|  |
| --- |
| Western Governors University |
| Data Mining II |
| Task 3 D212 |

|  |
| --- |
| Allison Casey  11-30-2024 |

**Part 1: Research Question**

**A1: PROPOSAL OF QUESTION**

What items are likely to be bought together?

**A2: DEFINED GOAL**

One goal of the analysis is to determine customer purchasing patterns to identify what items are likely to be purchased together.

**Part 2: Market Basket Justification**

**B1: EXPLANATION OF MARKET BASKET**

Market basket analysis analyzes the selected data by looking at patterns of items that are bought together and generates rules for the associations. There are three metrics associated with these rules: support, confidence, and lift. Support is the number of transactions divided by the total of transactions made. The confidence metric is the number of combined transactions divided by the individual transactions. The lift is calculated as a combined metric of the confidence percentage divided by the lift percentage. The expected outcome of performing this analysis on the data set would be to have rules generated that can be evaluated based on their metrics to help make business decisions such as product placement and promotions.

**B2: TRANSACTION EXAMPLE**

In row 5 of the original data set a customer purchased three items:

1. Apple Lightning to Digital AV Adapter
2. TP-Link AC1750 Smart WiFi Router
3. Apple Pencil

**B3: MARKET BASKET ASSUMPTION**

One assumption of market basket analysis is that every entry in the data set represents an independent transaction. (Deniran, 2023)

**Part 3: Data Preparation and Analysis**

**C1: TRANSFORMING THE DATA SET**

1. A black screen with white text

   Description automatically generated
2. A computer screen shot of a black background

   Description automatically generated
3. A screenshot of a computer program

   Description automatically generated
4. A screenshot of a computer

   Description automatically generated

Cleaned data set: Prepared\_df\_task3.csv

For full code see D212\_Task3.ipynb

**C2: CODE EXECUTION**

A screenshot of a computer

Description automatically generated

**C3: ASSOCIATION RULES TABLE**

**A screenshot of a computer program

Description automatically generated**

**C4: TOP THREE RULES**

A screenshot of a computer

Description automatically generated

1. Rule 63
   1. Antecedent: VIVO Dual LCD Monitor Desk mount
   2. Consequent: Dust-Off Compressed Gas 2 pack
   3. Support: 0.059725
   4. Confidence: 0.343032
   5. Lift: 1.439085
2. Rule 62
   1. Antecedent: Dust-Off Compressed Gas 2 pack
   2. Consequent: VIVO Dual LCD Monitor Desk mount
   3. Support: 0.059725
   4. Confidence: 0.250559
   5. Lift: 1.439085
3. Rule 41
   1. Antecedent: HP 61 ink
   2. Consequent: Dust-Off Compressed Gas 2 pack
   3. Support: 0.052660
   4. Confidence: 0.321400
   5. Lift: 1.348332

Where the antecedent is the item that leads to buying the consequent, support is the joint probability that the two will be purchased in the same transaction, confidence is the probability that the consequent will be purchased if the antecedent is purchased, and the lift is how much more the two are likely to be bought together versus by themselves.

**Part 4: Data Summary and Implications**

**D1: SIGNIFICANCE OF SUPPORT, LIFT, AND CONFIDENCE SUMMARY**

Support, lift, and confidence are the metrics that help describe the relationship and the strength of the relationship between the antecedents and consequents for each rule. Support is the number of transactions of the two occurring compared to the total number of transactions. For the top three rules this means that:

1. The probability that both the VIVO Dual LCD Monitor Desk mount and the Dust-Off Compressed Gas 2 pack are purchased together is 0.059725
2. The probability that both the Dust-Off Compressed Gas 2 pack and the VIVO Dual LCD Monitor Desk mount pack are purchased together is 0.059725
3. The probability that both the HP 61 ink and the Dust-Off Compressed Gas 2 pack are purchased together is 0.052660

Confidence is the probability of the consequent being purchased if the antecedent is purchased so the individual sale compared to the sale of both together. For the top three rules this means that:

1. If the VIVO Dual LCD Monitor Desk mount is purchased, then the probability that the Dust-Off Compressed Gas 2 pack is also purchased is 0.343032
2. If the Dust-Off Compressed Gas 2 pack is purchased, then the probability that the VIVO Dual LCD Monitor Desk mount is also purchased is 0.250559
3. If the HP 61 ink mount is purchased, then the probability that the Dust-Off Compressed Gas 2 pack is also purchased is 0.321400

Lift is a ratio of these two metrics so the likelihood of the items being purchased together as opposed to by themselves. This is significant because it can help drive decision making based off the rules. For the top three rules this means that:

1. The VIVO Dual LCD Monitor Desk mount increases the likelihood that the Dust-Off Compressed Gas 2 pack is also purchased by 1.439085
2. The Dust-Off Compressed Gas 2 pack mount increases the likelihood that the VIVO Dual LCD Monitor Desk mount is also purchased by 1.439085
3. The HP 61 ink increases the likelihood that the Dust-Off Compressed Gas 2 pack is also purchased by 1.348332

**D2: PRACTICAL SIGNIFICANCE OF FINDINGS**

The practical significance of the findings is that three rules have been identified for items that have a statistically significant chance of being purchased together. This is useful in terms of making business decisions such as deciding where to place these products within the store, in advertisements, or when creating promotions.

**D3: COURSE OF ACTION**

A recommended course of action based off the support, lift, and confidence would be to feature the VIVO Dual LCD Monitor Desk mount next to or near the Dust-Off Compressed Gas 2 pack since there is support metrics for both leading to buying the other. Another product placement that may be beneficial is placing the HP 61 ink near the Dust-Off as the lift indicates a 1.348332 higher likelihood of purchasing Dust-Off if HP 61 ink is purchased.

**Part 5: Attachments**

**E: PANOPTO VIDEO OF CODE**

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=a7d35778-c9b5-47c5-914e-b2390045f737

**Sources**

Contributor, TechTarget. “What Is Market Basket Analysis? Definition from Whatis.Com.” *Search Customer Experience*, TechTarget, 31 Jan. 2023, www.techtarget.com/searchcustomerexperience/definition/market-basket-analysis#:~:text=Market%20basket%20analysis%20is%20a,likely%20to%20be%20purchased%20together. Accessed 30 Nov. 2024.

Deniran, Oluwakemi Helen. “Boosting Sales with Data: The Power of Market Basket Analysis in Retail.” *Medium*, 27 Nov. 2023, medium.com/@chemistry8526/boosting-sales-with-data-the-power-of-market-basket-analysis-in-retail-c79cc10a14df#:~:text=Transaction%20Independence%3A%20It%20is%20assumed,item%20associations%20within%20individual%20transactions. Accessed 30 Nov. 2024.

GeeksforGeeks. “Market Basket Analysis in Data Mining.” *GeeksforGeeks*, 16 Aug. 2024, www.geeksforgeeks.org/market-basket-analysis-in-data-mining/. Accessed 30 Nov. 2024.