Teleco Association Rules and Lift Analysis

Javier Lopez

D212 - Data Mining II

Table of Contents

Part I: Research Question

A1. Research Question

What item pairings can Teleco leverage when recommending add-ons on their e-commerce site and can we identify any patterns or associations between these items?

A2. Objective

Our goal is to analyze Teleco's e-commerce transactions using MLXtend's Apriori and Association Rules algorithms to identify item associations and frequent itemsets that customers tend to purchase together. We expect to use the confidence and lift scores to understand the strength of the resulting rules from our analysis. Once identified, Teleco can use the item airing to design targeted marketing campaigns and create special offers or bundles.

Part 2: Market Basket Justification

B1. Market Basket Analysis

Market basket analysis is a data mining technique that analyzes transaction data to find associations between items. The analysis identifies frequent itemsets and creates association rules that show which items are often purchased together. The expected outcomes from the market basket analysis include discovering meaningful relationships and patterns between items that can be leveraged to improve marketing strategies, cross-selling, and upselling opportunities.

B2. Transaction Example

Below is an example of transactions in the Teleco transaction dataset.

	Transaction 4	Transaction 5	Transaction 6	Transaction 7
Item 1	Apple USB-C Charger cable	TopMate C5 Laptop Cooler pad	Anker 2-in-1 USB Card Reader	10ft iPHone Charger Cable 2 Pack
Item 2	Screen Mom Screen Cleaner kit		USB 2.0 Printer cable	Apple Lightning to USB cable
Item 3	Moread HDMI to VGA Adapter			HP952XL ink
Item 4	HP 62XL Tri-Color ink			
Item 5	Dust-Off Compressed Gas 2 pack			

B3. Assumption

One assumption of market basket analysis is that each transaction is independent of the other transactions. This means that each transaction is unique and not influenced by any purchases in another transaction. For example, there should be no transaction within the dataset that contains information dependent on data from another transaction within the dataset.

Part III. Data Preparation and Analysis

C2. Apriori Algorithm Execution

```
## Use apriori to get list of most frequently paired items
frq_items = apriori(df, min_support = 0.05, use_colnames = True)
```

```
## Review list
frq_items
```

	support	itemsets
0	0.079323	(HP 63XL Ink)
1	0.076523	(TopMate C5 Laptop Cooler pad)
2	0.179709	(Apple Pencil)
3	0.238368	(Dust-Off Compressed Gas 2 pack)
4	0.065858	(FEIYOLD Blue light Blocking Glasses)
5	0.071457	(Logitech M510 Wireless mouse)
6	0.050527	(10ft iPHone Charger Cable 2 Pack)
7	0.063325	(SanDisk Ultra 128GB card)
8	0.087188	(Apple Lightning to Digital AV Adapter)
9	0.098254	(SanDisk Ultra 64GB card)
10	0.081056	(Syntech USB C to USB Adapter)
11	0.174110	(VIVO Dual LCD Monitor Desk mount)
12	0.058526	(HP 62XL Tri-Color ink)
13	0.051060	(Premium Nylon USB Cable)
14	0.080389	(USB Type C to USB-A Charger cable)
15	0.062525	(Cat8 Ethernet Cable)
16	0.170911	(USB 2.0 Printer cable)
17	0.059992	(Falcon Dust Off Compressed Gas)
18	0.068391	(Anker USB C to HDMI Adapter)
19	0.095321	(Nylon Braided Lightning to USB cable)
20	0.129583	(Screen Mom Screen Cleaner kit)
21	0.052393	(SAMSUNG EVO 32GB card)
22	0.132116	(Apple USB-C Charger cable)
23	0.163845	(HP 61 ink)
24	0.095054	(Stylus Pen for iPad)
25	0.050927	(Apple Pencil, Dust-Off Compressed Gas 2 pack)
26	0.059725	(Dust-Off Compressed Gas 2 pack, VIVO Dual LCD...
27	0.052660	(Dust-Off Compressed Gas 2 pack, HP 61 ink)

```
## Collect rules and sort by confidence, lift
rules = association_rules(frq_items, metric='lift')
rules = rules.sort_values(['lift', 'support', 'confidence'], ascending =[False, False, False]).reset_index(drop=True)
## Display top 3 rules
rules
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314
1	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357
3	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256
4	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.179709	0.238368	0.050927	0.283383	1.188845	0.008090	1.062815
5	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.238368	0.179709	0.050927	0.213647	1.188845	0.008090	1.043158

C3. Support, Lift, and Confidence Values

Below is a list of association rules and their respective support, lift, and confidence values.

	antecedents	consequents	support	lift	confidence
0	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.059725	1.439085	0.343032
1	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.059725	1.439085	0.250559
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.052660	1.348332	0.321400
3	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.052660	1.348332	0.220917
4	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.050927	1.188845	0.283383
5	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.050927	1.188845	0.213647

C4. Top Three Association Rules

Below is a list of the top three association rules and their respective summaries.

	antecedents	consequents	support	lift	confidence
0	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.059725	1.439085	0.343032
1	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.059725	1.439085	0.250559
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.052660	1.348332	0.321400

Part IV. Data Summary and Implications

D1. Significance of Results

The support, lift, and confidence values are important metrics in market basket analysis that help to determine the strength and significance of the association rules. The support measures the frequency in which a particular itemset, or rule, occurs within the dataset. The higher the support value, the more frequent that the itemset is observed in the dataset and can be considered more significant. The confidence measures the likelihood that a consequent item is purchased with the antecedent item. The higher the confidence value, the stronger the association between the items in the rule. A higher confidence can also imply that the rule is likely to occur more often.

The lift is the ratio of the observed confidence to the expected confidence if the antecedent and consequent items were independent. Lift values greater than one indicate that a positive association exists between the items in the rule, meaning the items are more likely to be purchased together. Lift values less than one indicate that a negative association exists between the items in the rule, meaning that if one item is purchased, the other is likely to not be purchased with it. A lift value close to one implies that the two items are independent of each other.

Regarding our analysis of the top three rules from the Apriori algorithm, we found that VIVO Dual LCD Monitor Desk Mount and Dust-Off Compressed Gas 2 pack were sold the most together with a support value of 0.059, or 5.97% of the transactions. We can also see that the lift is well over one at a 1.43, meaning both items have a pretty high correlation, further corroborating their likelihood of being purchased in the same transaction. However, the item pairing has a higher confidence of 0.34 when the antecedent occurs before the consequent,

meaning that there is a higher likelihood that the Dust-Off Compressed Gas 2 pack gets added on to the VIVO Dual LCD Monitor Desk Mount than the other way around, which has a confidence value of 0.25.

D2. Practical Significance

Our findings show that the itemsets most likely to be purchased together are as follows.

VIVO Dual LCD Monitor Desk Mount	Dust-Off Compressed Gas 2 pack
HP 61 ink	Dust-Off Compressed Gas 2 pack
Apple Pencil	Dust-Off Compressed Gas 2 pack

The practical significance of our findings from the market basket analysis has several implications in a real-world situation. Based on our findings, we can conclude that the Dust-Off Compressed Gas 2 pack is the most popular add-on amongst all of Teleco's items. When putting these findings into practice, Teleco can adjust their product placement to ensure that the add-on does not get missed by the customer. Additionally, Teleco can also ensure that when managing their inventory that they always have the compressed gas available. Another approach that Teleco can take with the findings is creating product bundles where they offer a discount when customers purchase both items together.

D3. Recommendation

The research question for our real-world organizational situation asked what item pairings Teleco could leverage when recommending add-ons on their e-commerce site and sought to identify any patterns or associations between the items. We discovered that the Dust-Off Compressed Gas 2 pack was the most frequently purchased item with the VIVO Dual LCD Monitor Desk Mount, HP 61 ink, and Apple Pencil. Based on these findings, we

recommend that Teleco add a message in the user's cart before checkout offering the Dust-Off Compressed Gas 2 pack whenever any of the three items appear in the cart. Additionally, we recommend that Teleco revise their inventory planning to ensure that they always have the Dust-Off Compressed Gas 2 pack on-hand to avoid missed cross-selling opportunities.

Part V. Attachments

F. Web Sources

No third-party code was used in the completion of this project.

G. Sources

No in-text citations or references were used in the completion of this project.