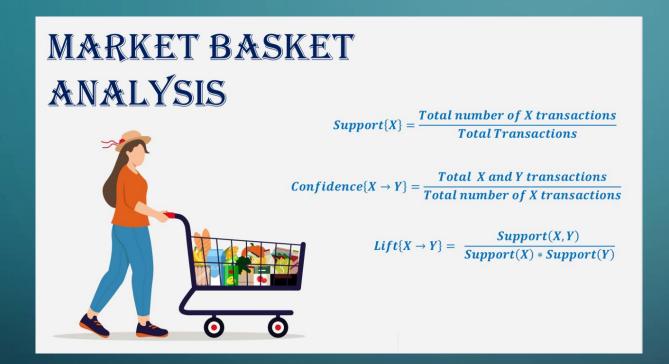
## MARKET RESEARCH ANALYSIS



MARKET BASKET ANALYSIS

-BY SUKANYA

#### DAGENDA - CONTENT

- Data Dictionary
- Executive Summary
- Problem Statement
- About the data
- Assumptions about the data
- Exploratory Data Analysis
  - Univariate Analysis
  - Bivariate Analysis
  - Inferences from EDA
  - Are there trends across months/years/quarters/days?

- Use of Market Basket Analysis Association Rules
  - What are association rules?
  - KNIME Workflow Image
  - Threshold Values
- Associations Identified Put the associations in a tabular manner
- Explain about Support, Confidence and Lift values that are calculated.
- Suggestions of possible combos with lucrative offers

### DATA DICTIONARY

Date	Order_id	Product
01-01-2018	1	yogurt
01-01-2018	1	pork
01-01-2018	1	sandwich bags
01-01-2018	1	lunch meat
01-01-2018	1	all- purpose
01-01-2018	1	flour
01-01-2018	1	soda
01-01-2018	1	butter
01-01-2018	1	beef
01-01-2018	1	aluminum foil
01-01-2018	1	all- purpose
01-01-2018	1	dinner rolls
01-01-2018	1	shampoo
01-01-2018	1	all- purpose
01-01-2018	1	mixes
01-01-2018	1	soap
01-01-2018	1	laundry detergent
01-01-2018	1	ice cream
01-01-2018	1	dinner rolls
01-01-2018	2	toilet paper
01-01-2018	2	shampoo

#### PROBLEM STATEMENT

 A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

#### ABOUT THE DATA

• SHAPE OF THE DATA: 20641 rows and 3 columns

• INFO ABOUT THE DATA: 1 Numerical field and 2 Categorical fields

NO MISSING VALUES

• SUMMARY STATS:

	count	mean	std	min	25%	50%	75%	max
Order_id	20641.0	575.986289	328.557078	1.0	292.0	581.0	862.0	1139.0

For Numerical Fields

	count	unique	top	freq
Date	20641	603	08-02-2019	183
Product	20641	37	poultry	640

For Categorical Fields

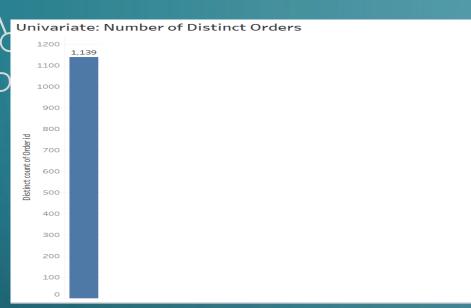
#### **ASSUMPTIONS ABOUT THE DATA**

The assumption made in Market Basket Analysis is that each entry in the dataset represents an independent transaction, where a member makes purchases on a specific date.

This analysis looks for relationships among entities and objects that frequently appear together.



## EXPLORATORY DATA ANALYSIS UNIVARIATE ANALYSIS

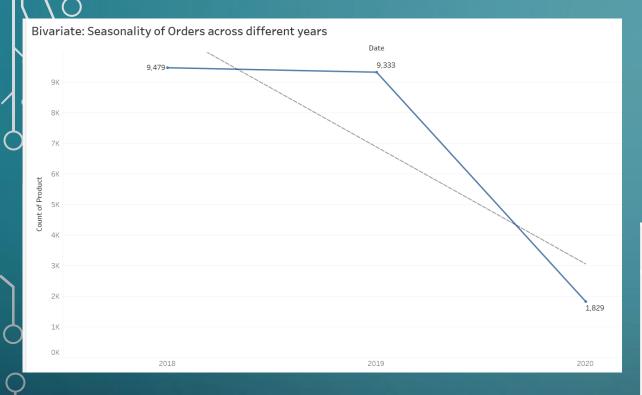


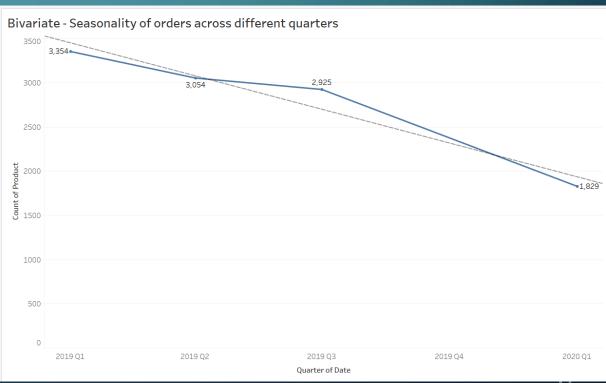


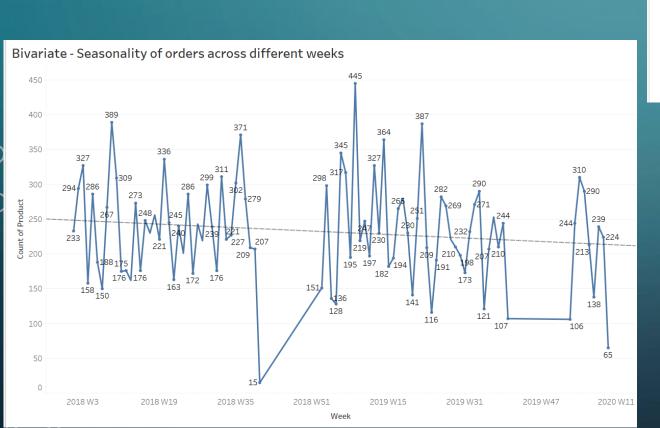


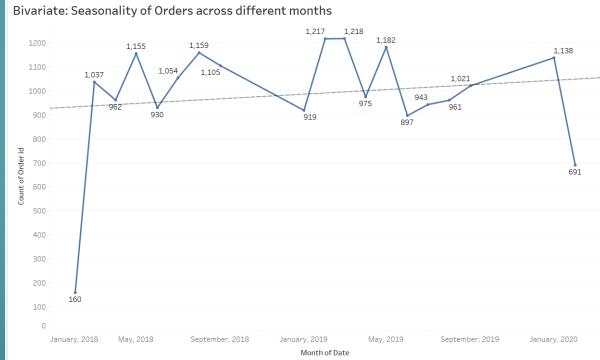
#### Total Number of different products sold Product poultry 597 soda cereals 579 ice cream cheeses waffles 575 soap lunch meat 573 bagels 570 juice eggs toilet paper 569 dinner rolls aluminum foil 566 coffee/tea shampoo 562 beef 556 paper towels milk 555 flour butter mixes 554 dishwashing liquid/.. 551 all-purpose ketchup yogurt 545 individual meals 543 tortillas pasta laundry detergent 542 spaghetti sauce 50 100 150 200 250 300 350 400 450 500 550 600 650 0 Count of dataset\_group.csv

#### **BIVARIATE ANALYSIS**









#### Number of distinct order Ids for each product Product **=** poultry 480 454 ice cream cereals lunch meat 450 waffles soda 445 cheeses 444 eggs dinner rolls dishwashing liquid/.. 442 bagels 438 yogurt aluminum foil milk 433 soap coffee/tea 432 toilet paper 431 laundry detergent juice 428 mixes individual meals beef 427 all-purpose 425 spaghetti sauce pasta ketchup 423 fruits tortillas 421 shampoo 419 sandwich bags butter 50 100 500 0 150 200 250 300 350 400 450 Distinct count of Order id =

Bivariate - Number of Products per OrderId Order id Count of Product 

#### **INFERENCES FROM EDA**

- The total number of orders are 20641 while the number of distinct orders are 1139.
- The total number of distinct products are 37.
- Maximum number of products sold is poultry with 640 units of poultry being sold, closely followed by soda with 597 units and cereals with 579 units.
- Hand soap is sold the least with 502 units.
- Order ID 1071, 1013, 957 and 226 give maximum number of products as the order.

#### Are there trends across months/years/quarters/days?

- 2018 and 2019 show almost the same number of orders placed with the graph line being straight but dips steeply in 2020, showing a major dip in the number of orders placed. The orders dip from around 9500 to 1829. This can be attributed to Covid.
- We see a downward trend in the number of orders from Q1 to Q2 to Q3 to Q4.
- Monthly trend shows a dip in the months of January and February and slowly picks up in the subsequent months in the number of orders placed.
- May sees a peak and then again the orders dip in June and July.
- Weekly trend shows a good variation in the seasonality of number of orders placed. The maximum dip happens in Week 34-35 in 2018. The peak number of orders is in Week 12 of 2019.

# USE OF MARKET BASKET ANALYSIS (ASSOCIATION RULES)

#### WHAT ARE ASSOCIATION RULES?

Association rules in Market Basket Analysis are used to predict the likelihood of products being purchased together. They do this by counting together how often items occur together and look for patterns that happen more often than expected.

This technique searches for association/relationship between items in a dataset.

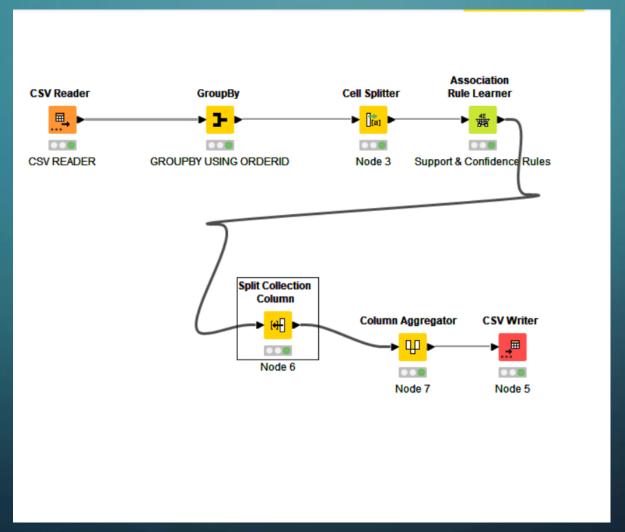
SUPPORT: Support is an indication of how frequently the item set appears in the data set. In other words, it's the number of transactions with both X and Y divided by the total number of transactions.

CONFIDENCE: For a rule,  $X \to Y$ , confidence shows the percentage in which Y is bought with X.

UIFT: The lift of a rule is the ratio of the observed support to that expected if X and Y were independent.

#### KNIME WORKFLOW IMAGE

• KNIME is used for Market Basket Analysis. Here is the KNIME workflow.



#### • THRESHOLD VALUES:

SUPPORT: 0.05

CONFIDENCE: 0.5

5% support means the fraction of transactions with probability of item A and B occurring together as compared to the total number of transactions is 5 out of 100.

We have taken 50% confidence level as the threshold, which means 50% of the people who are buying item A are also buying item B.

## ASSOCIATIONS IDENTIFIED - PUT THE ASSOCIATIONS IN A TABULAR MANNER

\ \ 	Support	Confidence	Lift	Recommended Items	implies	Basket Items
O						
	0.055312	0.649484536	1.79119343	paper towels	<	eggs, ice cream, pasta
	0.055312	0.642857143	1.73100304	pasta	<	paper towels, eggs, ice cream
	0.050922	0.674418605	1.726208518	cheeses	<	bagels, cereals, sandwich bags
	0.050044	0.640449438	1.700400723	juice	<	yogurt, toilet paper, aluminum foil
P	0.050922	0.630434783	1.677722471	mixes	<	yogurt, poultry, aluminum foil
7						
/	0.050922	0.610526316	1.659640749	sandwich bags	<	cheeses, bagels, cereals
	0.053556	0.642105263	1.650920756	dinner rolls	<	spaghetti sauce, poultry, laundry detergent

## EXPLAIN ABOUT SUPPORT, CONFIDENCE AND LIFT VALUES THAT ARE CALCULATED

In market basket analysis, support, confidence, and lift values are calculated to help understand customer behavior.

Support	Confidence	Lift	Consequent	implies	Rule
0.06	0.65	1.79	paper towels	<	eggs, ice cream, pasta
0.06	0.64	1.73	pasta	<	paper towels, eggs, ice cream
0.05	0.67	1.73	cheeses	<	bagels, cereals, sandwich bags
0.05	0.64	1.70	juice	<	yogurt, toilet paper, aluminum foil

Association rule: {eggs, ice cream, pasta, paper towels} — This means if customers buy eggs, ice cream and pasta, they are likely to buy paper towels as well.

Support: The percentage of transactions that contain a specific combination of items, relative to the total number of transactions in the database. We have set a support of 0.06. This means that 6% of all the transactions have these items – eggs, icecream, pasta and paper towels together.

confidence: The probability that a transaction that contains certain items also contains a second item. Confidence measures how often the basket containing eggs, ice cream and pasta contains paper towels holds true.

Confidence is 0.65.

The formula for confidence is Freq(X,Y)/Freq(X). This means that 65% of the transactions that contains eggs, ice cream and pasta also contains paper towels.

Lift: The formula for lift is the confidence percent divided by the support percent. Lift measures how much more likely the consequent (paper towels) is given to the antecedants (eggs, ice cream and pasta) than it would be by chance. If the value of lift is less than 1, the combination is not bought by consumers frequently. If the value of lift is greater than 1, the combination is bought frequently. If the value of lift is equal to 1, then the purchase of the pantecedent makes no difference on the consequent.

Lift is 1.79.

This means customers are 1.79 times more likely to buy paper towels when they buy eggs, ice cream and pasta.

# SUGGESTION OF POSSIBLE COMBOS WITH LUCRATIVE OFFERS

COMBO 1	COMBO 2	COMBO 3	COMBO 4
			Yogurt, toilet paper, aluminium foil,
Eggs, icecream, pasta & paper towels	Paper towels, eggs, icecream & pasta	Bagels, cereals, sandwich bags, cheeses	juice
	<b>OFFER:</b> Buy paper towels, eggs and	<b>OFFER:</b> Buy any two of the combo of bagels,	<b>OFFER:</b> Buy yogurt, toilet paper,
<b>OFFER:</b> Buy eggs, icecream and pasta	icecream together and get pasta at 15%	cereals and sandwich bags and get cheeses	aluminium foil together and get juice
and get paper towels for free!	discount!	at 10% discount!	free.
<b>DISCOUNT:</b> Get paper towels free if			
you buy any two items from egg,			
icecream and pasta.	<b>DISCOUNT:</b> Discount of 15% on pasta!	<b>DISCOUNT:</b> Discount of 10% on cheese!	DISCOUNT: BUY 3, GET 1 FREE!
			RECOMMENDATION: Free juice with
RECOMMENDATION: Sale of items in		<b>RECOMMENDATION:</b> The discount of 10% on	combo 4 will increase the sales of
the basket can increase by giving paper	<b>RECOMMENDATION:</b> Offering pasta at 15%	cheese will push up the sales of the combo	juice which is a perishable item and
towels free with any two products.	discount will boost the sales of pasta.	items too.	also increase sales of combo items.

