PGP-DSBA PROJECT REPORT

MRA – MAIN PROJECT
PART 2

BY
ISHAAN SHAKTI JAYARAMAN
PGPDSBA.O.JULY24.A

AGENDA

- PROBLEM STATEMENT
- DATA CONTENTS
- DATA DESCRIPTION
- EXPLORATORY DATA ANALYSIS
  - UNIVARIATE ANALYSIS
  - BIVARIATE ANALYSIS
  - MULTIVARIATE ANALYSIS
- MARKET BASKET ANALYSIS
  - MARKET BASKET ANALYSIS
  - ASSOCIATION RULES
  - KNIME WORKFLOW
  - FINAL OUTPUT & EXPLANATION OF VALUES
- INFERENCES & RECOMMENDATIONS

PROBLEM STATEMENT

In the competitive grocery retail industry, identifying frequently purchased item combinations is essential for boosting sales and customer satisfaction. By analyzing Point of Sale (POS) data using association rule mining, businesses can uncover key purchasing patterns. These insights enable stores to create targeted combo offers and discounts, increasing basket size and revenue. Aligning promotions with customer preferences enhances retention and profitability. Data-driven strategies also help optimize inventory and reduce operational inefficiencies.

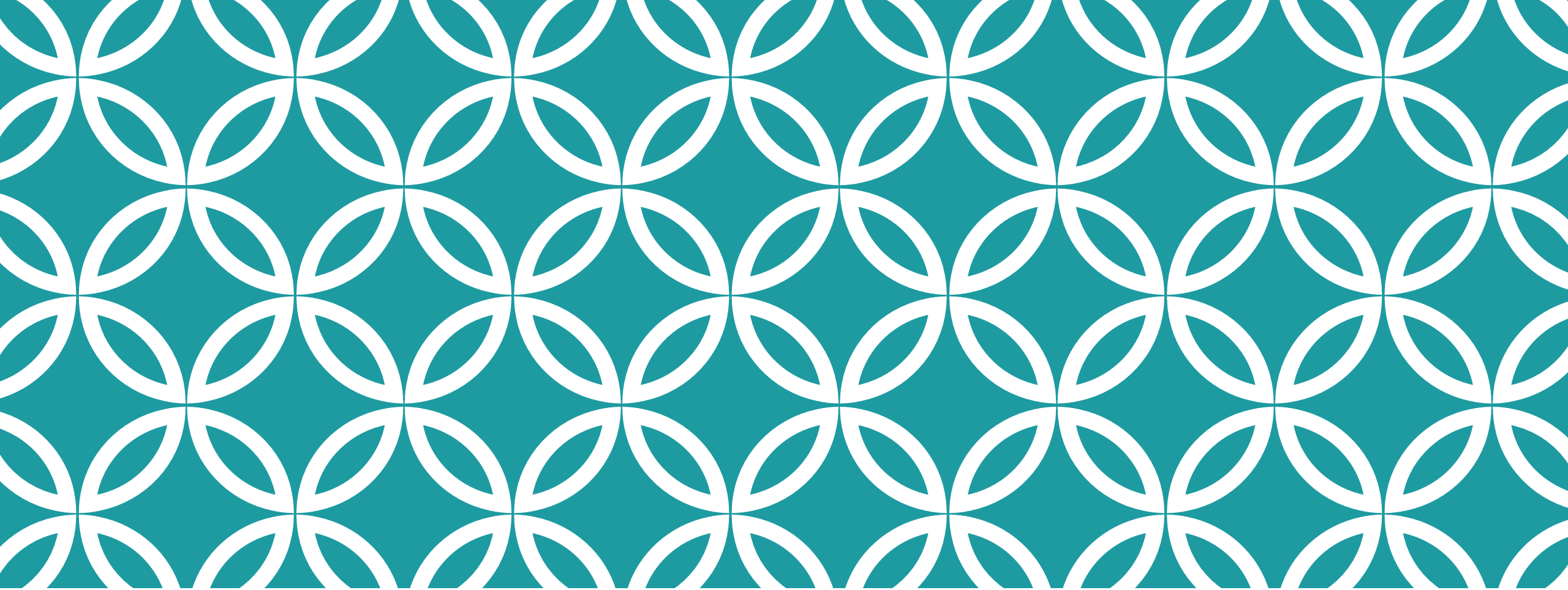
DATA CONTENTS

The dataset (dataset_group.csv) contains transactional data from a grocery store, where each row represents a product purchased in a specific order..

- There are 20641 observations in the dataset
- There are 3 columns of which 1 is a numerical variable, 2 are categorical variables.
- There are no null values in the dataset.
- There are 4730 duplicate entries in the dataset however these are duplicate entries of Order ID with each containing different product purchased, There is no need to delete these duplicate entries.
- The data of orders ranges from 01/01/18 to 26/02/20.

DATA DESCRIPTION

- **Date:** The date when the transaction took place.
- **Order_id:** A unique identifier for each customer order.
- **Product:** The individual item purchased in the transaction.

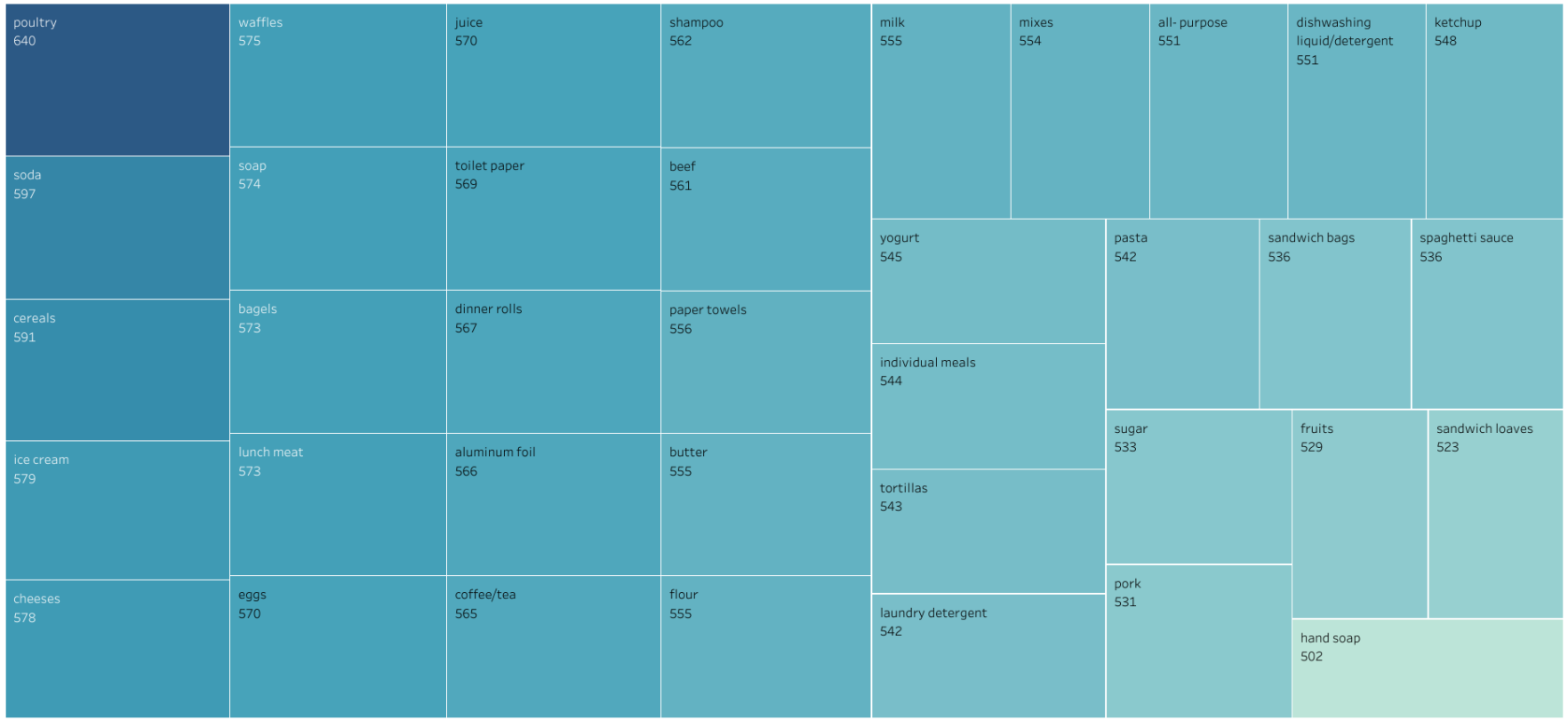


EXPLORATORY DATA ANALYSIS

Univariate analysis
Bivariate Analysis
Multivariate Analysis

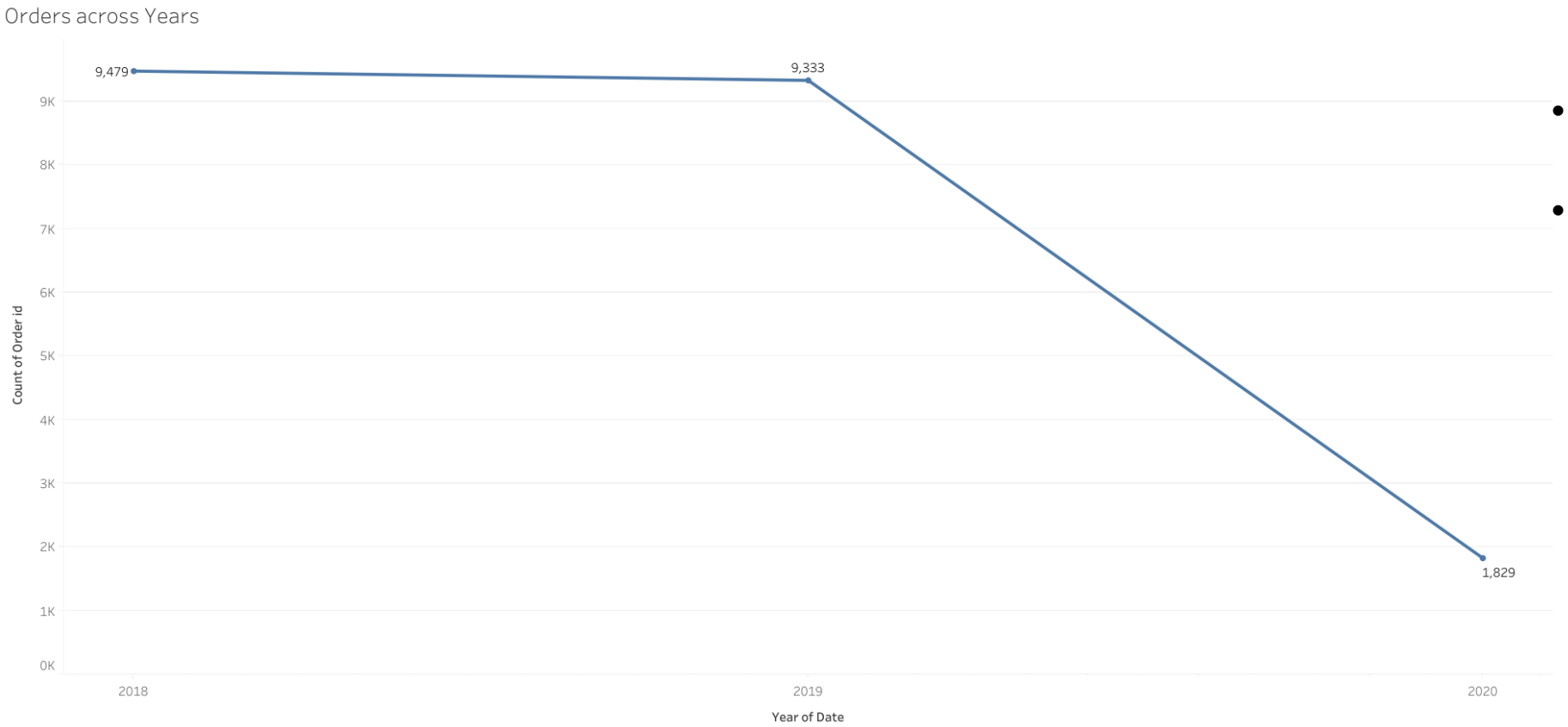
PRODUCT TREEMAP

Product Treemap



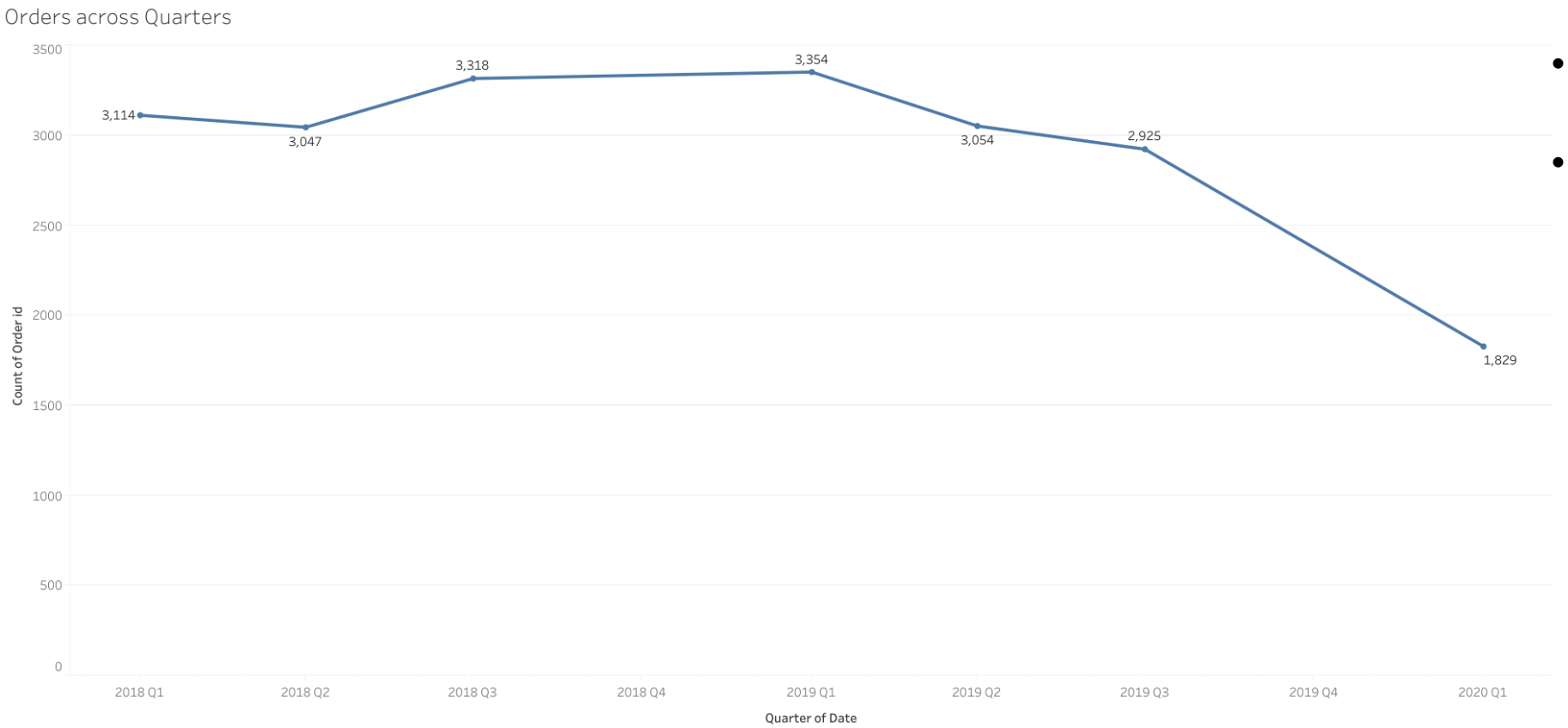
- Poultry is the most ordered product followed by soda, cereals, ice cream and cheeses.
- The distribution of all products ordered are slightly uniform.
- The lowest ordered product is hand soap.

ORDERS ACROSS YEARS



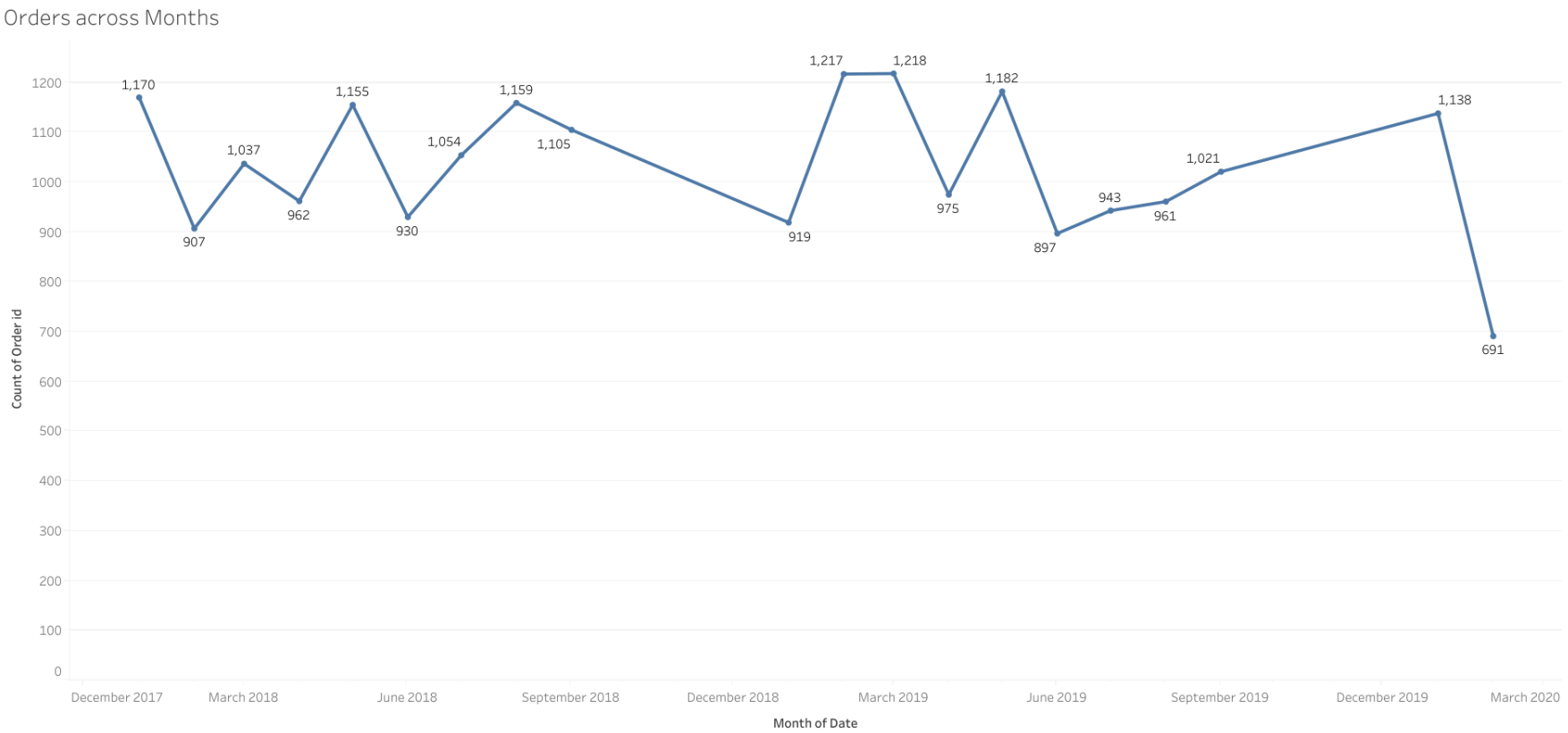
- Orders from the year 2018 to 2019 have remained stable.
- There is a decline in the year 2020, but this is due to only partial data present for the entire year.

ORDERS ACROSS QUARTERS



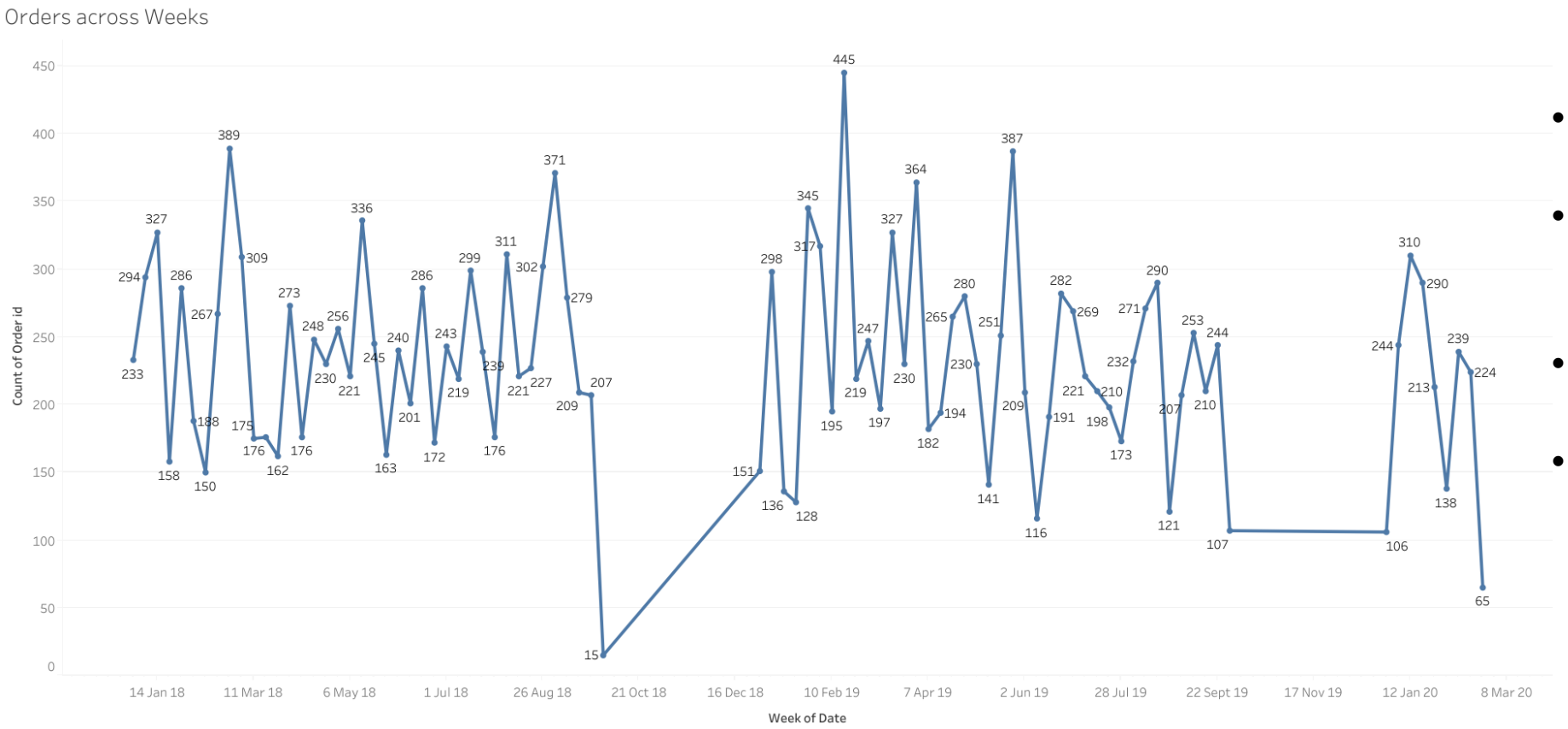
- 2019 saw decreases in sales for Q3 and Q4 compared to 2018.
- Q1 and Q2 sales for 2019 increased compared to 2018.

ORDERS ACROSS MONTHS



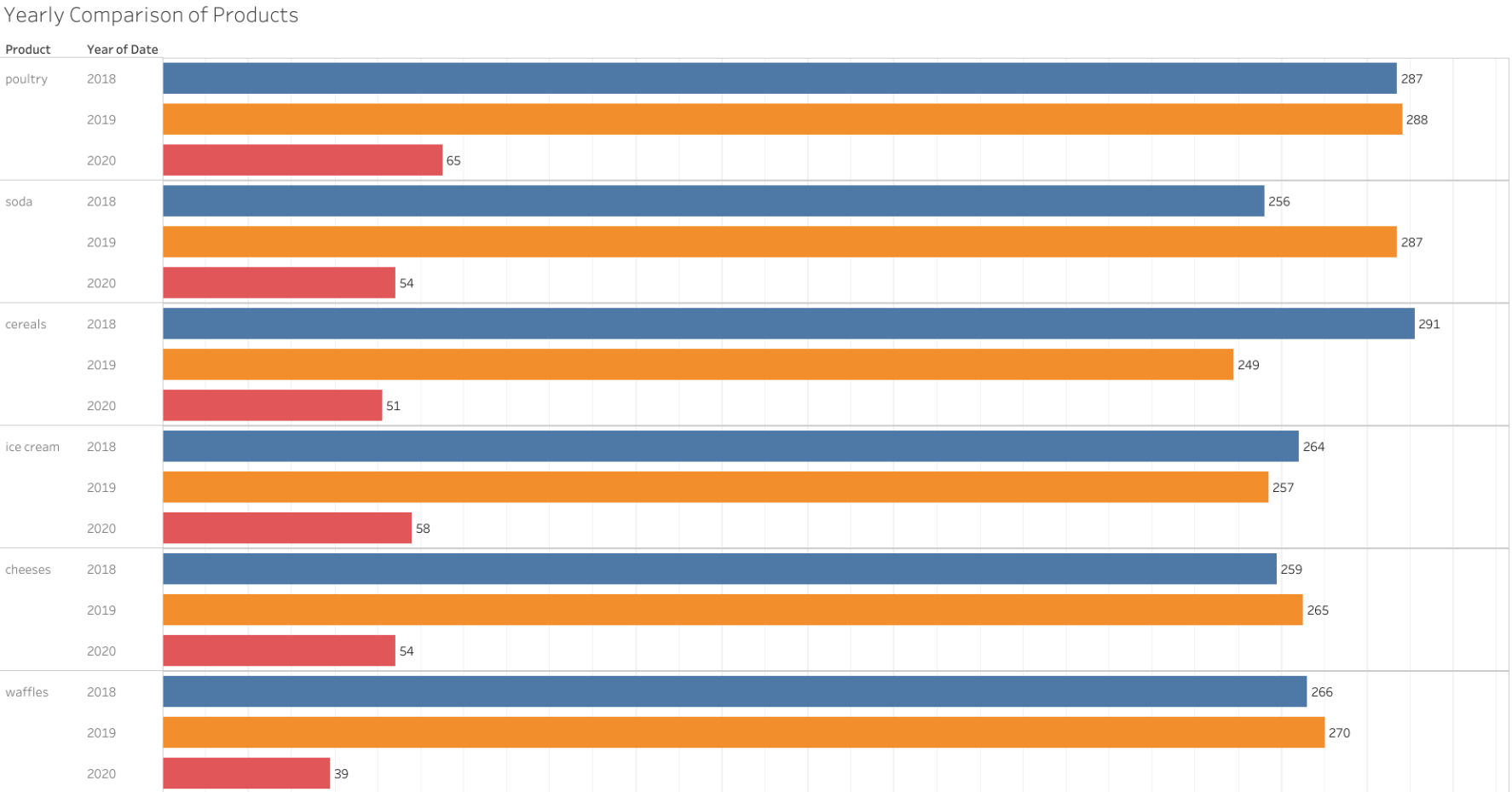
- There are increase in sales during certain months specifically around end of each quarter.
- Lowest order count is March 2020 followed by June 2019.

ORDERS ACROSS WEEKS

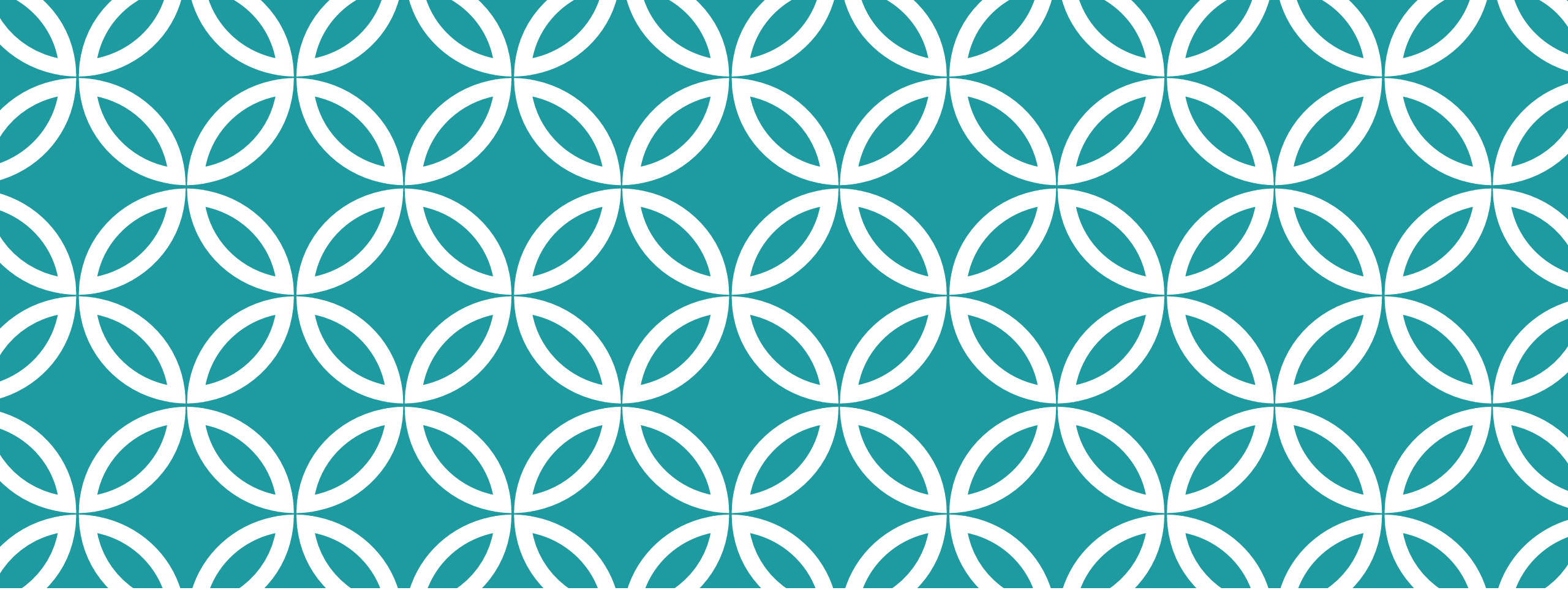


- There are dips and increases in orders seen after each week.
- Lowest order volume was seen during the last week of September 2018.
- End of Feb 2020 saw the total orders at 65.
- Highest number of orders was during the 3rd week of February 2019.

YEARLY COMPARISON OF PRODUCTS



- From this snippet of the top 6 products, we can see that there is an increase of orders of poultry, soda, cheeses and waffles for the year 2019 compared to 2018.
- Orders for cereals, ice cream have reduced compared the previous year.
- 2020 has the lowest orders but this is due to the data ending 26 feb 2020 which only accounts for around 2 months.



MARKET BASKET ANALYSIS

MARKET BASKET ANALYSIS

Market basket analysis is a technique used to uncover relationships between items purchased together by analysing transactional data. It helps businesses understand customer buying behavior and optimize product placements or recommendations. The approach relies on association rules, identifying patterns.

Retailers and e-commerce platforms use this method to enhance sales strategies and targeted marketing.

Ultimately, it improves decision-making based on real transaction data, boosting revenue and customer satisfaction.

ASSOCIATION RULES

Through the process of association rules, we can identify the relationships between all the products in every transaction.

Within the context of the dataset, these rules will help us analyse the different products purchased by each customer and accordingly we can identify which product to recommend for sale through product placement or strategies to boost sales.

Measures such as Support, Confidence and Lift are used to evaluate the relationship between the products.

Support – The proportion of transactions containing both the antecedent and consequent, indicating how frequently the rule appears in the dataset.

Confidence - The likelihood that a transaction containing the antecedent also contains the consequent, measuring the rule's reliability.

Lift - The ratio of observed confidence to expected confidence if antecedent and consequent were independent, showing how much one item boosts the likelihood of another.

KNIME WORKFLOW

To create a workflow for Market Basket Analysis, KNIME Analytics Platform is used. The data is prepared and the association rule learner node is utilised to generate the values for the measures support, confidence and lift used to identify relationships between different products.

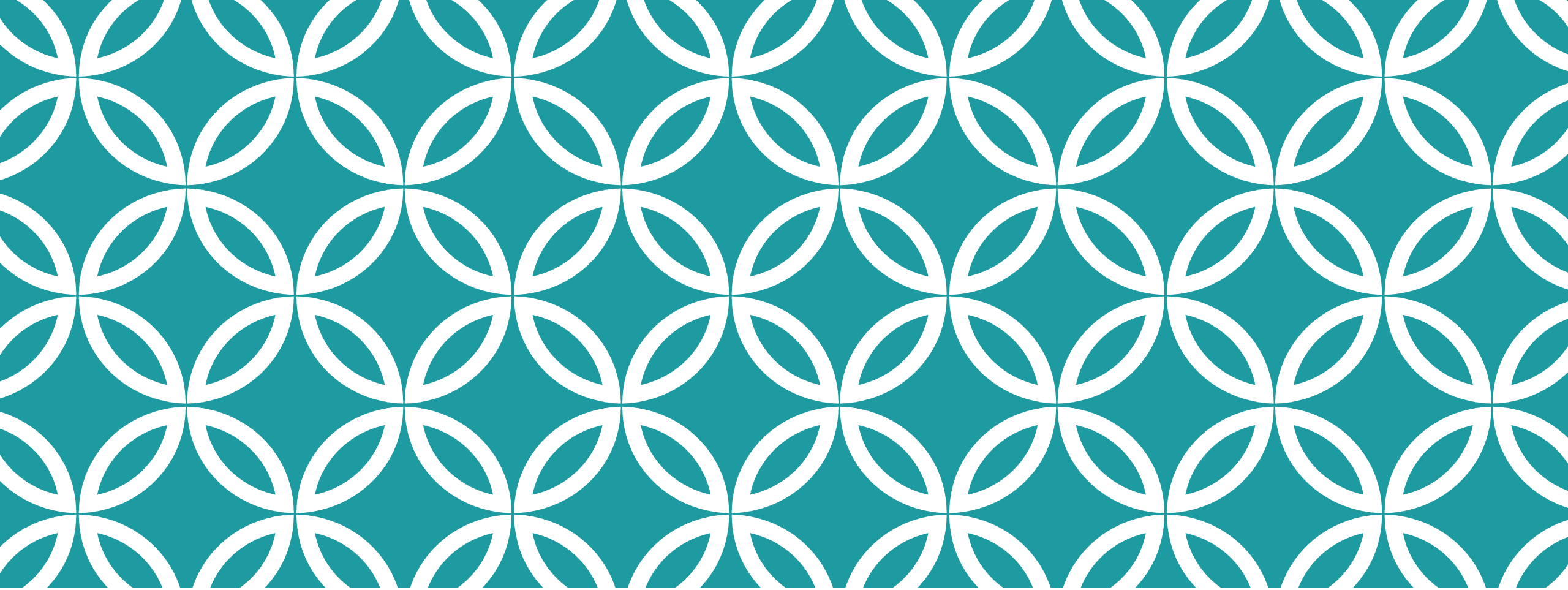
The threshold value considered for minimum support is 5%.
The threshold value considered for minimum confidence is 60%

With these parameters, 24 rules have been created.

FINAL OUTPUT & EXPLANATION OF VALUES

#	RowID	Support *Number (double)*	Confidence *Number (double)*	Lift *Number (double)*	Recommended Item *String*	implies *String*	Basket Items *Set*
1	rule0	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
2	rule1	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
3	rule2	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
4	rule3	0.05	0.6	1.424	poultry	<---	[dishwashing liquid/detergent,laundry
5	rule4	0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
6	rule5	0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
7	rule6	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
8	rule7	0.051	0.617	1.558	cereals	<---	[cheeses,bagels,sandwich bags]
9	rule8	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce,poultry,cereals]
10	rule9	0.051	0.637	1.512	poultry	<---	[dinner rolls,spaghetti sauce,cereals]
11	rule10	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cereals]

- A support value of 0.05 indicates the item appears in 5% of all transactions. The higher the support score the higher it appears in the dataset
- Confidence value of 0.64 indicates that 64% of the time when the basket items are bought the recommended item is bought as well.
- A positive lift value indicates that there is correlation between the products and there is a higher chance of the customer purchasing the recommended item. The higher the lift value, the higher the likelihood of additional purchase



INFERENCES & RECOMMENDATIONS

INFERENCES & RECOMMENDATIONS

- **Exclusive Combo Deal**: Bundle yogurt, poultry, and aluminum foil with juice at a discounted price or offer a free small item to increase perceived value.
- **Buy-Two-Get-One-Free Offer**: Apply this promotion to dinner rolls, spaghetti sauce, and ice cream to encourage bulk purchases.
- **Household Essentials Bundle**: Offer paper towels, eggs, and pasta as a discounted set to attract families stocking up on necessities.
- **Cereal-Based Cross-Selling**: Give discounts on cheese, bagels, and sandwich bags when purchasing cereals, aligning with common breakfast habits.
- **Limited-Time Savings Promotion**: Introduce short-term discounts on poultry, laundry detergent, and mixes to drive impulse buys.
- **Loyalty-Based Rewards**: Develop a loyalty program where frequent buyers of combo deals and associated items earn exclusive perks.
- **Breakfast Combo Discount**: Provide a special offer on coffee/tea when customers purchase cereals and cheeses together.
- **Meal Prep Kit**: Bundle poultry with essential spices, marinades, and a bonus bottle of dishwashing liquid to simplify meal preparation.
- **Rotating Weekly Deals**: Rotate buy-two-get-one-free promotions on high-support items like yogurt and poultry to maintain engagement.
- **Customizable Mix-and-Match Deal**: Let customers pick any three associated items (yogurt, poultry, aluminum foil, juice, mixes) at a flat discounted rate.

END OF PART 2

PGP-DSBA MRA Main Project

BY
ISHAAN SHAKTI JAYARAMAN
PGPDSBA.O.JULY24.A