# 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:Shahina Athar**

**Batch Id:**  10122020

**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.

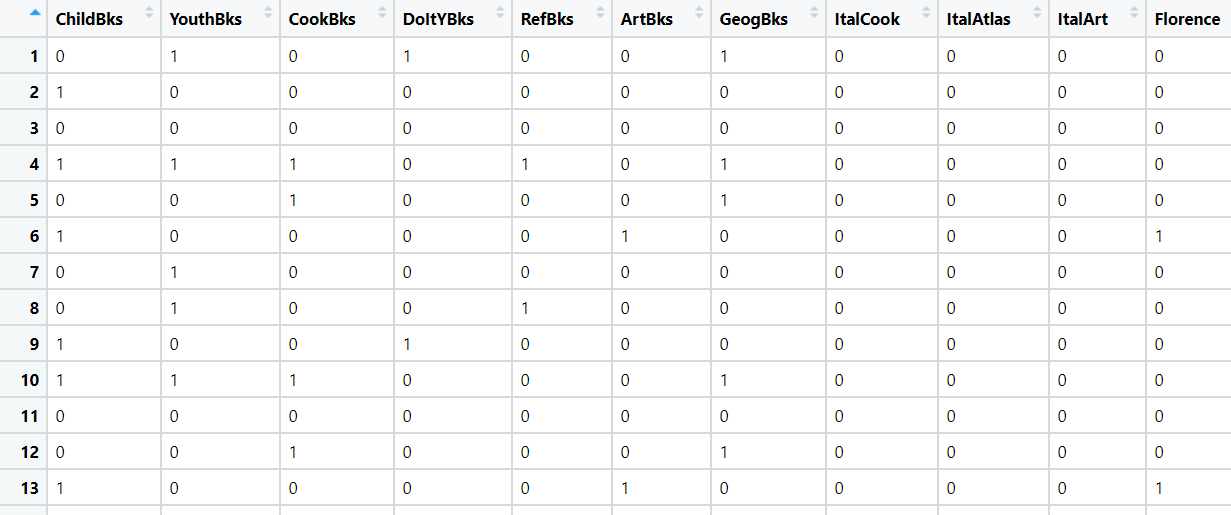
**Note:**

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**



**1) Business Problem:**

**Objective:** 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, we 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.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Types** | **Relevance** |
| ChildBks | Children Books | Quantitative, Discrete, Binary | Relevant |
| YouthBks | Youth Books | Quantitative, Discrete, Binary | Relevant |
| CookBks | Cook Books | Quantitative, Discrete, Binary | Relevant |
| DoItYBks | DoltY Books | Quantitative, Discrete, Binary | Relevant |
| RefBks | Ref Books | Quantitative, Discrete, Binary | Relevant |
| ArtBks | Art Books | Quantitative, Discrete, Binary | Relevant |
| GeogBks | Geography Books | Quantitative, Discrete, Binary | Relevant |
| ItalCook | Ital Cook Books | Quantitative, Discrete, Binary | Relevant |
| ItalAtlas | Ital Atlas Books | Quantitative, Discrete, Binary | Relevant |
| ItalArt | Ital Art Books | Quantitative, Discrete, Binary | Relevant |
| Florence | Florence Books | Quantitative, Discrete, Binary | Relevant |

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

**We generate top 10 Association Rules of books which is more purchased, because of this combinations we will again make a profit and their lift ratio is also more that 1 so that is the best rule we can say. See more in the excel sheet which I created and attach in the folder.**

|  |  |
| --- | --- |
| Antecedents | Consequents |
| {'ItalAtlas', 'GeogBks', 'ArtBks'} | {'ItalArt'} |
| {'ItalAtlas', 'ArtBks', 'DoItYBks'} | {'ItalArt'} |
| {'ChildBks', 'ItalAtlas', 'ArtBks'} | {'ItalArt'} |
| {'ItalAtlas', 'CookBks', 'ArtBks'} | {'ItalArt'} |
| {'ItalAtlas', 'ArtBks'} | {'ItalArt'} |
| {'ChildBks', 'ItalAtlas', 'GeogBks'} | {'ItalArt'} |
| {'ItalAtlas', 'GeogBks', 'CookBks'} | {'ItalArt'} |
| {'ItalAtlas', 'GeogBks'} | {'ItalArt'} |
| {'ItalAtlas', 'CookBks', 'DoItYBks'} | {'ItalArt'} |
| {'ChildBks', 'ItalAtlas', 'CookBks'} | {'ItalArt'} |

**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.**

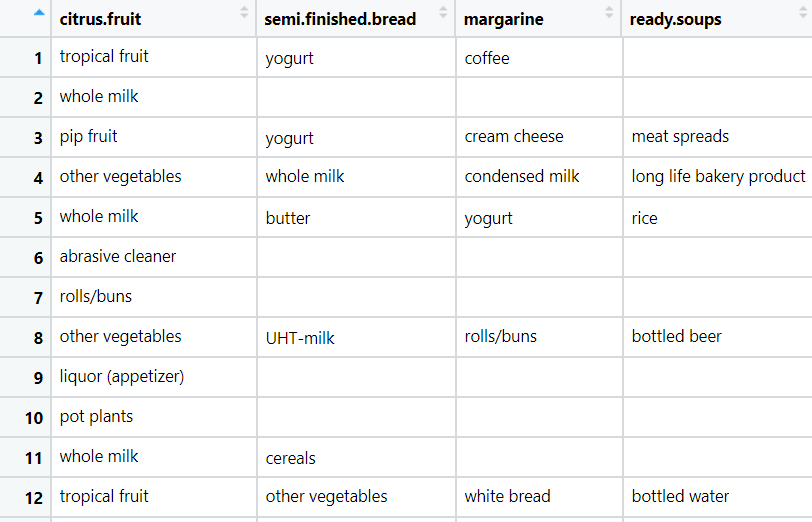
**1) Business Problem:**

**Objective:** The Departmental Store, has gathered the data of the products it sells on a Daily basis, we as a Data Scientist help The Departmental Store find the pattern of the products which is more sale either individually or together using Association Rules.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Types** | **Relevance** |
| citrus fruit | All Types of Citrus Fruit | Qualitative, Discrete, Nominal | Relevant |
| semi-finished bread | All Types of Semi-Finished Bread | Qualitative, Discrete, Nominal | Relevant |
| margarine | All Types of Margarine | Qualitative, Discrete, Nominal | Relevant |
| ready soups | All Types of ready soups | Qualitative, Discrete, Nominal | Relevant |

**2.) Groceries.csv**



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

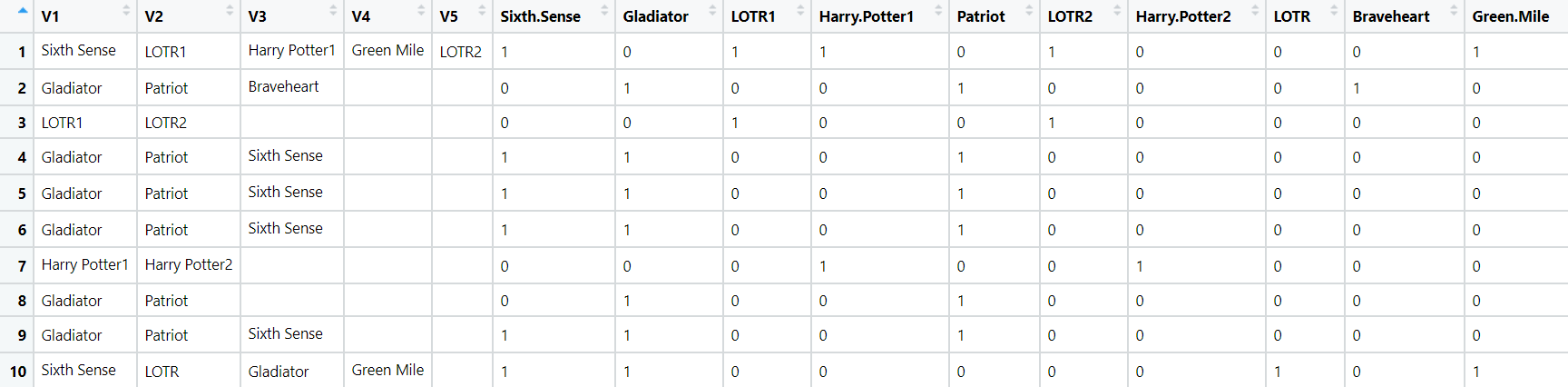
**# We generate top 10 Association Rules of groceries which is more purchased, because of this combinations we as a Data Scientist help The Departmental Store find the pattern of the products which is more sale either individually or together using Association Rules. See more in the excel sheet which I created and attach in the folder.**

|  |  |
| --- | --- |
| antecedents | consequents |
| {'berries'} | {'whipped/sour cream'} |
| {'other vegetables', 'tropical fruit'} | {'pip fruit'} |
| {'beef', 'whole milk'} | {'root vegetables'} |
| {'other vegetables', 'tropical fruit'} | {'root vegetables'} |
| {'other vegetables', 'tropical fruit'} | {'citrus fruit'} |
| {'beef'} | {'root vegetables'} |
| {'yogurt', 'whole milk'} | {'butter'} |
| {'whipped/sour cream', 'tropical fruit'} | {'other vegetables'} |
| {'other vegetables', 'pip fruit'} | {'root vegetables'} |
| {'other vegetables', 'whole milk'} | {'root vegetables'} |

**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**



**1) Business Problem:**

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

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Types** | **Relevance** |
| V1 | Movie Names | Qualitative, Discrete, Nominal | Relevant |
| V2 | Movie Names | Qualitative, Discrete, Nominal | Relevant |
| V3 | Movie Names | Qualitative, Discrete, Nominal | Relevant |
| V4 | Movie Names | Qualitative, Discrete, Nominal | Relevant |
| V5 | Movie Names | Qualitative, Discrete, Nominal | Relevant |
| Sixth Sense | Sixth Sense movie watch or not (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| Gladiator | Gladiator movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| LOTR1 | LOTR1  movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| Harry Potter1 | Harry Potter1  movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| Patriot | Patriot movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| LOTR2 | LOTR2  movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| Harry Potter2 | Harry Potter2  movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| LOTR | LOTR  movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| Braveheart | Braveheart  movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |
| Green Mile | Green Mile  movie watch or noy (0’s or 1’s) | Quantitative, Discrete, Binary | Relevant |

1. Result Share the benefits/impact of the solution - how or in what way the business (client):

**We generate top 10 Association Rules of movies list, because of this combinations we as a Data Scientist Analyze the data and come up with different rules of movie list so that the business objective is achieved. See more in the excel sheet which I created and attach in the folder.**

**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**



**1) Business Problem:**

**Objective:** **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 we as an Data Scientist to help them out, use the Association rules concept and provide insights to the company’s marketing team to improve its sales.**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Types** | **Relevance** |
| V1 | Product Names | Qualitative, Discrete, Nominal | Relevant |
| V2 | Product Names | Qualitative, Discrete, Nominal | Relevant |
| V3 | Product Names | Qualitative, Discrete, Nominal | Relevant |
| red | Red purchased | Quantitative, Discrete, Binary | Relevant |
| white | White purchased | Quantitative, Discrete, Binary | Relevant |
| green | Green purchased | Quantitative, Discrete, Binary | Relevant |
| yellow | Yellow purchased | Quantitative, Discrete, Binary | Relevant |
| orange | orange purchased | Quantitative, Discrete, Binary | Relevant |
| blue | blue purchased | Quantitative, Discrete, Binary | Relevant |

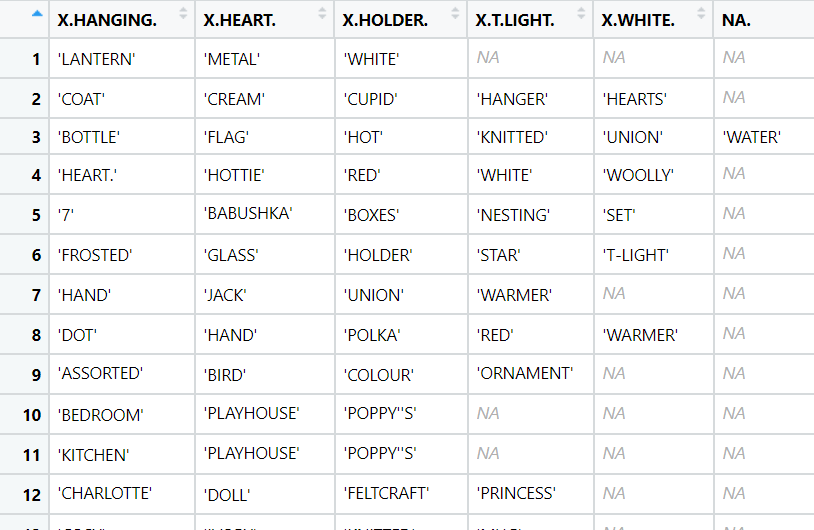
6.Result Share the benefits/impact of the solution - how or in what way the business (client):

**#We generate top 10 Association Rules on how Mobile Phone manufacturing company will launch its three brand new phone into the market by analyze the data of its previous model sales in different regions, and provide insights to the company’s marketing team to improve its sales. See more in the excel sheet which I created and attach in the folder.**

**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**



**1) Business Problem:**

**Objective:** **A retail store in India, has its transaction data, and it would like to know the buying pattern of the consumers in its locality, we as a Data Scientist 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.**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Types** | **Relevance** |
| X.HANGING. | Product Names | Qualitative, Discrete, Nominal | Relevant |
| X.HEART. | Product Names | Qualitative, Discrete, Nominal | Relevant |
| X.HOLDER. | Product Names | Qualitative, Discrete, Nominal | Relevant |
| X.T.LIGHT. | Red purchased | Qualitative, Discrete, Nominal | Relevant |
| X.WHITE. | White purchased | Qualitative, Discrete, Nominal | Relevant |
| NA. | Green purchased | Qualitative, Discrete, Nominal | Relevant |

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

**# So, we see top 10 Association 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, See more in the excel sheet which I created and attach in the folder.**