D212: Data Mining II Task 3

Justin Huynh

Student ID: 012229514

M.S. Data Analytics

A1. Proposal of Question

My research question for this project is "What are the common product combinations purchased by customers?" This question is focused on identifying frequent product sets in customer transactions. We will use the market basket analysis to answer our research question.

A2. Defined Goal

The goal of the data analysis is to identify frequently co-purchased items within the telecom retail sector using market basket analysis. By understanding these findings, an organization can create more effective cross-selling and bundling strategies, targeting customers with product combinations that are likely to appeal to them based on their past purchase behavior. This could potentially lead to increased revenue and improve repeat purchase rate.

B1. Explanation of Market Basket

The Market Basket Analysis (MBA) is a technique used to identify associations between items purchased together within a set of transactions. It analyzes the data by examining how frequently certain items are bought together, which helps identify common item sets (i.e. products that tend to co-occur in transactions).

In the context of the provided dataset, MBA will scan through the rows of transactions and identify item combinations that appear together more frequently than expected by chance. These combinations are often called "frequent item sets." The analysis will also calculate the following key metrics:

- Support: The proportion of transactions that contain a specific item or combination of items.

- Confidence: The likelihood that a transaction containing one item will also contain another item.
- Lift: A measure of how much more likely two items are to be purchased together than if they were independent.

Some expected outcomes are identification of frequent item sets (items that are often bought together), insights into customer purchasing behavior that can guide marketing and sales strategies such as product bundling or targeted promotions, and the creation of association rules that highlight potential cross-selling opportunities (e.g. if a customer buys item A, they are likely to buy item B as well).

B2: Transaction Example

Here's an example of what a transaction may look like in the data set:

- Transaction ID: 001
    - CustomerID: 123
    - Items Purchased:
        - Macbook Pro 13-inch
        - Airpods Pro (2nd Generation)
        - Logitech Wireless Keyboard

In this example, the transaction includes three items that were purchased together in a single transaction. MBA would look for other transactions where similar items are bought together to determine if this combination is common.

B3: Market Basket Assumption

One assumption of Market Basket Analysis is transaction independence. This means that the analysis assumes that the purchase of one item does not affect the

likelihood of purchasing another item, except in the context of the association rules. For example, MBA assumes that if "Macbook Pro 13-inch" and "Airpods Pro (2nd Generation)" are frequently bought together, this pattern reflects a genuine association rather than random chance or external factors.

C1: Transforming the Data Set

A copy of the fully prepared data set will be submitted as 'prepared_data_d212_task3.csv'.

C2: Code Execution

```
# generate frequent itemsets using Apriori
frequent_itemsets = apriori(basket_df, min_support=0.02, use_colnames=True)
```

```
# generate association rules from the frequent itemsets
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
```

```
# display the generated rules
rules.head()
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.089855	0.119184	0.025463	0.283383	2.377689	0.014754	1.229130	0.636628
1	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.119184	0.089855	0.025463	0.213647	2.377689	0.014754	1.157425	0.657826
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.081922	0.119184	0.026330	0.321400	2.696664	0.016566	1.297989	0.685314
3	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.119184	0.081922	0.026330	0.220917	2.696664	0.016566	1.178408	0.714305
4	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.049127	0.119184	0.020464	0.416554	3.495043	0.014609	1.509677	0.750763

See code attached in WGU_D212_Task_3.ipynb.

C3: Association Rules Table

	antecedents	consequents	support	confidence	lift
5	(Dust-Off Compressed Gas 2 pack)	(SanDisk Ultra 64GB card)	0.020464	0.171700	3.495043
4	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.020464	0.416554	3.495043
6	(Screen Mom Screen Cleaner kit)	(Dust-Off Compressed Gas 2 pack)	0.023997	0.370370	3.107548
7	(Dust-Off Compressed Gas 2 pack)	(Screen Mom Screen Cleaner kit)	0.023997	0.201342	3.107548
8	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.029863	0.343032	2.878170
9	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.029863	0.250559	2.878170
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.026330	0.321400	2.696664
3	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.026330	0.220917	2.696664
0	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.025463	0.283383	2.377689
1	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.025463	0.213647	2.377689

See code attached in WGU_D212_Task_3.ipynb.

C4: Top Three Rules

The top three rules in the associated rules table can be seen here, having a lift of over 2.8 and confidence of 0.34 or higher. Lift > 2.8 and Confidence > 0.34 are chosen as practical thresholds to ensure that the rules identified are both strong in terms of association and reliable in terms of prediction.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
4	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.049127	0.119184	0.020464	0.416554	3.495043	0.014609	1.509677	0.750763
6	(Screen Mom Screen Cleaner kit)	(Dust-Off Compressed Gas 2 pack)	0.064791	0.119184	0.023997	0.370370	3.107548	0.016275	1.398943	0.725189
8	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.087055	0.119184	0.029863	0.343032	2.878170	0.019487	1.340729	0.714782

The lift values in these rules indicate a strong association between the antecedents and the consequent, suggesting that the consequent items are significantly more likely to be purchased when the antecedent is purchased. If a customer buys a SanDisk Ultra 64GB card, they are likely to also buy a Dust-Off Compressed Gas 2

pack. The lift value of 3.49 indicates that this combination is 3.5 times more likely to occur together than randomly.

The confidence score shows the likelihood that the consequent is purchased when the antecedent is purchased. A confidence of over 0.34 (34%) shows a relatively strong association, making these rules actionable for cross-selling opportunities. Customers who bought a Screen Mom Screen Cleaner kit are often also purchasing a Dust-Off Compressed Gas 2 pack. This rule shows a strong association with a lift of 3.10.

The leverage represents the difference between the observed frequency of these pairs and what would be expected if the items were independent. Conviction indicates the reliability of these rules, with values greater than 1 showing that the rules are statistically significant. The purchase of a VIVO Dual LCD Monitor Desk mount is associated with buying a Dust-Off Compressed Gas 2 pack, with a lift of 2.87.

See code attached in WGU_D212_Task_3.ipynb.

D1: Significance of Support, Lift, and Confidence Summary

In the context of this analysis, support helps identify item pairs that are commonly purchased together. A higher support value means that the itemset is more prevalent across all transactions, indicating that it is a popular combination among customers. The rules with higher support are more relevant to a larger portion of the customer base.

The significance of the lift is crucial for identifying strong associations between items that are not just popular on their own but are particularly likely to be purchased together. In this analysis, the lift values of over 2.8 suggest that the identified item pairs

are significantly more likely to be bought together than by random chance, making them valuable for targeted promotions or bundling strategies.

Confidence helps determine the reliability of the association rule. Higher confidence values indicate that the presence of the antecedent strongly predicts the presence of the consequent. In this analysis, confidence values above 0.34 suggest a moderate to high likelihood that when customers buy one item, they will also buy the associated item, making these rules useful for predicting customer behavior.

D2: Practical Significance of Findings

The analysis has identified several reliable association rules that provide insights into customer purchasing behavior. For example, the strong association between the SanDisk Ultra 64GB card and the Dust-Off Compressed Gas 2 pack suggests that customers who purchase storage solutions are also concerned with maintaining their electronic devices, which leads them to buy cleaning supplies like compressed gas. Similarly, the association between the Screen Mom Screen Cleaner kit and the Dust-Off Compressed Gas 2 pack indicates that customers interested in screen cleaning products are also likely to purchase general device cleaning tools.

These findings are practically significant as they can inform strategic marketing and sales decisions. By understanding these associations, a company can take actions such as cross-selling certain products based on customer purchase behavior, or bundle offers that include commonly purchased-together items to potentially increase the overall transaction value.

D3: Course of Action

Based on the analysis results, a company can take the following recommended course of action:

- Implement Cross-Selling Strategies
    - Use the identified rules to suggest complementary products to customers during the online checkout process or at the point of sale in physical stores. For example, when a customer adds a SanDisk Ultra 64GB card to their cart, recommend the Dust-Off Compressed Gas 2 pack as an add-on purchase.
- Promotional Bundles
    - Create special offers or discounted bundles that include the frequently associated items. For example, create a "Device Maintenance Bundle" that includes a Screen Mom Screen Cleaner kit and a Dust-Off Compressed Gas 2 pack. This can increase sales of both items and enhance customer satisfaction by providing a convenient package.
- Optimize Product Placements
    - In physical stores, place the associated items near each other or in dedicated sections, which makes it easier for customers to find and purchase them together. If it's online, make sure that these items appear together in recommendation sections or collections.

E. Panopto Video of Code

The URL link will be submitted in the PA task submission.

E1. Panopto Video of Programs

The URL link will be submitted in the PA task submission.

F: Sources for Third-Party Code

1. “Mlxtend 0.1.7” Retrieved from https://pypi.org/project/mlxtend/0.1.7/
2. “Association Rules with Python” Retrieved from https://www.kaggle.com/code/mervetorkan/association-rules-with-python

G: Sources

1. “Association Rules with Python” Retrieved from https://medium.com/@mervetorkan/association-rules-with-python-9158974e761a
2. “A Gentle Introduction on Market Basket Analysis - Association Rules” Retrieved from https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce