

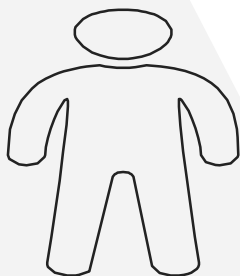
Market Basket Analysis



IDEA BEHIND THE PROJECT

MARKET BASKET ANALYSIS : Tool to Predict Customer Buying Pattern

Customers



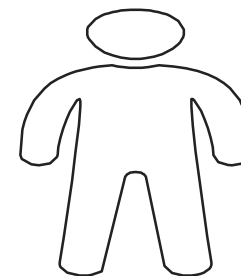
- No need to go grocery.
- What items should be in my cart ?



Frequently bought together



Retailers



- Generate higher sales volume to remain profitable.
- Customer Retention.

WORKFLOW

**Business Problem &
Datasets**

Exploratory Data Analysis

Objectives:

To find relationships and establish patterns across purchases.

To predict which previously purchased products will be in a user's next order.

**Recommendation:
Association Rules**

Predictive Modelling



Project Details-1 : Datasets

- ❖ Instacart kaggle Dataset 2017: <https://www.kaggle.com/competitions/instacart-market-basket-analysis/data>
- ❖ Contains Seven different dataframe files

	order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
0	2539329	1	prior	1	2	8	NaN
1	2398795	1	prior	2	3	7	15.0
2	473747	1	prior	3	3	12	21.0
3	2254736	1	prior	4	4	7	29.0
4	431534	1	prior	5	4	15	28.0
5	3367565	1	prior	6	2	7	19.0
6	550135	1	prior	7	1	9	20.0

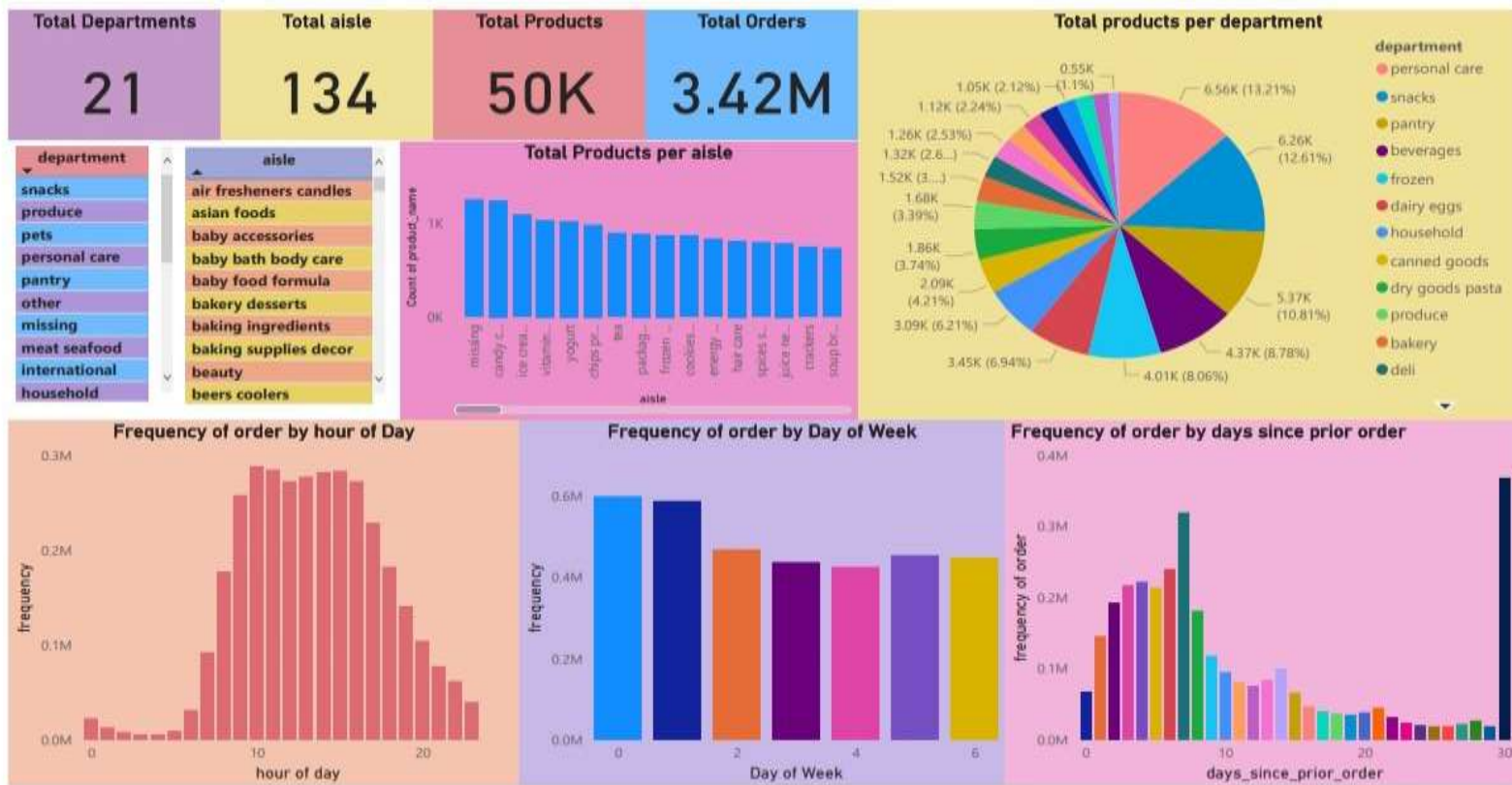
	product_id	product_name	aisle_id	department_id
0	1	Chocolate Sandwich Cookies	61	19
1	2	All-Seasons Salt	104	13
2	3	Robust Golden Unsweetened Oolong Tea	94	7
3	4	Smart Ones Classic Favorites Mini Rigatoni Wit...	38	1
4	5	Green Chile Anytime Sauce	5	13

	order_id	product_id	add_to_cart_order	reordered
0	2	33120	1	1
1	2	28985	2	1
2	2	9327	3	0
3	2	45918	4	1
4	2	30035	5	0

	aisle_id	aisle
0	1	prepared soups salads
1	2	specialty cheeses
2	3	energy granola bars
3	4	instant foods
4	5	marinades meat preparation

	department_id	department
0	1	frozen
1	2	other
2	3	bakery
3	4	produce
4	5	alcohol

PROJECT DETAILS-2: Dashboard



Frequency of order by hour of Day

Frequency of order by Day of Week

Frequency of order by days since prior order

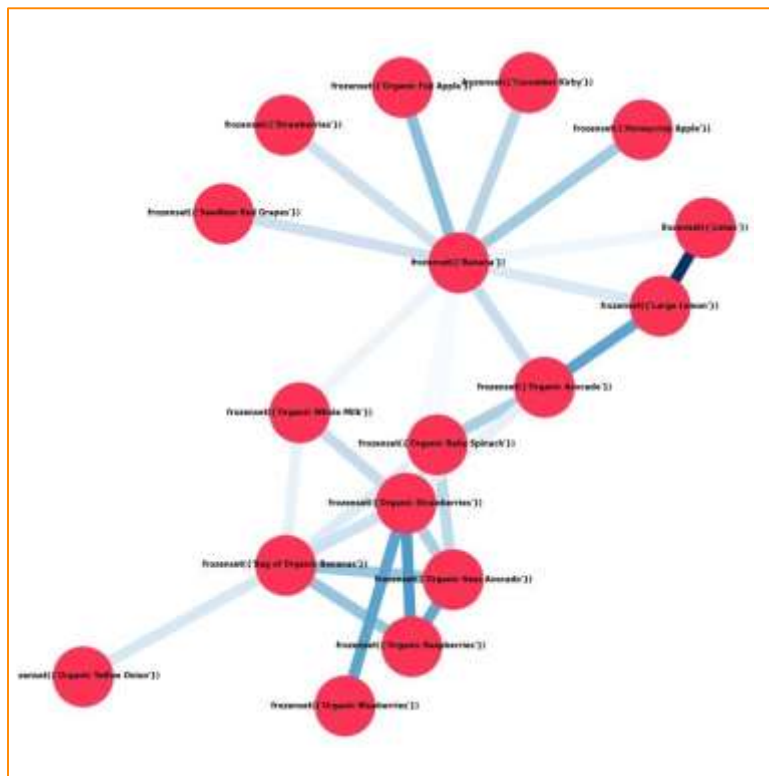


Rule	Support	Confidence	Lift
$A \Rightarrow D$	20	20	100
$C \Rightarrow A$	35	34	58
$A \Rightarrow C$	25	20	54
$B \& C \Rightarrow D$	15	10	56

Purchasing Patterns: Apriori Algorithm

Reordered or not:
LGBM Classification
model
(Avg F1-score=0.3284)

	antecedents	consequents	antecedent support	consequent support	support	confidence	$\frac{support}{\min(support)}$	leverage	conviction
35	(Limes)	(Large Lemon)	0.009364	0.005764	0.011880	0.107723	3.036544	0.007915	1.104485
34	(Large Lemon)	(Limes)	0.005764	0.005884	0.011560	0.180040	3.000544	0.037915	1.149343
83	(Organic Strawberries)	(Organic Raspberries)	0.112711	0.058325	0.045333	0.128840	2.210731	0.007988	0.001068
52	(Organic Raspberries)	(Organic Strawberries)	0.058325	0.112711	0.045333	0.249174	2.210731	0.007988	1.181751
36	(Organic Avocado)	(Large Lemon)	0.075348	0.005764	0.010538	0.130862	2.138728	0.005583	1.008547
37	(Large Lemon)	(Organic Avocado)	0.005764	0.075348	0.010538	0.160044	2.138728	0.005583	1.101097
47	(Organic Strawberries)	(Organic Blueberries)	0.112711	0.042958	0.010235	0.090609	2.114024	0.005384	0.052633
46	(Organic Blueberries)	(Organic Strawberries)	0.042958	0.112711	0.010235	0.238074	2.114024	0.005384	1.164840
48	(Organic Raspberries)	(Organic Hass Avocado)	0.008325	0.000338	0.010088	0.188018	2.081257	0.009697	1.120258
49	(Organic Hass Avocado)	(Organic Raspberries)	0.000338	0.008325	0.010088	0.121389	2.081257	0.009697	1.071777
24	(Banana)	(Organic Fuji Apple)	0.007988	0.037580	0.014378	0.071592	1.883367	0.006744	1.038147
25	(Organic Fuji Apple)	(Banana)	0.007988	0.200636	0.014378	0.378441	1.883367	0.006744	1.203076
0	(Bag of Organic Bananas)	(Organic Raspberries)	0.058325	0.058325	0.017294	0.107065	1.835662	0.007873	1.054584
4	(Organic Raspberries)	(Bag of Organic Bananas)	0.058325	0.161527	0.017294	0.286508	1.835662	0.007873	1.191874
3	(Organic Hass Avocado)	(Bag of Organic Bananas)	0.000338	0.161527	0.026487	0.283188	1.815175	0.011888	1.062594



SKILLS LEARNT

- Python Coding
- Python libraries like Pandas, Numpy, Matplotlib, Seaborn, Scikit Learn, mlxtend
- Apriori Algorithm and Performance metrics used like Support, Consequent and Lift.
- Various Machine Learning Algorithms for Classification.
- Power BI, SQL
- Business Intelligence
- Team Work



CHALLENGES FACED

- Big Data Handling of 3 million grocery order from 200000 users.
- Insufficient RAM on Colab GPU, so considered the fraction sample of data.
- Some NaN values and outliers in the data.



CONCLUSION / FUTURE PLAN

- Domain Knowledge is required for better validation of model.
- Increase sales and retention rate can be achieved via our model.

➤ **Future Work:**

- Extend this solution, to provide even more recommendations for different use cases.
- To find an end to end Deep Learning solution for this problem.
- Deploying this model using Flask.

THANKS!