# Topic: Association Rules

**Instructions**

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name: Shubham Sharma Batch Id:**  DS WDMCO 050121 Blr **Topic: - Association Rules.**

# Hints:

1. Business Problem
   1. Objective
   2. Constraints (if any)
2.  Work on each feature of the dataset to create a data dictionary as displayed in the below image**:**



**Using R and Python codes perform:**

3.Data Pre-processing

* 1. Data Cleaning, Feature Engineering, etc.

4.Model Building

4.1 Application of Apriori Algorithm.

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

5.Deployment

* 1. Deploy solutions using R shiny and Python Flask.

6. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

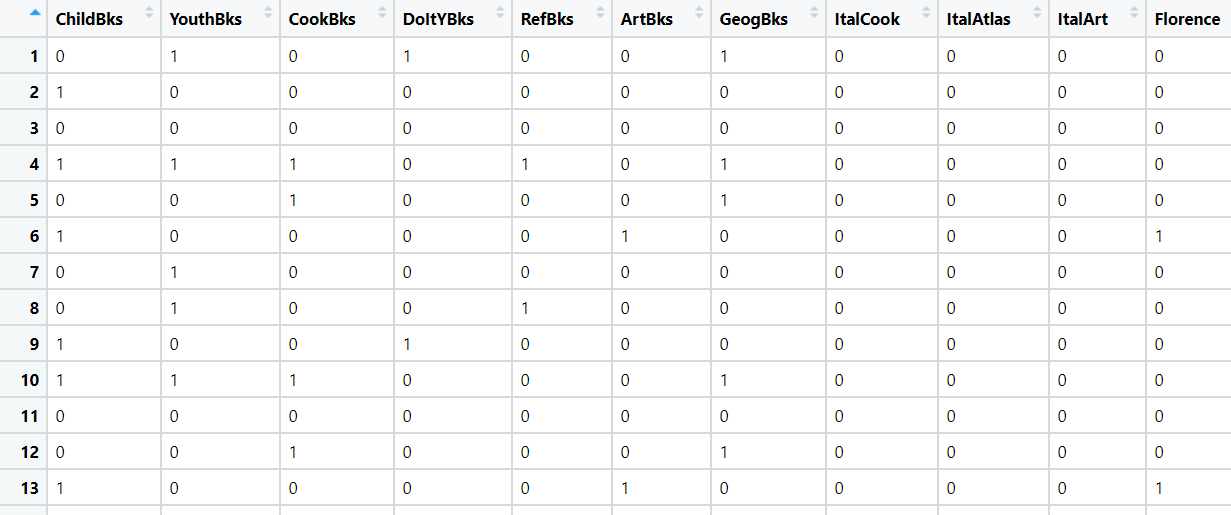
**Note:**

1. For each assignment, the solution should be submitted in the above format
2. Research and Perform all possible steps for improving the rules and also check if you can take out sub rules from main rules.
3. All the codes (executable programs) are running without errors
4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here that is commenting is necessary in the codes.
5. Please send all files at once whilst submitting assignments.

**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**



Business Problem

Objective :- maximize the sale

Constraints :- maximize footfall of customers

Python code steps :\_

Ans = Here data is in CSV format in “Binary Matrix Format” . Data Frame name is book. It has 2000 rows and 11 columns.

(**Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

Then we create a data frame that’s contain details of each columns ,like- description ,data types ,and save the details named as data\_details .all of them are important and all are categorical data type (Binary).

**.Data Pre-processing**

Dataframe has no missing values so. We also check for unique values for each columns named as “col\_uni” .

**Exploratory Data Analysis (EDA):**

. now we do exploratory data analysis for whole data frame and saved the details as “EDA”.

For graphical representation we have done scatter and histogram using seaborn libarary .

**####Model Building**

**Application of Apriori Algorithm.**

**Build most frequent item sets and plot the rules**

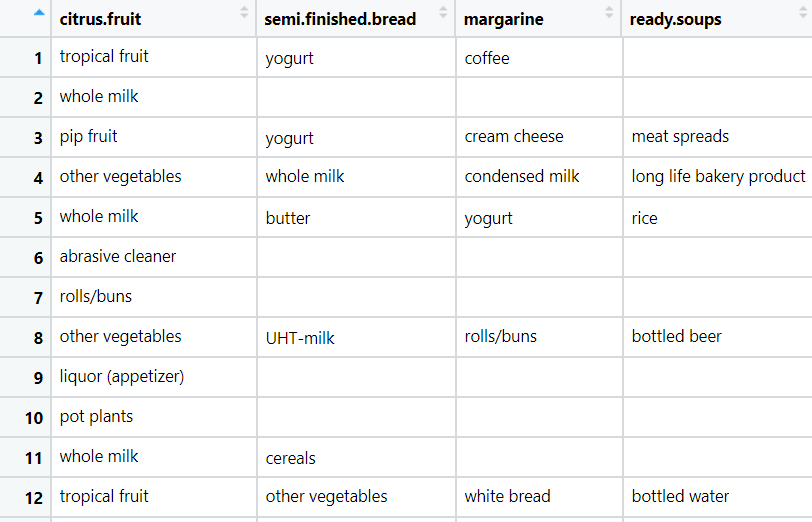
Now we are building our apriori algorithm on book data frame ,where min support should be grater then 0.01 and maximum length will be 4 .and after that we are selecting top 10 rules without any redundancy that contain maximum lift ratio and save them “top\_10\_rules”.

**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**

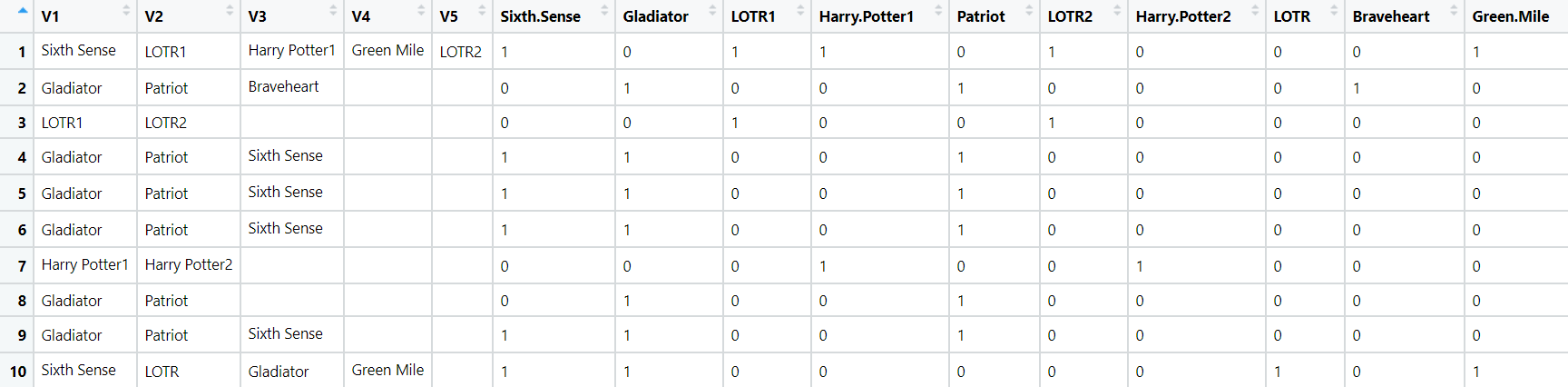


**This one is already explain in class .**

**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**



Business Problem

Objective :- maximize **audience**

Python code steps :\_

Ans = Here data is in CSV format in “List Format “ and the same data in “Binary Matrix Format” . Data Frame name is movie. It has 10 rows and 15 columns. And then I create a new dataframe only with Binary Matrix Format named as movie\_1 .

(**Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

Then we create a data frame that’s contain details of each columns ,like- description ,data types ,and save the details named as data\_details .all of them are important and all are categorical data type (Binary).

**.Data Pre-processing**

Dataframe has no missing values so. We also check for unique values for each columns named as “col\_uni” .

**Exploratory Data Analysis (EDA):**

. now we do exploratory data analysis for whole data frame and saved the details as “EDA”.

For graphical representation we have done scatter and histogram using seaborn libarary .

**####Model Building**

**Application of Apriori Algorithm.**

**Build most frequent item sets and plot the rules**

Now we are building our apriori algorithm on book data frame ,where min support should be grater then 0.02 and maximum length will be 3 .and after that we are selecting top 5 rules without any redundancy that contain maximum lift ratio and save them “top\_5\_rules”.

**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**



Business Problem

Objective :- maximize **sales**

Python code steps :\_

Ans = Here data is in CSV format in “List Format “ and the same data in “Binary Matrix Format” . Data Frame name is “mobail”. It has 11 rows and 9 columns. And then I create a new dataframe only with Binary Matrix Format named as mobail\_1 .

(**Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

Then we create a data frame that’s contain details of each columns ,like- description ,data types ,and save the details named as data\_details .all of them are important and all are categorical data type (Binary).

**.Data Pre-processing**

Dataframe has no missing values so. We also check for unique values for each columns named as “col\_uni” .

**Exploratory Data Analysis (EDA):**

. now we do exploratory data analysis for whole data frame and saved the details as “EDA”.

For graphical representation we have done scatter and histogram using seaborn libarary .

**####Model Building**

**Application of Apriori Algorithm.**

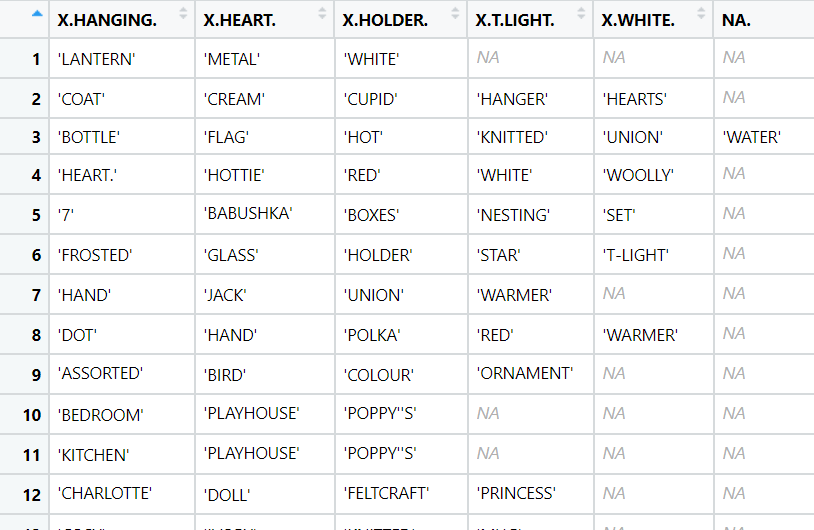
**Build most frequent item sets and plot the rules**

Now we are building our apriori algorithm on book data frame ,where min support should be grater then 0.009 and maximum length will be 2 .and after that we are selecting top 3 rules without any redundancy that contain maximum lift ratio and save them “top\_3\_rules”.

**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**



Business Problem

Objective :- maximize **customer footfall**

Python code steps :\_

Ans = Here data is in CSV format in “List Format”. Data Frame name is “mobail”. It has 557041 rows and 6 columns.

(**Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

Then we create a data frame that’s contain details of each columns ,like- description ,data types ,and save the details named as data\_details .all of them are important and all are categorical data type (Binary).

**.Data Pre-processing**

Dataframe has no missing values so. We also check for unique values for each columns named as “col\_uni” .

**Exploratory Data Analysis (EDA):**

. now we do exploratory data analysis for whole data frame and saved the details as “EDA”.

For graphical representation we have done scatter and histogram using seaborn libarary .

**####Model Building**

**Application of Apriori Algorithm.**

**Build most frequent item sets and plot the rules**

Now we are building our apriori algorithm on book data frame ,where min support should be grater then 0.0075 and maximum length will be 4 .and after that we are selecting top 3 rules without any redundancy that contain maximum lift ratio and save them “top\_10\_rules”.