# Association Rules

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: Prajay B. Urkude Batch ID: 16092021**

**Topic: Association Rules**

# 

Hints:

1. Business Problem
   1. **What is the business objective?**
   2. **Are there any constraints?**
2. Work on each feature of the dataset to create a data dictionary as displayed in the below image**:**



1. Data Pre-processing
   1. Data Cleaning, Feature Engineering, etc.
2. Model Building

4.1 Application of Apriori Algorithm

* 1. Build most frequent item sets and plot the rules
  2. Work on Codes

5.Deployment

5.1 Deploy solutions

6. Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?

**Problem Statement: -**

Kitabi Duniya, a famous book store in India, which was established before Independence, the growth of the company was incremental year by year, but due to online selling of books and wide spread Internet access its annual growth started to collapse, seeing sharp downfalls, you as a Data Scientist help this heritage book store gain its popularity back and increase footfall of customers and provide ways the business can improve exponentially, apply Association Rule Algorithm, explain the rules, and visualize the graphs for clear understanding of solution.

**1.) Books.csv**

# To capture the different set of rule values for Books Dataset using apriori algorithm.

# Also Observe the change in number of rules for different support,confidence values

Ans: **Business Objectives:**

* store gain its popularity back and increase footfall of customers and provide ways the business can improve exponentially.

**Business Constraints:**

* To tackle the e-commerce business of books

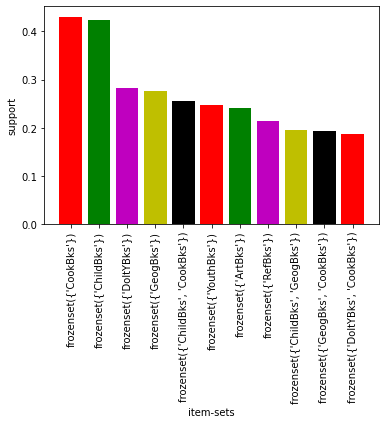
|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Features** | **Description** | **Type** | **Relevance** |
| ChildBks | Child books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| YouthBks | youth books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| CookBks | Cook books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| DoItYBks | Do it yourself books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| RefBks | Reference books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
|  |  |  |  |
| ArtBks | Art books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| GeogBks | Geography books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| ItalCook | Italian cook books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |  | Relevat, It gives the useful information |
| ItalAtlas | Italian Atlas books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| ItalArt | Italian art books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| Florence | Florence books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |

1 Import the libraries such as panda for data manipulation, matplotlib for data visualization and from mlxtend.frequent\_patterns package import apriori function and association\_rules function

2. Load the data

3. Data preprocessing : Checking the null values present or not, checking the 1st moment, 2nd moment business decisions.

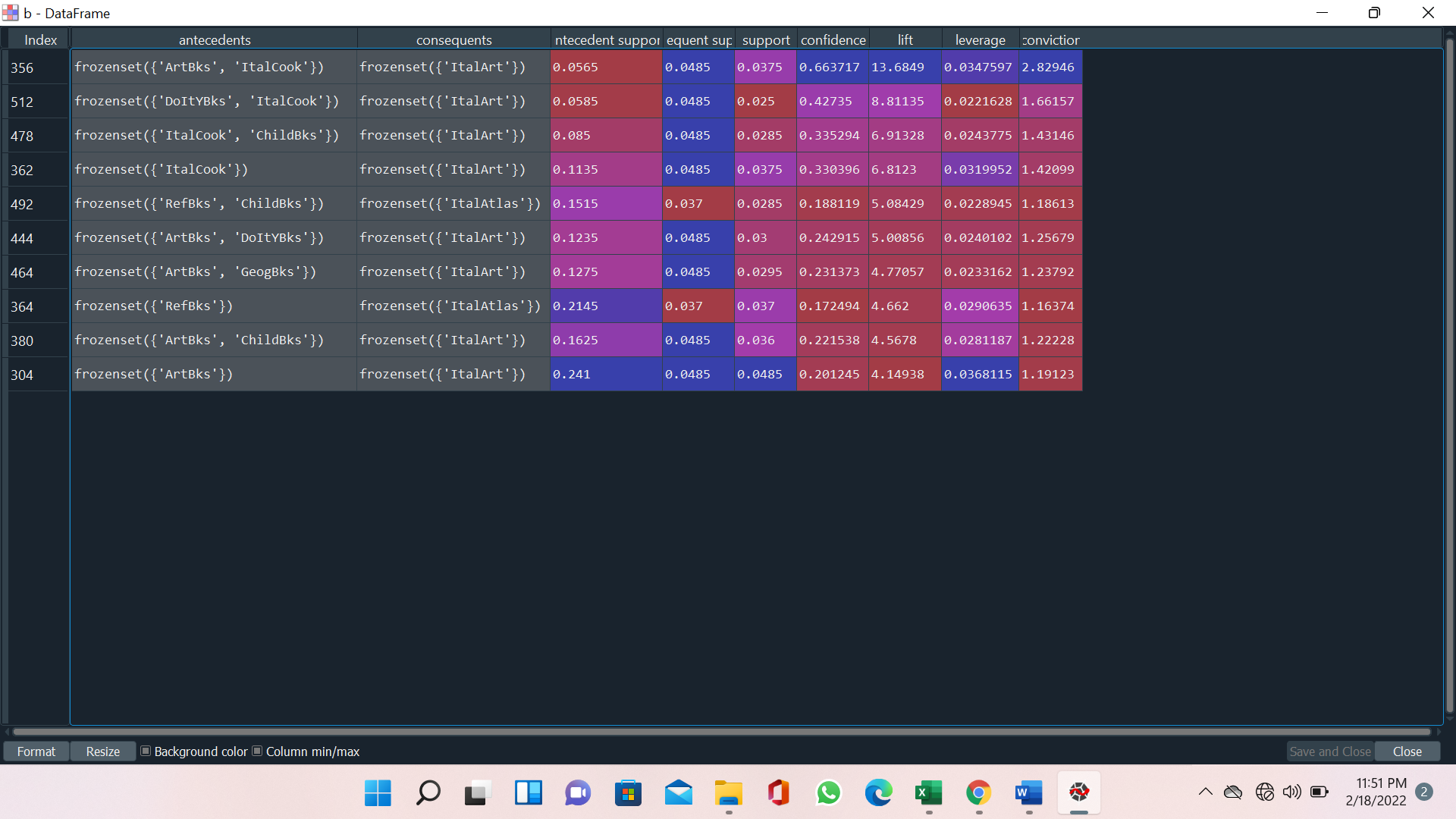
4. Application of the Apriori algorithm to find most frequent items by taking minimum support is 50 and maximum length is 3 and sort them in descending order by the support

5. plot the Bar plot by taking the itemsets on X axis and support on Y axis .

From the graph, Cooks books and child books has the highest supports and this are the best selling books from the store. And we should check the relation of this best selling books with another books .

6. Making the rules based on the lift calculation by using the association rules function.

7. Apply the profusion rule to remove the duplicate rules where the same product repeats.



Abovetable shows the top ten association rules for the books purchased in the book store.

From this, Antecedent: Artbooks , ItalCook books and Consequent: ItalArt, so people purchase the ItalArt book mostly when they purchase Artbks and ItalCook books and like this we can make the decisions for another books also by seeing the above table

So to grow the business we should use the this association rules, means books which are dependent on the another books we should keep them in the Same shelves or nearby shelves by doing the proper arrangement in the selves according to the demand and dependency.

The products which have high support, apply the better marketing strategies for selling this books with better offer.

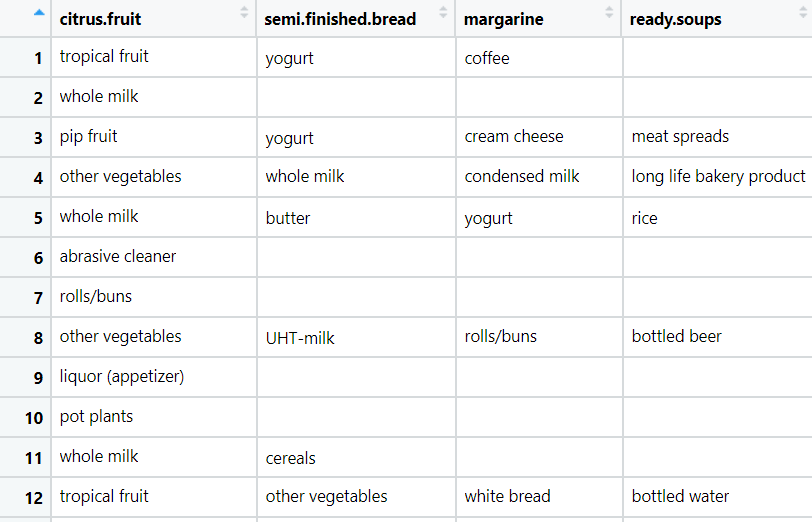
Add the Book sections which includes used books.

**Problem Statement: -**

The Departmental Store, has gathered the data of the products it sells on a Daily basis.

Using Association Rules concepts, provide the insights on the rules and the plots.

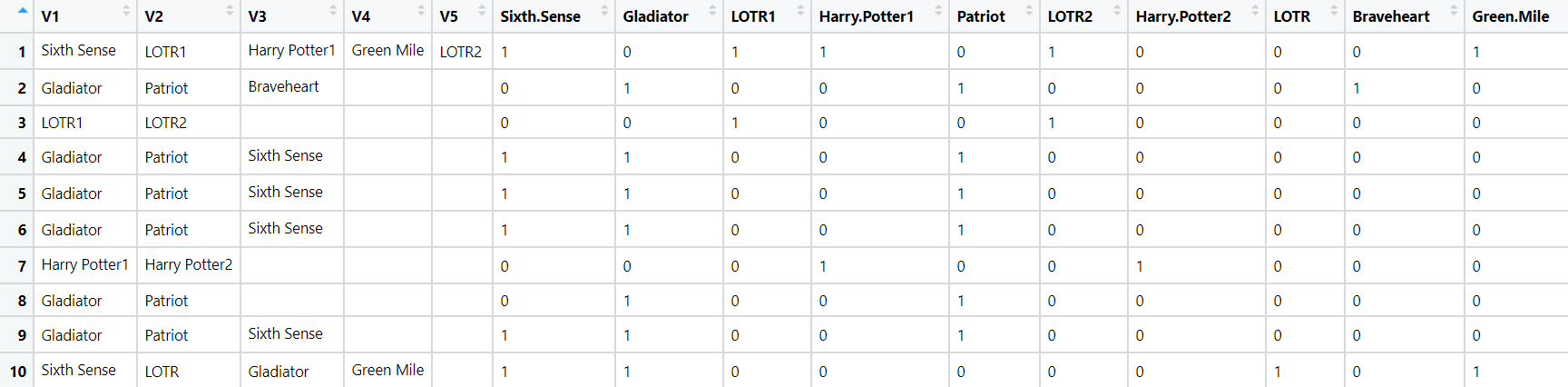
**2.) Groceries.csv**



**Problem Statement: -**

A film distribution company wants to target audience based on their likes and dislikes, you as a Chief Data Scientist Analyze the data and come up with different rules of movie list so that the business objective is achieved.

**3.) my\_movies.csv**



**Problem Statement: -**

A Mobile Phone manufacturing company wants to launch its three brand new phone into the market, but before going with its traditional marketing approach this time it want to analyze the data of its previous model sales in different regions and you have been hired as an Data Scientist to help them out, use the Association rules concept and provide your insights to the company’s marketing team to improve its sales.

**4.) myphonedata.csv**



**Problem Statement: -**

A retail store in India, has its transaction data, and it would like to know the buying pattern of the

consumers in its locality, you have been assigned this task to provide the manager with rules

on how the placement of products needs to be there in shelves so that it can improve the buying

patterns of consumes and increase customer footfall.

**5.) transaction\_retail.csv**

