# MODULE 23

1. **Book Dataset**

**Package**

**Package used in R**

arules- To run the apriori algorithm

arulesViz- Plot and Visualization

**Package used in python**

Pandas - Data Manipulation

Matplotlib - Plot and Visualization

mlxtend.frequent\_patterns

Apriori - To Run the apriori algorithm

association\_rules- To Run the Association Rules

**Loading the data**

Loading the book Dataset in R and Python

**EDA**

1)No NaN Value Found

**Modeling**

**Apriori = Min\_support = 0.08 , Max\_len = 4, Use\_colnames = True, Verbose=1 (254 Rules Created)**

**Top 5 Rules on High Lift and then High Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'ItalCook'}) | frozenset({'CookBks', 'ChildBks'}) | 0.09 | 0.75 | 2.93 |
| frozenset({'DoItYBks', 'YouthBks'}) | frozenset({'CookBks', 'ChildBks'}) | 0.08 | 0.71 | 2.77 |
| frozenset({'YouthBks', 'GeogBks'}) | frozenset({'CookBks', 'ChildBks'}) | 0.08 | 0.69 | 2.69 |
| frozenset({'DoItYBks', 'GeogBks'}) | frozenset({'CookBks', 'ChildBks'}) | 0.09 | 0.67 | 2.62 |
| frozenset({'DoItYBks', 'ArtBks'}) | frozenset({'CookBks', 'ChildBks'}) | 0.08 | 0.66 | 2.59 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'RefBks'}) | frozenset({'DoItYBks', 'ChildBks'}) | 0.09 | 0.42 | 2.28 |
| frozenset({'CookBks', 'ChildBks'}) | frozenset({'ItalCook'}) | 0.09 | 0.33 | 2.93 |
| frozenset({'CookBks'}) | frozenset({'ItalCook'}) | 0.11 | 0.26 | 2.32 |
| frozenset({'CookBks', 'ChildBks'}) | frozenset({'DoItYBks', 'GeogBks'}) | 0.09 | 0.35 | 2.62 |
| frozenset({'ChildBks', 'ArtBks'}) | frozenset({'GeogBks'}) | 0.10 | 0.63 | 2.27 |

**Top 5 Rules Without Reducdancies on high Lift and then high Support**

1. **Groceries Dataset**

**Package**

**Package used in R**

arules- To run the apriori algorithm

arulesViz- Plot and Visualization

**Package used in python**

Pandas - Data Manipulation

Matplotlib - Plot and Visualization

mlxtend.frequent\_patterns

Apriori - To Run the apriori algorithm

association\_rules- To Run the Association Rules

**Loading the data**

Loading the Groceries Dataset in R and Python

**EDA**

1)No NaN Value Found

**Modeling**

**Apriori = Min\_support = 0.003, Max\_len = 5, Use\_colnames = True, Verbose=1 (8812 Rules Created )**

**Top 5 Rules on High Lift and then High Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'flour'}) | frozenset({'sugar'}) | 0.005 | 0.287 | 8.463 |
| frozenset({'tropical fruit', 'root vegetables'}) | frozenset({'citrus fruit', 'other vegetables'}) | 0.004 | 0.213 | 7.361 |
| frozenset({'citrus fruit', 'root vegetables'}) | frozenset({'tropical fruit', 'other vegetables'}) | 0.004 | 0.253 | 7.045 |
| frozenset({'tropical fruit', 'curd'}) | frozenset({'whole milk', 'yogurt'}) | 0.004 | 0.386 | 6.892 |
| frozenset({'butter', 'other vegetables'}) | frozenset({'whole milk', 'whipped/sour cream'}) | 0.004 | 0.198 | 6.142 |

**Top 5 Rules Without Reducdancies on high Lift and then high Support and Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'flour'}) | frozenset({'sugar'}) | 0.00 | 0.29 | 8.46 |
| frozenset({'liquor'}) | frozenset({'bottled beer'}) | 0.00 | 0.42 | 5.24 |
| frozenset({'herbs'}) | frozenset({'root vegetables'}) | 0.01 | 0.43 | 3.96 |
| frozenset({'tropical fruit', 'citrus fruit'}) | frozenset({'pip fruit'}) | 0.01 | 0.28 | 3.71 |
| frozenset({'tropical fruit', 'curd'}) | frozenset({'yogurt'}) | 0.01 | 0.51 | 3.69 |

1. **My Movies Dataset**

**Package**

**Package used in R**

arules- To run the apriori algorithm

arulesViz- Plot and Visualization

**Package used in python**

Pandas - Data Manipulation

Matplotlib - Plot and Visualization

mlxtend.frequent\_patterns

Apriori - To Run the apriori algorithm

association\_rules- To Run the Association Rules

**Loading the data**

Loading the My Movies Dataset in R and Python

**EDA**

1. Removed the Categorical variable’s and used only the Dummy Variables.
2. No NaN Value Found

**Modeling**

**Apriori = Min\_support = .08 , Max\_len = 4, Use\_colnames = True, Verbose=1 (208 Rules Created )**

**Top 5 Rules on High Lift and then High Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'LOTR2'}) | frozenset({'LOTR1'}) | 0.2 | 1 | 5 |
| frozenset({'LOTR1'}) | frozenset({'LOTR2'}) | 0.2 | 1 | 5 |
| frozenset({'Gladiator', 'Green Mile'}) | frozenset({'LOTR'}) | 0.1 | 1 | 10 |
| frozenset({'Sixth Sense', 'LOTR1'}) | frozenset({'Harry Potter1', 'Green Mile'}) | 0.1 | 1 | 10 |
| frozenset({'Harry Potter1', 'Sixth Sense'}) | frozenset({'Green Mile', 'LOTR1'}) | 0.1 | 1 | 10 |
| frozenset({'Harry Potter1', 'Green Mile'}) | frozenset({'Sixth Sense', 'LOTR1'}) | 0.1 | 1 | 10 |

**Top 5 Rules Without Reducdancies on high Lift and then high Support and Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'Green Mile', 'LOTR1'}) | frozenset({'LOTR2', 'Harry Potter1'}) | 0.1 | 1 | 10 |
| frozenset({'Gladiator', 'Green Mile'}) | frozenset({'LOTR'}) | 0.1 | 1 | 10 |
| frozenset({'Sixth Sense', 'LOTR1'}) | frozenset({'Harry Potter1', 'Green Mile'}) | 0.1 | 1 | 10 |
| frozenset({'LOTR2', 'Green Mile'}) | frozenset({'Sixth Sense', 'LOTR1'}) | 0.1 | 1 | 10 |
| frozenset({'LOTR2', 'Green Mile'}) | frozenset({'Harry Potter1', 'Sixth Sense'}) | 0.1 | 1 | 10 |
| frozenset({'LOTR'}) | frozenset({'Gladiator', 'Sixth Sense', 'Green Mile'}) | 0.1 | 1 | 10 |

**4) My Phone Dataset**

**Package**

**Package used in R**

arules- To run the apriori algorithm

arulesViz- Plot and Visualization

**Package used in python**

Pandas - Data Manipulation

Matplotlib - Plot and Visualization

mlxtend.frequent\_patterns

Apriori - To Run the apriori algorithm

association\_rules- To Run the Association Rules

**Loading the data**

Loading the My Phone Dataset in R and Python

**EDA**

1. Categorical Variable Removed and Used Dummy variables to proceed Further
2. No NaN Value Found

**Modeling**

**Apriori = Min\_support = .05, Max\_len = 3, Use\_colnames = True, Verbose=1 ( 18 Rules Created)**

**Top 5 Rules on High Lift and then High Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'orange'}) | frozenset({'white'}) | 0.18 | 1.00 | 1.57 |
| frozenset({'white'}) | frozenset({'orange'}) | 0.18 | 0.29 | 1.57 |
| frozenset({'green', 'white'}) | frozenset({'red'}) | 0.09 | 1.00 | 1.83 |
| frozenset({'orange', 'red'}) | frozenset({'white'}) | 0.09 | 1.00 | 1.57 |
| frozenset({'green', 'red'}) | frozenset({'white'}) | 0.09 | 1.00 | 1.57 |

**Top 5 Rules Without Reducdancies on high Lift and then high Support and Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
| frozenset({'white'}) | frozenset({'red'}) | 0.36 | 0.57 | 1.05 |
| frozenset({'white'}) | frozenset({'orange'}) | 0.18 | 0.29 | 1.57 |
| frozenset({'white'}) | frozenset({'orange', 'red'}) | 0.09 | 0.14 | 1.57 |
| frozenset({'white'}) | frozenset({'blue'}) | 0.36 | 0.57 | 1.05 |
| frozenset({'red'}) | frozenset({'green', 'white'}) | 0.09 | 0.17 | 1.83 |
| frozenset({'red'}) | frozenset({'blue'}) | 0.36 | 0.67 | 1.22 |

1. **Transaction Retail Dataset**

**Package**

**Package used in R**

arules- To run the apriori algorithm

arulesViz- Plot and Visualization

**Package used in python**

Pandas - Data Manipulation

Matplotlib - Plot and Visualization

mlxtend.frequent\_patterns

Apriori - To Run the apriori algorithm

association\_rules- To Run the Association Rules

**Loading the data**

Loading the My Phone Dataset in R and Python

**EDA**

1. No NaN Value Found

**Modeling**

**Apriori = Support = 0.005, Confidence = 0.055, Maxlen = 2**

**Top 5 Rules on High Lift and then High Confidence**

|  |  |  |  |
| --- | --- | --- | --- |
| Support | Confidence | Coverage | Lift |
| {X.HANGING.='BAG'} => {X.HOLDER.='LUNCH'} | 0.009 | 0.123 | 0.075 |
| {X.HANGING.='BAG'} => {X.HEART.='JUMBO'} | 0.018 | 0.246 | 0.075 |
| {X.HANGING.='BAG'} => {X.HEART.='DESIGN'} | 0.009 | 0.122 | 0.075 |
| {X.HANGING.='BAG'} => {X.T.LIGHT.='RETROSPOT'} | 0.011 | 0.150 | 0.075 |
| {X.HANGING.='BAG'} => {X.HOLDER.='RED'} | 0.011 | 0.150 | 0.075 |