*Market Basket Analysis*

In the current project, by applying association rules on Instacart transactional data of the

customers it does not extract preferences of the individual customer rather it does specify the

connections at product levels of each transaction for all the customers. There by, mined rules

are used for the recommendation system. Perhaps, the strong association rules are supportive

for the recommender system in the approach of suggestions of most likely corelated with

similar products. Large datasets have been used for the analysis of the pattern and in order to

overcome the memory and computing issues, analysis of the data has been done in Google

Collab with the usage of GPU while unstacking the tractional data of the customers.

Recommendation system can be split into four stages:

1. The collection and pre-processing of raw data;

2. Convert pre-processed data into an easily achievable form using the selected machine

learning method such as Apriori or FP Growth algorithm;

3. Create a model of learning (training) using transformed data;

4. Use the previously developed set of association rules to report recommendations to the

User

Python libraries, frameworks and modules are very simple and powerful as compared

with the old days. Python has replaced many of the industry's languages, one of the reasons for

this is its vast collection of libraries, and is one of the most popular programming languages

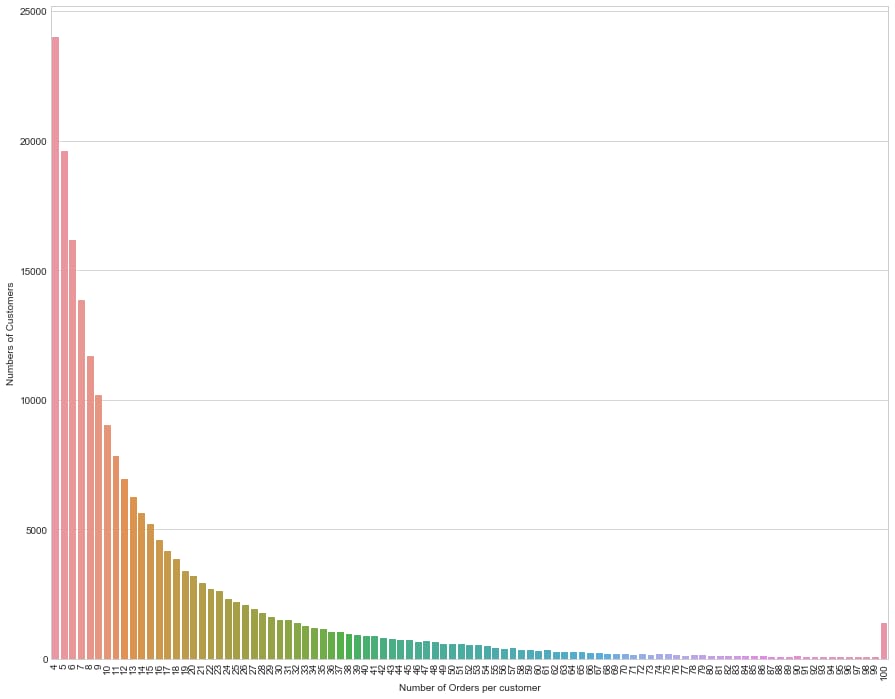
for this task today. Python libraries used in the project are:

* Pandas
* Numpy
* Matplotlib
* Seaborn
* MLxtend

Market basket analysis is a data mining technique used to discover associations and patterns in transactional data, such as customer purchase history. Data preparation is a crucial step in the process of conducting market basket analysis, as it involves getting the data ready for analysis. Here are the key steps involved in data preparation for market basket analysis:

* Data Collection
* Data Cleaning
* Data Transformation
* Data Encoding
* Support and Confidence Thresholds
* Apriori Algorithm (or other association rule mining algorithms)
* Rule Selection
* Post-Processing
* Visualization
* Interpretation

What are the minimum and maximum orders received from customers?



**Frequency Of Orders per customer**

Here, there are only 4 orders from 23,986 customers, only 5 orders from 19,590 customers. As

the number of customers orders increases, the number of customers ordering decreases.

Customers large percentage make 4 to 12 orders. When a company figures out a way to increase

the number of repeat customer orders, the sales will increase.

The Apriori algorithm works in two steps:

Prune and Join:

1.Generate all frequent item sets – A frequent item set is an item set that has transaction support

above minimum support.

2.Generate all confident association rules from frequent item sets – A confident association

rule is a rule with confidence above minimum confidence.

To apply Apriori algorithm on Instacart dataset, the Apriori class is applied that is imported

from the Apyori library.

• k-itemset is the itemset which contains: k element number.

• Lk refers to frequent sets of items with; k items.

• Ck corresponds with frequent sets of candidate items with elements;k

There are a variety of programming languages that can be used for the back end and front end of a market basket analysis application. Some popular choices include:

**Back end:**

Python

PySpark

MLlib

**Front end:**

JavaScript