D212 Task 3

August 1, 2024

DATA MINING II - D212

Task 3

Eric D. Otten

Student ID: 011399183

1 A1. Propose one question relevant to a real-world organizational situation that you will answer using market basket analysis.

My research question for D212 Task 3 is: "How can we identify key associations of customer purchases from the telecommunications company using its market basket dataset?"

2 A2. Define one goal of the data analysis. Ensure your goal is reasonable within the scope of the selected scenario and is represented in the available data.

My primary goal is to uncover the key associations between customer purchases to enhance our understanding of customer behavior.

By leveraging market basket analysis, I aim to identify patterns and relationships among the products and services customers frequently buy together. This analysis will provide valuable insights into customer preferences and purchasing habits, enabling us to develop targeted marketing strategies and personalized discount offers.

Ultimately, these findings will help improve customer satisfaction and loyalty, ensuring that the telecommunication company's offerings align more closely with customer needs and preferences. This goal is both reasonable and attainable, given the rich data available in our market basket dataset, which captures detailed transactional information across a diverse customer base.

3 B1. Explain the reasons for using market basket analysis by doing the following: 1. Explain how market basket analyzes the selected data set. Include expected outcomes.

Market basket analysis is a powerful tool for understanding customer purchasing patterns by examining the items they buy together. For our telecommunications company, it involves analyzing the

7,501 customer purchase histories, each consisting of up to 20 different items related to technological areas.

This method uses association rules to identify relationships between items, which can then highlight frequently co-purchased products. By examining these associations, we can uncover trends and patterns that may not be immediately obvious. **The expected outcomes include discovering which products are often bought together, identifying complementary products, and understanding customer preferences more deeply.**

Imagine, if you will, a customer's purchase history includes items such as a smartphone, a Bluetooth headset, a screen protector, and a charging cable.

Using market basket analysis, we can identify that customers who buy smartphones are also likely to purchase screen protectors and Bluetooth headsets. This information allows us to create promotional bundles or targeted marketing campaigns that highlight these commonly co-purchased items, thereby increasing the likelihood of additional sales and improving the customer experience.

4   B2. Provide one example of transactions in the data set.

The code below performs some necessary data preparation such as removing empty rows before identifying the 17th (index: 16) transaction.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from mlxtend.frequent_patterns import association_rules, apriori
from mlxtend.preprocessing import TransactionEncoder

df = pd.read_csv("teleco_market_basket.csv")

df = df[df["Item01"].notna()]
df.reset_index(drop=True, inplace=True)
df.iloc[16]
```

```
[1]: Item01        Logitech M510 Wireless mouse
     Item02                          HP 61 ink
     Item03    Falcon Dust Off Compressed Gas
     Item04       HyperX Cloud Stinger Headset
     Item05                         HP 952 ink
     Item06           Premium Nylon USB Cable
     Item07       TopMate C5 Laptop Cooler pad
     Item08                                NaN
     Item09                                NaN
     Item10                                NaN
     Item11                                NaN
     Item12                                NaN
     Item13                                NaN
     Item14                                NaN
```

```
Item15                              NaN
Item16                              NaN
Item17                              NaN
Item18                              NaN
Item19                              NaN
Item20                              NaN
Name: 16, dtype: object
```

5 B3. Summarize one assumption of market basket analysis.

One key assumption of market basket analysis is that the purchasing behavior of customers in the dataset is representative of the broader customer base. This means we assume that the patterns and associations identified from the historical purchase data of the 7,501 customers are indicative of the general purchasing behavior of all our customers.

This assumption is crucial because it underpins the validity of the insights and recommendations derived from the analysis. By ensuring our data is comprehensive and accurately reflects customer behavior, we can confidently use these insights to inform our marketing and sales strategies, ultimately driving better business outcomes.

6 C1. Prepare and perform market basket analysis by doing the following: 1. Transform the data set to make it suitable for market basket analysis. Include a copy of the cleaned data set.

```
[2]: # Convert df into lists
     transactions = []
     for i in range(len(df)):
         transaction = [str(df.values[i, j]) for j in range(20) if str(df.values[i,
         ↪j]) != "nan"]
         transactions.append(transaction)

     te = TransactionEncoder()

     te_array = te.fit(transactions).transform(transactions)
     encoded_df = pd.DataFrame(te_array, columns=te.columns_)
     encoded_df.to_csv("encoded_df.csv")
     encoded_df
```

```
[2]:       10ft iPHone Charger Cable  10ft iPHone Charger Cable 2 Pack  \
     0                          True                             False
     1                         False                             False
     2                         False                             False
     3                         False                             False
     4                         False                             False
     ...                         ...                               ...
     7496                      False                             False
```

```
7497                     False                               False
7498                     False                               False
7499                     False                               False
7500                     False                               False

        3 pack Nylon Braided Lightning Cable  3A USB Type C Cable 3 pack 6FT  \
0                                      False                            True
1                                      False                           False
2                                      False                           False
3                                      False                           False
4                                      False                           False
...                                      ...                             ...
7496                                   False                           False
7497                                   False                           False
7498                                   False                           False
7499                                   False                           False
7500                                   False                           False

        5pack Nylon Braided USB C cables  ARRIS SURFboard SB8200 Cable Modem  \
0                                  False                               False
1                                  False                               False
2                                  False                               False
3                                  False                               False
4                                  False                               False
...                                  ...                                 ...
7496                               False                               False
7497                               False                                True
7498                               False                               False
7499                               False                               False
7500                               False                               False

        Anker 2-in-1 USB Card Reader  Anker 4-port USB hub  \
0                              False                 False
1                              False                 False
2                              False                 False
3                              False                 False
4                              False                 False
...                              ...                   ...
7496                           False                 False
7497                           False                 False
7498                           False                 False
7499                           False                 False
7500                           False                 False

        Anker USB C to HDMI Adapter  Apple Lightning to Digital AV Adapter  …  \
0                             False                                  False  …
1                             False                                   True  …
```

4

```
2                         False                          False  …
3                         False                          False  …
4                         False                          False  …
…                           …                              …    …
7496                      False                          False  …
7497                      False                           True  …
7498                      False                          False  …
7499                      False                          False  …
7500                      False                          False  …

      hP 65 Tri-color ink  iFixit Pro Tech Toolkit  iPhone 11 case  \
0                   False                    False           False
1                   False                    False           False
2                   False                    False           False
3                   False                    False           False
4                   False                    False           False
…                     …                        …               …
7496                False                    False           False
7497                False                    False           False
7498                False                    False           False
7499                False                    False           False
7500                False                    False           False

      iPhone 12 Charger cable  iPhone 12 Pro case  iPhone 12 case  \
0                       False               False           False
1                       False               False           False
2                       False               False           False
3                       False               False           False
4                       False               False           False
…                         …                   …               …
7496                    False               False           False
7497                    False               False           False
7498                    False               False           False
7499                    False               False           False
7500                    False               False           False

      iPhone Charger Cable Anker 6ft  iPhone SE case  \
0                             False           False
1                             False           False
2                             False           False
3                             False           False
4                             False           False
…                               …               …
7496                          False           False
7497                          False           False
7498                          False           False
7499                          False           False
```

```
7500                              False          False


      nonda USB C to USB Adapter  seenda Wireless mouse
0                           True                 False
1                          False                 False
2                          False                 False
3                          False                 False
4                          False                 False
…                            …                     …
7496                       False                 False
7497                       False                 False
7498                       False                 False
7499                       False                 False
7500                       False                 False

[7501 rows x 119 columns]
```

[3]:
```python
frequent_itemsets = apriori(encoded_df, min_support = 0.02, use_colnames = True)
frequent_itemsets
```

[3]:
```
      support                                     itemsets
0    0.050527          (10ft iPHone Charger Cable 2 Pack)
1    0.042528            (3A USB Type C Cable 3 pack 6FT)
2    0.029463              (Anker 2-in-1 USB Card Reader)
3    0.068391               (Anker USB C to HDMI Adapter)
4    0.087188      (Apple Lightning to Digital AV Adapter)
..        …                                          …
98   0.023730  (Screen Mom Screen Cleaner kit, USB 2.0 Printe…
99   0.035462  (VIVO Dual LCD Monitor Desk mount, Screen Mom …
100  0.020131      (Stylus Pen for iPad, USB 2.0 Printer cable)
101  0.025197  (VIVO Dual LCD Monitor Desk mount, Stylus Pen …
102  0.027596  (VIVO Dual LCD Monitor Desk mount, USB 2.0 Pri…

[103 rows x 2 columns]
```

7 C2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate that the code is error free.

[4]:
```python
rules = association_rules(frequent_itemsets, metric = "lift", min_threshold = 1.
↪0)
rules
```

[4]:
```
                          antecedents                          consequents  \
0    (10ft iPHone Charger Cable 2 Pack)    (Dust-Off Compressed Gas 2 pack)
1      (Dust-Off Compressed Gas 2 pack)  (10ft iPHone Charger Cable 2 Pack)
```

```
2         (Anker USB C to HDMI Adapter)      (Dust-Off Compressed Gas 2 pack)
3      (Dust-Off Compressed Gas 2 pack)         (Anker USB C to HDMI Adapter)
4   (VIVO Dual LCD Monitor Desk mount)         (Anker USB C to HDMI Adapter)
..                                   …                                     …
89      (Screen Mom Screen Cleaner kit)  (VIVO Dual LCD Monitor Desk mount)
90               (Stylus Pen for iPad)              (USB 2.0 Printer cable)
91              (USB 2.0 Printer cable)               (Stylus Pen for iPad)
92  (VIVO Dual LCD Monitor Desk mount)               (Stylus Pen for iPad)
93               (Stylus Pen for iPad)  (VIVO Dual LCD Monitor Desk mount)

    antecedent support  consequent support   support  confidence      lift  \
0             0.050527            0.238368  0.023064    0.456464  1.914955
1             0.238368            0.050527  0.023064    0.096756  1.914955
2             0.068391            0.238368  0.024397    0.356725  1.496530
3             0.238368            0.068391  0.024397    0.102349  1.496530
4             0.174110            0.068391  0.020931    0.120214  1.757755
..                 …                   …        …          …         …
89            0.129583            0.174110  0.035462    0.273663  1.571779
90            0.095054            0.170911  0.020131    0.211781  1.239135
91            0.170911            0.095054  0.020131    0.117785  1.239135
92            0.174110            0.095054  0.025197    0.144717  1.522468
93            0.095054            0.174110  0.025197    0.265077  1.522468

    leverage  conviction  zhangs_metric
0   0.011020    1.401255       0.503221
1   0.011020    1.051182       0.627330
2   0.008095    1.183991       0.356144
3   0.008095    1.037830       0.435627
4   0.009023    1.058905       0.521973
..       …           …             …
89  0.012900    1.137061       0.417935
90  0.003885    1.051852       0.213256
91  0.003885    1.025766       0.232768
92  0.008647    1.058066       0.415518
93  0.008647    1.123778       0.379218

[94 rows x 10 columns]
```

8 C3. Provide values for the support, lift, and confidence of the association rules table.

```
[5]: rules
```

```
[5]:                          antecedents                           consequents  \
     0   (10ft iPHone Charger Cable 2 Pack)      (Dust-Off Compressed Gas 2 pack)
     1      (Dust-Off Compressed Gas 2 pack)  (10ft iPHone Charger Cable 2 Pack)
```

```
2          (Anker USB C to HDMI Adapter)      (Dust-Off Compressed Gas 2 pack)
3      (Dust-Off Compressed Gas 2 pack)          (Anker USB C to HDMI Adapter)
4    (VIVO Dual LCD Monitor Desk mount)          (Anker USB C to HDMI Adapter)
..                                  …                                        …
89       (Screen Mom Screen Cleaner kit)  (VIVO Dual LCD Monitor Desk mount)
90                  (Stylus Pen for iPad)              (USB 2.0 Printer cable)
91                (USB 2.0 Printer cable)                (Stylus Pen for iPad)
92   (VIVO Dual LCD Monitor Desk mount)                (Stylus Pen for iPad)
93                  (Stylus Pen for iPad)  (VIVO Dual LCD Monitor Desk mount)
```

	antecedent support	consequent support	support	confidence	lift
0	0.050527	0.238368	0.023064	0.456464	1.914955
1	0.238368	0.050527	0.023064	0.096756	1.914955
2	0.068391	0.238368	0.024397	0.356725	1.496530
3	0.238368	0.068391	0.024397	0.102349	1.496530
4	0.174110	0.068391	0.020931	0.120214	1.757755
..	…	…	…	…	…
89	0.129583	0.174110	0.035462	0.273663	1.571779
90	0.095054	0.170911	0.020131	0.211781	1.239135
91	0.170911	0.095054	0.020131	0.117785	1.239135
92	0.174110	0.095054	0.025197	0.144717	1.522468
93	0.095054	0.174110	0.025197	0.265077	1.522468

	leverage	conviction	zhangs_metric
0	0.011020	1.401255	0.503221
1	0.011020	1.051182	0.627330
2	0.008095	1.183991	0.356144
3	0.008095	1.037830	0.435627
4	0.009023	1.058905	0.521973
..	…	…	…
89	0.012900	1.137061	0.417935
90	0.003885	1.051852	0.213256
91	0.003885	1.025766	0.232768
92	0.008647	1.058066	0.415518
93	0.008647	1.123778	0.379218

```
[94 rows x 10 columns]
```

9 C4. Explain the top three relevant rules generated by the Apriori algorithm. Include a screenshot of the top three relevant rules.

```python
top_3_rules = rules.sort_values(by="lift", ascending=False).head(3)
top_3_rules
```

```
[6]:                             antecedents                              consequents  \
    84        (VIVO Dual LCD Monitor Desk mount)            (SanDisk Ultra 64GB card)
    85                 (SanDisk Ultra 64GB card)   (VIVO Dual LCD Monitor Desk mount)
    65   (FEIYOLD Blue light Blocking Glasses)    (VIVO Dual LCD Monitor Desk mount)

        antecedent support   consequent support    support   confidence      lift  \
    84             0.174110             0.098254   0.039195     0.225115  2.291162
    85             0.098254             0.174110   0.039195     0.398915  2.291162
    65             0.065858             0.174110   0.022930     0.348178  1.999758


        leverage   conviction   zhangs_metric
    84  0.022088     1.163716        0.682343
    85  0.022088     1.373997        0.624943
    65  0.011464     1.267048        0.535186
```

Firstly, we have a strong association between the VIVO Dual LCD Monitor Desk Mount and the SanDisk Ultra 64GB card. This rule indicates that customers who purchase the desk mount have a 22.5% probability of also buying the memory card, with a lift value of 2.29, suggesting a positive relationship that's over twice as likely compared to random chance.

Secondly, the relationship flips in the next rule, showing that purchasers of the SanDisk Ultra 64GB card also often buy the VIVO Dual LCD Monitor Desk Mount. This rule has an even higher confidence at 39.89%, underlining a stronger likelihood of the desk mount being purchased alongside the memory card. The lift remains the same at 2.29, reinforcing the strength of this association.

Lastly, the association between the FEIYOLD Blue light Blocking Glasses and the VIVO Dual LCD Monitor Desk Mount highlights a significant crossover in the customer base interested in office ergonomics and eye protection. With a confidence of 34.82% and a lift of nearly 2.00, customers who consider blue light glasses are quite likely to consider enhancing their workstation setup with the desk mount, indicating a strategic opportunity to bundle these products for promotions.

10 D1. Summarize your data analysis by doing the following: Summarize the significance of support, lift, and confidence from the results of the analysis.

Support: The support metric tells us how often an itemset appears in the dataset. For instance, the rule involving the VIVO Dual LCD Monitor Desk Mount and SanDisk Ultra 64GB card has a support of 0.039195, indicating these items are purchased together in approximately 3.92% of all transactions. This suggests a modestly frequent occurrence worth noting for marketing strategies.

Confidence: Confidence measures the likelihood that the consequent is bought when the antecedent is purchased. The confidence of 22.51% in the first rule means that there is a 22.51% chance of selling a SanDisk Ultra 64GB card when a VIVO Dual LCD Monitor Desk Mount is purchased. The second rule shows a higher confidence of 39.89%, indicating a stronger likelihood of the desk mount being purchased when customers buy the memory card.

Lift: Lift indicates the strength of the association between items beyond coincidence. A lift value greater than 1, such as 2.29 in these rules, suggests a positive association. This high lift value

reveals that customers are more than twice as likely to purchase these items together compared to buying them independently.

11 D2. Discuss the practical significance of your findings from the analysis.

The observed relationships suggest that customers interested in office or tech products like monitor mounts and memory cards exhibit specific bundled purchasing behaviors.

This insight allows for strategic product placement and promotional bundling, which could enhance the shopping experience and increase sales. Furthermore, the association between FEIYOLD Blue light Blocking Glasses and the VIVO Dual LCD Monitor Desk Mount with a lift of nearly 2.00 indicates a significant cross-selling opportunity, especially among customers looking to optimize their workstation setups.

12 D3. Recommend a course of action for the real-world organizational situation from part A1 based on the results from part D1.

Given these insights, I recommend the following actions for the telecommunications company:

Bundle the VIVO Dual LCD Monitor Desk Mount with SanDisk Ultra 64GB cards and possibly include FEIYOLD Blue light Blocking Glasses in a special offer. This could attract more customers looking to purchase comprehensive workstation setups.

Use the data from these rules to target advertisements more precisely. Customers buying any of these products could be shown ads for the other items, potentially in real-time as they shop online or through follow-up marketing campaigns.

Optimize in person and e-commerce store layouts. In physical stores, placing these items near each other can encourage impulse buys, capitalizing on the revealed purchasing patterns.

13 F. Web Sources

TransactionEncoder docs - https://rasbt.github.io/mlxtend/user_guide/preprocessing/TransactionEncoder/

14 G. Works Consulted