# 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) 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.

**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

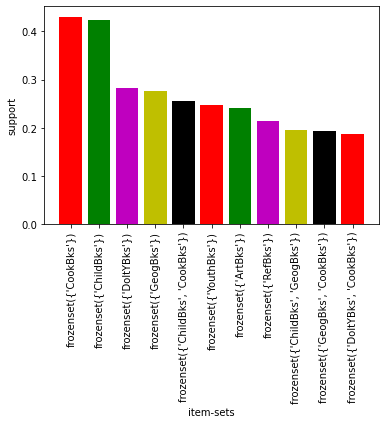
|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Features** | **Description** | **Type** | **Relevance** |
| ChildBks | Child books purchased or not | Qualitative, Nominal | Relevant, It gives the useful information |
| YouthBks | youth books purchased or not | Qualitative, Nominal | Relevant, It 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.

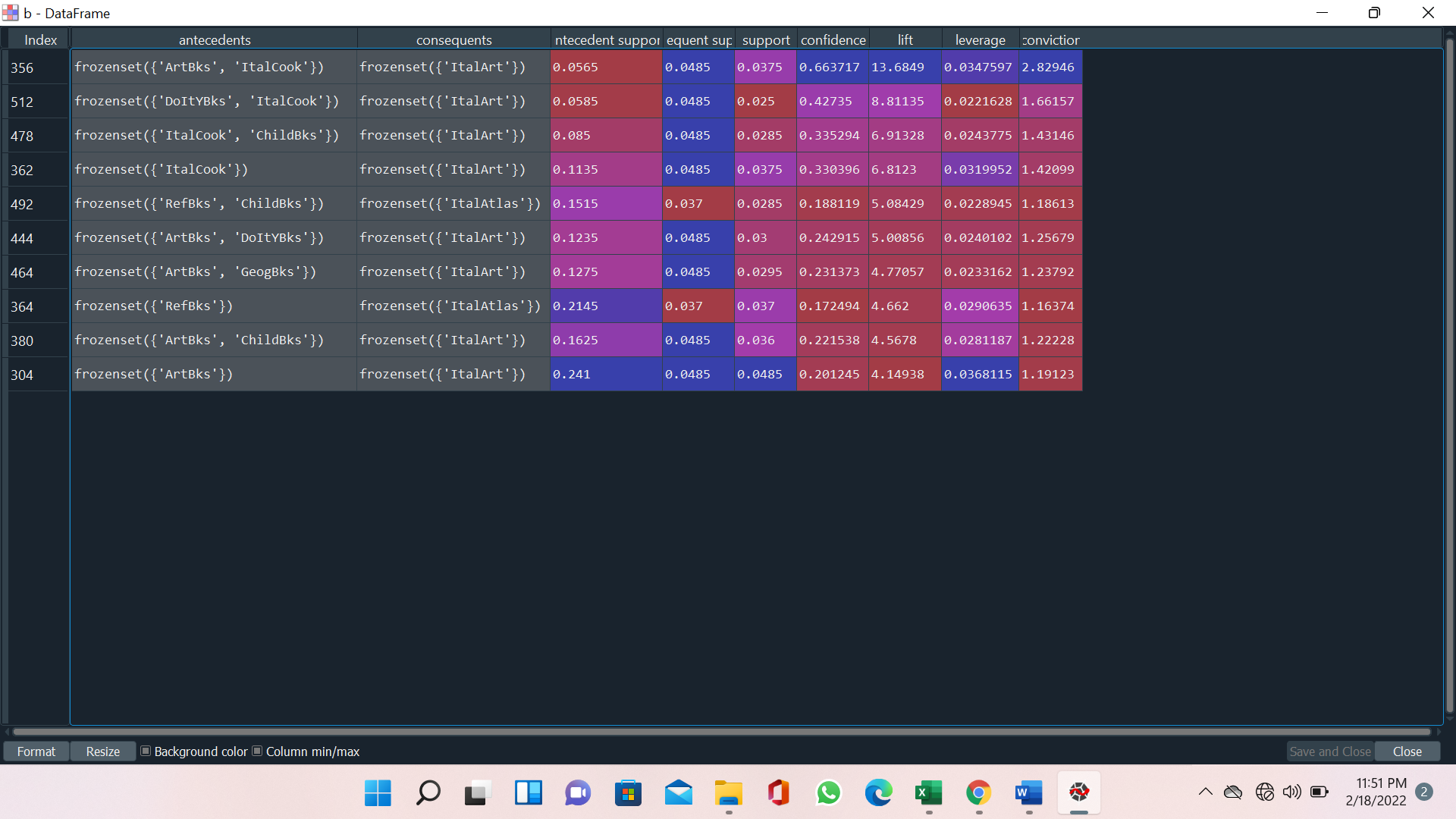
4. Application of the Apriori algorithm to find most frequent items by taking minimum support is 50 and maximum length is 3 and sort them in descending order by the support

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.



Lift >1 tells us 2 items occur in transactions together more often than we would expect based on their individual support values. This means the relationship is unlikely to be explained by random chance. This natural threshold is convenient for filtering purposes.

Lift <1 tells us 2 items are paired together less frequently in transactions than we would expect if the pairings occurred by random chance

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

Rule 1 indicates that if Art books , ItalCook book are purchased, then with 67% confidence Italart book will also be purchased.

In other words 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.

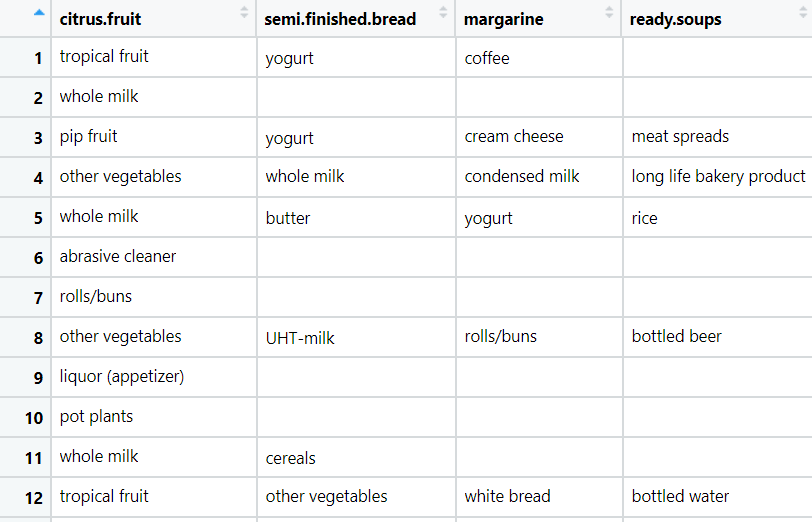
Add the Book sections which includes used books.

**2 ) 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.

**Groceries.csv**



**Ans:-** : **Business Objectives:**

* To produce the meaningful rules for the products for the better business.

**Business Constraints:**

* Application of the AI algorithms for the smart data solutions.

Steps to produce the rules for groceries.csv datasets :-

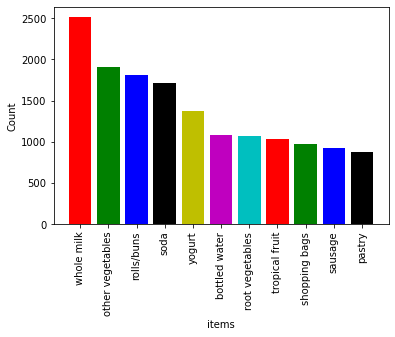
1. Import the libraries such as panda for data manipulation, matplotlib for data visualization and from mlxtend.frequent\_patterns package import apriori function and association\_rules function.

2. loading the data, as the file is transaction data, we will be reading data directly by creating empty list and read the data in this empty list.

3. Splitting the data into separate transactions using separator as "\n".

4. create different list for frequency of items and and items by doing the iterations and by splitting the data and sort it I decending order.

4. Draw the bar plot for 1st 10 most frequent items by plotting items on X-axis and counts of items on Y-axis

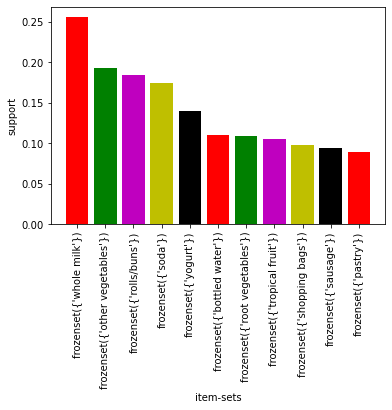


From the graph, Item whole milk has purchased maximum no. of times followed by other vegetables, rolls buns, soda, bottled water etc. and we have make the rules for the above items from which we can find which items is purchased together most of the times.

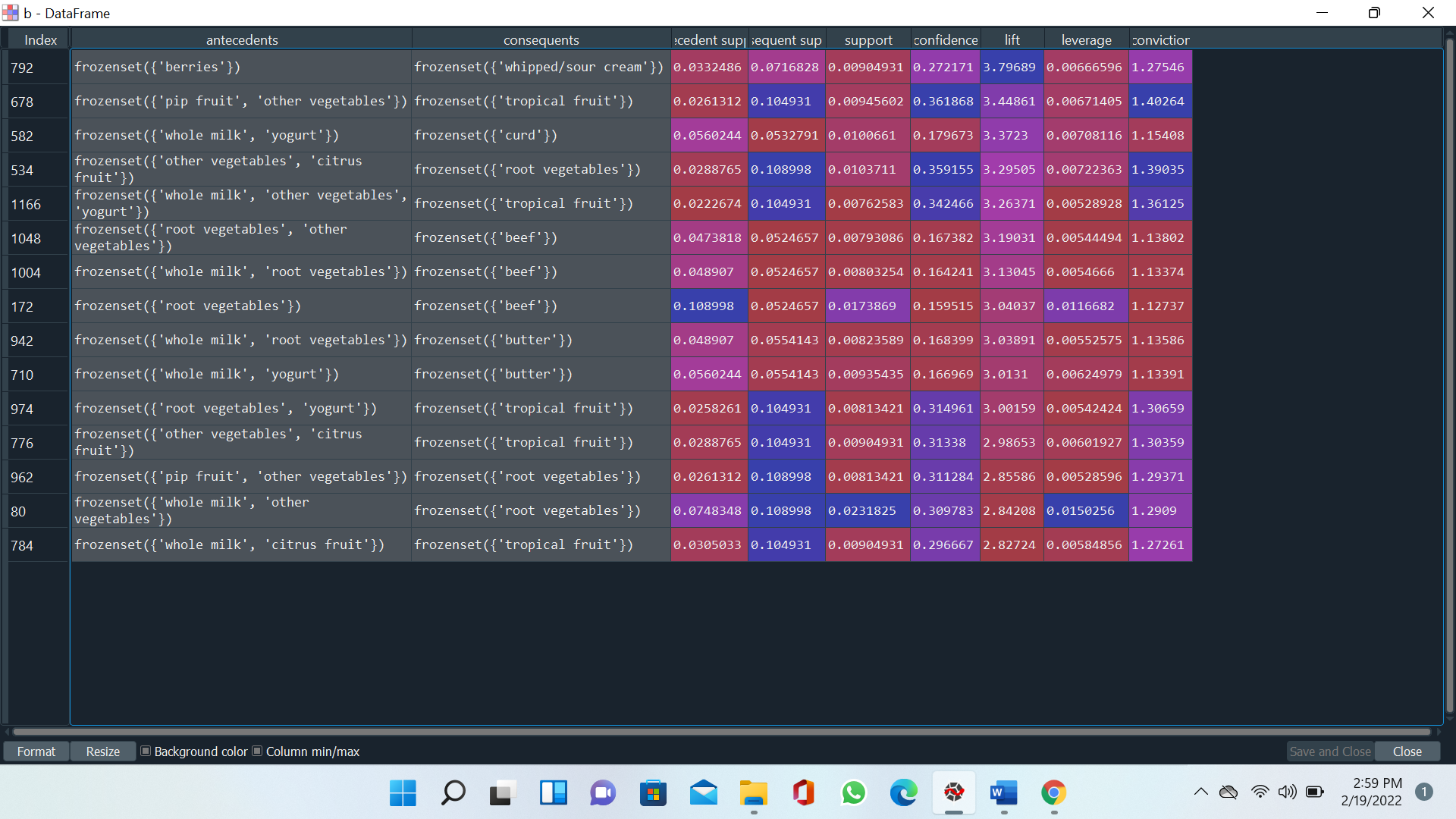
5. Convert the list of items into data frame and convert the nonnumeric data into numeric data

By get\_dummies() functions or one hot encoding.

6. calculate the most frequent item sets based on the support by taking minimum support is 75 and maximum length is 4 and draw the bar plot between item sets and the support for it for first 10 more frequent items sets .



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

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

Lift > 1 tells us 2 items occur in transactions together more often than we would expect based on their individual support values. This means the relationship is unlikely to be explained by random chance. This natural threshold is convenient for filtering purposes.

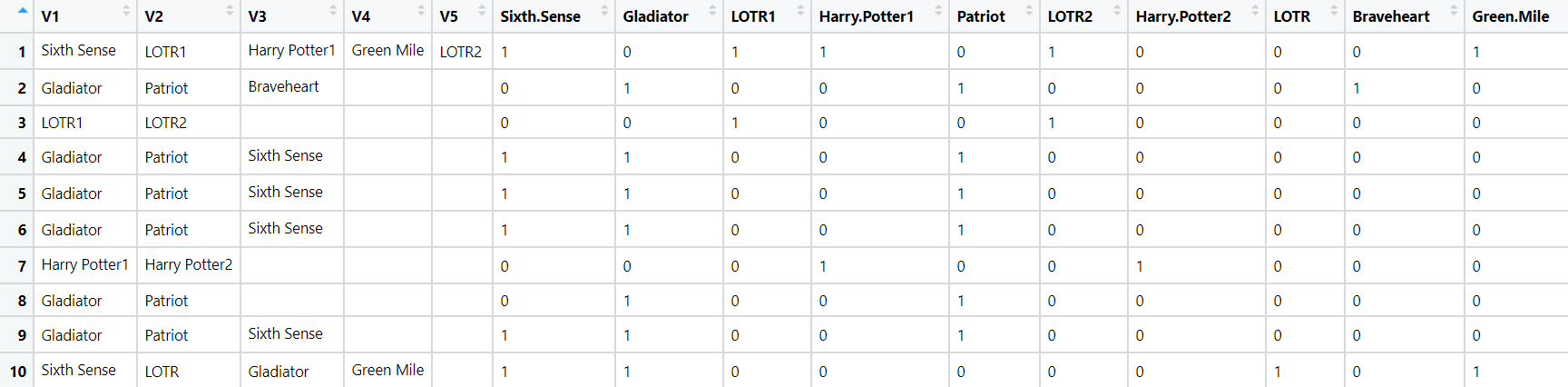
Lift <1 tells us 2 items are paired together less frequently in transactions than we would expect if the pairings occurred by random chance.

So from the above table we can say that as the lift ratio for the items sets (berries, whipped/source cream) is very high then this two product purchasing combinly and this products are dependent on each other and we put such products in the one shelves or side by side shelves and same for the other products we should apply the rule

**3) 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.

**Dataset : my\_movies.csv**



**Ans :- Business Objectives:**

* To target the audience for watching the movies based on there likes and dislikes.

**Business Constraints:**

* Application of the AI algorithms for the smart data solutions.

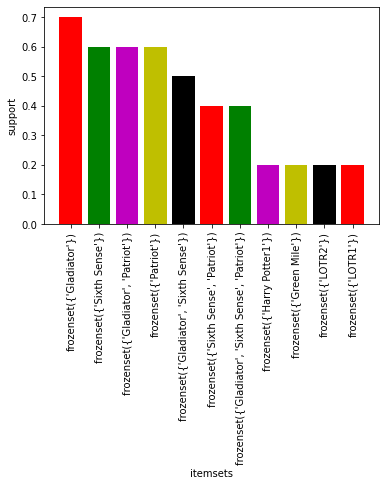
1 Import the libraries such as panda for data manipulation, matplotlib for data visualization and from mlxtend.frequent\_patterns package import apriori function and association\_rules function

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

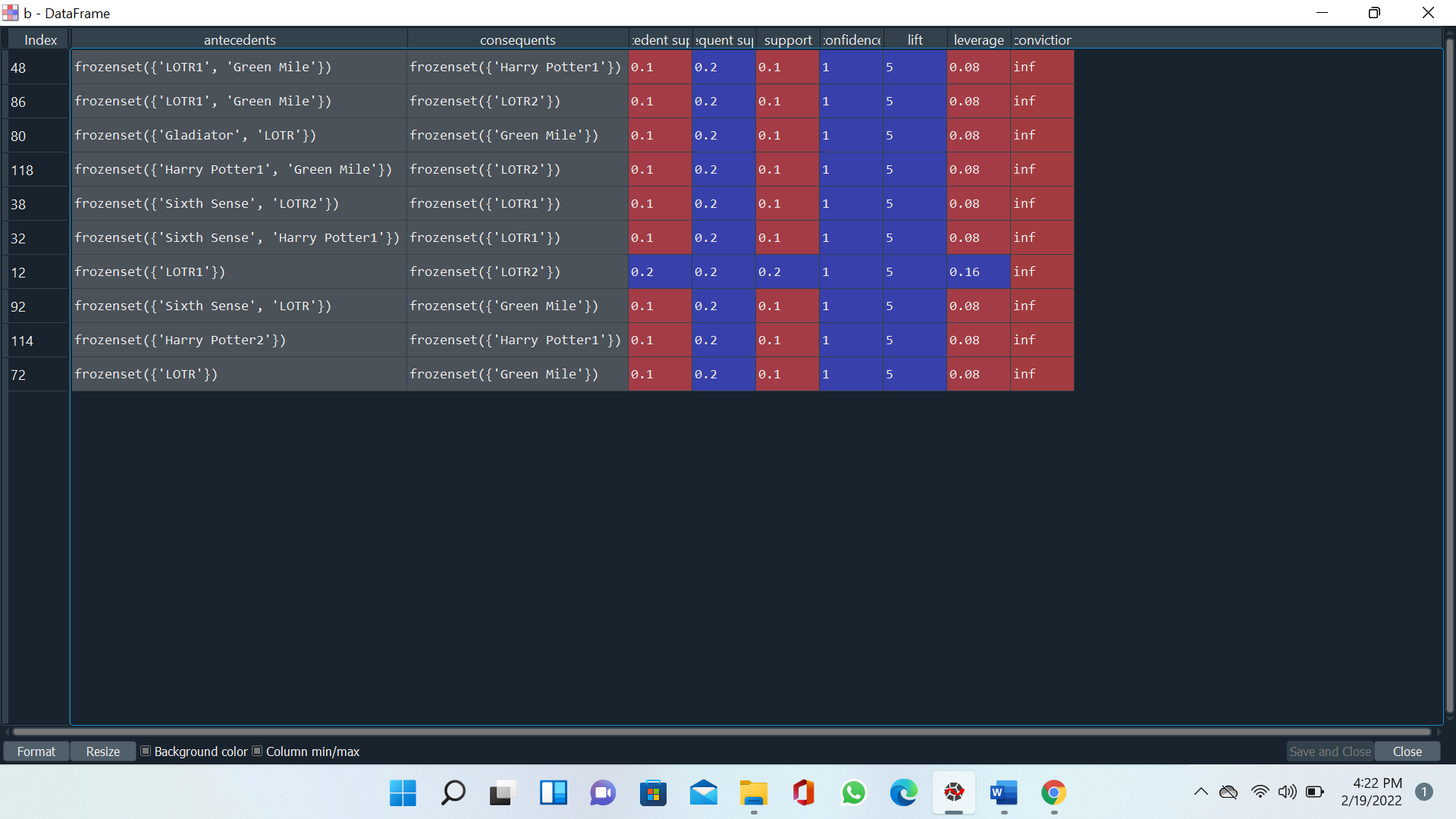
5. plot the Bar plot by taking the item sets on X axis and support on Y axis .



From the graph, Gladiator, sixth sense, patriot has the highest supports and this are the most watchable movies. We should check the relation of this movies with another movies .

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.



From the above table movies LOTR1, Green Mile, Harry Potter1 has high lift ratio and this movies are dependent on each other so we can say that people are preferred in watching movies like LOTR1, Green Mile and then Harry Potter one after other which is same for the other results also.

**4) 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 wants to analyze the data of its previous model sales in different regions and you have been hired as a 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.

**myphonedata.csv**



**Ans:-**

**Business Objectives:**

* To determine the demand of the variants of the phones based on the color and increase the sells of the mobile.

**Business Constraints**

* Lack of data coverage for all the customers.

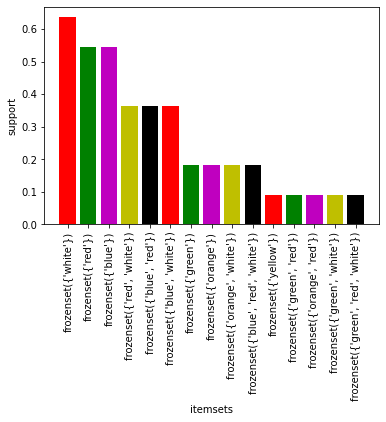
1 Import the libraries such as panda for data manipulation, matplotlib for data visualization and from mlxtend.frequent\_patterns package import apriori function and association\_rules function

2. Load the data

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

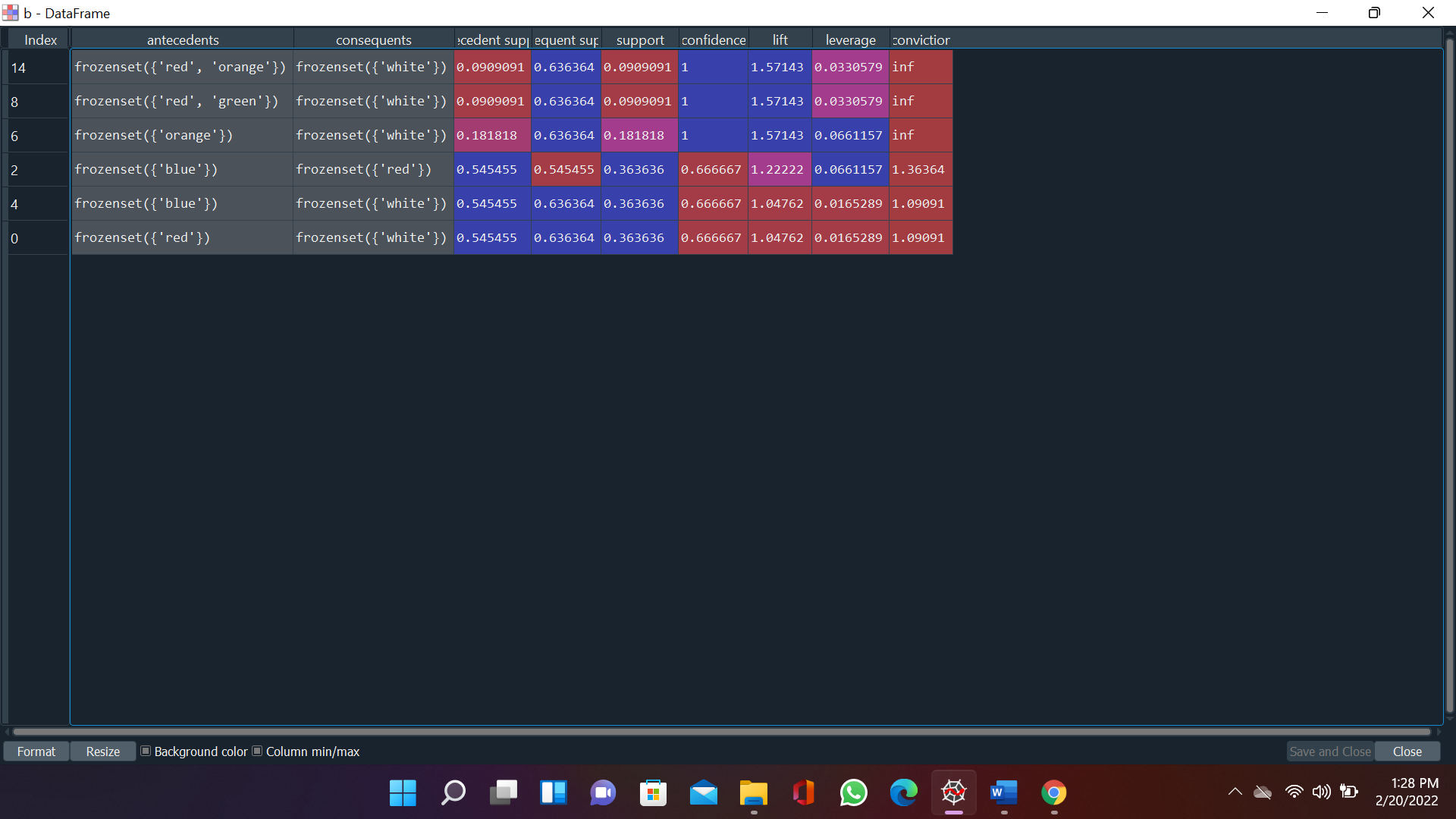
4. Application of the Apriori algorithm to find most frequent items by taking minimum support is 50 and maximum length is 3 and sort them in descending order by the support

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



From the graph, white, red, blue has the highest supports and this are the most purchased mobiles. We should check the relation of this mobiles with other mobile.

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

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

From the table, red color, orange color mobiles purchased then with 100% confidence white color mobiles will also be purchased.

The support for red and orange color is around 9% and for white is 63% indicates that 9% rule has the support of 9% transaction for red and orange color and 63% transaction for only for white color.

As the lift ratio > 1 tells us 2 items occur in transactions together more often than we would expect based on their individual support values. This means the relationship is unlikely to be explained by random chance. This natural threshold is convenient for filtering purposes

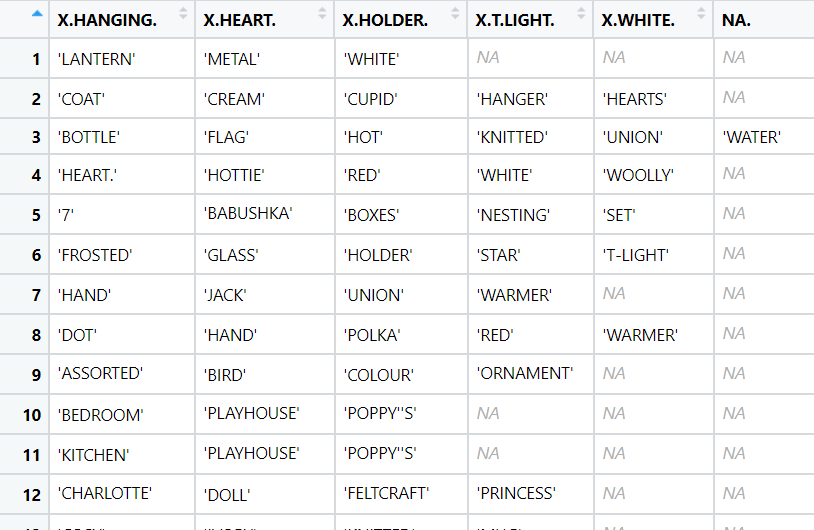
**5.) Problem Statement: -**

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.

**transaction\_retail.csv**

**Ans: Business Objectives:**

* To produce the meaningful rules for the product placements improve the buying patterns of consumes and increase customer

footfall.

**Business Constraints:**

* Application of the AI algorithms for the smart data solutions.

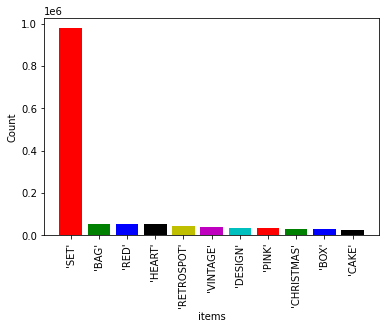
Steps to produce the rules for groceries.csv datasets :-

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. loading the data, as the file is transaction data, we will be reading data directly by creating empty list and read the data in this empty list.

3. Splitting the data into separate transactions using separator as "\n".

4. create different list for frequency of items and and items by doing the iterations and by splitting the data and sort it I decending order.

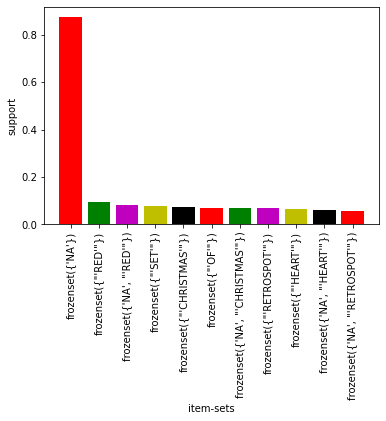
4. Draw the bar plot for 1st 10 most frequent items by plotting items on X-axis and counts of items on Y-axis

From the graph, Item SET has purchased maximum no. of times followed by bag, etc. and we have make the rules for the above items from which we can find which items is purchased together most of the times.

5. Convert the list of items into data frame and convert the nonnumeric data into numeric data

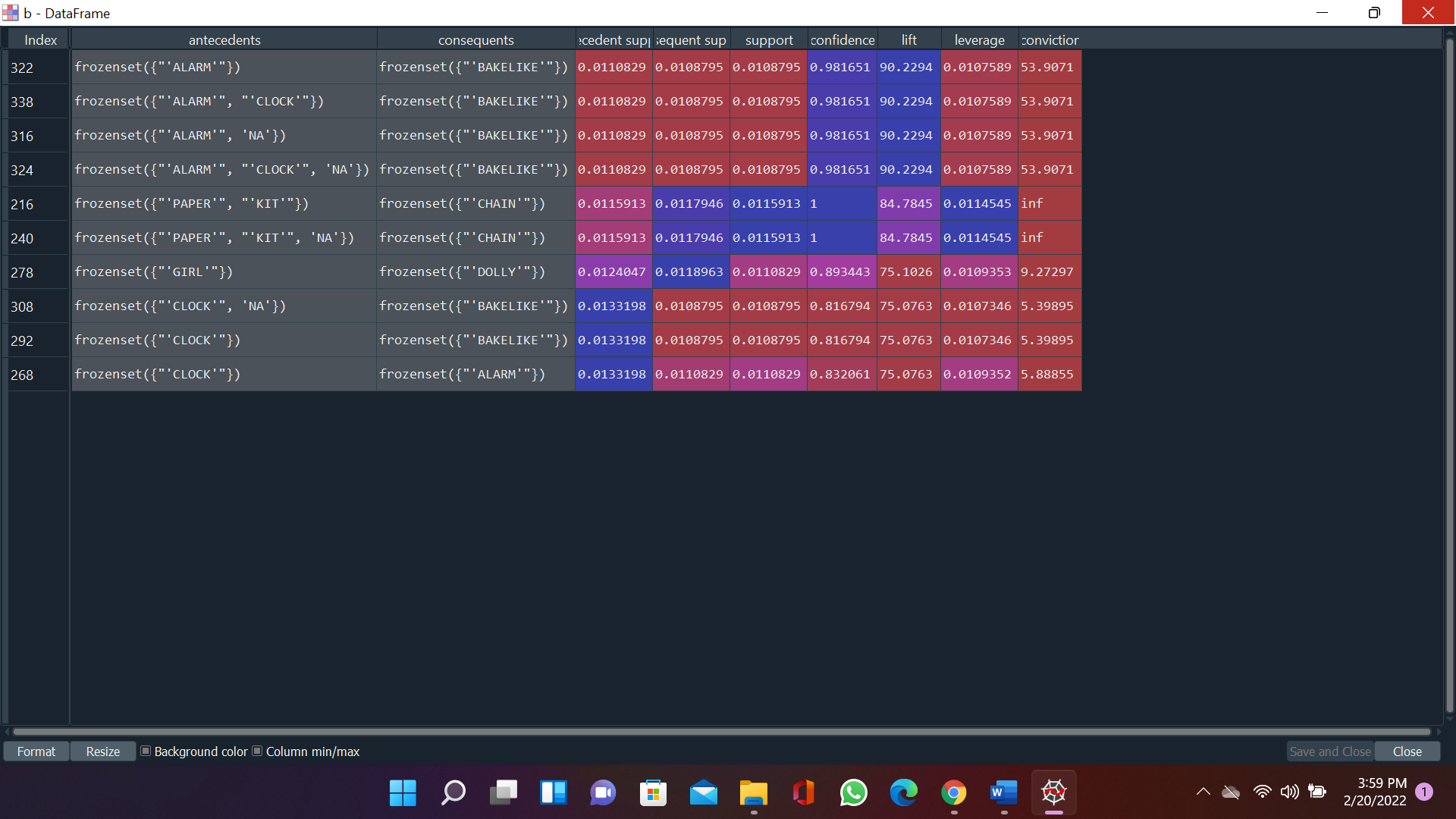
By get\_dummies() functions or one hot encoding.

6. calculate the most frequent item sets based on the support by taking minimum support is 75 and maximum length is 4 and draw the bar plot between item sets and the support for it for first 10 more frequent items sets .



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

8. Apply the profusion rule to remove the duplicate



From the above table we can say that Alarm watch when purchased then Bakelike also purchased with 98% of confidence but as lift > 1 items occur in transactions together more often than we would expect based on their individual support values. Same for the other rules.