Problem Definition and Design Thinking

Problem Definition: The problem is to perform market basket analysis on a provided dataset to unveil hidden patterns and associations between products. The goal is to understand customer purchasing behavior and identify potential cross-selling opportunities for a retail business. This project involves using association analysis techniques, such as Apriori algorithm, to find frequently co-occurring products and generate insights for business optimization.

Design Thinking:

1. Data Source: Choose a dataset containing transaction data, including lists of purchased products.
2. Data Preprocessing: Prepare the transaction data by transforming it into a suitable format for association analysis.
3. Association Analysis: Utilize the Apriori algorithm to identify frequent itemsets and generate association rules.
4. Insights Generation: Interpret the association rules to understand customer behavior and cross-selling opportunities.
5. Visualization: Create visualizations to present the discovered associations and insights.
6. Business Recommendations: Provide actionable recommendations for the retail business based on the insights.

Market Basket insights

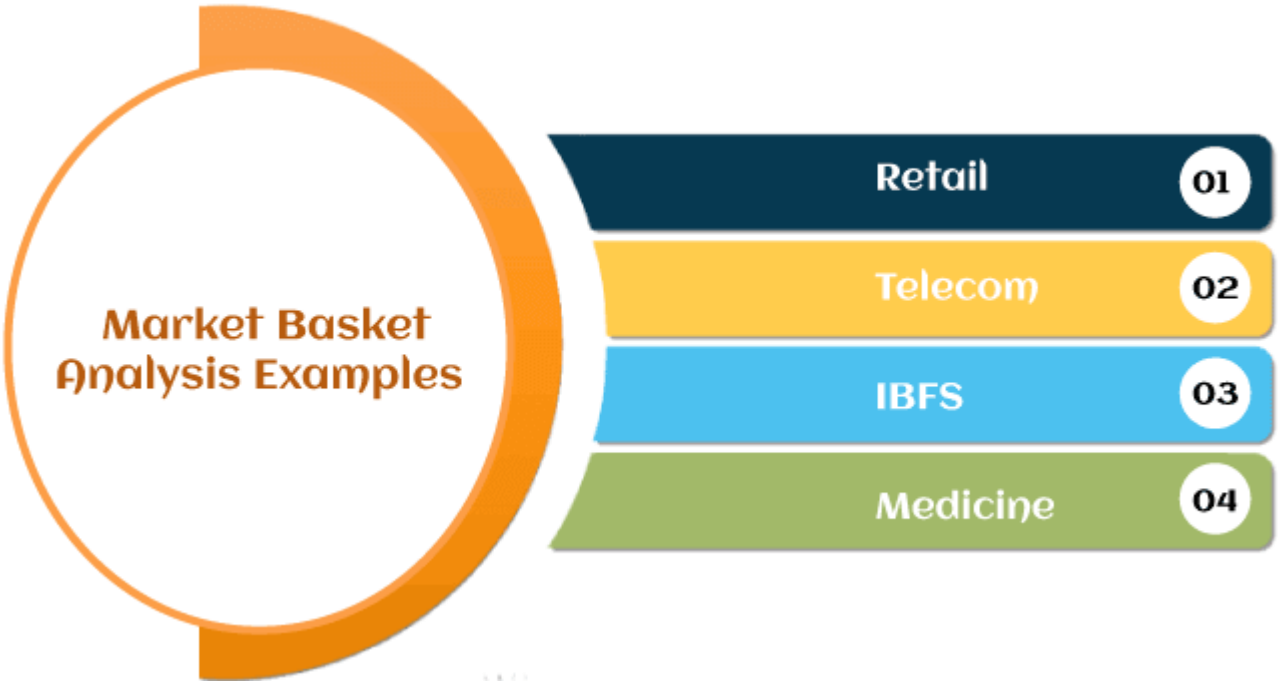
Market basket analysis technique used by retailers to increase sales by better understanding customer purchasing patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings and products that are likely to be purchased together.

The adoption of market basket analysis was aided by the advent of electronic point-of-sale (POS) systems. Compared to handwritten records kept by store owners, the digital records generated by POS systems made it easier for applications to process and analyze large volumes of purchase data.

Implementation of market basket analysis requires a background in statistics and data science and some algorithmic computer programming skills. For those without the needed technical skills, commercial, off-the-shelf tools exist.

Descriptive market basket analysis: This type only derives insights from past data and is the most frequently used approach. The analysis here does not make any predictions but rates the association between products using statistical techniques. For those familiar with the basics of Data Analysis, this type of modelling is known as unsupervised learning.

For example, suppose 5000 transactions have been made through a popular e-Commerce website. Now they want to calculate the support, confidence, and lift for the two products. For example, let's say pen and notebook, out of 5000 transactions, 500 transactions for pen, 700 transactions for notebook, and 1000 transactions for both.



I have analysed the dataset with visualizations and perform A rule mining with the help of Apriori algorithm. I have never realized or questioned myself why these items are kept closely in the supermarket, thought that it was for customer's convenience but little did I know that it had a business impact.

```
#BasicnecessaryLibraries

                        import numpy as np

                        import pandas as pd
```

```
#Visualization libraries

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

import altair as alt

import plotly.express as px

from wordcloud import WordCloud,STOPWORDS

import holoviews as hv

from holoviews import opts

hv.extension('bokeh')


#Apriori libraries

from mlxtend.frequent_patterns import apriori

from mlxtend.frequent_patterns import association_rules
```

Suppose, as manager of an *AllElectronics* branch, you would like to learn more about the buying habits of your customers. Specifically, you wonder, *"Which groups or sets of items are customers likely to purchase on a given trip to the store?"* To answer your question, market basket analysis may be performed on the retail data of customer transactions at your store. You can then use the results to plan marketing or advertising strategies, or in the design of a new catalog. For instance, market basket analysis may help you design different store layouts. In one strategy, items that are frequently purchased together can be placed in proximity to further encourage the combined sale of such items. If customers who purchase computers also tend to buy antivirus software at the same time, then placing the hardware display close to the software display may help increase the sales of both items.

In an alternative strategy, placing hardware and software at opposite ends of the store may entice customers who purchase such items to pick up other items along the way. For instance, after deciding on an expensive computer, a customer may observe security systems for sale while heading toward the software display to purchase antivirus software, and may decide to purchase a home security system as well. Market basket analysis can also help retailers plan which items to put on sale at reduced prices. If customers tend to purchase computers and printers together, then having a sale on printers may encourage the sale of printers *as well as* computers.