

# N.M.A.M. INSTITUTE OF TECHNOLOGY

(An Autonomous Institution under VTU, Belgaum) (AICTE approved, NBA Accredited, ISO 9001:2008 Certified) NITTE -574 110, Udupi District, KARNATAKA.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**SHOPPING RECOMMENDATION SYSTEM**

A Mini Project Report Submitted by

**Rashmitha(4NM18CS130)**

**Rishika R Palan(4NM18CS134)**

UNDER THE GUIDANCE OF

**Dr. Sarika Hegde**

Associate Professor

Department of Computer Science and Engineering

# in partial fulfilment of the requirements for the award of the Degree of

Bachelor of Engineering in

Computer Science & Engineering

from

# Visvesvaraya Technological University, Belagavi



(ISO 9001:2015 Certified), Accredited with ‘A’ Grade by NAAC

(: 08258 - 281039 – 281263, Fax: 08258 – 281265

Department of Computer Science and Engineering

B.E. CSE Program Accredited by NBA, New Delhi from 1-7-2018 to 30-6-2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

“Shopping recommendation system”

is a bonafide work carried out by

Rashmitha(4NM18CS130) Rishika R Palan(4NM18CS134)

in partial fulfilment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering prescribed by Visvesvaraya Technological University, Belagavi during the year 2020-2021.

It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report.

The Mini project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the Bachelor of Engineering Degree.

Signature of Guide Signature of HOD

**ABSTRACT**

Recommender systems are an integral part of e-commerce websites these days. Customer behaviors can be analyzed for improving the user experience of the customers on these e-commerce sites. Recommendation systems generally recommend products based on the order history or the ratings provided by the customer. These systems can boost sales and increase the profits incurred.

In this machine learning mini-project, the Apriori algorithm is used to recommend products to the customer based on their city and state. Apriori algorithm generates frequent itemset from the baskets formed using market basket analysis.

.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sno.** | **Titles** | **Page No** |
| 1 | Introduction | 5 |
| 2 | Literature Survey | 6 |
| 3 | Design | 7 |
| 4 | Implementation | 8-12 |
| 5 | Screenshots of result | 13-15 |
| 6 | Conclusion | 16 |
| 7 | References | 17 |

**INTRODUCTION**

The project aims at improving recommendation in E-Commerce websites. For that there will be User Interface development which acts as a kind of input module to the project. The existing E-Commerce websites aims at providing recommendation based on the transaction history of the user and sometimes they recommend latest products and also highly rated products.

In this project we aim at recommending products to the user based on the city and state of user. Apriori Algorithm is the main algorithm used in our project. Apriori Algorithm is the general algorithm which can be used by developers according to their need and implement it in their projects.

**LITERATURE SURVEY**

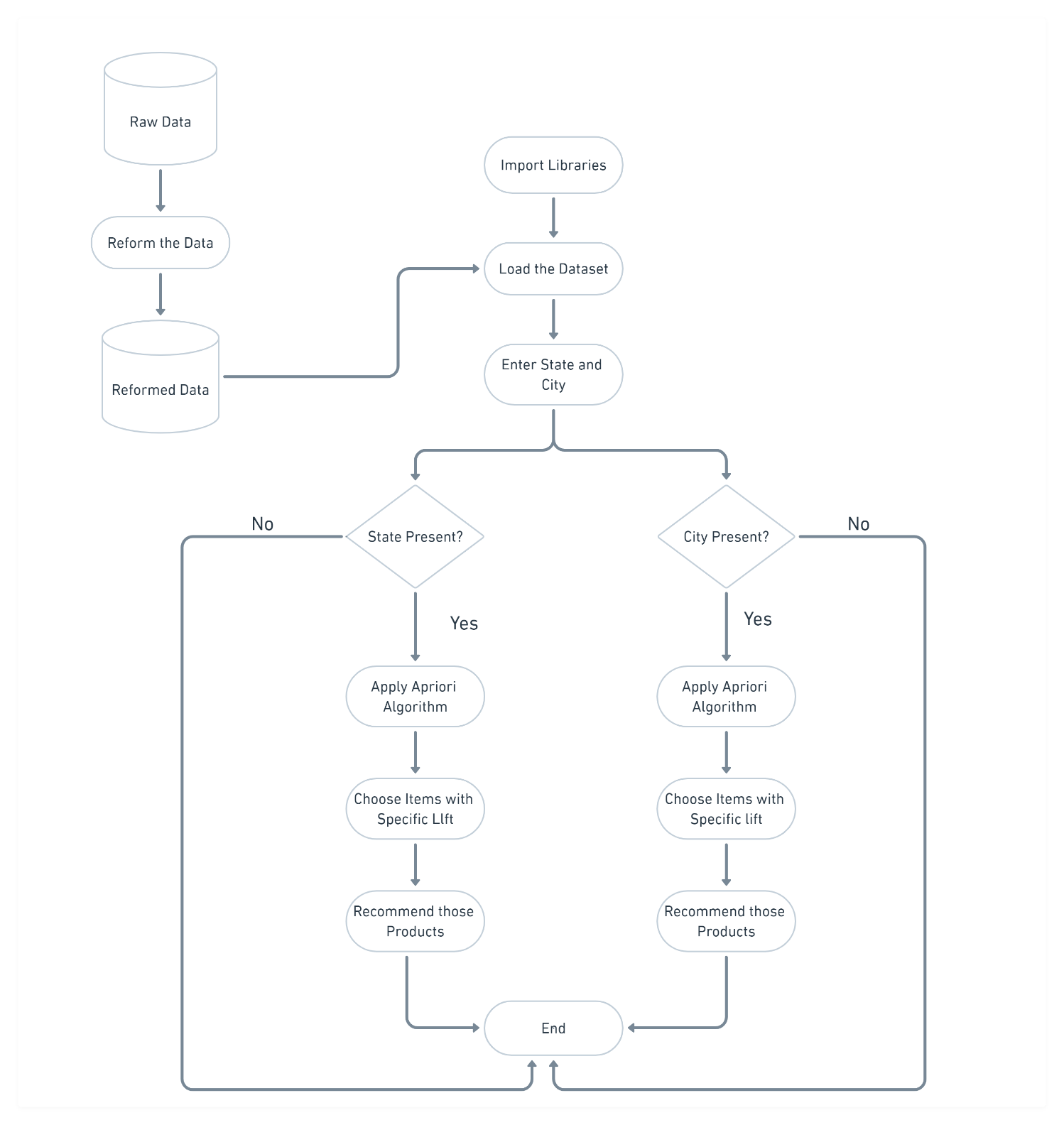
With the rapid development of information technology, the information overload problem in e-commerce sites is becoming increasingly serious. It is difficult for people to obtain their own needs from the massive items of information quickly. Recommendation systems contribute to alleviating the problem of information overload that exists on the e-commerce site.

A recommendation system is the use of statistical and knowledge discovery techniques to solve the interaction with the target customers to provide products recommended problem. Products that are recommended to consumers valuable by improving the data mining efficiency.

Apriori algorithm increases the efficiency of data mining to achieve the unity of real-time and recommendation accuracy. Apriori is an algorithm used for Association Rule Mining. It searches for a series of frequent sets of items in the datasets. It builds on associations and correlations between the itemsets. It is the algorithm behind “*You may also like”* where you commonly saw in recommendation platforms.

**DESIGN**

**Flowchart of the recommendation system**

****

**IMPLEMENTATION**

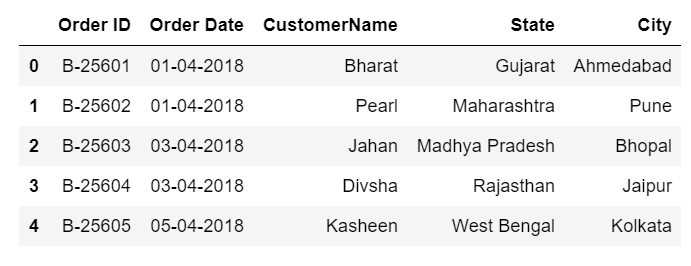
**Dataset explanation:**

Apriori Algorithm is the main algorithm used in our project. Apriori Algorithm is the general algorithm which can be used by developers according to their need and implement it in their projects. The next step of refinement is removal of the data which is not necessary for the analysis, here we have two CSV files Customer Details and Order Details we combine these two CSV files into one dataset and remove the unnecessary columns while recommending the products to the customer based and city and state.

**Screenshots of dataset:**

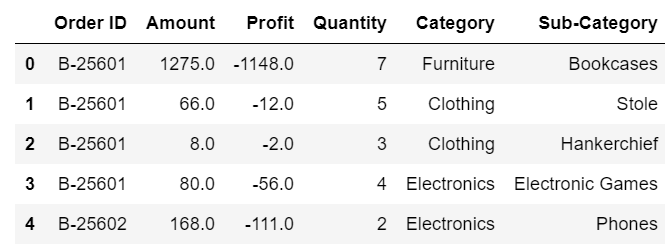
1. Customer Details

This contains five features Order ID, Order Date, Customer Name, State and City. This dataset contains 560 records under five attributes.

****

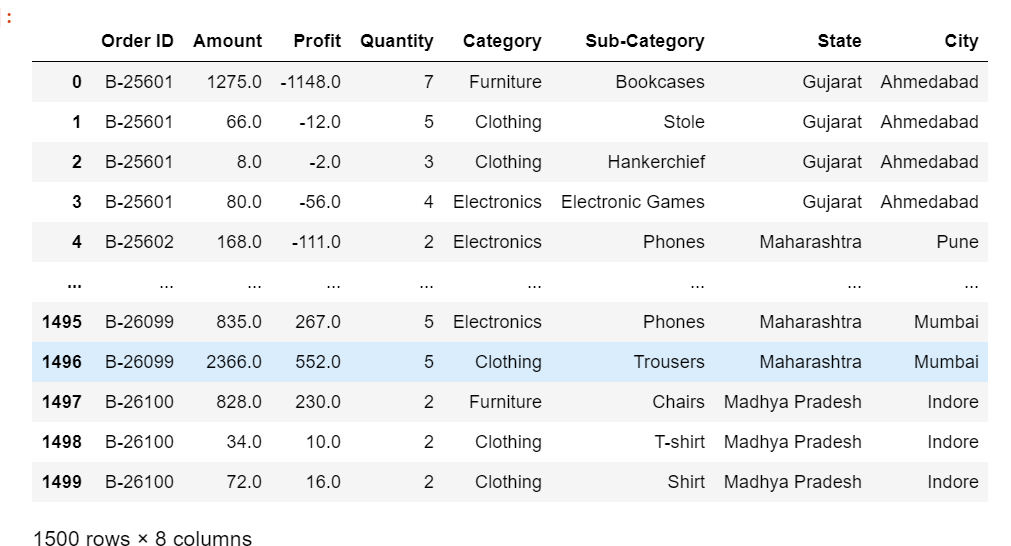
1. Order Details

Order Details dataset contains six attributes Order Id, Amount, Profit, Quantity, Category and Sub-Category. This dataset contains 1500 records under six attributes.

****

1. Merged Dataset

Merged Dataset is obtained by merging Customer Details and Order Details based on Order ID and it contains 1500 records under 10 attributes. Here we are recommending based on city and state so we are dropping the attributes Order Date and Customer Name and removed the extra spaces present in the state attribute. After cleaning the dataset, it contains 1500 records under 8 attributes.



**Algorithm Explanation:**

Apriori algorithm assumes that All subset of a frequent itemset must be frequent. If an itemset is infrequent, all its supersets will be infrequent. Association Rule is one of the very important concepts of machine learning being used in market basket analysis.

**Steps**

* Scan the opinion data set and determine the support(s) of each item.
* Generate L1 (Frequent one item set). Use Lk-1, join Lk-1 to generate the set of candidate k - item set.
* Scan the candidate k item set and generate the support of each candidate k – item set.
* Add to frequent item set, until C=Null Set.
* For each item in the frequent item set generate all nonempty subsets.
* For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence. Then add to Strong Association Rule.

Consider the following database, where each row is a transaction and each cell is an individual item of the transaction:

|  |  |  |
| --- | --- | --- |
| alpha | beta | epsilon |
| alpha | beta | theta |
| alpha | beta | epsilon |
| alpha | beta | theta |

The association rules that can be determined from this database are the following:

1. 100% of sets with alpha also contain beta
2. 50% of sets with alpha, beta also have epsilon
3. 50% of sets with alpha, beta also have theta

**Market Basket Analysis**

Market Basket Analysis is the study of customer transaction databases to determine dependencies between the various items they purchase at different times. **Association rule** **learning** is a rule-based machine learning method for discovering interesting relations between variables in large databases. It identifies frequent if-then associations called association rules which consists of an antecedent (if) and a consequent (then).

**Feature Analysis:**

* **Support** is an indication of how frequently the items appear in the data. Mathematically, support is the fraction of the total number of transactions in which the item set occurs. Mathematically,



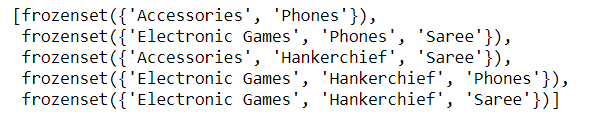
* **Confidence** indicates the number of times the if-then statements are found true. Confidence is the conditional probability of occurrence of consequent given the antecedent. Mathematically,



* **Lift** can be used to compare confidence with expected confidence. This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is. Mathematically,



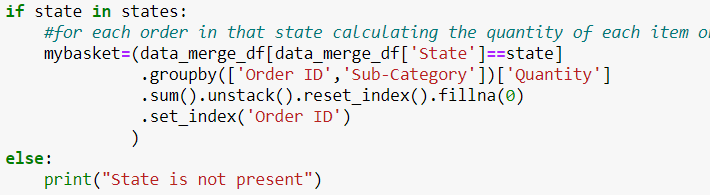
**Screenshot of the result**

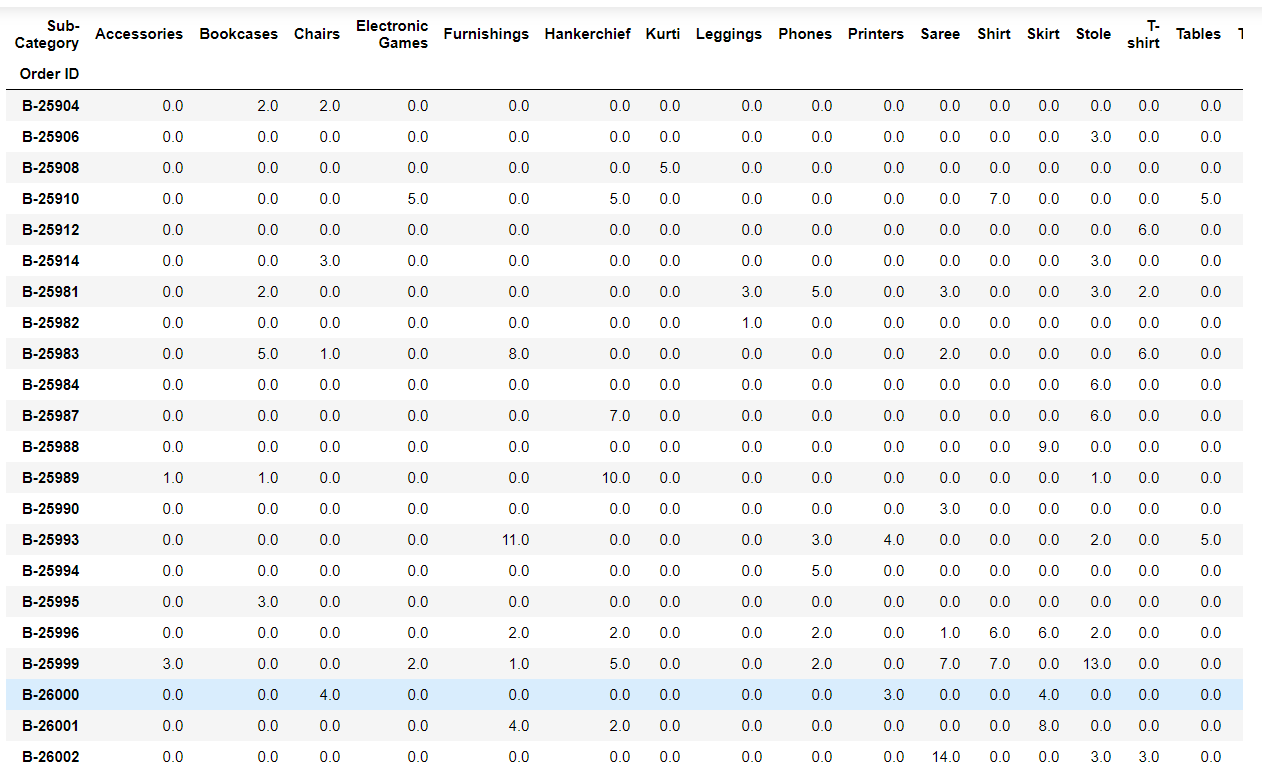


**SCREENSHOT OF RESULT**

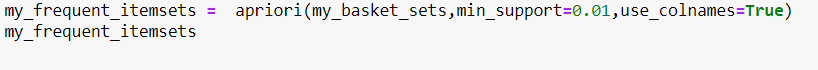
**Result Explanation:**

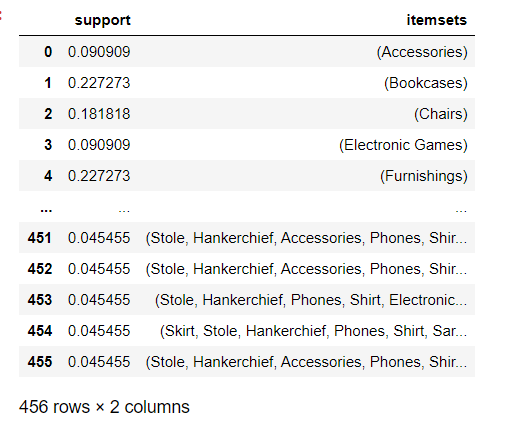
A basket is created for each state for analysis and calculating the quantity of each product ordered in that state.



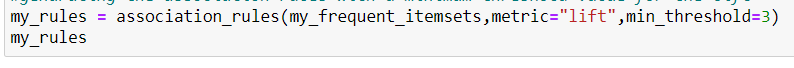


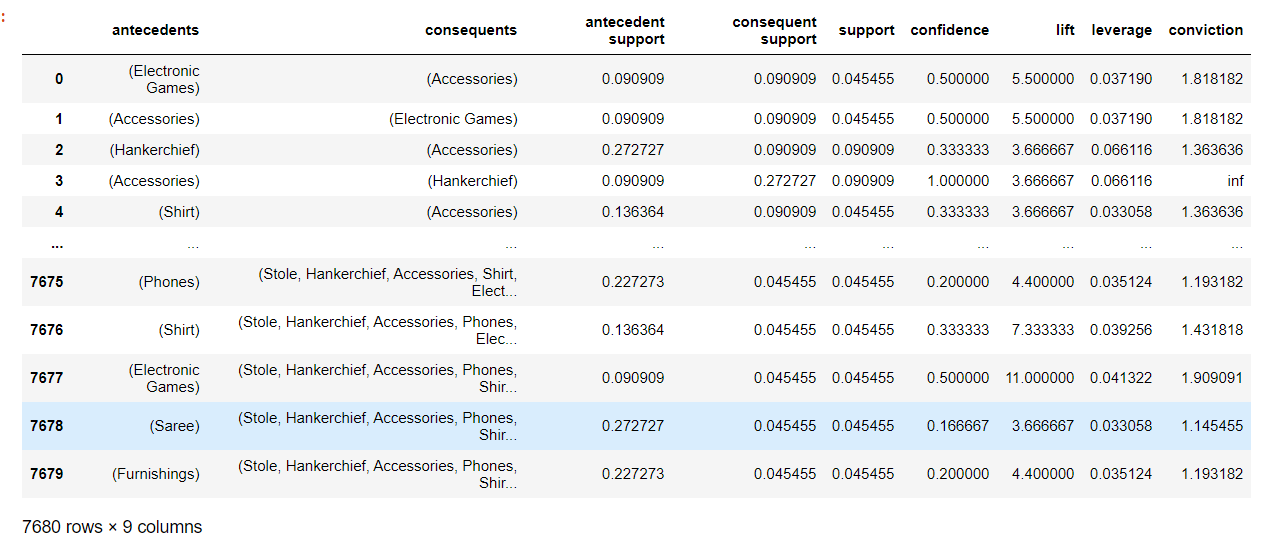
Apriori algorithm is used to generate the frequent item sets with minimum support value of 0.1



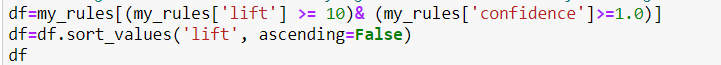


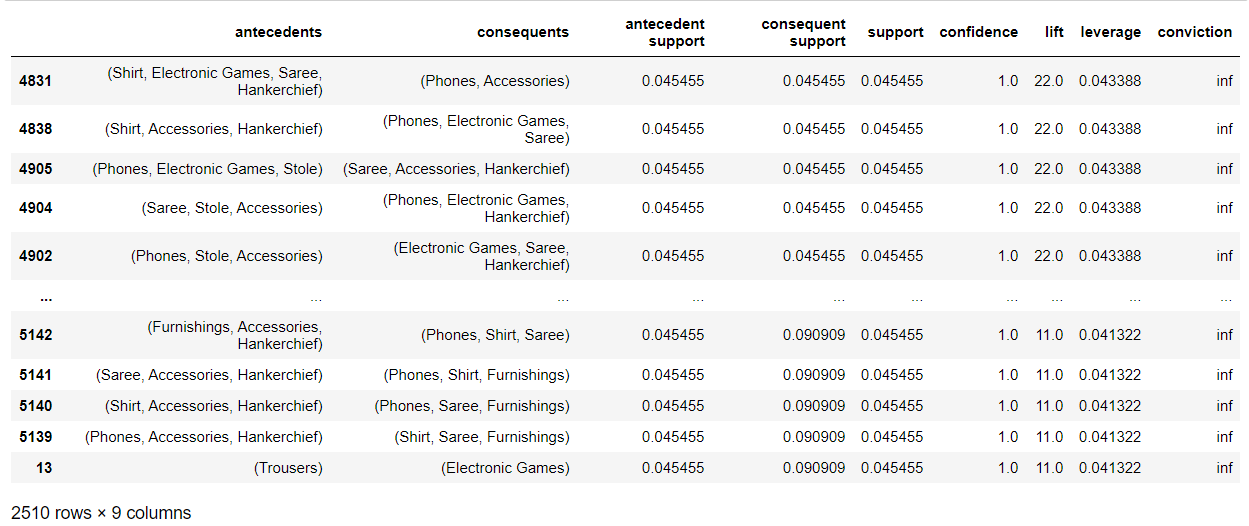
Generating the association rules with minimum threshold value for the lift



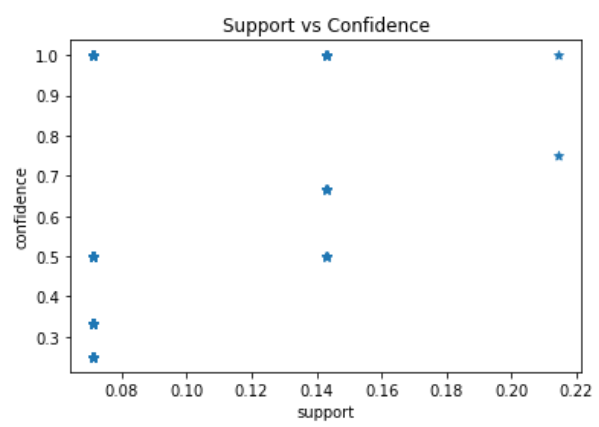


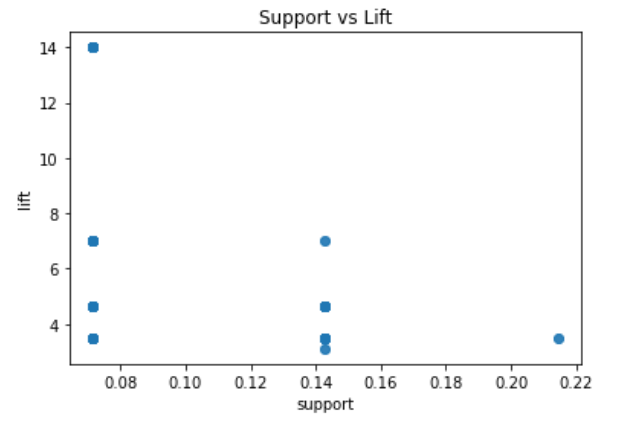
Choosing all the products that have lift greater than 3 and confidence greater then 0.7





**Graphs of support vs confidence and support vs lift**





**CONCLUSION**

The proposed system uses Apriori algorithm for finding item sets frequently bought together considering customer city and state. The customer spends more time due to better recommendation. This increases the profit for the organization and also increases the number of customers. Future work can involve incorporating many fields to make the recommendation more efficient and also integrating user feedbacks to provide recommendations.

**REFERENCES**

* StackOverflow<https://stackoverflow.com/>
* YouTube <https://www.youtube.com/>
* GeeksForGeeks <https://www.geeksforgeeks.org/>
* javatpoint<https://www.javatpoint.com/apriori-algorithm-in-machine-learning>
* Mediumhttps://towardsdatascience.com/underrated-machine-learning-algorithms-apriori-1b1d7a8b7bc

**Dataset**

* Kaggle <https://www.kaggle.com/benroshan/ecommerce-data>