

A1. Research Question

Using the provided transactions data set relating to the telecommunications churn data, I will answer the question, **“Can a Market Basket Analysis method be used to identify groups of products frequently purchased together?”**

A2. Analysis Goal

The **goal of this analysis** is to use a **Market Basket Analysis (MBA)** technique to identify patterns in **associated product purchases** for telecommunications customers. Understanding associations between specific products such as how often and likely they are to be purchased together will help **inform stakeholders on future business strategies and decisions, such as product discounts and promotions.**

B1. MBA Technique

The Market Basket Analysis technique will be applied to the transactional churn data set and used for analysis in the following manner:

- The data set will be **prepared** for MBA by transforming the rows of transaction logs into a matrix of items listed as “TRUE” if included in a transaction or “FALSE” if not included. This process is expanded on further in section C1.
- Using the prepared matrix, an **Apriori algorithm** will be conducted, which analyzes **combinations of items** in each transaction and creates **rules** that describe purchasing patterns. These rules are expressed as **conditional If-Then statements**, where the first item or group of items is the **antecedent** and the second item is the **consequent**. For example: “If this phone charger (antecedent) is included in the transaction, then what is the probability the transaction also includes this phone case (consequent)?”
- The following relevant values (among others) are calculated during the MBA method:
 - **Support** represents the relative frequency of an item or group of items within the overall transactions. For example, if the product combination of “HP 64 Ink,” “HOVAMP iPhone Charger”, and “Apple Pencil” was found in 1% of transactions, then the **support value for the rule “HP 64 Ink and HOVAMP iPhone Charger → Apple Pencil” would be 0.01.**
 - **Confidence** represents the relative frequency of an item within transactions that include another item or group of items. For example, if 45% of all transactions that include “HP 64 Ink” and “HOVAMP iPhone Charger” also include “Apple Pencil” then the **confidence value for the rule “HP 64 Ink and HOVAMP iPhone Charger → Apple Pencil” would be 0.45.**

- **Lift** represents the ratio of a rule's observed confidence to its expected confidence. This ultimately measures how much more likely a product is to be purchased based on the other products in the rule. For example, if the rule "HP 64 Ink and HOVAMP iPhone Charger → Apple Pencil" has a lift value of 2.5, this indicates that an Apple Pencil is **2.5 times more likely to be included in a purchase that includes HP 64 Ink and a HOVAMP iPhone Charger than expected.**
- These values will not only be used to compare resulting rules but also as parameters for selecting only the best performing rules. For this analysis, limits will be set at **support = 0.008** and **confidence = 0.4**. Any rule with either value less than its limit will be removed and not included in the final results.
- All rules as identified by the Apriori algorithm that exceed the support and confidence parameters will be **sorted by their lift values**. The **expected outcome** of this analysis is to use this list of rules to identify groups of products frequently purchased together. These results will be used to help inform future business strategies such as product discounts and promotions.

B2. Transactions Example

In the **initial** provided transactions data set, each row describes a group of products that have been purchased together in a telecom transaction. Blank rows are inserted in between each transaction, so the following example from the initial data set shows the 4th row, but **2nd transaction**, listed in the data, which summarizes the **purchase of three items: an Apple Lightning to Digital AV Adapter, a TP-Link AC1750 Smart WiFi Router, and an Apple Pencil.**

	Item01	Item02	Item03
3			
4	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil

After the transformation of the original data as described in section C1, this same information will be presented in the following manner:

	10ft iPhone Charger Cable	10ft iPhone Charger Cable 2 Pack	3 pack Nylon Braided Lightning Cable	3A USB Type C Cable 3 pack 6FT	5pack Nylon Braided USB C cables	Anker 2-in-1 USB Card Reader	Anker 4-port USB hub	Anker USB C to HDMI Adapter	Apple Lightning to Digital AV Adapter	Apple Lightning to USB cable
2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE

Note: this second screenshot only includes a subset of columns, since there are 119 total products, each of which has its own column. The "TRUE" value listed under Apple Lightning to Digital AV Adapter is also present in the TP-Link AC1750 Smart WiFi Router and Apple Pencil columns. This combination of TRUE values represents the same information that the preceding screenshot of the original data represented: **a customer's purchase of the three products listed.**

B3. MBA Assumption

To perform the MBA method, an **assumption** must be made that each of the transactions are made **independently** of each other. For this analysis, this can be assumed to be true, as there is **no indication that any transactions are inherently related**.

This assumption was clarified at the following source:

<https://medium.com/@chemistry8526/boosting-sales-with-data-the-power-of-market-basket-analysis-in-retail-c79cc10a14df>

C1. Data Transformation

To prepare the transactions data for MBA, the following **transformations** were made to the initial provided data set:

- **Blank rows** that separated each individual transaction were **removed**.
- A **transactional ID column** was added.
- The data set was **pivoted** such that each row represented one product from one transaction.
- The redundant **Item** column was removed.
- **Empty product values** resulting from previous steps were removed.
- The **data frame** was split into products and IDs and **transformed to a transactions object**.
- The **data was converted into a matrix** with each product as a column and True/False values representing the **presence or absence of products** in each transaction **as the rows**.

Please see the attached CSV file “D212_Task3.csv” for the final **prepared data set**.

C2. Association Rule Generation

With the prepared data, an **Apriori algorithm** was conducted to generate the association rules for MBA as described in section B1. The **input** and **output** from the execution of the code is provided:

```
> arules <- apriori(mba, parameter = list(supp = 0.008, conf = 0.4))
Apriori

Parameter specification:
 confidence minval  smax  arem  aval originalsupport  maxtime support  minlen maxlen target  ext
      0.4       0.1    1 none  FALSE             TRUE       5   0.008     1    10 rules TRUE

Algorithmic control:
 filter tree heap memopt load sort verbose
  0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 60

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[119 item(s), 7501 transaction(s)] done [0.00s].
sorting and recoding items ... [88 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [31 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
```

C3. Association Rule Metrics

The **association rules** table generated by the Apriori algorithm is provided below. Listed are all rules for which the parameters **support** = **0.008** and **confidence** = **0.4** were met, with 31 total rules meeting these requirements. The rules are ordered by **lift**, which is listed along with the other relevant metrics, **support** and **confidence**. A description of each metric is provided in section B1.

lhs	rhs	support	confidence	coverage	lift
[1] {Nylon Braided Lightning to USB cable, SanDisk Ultra 64GB card}	=> {VIVO Dual LCD Monitor Desk mount}	0.008665511	0.5118110	0.01693108	2.939582
[2] {Apple Pencil, SanDisk Ultra 64GB card}	=> {VIVO Dual LCD Monitor Desk mount}	0.008932142	0.4466667	0.01999733	2.565426
[3] {SanDisk Ultra 64GB card, Screen Mom Screen Cleaner kit}	=> {VIVO Dual LCD Monitor Desk mount}	0.009732036	0.4424242	0.02199707	2.541060
[4] {Dust-off compressed Gas 2 pack, SanDisk Ultra 64GB card}	=> {VIVO Dual LCD Monitor Desk mount}	0.017064391	0.4169381	0.04092788	2.394681
[5] {10ft iPhone charger cable 2 Pack, Screen Mom Screen Cleaner kit}	=> {Dust-off compressed Gas 2 pack}	0.008532196	0.5614035	0.01519797	2.355194
[6] {Apple Lightning to Digital AV Adapter, USB 2.0 Printer cable}	=> {Apple Pencil}	0.009065458	0.4121212	0.02199707	2.293265
[7] {Nylon Braided Lightning to USB cable, SanDisk Ultra 64GB card}	=> {Dust-off compressed Gas 2 pack}	0.009198773	0.5433071	0.01693108	2.279277
[8] {Apple Pencil, SanDisk Ultra 64GB card}	=> {Dust-off compressed Gas 2 pack}	0.010131982	0.5066667	0.01999733	2.125563
[9] {FEIYOLD Blue light Blocking Glasses, HP 61 ink}	=> {Dust-off compressed Gas 2 pack}	0.008265565	0.5040650	0.01639781	2.114649
[10] {SanDisk Ultra 64GB card, Screen Mom Screen Cleaner kit}	=> {Dust-off compressed Gas 2 pack}	0.011065191	0.5030303	0.02199707	2.110308
[11] {FEIYOLD Blue light Blocking Glasses, Screen Mom Screen Cleaner kit}	=> {Dust-off compressed Gas 2 pack}	0.008532196	0.5000000	0.01706439	2.097595
[12] {HP 61 ink, SanDisk Ultra 64GB card}	=> {Dust-off compressed Gas 2 pack}	0.010931876	0.4739884	0.02306359	1.988472
[13] {HP 61 ink, Stylus Pen for iPad}	=> {Dust-off compressed Gas 2 pack}	0.009320889	0.4697987	0.01986402	1.970895
[14] {Nylon Braided Lightning to USB cable, Screen Mom Screen Cleaner kit}	=> {Dust-off compressed Gas 2 pack}	0.011065191	0.4689266	0.02359685	1.967236
[15] {10ft iPhone charger cable 2 Pack}	=> {Dust-off compressed Gas 2 pack}	0.023063592	0.4564644	0.05052660	1.914955
[16] {Stylus Pen for iPad, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.011465138	0.4550265	0.02519664	1.908923
[17] {FEIYOLD Blue light Blocking Glasses, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.010265298	0.4476744	0.02293028	1.878079
[18] {Anker USB C to HDMI Adapter, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.009320889	0.4458599	0.02093054	1.870467
[19] {Screen Mom Screen Cleaner kit, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.015731236	0.4436090	0.03546194	1.861024
[20] {HP 61 ink, Screen Mom Screen Cleaner kit}	=> {Dust-off compressed Gas 2 pack}	0.013998134	0.4356846	0.03212905	1.827780
[21] {SanDisk Ultra 64GB card, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.017064391	0.4353741	0.03919477	1.826477
[22] {Nylon Braided Lightning to USB cable, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.011998400	0.4306220	0.02786295	1.806541
[23] {HP 61 ink, Nylon Braided Lightning to USB cable}	=> {Dust-off compressed Gas 2 pack}	0.009732036	0.4244186	0.02293028	1.780517
[24] {Apple Pencil, Screen Mom Screen Cleaner kit}	=> {Dust-off compressed Gas 2 pack}	0.013064925	0.4242424	0.03079589	1.779778
[25] {FEIYOLD Blue light Blocking Glasses}	=> {Dust-off compressed Gas 2 pack}	0.027596320	0.4190283	0.06585789	1.757904
[26] {Apple Pencil, Nylon Braided Lightning to USB cable}	=> {Dust-off compressed Gas 2 pack}	0.009065458	0.4171779	0.02173044	1.750141
[27] {SanDisk Ultra 64GB card}	=> {Dust-off compressed Gas 2 pack}	0.040927876	0.4165536	0.09825357	1.747522
[28] {Apple Pencil, HP 61 ink}	=> {Dust-off compressed Gas 2 pack}	0.013464871	0.4056225	0.03319557	1.701663
[29] {HP 61 ink, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.015864551	0.4047619	0.03919477	1.698053
[30] {Logitech M510 Wireless mouse, VIVO Dual LCD Monitor Desk mount}	=> {Dust-off compressed Gas 2 pack}	0.008532196	0.4025157	0.02119717	1.688630
[31] {3A USB type C cable 3 pack 6ft}	=> {Dust-off compressed Gas 2 pack}	0.017064391	0.4012539	0.04252766	1.683336

Summary statistics for the support, confidence, and lift of the 31 association rules are as follows:

summary of quality measures:			
support	confidence	coverage	lift
Min. :0.008266	Min. :0.4013	Min. :0.01520	Min. :1.683
1st Qu.:0.009132	1st Qu.:0.4181	1st Qu.:0.02046	1st Qu.:1.780
Median :0.010932	Median :0.4436	Median :0.02293	Median :1.909
Mean :0.013185	Mean :0.4525	Mean :0.02987	Mean :2.007
3rd Qu.:0.014865	3rd Qu.:0.4719	3rd Qu.:0.03433	3rd Qu.:2.120
Max. :0.040928	Max. :0.5614	Max. :0.09825	Max. :2.940

C4. Top Three Rules

The **top three rules** generated by the Apriori algorithm are summarized below:

lhs	rhs	support	confidence	coverage	lift
[1] {Nylon Braided Lightning to USB cable, SanDisk Ultra 64GB card}	=> {VIVO Dual LCD Monitor Desk mount}	0.008665511	0.5118110	0.01693108	2.939582
[2] {Apple Pencil, SanDisk Ultra 64GB card}	=> {VIVO Dual LCD Monitor Desk mount}	0.008932142	0.4466667	0.01999733	2.565426
[3] {SanDisk Ultra 64GB card, Screen Mom Screen Cleaner kit}	=> {VIVO Dual LCD Monitor Desk mount}	0.009732036	0.4424242	0.02199707	2.541060

- The **first-rated rule** according to the **lift** metric describes the purchase of a Nylon Braided Lightning to USB cable and a SanDisk Ultra 64GB card (antecedent) associated with the purchase of a VIVO Dual LCD Monitor Desk mount (consequent). This rule identifies that customers who purchase the cable and the card are likely to also purchase the desk mount. Relevant metrics can be described as follows:

- The **support** of this rule, approximately 0.0087, indicates that the combination of these products was found in about 0.87% of all transactions in the data set.
- The **confidence** of this rule, approximately 0.51, indicates that of all transactions that include the cable and the card, about 51% of them also include the desk mount.
- The **lift** of this rule, approximately 2.94, indicates that the desk mount is about 2.94 times more likely to be included in a transaction with the cable and the card than expected.
- The **second-rated rule** according to the **lift** metric describes the purchase of an Apple Pencil and a SanDisk Ultra 64GB card (antecedent) associated with the purchase of a VIVO Dual LCD Monitor Desk mount (consequent). This rule identifies that customers who purchase the pencil and the card are likely to also purchase the desk mount. Relevant metrics can be described as follows:
 - The **support** of this rule, approximately 0.0089, indicates that the combination of these products was found in about 0.89% of all transactions in the data set.
 - The **confidence** of this rule, approximately 0.45, indicates that of all transactions that include the pencil and the card, about 45% of them also include the desk mount.
 - The **lift** of this rule, approximately 2.57, indicates that the desk mount is about 2.57 times more likely to be included in a transaction with the pencil and the card than expected.
- The **third-rated rule** according to the **lift** metric describes the purchase of a SanDisk Ultra 64GB card and Screen Mom Screen Cleaner Kit (antecedent) associated with the purchase of a VIVO Dual LCD Monitor Desk mount (consequent). This rule identifies that customers who purchase the card and the kit are likely to also purchase the desk mount. Relevant metrics can be described as follows:
 - The **support** of this rule, approximately 0.0097, indicates that the combination of these products was found in about 0.97% of all transactions in the data set.
 - The **confidence** of this rule, approximately 0.44, indicates that of all transactions that include the card and the kit, about 44% of them also include the desk mount.
 - The **lift** of this rule, approximately 2.54, indicates that the desk mount is about 2.54 times more likely to be included in a transaction with the card and the kit than expected.

D1. Metrics Significance

The **significance** of the relevant metrics summarized in section C3 can be described as follows:

- **Support** measures the relative frequency of the combinations of products within each rule. While many of these values may appear to be quite low, with the top three relevant rules each representing less than 1% of all transactions, it is important to note that there are 7501 transactions in the data set, so even support values around 1% represent about 75 purchases. The summary statistics of support show that the mean and median are both greater than 1%, and the maximum support value is greater than 4%. Even with the overlap of products in many of the rules determined to be relevant to the MBA, there is a significant number of transactions for which useful information was found in this analysis.

- **Confidence** measures the probability of purchasing the consequent, given the purchase of the antecedent. The confidence values for the top three rules were around 45-50%, and the summary statistics for all relevant rules indicate measures of center around 45% for the confidence of the rules. While this may seem to suggest that the prediction of a product's purchase is less than a coin flip, given that there are 119 products to choose from, having confidence values this high could prove to be highly important in understanding customer transaction patterns.
- **Lift** measures the ratio of observed confidence to expected confidence. The top three relevant rules all have lift values greater than 2.5, which indicates that customers who purchase the antecedents for these rules are more than 2.5 times more likely to purchase the consequents. The summary statistics for all relevant rules show measures of center around 2, so for the 31 relevant rules generated in this analysis, on average they can help predict customer purchase patterns with twice as much accuracy as expected. This could be incredibly beneficial in informing future business decisions.

D2. Practical Significance

As a result of the metrics summarized in section C3 and their significance summarized in section D1, it has been determined that the **Market Basket Analysis was successful in identifying groups of products frequently purchased together**, and the results are **practically significant**. The analysis provided descriptions of customer purchasing patterns with specificity and accuracy and can be used to inform useful marketing campaigns such as product discounts and promotions.

D3. Recommended Course of Action

Based on the results of the significance of the Market Basket Analysis results, it is recommended that the telecommunications company **use the insights gained to market specific products to customers** who will be purchasing other specific products in their transactions. All 31 relevant rules (as determined by the Apriori algorithm) may be used as insights in a similar manner as the **examples given below for each of the top three rules**:

- Customers who are about to purchase a Nylon Braided Lightning to USB cable and a SanDisk Ultra 64GB card **should be targeted to promote** a VIVO Dual LCD Monitor Desk mount, as it has been determined that they will be much more likely to purchase the desk mount than expected.
- Customers who are about to purchase an Apple Pencil and a SanDisk Ultra 64GB card **should be targeted to promote** a VIVO Dual LCD Monitor Desk mount, as it has been determined that they will be much more likely to purchase the desk mount than expected.
- Customers who are about to purchase a SanDisk Ultra 64GB card and Screen Mom Screen Cleaner Kit **should be targeted to promote** a VIVO Dual LCD Monitor Desk mount, as it has been determined that they will be much more likely to purchase the desk mount than expected.

E. Panopto Video

A link to a Panopto video has been provided which shows the execution of the code used to prepare for and conduct the described Market Basket Analysis process.

F. Third-Party Code References

WGU Courseware was used as a resource to learn the methods, concepts, and functions used to create the codes in this project, including DataCamp course tracks (datacamp.com) and Dr. Kesselly Kamara's D212 Panopto videos. There are no codes that have been taken directly from any other resources.

G. Content References

The Market Basket Analysis assumption was clarified at the following source:

<https://medium.com/@chemistry8526/boosting-sales-with-data-the-power-of-market-basket-analysis-in-retail-c79cc10a14df>