MRA Project
Milestone 2

Presented By: Tejas Padekar
Batch: PGP DSBA – G1
Email: tnpadekar@gmail.com

AGENDA:

- About data (Continuous and Categorical)
- EDA
- Inferences
- Market Basket Analysis
- Recommendations

PROBLEM STATEMENT:

- A Grocery Store shared the transactional data with you. Your job is to identify the most popular combos that can be suggested to the Grocery Store chain after a thorough analysis of the most commonly occurring sets of menu items in the customer orders. The Store doesn't have any combo meals. Can you suggest the best combo meals?
- Below is head of the dataset

	Date	Order_id	Product
0	2018-01-01	1	yogurt
1	2018-01-01	1	pork
2	2018-01-01	1	sandwich bags
3	2018-01-01	1	lunch meat
4	2018-01-01	1	all- purpose

ABOUT DATA

- There are 20641 rows and 3 columns.
- There are 1 Integer and 2 object data types.
- There are no null values in the dataset.
- There are 4730 duplicate rows in the dataset.
- Data is available for 3 years 2018, 2019 and 2020
- Data for each year 2018 and 2019 is provided until Q3 ie; September only and for 2020 it is provided for January and February only
- A total of 1139 orders were received during this timeline.
- There are 37 unique Products as of today with the store.

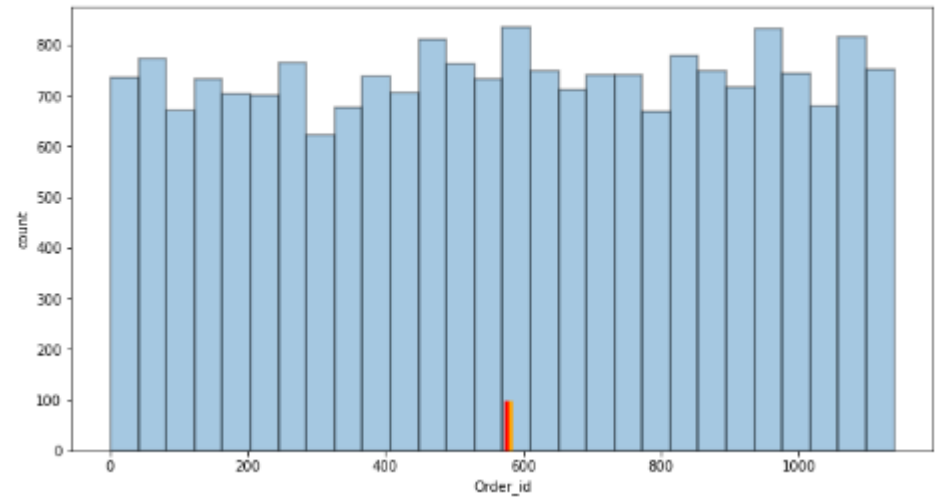
EXPLORATARY DATA ANALYSIS

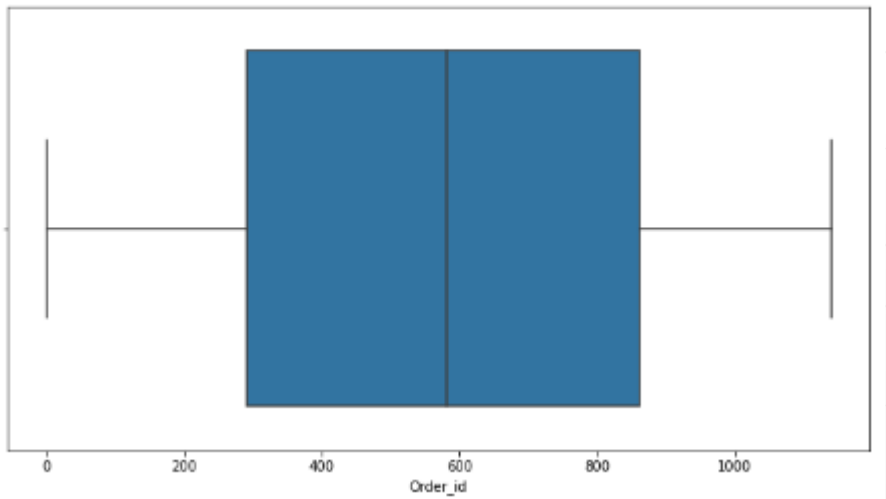
We have used Python and Tableau for performing EDA

UNIVARIATE ANALYSIS
DATA DISTRIBUTION & OUTLIERS

- There are no outliers, no high skewness in Order_id and the data seems to be normally distributed as well.

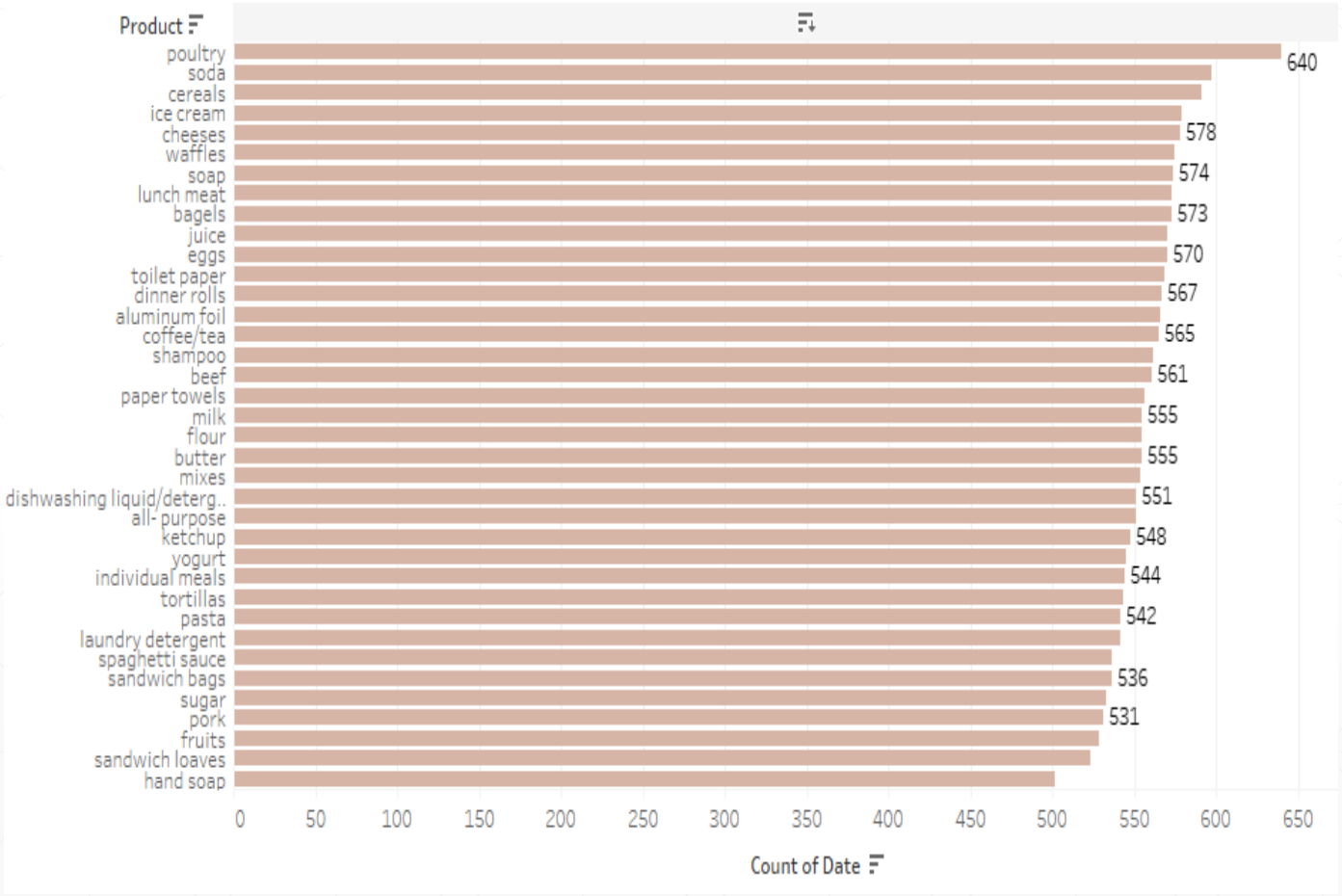
```
Order_id
Skew: -0.03
```





PRODUCT QUANTITY ORDERED MOST

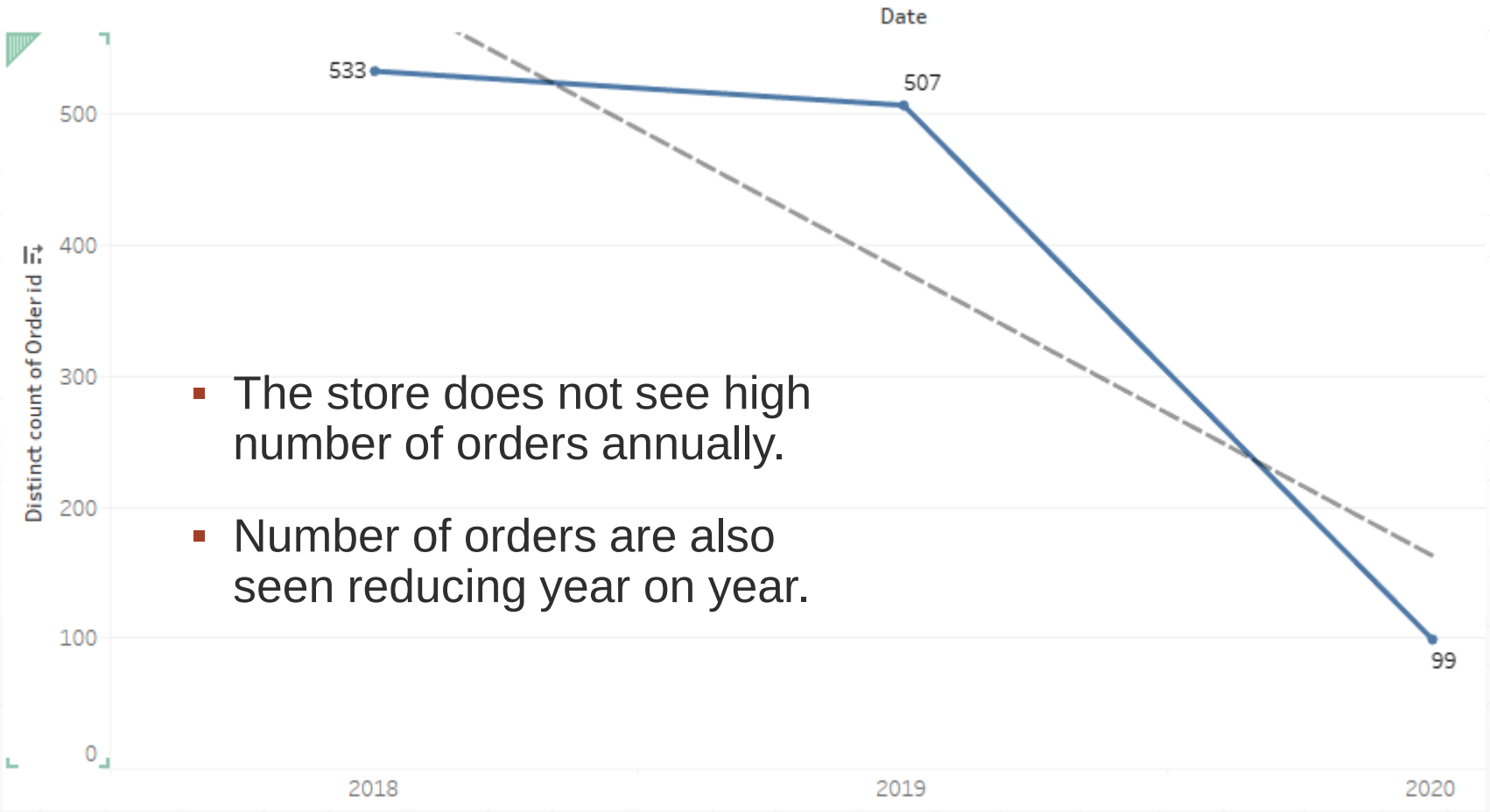
- Poultry accounts for highest purchased product quantity followed by Soda in 2nd and Cereal in the 3rd position.
- Hand Soup is the least purchased product quantity followed by Sandwich loaves and Fruits.
- There is not a drastic difference between the range in which all the products quantity are ordered.



PRODUCT ORDERED MOST

- Poultry accounts for highest purchased product followed by Ice Cream in 2nd and Cheeses in the 3rd position.
- Flour is the least purchased product followed by Pork and Hand Soaps.
- There is not a drastic difference between the range in which all the products are purchased.

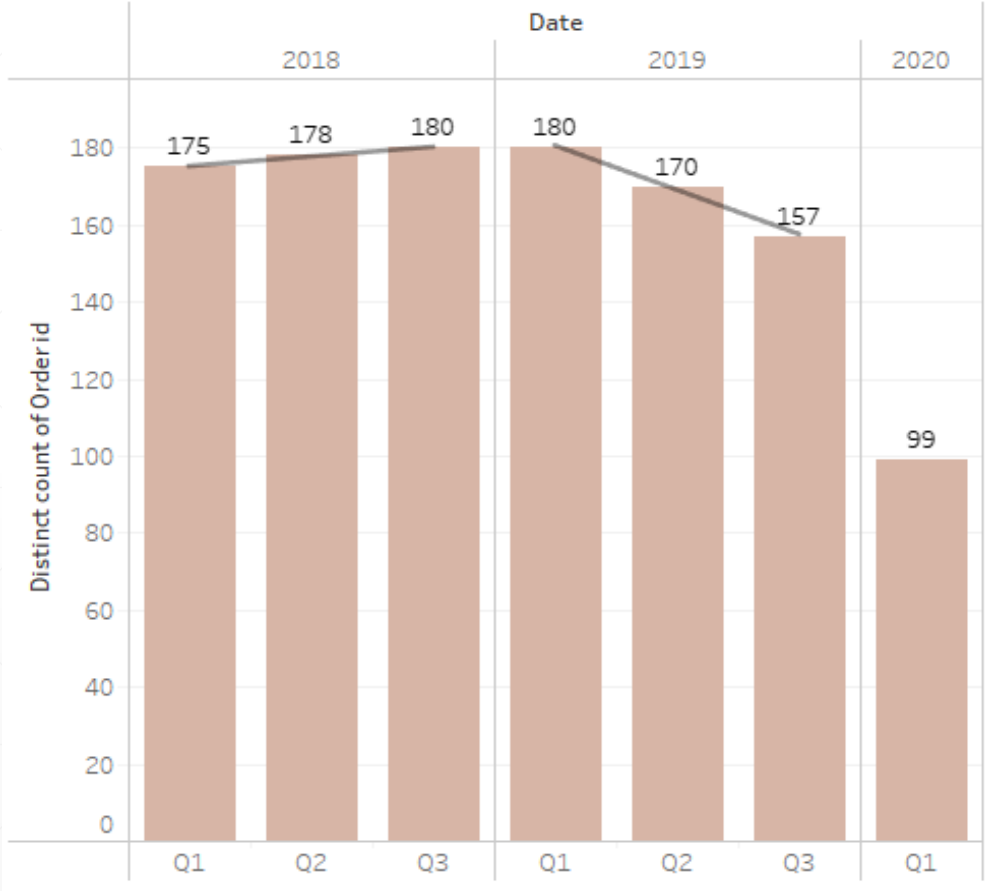
YEARLY ORDERS AND TREND



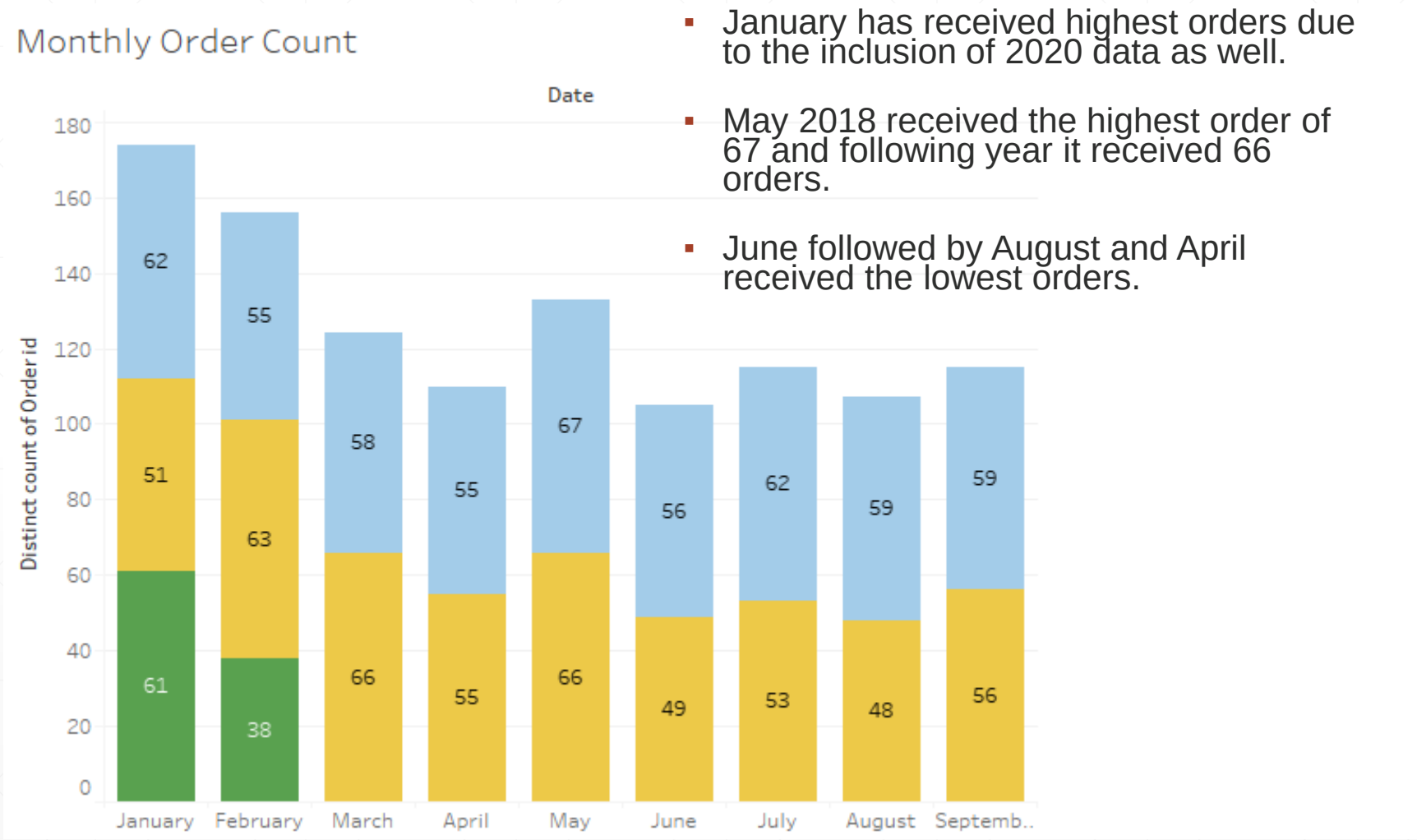
- The store does not see high number of orders annually.
- Number of orders are also seen reducing year on year.

QUARTERLY ORDERS AND TREND

- In 2018, Q1 to Q3 order count has slightly increased with Q3 receiving the highest order for 180.
- In contrast to 2018, Q1 to Q3 order count has decreased significantly with Q3 receiving the lowest orders for 157.
- Data shows inconsistency of customer visits to the store.
- No data available for Q4 for both 2018 and 2019.

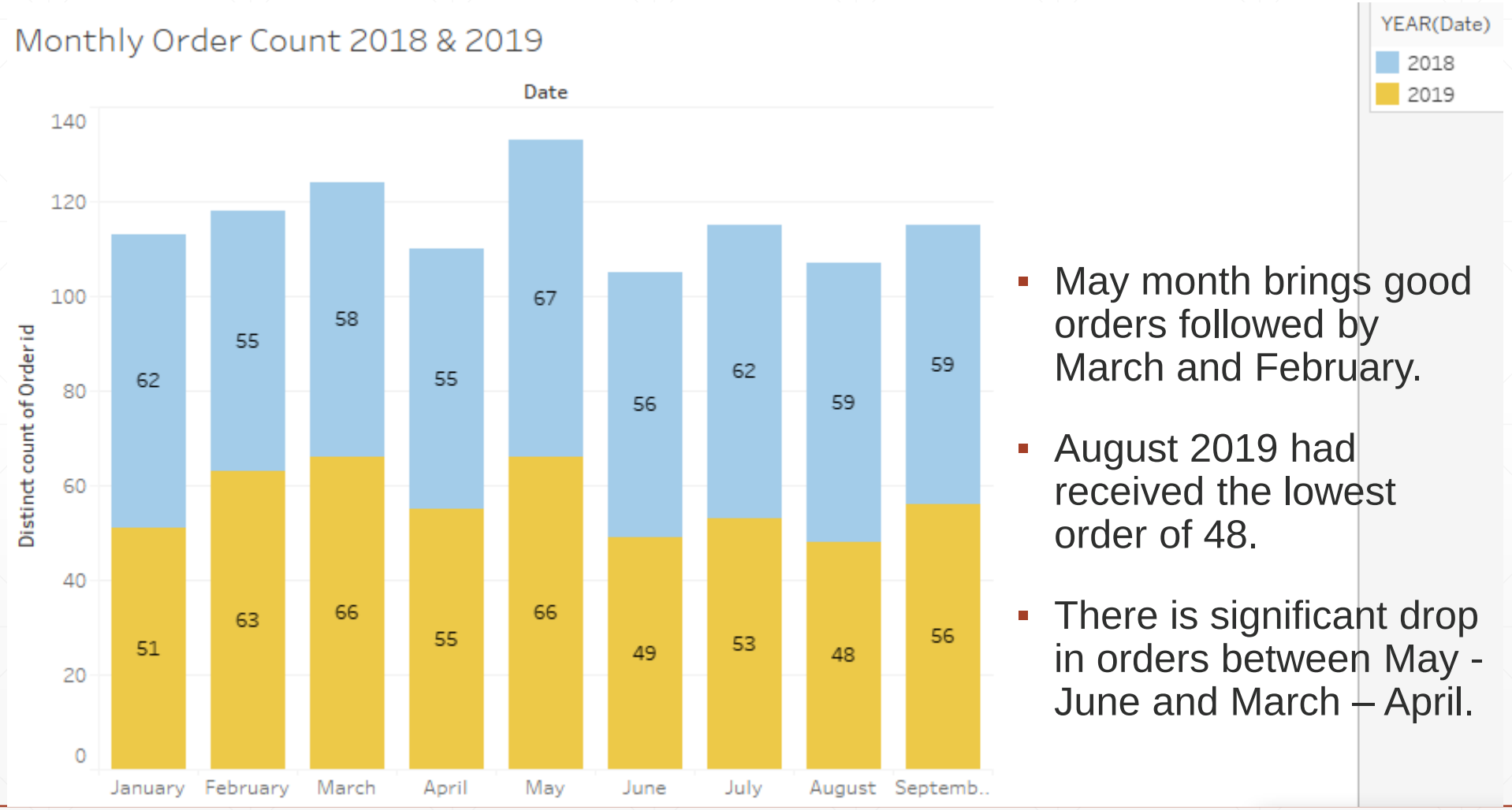


MONTHLY ORDERS



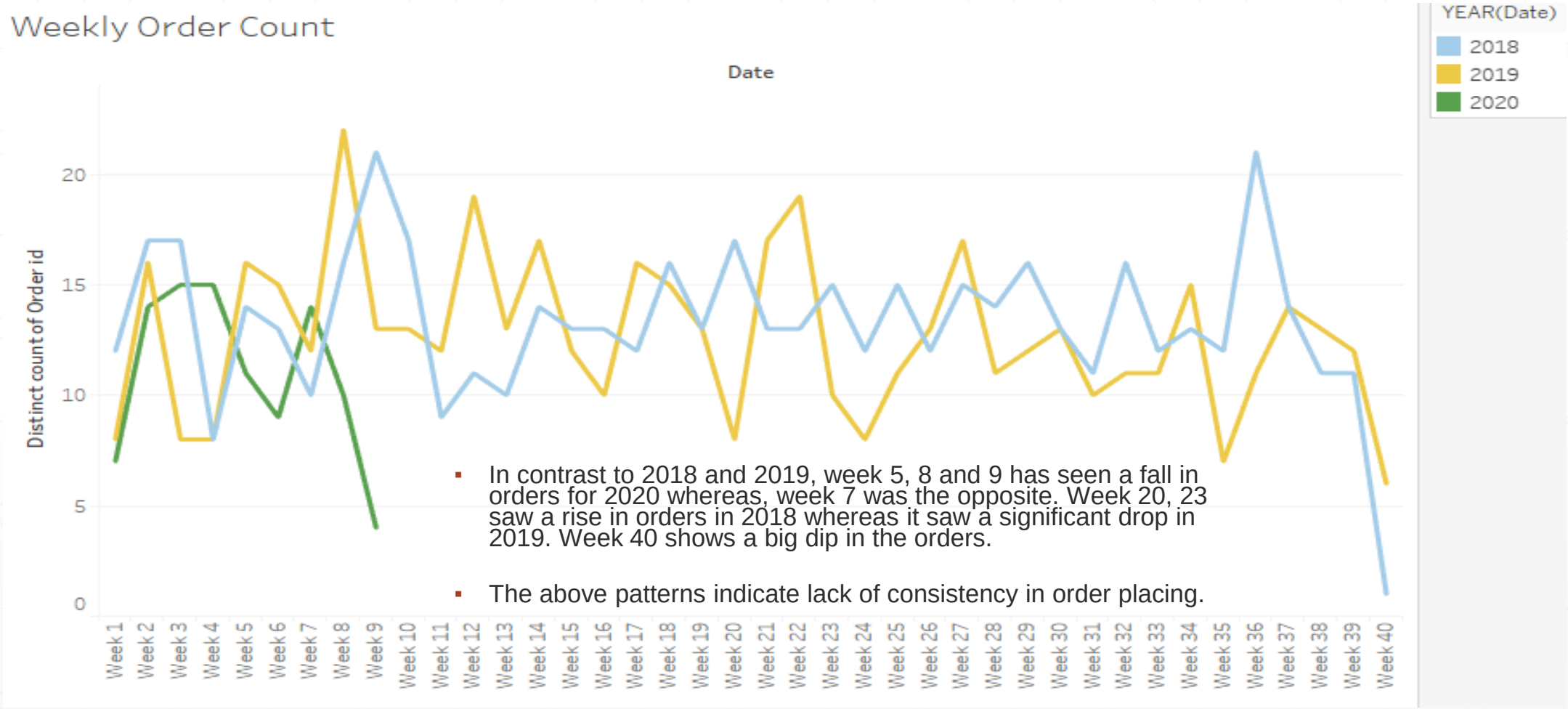
- January has received highest orders due to the inclusion of 2020 data as well.
- May 2018 received the highest order of 67 and following year it received 66 orders.
- June followed by August and April received the lowest orders.

MONTHLY ORDERS EXCLUDING 2020



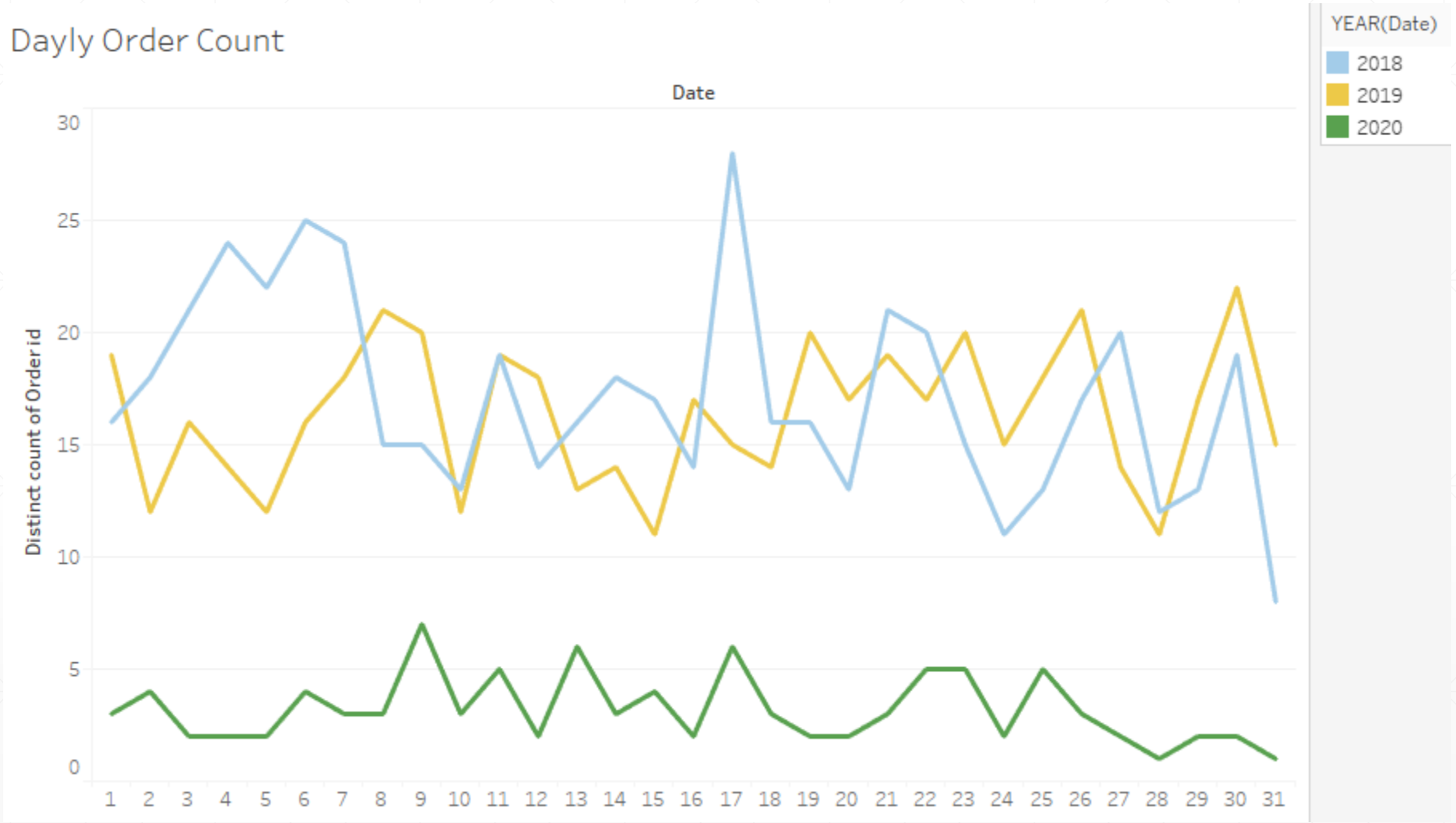
- May month brings good orders followed by March and February.
- August 2019 had received the lowest order of 48.
- There is significant drop in orders between May - June and March – April.

WEEK-WISE ORDERS TREND



- In contrast to 2018 and 2019, week 5, 8 and 9 has seen a fall in orders for 2020 whereas, week 7 was the opposite. Week 20, 23 saw a rise in orders in 2018 whereas it saw a significant drop in 2019. Week 40 shows a big dip in the orders.
- The above patterns indicate lack of consistency in order placing.

DAY-WISE ORDERS TREND



INFERENCES:

- There are 20641 rows and 3 columns.
- Data is available for 3 years 2018, 2019 and 2020.
- Data for each year 2018 and 2019 is provided until Q3 ie; September only and for 2020 it is provided for January and February only.
- A total of 1139 orders were received during this timeline.
- There are 37 unique Products with the store.
- Poultry accounts for highest purchased product quantity followed by Soda in 2nd and Cereal in the 3rd position.
- Hand Soup is the least purchased product quantity followed by Sandwich loaves and Fruits.
- There is not a drastic difference between the range in which all the products quantity are ordered.
- Poultry accounts for highest purchased product followed by Ice Cream in 2nd and Cheeses in the 3rd position.
- Flour is the least purchased product followed by Pork and Hand Soaps.
- There is not a drastic difference between the range in which all the products are purchased.

INFERENCES:

- The store does not see high number of orders annually.
- Number of orders are also seen reducing year on year.
- In 2018, Q1 to Q3 order count has slightly increased. However, in 2019, Q1 to Q3 order count has decreased significantly.
- Q3 was best in terms of orders received for 2018 but it received the lowest orders in the same quarter for 2019.
- Data shows inconsistency of customer visits to the store.
- January has received highest orders due to the inclusion of 2020 data as well
- May 2018 received the highest order of 67 and following year it received 66 orders.
- June followed by August and April received the lowest orders .
- May month brings good orders followed by March and February.
- August 2019 had received the lowest order of 48.

INFERENCES:

- There is significant drop in orders between May - June and March – April.
- In contrast to 2018 and 2019, week 5, 8 and 9 has seen a fall in orders for 2020 whereas, week 7 was the opposite.
- Week 20, 23 saw a rise in orders in 2018 whereas it saw a significant drop in 2019 during these weeks. Week 40 shows a big dip in the orders.
- Week 17, 26, 30, 37 saw orders increasing in 2019 whereas in 2018 orders had decreased during these weeks.
- Weekly Order Trend shows inconsistency in order placing or demand by the consumer.
- There is presence of seasonality in the dataset.
- Day-Wise we can see that 17th and 30th see highest orders.
- Orders range typically between 10 to 20 orders per day.

MARKET BASKET ANALYSIS

MARKET BASKET ANALYSIS AND ADVANTAGES

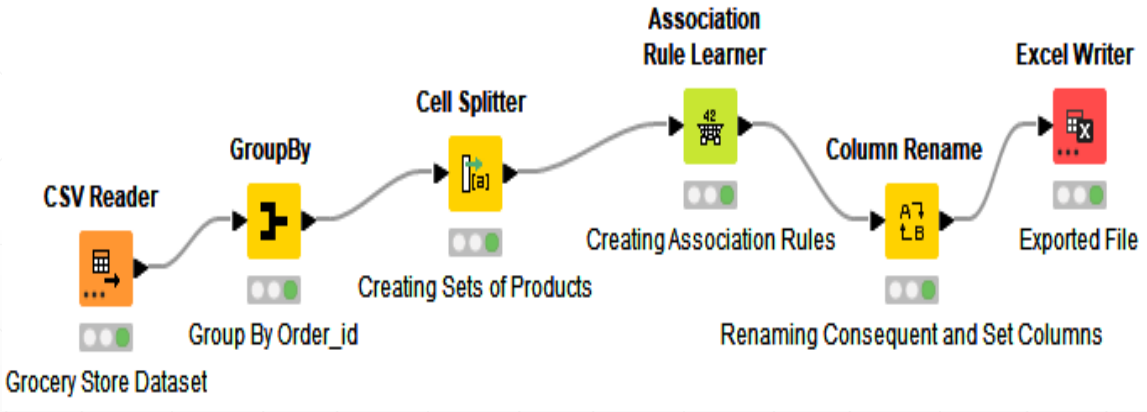
- Market Basket Analysis (MBA) is a data mining technique used by retailers and businesses to understand consumer buying patterns/behaviors to design strategies for increasing revenue and profits.
- Its main purpose is to identify the products which are purchased frequently in combination with other items and how strongly these combination of purchases can help increase the revenues. For eg; If consumer buys products A & B then he/she would buy C as well.
- **Advantages of MBA:**
  - A. Increases sales, cross-selling, consumer loyalty and CLV
  - B. Helps in designing relevant promotional campaigns
  - C. Create online recommendation engines
  - D. Design inventory procurement priorities
  - E. Helps with in-store product placement

THRESHOLDS

- SUPPORT: This helps in identifying the frequency of occurrence for a set of products in an order or which set of products are purchased together the most out of the total purchases made. Set of products with higher support are likely to be purchased more frequently than the one's with a lower support.
- **For our project we have set minimum support as 0.05**
- CONFIDENCE: This denotes the probability of purchases containing recommended products (Consequent) in the same purchases already containing the set of products (Antecendent). For eg; How many purchases which contain product A (set of products) also contain product B (consequent). A higher confidence means its most likely that B will also be purchased if A is present in the order.
- **For our project we have set minimum confidence as 0.05**
- LIFT: This suggests the probability of purchasing recommended products if set of products are already present in order and if there is any association between the two. If lift is > 1, then the probability of purchasing recommended products increases if set of products are present in the order and it can be said that there is an association between the two and vice versa. A higher lift indicates higher level of association.

KNIME MBA WORKFLOW

- Data is Grouped by Order_id
- Products are concatenated
- Set of Products are created using Cell Splitter
- Support is set at minimum 0.05
- Confidence is set at minimum 0.5
- We got a total of 1247 rules
- Consequent and Set Columns are renamed to Recommended_Product and Set_of_Products respectively



ASSOCIATION RULES

- Association Rule determines the frequency of occurrence of an itemset in a purchase order where itemset represents a set of products.
- It helps show how strong or weak the associations between the products are. It determines that products having strong associations are often purchased together in most cases and products with weak associations are said to be independent of each another.
- Using the association rules we would be able to determine which products have strong associations and which are often purchased together helping our store owner to form cross-selling strategies, promotional or/and combo or/and discount offers, shuffle with in-store product placement etc.
- Moreover, the association rules are proven to be very useful for retailers which again makes it very relevant for our grocery store to make use of the same and increase revenue and profits.

ASSOCIATION RULE OUTPUT

- Top 20 rows of Association Rule Output sorted by Lift in descending order.

Table "default" - Rows: 1247 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	Support	Confide...	Lift	Recommended_Product	implies	Set_Of_Products
rule59	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]
rule58	0.055	0.643	1.731	pasta	<---	[paper towels, eggs, ice cream]
rule21	0.051	0.674	1.726	cheeses	<---	[bagels, cereals, sandwich bags]
rule3	0.05	0.64	1.7	juice	<---	[yogurt, toilet paper, aluminum foil]
rule18	0.051	0.63	1.678	mixes	<---	[yogurt, poultry, aluminum foil]
rule20	0.051	0.611	1.66	sandwich bags	<---	[cheeses, bagels, cereals]
rule52	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]
rule40	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]
rule7	0.05	0.62	1.645	juice	<---	[yogurt, poultry, aluminum foil]
rule43	0.052	0.686	1.628	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
rule49	0.052	0.634	1.627	eggs	<---	[paper towels, dinner rolls, pasta]
rule50	0.052	0.602	1.621	pasta	<---	[paper towels, eggs, dinner rolls]
rule24	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce, poultry, cereals]
rule57	0.055	0.63	1.616	eggs	<---	[paper towels, ice cream, pasta]
rule11	0.05	0.613	1.616	coffee/tea	<---	[yogurt, cheeses, cereals]
rule44	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce, poultry, juice]
rule35	0.052	0.628	1.61	eggs	<---	[dinner rolls, poultry, soda]
rule54	0.054	0.598	1.603	spaghetti sauce	<---	[dinner rolls, poultry, laundry detergent]
rule29	0.051	0.604	1.589	milk	<---	[poultry, laundry detergent, cereals]
rule42	0.052	0.59	1.581	spaghetti sauce	<---	[dinner rolls, poultry, ice cream]

INTERPRETATION OF ASSOCIATION RULES OUTPUT

Row ID	D Support	D Confide...	D Lift	S Recommended_Product	S implies	S Set_Of_Products
rule59	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]
rule58	0.055	0.643	1.731	pasta	<---	[paper towels, eggs, ice cream]
rule21	0.051	0.674	1.726	cheeses	<---	[bagels, cereals, sandwich bags]

- In Rule 59, there is support that 5.5% orders have eggs, ice cream, pasta and paper towels. If eggs, ice cream and pasta are purchased in an order, then there is confidence of 64% that paper towels will also be purchased in the same order or and there is 79% chance that wherever eggs, ice cream and pasta are in an order, paper towels will also be purchased.
- In Rule 58, there is support that 5.5% orders have eggs, ice cream, pasta and paper towels. If eggs, ice cream and paper towels are purchased in an order, then there is confidence of 64% that pasta will also be purchased in the same order and there is 73% chance that wherever eggs, ice cream and paper towels are in an order, pasta will also be purchased.
- In Rule 21, there is support that 5.5% orders have bagels, cereals, sandwich bags and cheeses. If bagels, cereals, sandwich bags are purchased in an order, then there is confidence of 67 % that cheeses will also be purchased in the same order and there is 73% chance that wherever bagels, cereals, sandwich bags are in an order, cheeses will also be purchased.

RECOMMENDATIONS – KNIME ANALYSIS

Row ID	Support	Confide...	Lift	Recommended_Product	implies	Set_Of_Products
rule59	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]
rule58	0.055	0.643	1.731	pasta	<---	[paper towels, eggs, ice cream]

- Can roll out Combo offers for eggs, ice cream, pasta and paper towels or can offer 5% discount on Pasta if paper towels, eggs and ice cream are already purchased as pasta has a lower rate of order.

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule21	0.050921861	0.674418605	1.726208518	cheeses	<---	[bagels, cereals, sandwich bags]
rule20	0.050921861	0.610526316	1.659640749	sandwich bags	<---	[cheeses, bagels, cereals]

- Can roll out Combo offers for bagels, cereals, sandwich bags and cheeses or can offer 15% discount on sandwich bags if bagels, cereals and cheeses are already purchased as sandwich bags have a lower rate of order.

RECOMMENDATIONS – KNIME ANALYSIS

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule18	0.050921861	0.630434783	1.677722471	mixes	<---	[yogurt, poultry, aluminum foil]

- Can roll out Combo offers for yogurt, poultry, aluminium foil and mixes

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule40	0.051799824	0.641304348	1.648861517	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule24	0.050921861	0.630434783	1.620914712	dinner rolls	<---	[spaghetti sauce, poultry, cereals]

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule52	0.053555751	0.642105263	1.650920756	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]

- Can roll out Combo offers for dinner rolls, spaghetti sauce, poultry and ice cream/cereals/laundry detergent or can also provide free dinner rolls if spaghetti sauce, poultry and ice cream/cereals/laundry detergent are purchased. This can help improve the sale of spaghetti sauce as the other items are purchased regularly and spaghetti sauce would cover the cost of dinner rolls.

RECOMMENDATIONS – KNIME ANALYSIS

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule11	0.050043898	0.612903226	1.615964755	coffee/tea	<---	[yogurt, cheeses, cereals]

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule8	0.050043898	0.587628866	1.528103375	yogurt	<---	[cheeses, cereals, coffee/tea]

- Can roll out Breakfast Combo for coffee/tea, yogurt, cheeses and cereals during the weeks and months with lower sales. Also these products can be placed close to each other to increase the chance of being purchased together. Can offer 10% discount on Yogurt if cheeses, cereals and coffee/tea is purchased.

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule870	0.082528534	0.562874251	1.497929375	individual meals	<---	[sandwich loaves, lunch meat]

- Can roll out Lunch Meal Combo offers for individual meals, sandwich loaves and lunch meat.

RECOMMENDATIONS – KNIME ANALYSIS

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule49	0.051799824	0.634408602	1.627458103	eggs	<---	[paper towels, dinner rolls, pasta]

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule35	0.051799824	0.627659574	1.610144719	eggs	<---	[dinner rolls, poultry, soda]

- Can roll out Combo offers for eggs, paper towels, dinner rolls and pasta as seen in rule49 or eggs, dinner rolls, towels, poultry and soda as seen in rule35. Can also offer 10% discount on eggs if purchased with either of the two rules.

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule28	0.050921861	0.574257426	1.517585169	laundry detergent	<---	[poultry, milk, cereals]

- Can offer 10% discount on laundry detergent if poultry, milk and cereals are already purchased. This way, we can try and improve sales of laundry detergent using the combo of the other products. We can also offer all four products at a reasonable price.

RECOMMENDATIONS – KNIME ANALYSIS

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule909	0.082528534	0.513661202	1.470000275	sandwich loaves	<---	[cheeses, ketchup]

- Can roll out Combo offers for sandwich loaves, cheeses and ketchup while also placing them near each other.

Row ID	Support	Confidence	Lift	Recommended_Product	implies	Set_Of_Products
rule32	0.051799824	0.567307692	1.452052723	soda	<---	[eggs, dinner rolls, poultry]

- Can roll out Combo offers for soda, eggs, dinner rolls and poultry.

THANK YOU