Market Basket Analysis

Marketing & Retail Analytics

Prepared By
Dhruv Dosad

Agenda

1. **Executive Summary**
   - Problem Statement
2. **About data**
   - Data Characteristics
   - Exploratory data analysis
   - Summary & Insights
3. **About Market Basket**
   - What is Market Basket?
   - Approach
   - KNIME Workflow
4. **Associations**
   - Association Rule
   - Market Basket Recommendations
5. **Overall Recommendations**

Executive Summary

- **Sales Overview:**
  - Over a **2-year and 2-month period** starting from **01/01/2018.**
  - Total products sold: **15911** through **1,139** orders.
  - No data available for the **fourth quarter** in both 20218 & 2019..
- **Yearly Comparisons** (2018 & 2019):
  - In **2019,** a **decrease of 26** in the total number of **orders** compared to **2018**.
  - A **decrease of 155 units** in the total number of **products sold**.
- **Trend Analysis:**
  - **Consistent decreasing trend** in the number of orders.
  - A **mild decline** in the **total number of products** sold.
  - **Orders peaked** in **May** **month for both years**
  - A gradual **decline in sales** observed **after May**.
- **Product Popularity:**
  - Among 37 products:
    - **Most popular**: Poultry, ice cream, Cereals.
    - **Least popular**: Hand Soap, Sandwich Loaves, flour, pork.
- **Weekly Demand Trends:**
  - **Most products** show **decreasing demand** over time for all products except 1.
  - Only Yogurt shows a slight positive increase in demand.

Problem Statement

The Company's Data Challenge

A grocery store has entrusted us with their transactional data, seeking a solution to enhance their revenue-generation strategies. They are grappling with optimizing their customer offerings and need data-driven insights to overcome this hurdle.

Objective:

Our objective is to :

- **Analyse data**: Analyze POS data to identify common item combinations in customer orders.
- **Recommend**: Develop data-driven strategies for **popular combo offers and discounts**.
- **Increase Revenue**: Use insights to boost the grocery store's revenue through tailored customer incentives.

ABOUT DATA

About Data : Data Characteristics

Sample of 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
...	...	...	...
20636	2020-02-25	1138	soda
20637	2020-02-25	1138	paper towels
20638	2020-02-26	1139	soda
20639	2020-02-26	1139	laundry detergent
20640	2020-02-26	1139	shampoo

20641 rows × 3 columns

```
RangeIndex: 20641 entries, 0 to 20640
Data columns (total 3 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       20641 non-null  datetime64[ns]
 1   Order_id   20641 non-null  int64
 2   Product    20641 non-null  object
dtypes: datetime64[ns](1), int64(1), object(1)
```

Column	Dictionary
Date	Date of Order/Transaction
Order_id	Order ID
Product	Product purchased

- **Shape of the data**: The dataset contains **20641** rows and 3 columns.
- **Data types** : We have the columns with data type as datetime64(1), int64(1), object(1)
- Key column includes the Product & Order id

Note: The above data is before Exploratory data analysis & data cleaning

About Data : Data Cleaning

Duplicate Values

```
Total duplicate values: 4730
```

Duplicate Value check

	Date	Order_id	Product
4	2018-01-01	1	all- purpose
10	2018-01-01	1	all- purpose
11	2018-01-01	1	dinner rolls
13	2018-01-01	1	all- purpose
18	2018-01-01	1	dinner rolls
...	...	...	...
20632	2020-02-25	1138	sandwich bags
20633	2020-02-25	1138	toilet paper
20634	2020-02-25	1138	soda
20635	2020-02-25	1138	soda
20636	2020-02-25	1138	soda

8613 rows × 3 columns

Missing values

```
Date          0
Order_id      0
Product       0
dtype: int64
```

Insights

- **No Missing** values in the data
- **4730 duplicates** found, that were removed.
- Total Unique rows: **15911**
- No other anomalies found

About Data : Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max
Order_id	15911.0	574.15	328.54	1.0	289.5	579.0	859.0	1139.0

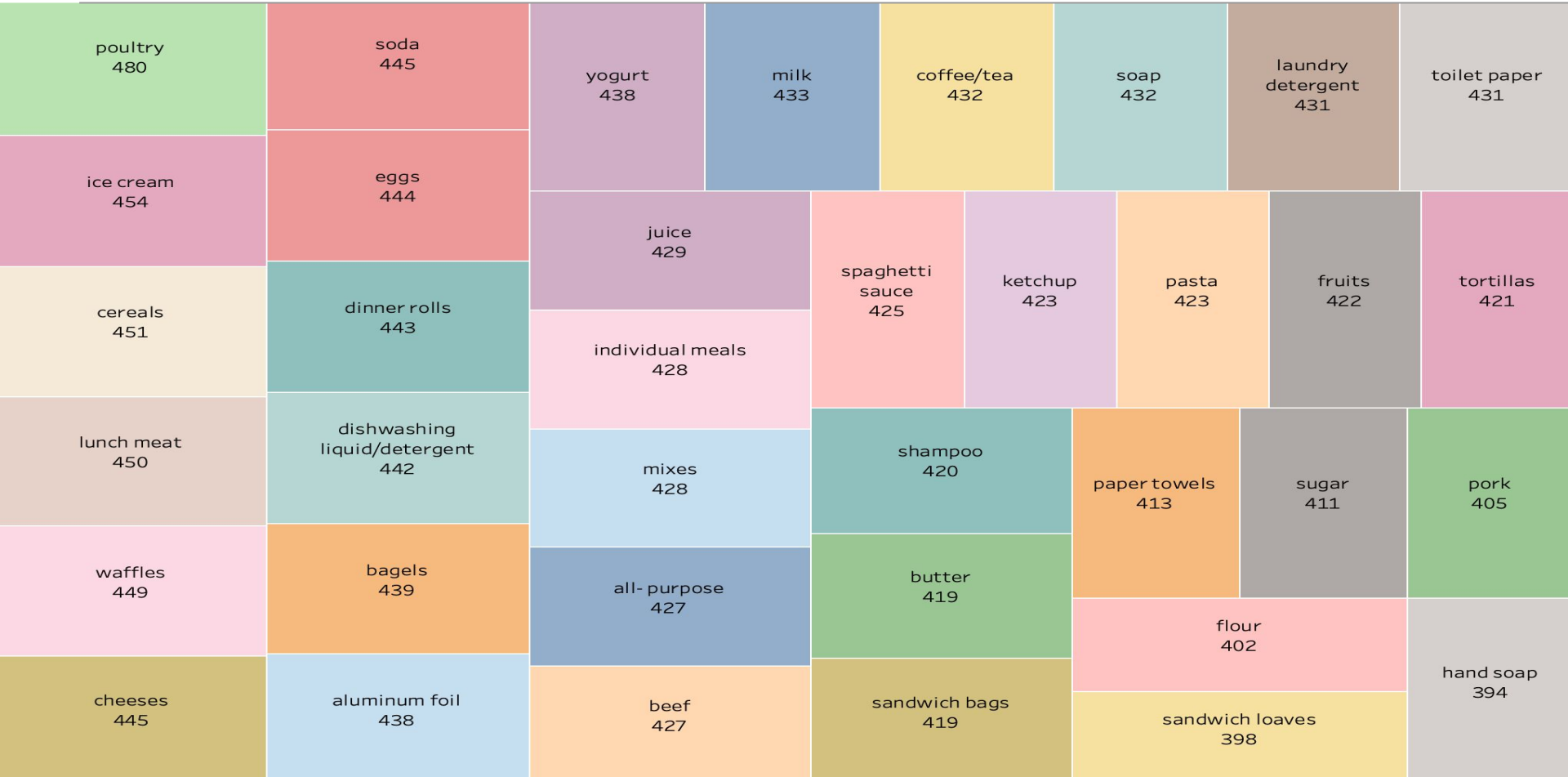
	count	unique	top	freq
Product	15911	37	poultry	480

- Total **1139 orders** containing **15911** items sold
- **37 unique products** with **Poultry** as **most purchased** product having **frequency 480**
- The data is ranging from **01/01/2018** to **26/02/2020** , approximately **2 Years & 2 Months**

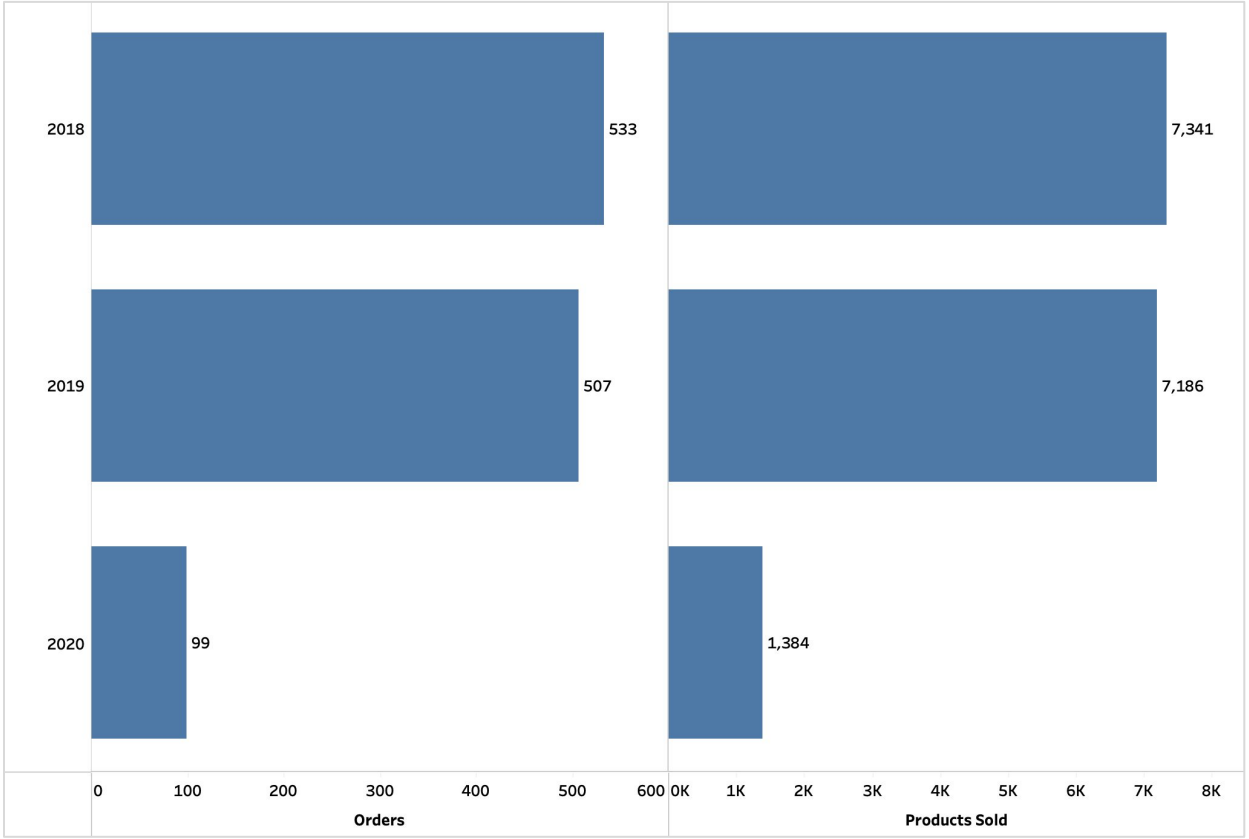
Note: The above data is before Exploratory data analysis & data cleaning

EXPLORATORY DATA ANALYSIS

Frequency of Products Sold



Yearly Orders & Products Sold

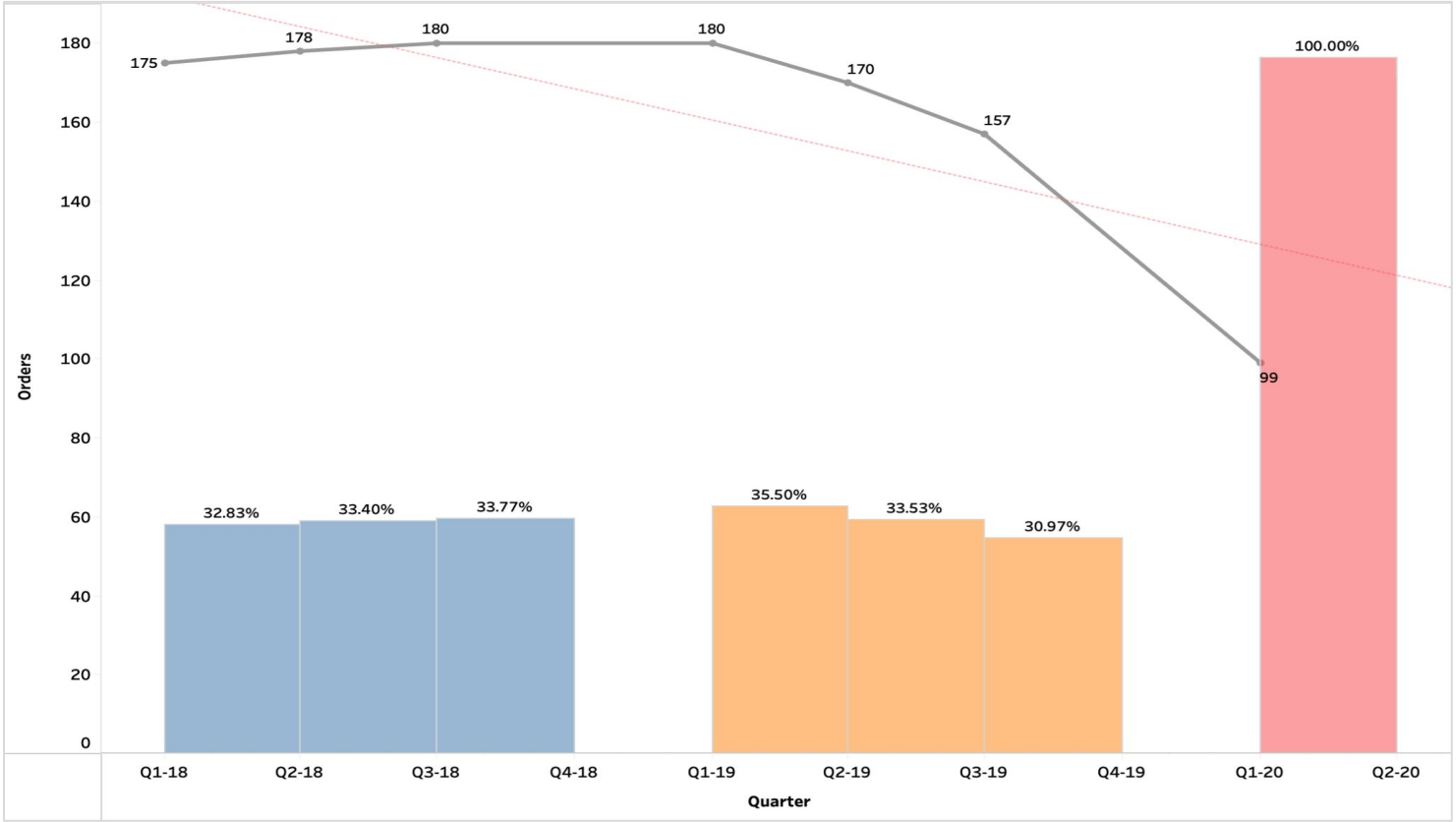


Insights

- **Decrease** in the **Total orders and count of products sold** in **2019 compared to 2018**
- The **Orders decreased** to **507** in **2019** as compared to **533 in 2018**
- Similarly the **Product sold** count also **reduced to 7186 in 2019 from 7341 in 2018.**

*2020 has not been considered as it has only 2 months of data

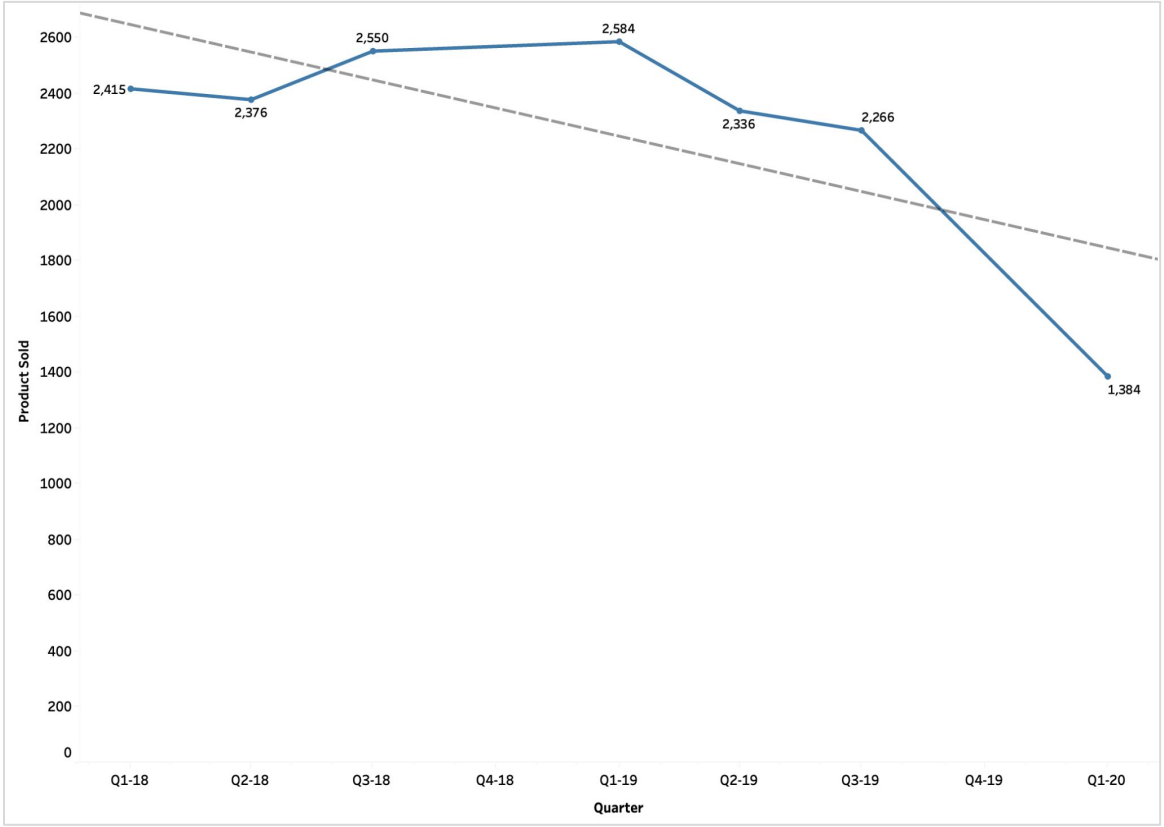
Orders Over Time: Quarterly Orders Placed

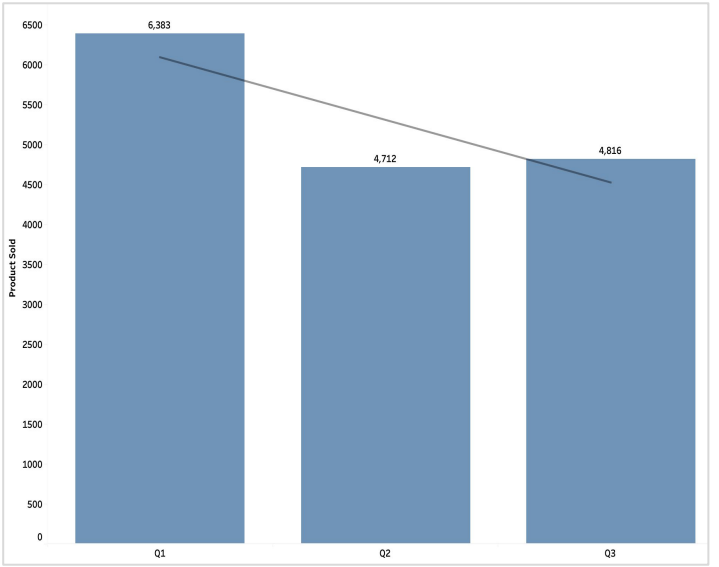


Insights

- The Order Trend shows **decrease in Sales** over time
- In **2018**, the **quarterly trend** was **slightly increasing,** however, in **2019, the sales decreased from Q1 to Q3**
- **No data found** for **quarter 4** in both **2018 & 2019**.This is critical & business need to check immediately for reason
- **For 2020, only Jan & Feb** month data available

Orders Over Time: Quarterly Products Sold

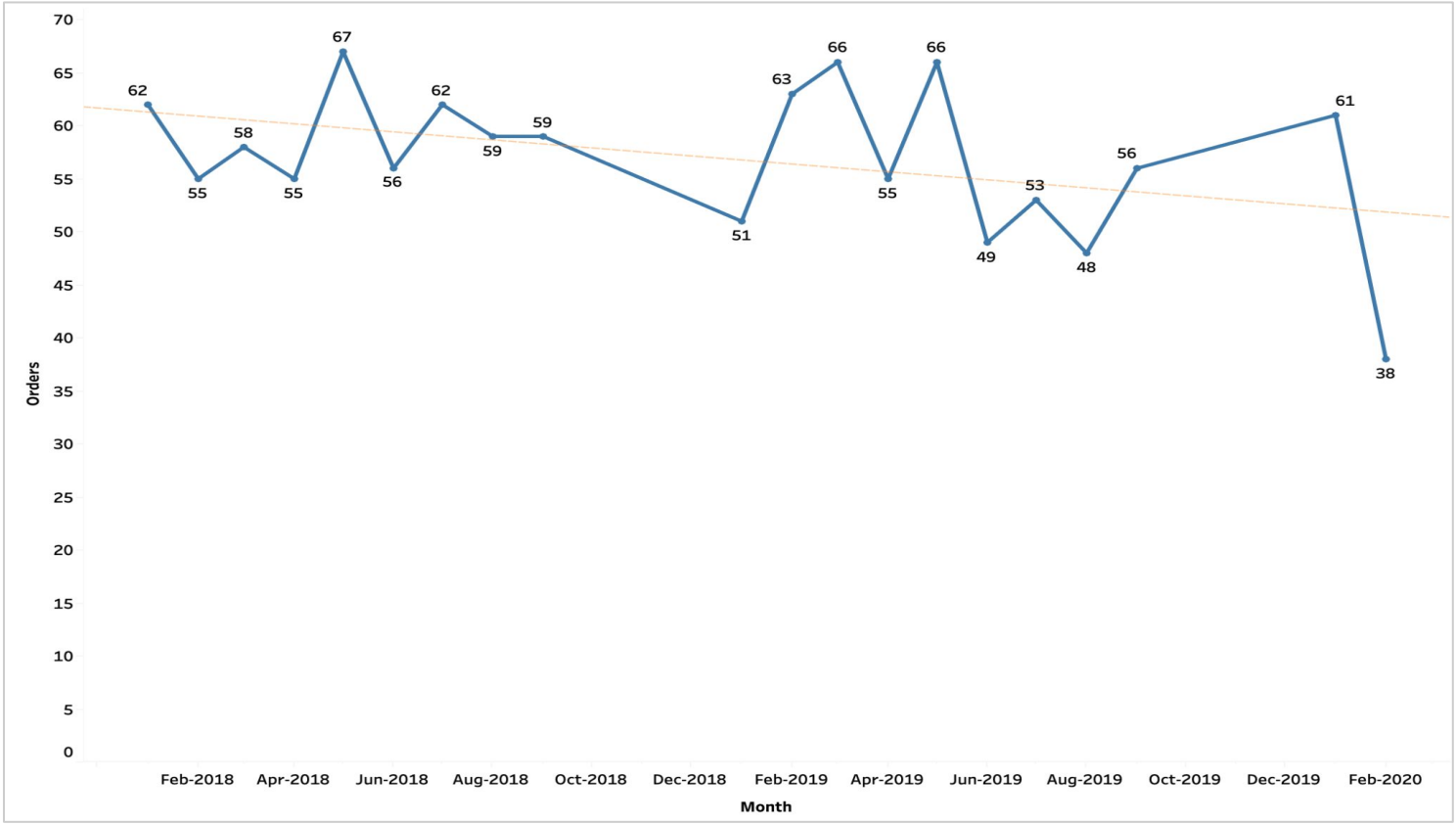




Insights

- The **overall Order Trend** is in **decreasing** Quarter on Quarter
- On the entire data, **Quarter 1** has the **highest sale** & **Quarter 2 has the lowest**
- **No data found** for **Q4** in both **2018 & 2019**

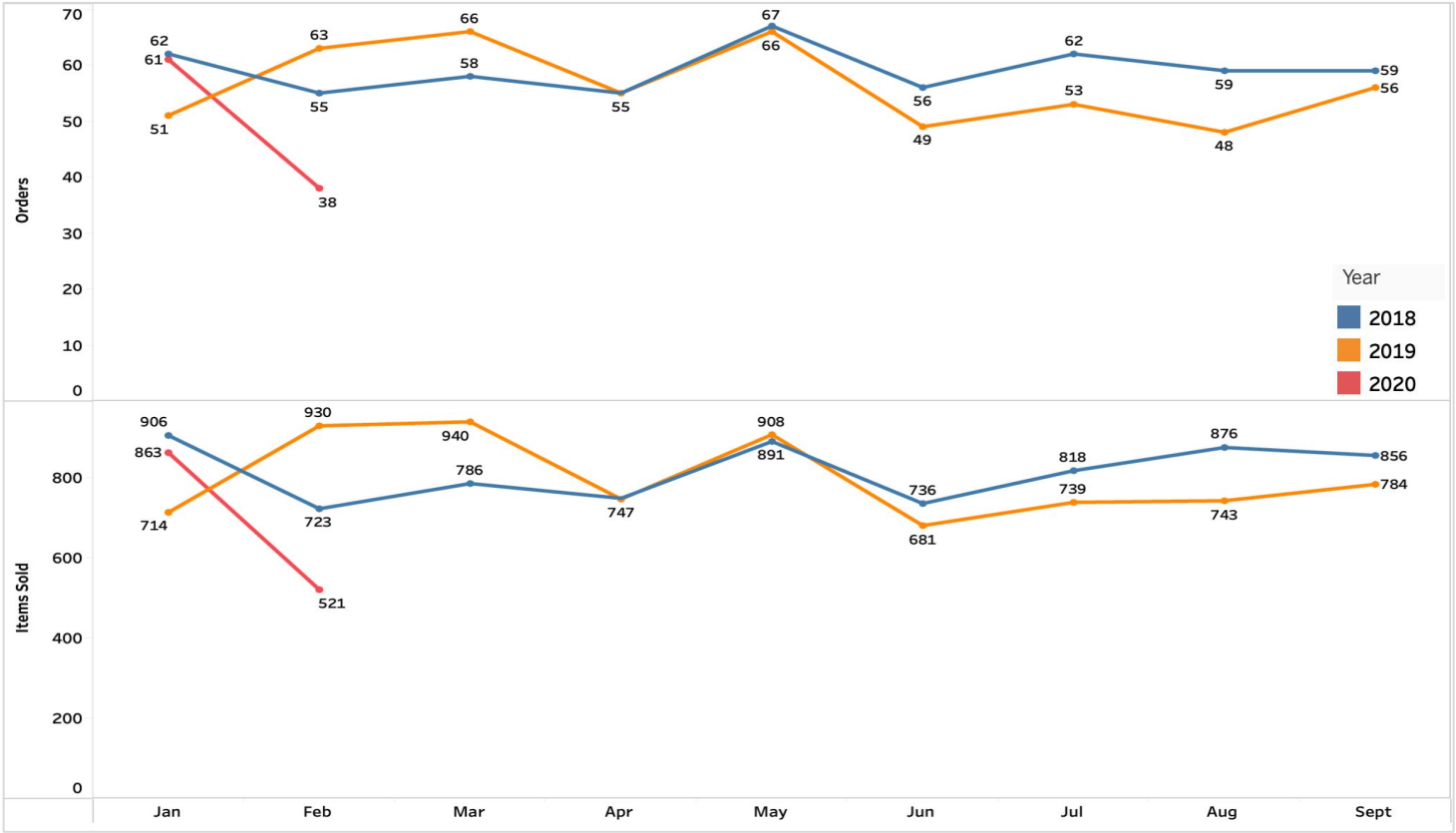
Orders Over Time: Monthly Order Trend



Insights

- **Highest sales** observed in **May 2018 (67 Orders)**.
- This is followed by **Mar 2019 & May 2019 (66 Orders each**

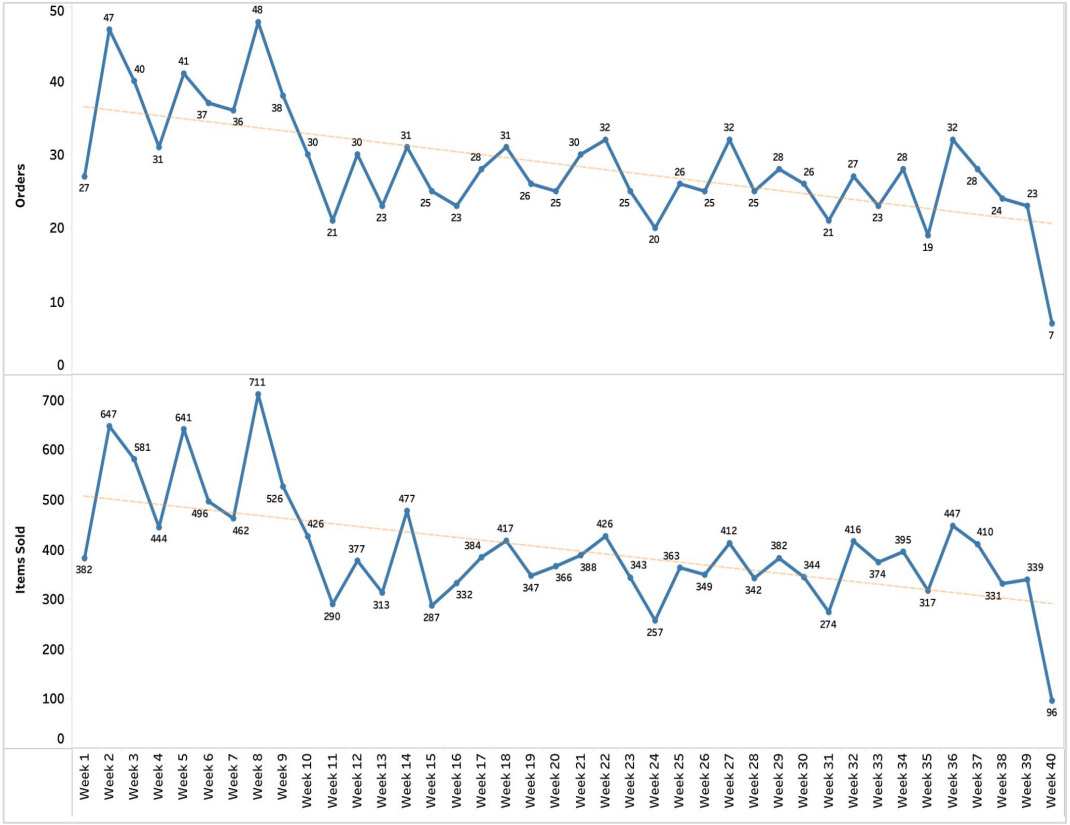
Orders Over Time: YOY Comparison

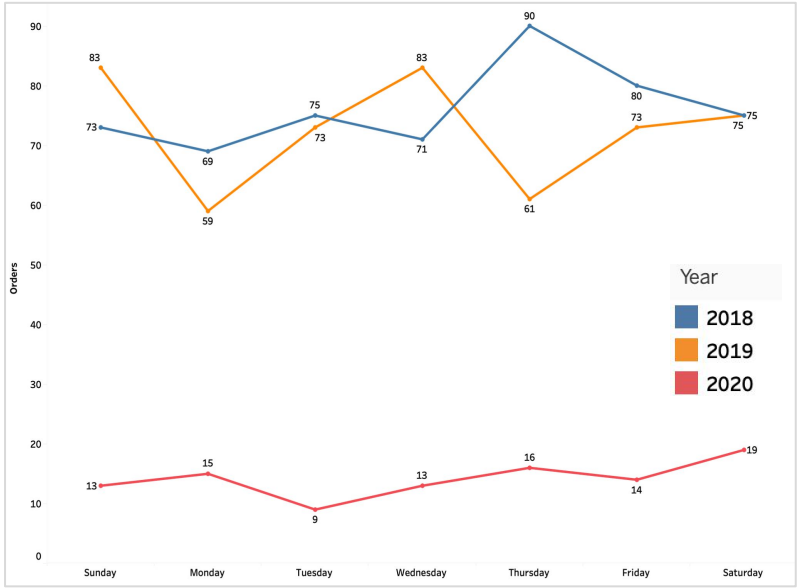


Insights

- Orders were highest in **May month** for both **2018 & 2019**
- Item sold hd a different peak than order count for both years.
- **Apr & June** shows the **dip** for orders.
- For 2020, only 2 months data available, However, Feb'20 shows huge dip

Orders Over Time: WOW Order Trend

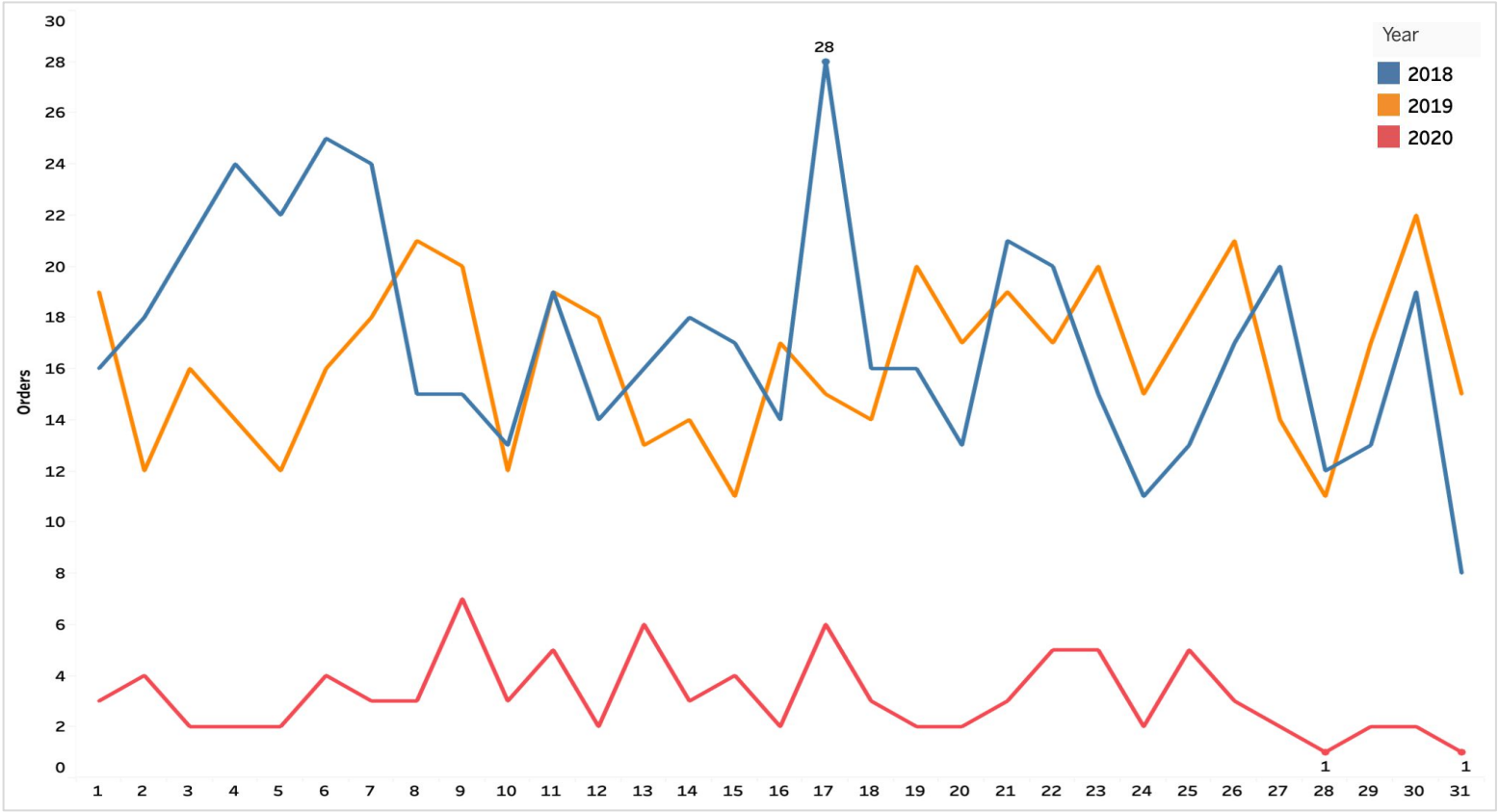




Insights

- The Week Wise Trend is **decreasing**
- **Highest sales** are in **Week 8.**
- No particular pattern found in basis Weekday
- For 2020, only 2 months data available

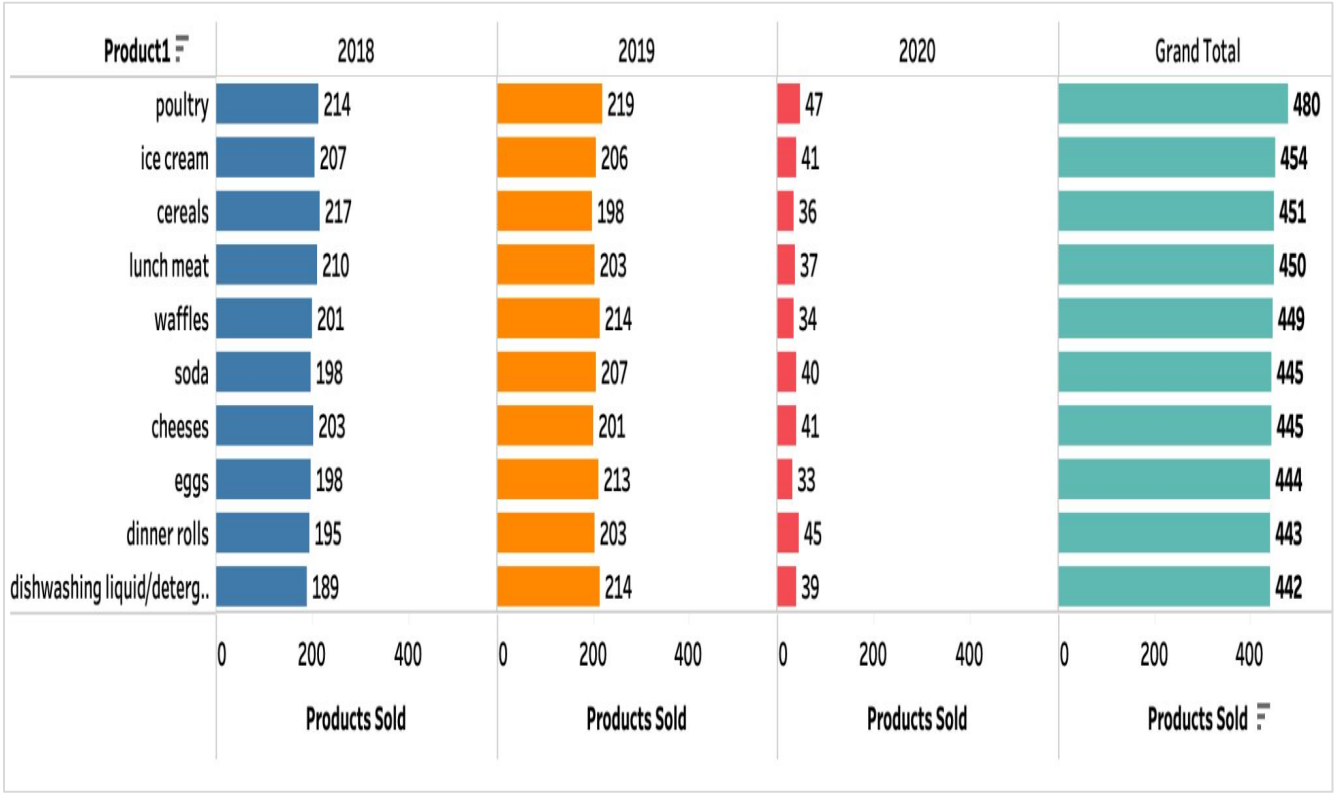
Orders Over Time: Daily Orders



Insights

- The order count on daily shows no specific pattern, it varies throughout the month

Top 10 Items Sold

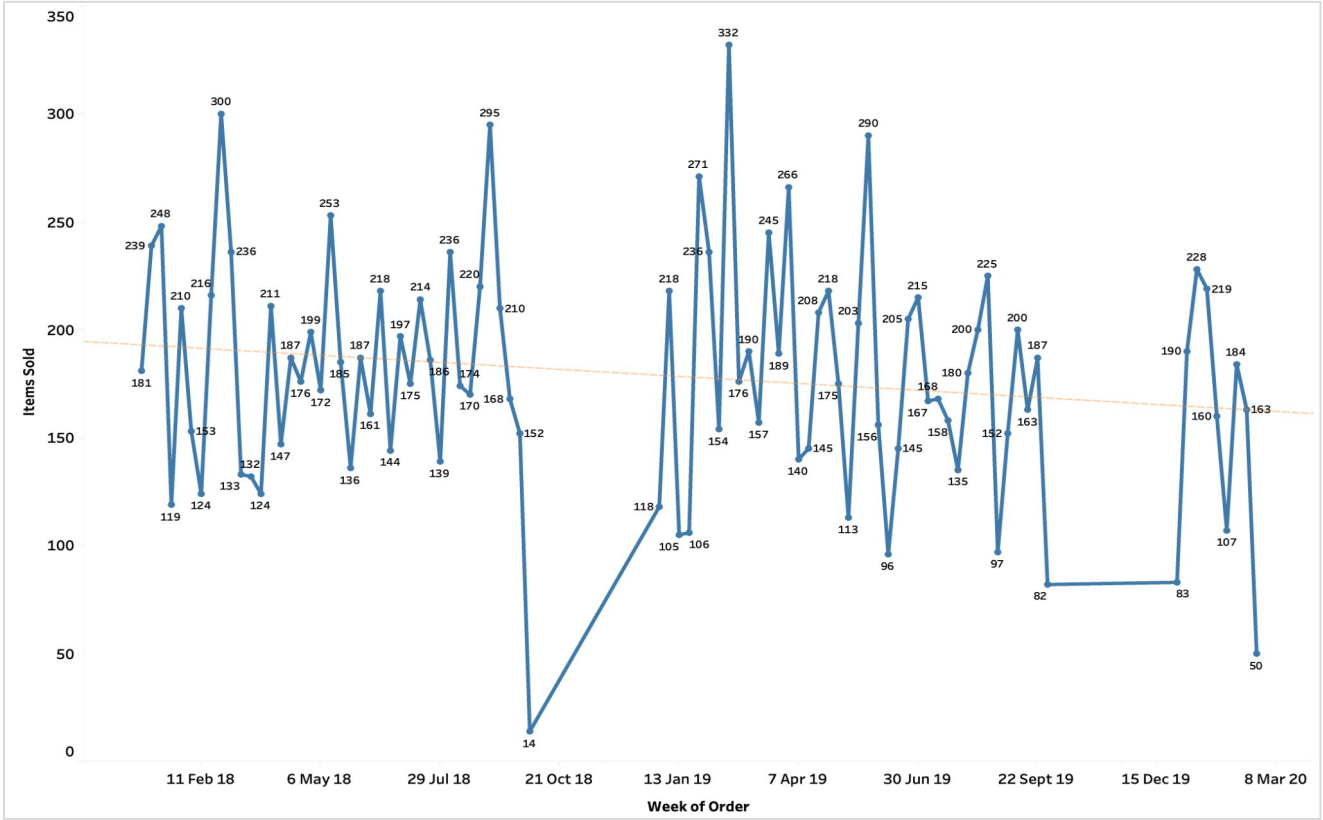


Insights

- **Poultry, ice cream, Cereals** are the most popular products.
- **Hand Soap, Sandwich Loaves, flour, pork** are the **least popular** products.

Top 10 is on Grand Total

Weekly Product Demand



Insights

- Product sold is in decreasing trend for most of the items.
- On Analysing the Trend line, coefficient of slope for all products are in negative.
- This suggests decreasing trend, **except Yogurt** which has a slight positive Slope coefficient .
- Hence, we can say that, except Yogurt, all products demand is **decreasing**

Top 10 is on Grand Total

Summary of Exploratory Analysis

Sales Overview:

- The data is, starting from January 1, 2018 having **Over a 2-year and 2-month period**. A total of **15,911 products** were sold through **1,139 orders**.
- Unfortunately, there's no data available for the fourth quarter in both years.

Yearly Comparisons:

- In 2019, the total number of orders **decreased by 26** compared to the previous year (2018).
- Additionally, the total number of products sold experienced a **decrease of 155 units in 2019** compared to 2018.

Trend Analysis:

- A **consistent decreasing** trend was observed in the number of orders over the analyzed period.
- There was a **mild decline in the total number of products sold**, although not as pronounced.
- A gradual decline was noted in sales and orders **after May**.

Product Popularity

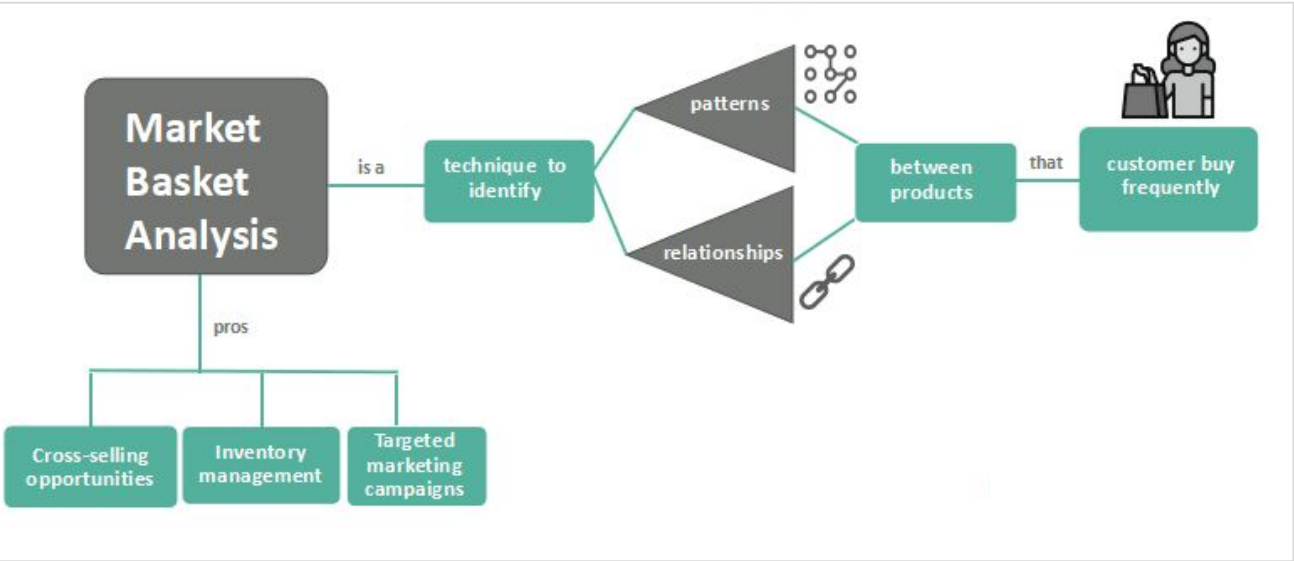
- Among **37 products**, clear preferences emerged.
- **Poultry, ice cream, and cereals** were the **most popular** choices among customers.
- **Conversely, hand soap, sandwich loaves, flour, and pork** ranked as the **least popular** products during this period.

Weekly Demand Trends:

- Weekly demand trends for most products displayed a **consistent decrease** over time.
- One product, yogurt, showed a slight positive increase in demand.

MARKET BASKET ANALYSIS

Market Basket Analysis : What is it?



Applications:

- **Recommendation Systems**: Enhance customer experience by suggesting complementary products.
- **Personalized Shopping Assistant**: Tailor product recommendations based on individual preferences.

Benefits

- **Enhances Customer Experience**: Provide personalized product recommendations, enhancing satisfaction.
- **Saves Customer Time:** Streamline the shopping process with relevant suggestions.
- **Increases Customer Engagement:** Encourage additional purchases and customer loyalty.
- **Expands Product Exposure:** Introduce customers to a broader range of offerings.
- **Boosts Profitability:** Drive more sales and revenue through data-driven insights

** Image used from internet*

Market Basket Analysis : How is it Calculated?

- The first step in Market Basket Analysis is the formation of two sets of products for all product combinations: -
  - Set A being the Group of products
  - Set B being the recommended product, when Set A is purchased.
- The Rule being: If people buy Set A, they also buy Set B.

- After formation of sets - 2 parameters namely Support & Confidence are computed for each product combinations.
- Support tells us about the popularity of the products.
- Support is computed for Set A and higher the value of Support means more the number of times the products in Set A appear together in multiple orders.

  Support = No. of Baskets Set A appears in/ Total No. of Baskets

- After computing support, we compute the confidence.

- Confidence is the conditional probability that a customer will purchase the recommended product in Set B provided they have already purchased or are about to purchase the group of products in Set A.

  Confidence = No. of Baskets Set (A&B) appear in / No. of Baskets Set A appears in

- A threshold value for both Support & Confidence is then decided iteratively. Lower the value of these thresholds, More will be the volume of the association rules but the accuracy will be low. We need both an adequate volume of association rules & a better accuracy.

Market Basket Analysis : How is it Calculated?

- Lower the Support, more will be the volume of different combination of products in Set A & Lower the confidence, more will be the volume of recommended products.
- The effectiveness of the association rules are then computed by calculating the Lift.
- Lift is the conditional probability that tells us about the lift in the probability of purchasing the recommended product in Set B provided someone has already purchased or is about to purchase the group of products in Set A.

$$\textbf{Lift} = \textbf{Confidence} / \textbf{Support of Set B}$$

$$\Rightarrow \textbf{Lift} = \frac{(\textbf{No.of Baskets Set (A\&B)appear in} / \textbf{No.of Baskets Set A appears in})}{(\textbf{No.of Baskets Set B appears in} / \textbf{Total No.of Baskets})}$$

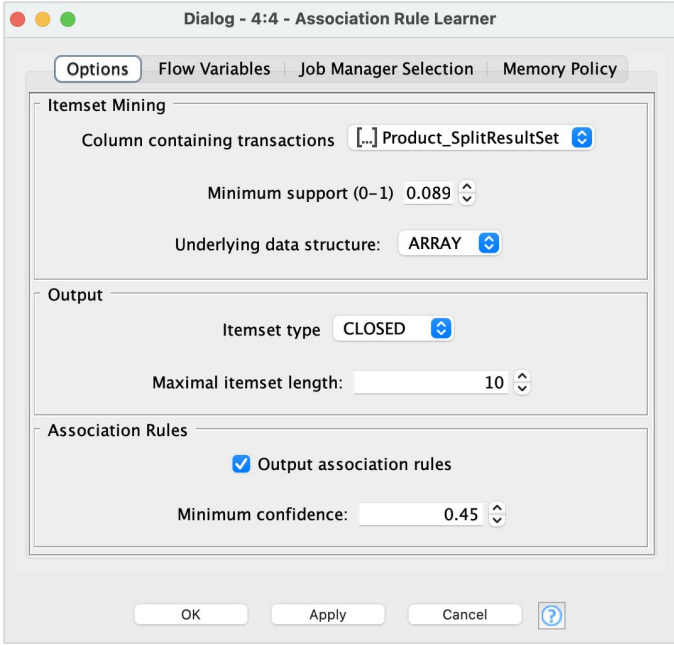
- In simpler context, if a customer purchases the group of products in Set A, what is the probability that the customer purchasing the recommended product in Set B will increase.

Market Basket Analysis : Workflow

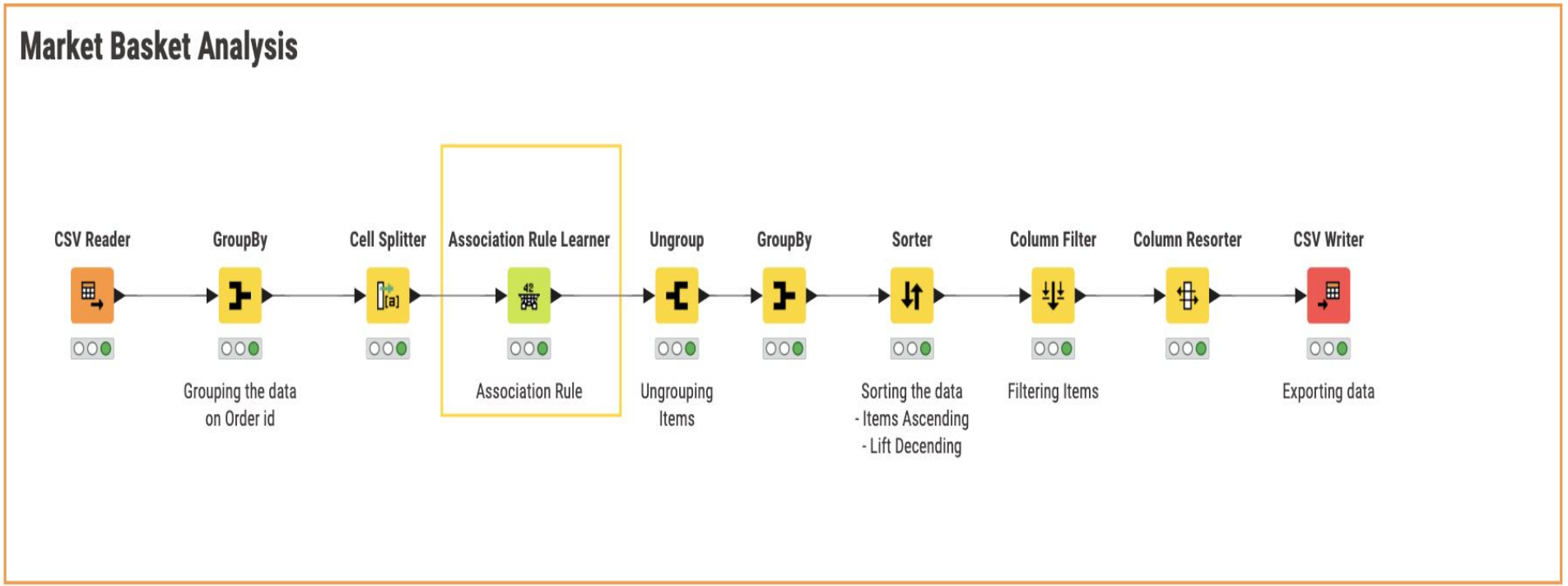
- **Data Grouping**: Initially, transactions are grouped by Order ID to organize the dataset effectively.
- **Product Sets**: Within each order, duplicate product transactions are eliminated, resulting in sets of unique products purchased.
- **Association Rules**: Our Association Rule Learner discovers relationships among products based on user-defined thresholds for support and confidence.
- **Threshold Balancing**: Finding the right balance between data volume and accuracy is essential. Lower thresholds yield more rules but lower accuracy. We iteratively fine-tune these thresholds based on business requirements.
- **Threshold Values**: Currently, we employ minimum support (0.089) and minimum confidence (0.45), generating 165 association rules for 37 unique products. Adjusting these thresholds allows us to tailor the analysis for more data volume or increased accuracy.
- **Association**: The associations identified are then evaluated by computing the lift. Higher the Lift more will be the probability of purchasing the recommended products.

The results from the analysis are then used to design appropriate marketing strategies.

Association Rules Parameters



Market Basket Analysis : KNIME WORKFLOW



Associations

Associations : KNIME Output

#	Row...	Consequent *String*	implies *String*	Items (#1) *String*	Support *Number (double)*	Confidence *Number (double)*	Lift *Number (double)*
1	Row85	poultry	<---	all- purpose	0.176	0.468	1.111
2	Row...	yogurt	<---	aluminum foil	0.177	0.461	1.199
3	Row49	ice cream	<---	aluminum foil	0.176	0.459	1.151
4	Row86	poultry	<---	aluminum foil	0.176	0.457	1.084
5	Row87	poultry	<---	beef	0.17	0.454	1.078
6	Row88	poultry	<---	butter	0.166	0.451	1.07
7	Row89	poultry	<---	cereals	0.181	0.457	1.084
8	Row50	ice cream	<---	cheeses, aluminum foil	0.09	0.534	1.339
9	Row51	ice cream	<---	cheeses	0.179	0.458	1.15
10	Row90	poultry	<---	cheeses	0.181	0.463	1.098
11	Row...	waffles	<---	coffee/tea	0.172	0.454	1.151
12	Row91	poultry	<---	coffee/tea	0.175	0.461	1.093
13	Row92	poultry	<---	dinner rolls, cereals	0.092	0.538	1.278
14	Row93	poultry	<---	dinner rolls, laundry detergent	0.09	0.543	1.287
15	Row94	poultry	<---	dinner rolls, lunch meat	0.091	0.562	1.334
16	Row95	poultry	<---	dinner rolls, mixes	0.09	0.557	1.321
17	Row38	eggs	<---	dinner rolls, pasta	0.091	0.528	1.354
18	Row...	soda	<---	dinner rolls, pasta	0.09	0.523	1.338
19	Row...	spaghetti sauce	<---	dinner rolls, poultry	0.099	0.509	1.364

Strongest Associations : KNIME Output

#	Row...	Consequent *String*	implies *String*	Items (#1) *String*	Support *Number (double)*	Confidence *Number (double)*	Lift ↓ *Number (double)*
141	Row30	dinner rolls	<---	spaghetti sauce, poult...	0.099	0.579	1.49
101	Row60	juice	<---	poultry, aluminum foil	0.096	0.545	1.447
26	Row81	pasta	<---	dinner rolls, soda	0.09	0.528	1.422
27	Row40	eggs	<---	dinner rolls, soda	0.095	0.554	1.421
47	Row...	soda	<---	eggs, soap	0.092	0.553	1.414
39	Row82	pasta	<---	eggs, dinner rolls	0.091	0.525	1.414
109	Row4	aluminum foil	<---	poultry, juice	0.096	0.542	1.41
68	Row...	yogurt	<---	juice, aluminum foil	0.093	0.541	1.406
40	Row...	soda	<---	eggs, dinner rolls	0.095	0.545	1.396
157	Row64	juice	<---	yogurt, aluminum foil	0.093	0.525	1.393
43	Row18	dinner rolls	<---	eggs, pasta	0.091	0.539	1.385
42	Row...	soda	<---	eggs, ice cream	0.09	0.537	1.374
94	Row21	dinner rolls	<---	pasta, soda	0.09	0.534	1.372
158	Row6	aluminum foil	<---	yogurt, juice	0.093	0.527	1.371
102	Row79	mixes	<---	poultry, aluminum foil	0.09	0.515	1.371
28	Row96	poultry	<---	dinner rolls, spaghetti ...	0.099	0.577	1.368
19	Row...	spaghetti sauce	<---	dinner rolls, poultry	0.099	0.509	1.364
63	Row41	eggs	<---	ice cream, soda	0.09	0.531	1.363
64	Row...	waffles	<---	ice cream, soda	0.09	0.536	1.361

Associations

The generated association rules serve as a recommendation system during customer shopping experiences.

Recommendation Logic: If a customer shows interest in or has items from Set A in their cart, they will be recommended products from Set B.
Priority Order: When multiple products (consequents) are associated with an item in Set A, the recommendation order is determined by the lift value. The product with the highest lift is recommended first, followed by others.
Example: As an illustration, consider the association rules pertaining to Yogurt

Item	Implies	Consequent	Support	Confidence	Lift
Yogurt	=>	Juice	0.176	0.459	1.218
Yogurt	=>	Aluminum foil	0.177	0.461	1.199
Yogurt	=>	Eggs	0.175	0.454	1.166
Yogurt	=>	Waffles	0.174	0.452	1.147
Yogurt	=>	Poultry	0.181	0.470	1.116

- If a customer purchases Yogurt, the first recommendation will be of Juice as it has a higher probability of being purchased along with Juice based on past data (higher lift).
- Then Aluminum Foil, Eggs in that order.

Associations

Utilizing Associations: The derived associations also facilitate the creation of product combinations frequently purchased together. However, it's crucial to consider product types before forming these combinations.

Recommendation Logic: If customers exhibit interest in or have items from Set A in their cart, they will receive recommendations for product combinations that include Set A and the Consequent product from Set B.

Marketing Opportunities: These product combinations can be bundled with enticing offers and incorporated into various marketing strategies to boost the sales of both associated products.

Example: For instance, consider the scenario where a customer's cart includes items from Set A, opening the door to potential offers and strategies

Item	Implies	Consequent	Support	Confidence	Lift
Spaghetti sauce, Poultry	=>	Dinner rolls	0.099	0.579	1.490
Dinner rolls, Poultry	=>	Spaghetti sauce	0.099	0.509	1.364
Dinner rolls	=>	Poultry	0.195	0.501	1.189
Poultry	=>	Dinner rolls	0.195	0.463	1.189
Spaghetti sauce	=>	Dinner rolls	0.172	0.461	1.186
Spaghetti sauce	=>	Poultry	0.171	0.459	1.089

- Basis the association, the store can create a combo offer for Dinner Rolls, Poultry & Spaghetti Sauce.

Recommendations & Combos

Recommendation - 1

Item	Implies	Consequent	Support	Confidence	Lift
Spaghetti sauce, Poultry	=>	Dinner rolls	0.099	0.579	1.490
Dinner rolls, Poultry	=>	Spaghetti sauce	0.099	0.509	1.364
Dinner rolls	=>	Poultry	0.195	0.501	1.189
Poultry	=>	Dinner rolls	0.195	0.463	1.189
Spaghetti sauce	=>	Dinner rolls	0.172	0.461	1.186
Spaghetti sauce	=>	Poultry	0.171	0.459	1.089

- **Smart Pairing**: **Combine Dinner Rolls, Spaghetti Sauce, and Poultry** to create attractive product bundles. **Offer** discounts for customers **buying all three or a "buy 2 get 1 free" deal**.
  - **For example**: purchasing Dinner Rolls and Poultry would include Spaghetti Sauce at no extra cost.
  -
- **Demand Insights**: While Spaghetti Sauce demand is declining, Poultry and Dinner Rolls show steady or slightly increasing weekly demands, with Poultry being the top seller.
-
- **Boost Sales**: **Pairing Spaghetti Sauce with Poultry and Dinner Rolls** can rejuvenate Spaghetti **Sauce sales**. This approach not only drives sauce sales but also encourages customers to choose the combo, **boosting overall sales** and customer satisfaction.

Recommendation - 2

Item	Implies	Consequent	Support	Confidence	Lift
Juice, Aluminum Foil	=>	Yogurt	0.093	0.541	1.406
Yogurt, Aluminum Foil	=>	Juice	0.093	0.525	1.393
Yogurt, Juice	=>	Aluminum Foil	0.093	0.527	1.371
Juice	=>	Yogurt	0.176	0.469	1.218
Yogurt	=>	Juice	0.176	0.459	1.218
Aluminum Foil	=>	Yogurt	0.177	0.461	1.199
Yogurt	=>	Aluminum Foil	0.177	0.461	1.199
Juice	=>	Aluminum Foil	0.172	0.457	1.188

- **Strategic Bundles**: Consider **bundling Yogurt, Juice, and Aluminum Foil** to **provide cost-effective** options. Implement combo deals where customers purchasing these items together enjoy discounts or a "**buy 2 get 1" offer.**
    - **For example**, buying Yogurt or Juice with Aluminum Foil includes an extra Yogurt or Juice for free.
- **Demand Synergy**: With **increasing weekly demand for Yogurt**, these bundles can significantly **boost Juice** and **Aluminum Foil sales**. Leveraging Yogurt's popularity will positively impact these complementary products.

Recommendation - 3

Item	Implies	Consequent	Support	Confidence	Lift
Ice Cream, Soda	=>	Waffles	0.090	0.536	1.361
Ice Cream, Waffles	=>	Soda	0.090	0.523	1.338
Waffles, Soda	=>	Ice Cream	0.090	0.510	1.279
Soda	=>	Waffles	0.177	0.454	1.152

- **Combo Deals**: Explore **bundled offers** for **Ice Cream, Soda, and Waffles**, allowing customers to purchase these products **together at a reduced price**. Consider implementing a "**buy 2 get 1" promotion,** where buying **Ice Cream and Soda includes a complimentary pack of Waffles**.
- **Demand Boost**: While Waffles face decreasing demand, Ice Cream and Sodas show steady to slightly increasing weekly demand. By pairing them together, you can **revitalize Waffles sales** and entice customers to opt for the combo.

Recommendations - Summary

Boost Sales with Smart Strategies:

- **Combo Deals**: Offer discounted schemes for product combinations:
    - Dinner Rolls, Spaghetti Sauce & Poultry
    - Yogurt, Juice & Aluminum Foil
    - Ice Cream, Soda & Waffles
- **Promotional Sales**: Implement frequent sale offers for slower-selling products such as Hand Soap, Sandwich Loaves, and fruits.
- **Leverage Associations**: Explore additional product associations to enhance sales:
    - Soda & Eggs
    - Dinner Rolls & Eggs
    - Ice Cream & Cheeses
    - Yogurt & Poultry
    - Lunch Meat & Poultry
- Consider bundled offers for these product pairs, providing customers with a discounted rate when purchasing them together. This strategy can boost sales for these products and create a mutually beneficial sales impact.

Thank You