PERFORMANCE ASSESSMENT

Task 3 | Association Rules & Lift Analysis

KAILI HAMILTON

Masters of Science in Data Analytics, Western Governors University

Course: D212

Instructor: Dr. Kesselly Kamara

Program Mentor: Krissy Bryant

February, 2024

TABLE OF CONTENTS

A1

One research question for this performance assessment is, “How can we use market basket analysis to find out what items are frequently purchased together?”

A2

One goal of this data analysis is to use market basket analysis to find out associations between items, that is, what items are most likely to be purchased together?

B1

Market Basket Analysis is centered around the identification and analysis of rules, or associations. “Market Basket Analysis (MBA) is a data mining method used to define the strength of relationships between pairs of items bought together. The discovery of association between pairs of items bought together can uncover patterns. Recurrent itemset mining leads to the finding of associations and links between items in large transaction datasets. MBA captures data during transaction and lists all purchased items by a single buyer. MBA determines a relationship of what items were bought with another item(s). The collected transactions records are used to create purchasing profiles containing IF=THEN rules of purchased items. The IF part of the rule is known as the antecedent and the THEN part of the rule is known as the consequent” (Kamara, n.d.).

“MBA relies on Association Rules to identify relationships between items from a logged dataset. The goal of association analysis is to identify relationships between products by considering co-occurences of purchases in past transactions. Association analysis is an unsupervised machine learning method where there is no target variable to predict. The algorighm reviews every sisngle transaction containing serveral items and extract useful relationship patterns between the items in a form of rules. [The] Apriori Algorithm provides efficient methods to extract rules from large datasets in logged transactions” (Kamara, n.d.).

Expected outcomes of MBA include using the Apriori Algorithm to generate association rules and identify the top three of these rules. We also expect to calculate metrics to indicate the strength of these associates, which are support, confidence, and lift.

B2

One example of transactions in the dataset can be seen in the screenshot. Each row of the dataset is a transaction for a single customer. Take, for example, row 3. This customer purchased three items: Apple Lighting to Digital AV Adapter, TP-Link AC1750 Smart WiFi Router, and Apple Pencil.

	Item01	Item02	Item03	Item04	Item05	Item06
0	NaN	NaN	NaN	NaN	NaN	NaN
1	Logitech M510 Wireless mouse	HP 63 Ink	HP 65 ink	nonda USB C to USB Adapter	10ft iPHone Charger Cable	HP 902XL ink
2	NaN	NaN	NaN	NaN	NaN	NaN
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN

B3

One assumption of MBA is called the “Apriori principle” which states that subsets of frequent sets are frequent. This means that the algorithms retains sets known to be frequent and prune sets not known to be frequent (Hull, n.d.).

C1

The dataset has been transformed and made suitable for market basket analysis. Data preparation and transformation included removing blank rows and blank columns, turning the dataframe into a list of lists, feed the list to TransactionEncoder to create a NumPy Boolean array, convert this array back to a DataFrame, and remove blank columns again (Kamara, n.d.).

A copy of the cleaned data set is included in the submission of this performance assessment.

C2

Once the dataset has been transformed, association rules were generated with the Apriori Algorithm. Screenshots are included here.

Apriori Algorithm

[16]:
```
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules
```

[18]:
```
rules_df = apriori(df, min_support=0.02, use_colnames = True)
rules_df.head(5)
```

[18]:

	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)

[19]:
```
rule_table = association_rules(rules_df, metric = 'lift', min_threshold = 1)
rule_table.head()
```

[19]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
1	(Dust-Off Compressed Gas 2 pack)	(10ft iPHone Charger Cable 2 Pack)	0.238368	0.050527	0.023064	0.096756	1.914955	0.011020	1.051182	0.627330
2	(Anker USB C to HDMI Adapter)	(Dust-Off Compressed Gas 2 pack)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144
3	(Dust-Off Compressed Gas 2 pack)	(Anker USB C to HDMI Adapter)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
4	(Anker USB C to HDMI Adapter)	(VIVO Dual LCD Monitor Desk mount)	0.068391	0.174110	0.020931	0.306043	1.757755	0.009023	1.190117	0.462740

[22]:
```
top_three_rules = rule_table.sort_values('confidence', ascending=False).head(3)
top_three_rules
```

[22]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
36	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.065725	0.238368	0.027463	0.417850	1.752960	0.011796	1.308308	0.459753
52	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.098254	0.238368	0.040928	0.416554	1.747522	0.017507	1.305401	0.474369

[23]:
```
top_three_rules = rule_table.sort_values('lift', ascending=False).head(3)
top_three_rules
```

[23]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
85	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	0.174110	0.098254	0.039195	0.225115	2.291162	0.022088	1.163716	0.682343
84	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	0.098254	0.174110	0.039195	0.398915	2.291162	0.022088	1.373997	0.624943
64	(FEIYOLD Blue light Blocking Glasses)	(VIVO Dual LCD Monitor Desk mount)	0.065725	0.174110	0.022930	0.348884	2.003815	0.011487	1.268423	0.536193

[24]:
```
top_three_rules = rule_table.sort_values('support', ascending=False).head(3)
top_three_rules
```

[24]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
62	(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	0.369437
63	(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	0.400606
41	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256	0.339197

[25]:
```
sorted_rules = rule_table[(rule_table['lift'] > 0.02)]
sorted_rules.head(3)
```

[25]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
1	(Dust-Off Compressed Gas 2 pack)	(10ft iPHone Charger Cable 2 Pack)	0.238368	0.050527	0.023064	0.096756	1.914955	0.011020	1.051182	0.627330
2	(Anker USB C to HDMI Adapter)	(Dust-Off Compressed Gas 2 pack)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144

C3

Values for support, lift, and confidence of the association rules table are displayed in the screenshot, which displays the head of the dataframe.

```
[19]: rule_table = association_rules(rules_df, metric = 'lift', min_threshold = 1)
rule_table.head()
```

[19]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
1	(Dust-Off Compressed Gas 2 pack)	(10ft iPHone Charger Cable 2 Pack)	0.238368	0.050527	0.023064	0.096756	1.914955	0.011020	1.051182	0.627330
2	(Anker USB C to HDMI Adapter)	(Dust-Off Compressed Gas 2 pack)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144
3	(Dust-Off Compressed Gas 2 pack)	(Anker USB C to HDMI Adapter)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
4	(Anker USB C to HDMI Adapter)	(VIVO Dual LCD Monitor Desk mount)	0.068391	0.174110	0.020931	0.306043	1.757755	0.009023	1.190117	0.462740

The top three relevant rules generated by the Apriori Algorithm are explained below.

The top three rules for using the support metric are displayed. Support measures the proportion of transactions in which the rules occur (Hull, n.d.). The support for the top three rules ranges from approximately 5-6%.

```
[24]: top_three_rules = rule_table.sort_values('support', ascending=False).head(3)
top_three_rules
```

[24]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
62	(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	0.369437
63	(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	0.400606
41	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256	0.339197

Next, the top three rules for using the confidence metric are displayed. Confidence measures the probability that a customer purchases the consequent item given that they purchased the antecedent item (Hull, n.d.). The confidence of the top three rules ranges from approximately 41-46%.

```
[22]: top_three_rules = rule_table.sort_values('confidence', ascending=False).head(3)
top_three_rules
```

[22]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
36	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.065725	0.238368	0.027463	0.417850	1.752960	0.011796	1.308308	0.459753
52	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.098254	0.238368	0.040928	0.416554	1.747522	0.017507	1.305401	0.474369

Lastly, lift measures the strength of the association between items. This ratio is found by obtaining the proportion of transactions that contain both the antecedent and consequent and dividing it by the proportion of transactions if the antecedent and consequent were assigned randomly and independently. So, when the lift metric is greater than one, this indicates that the association is strong (Hull, n.d.). The lift metric for the top three rules is between 2 and 2.3.

```
[23]: top_three_rules = rule_table.sort_values('lift', ascending=False).head(3)
top_three_rules
```

[23]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
85	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	0.174110	0.098254	0.039195	0.225115	2.291162	0.022088	1.163716	0.682343
84	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	0.098254	0.174110	0.039195	0.398915	2.291162	0.022088	1.373997	0.624943
64	(FEIYOLD Blue light Blocking Glasses)	(VIVO Dual LCD Monitor Desk mount)	0.065725	0.174110	0.022930	0.348884	2.003815	0.011487	1.268423	0.536193

C4

The top three relevant rules generated by the Apriori Algorithm are explained below.

The top three rules for using the support metric are displayed. Support measures the proportion of transactions in which the rules occur (Hull, n.d.). The support for the top three rules ranges from approximately 5-6%.

```
[24]: top_three_rules = rule_table.sort_values('support', ascending=False).head(3)
      top_three_rules
```

[24]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
62	(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	0.369437
63	(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	0.400606
41	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256	0.339197

Next, the top three rules for using the confidence metric are displayed. Confidence measures the probability that a customer purchases the consequent item given that they purchased the antecedent item (Hull, n.d.). The confidence of the top three rules ranges from approximately 41-46%.

```
[22]: top_three_rules = rule_table.sort_values('confidence', ascending=False).head(3)
      top_three_rules
```

[22]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
36	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.065725	0.238368	0.027463	0.417850	1.752960	0.011796	1.308308	0.459753
52	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.098254	0.238368	0.040928	0.416554	1.747522	0.017507	1.305401	0.474369

Lift measures the strength of the association between items. This ratio is found by obtaining the proportion of transactions that contain both the antecedent and consequent and dividing it by the proportion of transactions if the antecedent and consequent were assigned randomly and independently. So, when the lift metric is greater than one, this indicates that the association is strong (Hull, n.d.). The lift metric for the top three rules is between 2 and 2.3.

```
[23]: top_three_rules = rule_table.sort_values('lift', ascending=False).head(3)
      top_three_rules
```

[23]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
85	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	0.174110	0.098254	0.039195	0.225115	2.291162	0.022088	1.163716	0.682343
84	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	0.098254	0.174110	0.039195	0.398915	2.291162	0.022088	1.373997	0.624943
64	(FEIYOLD Blue light Blocking Glasses)	(VIVO Dual LCD Monitor Desk mount)	0.065725	0.174110	0.022930	0.348884	2.003815	0.011487	1.268423	0.536193

D1

The significance of the support, confidence, and lift metrics is that they help us to determine if the association rules are valuable and significant. Support identifies frequent itemsets, confidence measures the strength of association between items, and lift helps distinguish between spurious and meaningful associations between items in market basket analysis.

Support measures the proportion of transactions in which the rules occur (Hull, n.d.). Based on this metric, the top three rules state:

1. If (VIVO Dual LCD Monitor Desk mount) then (Dust-Off Compressed Gas 2 pack)
2. If (Dust-Off Compressed Gas 2 pack) then (VIVO Dual LCD Monitor Desk mount)
3. If (Dust-Off Compressed Gas 2 pack) then (HP 61 ink)

Confidence measures the probability that a customer purchases the consequent item given that they purchased the antecedent item (Hull, n.d.). Based on this metric, the top three rules state:

1. If (10ft iPhone Charger Cable 2 Pack) then (Dust-Off Compressed Gas 2 pack)
2. If (FEIYOLD Blue light Blocking Glasses) then (Dust-Off Compressed Gas 2 pack)
3. If (SanDisk Ultra 64GB card) then (Dust-Off Compressed Gas 2 pack)

Lift measures the strength of the association between items. This ratio is found by obtaining the proportion of transactions that contain both the antecedent and consequent and dividing it by the proportion of transactions if the antecedent and consequent were assigned randomly and independently. So, when the lift metric is greater than one, this indicates that the association is strong (Hull, n.d.). Based on this metric, the top three rules state:

1. If (VIVO Dual LCD Monitor Desk mount) then (SanDisk Ultra 64GB card)
2. If (SanDisk Ultra 64GB card) then (VIVO Dual LCD Monitor Desk mount)
3. If (FEIYOLD Blue light Blocking Glasses) then (VIVO Dual LCD Monitor Desk mount)

D2

The practical significance of the findings from this analysis help us to find actionable insights for this business. The high support itemsets are the most frequently occuring. This information can be used to optimize product placement and promotional strategies. The high confidence itemsets have a strong association, that when one item is purchased in the set, the other item is highly likely to be purchased. This information can be used to make personalized recommendations and cross-selling. The high lift itemsets evaluate the importance of the association between items while considering their individual popularity. This information can be used to drive marketing strategies such as bundling (Kamara, n.d.).

D3

A course of action recommended by this analysis is as follows:

Place these items next to each other in the physical store: VIVO Dual LCD Monitor Desk mount, Dust-Off Compressed Gas 2 pack, and HP 61 ink.

Additionally, for the online stores, make personalized recommendations when one of the items is added to a customer's cart: 10ft iPhone Charger Cable 2 Pack, Dust-Off Compressed Gas 2 pack, FEIYOLD Blue light Blocking Glasses, SanDisk Ultra 64GB card.

Lastly, conduct a bundling promotion that includes these items: VIVO Dual LCD Monitor Desk mount, SanDisk Ultra 64GB card, FEIYOLD Blue light Blocking Glasses.

E

A Panopto video presentation is included with the submission of this performance assessment.

F

Hull, Isaiah. "Market Basket Analysis in Python". Datacamp, n.d., app.datacamp.com/learn/courses/market-basket-analysis-in-python

Kamara, Kesselly. "D212 Recommended Study Plan". D212 Western Governors University, n.d., srm.file.force.com/servlet/fileField?id=0BE3x000000c9YS.

G

Hull, Isaiah. "Market Basket Analysis in Python". Datacamp, n.d., app.datacamp.com/learn/courses/market-basket-analysis-in-python

Kamara, Kesselly. "D212 Recommended Study Plan". D212 Western Governors University, n.d., srm.file.force.com/servlet/fileField?id=0BE3x000000c9YS.

H

Professional communication is demonstrated throughout the content and presentation of this PA.