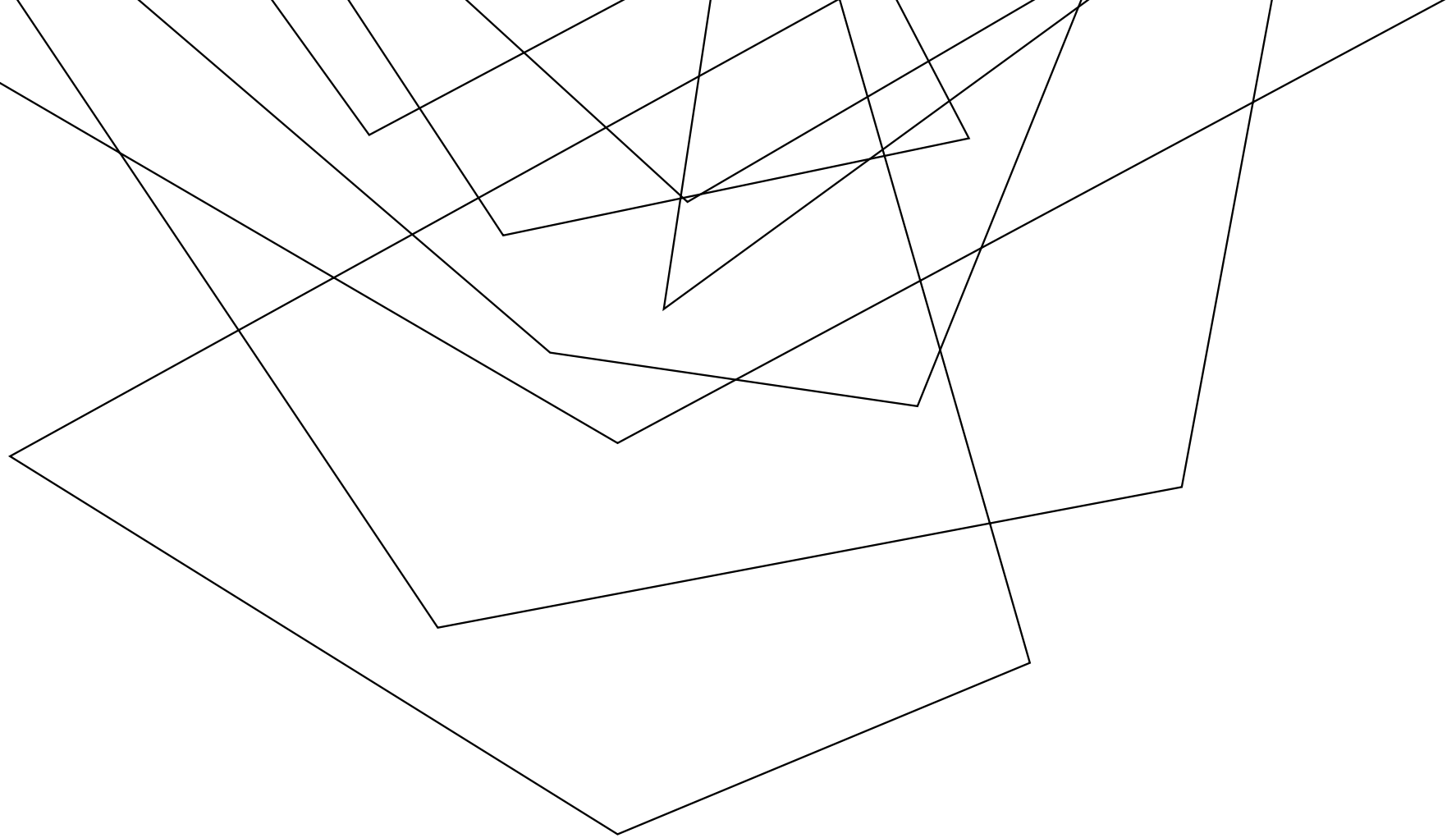
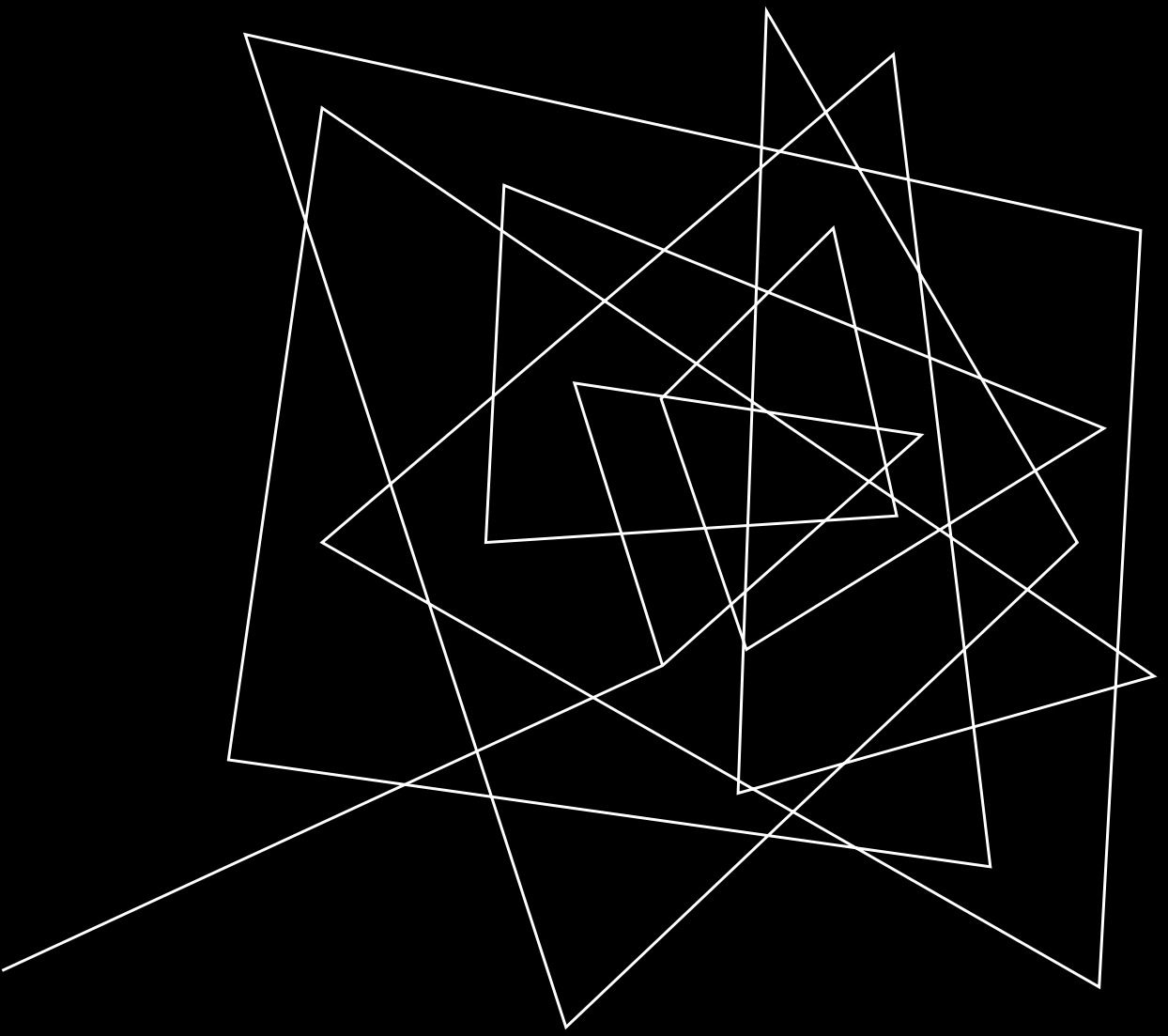


THE FOOD RETAIL

Jumoponpatha Chaimongkonrojna

January 15th, 2024



PROBLEM STATEMENT

In order to improve the purchase of products from the food retail, which products can be the target for cross-selling?

AGENDA

- Problem Statement
- Data Collection
- Data Preparation
- Exploratory Analysis (EDA)
- Implementing and Interpreting Results of the Association Rule Analysis
- Implication and Conclusion

DATA COLLECTION

DATA COLLECTION

TRANSACTION DATASET (5,197,681 ROWS × 11 COLUMNS)

	upc	dollar_sales	units	time_of_transaction	geography	week	household	store	basket	day	coupon
0	7680850106	0.80	1	1100	2	1	125434	244	1	1	0
1	3620000470	3.59	1	1100	2	1	125434	244	1	1	0
2	1800028064	2.25	1	1137	2	1	108320	244	2	1	0
3	9999985067	0.85	1	1148	2	1	162016	244	3	1	0
4	9999985131	2.19	1	1323	2	1	89437	244	4	1	0
...	...	...	...	...	...	...	...	...	...	...	...
5197676	9999985001	0.39	1	2354	1	104	435465	199	3316346	728	0
5197677	9999966720	1.05	1	2354	1	104	435465	199	3316346	728	0
5197678	9999985027	0.99	1	2311	1	104	352690	218	3316347	728	0
5197679	3620000300	1.53	1	2245	1	104	55530	93	3316348	728	0
5197680	1510000073	0.99	1	2209	1	104	435755	18	3316349	728	0

DATA COLLECTION

PRODUCT DATASET (927 ROWS × 11 COLUMNS)

	upc	product_description	commodity	brand	product_size
0	111112360	VINCENT S ORIG MARINARA S	pasta sauce	Vincent's	25 OZ
1	566300023	PINE MOUNTAIN SYRUP	syrups	Pine Mountain	40 OZ
2	566300028	MILLER CANE SYRUP	syrups	Miller	19 OZ
3	566300029	MILLER CANE SYRUP	syrups	Miller	12 OZ
4	566300035	PINE MOUNTAIN SYRUP	syrups	Pine Mountain	19 OZ
...	...	...	...	...	...
922	9999985217	PRIVATE LABEL ALPHABETS	pasta	Private Label	16 OZ
923	9999985260	PRIVATE LABEL COMPLETE PANCAKE MIX	pancake mixes	Private Label	32 OZ
924	9999985261	PRIVATE LABEL COMPLETE PANCAKE MIX	pancake mixes	Private Label	2 LB
925	9999985488	PRIVATE LABEL ITAL NESTED ANGEL HAIR	pasta	Private Label Premium	16 OZ
926	9999985766	PRIVATE LABEL ITALIAN TRI COLORE CRE	pasta	Private Label Premium	16 OZ

DATA COLLECTION

STORE DATASET (387 ROWS × 11 COLUMNS)

	store	store_zip_code
0	1	37865
1	2	30084
2	3	30039
3	4	31210
4	5	30044
...	...	...
382	383	40502
383	384	42367
384	385	37160
385	386	29803
386	387	37748

DATA COLLECTION

CASUAL DATASET (351,371 ROWS × 11 COLUMNS)

	upc	store	week	feature_desc	display_desc	geography
0	7680850108	1	68	Wrap Interior Feature	Not on Display	1
1	5100001212	1	66	Wrap Back Feature	Not on Display	1
2	5100002792	1	72	Interior Page Feature	Not on Display	1
3	3620000300	1	55	Wrap Interior Feature	Not on Display	1
4	4112907742	1	68	Wrap Interior Feature	Not on Display	1
...	...	...	...	...	...	...
351367	9999966721	387	101	Wrap Front Feature	Not on Display	1
351368	9999966720	387	101	Wrap Front Feature	Not on Display	1
351369	3620000507	387	104	Interior Page Feature	Not on Display	1
351370	3620000493	387	104	Interior Page Feature	Not on Display	1
351371	9999966715	387	101	Wrap Front Feature	Not on Display	1

DATA PREPARATION

DATA PREPARATION

No Null/NAN/missing value

```
1 transactions.isnull().sum()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/
  and should_run_async(code)
upc                  0
dollar_sales         0
units                0
time_of_transaction  0
geography            0
week                 0
household            0
store                0
basket               0
day                  0
coupon               0
dtype: int64
```

DATA PREPARATION

No duplicated rows

```
1 transactions.duplicated().sum()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ip
  and should_run_async(code)
0
```

DATA PREPARATION

Remove the non-positive value at ‘dollar_sales’ column

	upc	dollar_sales	units	time_of_transaction	geography	week	household	store	basket	day	coupon
591	3340060109	-0.01	1	1054	2	1	443392	241	385	1	1
1702	3340060108	-0.01	1	2038	2	1	20649	280	1066	1	1
8840	3340060109	-0.01	1	1831	2	1	410790	257	5659	4	1
9758	7680851829	-0.18	1	2011	2	1	44994	276	6254	3	0
10751	3340060110	-0.01	1	2251	2	1	125145	359	6879	4	1
...	...	...	...	...	...	...	...	...	...	...	...
5155492	1510000043	-0.10	1	2220	2	104	463624	304	3290294	723	0
5157872	2920000212	-0.01	1	1153	1	104	361212	73	3291714	723	1
5179958	9999985004	-0.50	1	1323	1	104	218647	201	3305426	727	1
5186586	9999981581	-0.09	1	1811	1	104	179262	28	3309544	728	1
5186587	9999985005	-0.14	1	1811	1	104	179262	28	3309544	728	1

DATA PREPARATION

Convert ‘time_of_transaction’ to be Time Index

	upc	dollar_sales	units	time_of_transaction	geography	week	household	store	basket	day	coupon
0	7680850106	0.80	1	11:00	2	1	125434	244	1	1	0
1	3620000470	3.59	1	11:00	2	1	125434	244	1	1	0
2	1800028064	2.25	1	11:37	2	1	108320	244	2	1	0
3	9999985067	0.85	1	11:48	2	1	162016	244	3	1	0
4	9999985131	2.19	1	13:23	2	1	89437	244	4	1	0
...	...	...	...	...	...	...	...	...	...	...	...
5197676	9999985001	0.39	1	23:54	1	104	435465	199	3316346	728	0
5197677	9999966720	1.05	1	23:54	1	104	435465	199	3316346	728	0
5197678	9999985027	0.99	1	23:11	1	104	352690	218	3316347	728	0
5197679	3620000300	1.53	1	22:45	1	104	55530	93	3316348	728	0
5197680	1510000073	0.99	1	22:09	1	104	435755	18	3316349	728	0

DATA PREPARATION

54% of Transaction was from Area 1

	geography
1	0.540566
2	0.459434

	upc	dollar_sales	units	time_of_transaction	geography	week	household	store	basket	day	coupon
343959	9999985261	1.39	1	00:18	1	1	350145	133	220760	1	0
343960	9999985004	0.78	2	00:29	1	1	340936	133	220761	1	0
343961	5150002683	2.39	1	07:52	1	1	282472	133	220762	1	0
343962	3000003690	1.99	1	07:52	1	1	282472	133	220762	1	0
343963	1800028067	2.19	1	07:52	1	1	282472	133	220762	1	0
...	...	...	...	...	...	...	...	...	...	...	...
5197676	9999985001	0.39	1	23:54	1	104	435465	199	3316346	728	0
5197677	9999966720	1.05	1	23:54	1	104	435465	199	3316346	728	0
5197678	9999985027	0.99	1	23:11	1	104	352690	218	3316347	728	0
5197679	3620000300	1.53	1	22:45	1	104	55530	93	3316348	728	0
5197680	1510000073	0.99	1	22:09	1	104	435755	18	3316349	728	0

Note: In order to serve the need of demonstration, only data of Area 1 was used in the Apriori algorithm Analysis.

DATA PREPARATION

Top Transaction was from Store 71 (1.1% of Total Transaction)

	store
71	0.011040
196	0.009079
16	0.007870
17	0.007798
186	0.007772
162	0.007561
172	0.007381
33	0.007364
215	0.007173
35	0.007153

	upc	dollar_sales	units	time_of_transaction	geography	week	household	store	basket	day	coupon
346502	7130000035	0.99	1	07:42	1	1	218169	71	222387	1	0
346503	7173000722	0.89	1	08:42	1	1	315685	71	222388	1	0
346504	1800000956	3.49	1	10:35	1	1	467570	71	222389	1	0
346505	2920000213	1.09	1	11:05	1	1	231000	71	222390	1	0
346506	7151800024	1.59	1	11:14	1	1	341073	71	222391	1	0
...	...	...	...	...	...	...	...	...	...	...	...
5196230	9999982584	1.76	2	21:01	1	104	482810	71	3315452	726	0
5196231	1800028035	1.99	1	21:01	1	104	482810	71	3315452	726	0
5196232	3620000468	1.58	1	23:24	1	104	172500	71	3315453	726	0
5196233	9999971884	1.27	1	23:24	1	104	172500	71	3315453	726	0
5196234	3620001376	1.58	1	23:24	1	104	172500	71	3315453	726	0

Note: In order to serve the need of demonstration, only data of Store 71 was used in the Apriori algorithm Analysis.

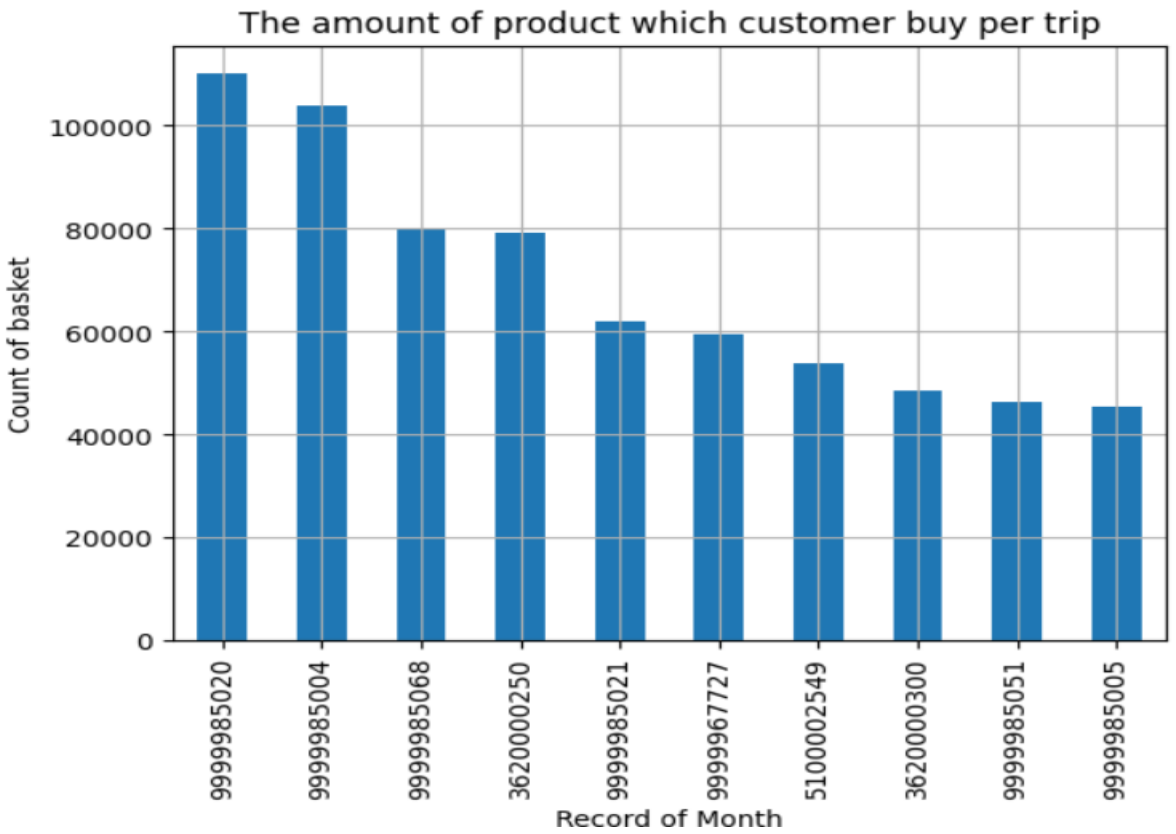
DATA PREPARATION

From total 5,197,681 transactional records, there were 19,538 rows left for implementing association rule analysis.

	upc	dollar_sales	units	time_of_transaction	geography	week	household	store	basket	day	coupon
346508	3000005300	2.29	1	11:36	1	1	286210	71	222393	1	0
346509	3000005970	3.49	1	11:36	1	1	286210	71	222393	1	0
346513	3000005040	2.29	1	13:55	1	1	191228	71	222397	1	0
346514	7130000022	1.98	2	13:55	1	1	191228	71	222397	1	0
346515	2700042238	1.39	1	14:08	1	1	265252	71	222398	1	0
...	...	...	...	...	...	...	...	...	...	...	...
5196230	9999982584	1.76	2	21:01	1	104	482810	71	3315452	726	0
5196231	1800028035	1.99	1	21:01	1	104	482810	71	3315452	726	0
5196232	3620000468	1.58	1	23:24	1	104	172500	71	3315453	726	0
5196233	9999971884	1.27	1	23:24	1	104	172500	71	3315453	726	0
5196234	3620001376	1.58	1	23:24	1	104	172500	71	3315453	726	0

EXPLORATORY DATA ANALYSIS (EDA)

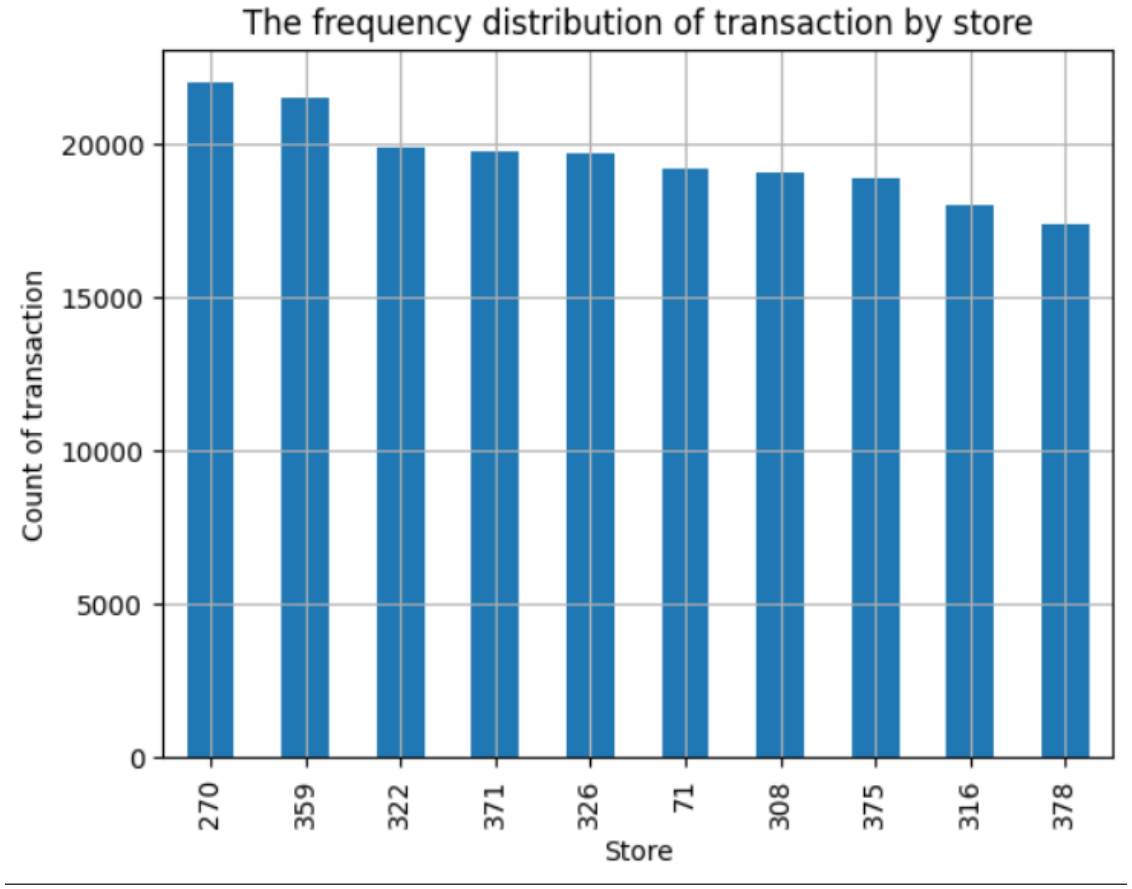
WHAT ITEMS DO CUSTOMERS BUY MORE OFTEN?



Result:

The top 10 most sold products represented 13% of total items sold by the company.

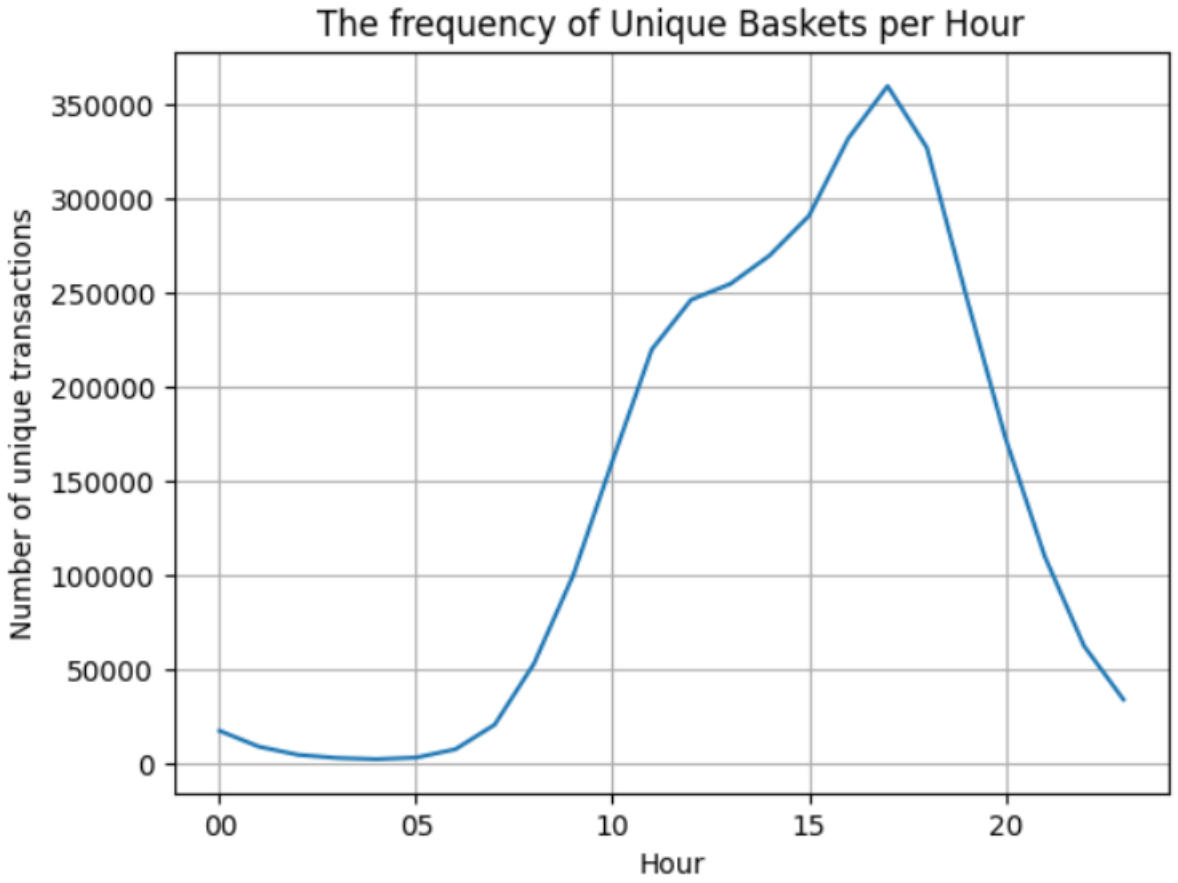
WHICH ARE THE MOST POPULAR STORE?



Result:

Store 71 is in the Top 10 most popular stores

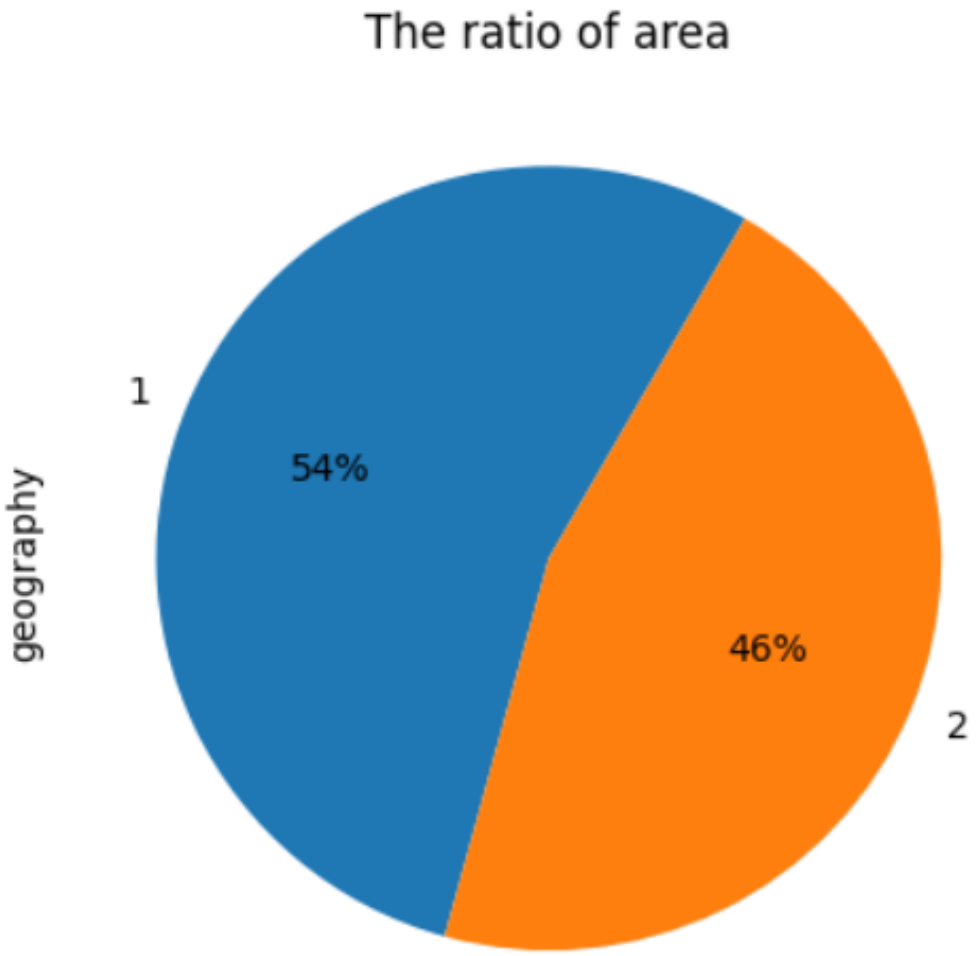
WHAT TIME OF DAY WOULD CUSTOMERS BUY MORE OFTEN?



Result:

Majority of the transactions occurred between 10am. to 8pm. with the peak at a 5pm.

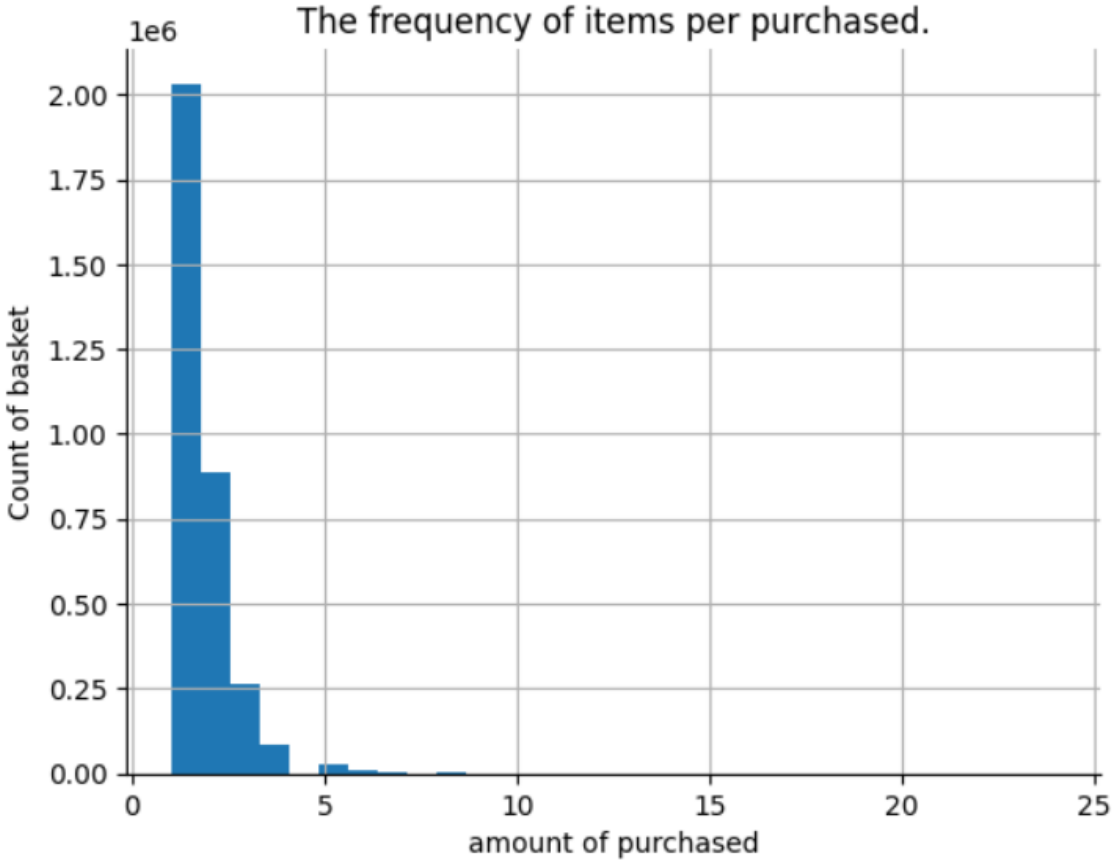
WHAT IS TRANSACTION BY GEOGRAPHICAL AREA?



Result:

54% of total transactions was from Area 1.

HOW MANY ITEMS DOES EACH CUSTOMER BUY?



Result:

A large majority (96%) of customers typically purchased between 1 to 4 items, with a peak at 1 items(61%).

There is a huge opportunity to identify items that can be potentially purchased with.

IMPLEMENTING ASSOCIATION RULE ANALYSIS

Apriori algorithm was used for the analysis

Result of the Top Ten Association Rules based on Lift Index

	antecedents	consequents	support	confidence	lift
0	(7130030013)	(7130030011)	0.001164	0.346154	70.451417
1	(7130030011)	(7130030013)	0.001164	0.236842	70.451417
2	(4112900005)	(4112900006)	0.001552	0.230769	55.774038
3	(4112900006)	(4112900005)	0.001552	0.375000	55.774038
4	(7680850295)	(7151800016, 601011296)	0.001164	0.069767	49.052854
5	(7151800016, 601011296)	(7680850295)	0.001164	0.818182	49.052854
6	(601011296, 7680850295)	(1800000956)	0.001939	0.306122	45.529827
7	(1800000956)	(601011296, 7680850295)	0.001939	0.288462	45.529827
8	(601011296)	(7151800016, 7680850295)	0.001164	0.047619	36.828571
9	(7151800016, 7680850295)	(601011296)	0.001164	0.900000	36.828571

Interpretation of an example:

- There was a strong support of 0.1164% transactions for upc: 7130030013 and upc: 7130030011 in the same basket.
- We have 34.6154% confidence that upc: 7130030011 will be bought whenever upc: 7130030013 is bought.
- The purchased of product upc: 7130030011 was 70.45 times more whenever upc: 7130030013 is purchased than when upc: 7130030011 is purchased alone.

IMPLEMENTING AND INTERPRETING RESULTS OF THE ASSOCIATION RULE ANALYSIS

Result of the Top Ten Association Rules based on Lift Index

	antecedents	consequents	support	confidence	lift
0	(7130030013)	(7130030011)	0.001164	0.346154	70.451417
1	(7130030011)	(7130030013)	0.001164	0.236842	70.451417
2	(4112900005)	(4112900006)	0.001552	0.230769	55.774038
3	(4112900006)	(4112900005)	0.001552	0.375000	55.774038
4	(7680850295)	(7151800016, 601011296)	0.001164	0.069767	49.052854
5	(7151800016, 601011296)	(7680850295)	0.001164	0.818182	49.052854
6	(601011296, 7680850295)	(1800000956)	0.001939	0.306122	45.529827
7	(1800000956)	(601011296, 7680850295)	0.001939	0.288462	45.529827
8	(601011296)	(7151800016, 7680850295)	0.001164	0.047619	36.828571
9	(7151800016, 7680850295)	(601011296)	0.001164	0.900000	36.828571

Interpretation of an example:

- There was a strong support of 0.1164% transactions for upc: 7130030013 and upc: 7130030011 in the same basket.
- We have 34.6154% confidence that upc: 7130030011 will be bought whenever upc: 7130030013 is bought.
- The purchased of product upc: 7130030011 was 70.45 times more whenever upc: 7130030013 is purchased than when upc: 7130030011 is purchased alone.

Among the Top Ten Association Rules, there are only 6 itemsets suitable for cross-selling [1]

- Upc:7130030013 will be bought whenever upc:7130030011 is bought (34.61% confidence)
- Upc:4112900006) will be bought whenever upc:4112900005 is bought (37.50% confidence)
- Upc:(7151800016, 601011296) will be bought whenever upc:7680850295 is bought (81.81% confidence)
- Upc:(601011296, 7680850295) will be bought whenever (1800000956) is bought (30.61% confidence)
- Upc:1800000956 will be bought whenever upc:(601011296, 7680850295) is bought (28.84% confidence)
- Upc:(7151800016, 7680850295) will be bought whenever upc:601011296 is bought, (90% confidence)

Note [1] Cofidence threshold was set at least 25%

IMPLICATION AND CONCLUSION

IMPLICATION AND CONCLUSION

Due to time constraints, I can only run the association rule analysis to identify the items that are frequently purchased together which is potential for cross-selling. In order to improve the purchase of product from the food retail, I recommend the store manager to rearrange the placement of the 6 pairs of product items next to each other. Furthermore, I would recommend the further anlaysis to improve promotion strategies based on the dataset in the causal file.

THANK YOU

Jumoponpatha Chaimongkonrojna

jumponpat59@gmail.com

0865639118

Google Colab:

https://colab.research.google.com/drive/1BvrV_8HY1NtShJpi4ZugI927B9LOV-hB?usp=sharing