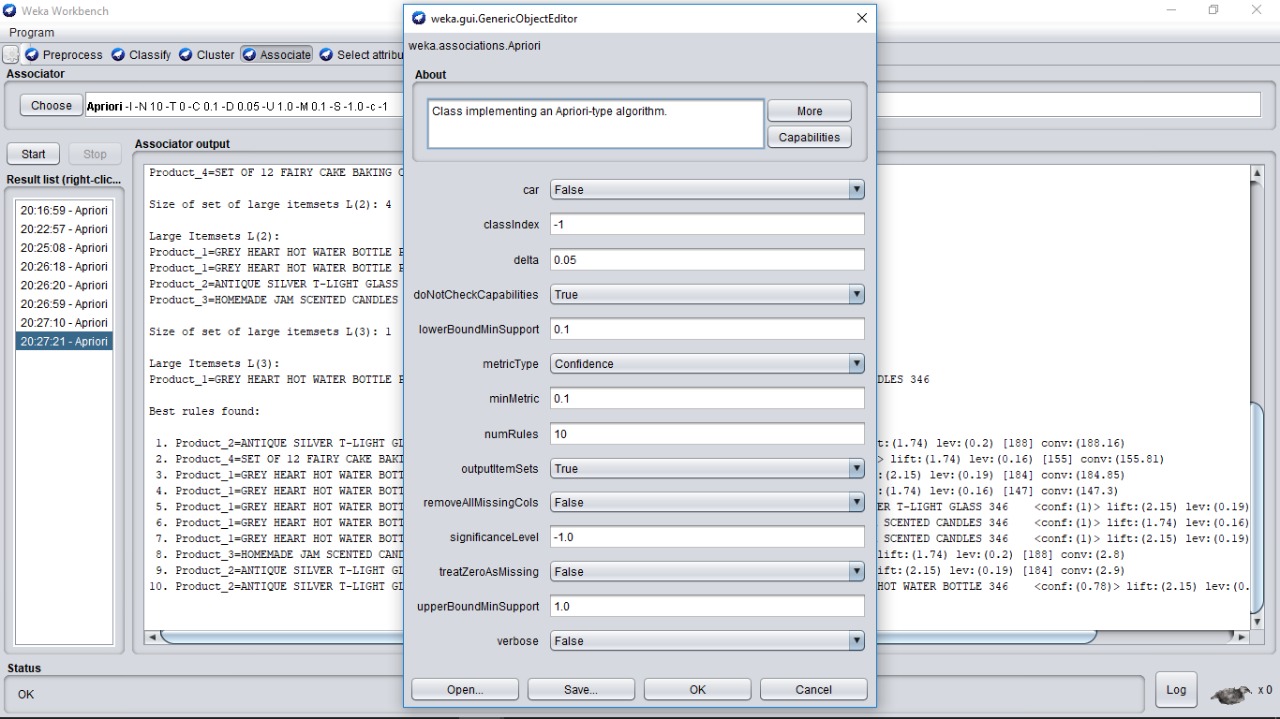
Outcome – 5: Items purchased together frequently

To analyze which of the items have been frequently brought by the customers, we have used Apriori Algorithm. The outcome has been obtained by using Weka. The product is divided into number of columns based on customer ID and considered the product only which had the highest count.







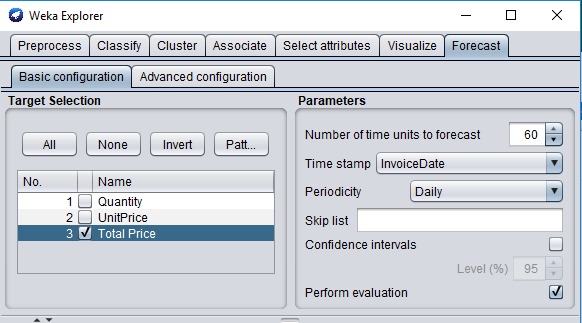


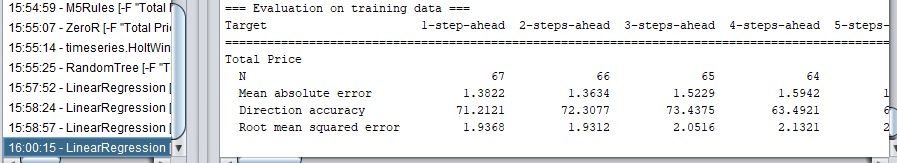
The most frequently purchased items are

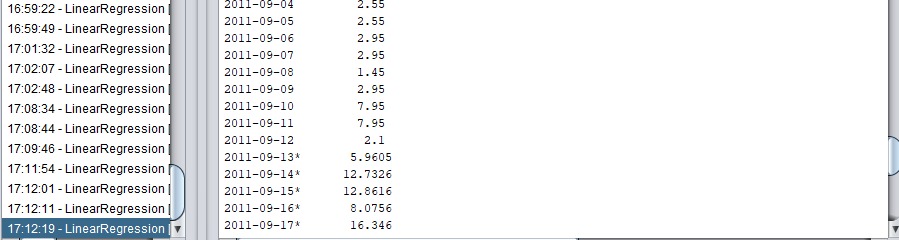
* + Antique Silver T-Light Glass and Homemade Jam Scented Candles
  + Set of 12 Fairy Cake Baking Cases and Homemade Jam Scented Candles
  + Grey Heart Hot Water Bottle and Antique Silver T-Light Glass

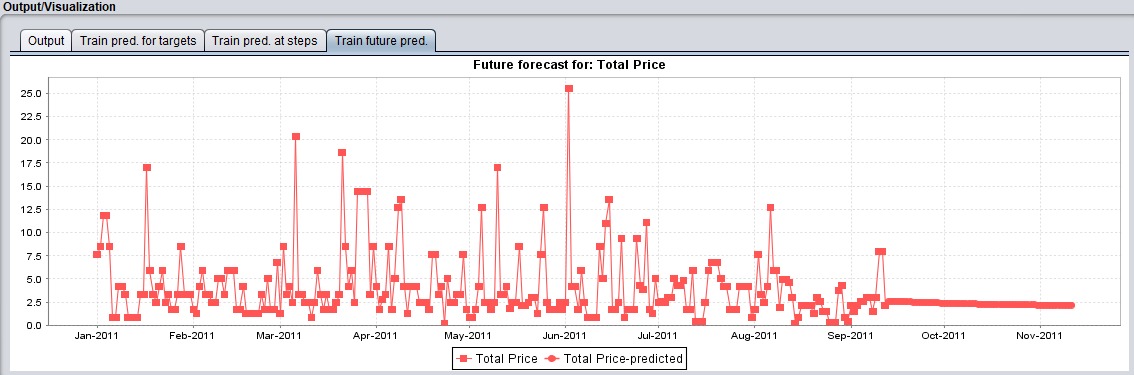
Outcome – 6: Predicting the sales for the next two months

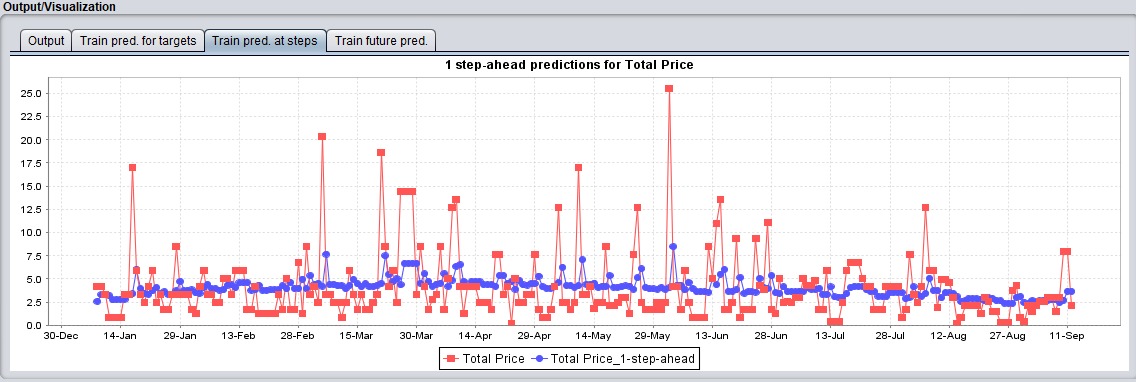
To predict the sales for the next two months considering the daily sales, we used a number of classification algorithms. The output has been visualized through Weka. This analysis helps in finding out a trend for the next two months, September and October 2011. The data considered for training is from 12/01/2010 to 09/12/2011.

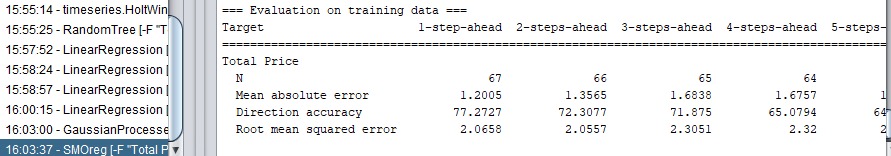
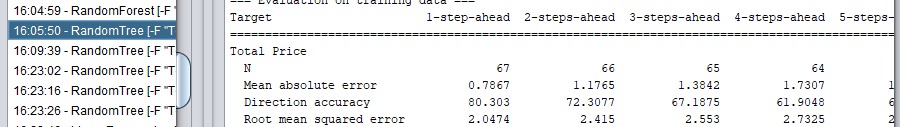
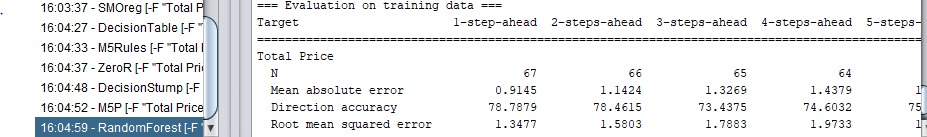










From the results obtained from the various classification algorithms, we have concluded that the best algorithm that suits our dataset is the Random tree. It has the highest accuracy and hence suits the best for classification.

1. Conclusion

A detailed analysis of the retail domain teaches us many dimensions of the business. To help a retail store grow, get profits and expand its customer base, it needs to have a report and timely update of the sale. We have used many a type of charts to show our results and depict the same. Data visualization is a very important of data analysis to understand the various trends and business ideas. Study of change in business sales is very important as it paves a path to the future of the business.

1. References

http://archive.ics.uci.edu/ml/datasets/Online+Retail

http://link.springer.com/article/10.1057/dbm.2012.17

http://www.cs.waikato.ac.nz/ml/weka/downloading.html

https://www.rstudio.com/products/rstudio2/

https://www.tableau.com/