Business Understanding :

* Will ask about Current product Shelving and inventory process.?
* Why do they want to optimize ?
* Does Market Basket Analysis required for Subset of products or Whole Products ?
* What kind of data do they have about transaction ?
* Are there any regulatory or compliance considerations that may impact how you can use market basket analysis to optimize product shelving and inventory ?
* What is your timeline for implementing the change ?
* Have you conducted market basket analysis before, and if so, what insights did you gain from it?
* Do you have any concerns about data privacy or security when it comes to analyzing customer transaction data for market basket analysis purposes?

Data Understanding:

* What kind of data would you need ?

Market basket analysis requires transactional data, which is a type of data that describes the purchase history of customers. Specifically, the data should include a record of all the transactions that have taken place in a particular store or retail outlet. Each transaction should be associated with a unique identifier, such as a receipt number, and a list of the items that were purchased in that transaction.

* What will be the target variable ?

Market basket analysis is an unsupervised learning technique, which means that there is no target variable in the traditional sense. The goal of market basket analysis is to uncover patterns and relationships in transactional data, and to identify which items are frequently purchased together.

The output of a market basket analysis is typically a set of rules or associations, which describe the relationships between the items. These rules are often expressed in the form of if-then statements, such as "If a customer buys item A, they are likely to buy item B as well."

In summary, market basket analysis does not involve the use of a target variable, but instead aims to uncover relationships and patterns in transactional data.

* What will be the different kind of features you needed ?

Transaction ID ,Item ID , Time stamp,CustomerID,Item description,Quantity , category,Price.

* What ares some of the example you think might be useful ?

Transaction ID , Item ID

* What are the Quality Checks that can be done ?

Check for anomalies : Example If we see sales for particular day is very very high compared to other days . suppose in a single transaction , Lots of products are there . In some transaction , there were no products purchased . If such cases happen , we should ask the Client Whether those informations are correct .

Check for null Values : If there are lots of null values , ask the client why it happened .

Duplicates:

Data Integrity:

* How would you approach Data Exploration

Item Frequency : Most bought and Least bought

Market Basket Segmentation :customer purchase pattern based on demographic

Time Series Based Analysis

* What are the preprocessing steps would you apply for different types of Variables?

For Null Values : For market masket analysis , the important columns are products they bought . so there wont be null values . if the entire rows contain 0 , then delete that row.

* What kind of train-test split would you do ?

No, it is not necessary to split data into training and testing sets when doing market basket analysis, aka. association rules mining.Market basket analysis does not build a model. Instead, it finds piecewise of rules of associations (measured in support, confidence, lift and etc.) between items.

* What kind of models and metrics would you use ?

Apriori Algorithm

AIS

SETM

FP Growth

For instance, the Apriori algorithm is an efficient and widely used algorithm for mining frequent itemsets and generating association rules, but it may not be the best choice for datasets with a large number of items or transactions. In such cases, algorithms like FP-growth or Eclat may be more efficient

* Metrics

Support: Support is the percentage of transactions that contain a particular itemset. It is a measure of how frequently the itemset appears in the dataset.

Confidence: Confidence is the conditional probability that an item Y is purchased given that item X is purchased. It measures the strength of the association between the two items.

Lift: Lift is the ratio of the observed support of an itemset to the expected support of the itemset if the items were independent. It measures the strength of the association between two items while taking into account their individual frequencies.

Conviction: Conviction is a measure that indicates how much more likely it is that an association rule is true compared to if it were simply due to chance.

Interest: Interest measures the difference between the observed and expected support of the consequent in an association rule.

Kulczynski measure: The Kulczynski measure is a combination of support and confidence that measures the similarity between two itemsets.

Jaccard similarity coefficient: The Jaccard similarity coefficient measures the similarity between two sets of items based on the size of their intersection and union.

Cosine similarity: The cosine similarity measures the similarity between two sets of items based on the angle between their vectors in a multi-dimensional space.