

# User Guide for Interactive Network Visualization Application

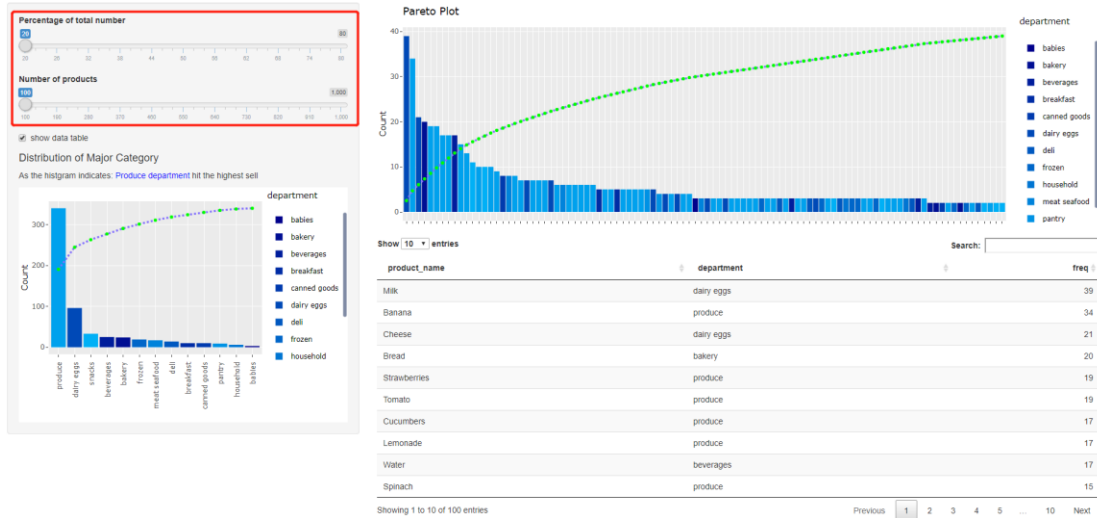
## Data Upload

ID	order_id	product_id	RACE	product_name	user_id	department_id	CLASS
4	4	NA		Plain Pre-Sliced Bagels	178520	3	Train
4	4	NA		Plain Pre-Sliced Bagels	178520	3	Train
4	4	NA		Plain Pre-Sliced Bagels	178520	3	Train
4	4	NA		Plain Pre-Sliced Bagels	178520	3	Train
4	4	NA		Tiny Twists Pretzels	178520	19	Train
4	4	NA		Tiny Twists Pretzels	178520	19	Train
4	4	NA		Tiny Twists Pretzels	178520	19	Train
4	4	NA		Tiny Twists Pretzels	178520	19	Train
4	4	NA		Cracker	178520	19	Train
4	4	NA		Cracker	178520	19	Train
4	4	NA		Cracker	178520	19	Train
4	4	NA		Cracker	178520	19	Train
4	4	NA		Cracker	178520	19	Train
4	4	NA		Orange	178520	7	Train
4	4	NA		Orange	178520	7	Train
4	4	NA		Orange	178520	7	Train
4	4	NA		Orange	178520	7	Train
4	4	NA		Orange	178520	7	Train

In this section, the user can upload at most two csv files for further analysis. If there is no file uploaded, the **default(supermarket) dataset** would be in use. Below are functions provided for data preprocessing.

- Header
  - Tick the box if the first row of the sheet is column name/header.
- Separator
  - Choose “Comma”, “Semi-colon” or “Tab” to delimit columns.
- Join Type
  - Inner Join: Select all records from Table A and Table B, where the join condition is met.
  - Outer Join: Select all records from Table A and Table B, regardless of whether the join condition is met or not.
  - Left Join: Select all records from Table A, along with records from Table B for which the join condition is met (if at all).
  - Right Join: Select all records from Table B, along with records from Table A for which the join condition is met (if at all).
- Sorting
  - The default sorting is descending. Tick the box to change the sorting to be ascending order

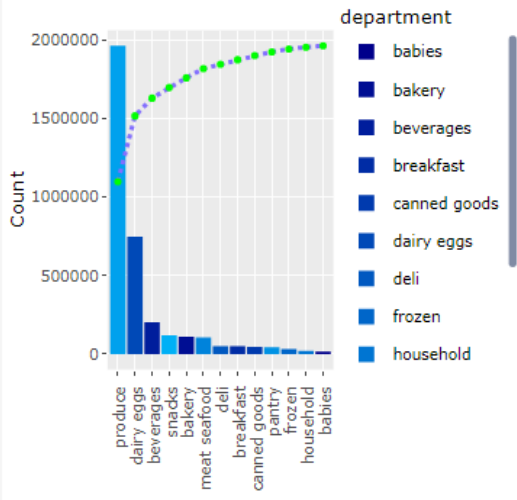
# Pareto Chart



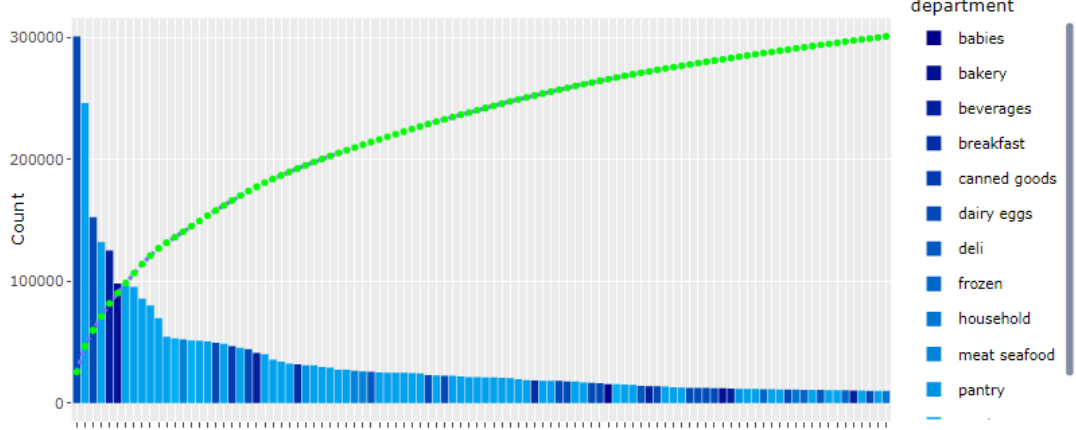
- Data Selection
  - The selection panel is located at the upper left corner. It includes two sliders, one controls the percentage of the total number of products and another controls the number of products.
    - The charts and data table would update accordingly as the user changes the parameter(s).
- Checkbox  show data table
  - Tick the box if you would like to show the data table at the lower right corner.
- Pareto Chart for Department/Product
  - Below is the simple explanation of the different items in the chart:
    - Bar Chart: Frequency of the department/product.
    - Line: Running total percentage of the products in the transaction data
    - Bar Color: By department.
    - Interactivity: Detailed information pops when hovering on it.

### Distribution of Major Category

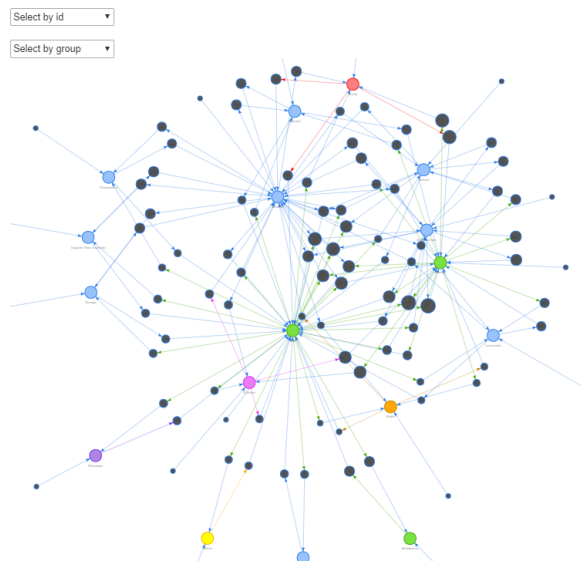
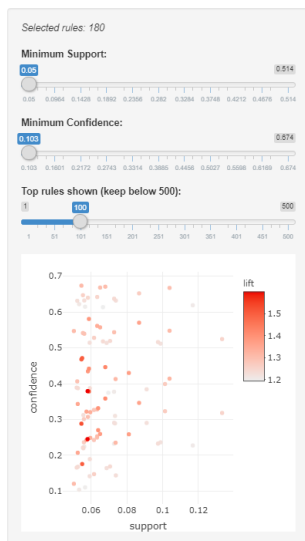
As the histogram indicates: Produce department hit the highest sell



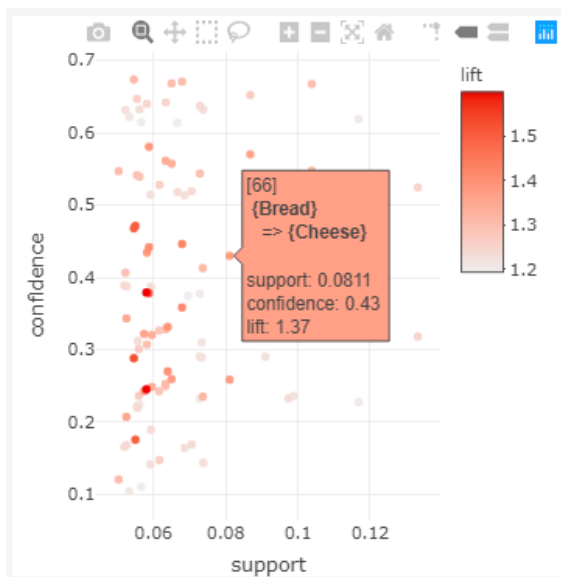
### Pareto Plot



## Market Basket Analysis



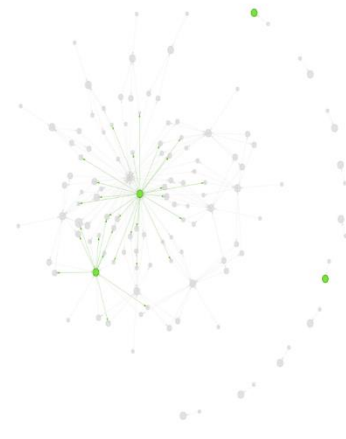
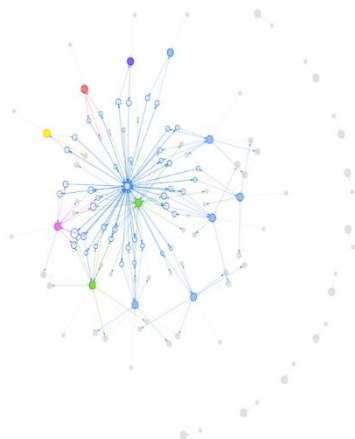
- Parameters for market basket analysis
  - Minimum Support: The minimum support required to be included in the association rules
  - Minimum Confidence: The minimum confidence required to be included in the association rules
  - Top Rules Shown: Number of top confidence rules to show.
    - Note that the rules are sorted by support descending in the algorithm
- Scatter Plot
  - Scatter plot of rules with confidence as the y-axis and support as the x-axis
  - The diverging color is used to differentiate the count of the transaction.
    - The darker the red color indicates the higher the transaction count is
  - Hovering at scatter points shows detail of the specific rule



- Network Analysis
  - To highlight the association rules, following functions are included in the visualization.
    - To highlight the individual product, select from the dropdown under **id**
    - To highlight products from a department, select from the dropdown under **group**

Select by group  
Banana

Select by id  
dairy eggs



- To focus on a specific section on the association rules, you can zoom in & out

