

# MRA Project

Proprietary

# Agenda

- Problem Statement
- Exploratory Analysis
- Market Basket Analysis using KNIME
- Inferences & Recommendations

Proprietary



# 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 items in the customer orders. The Store doesn't have any combo offers. Can you suggest the best combos & offers?

## Exploratory Analysis

Exploratory Analysis of data & an executive summary (in PPT) of your top findings, supported by graphs.  
Are there trends across months/years/quarters/days etc. that you are able to notice?

## Use of Market Basket Analysis (Association Rules)

Write Something about the association rules and its relevance in this case  
Add KNIME workflow Image or Python package used  
Write about threshold values of Support and Confidence

## Associations Identified

Put the associations in a tabular manner  
Explain about support, confidence, & lift values that are calculated

## Suggestion of Possible Combos with Lucrative Offers

Write recommendations  
Make discount offers or combos (or buy two get one free) based on the associations and your experience



# Exploratory Analysis

Proprietary



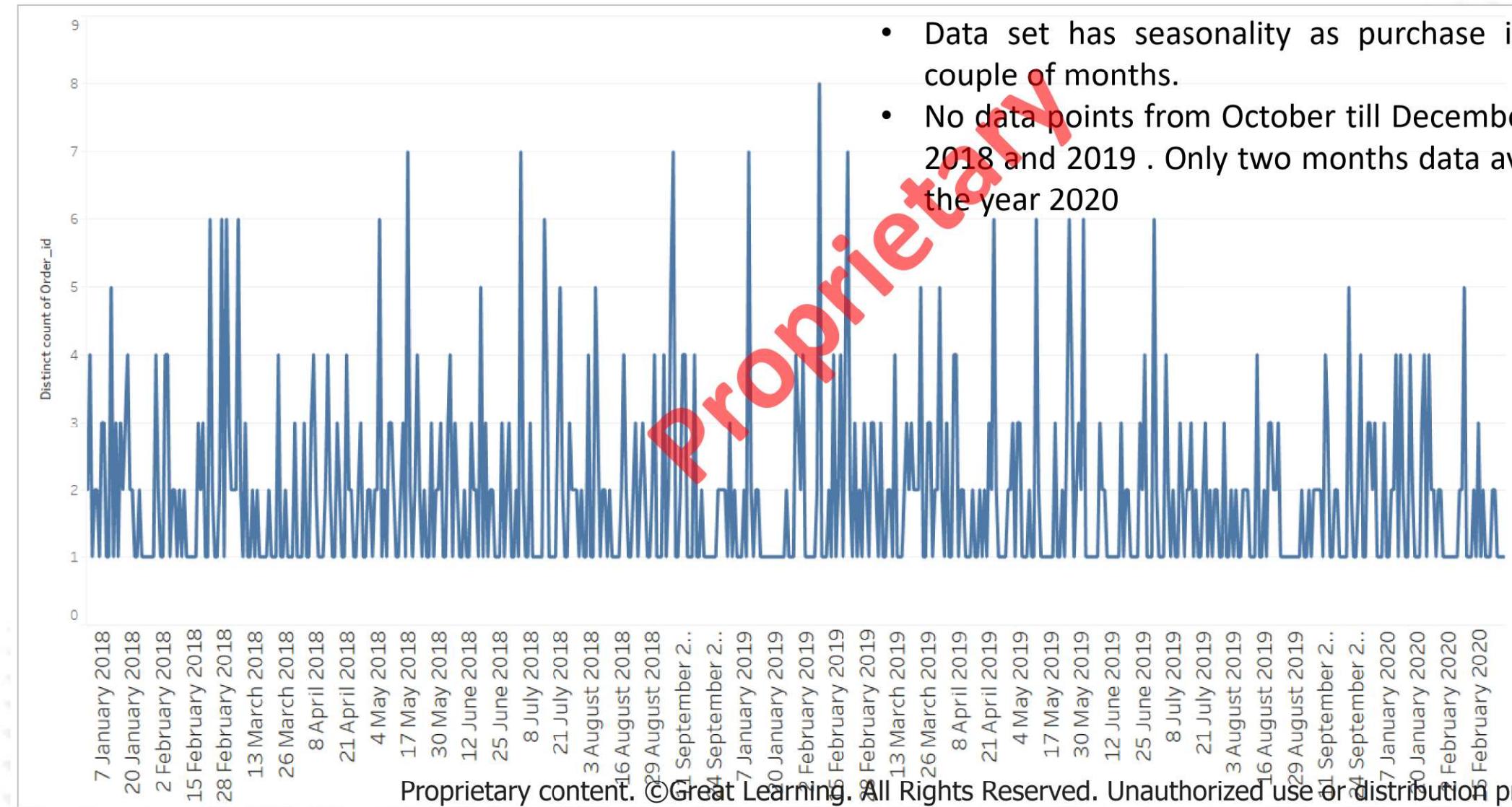
# About Data

- There are 20641 rows and 3 Columns in the dataset provided
- There are no missing values in the data
- There are about 4730 duplicate rows which are dropped
- There are 1139 Unique Transactions and 37 Unique products
- The data is for 3 years 2018,2019 and 2020
- The data is for 9 months of 2018, 9 months of 2019 and 2 months of 2020
- The Start date is 1st Jan, 2018 and End date is 26th Feb 2020

Proprietary



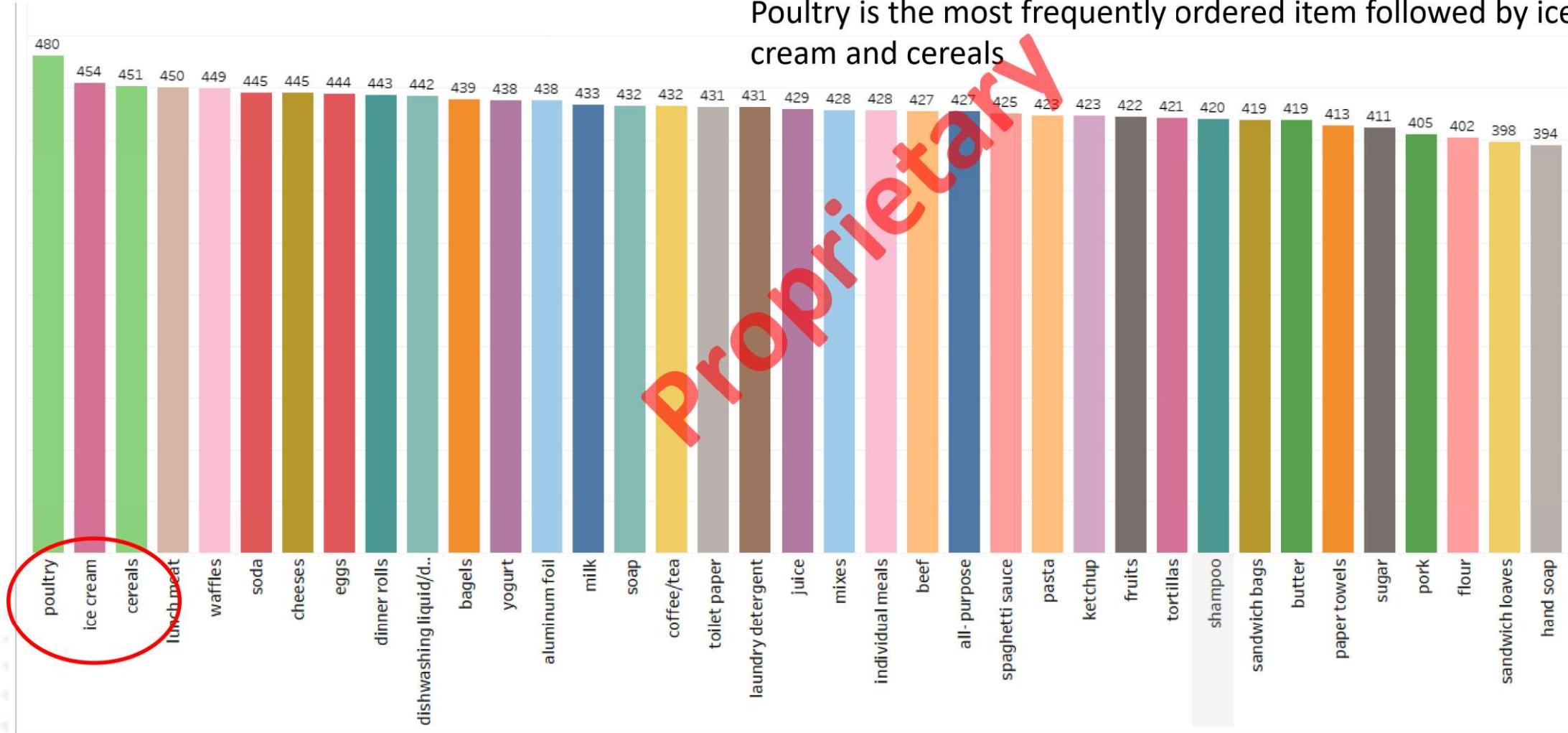
# Order Counts Trends w.r.t date



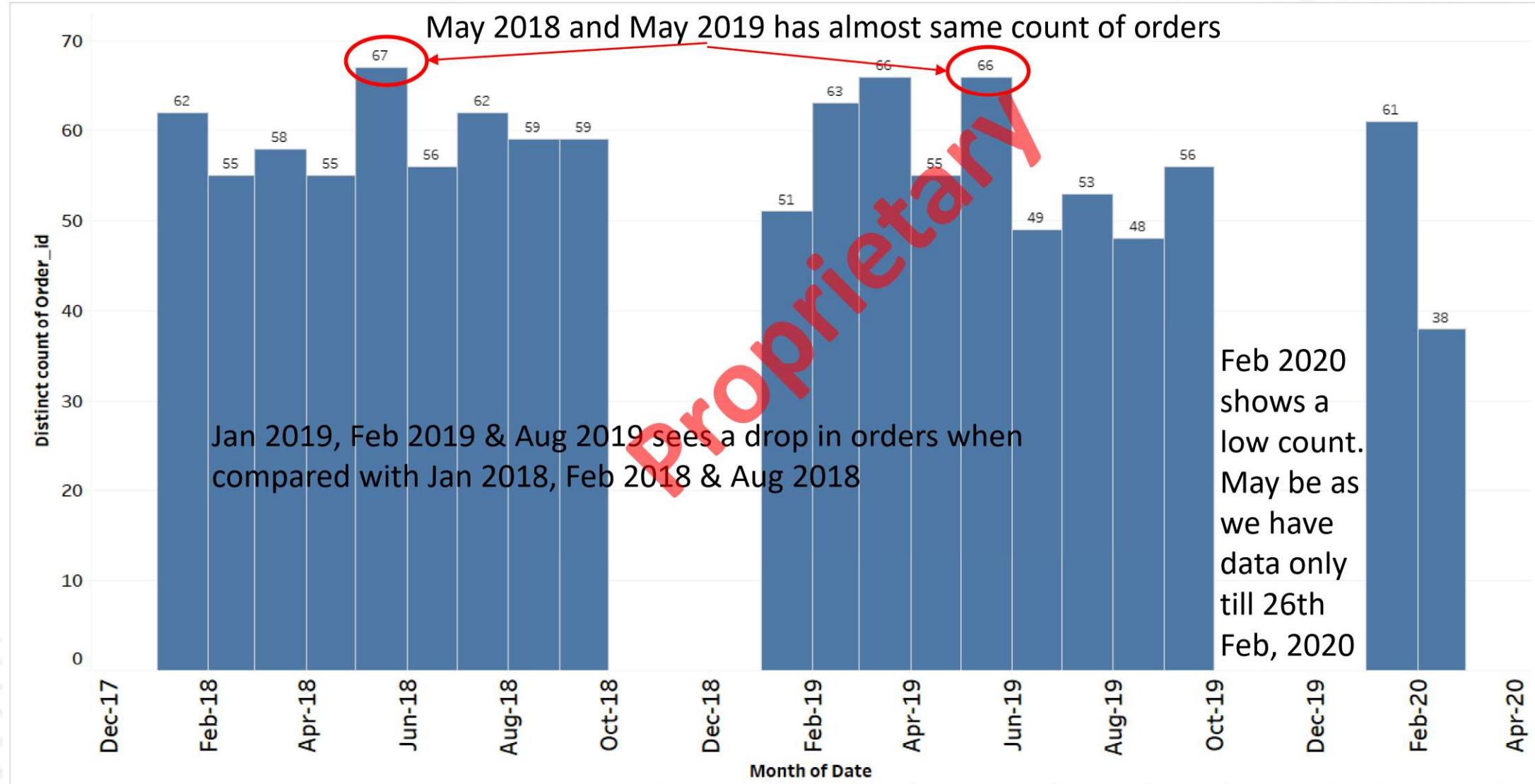
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



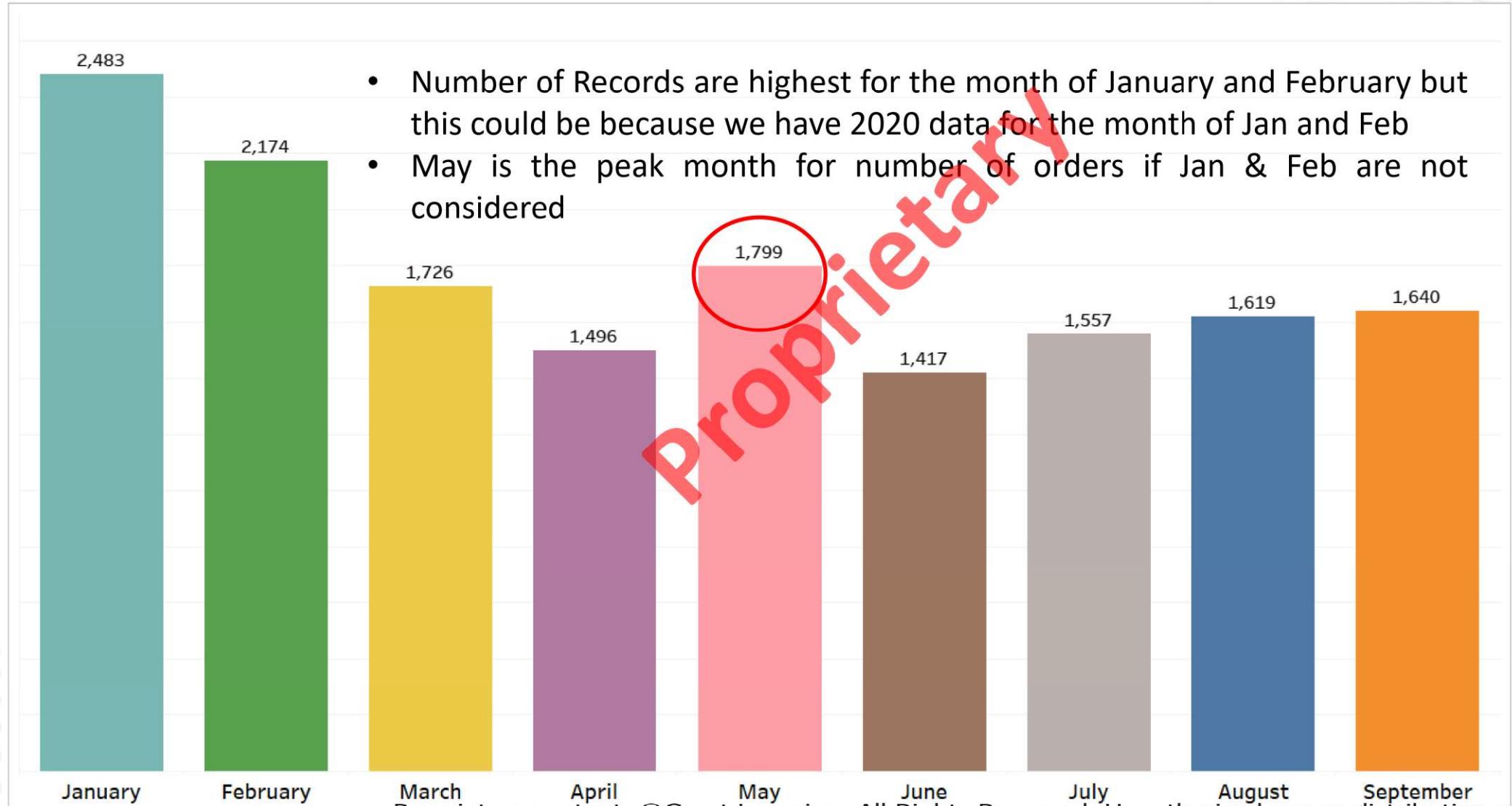
# Products wise Distribution



# Month-Yearly Order Counts



# Monthly Records



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



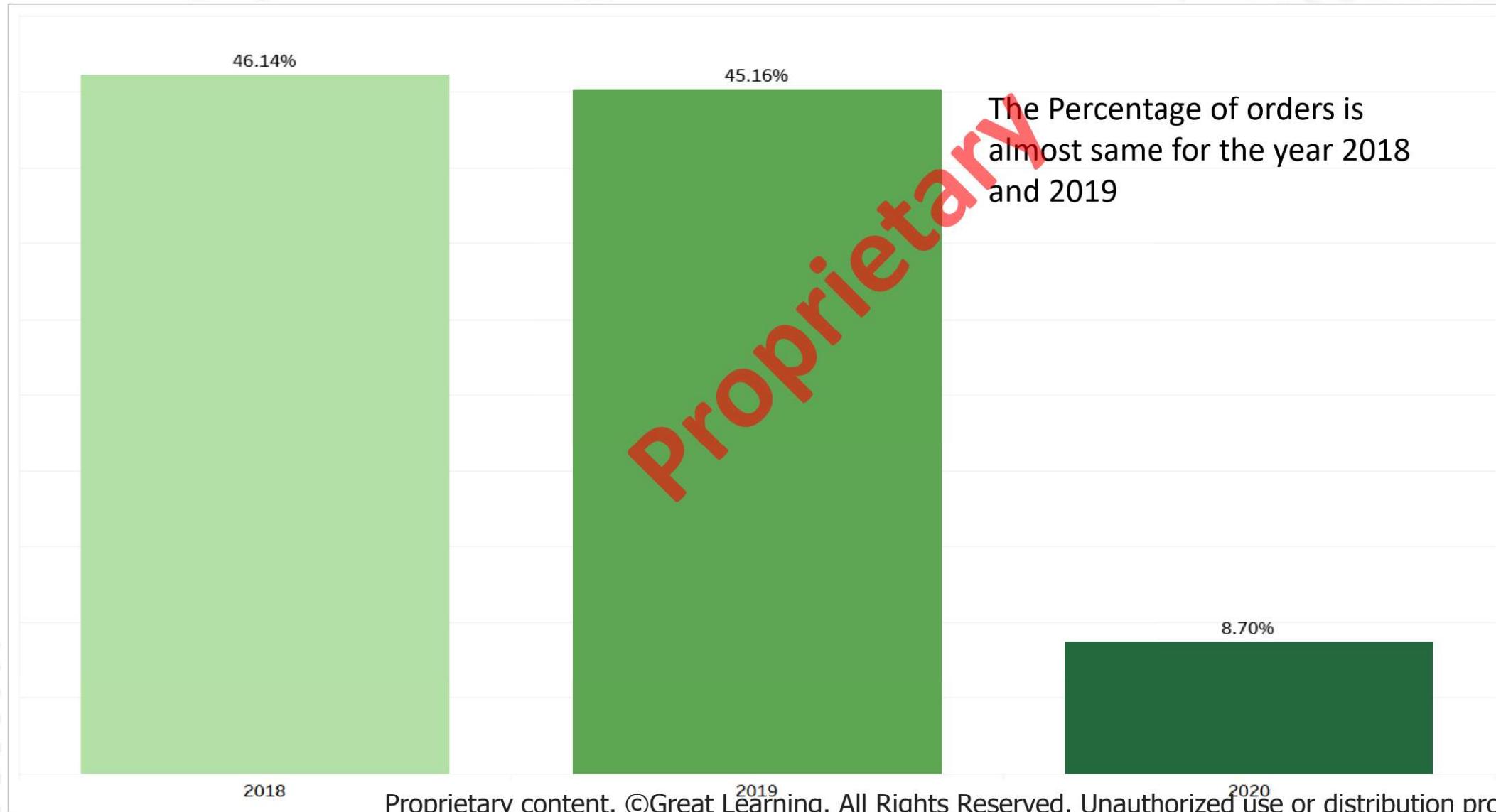
# Season wise percentage of orders



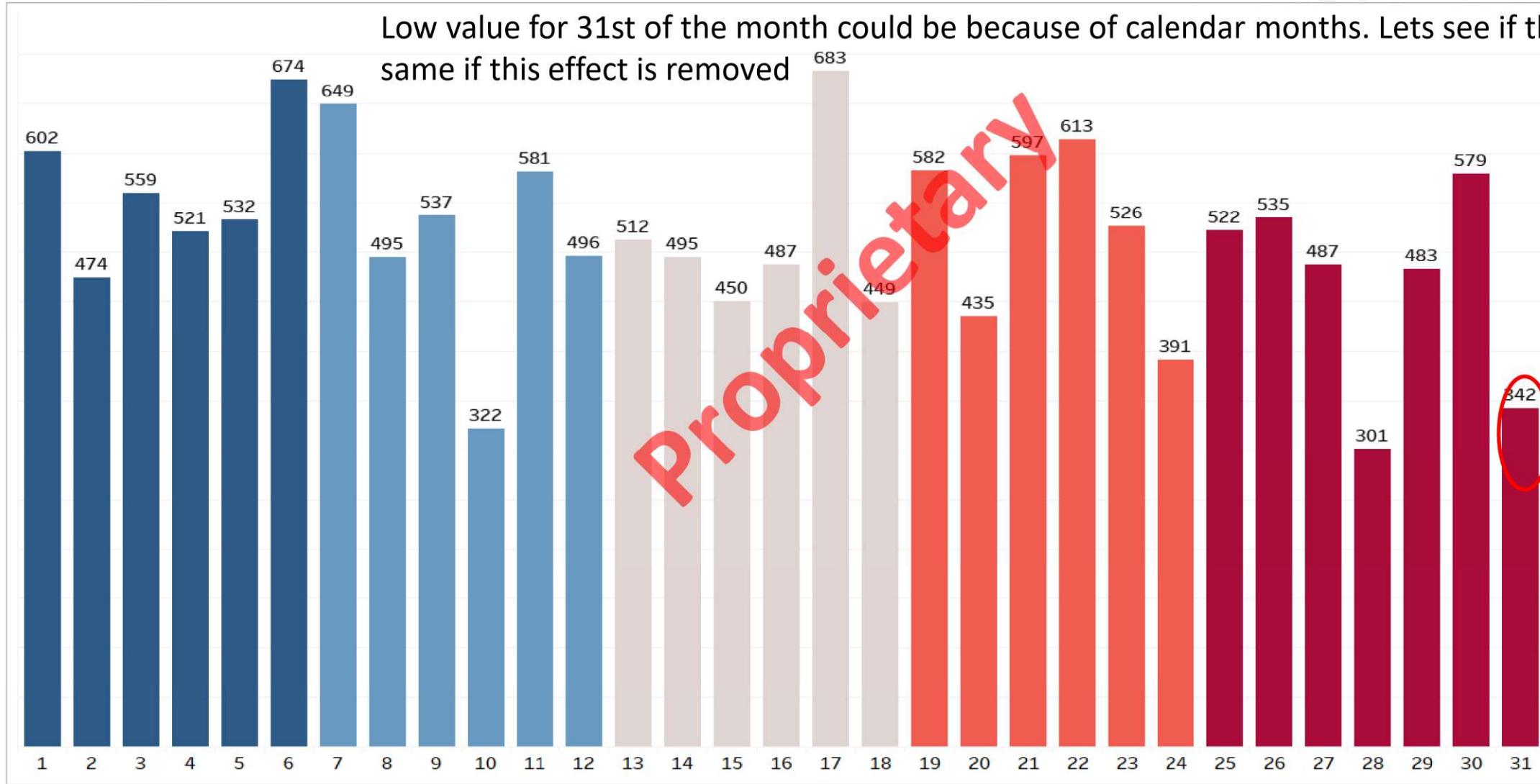
# Quarter wise percentage of orders



# Yearly percentage of orders

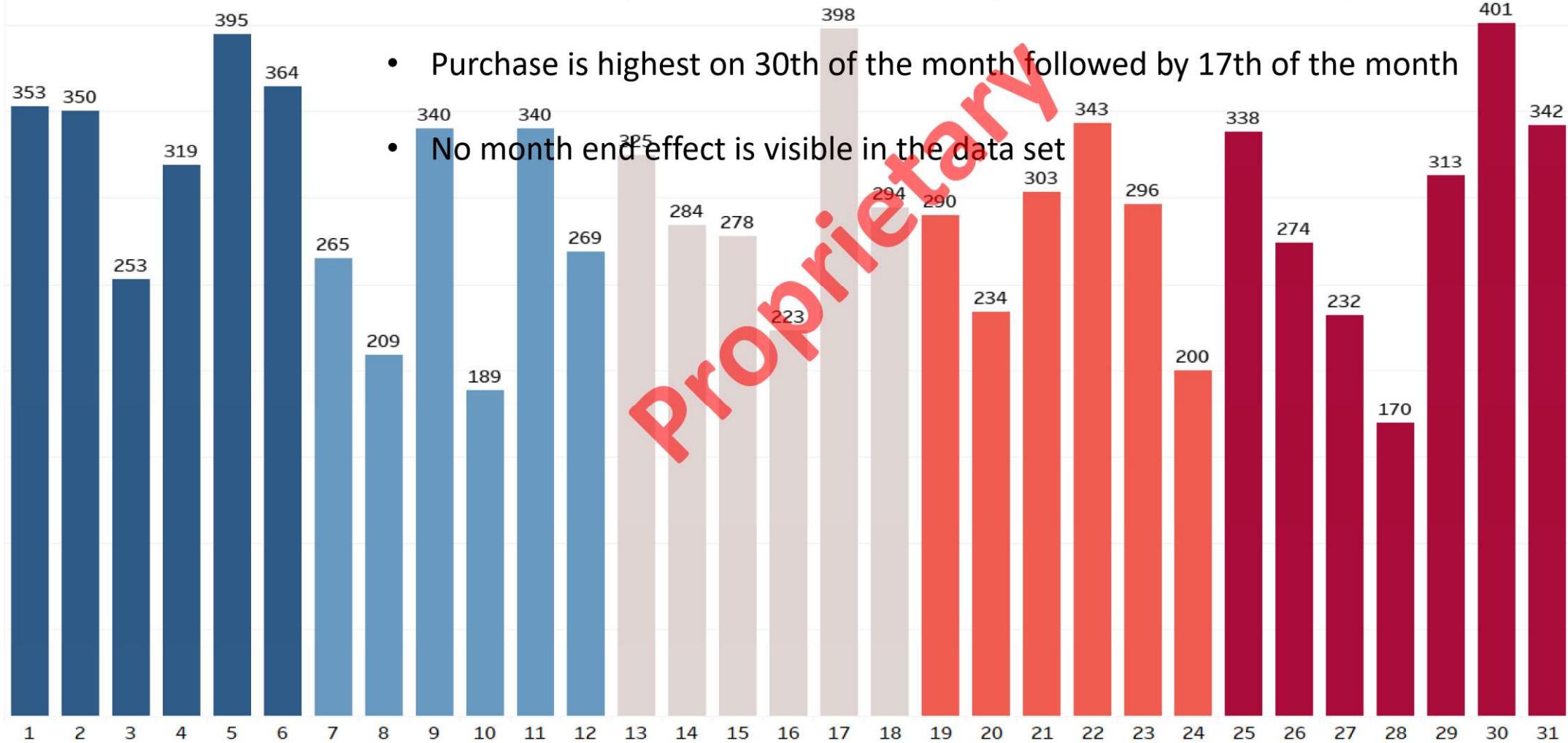


# Daily Order Counts



# Daily Order Counts (Jan, Mar, May, July & Aug)

- Purchase is high even on month end days for months with 31 days.



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



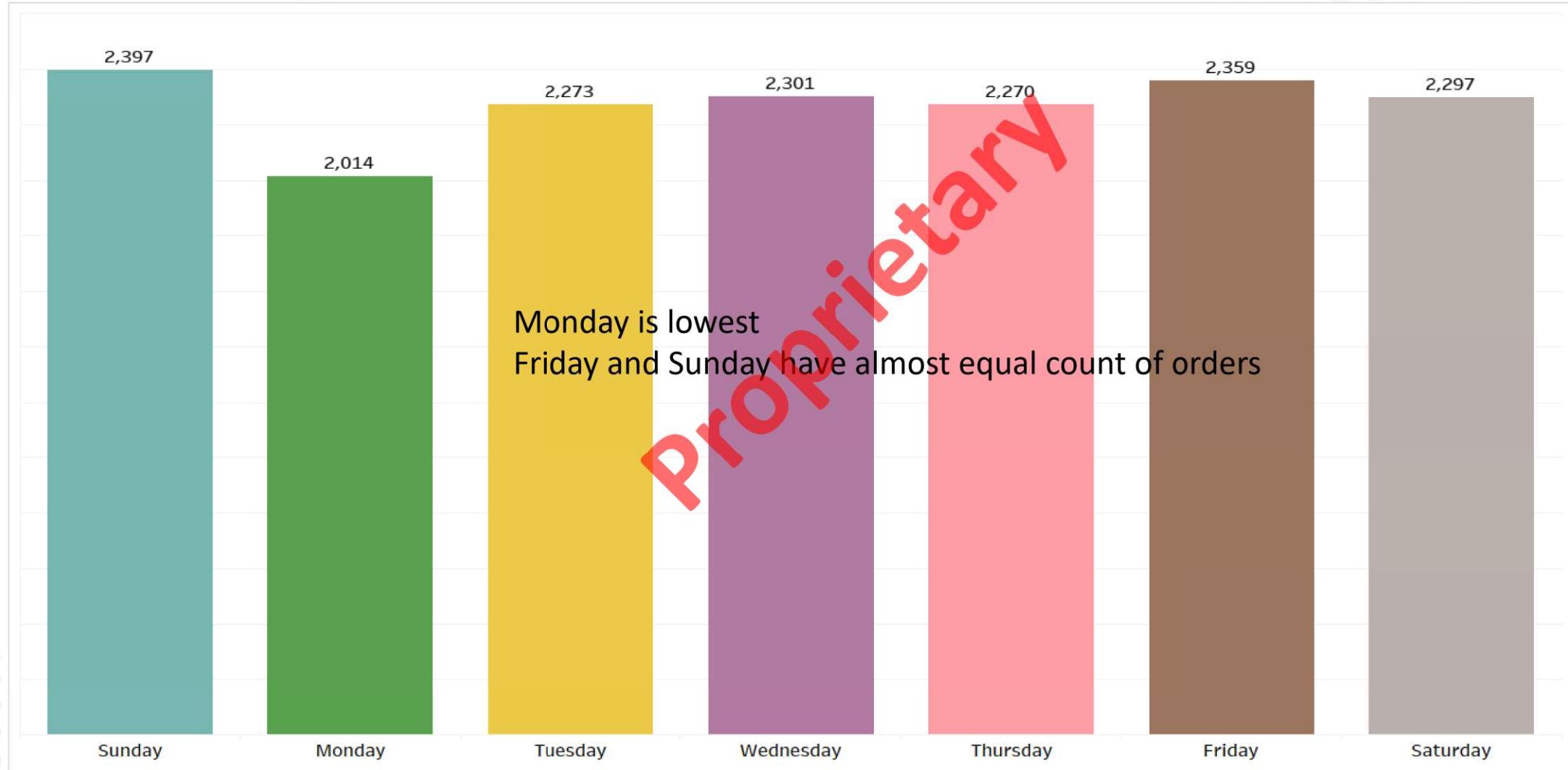
# Daily Order Counts (Apr, June, Sep)



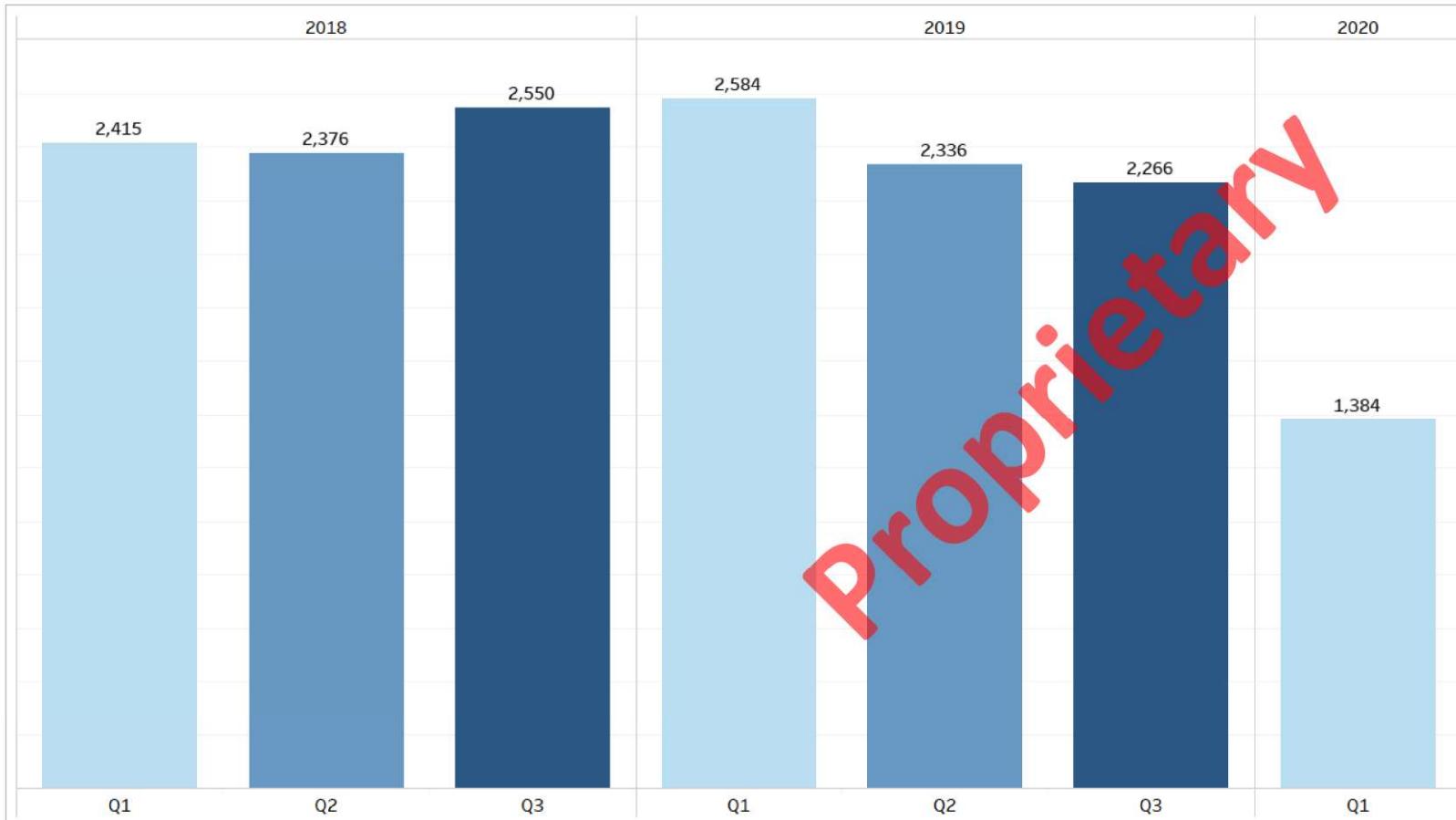
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



# Weekday wise Orders count



# Year-Quarter Order Counts



- Quarter 1 of 2019 had high count of order when compared with Q1 of 2018. Q3 of 2018 have a higher count than Q3 of 2019.
- Quarter 2 of both 2018 and 2019 was almost equal.
- Low value for Q1 of 2020 could be as we have data points only till 26th Feb, 2020 .



# Year-Quarter Product wise Order Counts – Q1



- Poultry and Icecream had more order in Q1 2019 when compared with Q1 2018
- Flour had the maximum from in 2019 when compared with 2018 of Q1



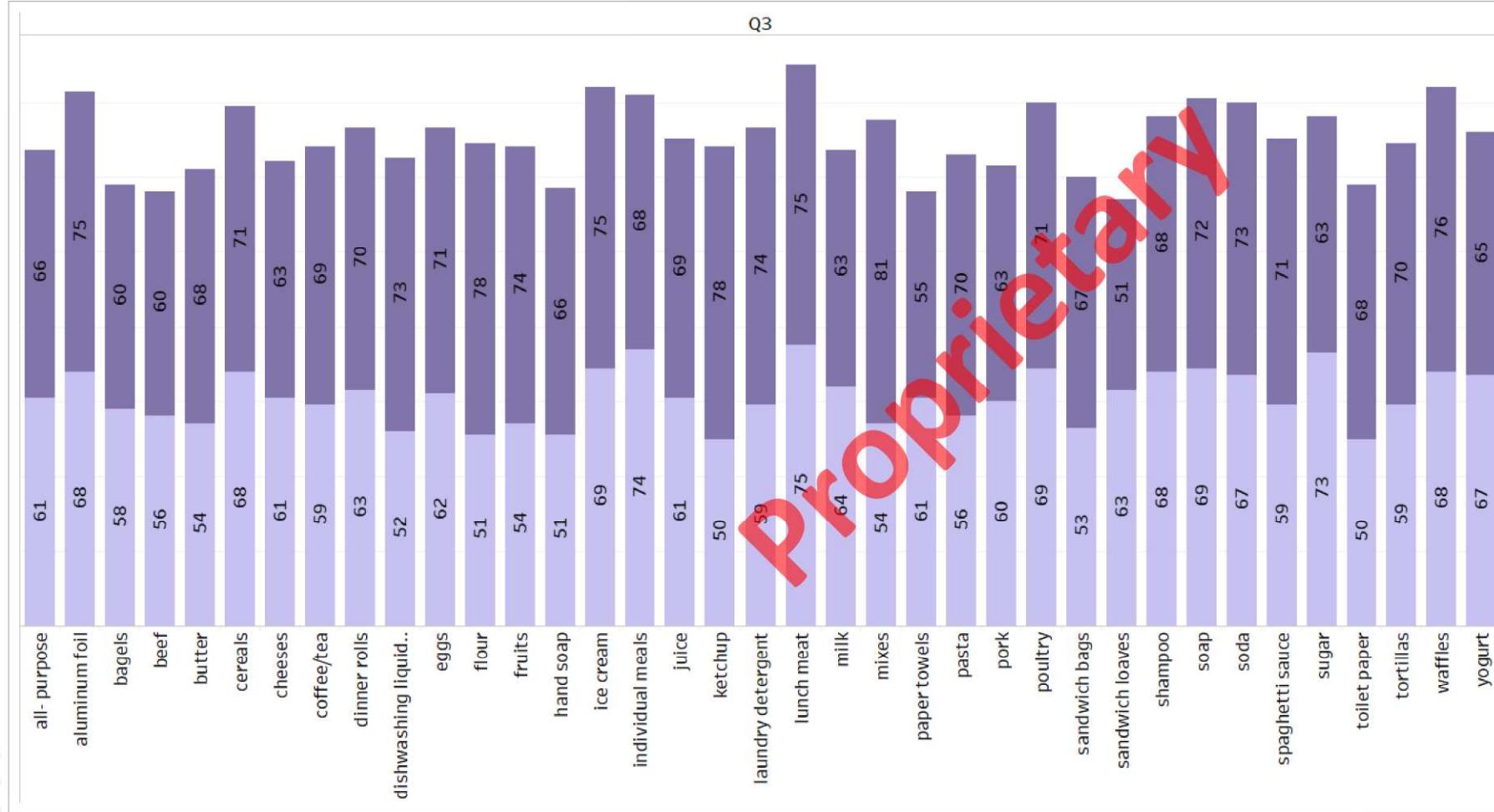
# Year-Quarter Product wise Order Counts – Q2



- For Poultry, trend observed between Q2 2018 and Q2 2019 is similar to Q1 ((more orders in Q2 2019 ).
- However for ice cream order count is higher for Q2 2018 when compared with Q2 2019
- Ketchup and mixes had the maximum drop in 2019 when compared to 2018 for Q2



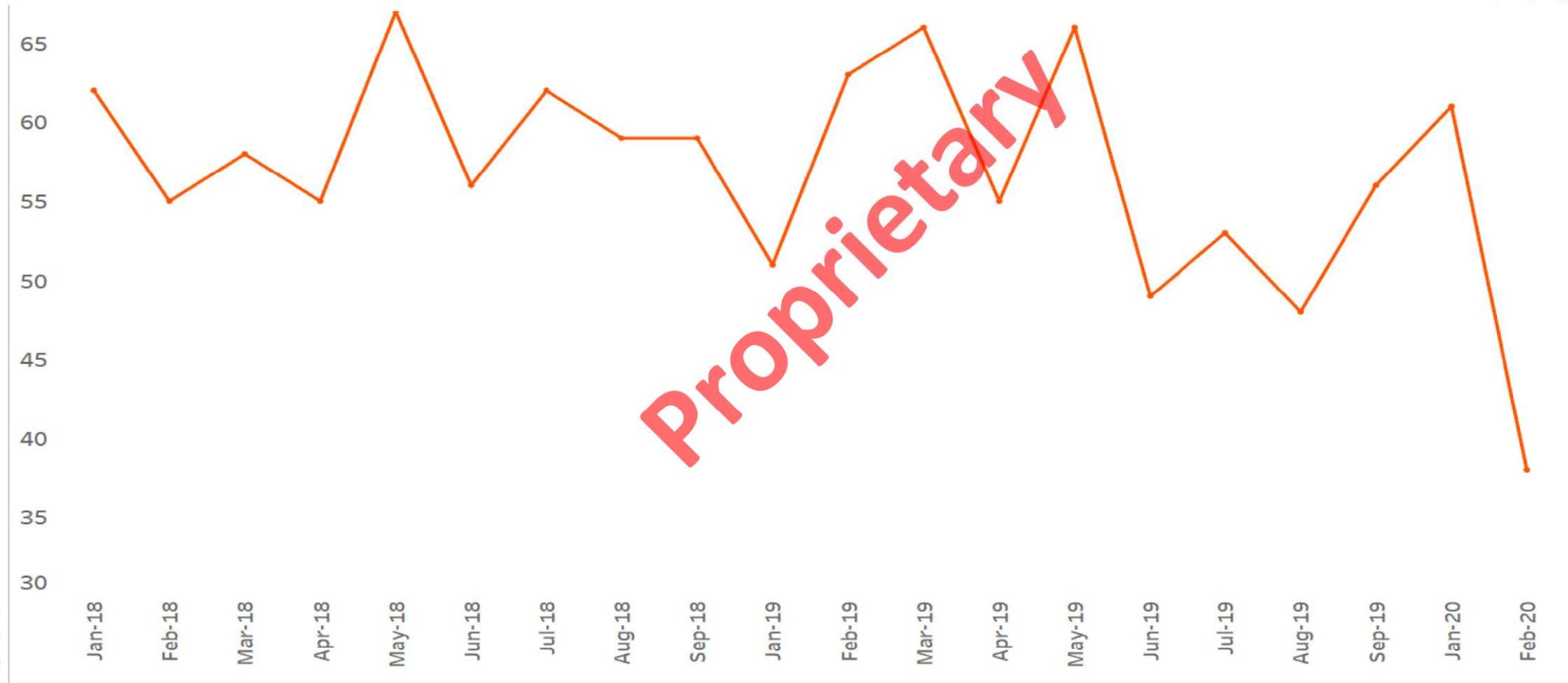
# Year-Quarter Product wise Order Counts – Q3



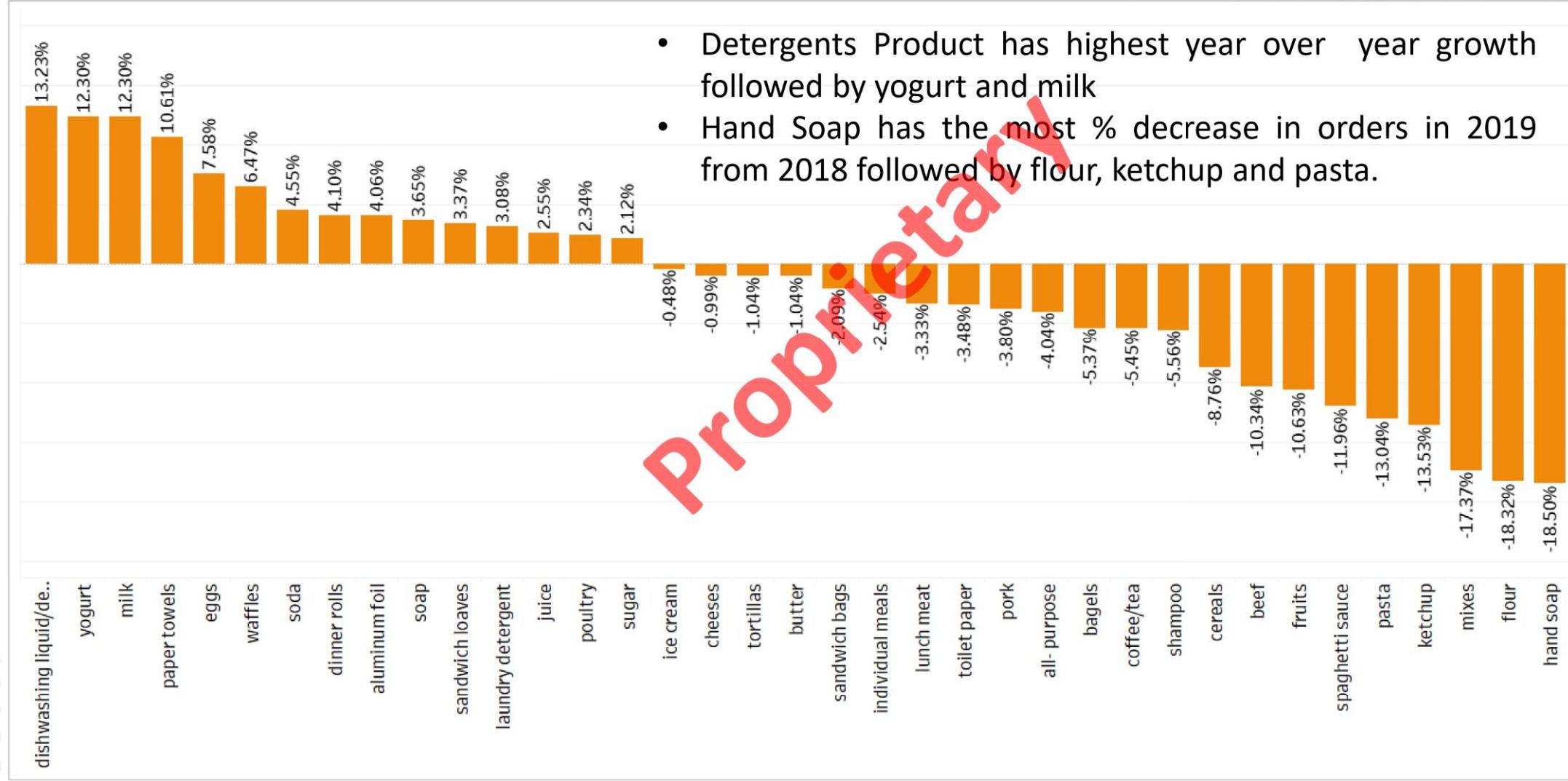
- For Poultry, trend observed between Q3 2018 and Q3 2019 is different from Q1 and Q2 (more orders in Q3 2018).
- For ice cream trend is similar to Q2.( More orders in Q3 2018 than Q3 2019).
- Ketchup and toilet paper had the maximum drop in 2019 when compared to 2018 for Q2



# Order Count Trend



# Year on Year Growth Products wise (2018 to 2019)



# EDA Inferences

Proprietary



# Key Observations (1)

- (1) Dataset has 20641 rows and 3 features. Data is arranged based on order\_id on a specific date.
- (2) Dataset has 1139 unique Orders.
- (3) Data set has seasonality as purchase is high for couple of months.  
No data points from October till December for both 2018 and 2019 . Only two months data available for the year 2020
- (4) Data corresponds to 37 unique Products.
- (5) Poultry is the most frequently ordered item followed by ice cream and cereals
- (6) Data set is for 9 months of 2018, 9 months of 2019 and 2 months of 2020. Start date is 1st Jan, 2018 and End date is 26th Feb 2020
- (7) May 2018 and May 2019 has almost same count of orders.. Jan 2019, Feb 2019 & Aug 2019 sees a drop in orders when compared with Jan 2018, Feb 2018 & Aug 2018.. May is the peak month for both 2018 and 2019.
- (8) Winter shows the highest count of orders.
- (9) Sale is highest in Quarter 1. Quarter 2 and Quarter 3 have equal transaction count
- (10) Purchase is high even on month end days for months with 31 days. Purchase is highest on 30th of the month followed by 17th of the month  
No month end effect is visible in the data set



## Key Observations (2)

- (11) Purchase is high in the first week for the months with 30 or less than 30 days.
- (12) Monday is lowest. Friday and Sunday have almost equal count of orders.
- (13) Quarter 1 of 2019 had high count of order when compared with Q1 of 2018. Q3 of 2018 have a higher count than Q3 of 2019. Quarter 2 of both 2018 and 2019 was almost equal. Low value for Q1 of 2020 could be as we have data points only till 26th Feb, 2020
- (14) Poultry and Ice-cream had more order in Q1 2019 when compared with Q1 2018. flour had the maximum from in 2019 when compared with 2018 of Q1.
- (15) For Poultry, trend observed between Q2 2018 and Q2 2019 is similar to Q1 (more orders in Q2 2019). However for ice cream order count is higher for Q2 2018 when compared with Q2 2019. Ketchup and mixes had the maximum drop in 2019 when compared to 2018 for Q2.
- (16) For Poultry, trend observed between Q3 2018 and Q3 2019 is different from Q1 and Q2 (more orders in Q3 2018). For ice cream trend is similar to Q2. (More orders in Q3 2018 than Q3 2019). Ketchup and toilet paper had the maximum drop in 2019 when compared to 2018 for Q2.
- (17) Detergents Product has highest year over year growth followed by yogurt and milk. Hand Soap has the most % decrease in orders in 2019 from 2018 followed by flour, ketchup and pasta.



# Market Basket Analysis

Proprietary



# Market Basket Analysis

Market Basket Analysis is a techniques used by large retailers to uncover the association between items or to identify the relationship between items which are bought together more frequently .

**Support** - It is a measure of how frequently the collection of items occur together as a percentage of all transactions

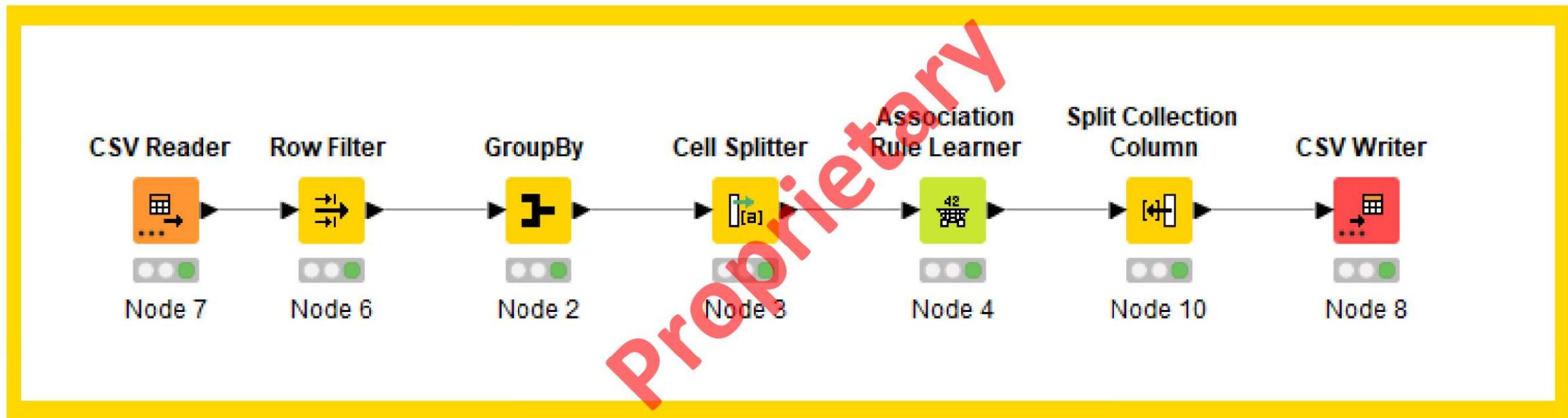
**Lift** – It is the ratio of confidence to expected confidence. Higher lift values indicate stronger associations.

**Confidence** – It is the ratio of the number of transactions that include all items in {B} as well as the number of transactions that include all items in {A} to the number of transactions that include all items in {A}. (Assume A & B are items in some transaction)

Proprietary



# MBA using KNIME



Support Value set to 0.049 while Confidence is set to 0.60 and gives us 37 rules  
Grouped by Order Id and Products are concatenated



# Association Rule Learner Output

row ID	Support	Confidence	Lift	Consequent	implies	Ante 1	Ante 2	Ante 3
rule37	0.055311677	0.649484536	1.79119343	paper towels	<---	eggs	ice cream	pasta
rule36	0.055311677	0.642857143	1.73100304	pasta	<---	paper towels	eggs	ice cream
rule20	0.050921861	0.674418605	1.726208518	cheeses	<---	bagels	cereals	sandwich bags
rule14	0.050043898	0.640449438	1.700400723	juice	<---	yogurt	toilet paper	aluminum foil
rule4	0.049165935	0.615384615	1.697150307	paper towels	<---	eggs	ice cream	lunch meat
rule18	0.050921861	0.630434783	1.677722471	mixes	<---	yogurt	poultry	aluminum foil
rule19	0.050921861	0.610526316	1.659640749	sandwich bags	<---	cheeses	bagels	cereals
rule32	0.053555751	0.642105263	1.650920756	dinner rolls	<---	spaghetti sauce	poultry	laundry detergent
rule26	0.051799824	0.641304348	1.648861517	dinner rolls	<---	spaghetti sauce	poultry	ice cream
rule15	0.050043898	0.619565217	1.644952873	juice	<---	yogurt	poultry	aluminum foil
rule13	0.049165935	0.622222222	1.636746215	milk	<---	poultry	ice cream	cereals
rule0	0.049165935	0.629213483	1.636242368	yogurt	<---	eggs	juice	aluminum foil
rule3	0.049165935	0.651162791	1.633644094	ice cream	<---	paper towels	eggs	lunch meat
rule27	0.051799824	0.686046512	1.627931202	poultry	<---	dinner rolls	spaghetti sauce	ice cream

Sorted by Lift (Highest to Lowest)



# Interpreting rules (rule 0 and rule 1)

row ID	Consequent	implies	Antecedent 1	Antecedent 2	Antecedent 3	Support	Confidence	Lift
rule0	yogurt	<--	eggs	juice	aluminum foil	0.049	0.629	1.636
rule1	lunch meat	<--	shampoo	waffles	juice	0.049	0.615	1.558
rule2	waffles	<--	shampoo	lunch meat	juice	0.049	0.609	1.544
rule3	ice cream	<--	paper towels	eggs	lunch meat	0.049	0.651	1.634
rule4	paper towels	<--	eggs	ice cream	lunch meat	0.049	0.615	1.697
rule5	eggs	<--	poultry	pasta	soda	0.049	0.629	1.614

## Rule 0:

- Support** says 5% customers bought eggs, yogurt, juice and aluminium foil
- Confidence** is that 63% customers who bought eggs, juice, aluminium foil also bought yogurt
- Lift** is that there is 64% increase in expectation that customer will buy yogurt when we know that they bought eggs, juice, aluminium foil

## Rule 1:

- Support** says 5% customers bought lunch meat, waffles, juice and shampoo
- Confidence** is that 61.5% customers who bought waffles, juice and shampoo also bought lunch meat
- Lift** is that there is 56% increase in expectation that customer will buy lunch meat when we know that they bought waffles, juice and shampoo



# Interpreting rules (rule 2 and rule 3)

row ID	Consequent	implies	Antecedent 1	Antecedent 2	Antecedent 3	Support	Confidence	Lift
rule0	yogurt	<--	eggs	juice	aluminum foil	0.049	0.629	1.636
rule1	lunch meat	<--	shampoo	waffles	juice	0.049	0.615	1.558
rule2	waffles	<--	shampoo	lunch meat	juice	0.049	0.609	1.544
rule3	ice cream	<--	paper towels	eggs	lunch meat	0.049	0.651	1.634
rule4	paper towels	<--	eggs	ice cream	lunch meat	0.049	0.615	1.697
rule5	eggs	<--	poultry	pasta	soda	0.049	0.629	1.614

## Rule 2:

- **Support** says 5% customers bought lunch meat, waffles, juice and shampoo
- **Confidence** is that 61% customers who bought lunch meat, juice and shampoo also bought waffles
- **Lift** is that there is 54% increase in expectation that customer will buy waffles when we know that they bought lunch meat, juice and shampoo

## Rule 3:

- **Support** says 5% customers bought eggs, ice cream, lunch meat and paper towels
- **Confidence** is that 65% customers who bought eggs, lunch meat, and paper towels also bought ice cream
- **Lift** is that there is 63% increase in expectation that customer will buy ice cream when we know that they bought eggs, lunch meat, and paper towels



# Interpreting rules (rule 4 and rule 5)

row ID	Consequent	implies	Antecedent 1	Antecedent 2	Antecedent 3	Support	Confidence	Lift
rule0	yogurt	<--	eggs	juice	aluminum foil	0.049	0.629	1.636
rule1	lunch meat	<--	shampoo	waffles	juice	0.049	0.615	1.558
rule2	waffles	<--	shampoo	lunch meat	juice	0.049	0.609	1.544
rule3	ice cream	<--	paper towels	eggs	lunch meat	0.049	0.651	1.634
rule4	paper towels	<--	eggs	ice cream	lunch meat	0.049	0.615	1.697
rule5	eggs	<--	poultry	pasta	soda	0.049	0.629	1.614

## Rule 4:

- **Support** says 5% customers bought lunch meat, paper towels, eggs and ice cream
- **Confidence** is that 61.5% customers who bought lunch meat, eggs and ice cream also bought paper towels
- **Lift** is that there is 70% increase in expectation that customer will buy paper towels when we know that they bought lunch meat, eggs and ice cream

## Rule 5:

- **Support** says 5% customers bought eggs, poultry, pasta and soda
- **Confidence** is that 63% customers who bought poultry, pasta and soda also bought eggs
- **Lift** is that there is 61.4% increase in expectation that customer will buy eggs when we know that they bought poultry, pasta and soda



# Taking Top 2 rules w.r.t Lift >1.6 and Confidence >0.66

row ID	Support	Confidence	Lift	Consequent	implies	Ante 1	Ante 2	Ante 3
rule20	0.050921861	0.674418605	1.726208518	cheeses	<---	bagels	cereals	sandwich bags
rule27	0.051799824	0.686046512	1.627931202	poultry	<---	dinner rolls	spaghetti sauce	ice cream

**Interpretations:**

**Rule 20:**

Support says 6% customers bought cereals, sandwich bags, bagels and cheeses together

Confidence is that 67% customers who bought cereals, sandwich bags and bagels also bought cheese

Lift is that there is 72% increase in expectation that customer will buy cheese when we know that they bought cereals, sandwich bags and bagels

**Rule 27:**

Support says 5% customers bought dinner rolls, ice cream, spaghetti sauce and poultry together

Confidence is that 69% customers who bought dinner rolls, ice cream and spaghetti sauce also bought poultry.

Lift is that there is 63% increase in expectation that customer will buy poultry when we know that they bought dinner rolls, ice cream and spaghetti sauce



# Recommendations

## Combo Offers

1. Poultry with Eggs (Buy 2 poultry get a carton of eggs free)
2. As per rule 20, cheese with sandwich bags and Bagels can be combined at a competitive price
3. Combo offer of Poultry and Beef , Spaghetti sauce with dinner rolls and poultry
4. Lunch combo of Lunch Meat and Waffles with Juice

## Discount Offers

1. 10% off if you buy Poultry with One Eggs Carton (assuming one carton contains dozen eggs)
2. 5% off on Cheese with Bagels
3. 15% off on Poultry & Beef if bought together
4. 5% off on Milk and Ice cream if bought together

Proprietary



**THANK YOU**

Always remember, Sample Solution is not the only solution.

