Sales Analysis & Predicting

Muhammad Randa Yandika

Introduction

This project encompasses a diverse dataset comprising purchase, sales, and stock information.
It aims to conduct deep analysis, price forecasting, and customer segmentation clustering.
Through this integrated approach, we seek to uncover trends, optimize pricing, and categorize customers, enabling informed and tailored strategies.

Table of contents

01

Deep Analysis

Analysis anything from dataset

02

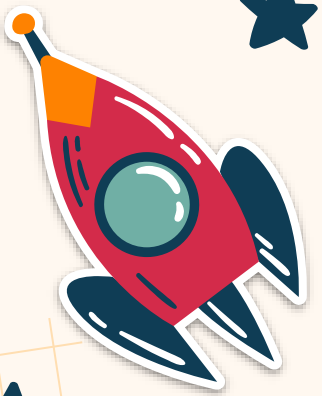
Forecasting

Sales/Price Forecasting in next 1 year

03

Clustering

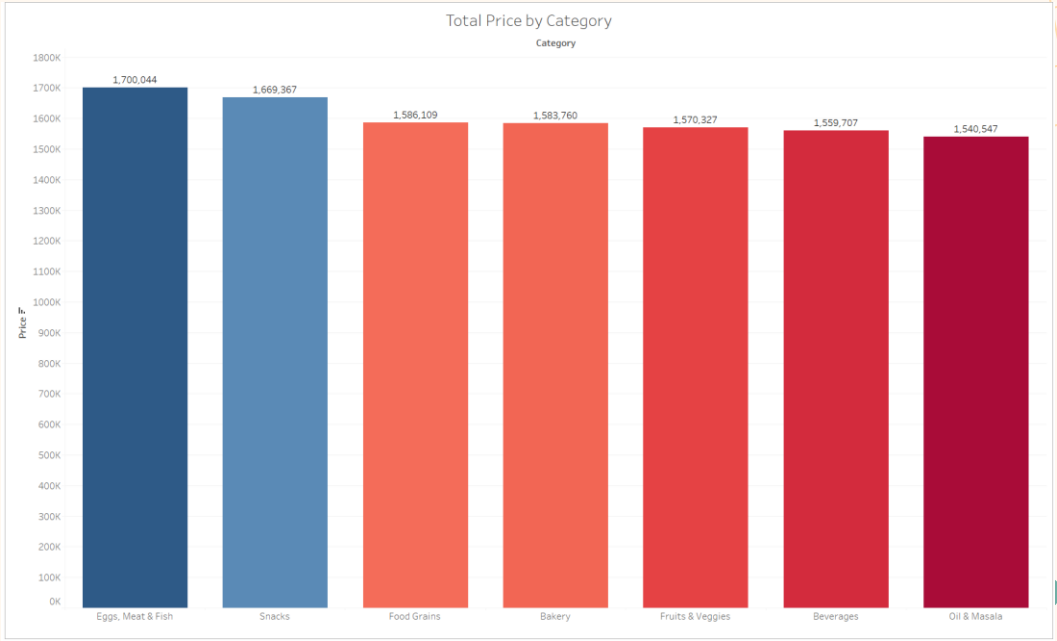
Customer Segmentation based on quantity and price



01
Deep Analysis

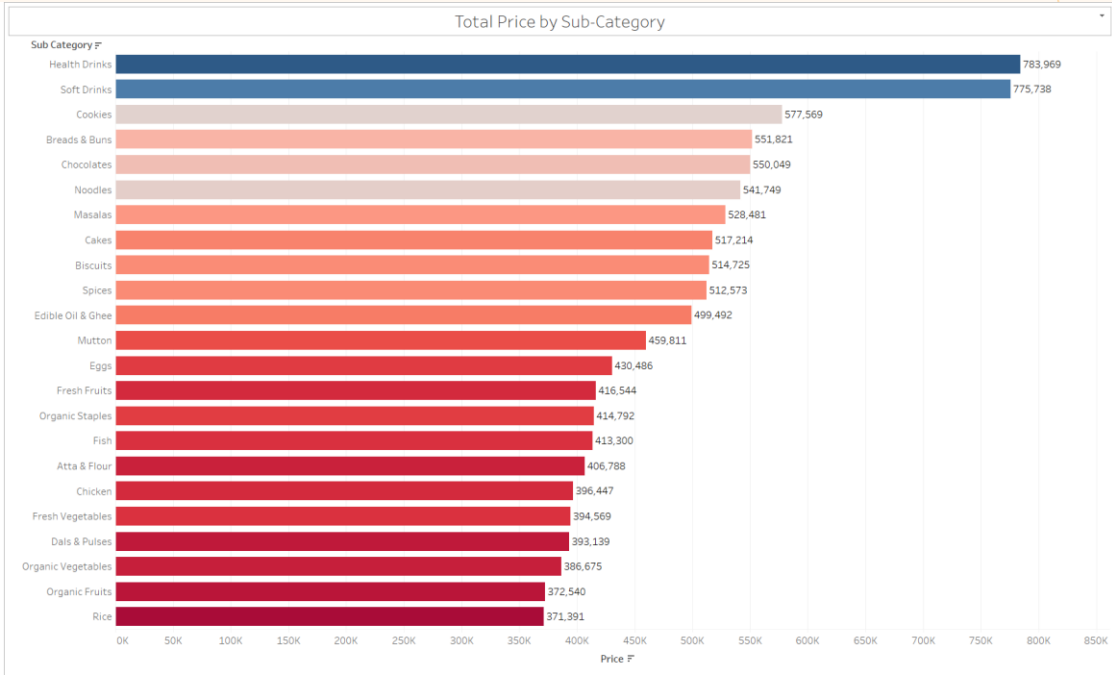
Which category generates the highest sales?

The **egg, meat and fish** categories generated the most sales which is **1.700.044**, while **oil and masala** generated the least which is **1.540.547**



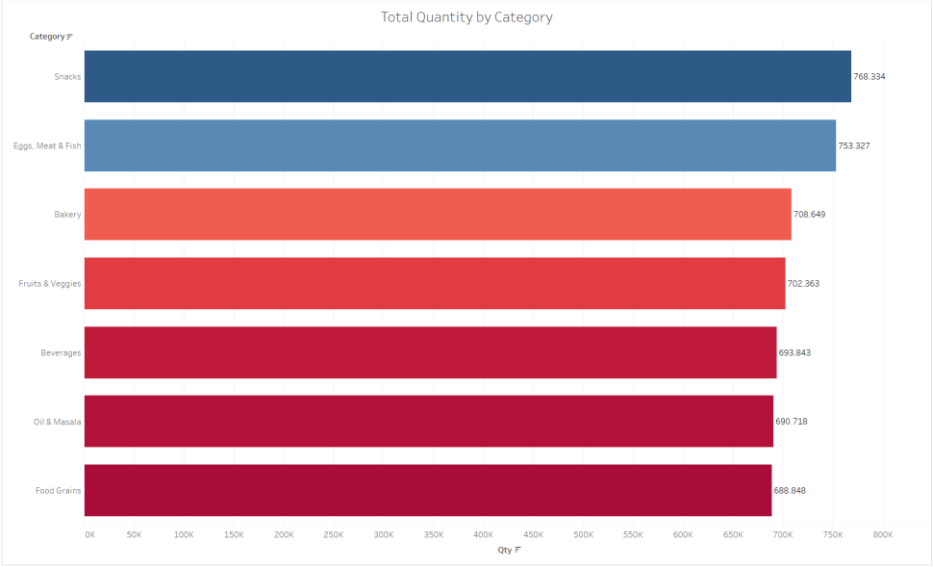
Which sub-category generates the highest sales?

Health & Soft drink is items generated the most sales which is **783.969 and 775.378**, while **Rice** generated the least which is **371.391**



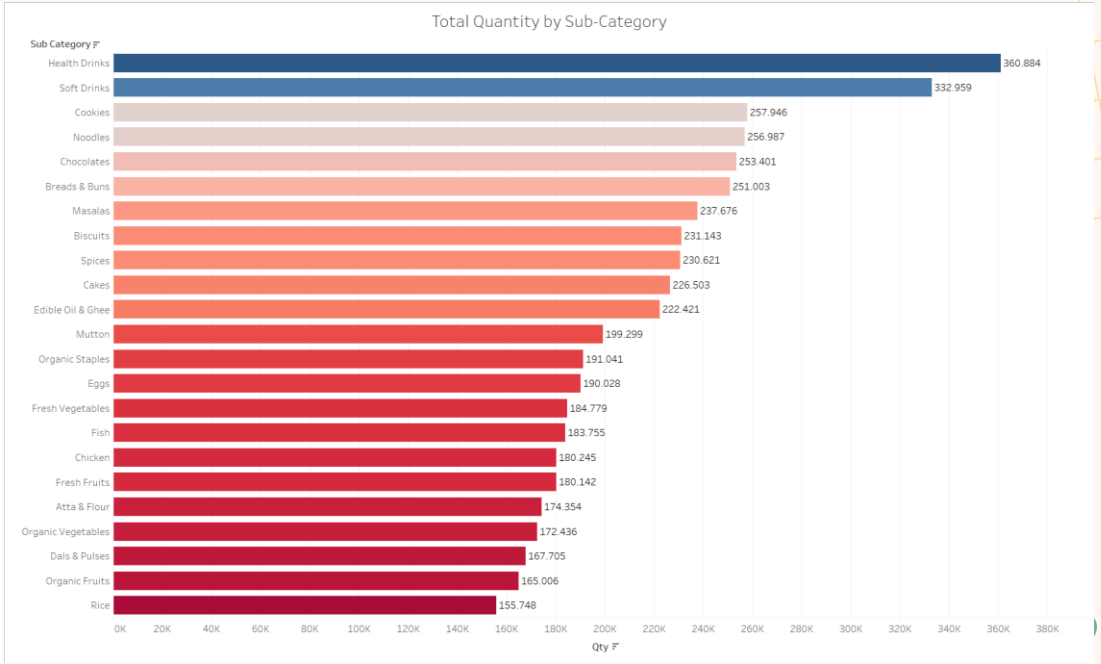
Which sub-category sold the most items?

Snack is the most popular items, selling **768.334** item. **Food grains** is the least popular items, selling 688.848 item.



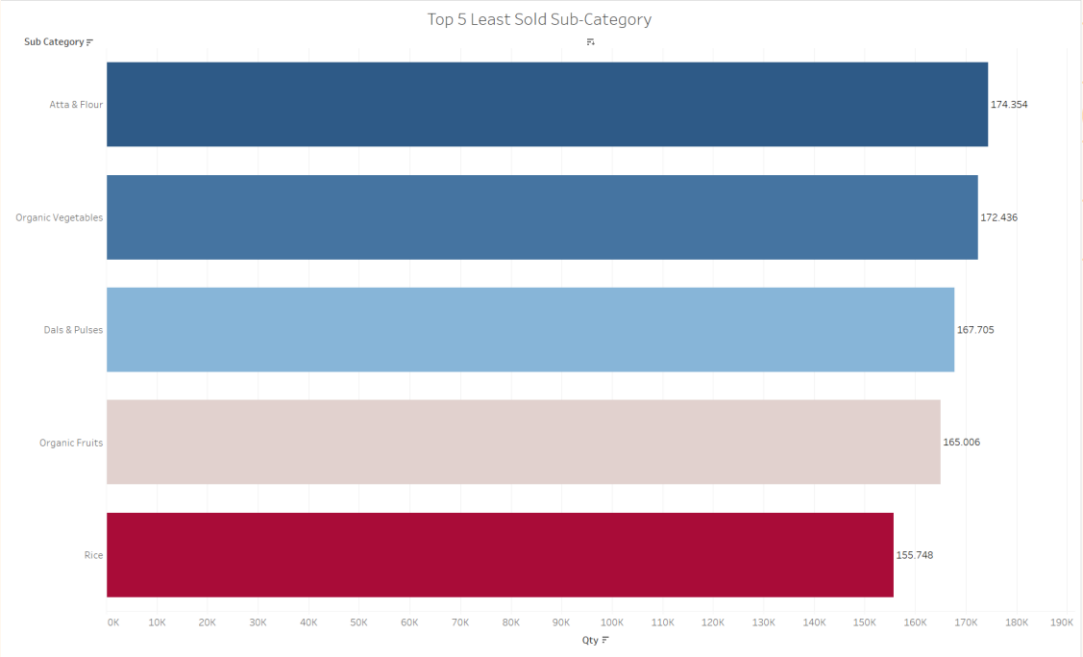
Which sub-category sold the most items?

Health & Soft drink is the most popular items, **360.884 and 332.959. Rice** is the least popular items, selling **155.748** item.



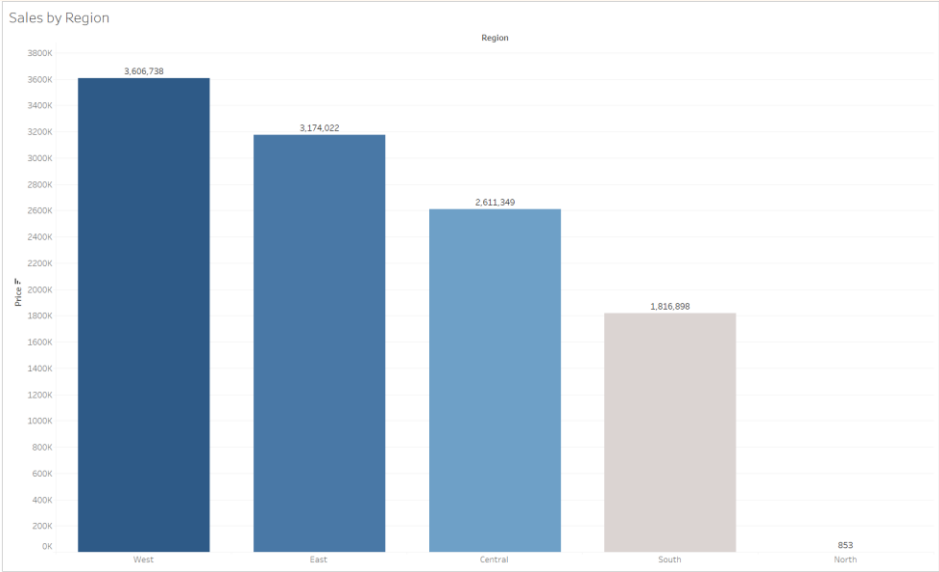
What are the 5 products that sold the least?

Rice, Organic Fruit, Dals & Pulses, Organic Vegetables, and Atta & Flour, is top 5 product with least sold in 4 years



Which region generates the highest sales?

The **west** region generates the most sales among the others as much **3.606.738**. And **North** just generate **853.**

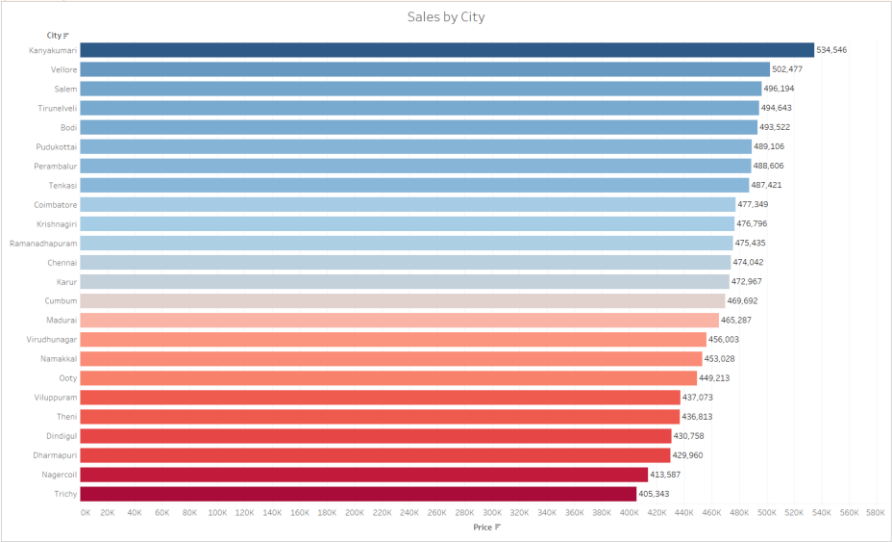


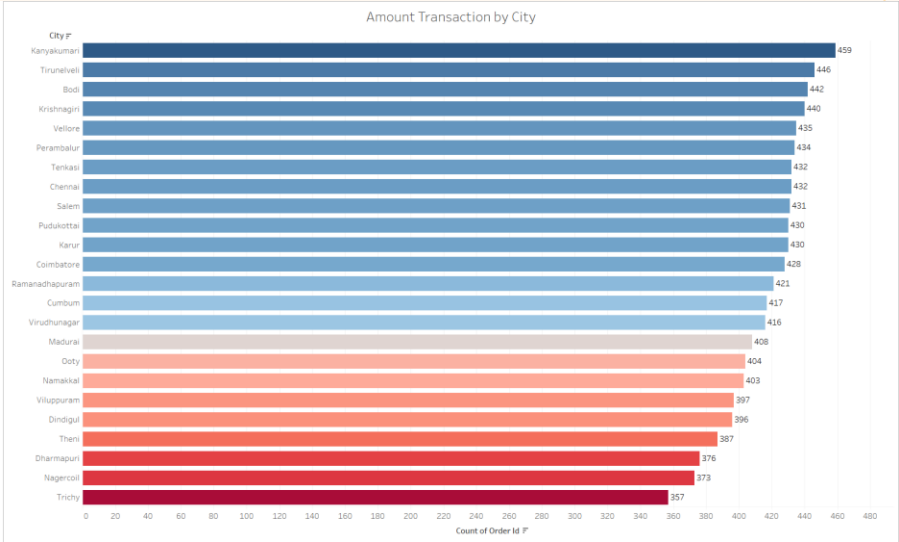
Which region made the most transactions?

West is the region with the most transactions, **3202 times** and **North just 1 transaction**

Which city makes the most transactions and generates the most sales?

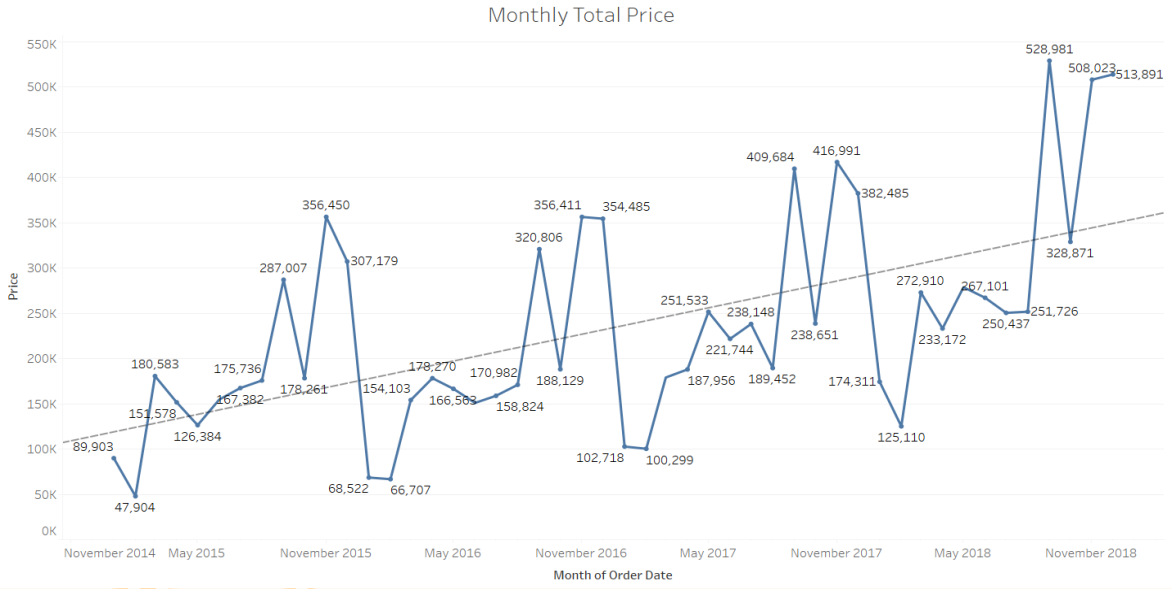
Kanyakumari generates **most sales** than others and also do the **most transactions.**
Trichy generates **least sales** and **least transaction.**





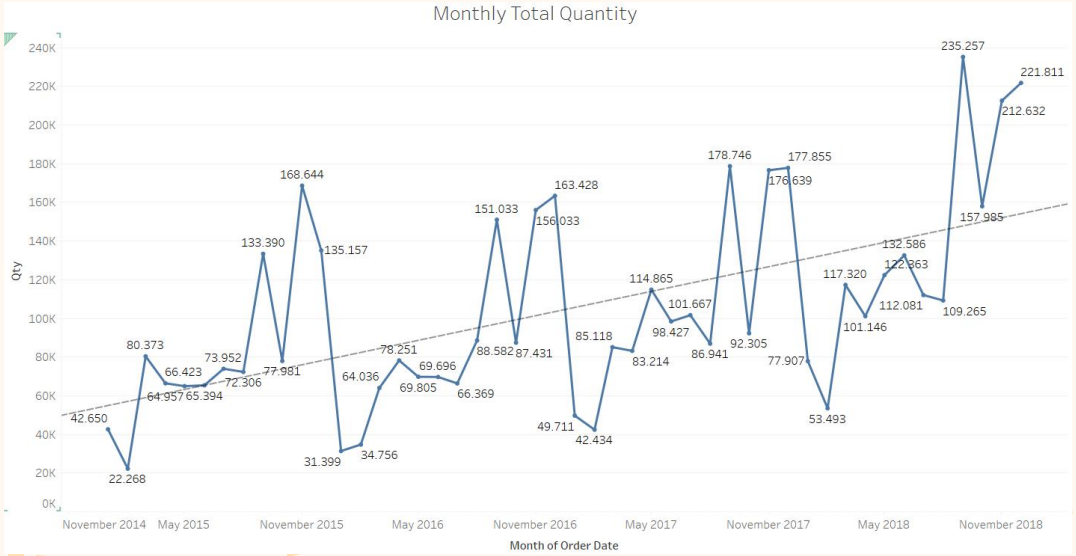
How sales trend in every month?

There is an **increasing trend** for sales every year, every year there is a **similar pattern of trends** that occur for sales. Ex. there was an **increase** in sales in **August - September** and a **decrease** in sales in **December - January**



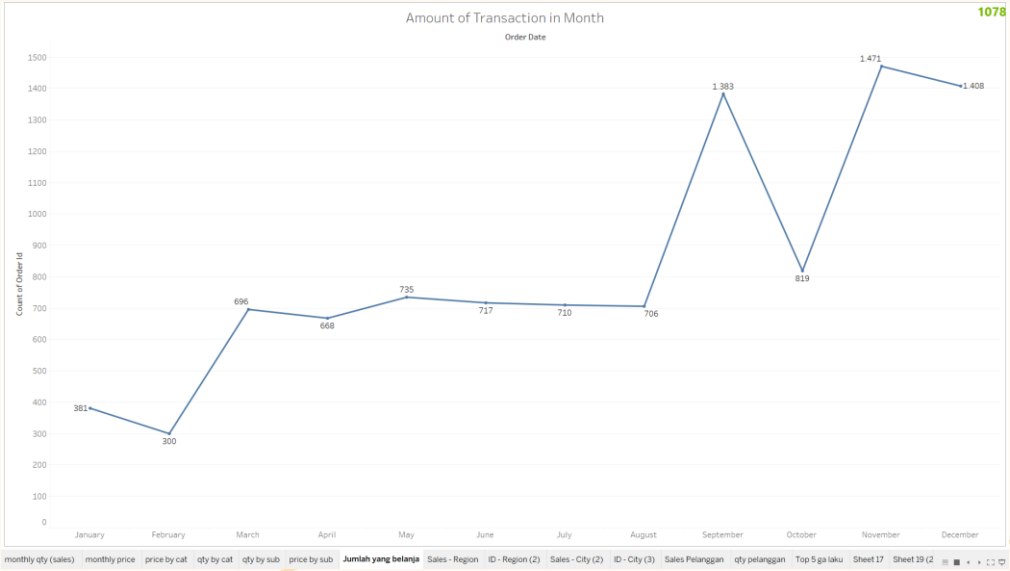
How quantity trend in every month?

The quantity trend in the last 4 years, **has a pattern that is almost the same** as the sales trend. Shows these two things have a **strong correlation**.
Ex. the highest quantity was in November 2018, as well as the sales.



In what month does the customer most often make transactions?

From this chart, we know **Most transactions occur in September** (1383 Times), **November** (1471 Times), **December** (1408 Times). From **march to august there were stable transactions. T**he fewest transactions occur in **January-February**

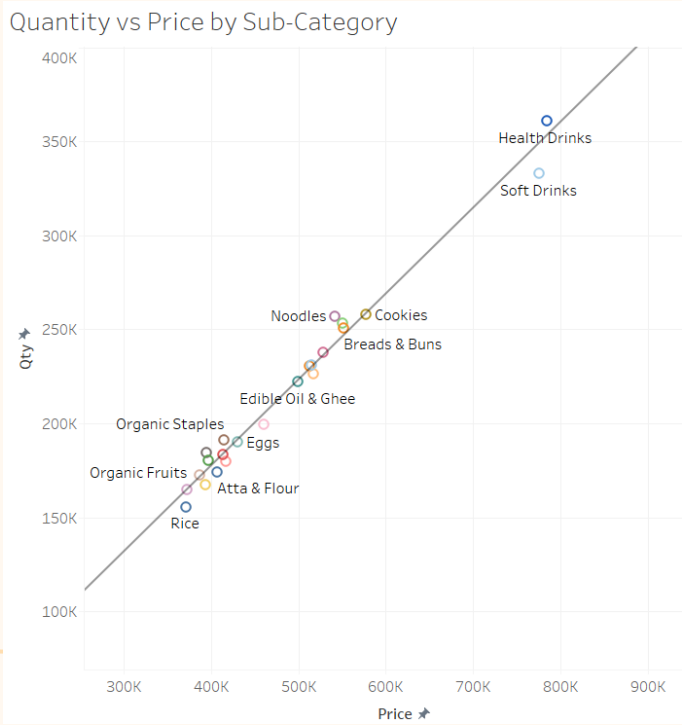


What recommendations can be made from these trends

- We can adjust the **purchased quantity for stock** inventory **based on demand** where **September and October** are the best-selling months.
- **Reducing purchased quantity in January-February** because customers have **spent their money at the end of the year**, and prefer to **save their money** more at the beginning of the year.
- **Conducting extensive campaigns** for the **most frequently purchased products** in January-February
- **Providing promos** for products that might sell well according to the season, for example **summer drink products**
- These things increase the possibility for us to increase sales in months that are quiet in transactions

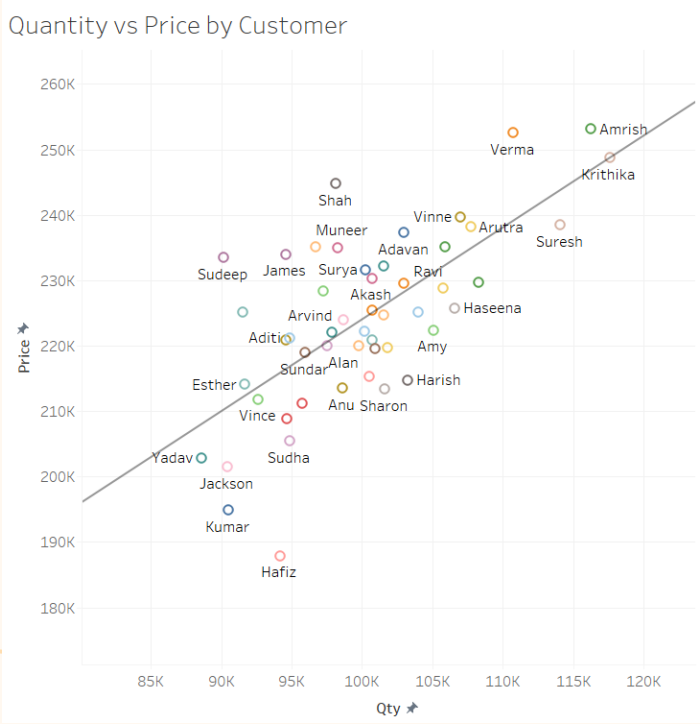
How is the correlation between Quantity and price based on sub-category?

This plot shows a positive correlation between the two. the more items sold, the more sales generated.
Health & soft drink sell the most items and generate the highest sales, and **Rice** sell the fewest items and generate the fewest sales as well



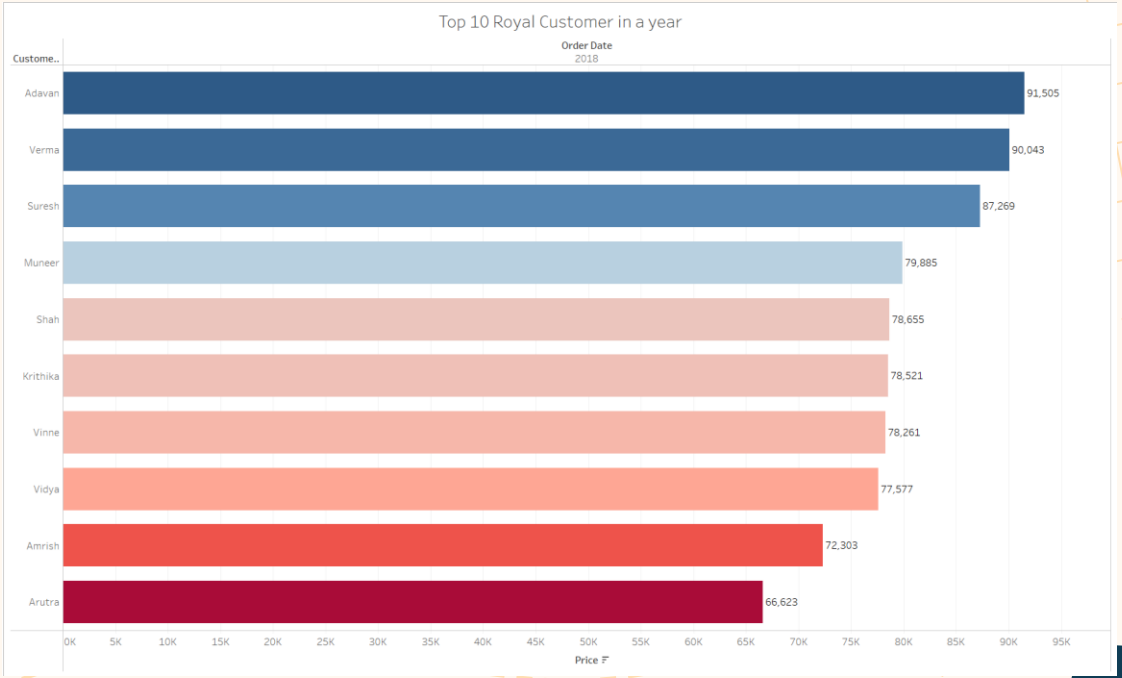
How is the correlation between Quantity and price based on Customer?

There are customers who buy more goods and there are also those who spend more money. Ex.
Krithika buy more items than **Amrish** but, **Amrish** spend more money than **Krithika.**



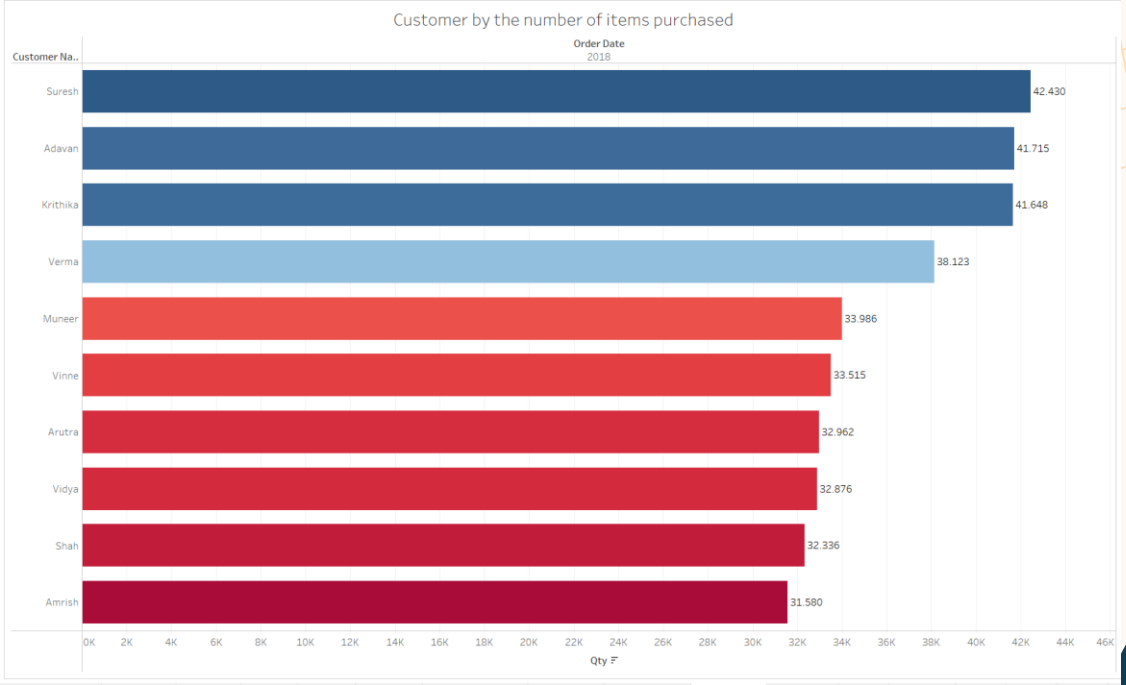
Who are the top 10 customers who spend the most money in 2018?

This is top 10 royal customer in 2018, **Advan** spends the **most money in 2018** as much as 91,505



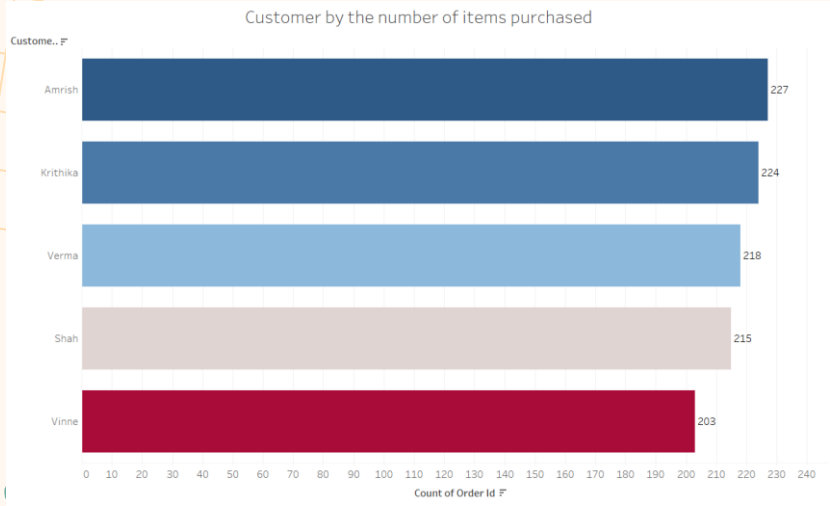
Who are the top 10 customers who bought the most items in 2018?

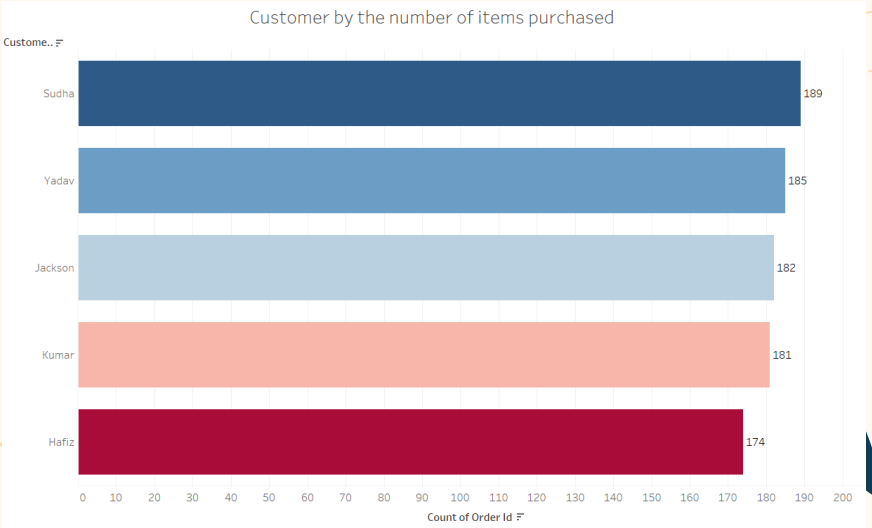
Suresh bought the most items in 2018, but he is **ranked 3rd for spend money.** This proves that **Suresh buys cheap goods more often.**
And **Adavan buys expensive items than Suresh.**



Who are the most loyal customer?

Amrsih is the most loyal customer in this case with **227 times transactions**, and **hafiz** is the **least transactions** with **174 times** in 4 years.
We can do some **clustering based on this analysis**, to **treat customer better** than before.





What can be done to retain customers?

1.Frequent Shoppers:

1. Loyalty Programs: Offer rewards, discounts, or exclusive access to frequent shoppers. Create a tiered loyalty program that provides greater benefits for higher levels of shopping frequency.
2. Early Access: Provide early access to new products or sales events to reward frequent shoppers for their loyalty.

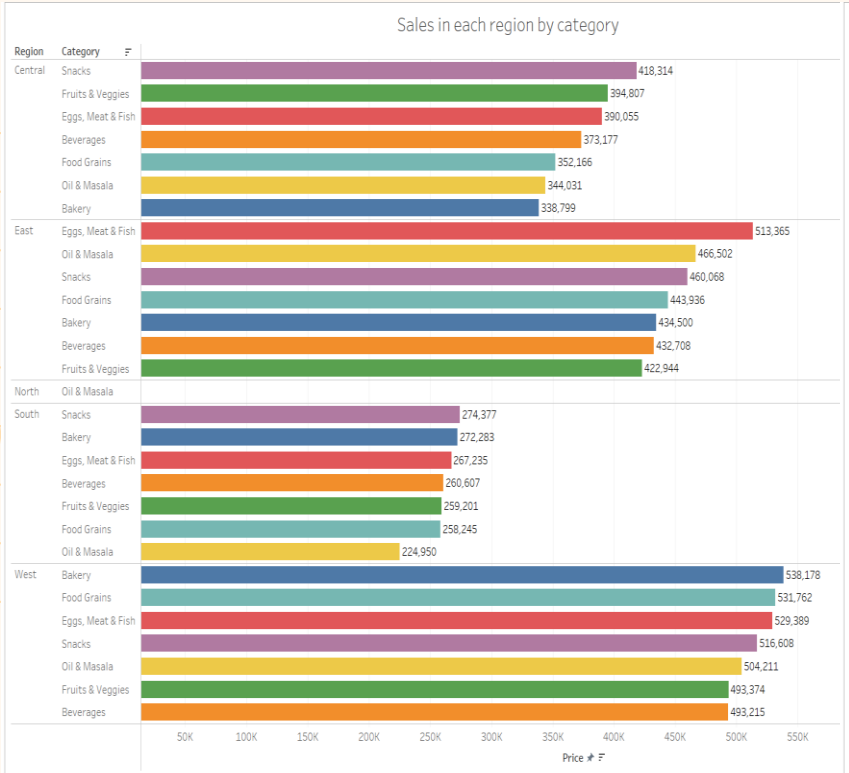
2.High Spenders:

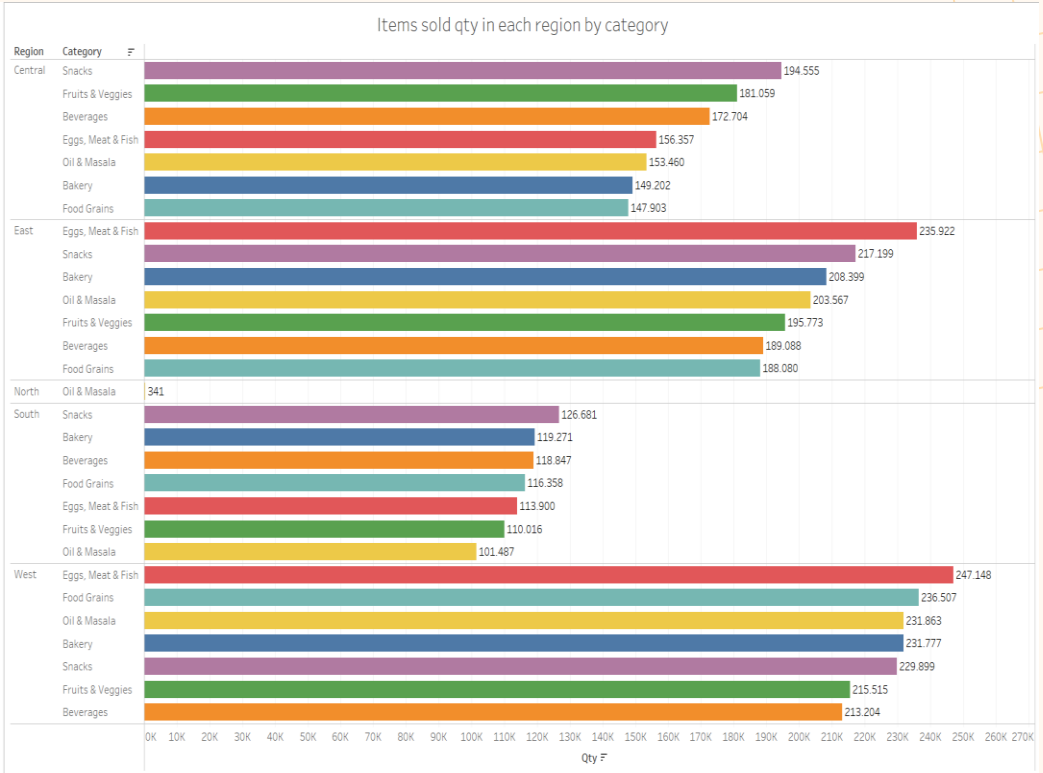
1. VIP Treatment: Offer premium customer service, priority support, and personalized assistance to high-spending customers.
2. Exclusive Offers: Provide special offers or discounts tailored to high-spending customers to make them feel valued.
3. Thank You Gifts: Send personalized thank-you gifts or handwritten notes to show appreciation for their substantial purchases.

3.High Quantity Buyers:

1. Bulk Discounts: Offer discounts for larger quantity purchases to incentivize high quantity buyers to continue purchasing in bulk.
2. Subscription Options: Introduce subscription models for products they frequently buy in large quantities.

What categories generate the most sales and the most sold in each region?





What categories generate the most sales and the most sold in each region?

- There are 2 categories with **most sold and earn high sales**, which is **Snacks & Egg, Meat, Fish**
- **Snacks** are items with the **highest income and sales** in the **central** and **south** regions.
- **Meat** is **top 3 category with highest generates sales** in **all regions** except north.
- **Bakery** is the category with the **highest income in the west region** even though the **amount sold is not as much as Egg, Meat, Fish, oil & masala, and food grain.**
- The **central** region is the region with the **highest sales and consumption** of **fruit/vegetables** among other regions.
- **Egg, Meat, Fish** is the most popular item in **east** region.
- The **west** region is the region with the **highest sales and consumption** of **food grains** among other regions.
- The **north** region just buy 1 category which is **Oil & masala.**

What can be done to increase sales and the quantity of items sold in each region?

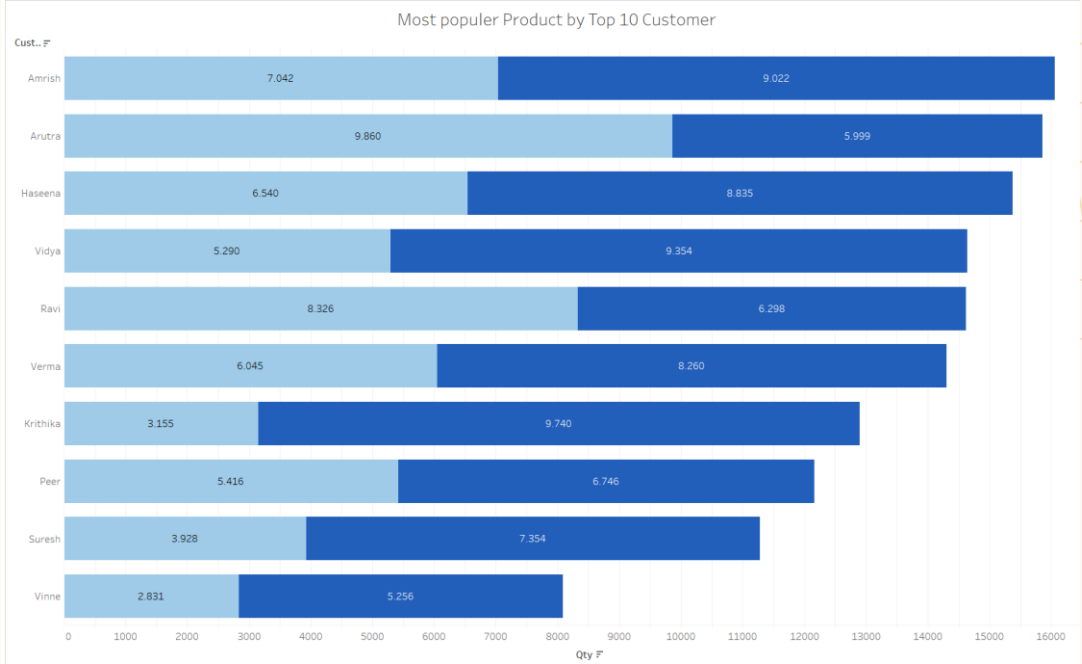
- Paying special attention to the **northern region** because there is **only 1 transaction in 4 years** by providing **discounts, promos, bundles** for **goods with low to medium prices** to see the market in the region.
- Conducting special campaigns in each region based on **the highest sales for several existing categories.** In order to further **increase sales and the amount sold** than before.
- Provide **bundle packages for 2-3 items** for the **5 most popular items** in each region.

What 2 products are the most popular in the top 10 customers who buy the most?

Health drink & soft drink is the **most popular item** in the 10 customers with the **most item purchases** and also for all customers

*We can **update or improve product packaging** to attract attention and **create a strong brand image.**

*Continue to **improve the taste and quality** of healthy/soft drink products so that they **remain delicious and satisfying for customers.**



What item sold the most in the 3 months with the fewest sales?

These items **sold the most** in the month with the **lowest sales.**
We can get the opportunity to **increase sales** if we focus on some of these items in each month, by **conducting campaigns and promotions.**

Sub Category	January
Soft Drinks	17.791
Health Drinks	14.122
Noodles	13.586
Edible Oil & Ghee	12.612
Masalas	12.557
Spices	11.592

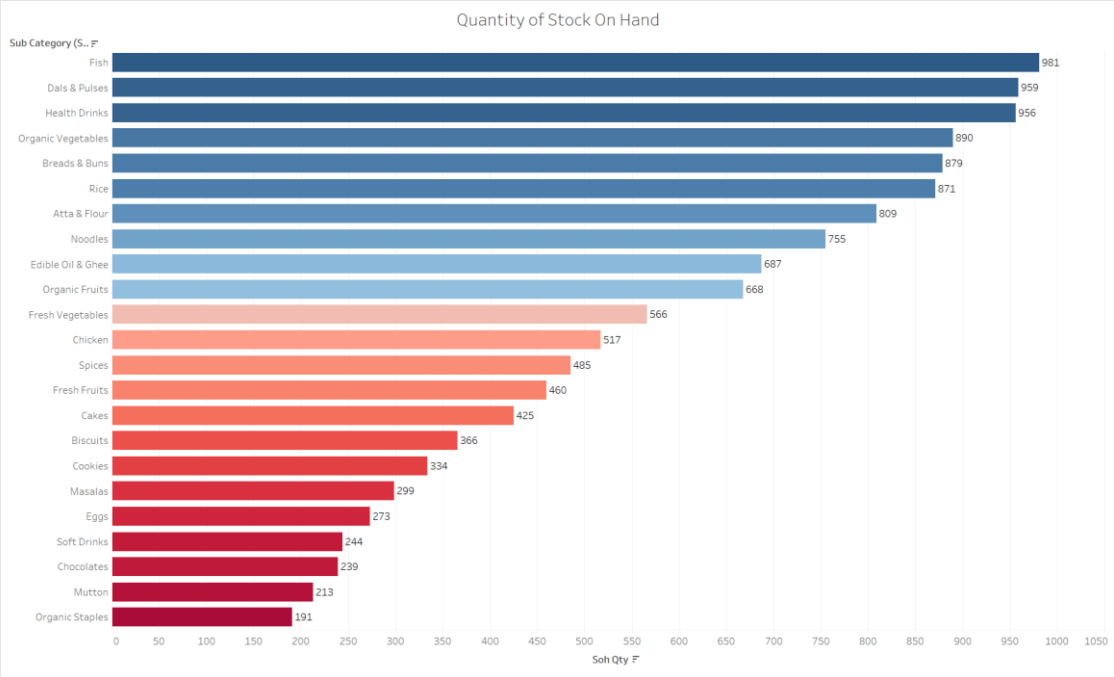
Sub Cat (Sub Category)	February
Chocolates	12.397
Biscuits	11.586
Soft Drinks	10.884
Fresh Fruits	9.513
Masalas	9.042

Sub Category	October
Health Drinks	32.471
Soft Drinks	26.284
Organic Staples	24.178
Masalas	24.020
Breads & Buns	22.459

How many quantity stock on hand?

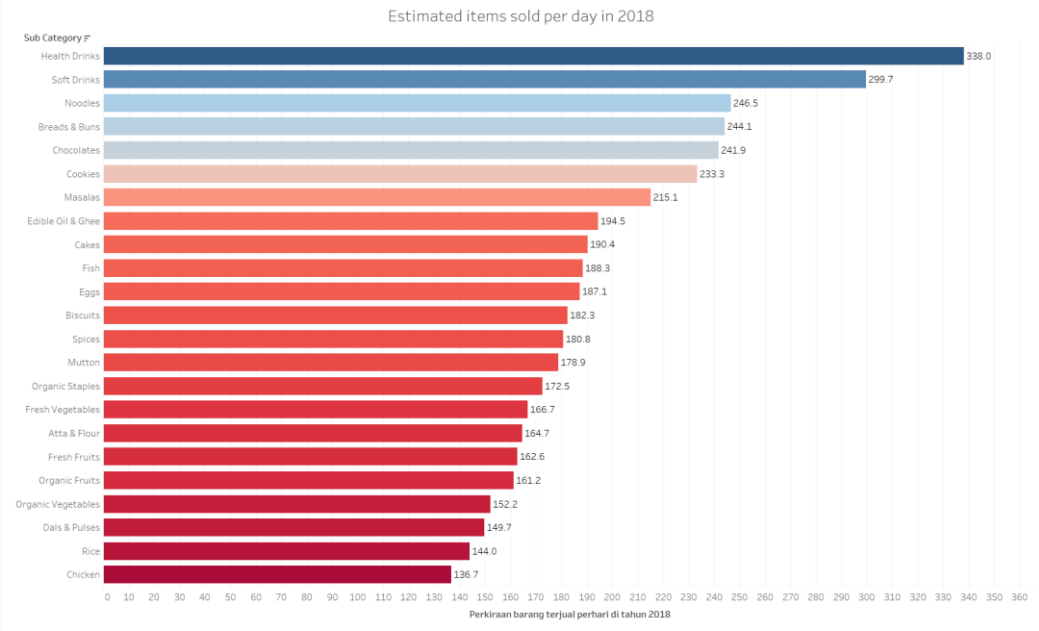
I assume that SOH is **stock that must be prepared** every day

From that chart, we have **Fish, Dals & Pulses, Health Drinks** is top 3 product with high quantity in Stock On Hand which is more than **900 item** per day.
And **organic staples** only prepared **191 item** per day



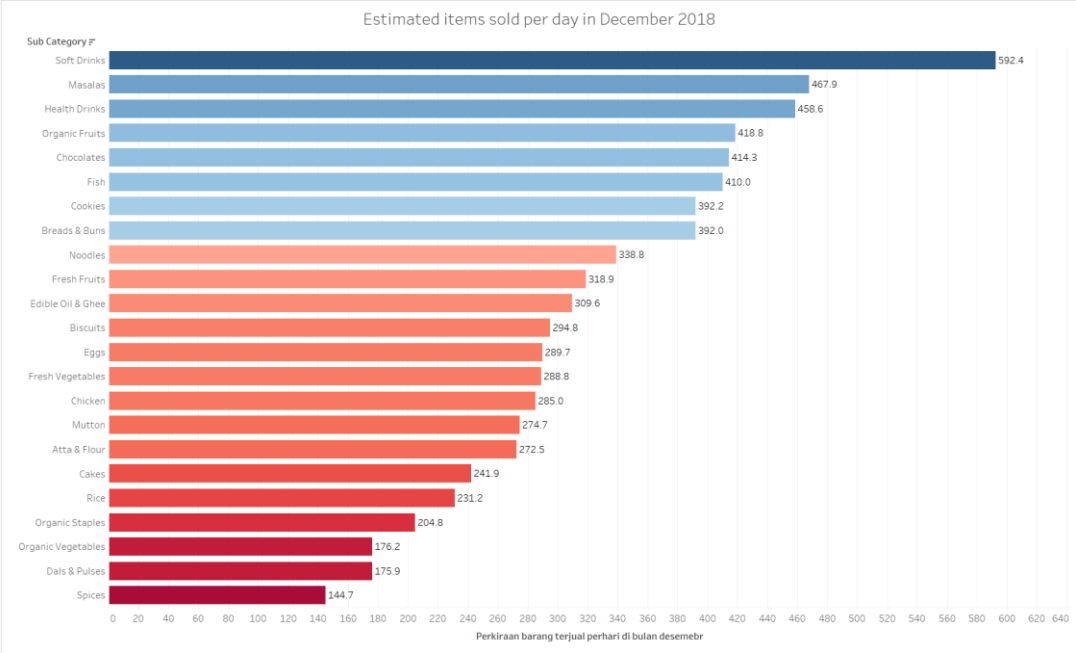
What is the estimated number of items sold in 2018?

From the **SOH data**, I looked up the **estimated number of items sold each day** in 2018, and the result is
Health drink and soft drink still in the **highest ranking for items sold** per day, approximately **300 items** per day.
Rice and chicken is the item with the **lowest estimated sales** each day, **less than 150** item per day



What is the estimated number of items sold in December 2018?

We compare with how many estimated sales in the last 1 month (December 2018). **Soft Drink & Health drink** still in top ranking, also **masalas** with more than **450 items in a day**.



What is the comparison between SOH quantity and sales quantity?

Had to adjust some of the existing items, because there were a lot of items in SOH but the sales were less than SOH.

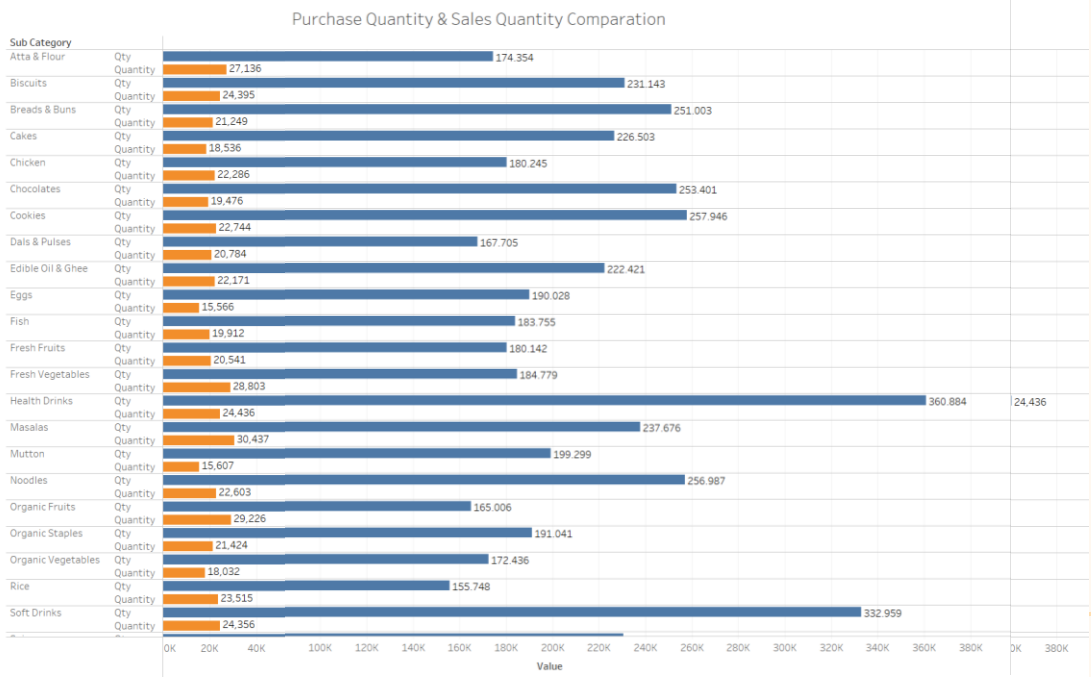
- During 2018, **beverages (Health & Soft drink)** were the **best-selling products**. However, in **stock on hand** the two products are **not balanced**, we have to **reduce** quantity of **health drink stocks and increase** quantity of **soft drink stocks.**
- Had to **reduce** quantity of **stocks of fish, Dals & pulses** because the stocks were **very large** compared to the estimated sales of these products. Where **Dals & Pulses** is one of the **most rarely purchased products.**
- **Organic vegetable** stock quantity also needs to be **reduced** because it is a product that is **rarely purchased**

What needs to be done to the condition of the difference in SOH quantity and Sales quantity?

- By **making adjustments** to product stock quantity, of course we can **save on purchasing** in the following days because there is no need to spend money to **purchase many items** that may **not necessarily sell well.**
- **Reducing perishable goods** (vegetables, fruit, bread) so that customers can buy the **freshest goods** and can also **reduce the burden of purchasing** if these goods are not selling.
- Make **good storage management** to store items that have been **purchased a lot** and **items that have not been sold**

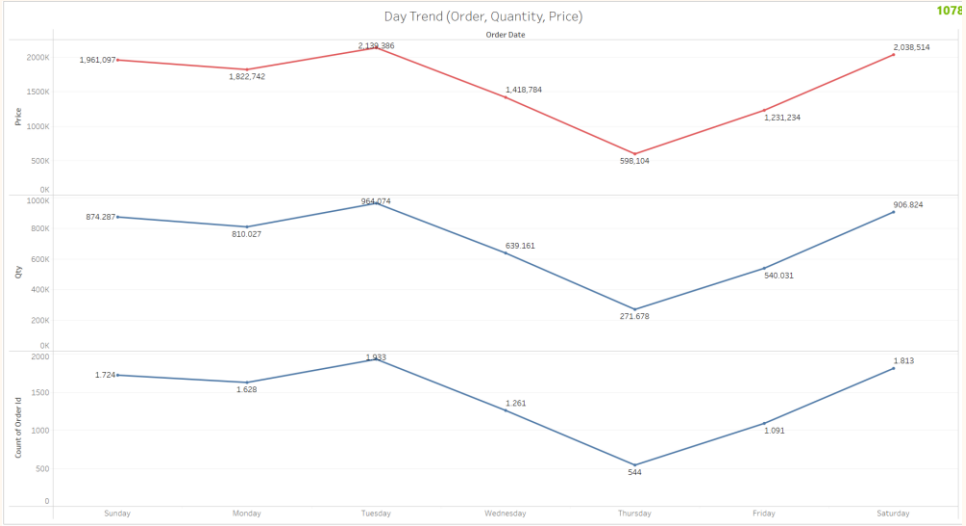
How is the comparison between purchase quantity and sales quantity?

- There is a very **significant difference** where the sales quantity is **not proportional** to the quantity purchased. Where the two quantities should not have much difference, these differences may have an **impact on stock** being **unavailable** if there is **high demand**



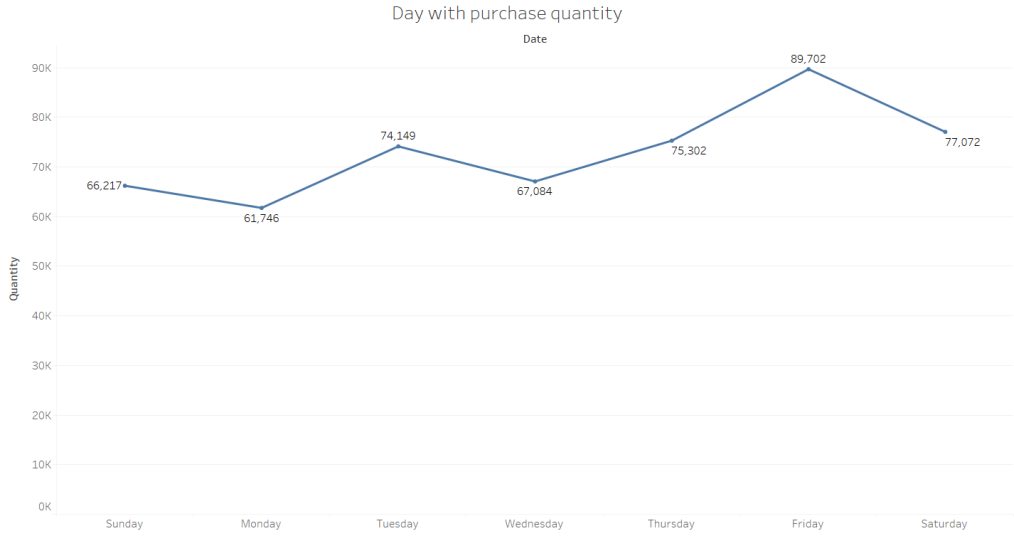
How the sales trends, quantity and transaction every day?

- From 3 variable, they have same trend pattern. Most **orders, quantity, and generates sales** occur on **Tuesdays** and **Saturdays.** That day is the best time to give promotions in increasing sales.



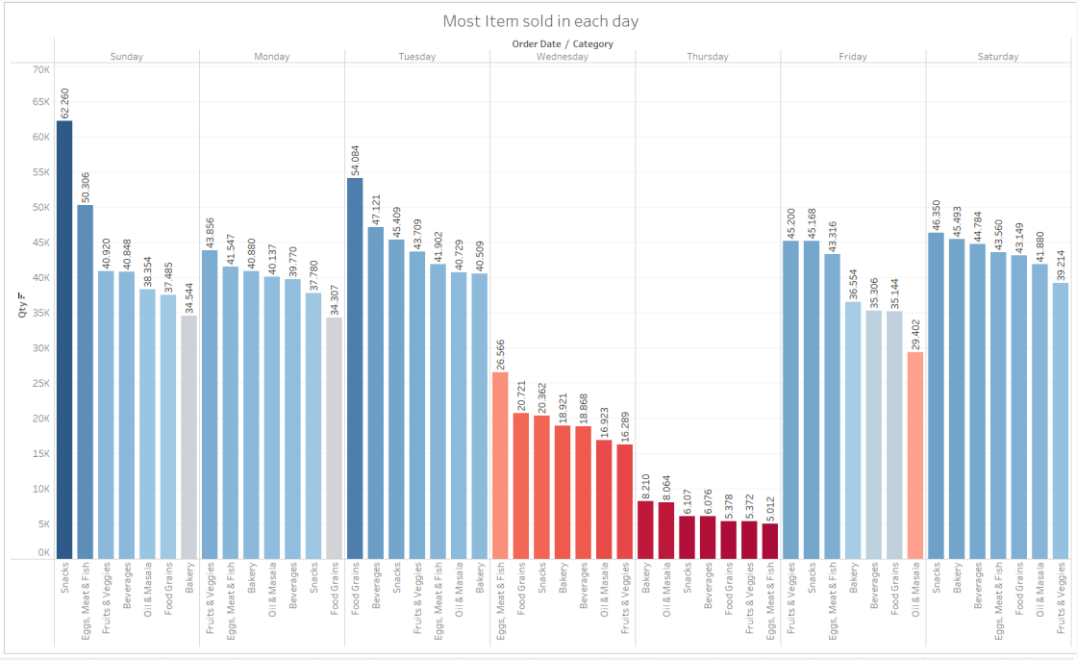
What days do you make purchases the most?

- If we look at to the quantity sold trend in every day, the best day for purchased quantity is Thursday to prepare on hand stock which will be sold on Friday-Tuesday.
- Meanwhile, we should reduce the purchased quantity on Tuesday because on Wednesday and Thursday sales will decrease than usual



What categories are selling the most on a daily basis in 2018

- Every day sold from different categories, **snacks** sell better on **weekends**. Fruits **and veggies** are mostly sold on **Monday and Friday**, **food grains** are mostly sold on **Tuesday** and **thursday** is the **fewest selling** days of the week.



What can we do from that Line Chart trend & Bar chart Before?

- On certain days, especially on weekends, you can run **a 'weekend sales' campaign** with the aim of **maintaining previous sales or increasing sales** because customers tend to shop on weekends. As well as the **Weekend Snacks campaign** because on **Saturday and Sunday customers will prefer snacks**

- Can provide **promos** for **categories, fruits & veggies, and meat on Mondays**. Then **food grains, beverages and snacks on Tuesday** to entice customers to spend their money.

- Adjusting the **time to be purchased** with the trends that have been obtained from sales trends and transaction trends and **adjusting what categories are purchased every day** for adjust **the stock and demand** from customers who shop

02
Forecasting

Process Step by Step

This is an outline of the forecasting process that will be carried out

Data Preprocessing

Data Cleaning, Choose the value to be forecasted

Stationary Test

Stationary test using Augmented Dickey-Fuller

Modelling

Create Model using ARIMA Method

Future Forecasting

Forecasting price/sales in future

Data Preprocessing

Data Cleaning

Make some observations on the data, and make sure there are no null or duplicated values

Adjustments to the order date column

Convert 'order_date' column to datetime, Set 'order_date' as the index

Aggregation

Resample data by month and calculate the sum of prices

Stationary Test

```
Observations of Dickey-fuller test
Test Statistic                  -3.622187
p-value                          0.005351
#lags used                       0.000000
number of observations used     47.000000
critical value (1%)             -3.577848
critical value (5%)             -2.925338
critical value (10%)            -2.600774
dtype: float64
```

Based on the results of the Dickey-Fuller test given, the conclusion that can be drawn is that the p-value is 0.005351.

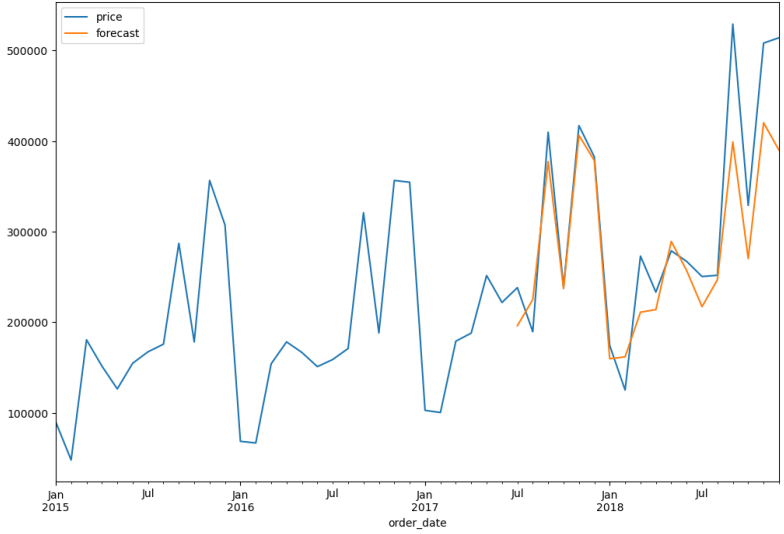
A small p-value (smaller than the specified significance level, for example 0.05) indicates that the null hypothesis can be rejected, so the data can be considered stationary.

Modelling Process

- After the data has been declared stationary, the modeling process can be directly carried out on the data.
- The data consists of 48 instances, divided into 30 data train and 18 data test

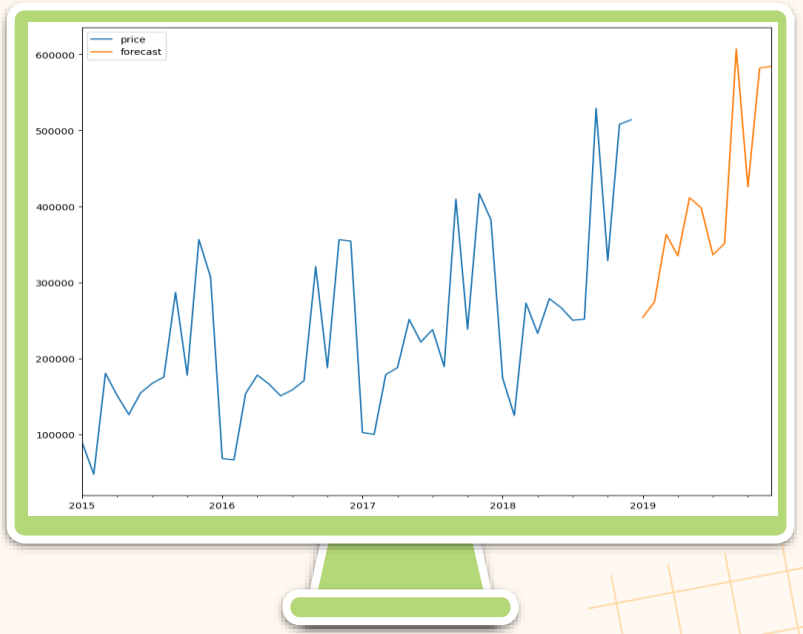
- Forecasting result for data test, has a trend that is not much different from the original data.

- From these results the model is considered good enough for forecasting



Future Forecasting

Based on forecasting results, in the next 1 year the sales trend will increase and also has the same pattern as the previous years.
September is again the month with the highest sales/price and January the lowest




```
2019-01-31     254135.031045
2019-02-28     274582.819762
2019-03-31     363243.670444
2019-04-30     335101.514864
2019-05-31     411596.492412
2019-06-30     398129.193616
2019-07-31     336503.581097
2019-08-31     351328.941273
2019-09-30     607227.470887
2019-10-31     425833.565136
2019-11-30     581958.116537
2019-12-31     584324.269133
Name: forecast, dtype: float64
```

Future Forecasting

Based on forecasting results, of course we can **make preparations** for **purchasing items** in the next 1 year.

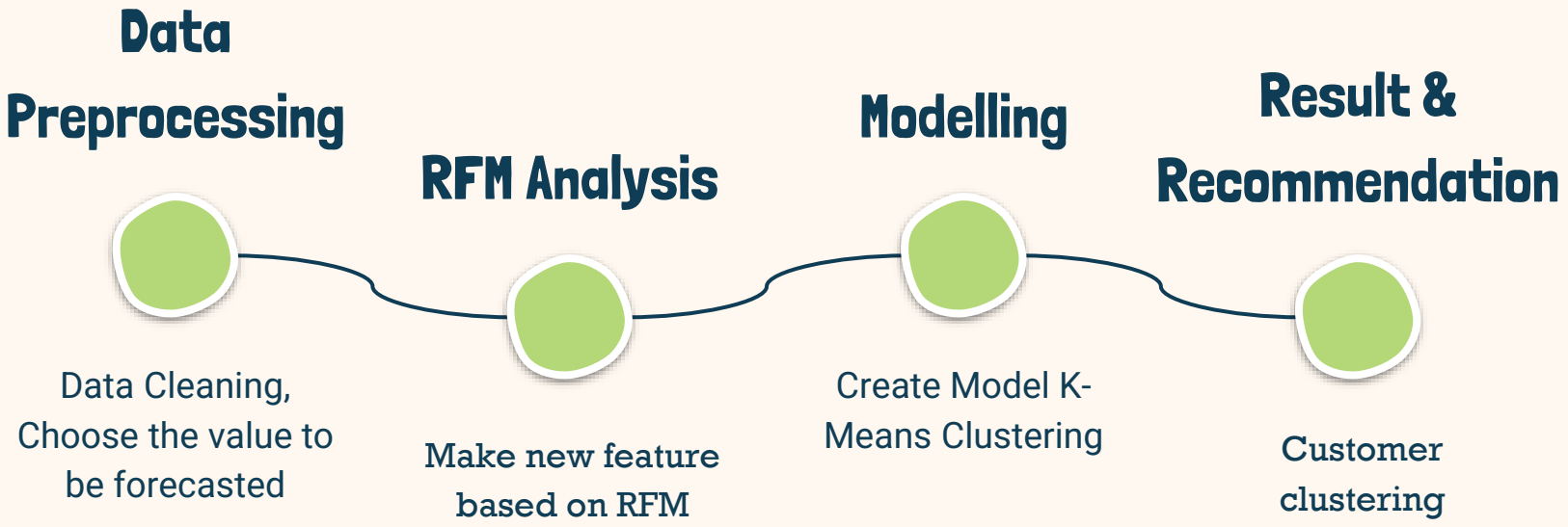
That way **we don't have to spend on purchased items** that don't sell that month, we can see based on the **previous analysis to prepare the purchasing budget.**

From this analysis we can also **prepare campaigns for each product** based on which **ones are selling best in that month** so that these predictions can be achieved

03
Clustering

Process Step by Step

This is an outline of the clustering process that will be carried out



Data Preprocessing

Create RFM Column

Find and Aggregate the related columns to get the RFM value

Data Cleaning

Make some observations on the data, and make sure there are no null or duplicated values and outliers in data.

Standardization

Using StandardScaler and look at the distribution.

RFM Analysis

Recency

Look for the distance of the day the last customer made a transaction and the last time in the data

Frequency

Looking for how many times to make transactions

Monetary

How much money has been spent

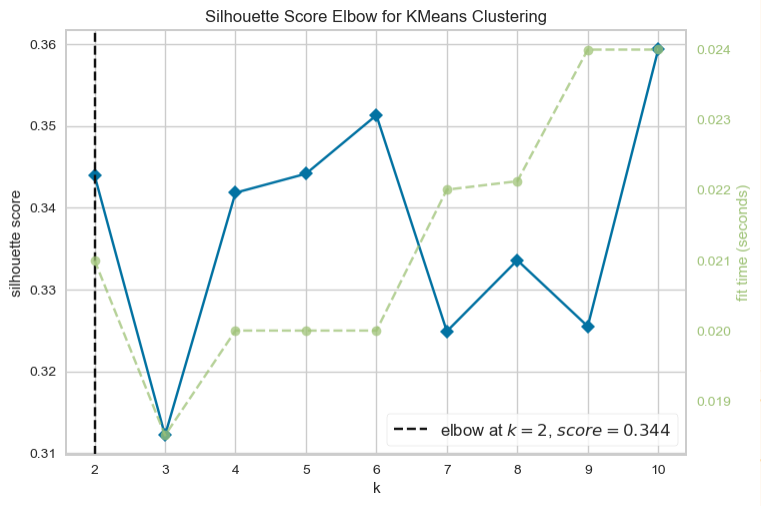
Result

	customer_name	frequency	monetary	recency
0	Adavan	205	237296.71	9
1	Aditi	187	221234.95	3
2	Akash	196	225366.74	4
3	Alan	198	219986.77	8
4	Amrish	227	253159.11	5

Modelling Process

- Find best K using silhouette score
- Best K=10, but we only have 50 data. I think 10 cluster is too much for 50 data. For this case, I will choose 6 cluster

- Build K-Means Model with 6 cluster



Clustering Result

Cluster 0

Frequent, high spending, short recency.

Cluster 1

Consistent buying, moderate spending, short recency.

Cluster 2

Faithful, high spending, longer recency intervals.

Cluster 3

Infrequent, moderate spending, longer recency intervals.

Cluster 4

Moderately Frequent, high spending, short recency, premium preferences.

Cluster 5

Infrequent, cautious spending, longer recency intervals.

Recommendation

Clusters 0:

- Provide a lot of special treatment in this cluster, such as special products that are only available for customers who have made purchases regularly and have shopped frequently.
- Offer Repeat Purchase Discounts: Provide additional discounts on subsequent purchases as an incentive for customers to shop again.

Cluster 1:

- Doing the same thing as in cluster 0 but slightly reduced the number of special products available and the discount is not as much as cluster 0.

Cluster 2:

- Send newsletters about the latest items or promos so that this cluster can shop more often than before.
- Offer Limited Time Discounts: Provide special discounts with a tight time limit, encouraging quick purchases.

Recommendation

Cluster 3:

- Can do the same things as cluster 2, but the newsletters and discounts offered are not as much as in cluster 2.

Cluster 4:

- Early Access Offers: Give customers in this cluster early access to new sales or products.
- Make offers for existing premium products, of course, included with other benefits.

Cluster 5:

- Offer Affordable Products: Provide a selection of products at lower prices to encourage more purchases.
- Give a special promo if this cluster will make a transaction after a long time of not making a transaction, you can also provide credit for this cluster transaction.

Thanks!