A

1)

One question that can be answered with market basket analyis is: Which items are frequently purchased together?

2)

What are the most frequently purchased groups of items in the transaction data set.

B

1)

Market basket analysis analyzes the selected data set by identifying item sets that are frequently purchased. Association rules are generated based on the frequency of item sets. Once the association rules are generated they can be evaluated using metrics such as confidence, lift, and support. An expected outcome would be for the analysis to identify all frequently purchased item sets and the relationship or association rules between the items. Each association rule will have a metric such as confidence, lift, and support applied to it. Finally, the metrics can be evaluated and decisions made concerning product placement, promotions and others.

2)

From row 9 of the provided data set a customer purchases HP 65 ink, and Cat8 Ethernet Cable in a single transaction.

3)

One assumption of market basket analysis is that the data is transaction based. This means that each data point represents a single transaction.

C)

1

Transform the data set.

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

file_path = '/home/dj/skewl/D212/3/teleco_market_basket.csv'
pd.set_option('display.max_columns', None)
# Read the data from the CSV file into a DataFrame
df = pd.read_csv(file_path)
# drop empty rows
df = df.dropna(thresh=1)
# reset index
df = df.reset_index(drop=True)
#replace nans
df = df.fillna("unknown")
# change each row to a list
df = df.apply(lambda row: row.tolist(), axis=1)
# drop empty rows
df = df.dropna()
#convert data frame to list of lists
df = list(df)
# Remove the unknown value from each list
df = [[item for item in inner_list if item != "unknown"] for inner_list in df]
# write to file
pd.DataFrame(df).to_csv('prepared_data.csv', index=False)
```

2)

Code to generate association rules.

In [2]:
```
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
# Convert the list of lists to a DataFrame with TransactionEncoder
te = TransactionEncoder()
#one hot encode
transactions_encoded = te.fit_transform(df)
#make dataframe
df_transactions = pd.DataFrame(transactions_encoded, columns=te.columns_)
# Find frequent itemsets with a minimum support of 0.5
frequent_itemsets = apriori(df_transactions, min_support=0.01, use_colnames=True)
# Generate association rules with a minimum lift of 0.01
rules = association_rules(frequent_itemsets, metric='lift', min_threshold=0.01)
```

3)

Support, lift, and confidence values of the association rules table.

In [3]:
```
print(rules[['support','confidence','lift']])
```

```
      support  confidence      lift
0    0.023064    0.456464  1.914955
1    0.023064    0.096756  1.914955
2    0.010132    0.200528  1.223888
3    0.010132    0.061839  1.223888
4    0.015198    0.300792  2.321232
..        ...         ...       ...
427  0.010932    0.340249  1.954217
428  0.010932    0.308271  1.881480
429  0.010932    0.066721  1.881480
430  0.010932    0.062787  1.954217
431  0.010932    0.084362  2.152382

[432 rows x 3 columns]
```

Top 3 relevant association rules

In [4]:
```
#sort itemsets by relevance
rules = rules.sort_values(by=['support','lift','confidence'], ascending=False)
print("Association Rules:")
print(rules.head(3))
```

```
Association Rules:
                          antecedents                         consequents  \
178  (VIVO Dual LCD Monitor Desk mount)    (Dust-Off Compressed Gas 2 pack)
179    (Dust-Off Compressed Gas 2 pack)  (VIVO Dual LCD Monitor Desk mount)
142                        (HP 61 ink)    (Dust-Off Compressed Gas 2 pack)

     antecedent support  consequent support   support  confidence      lift  \
178            0.174110            0.238368  0.059725    0.343032  1.439085
179            0.238368            0.174110  0.059725    0.250559  1.439085
142            0.163845            0.238368  0.052660    0.321400  1.348332

     leverage  conviction  zhangs_metric
178  0.018223    1.159314       0.369437
179  0.018223    1.102008       0.400606
142  0.013604    1.122357       0.308965
```

4)

Rule 1.

Antecedent = (VIVO Dual LCD Monitor Desk mount). Buying this item leads to buying the consequent.

Consequent = (Dust-Off Compressed Gas 2 pack)

Support = 0.059725. This is the joint probability that the antecedent and consequents will be purchased in the same transaction.

Confidence = 0.34302. This is the probability the consequent will be purchased if the antecedent is purchased.

Lift = 1.439085. This indicates how much more likely the antecedent and consequents are to be purchased together rather than by themselves

Rule 2.

Antecedent = (Dust-Off Compressed Gas 2 pack). Buying this item leads to buying the consequent.

Consequent = (VIVO Dual LCD Monitor Desk mount)

Support = 0.059725. This is the joint probability that the antecedent and consequents will be purchased in the same transaction.

Confidence = 0.250559. This is the probability the consequent will be purchased if the antecedent is purchased.

Lift = 1.439085. This indicates how much more likely the antecedent and consequents are to be purchased together rather than by themselves

Rule 3.

Antecedent = (HP 61 ink). Buying this item leads to buying the consequent.
Consequent = (Dust-Off Compressed Gas 2 pack)

Support = 0.052660. This is the joint probability that the antecedent and consequents will be purchased in the same transaction.

Confidence = 0.321440. This is the probability the consequent will be purchased if the antecedent is purchased.

Lift = 1.34883. This indicates how much more likely the antecedent and consequents are to be purchased together rather than by themselves

D

1)

Lift, confidence, and support are key metrics that describe the strength and nature of the relationship between antecedents and consequents. Lift describes how much more likely the antecedent and consequents are to be purchased togther rather than by themselves. Confidence describes the probability of a consequent being purchased if an antecedent is purchased. Finally, support describes the joint probability of the consequent and the antecedent being purchased together. Knowing these metrics is significant because it provides more detail of the relationship between the items and can drive decision making for businesses.

2)

The practical significance of the findings from my analysis are that we have identified three sets of items that have a statistically significant probability of being purchased in the same transaction. This analysis can drive business decisions such as product placement, marketing, and promotion design. Since we know that VIVO Dual LCD Monitor Desk mount and Dust-Off Compressed Gas 2 pack have a support of .05975, stores can place these items together, or feature advertisements for the two products on the same page of the newspaper. This is a good decision because we know the probability of a customer buying the Dust-Off if the customer buys the LCD monitor. That is one practically significant finding of the analysis.

3)

A recommended course of action based on the the results of D1 is that businesses should feature product placement of VIVO Dual LCD Monitor Desk mount and Dust-Off Compressed Gas 2 pack near each other because of the support metric. The data shows that there is a higher probability of selling Dust-Off to a customer that is buying a LCD monitor. Inside the store Dust-Off should be placed next to the HP ink because of the lift metric. This is because customers have a 1.3488 higher likelihood of purchasing Dust-Off when purchasing HP ink compared to those customers who do not buy HP ink. These are a couple of actionable strategies based on the results of the analysis.

In [ ]: