MAJOR PROJECT REPORT

Submitted by : ABHIPSA JENA
Email : abhipsajena2k@gmail.com
Batch : JUNE 2021 - JULY 2021
Group Code : DS-06-BM2

WEBSITE : https://abhipsa-j.github.io/Market_Basket_Analysis/
(Deployed on)

Problem Statement

To find the association rules between the items using the Apriori Algorithm. Or in other words, you have to find out those items/itemset that customers bought together which helps the owner for store layout/marketing. And then show the relative Sales of the data in Tableau Dashboard

Tools Used

Python
Google Colab
Tableau

Algorithms

The algorithm used to find the association rules is Apriori Algorithm.

Apriori Algorithms

Apriori is an algorithm for frequent item set mining and association rule learning over relational databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

Association Rule Mining
Association Rule Mining is used when you want to find an association between different objects in a set, find frequent patterns in a transaction database, relational databases or any other information repository. The applications of Association Rule Mining are found in Marketing, Basket Data Analysis (or Market Basket Analysis) in retailing, clustering and classification. It can tell you what items do customers frequently buy together by generating a set of rules called Association Rules.

Approach

The provided dataset had transactions labeled as weekend or weekdays. We divided the dataset into two parts : Weekdays & Weekends.
And used Apriori Algorithm to train our model on both the datasets individually in order to find the association rules.

Steps involve preparation of data, separating the data into weekends and weekdays, transforming the table and using Apriori Algorithm to find association rules.

Result

5 rules were generated for Weekdays dataset and 5 rules were generated for Weekdays dataset

> WEEKDAYS RULES

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
15	(Toast)	(Coffee)	0.037266	0.485598	0.026526	0.711790	1.465802
10	(Pastry)	(Coffee)	0.087551	0.485598	0.049634	0.566914	1.167456
0	(Cake)	(Coffee)	0.095850	0.485598	0.052238	0.544992	1.122310
8	(Medialuna)	(Coffee)	0.052726	0.485598	0.028641	0.543210	1.118641
5	(Cookies)	(Coffee)	0.062002	0.485598	0.031082	0.501312	1.032361
12	(Sandwich)	(Coffee)	0.075346	0.485598	0.037592	0.498920	1.027434
7	(Hot chocolate)	(Coffee)	0.051424	0.485598	0.025386	0.493671	1.016625
2	(Cake)	(Tea)	0.095850	0.149390	0.022783	0.237691	1.591080
3	(Tea)	(Cake)	0.149390	0.095850	0.022783	0.152505	1.591080
1	(Coffee)	(Cake)	0.485598	0.095850	0.052238	0.107574	1.122310
11	(Coffee)	(Pastry)	0.485598	0.087551	0.049634	0.102212	1.167456
13	(Coffee)	(Sandwich)	0.485598	0.075346	0.037592	0.077413	1.027434
4	(Coffee)	(Cookies)	0.485598	0.062002	0.031082	0.064008	1.032361
9	(Coffee)	(Medialuna)	0.485598	0.052726	0.028641	0.058981	1.118641
14	(Coffee)	(Toast)	0.485598	0.037266	0.026526	0.054625	1.465802
6	(Coffee)	(Hot chocolate)	0.485598	0.051424	0.025386	0.052279	1.016625

> WEEKEND RULES

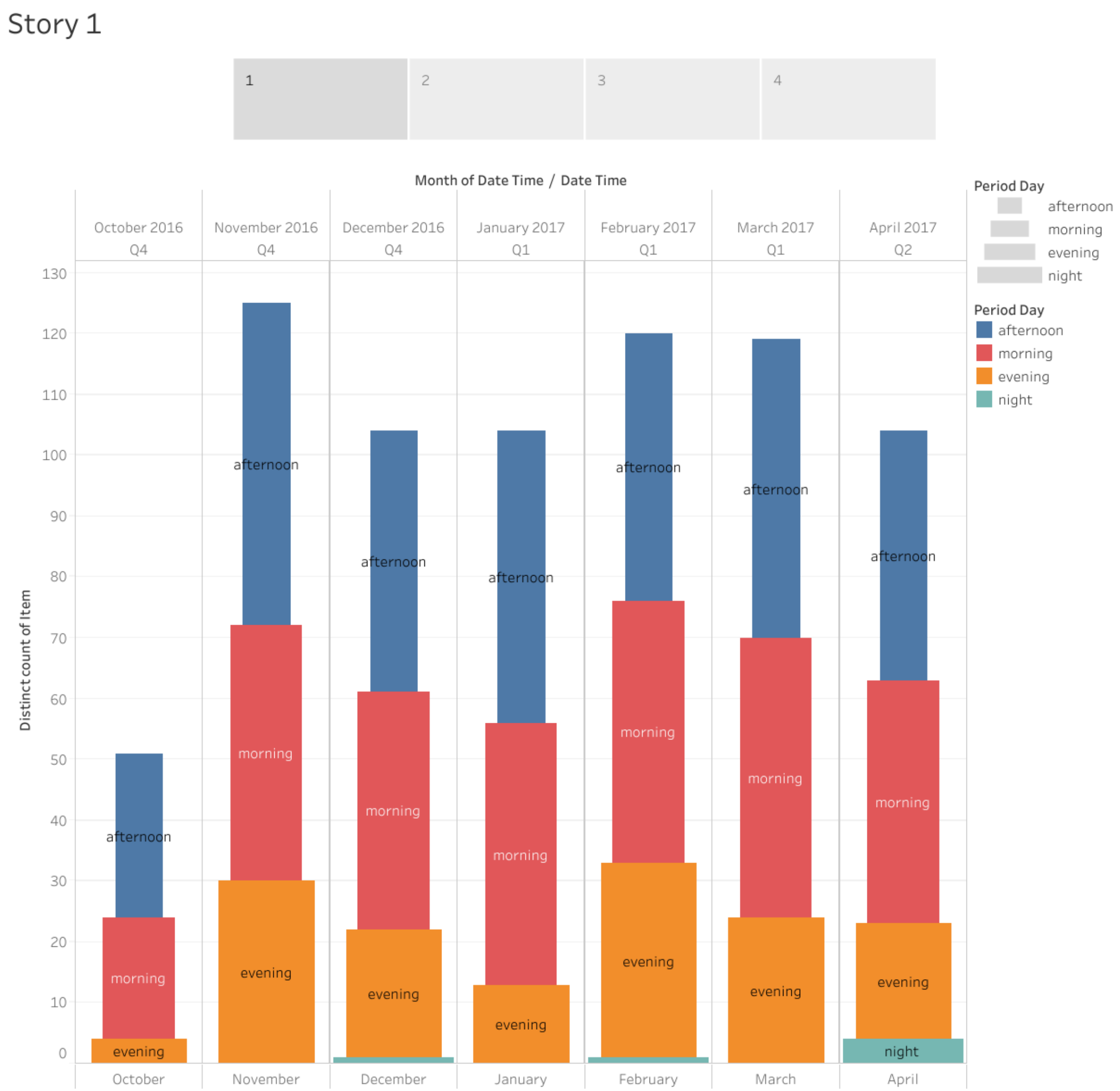
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(Alfajores)	(Coffee)	0.038554	0.465060	0.020181	0.523437	1.125526
1	(Coffee)	(Alfajores)	0.465060	0.038554	0.020181	0.043394	1.125526
2	(Pastry)	(Bread)	0.083434	0.344578	0.032229	0.386282	1.121027
3	(Bread)	(Pastry)	0.344578	0.083434	0.032229	0.093531	1.121027
4	(Coffee)	(Brownie)	0.465060	0.051807	0.028614	0.061528	1.187643
5	(Brownie)	(Coffee)	0.051807	0.465060	0.028614	0.552326	1.187643
6	(Cake)	(Coffee)	0.118675	0.465060	0.059337	0.500000	1.075130
7	(Coffee)	(Cake)	0.465060	0.118675	0.059337	0.127591	1.075130
8	(Cake)	(Tea)	0.118675	0.130120	0.025602	0.215736	1.657971
9	(Tea)	(Cake)	0.130120	0.118675	0.025602	0.196759	1.657971
10	(Coffee)	(Cookies)	0.465060	0.040361	0.022892	0.049223	1.219550
11	(Cookies)	(Coffee)	0.040361	0.465060	0.022892	0.567164	1.219550
12	(Coffee)	(Hot chocolate)	0.465060	0.071084	0.037349	0.080311	1.129797
13	(Hot chocolate)	(Coffee)	0.071084	0.465060	0.037349	0.525424	1.129797
14	(Coffee)	(Juice)	0.465060	0.041566	0.024096	0.051813	1.246527
15	(Juice)	(Coffee)	0.041566	0.465060	0.024096	0.579710	1.246527
16	(Medialuna)	(Coffee)	0.078614	0.465060	0.047289	0.601533	1.293451
17	(Coffee)	(Medialuna)	0.465060	0.078614	0.047289	0.101684	1.293451
18	(Pastry)	(Coffee)	0.083434	0.465060	0.043675	0.523466	1.125587
19	(Coffee)	(Pastry)	0.465060	0.083434	0.043675	0.093912	1.125587
20	(Sandwich)	(Coffee)	0.065361	0.465060	0.039458	0.603687	1.298083
21	(Coffee)	(Sandwich)	0.465060	0.065361	0.039458	0.084845	1.298083
22	(Coffee)	(Scone)	0.465060	0.061145	0.031928	0.068653	1.122795
23	(Scone)	(Coffee)	0.061145	0.465060	0.031928	0.522167	1.122795
24	(Coffee)	(Spanish Brunch)	0.465060	0.030422	0.021386	0.045984	1.511568
25	(Spanish Brunch)	(Coffee)	0.030422	0.465060	0.021386	0.702970	1.511568

Tableau

https://public.tableau.com/app/profile/ayush.shende/viz/Book2_16302178687840/Story1

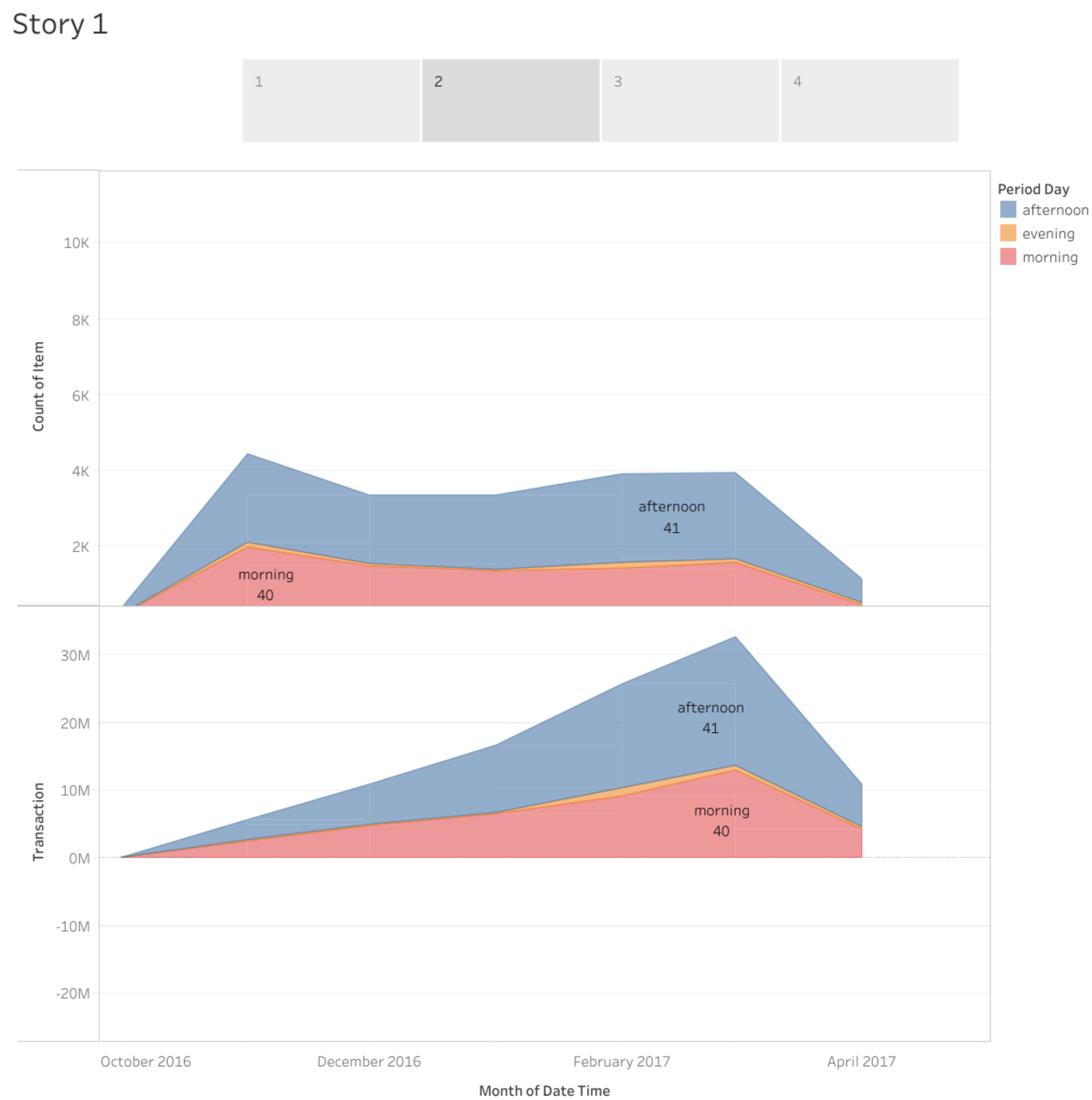
GRAPH 1

So the graph below shows the monthly distribution for number of items sold in different segments of day (Morning, Afternoon, Evening and Night)



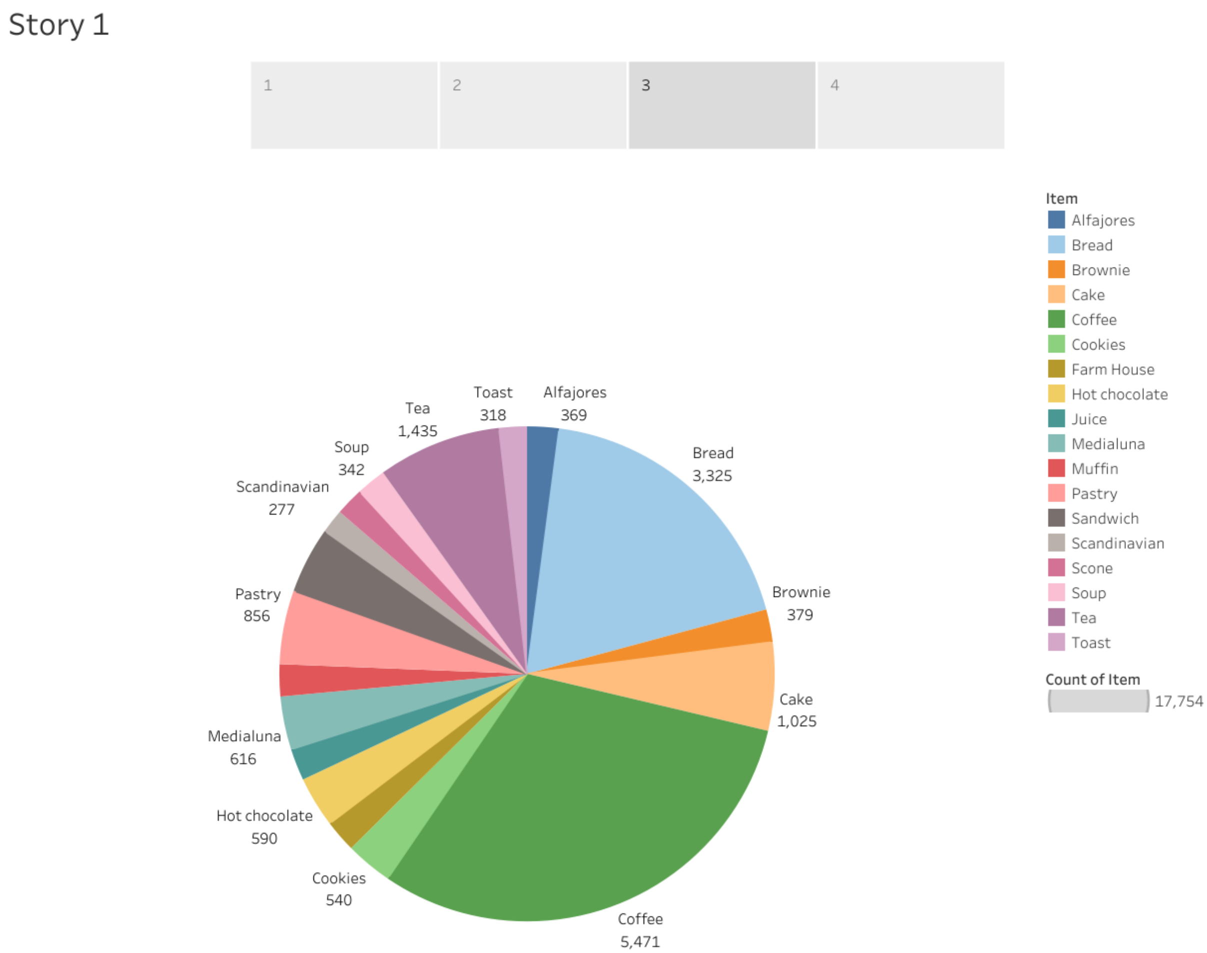
GRAPH 2

The below picture gives the visual overview of how the number of transactions and count of sold items varies over time in different segments of the day.



GRAPH 3

The pie diagram shows the distribution of 18 items occured in all time.



GRAPH 4

The bar diagram here shows the few of the top items sold and each bar is further divided into weekdays and weekends.