

Part I: Research Question

A. Describe the purpose of this data mining report by doing the following:

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

Which products should be grouped together to encourage sales of multiple items?

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

One goal of the analysis is to determine which products are frequently bought together.

Part II: Market Basket Justification

B. Explain the reasons for using market basket analysis by doing the following:

- 1. Explain how market basket analyzes the selected dataset. Include expected outcomes.**

Market basket analysis uses transaction data to determine items that are frequently bought together, or frequent itemsets. It then delivers rules about which items are associated with each other, called association rules. Expected outcomes would be lists of both the frequent itemsets and association rules. (Kadlaskar, 2021)

- 2. Provide one example of transactions in the dataset.**

'Logitech M510 Wireless mouse', 'HP 63 Ink', 'HP 65 ink', 'nonda USB C to USB Adapter' is one transaction in the dataset.

- 3. Summarize one assumption of market basket analysis.**

One assumption of market basket analysis is that the data is fed to an algorithm in the form of a list of lists.

Part III: Data Preparation and Analysis

C. Prepare and perform market basket analysis by doing the following:

- 1. Transform the dataset to make it suitable for market basket analysis. Include a copy of the cleaned dataset.**

The cleaned dataset is provided in the Keim D212 Task Three Cleaned Data document.

2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate the error-free functionality of the code.

```
[99]: # Apply apriori algorithm
association_rules = apriori(transactions, min_support=0.0045, min_confidence=0.2, min_lift=3, min_length=2)
association_results = list(association_rules)

100]: # Call list of rules
for i in range(0, len(association_results)):
    print(association_results[i][0])

FrozenSet({'5pack Nylon Braided USB C cables', 'HP 63XL Ink'})
FrozenSet({'FEIYOLD Blue light Blocking Glasses', 'Anker 2-in-1 USB Card Reader'})
FrozenSet({'Falcon Dust Off Compressed Gas', 'Apple Lightning to USB cable'})
FrozenSet({'AutoFocus 1080p Webcam', 'SanDisk Ultra 64GB card'})
FrozenSet({'HP 63XL Ink', 'iPhone 11 case'})
FrozenSet({'Logitech M510 Wireless mouse', 'iPhone 11 case'})
FrozenSet({'SanDisk Ultra 64GB card', 'SanDisk 128GB Ultra microSDXC card'})
FrozenSet({'FEIYOLD Blue light Blocking Glasses', 'Dust-Off Compressed Gas 2 pack', '10ft iPhone Charger Cable 2 Pack'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'Nylon Braided Lightning to USB cable', 'Anker USB C to HDMI Adapter'})
FrozenSet({'Logitech M510 Wireless mouse', 'Nylon Braided Lightning to USB cable', 'Dust-Off Compressed Gas 2 pack'})
FrozenSet({'SanDisk Ultra 64GB card', 'Dust-Off Compressed Gas 2 pack', 'SanDisk 128GB Ultra microSDXC card'})
FrozenSet({'FEIYOLD Blue light Blocking Glasses', 'Nylon Braided Lightning to USB cable', 'Screen Mom Screen Cleaner kit'})
FrozenSet({'FEIYOLD Blue light Blocking Glasses', 'Nylon Braided Lightning to USB cable', 'VIVO Dual LCD Monitor Desk mount'})
FrozenSet({'FEIYOLD Blue light Blocking Glasses', 'SanDisk Ultra 64GB card', 'Screen Mom Screen Cleaner kit'})
FrozenSet({'FEIYOLD Blue light Blocking Glasses', 'Screen Mom Screen Cleaner kit', 'VIVO Dual LCD Monitor Desk mount'})
FrozenSet({'FEIYOLD Blue light Blocking Glasses', 'VIVO Dual LCD Monitor Desk mount', 'Stylus Pen For iPad'})
FrozenSet({'Nylon Braided Lightning to USB cable', 'Logitech M510 Wireless mouse', 'HP 61 Ink'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'Nylon Braided Lightning to USB cable', 'Logitech M510 Wireless mouse'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'SanDisk Ultra 64GB card', 'Logitech M510 Wireless mouse'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'SanDisk Ultra 64GB card', 'Nylon Braided Lightning to USB cable'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'SanDisk Ultra 64GB card', 'Premium Nylon USB Cable'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'SanDisk Ultra 64GB card', 'SAMSUNG EVO 32GB card'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'SanDisk Ultra 64GB card', 'SanDisk 128GB Ultra microSDXC card'})
FrozenSet({'VIVO Dual LCD Monitor Desk mount', 'Nylon Braided Lightning to USB cable', 'Dust-Off Compressed Gas 2 pack', 'Screen Mom Screen Cleaner kit'})
```

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

```
[101]: # Call stats for rules
for item in association_results:
    # first index of the inner list
    # Contains base item and add item
    pair = item[0]
    items = [x for x in pair]
    print("Rule: " + items[0] + " -> " + items[1])
    # second index of the inner list
    print("Support: " + str(item[1]))
    # third index of the list located at 0th position
    # of the third index of the inner list
    print("Confidence: " + str(item[2][0][2]))
    print("Lift: " + str(item[2][0][3]))
    print("-----")

Rule: 5pack Nylon Braided USB C cables -> HP 63XL Ink
Support: 0.005732568990801226
Confidence: 0.3006993006993007
Lift: 3.790832696715049
-----
Rule: FEIYOLD Blue light Blocking Glasses -> Anker 2-in-1 USB Card Reader
Support: 0.007998933475536596
Confidence: 0.2714932126696833
Lift: 4.122410097642296
-----
Rule: Falcon Dust Off Compressed Gas -> Apple Lightning to USB cable
Support: 0.004532728969470737
Confidence: 0.29059829059829057
Lift: 4.84395061728395
-----
Rule: AutoFocus 1080p Webcam -> SanDisk Ultra 64GB card
Support: 0.005332622317024397
Confidence: 0.3773584905660377
Lift: 3.840659481324083
-----
Rule: HP 63XL Ink -> iPhone 11 case
Support: 0.005865884548726837
Confidence: 0.3728813559322034
Lift: 4.700811850163794
-----
Rule: Logitech M510 Wireless mouse -> iPhone 11 case
Support: 0.005065991201173177
Confidence: 0.3220338983050847
Lift: 4.506672147735896
-----
Rule: SanDisk Ultra 64GB card -> SanDisk 128GB Ultra microSDXC card
Support: 0.015997866951073192
Confidence: 0.3234501347708895
Lift: 3.2919938411349285
-----
```

Rule: SanDisk Ultra 64GB card -> Dust-Off Compressed Gas 2 pack
Support: 0.006665777896280496
Confidence: 0.3906250000000006
Lift: 3.975682666214383

Rule: FEIYOLD Blue light Blocking Glasses -> Nylon Braided Lightning to USB cable
Support: 0.004799360085321957
Confidence: 0.4235294117647058
Lift: 3.2684095860566447

Rule: FEIYOLD Blue light Blocking Glasses -> Nylon Braided Lightning to USB cable
Support: 0.005732568990801226
Confidence: 0.20574162679425836
Lift: 3.1240241752707125

Rule: FEIYOLD Blue light Blocking Glasses -> SanDisk Ultra 64GB card
Support: 0.004932675643247567
Confidence: 0.22424242424242427
Lift: 3.40494417862839

Rule: FEIYOLD Blue light Blocking Glasses -> Screen Mom Screen Cleaner kit
Support: 0.007199040127982935
Confidence: 0.20300751879699247
Lift: 3.0825089038385434

Rule: FEIYOLD Blue light Blocking Glasses -> VIVO Dual LCD Monitor Desk mount
Support: 0.005065991201173177
Confidence: 0.20105820105820105
Lift: 3.0529100529100526

Rule: Nylon Braided Lightning to USB cable -> Logitech M510 Wireless mouse
Support: 0.005332622317024397
Confidence: 0.29629629629629634
Lift: 3.1084175084175087

Rule: VIVO Dual LCD Monitor Desk mount -> Nylon Braided Lightning to USB cable
Support: 0.005999200106652446
Confidence: 0.21531100478468898
Lift: 3.0131489680782684

Rule: VIVO Dual LCD Monitor Desk mount -> SanDisk Ultra 64GB card
Support: 0.005999200106652446
Confidence: 0.5232558139534884
Lift: 3.005315360233627

Rule: VIVO Dual LCD Monitor Desk mount -> SanDisk Ultra 64GB card
Support: 0.008665511265164644
Confidence: 0.31100478468899523
Lift: 3.165328208890303

```

Rule: VIVO Dual LCD Monitor Desk mount -> SanDisk Ultra 64GB card
Support: 0.004799360085321957
Confidence: 0.5714285714285714
Lift: 3.2819951870487856
-----
Rule: VIVO Dual LCD Monitor Desk mount -> SanDisk Ultra 64GB card
Support: 0.005332622317024397
Confidence: 0.3225806451612903
Lift: 3.283144395325426
-----
Rule: VIVO Dual LCD Monitor Desk mount -> SanDisk Ultra 64GB card
Support: 0.006399146780429276
Confidence: 0.3934426229508197
Lift: 4.004359721511667
-----
Rule: VIVO Dual LCD Monitor Desk mount -> Nylon Braided Lightning to USB cable
Support: 0.004532728969470737
Confidence: 0.28813559322033894
Lift: 3.0228043143297376
-----

```

4. Identify the top three rules generated by the Apriori algorithm. Include a screenshot of the top rules along with their summaries.

```

[101]: # Call stats for rules
for item in association_results:
    # first index of the inner list
    # Contains base item and add item
    pair = item[0]
    items = [x for x in pair]
    print("Rule: " + items[0] + " -> " + items[1])
    # second index of the inner list
    print("Support: " + str(item[1]))
    # third index of the list located at 0th position
    # of the third index of the inner list
    print("Confidence: " + str(item[2][0][2]))
    print("Lift: " + str(item[2][0][3]))
    print("-----")

Rule: 5pack Nylon Braided USB C cables -> HP 63XL Ink
Support: 0.005732568990801226
Confidence: 0.3006993006993007
Lift: 3.790832696715049
-----
Rule: FEIYOLD Blue light Blocking Glasses -> Anker 2-in-1 USB Card Reader
Support: 0.007998933475536596
Confidence: 0.2714932126696833
Lift: 4.122410097642296
-----
Rule: Falcon Dust Off Compressed Gas -> Apple Lightning to USB cable
Support: 0.004532728969470737
Confidence: 0.29059829059829057
Lift: 4.84395061728395
-----

```

Part IV: Data Summary and Implications

D. Summarize your data analysis by doing the following:

1. Summarize the significance of support, lift, and confidence from the results of the analysis.

Support is the percentage of transactions containing the antecedent. For example, in the first rule above, the support is 0.0057, meaning that 5pack Nylon Braided USB C cables

appear in 0.57% of the transactions in the dataset. Lift is the likelihood that a consequent is purchased with a specific antecedent. The first transaction has a lift of 3.79, meaning that HP 63 XL Ink is 3.79 times more likely to be purchased with the cables as it is by itself. Confidence is a percentage of how many transactions with a specific antecedent contain a specific consequent. The first transaction has a confidence of 0.30, meaning that 30% of the purchases containing the cables also contain the ink.

2. Discuss the practical significance of the findings from the analysis.

The support, lift and confidence can be used to determine how items should be arranged in a physical store, and which items should be recommended based on in-cart items in an online shopping environment.

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

Based on the results of the analysis, I recommend that in a physical store, 5pack Nylon Braided USB C cables and HO 63XL Ink are placed next to each other, FEIYOLD Blue Light Blocking Glasses and Anker 2-in-1 USB Card Reader are placed next to each other, and Falcon Dust Off Compressed Gas and Apple Lightning to USB cable are placed next to each other. In an online store environment, I recommend that HP 63XL Ink is suggested when a customer has 5pack Nylon Braided cables in their shopping cart, Anker 2-in-1 USB Card Reader is suggested to customers with FEIYOLD Blue Light Blocking Glasses in their carts, and Apple Lightning to USB cable is suggested to customers buying Falcon Dust Off Compressed Gas.

Third Party Code Citations: (Kadlaskar, 2021) (Malik, 2021)

Works Cited

Kadlaskar, A. (2021, December 21). *A Comprehensive Guide on Market Basket Analysis*.

Retrieved from analyticsvidhya.com/: <https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-market-basket-analysis/>

Malik, U. (2021, December 21). *Association Rule Mining via Apriori Algorithm in Python*.

Retrieved from Stackabuse.com: <https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/>