1. **Problem Statement**

We need to do Market Analysis of store data with some food item categories such as almonds, avocado, green grapes, vegetables mix, cottage cheese, energy drink, mineral water, spinach, olive oil, honey, salad etc.

1. **Process**

We start with preprocessing step in which we first use the One-Hot Encoding technique to convert our categorical variable into numerical variable. After applying the One- Hot Encoding technique we end of getting 120 columns. We observe that there are so many ‘nan’ values present in our dataset, which may create an issue during our analysis so we just drop that ‘nan’ values. Then we create a table to find the frequency of each item and we observe that “Mineral Water” is most occurring or we can say most purchased item in the store.

1. **Exploratory Data Analysis**

After completing the preprocessing steps we move on to the Data Visualization part and try make our analysis more powerful by forming the graphs.

* We first make a bar graph for our analysis and analyze the frequency of top 10 items present in the store.
* Secondly we make the word cloud and with the help of word cloud we try to analyze which word occurs most frequently.
* Then we import the new library named “networkx” and try to represent our analysis in the form of nodes.
* Make 3 graphs using the technique stated above and try to get top 30 items which is mostly purchased.

With the help of EDA we are able make our selection between relevant and irrelevant attributes and we choose top 24 items/columns out of that 119 columns.

1. **Approach**

Till now we will get the idea which features are relevant that means these items are more purchased items and they directly affects the sale and profits of the store, so the store owner make sure that he/she need to keep these items on their store always. Now we are going apply Apriori Algorithm and try to find out the rules for making future decisions.

1. **Rules and Analysis**

Now we the help of Apriori Algorithm we are able to generate the rules and we that rules we are able to get the Support, Confidence and Rules.

Support: This says how popular an item set is, as measured by the proportion of transactions in which an item set appears.

Confidence: This says how likely item Y is purchased when item X is purchased, expressed as {X -> Y}. This is measured by the proportion of transactions with item X, in which item Y also appears.

Lift: This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is. A lift value greater than 1 means that item Y is likely to be bought if item X is bought, while a value less than 1 means that item Y is unlikely to be bought if item X is bought.

From the generated output we can filter out some insights such as:

* There are 28% of transactions containing mineral water also contain frozen vegetables.
* There are 29% of transactions containing chicken also light cream.
* There are 30% of transactions containing mushroom cream sauce also contains escalope.