Cafe Chain Data

BY- SAUMYA JAIN
(PGPDSBA.O.JUNE23.B)

Contents

Agenda:

- Executive Summary of the data
- Exploratory Data Analysis
- Market Basket Analysis
- Associations Identified
- Recommendation

Problem Statement

- The data set provided to you is the data set of a Cafe Chain for one of its restaurants. Do a thorough analysis of the data and come up with the following analysis. The owner of the restaurant wants you to use this data to come up with a set of recommendations that can help his Cafe Chain increase its revenues. He is able to provide you with a data set for POS (point of sale data) for one of his chains.

Executive Summary:

- Data: from 01-04-2010 to 0 9 - 07-2010

- Objective: Project involves conducting a thorough analysis of Dataset: 145831 Rows, 10 columns with a set of recommendations that can help his Cafe Chain increase its revenues with a data set for POS (point of sale data) for one of his chains

- Missing values : No missing values present

- Market basket analysis : Project involves conducting a thorough analysis of Dataset: 145831 Rows, 10 columns with a set of recommendations that can help his Cafe Chain increase its revenues with a data set for POS (point of sale data) for one of his chains

- We also refer to it as frequent itemset mining or association analysis.

Data Dictionary

DATE = DATE AT WHICH PRODUCT IS SOLD

BILL _NUMBER = BILL ASSOCIATED
 WITH THE ORDERED ITEMS

ITEM_DESC = ITEM
description of item
THAT IS SOLD

Assumptions:

- The data represents a list of items purchased at a restaurant on various dates.

- Each entry in the data represents a single item purchased.

- There is information provided about the quantity and total price

- The first column in the data represents the date the product was purchased.

- The second column represents the customer who has done the purchase.

- The third column represents the item which they have purchased.

- We have not dropped the duplicated values because of drop duplicate rows, it may inadvertently remove valid information from the dataset.

- The same item can be purchased by multiple customers on different dates.

Exploratory Analysis

- Weekly, Monthly, Quarterly, Yearly Weekday Trends in Sales count

- Products counts & Year Wise top products

- Summary and Recommendations

Yearly Count of Products Sold

- As we have data till july 2011 that's why the count of products sold in 2011 is low

- As compared for 2010 and 2011 the products sold in 2010 is more as compared to 2011.

Count of quantity per year
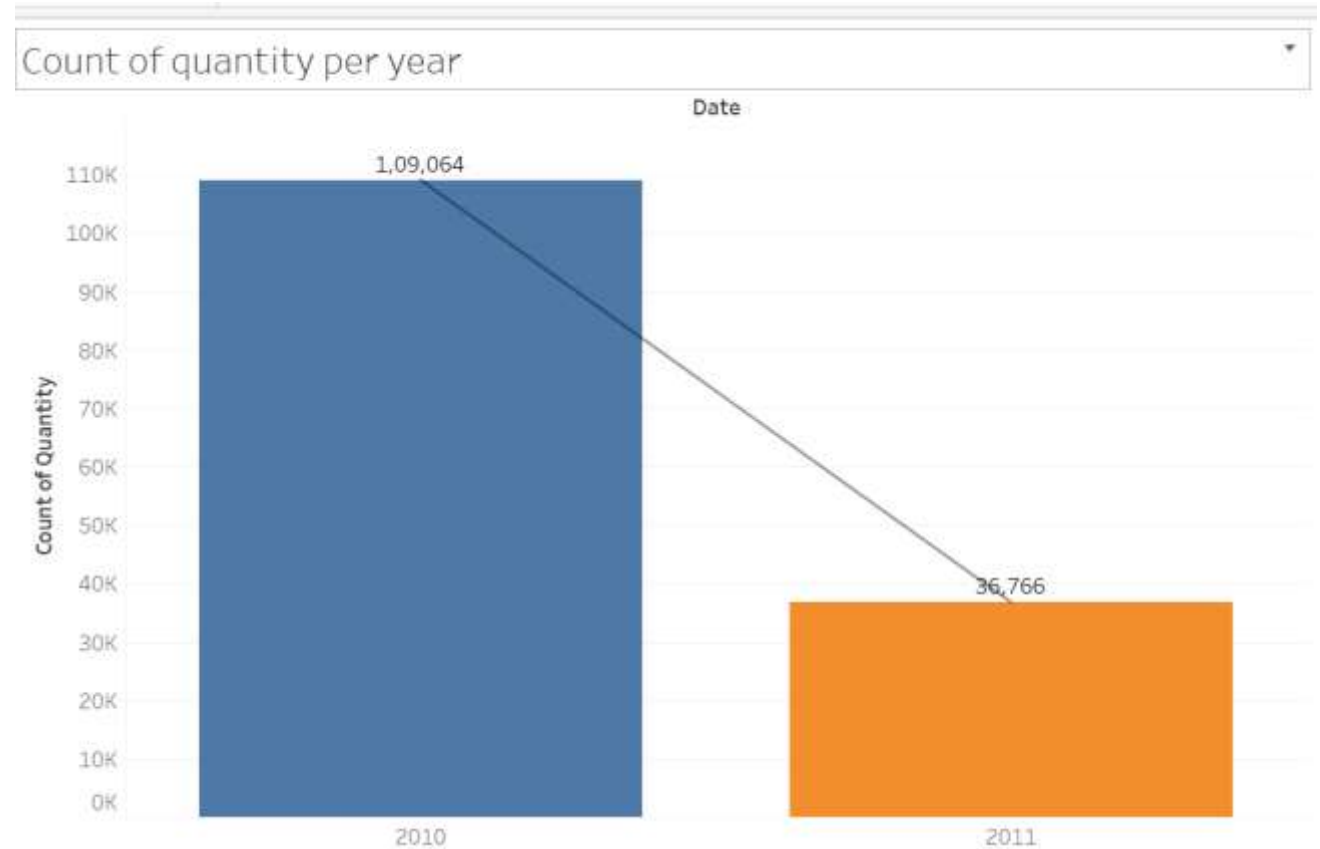
Category wise count

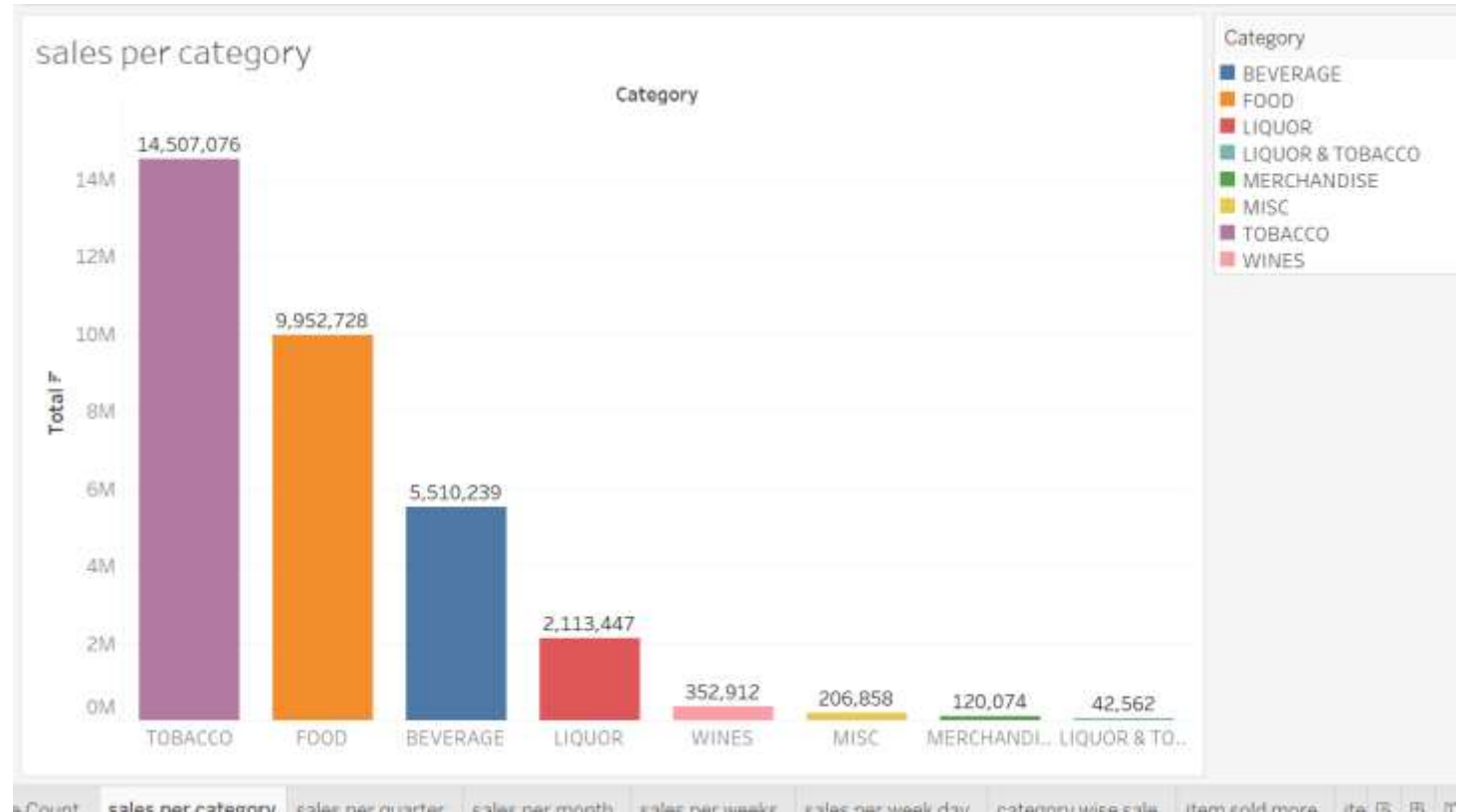Food Is more likely to be get order followed by beverage and Tabacco.

Least ordered is the category for liquor and Tabacco with a count of only 54 times only.
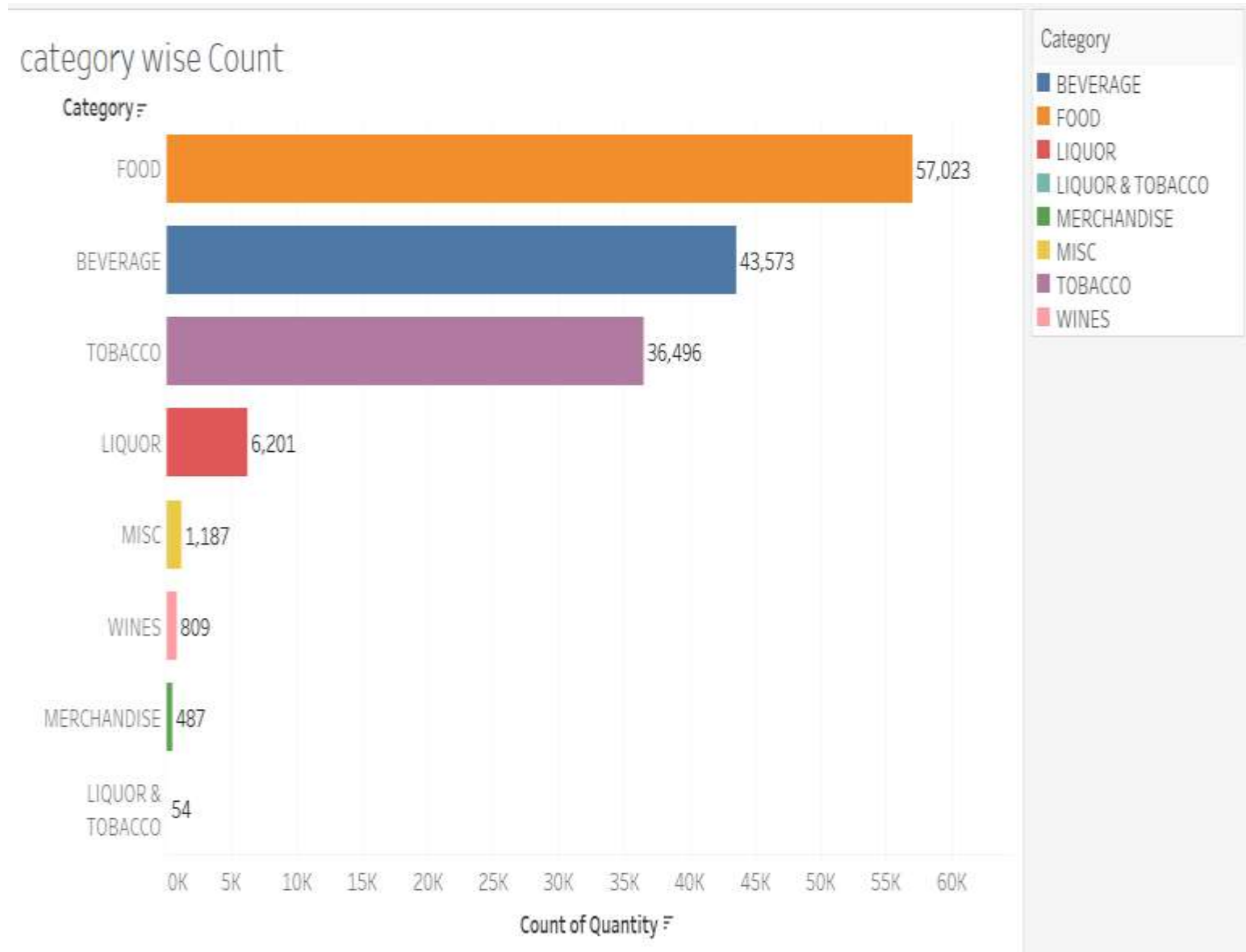
SALES PER CATEGORY

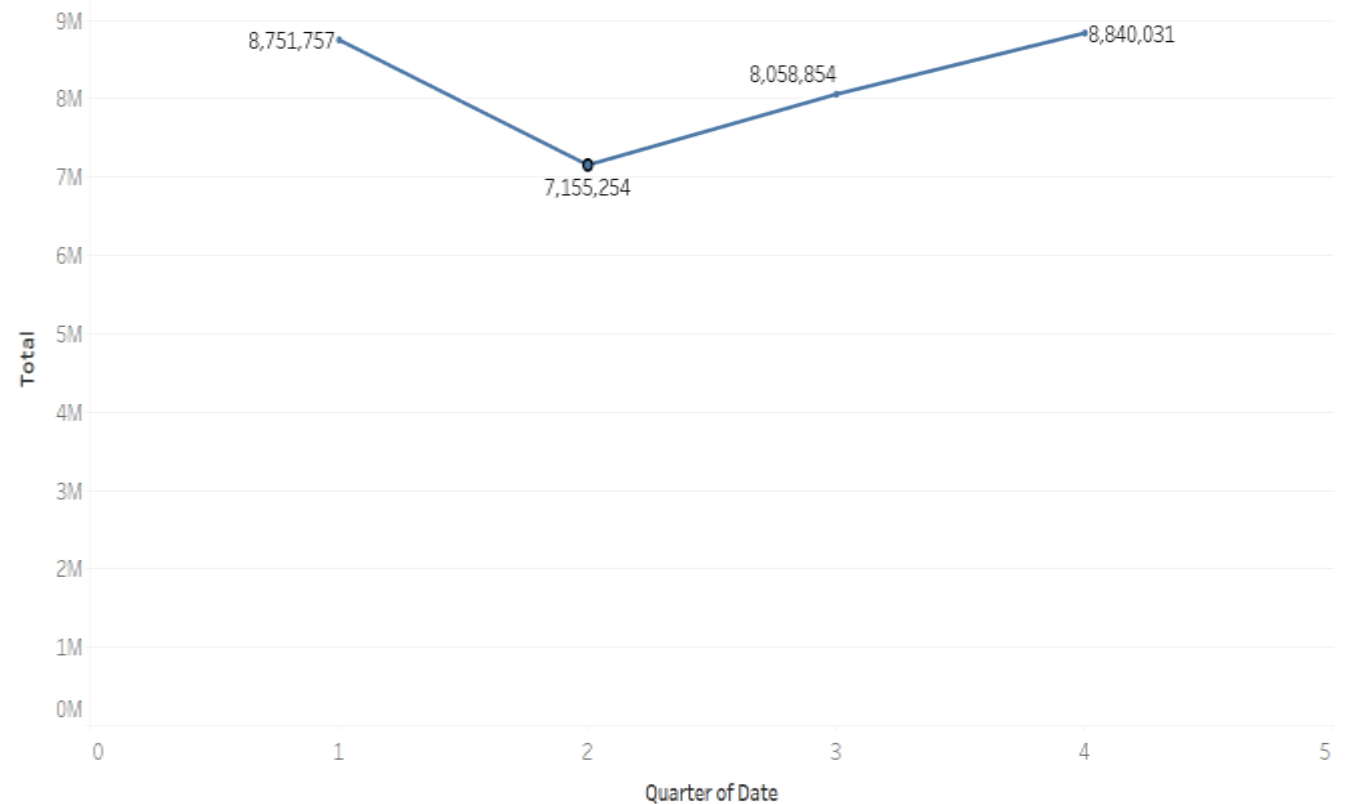Tabacco Is more likely to be get sales followed by food

Least ordered is the category for liquor and Tabacco with a count sales of only 42K.
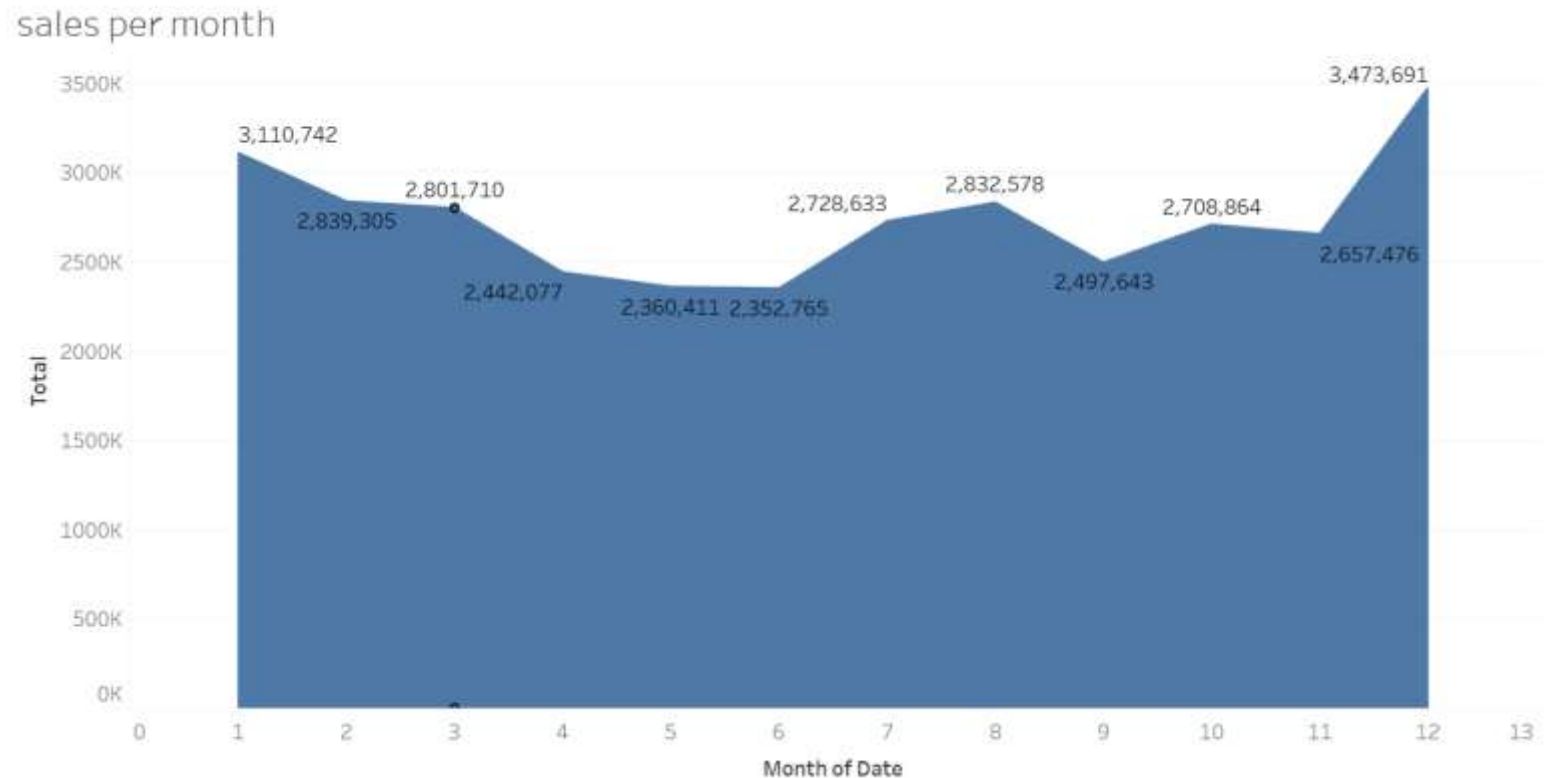
Sales per Quarter

- Q2 performed low as compared to all other quarters.
- Q4 performed very well followed by Q1.
- Q3 the sales is nearby around 80 lakh.

sales per quarter

Monthly Products sold

- Total sales is highest in December month.

  - 2nd highest sales has been done in January month
  - The minimum sales has been done in April and May month

sales per month

Weekly count of the Quantity sold

Sales per weekday

sales per week day

4,987,908
4,379,457
4,222,992
4,361,117
4,361,245
4,851,137
5,642,039

Total

Weekday of Date

Highest product sold count

Top 5 Product sold in year 2010 and 2011

SUMMARY

- Food Is more likely to be get order followed by beverage and Tabacco.

- Least ordered is the category for liquor and Tabacco with a count of only 54 times only

- Tabacco Is more likely to be get sales followed by food

- Least ordered is the category for liquor and Tabacco with a count sales of only 42K

- Q2 performed low as compared to all other quarters.

- Q4 performed very well followed by Q1.
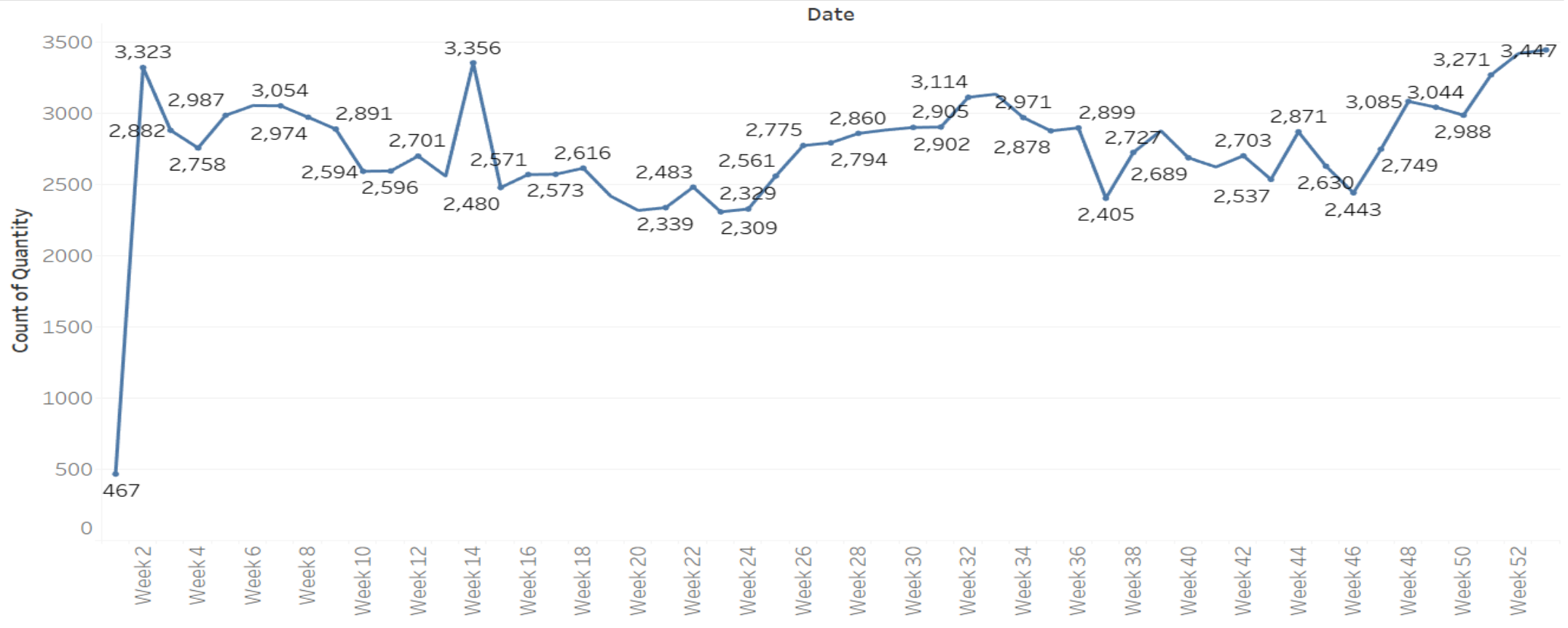
- Q3 the sales is nearby around 80 lakh.

- Total sales is highest in December month.

- 2nd highest sales has been done in January month

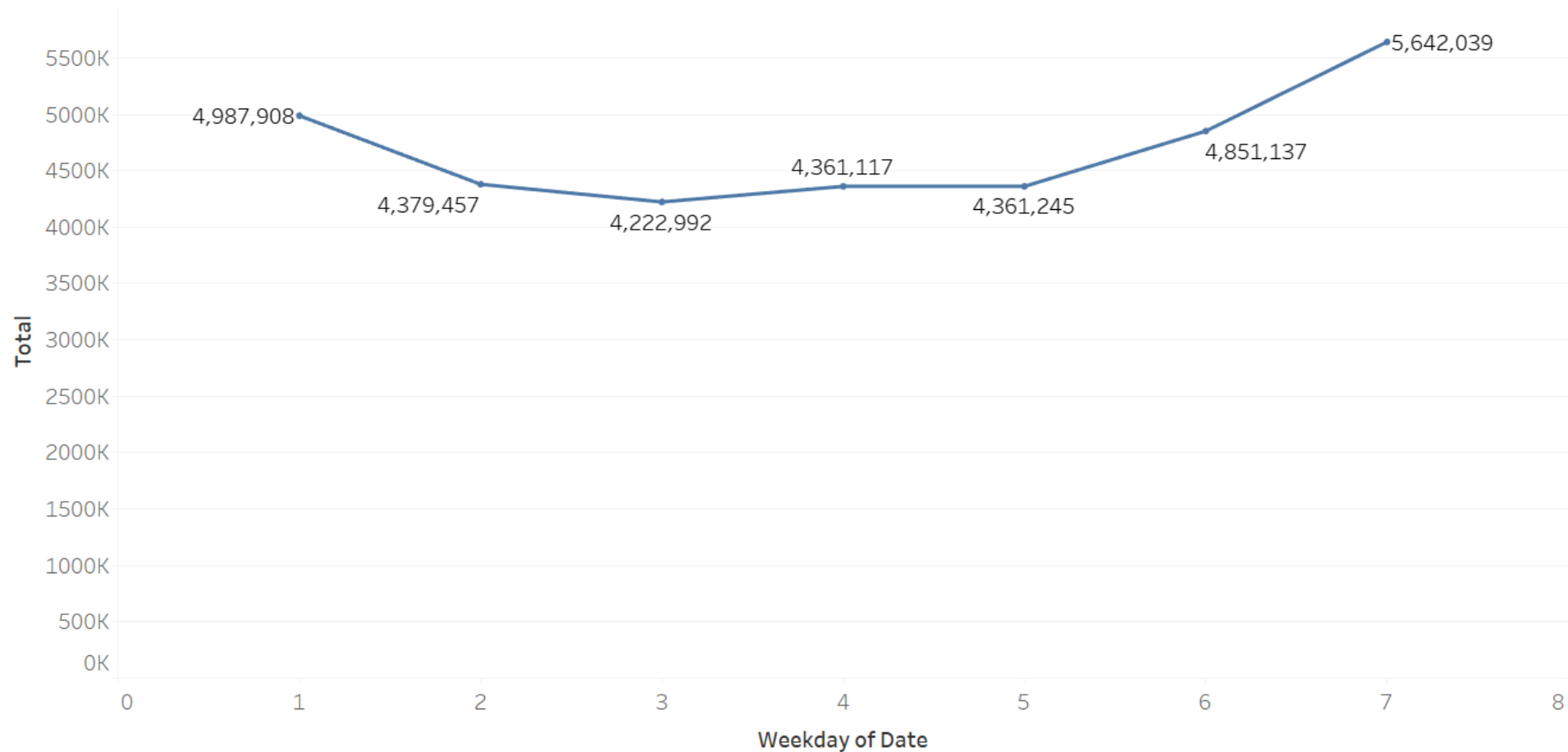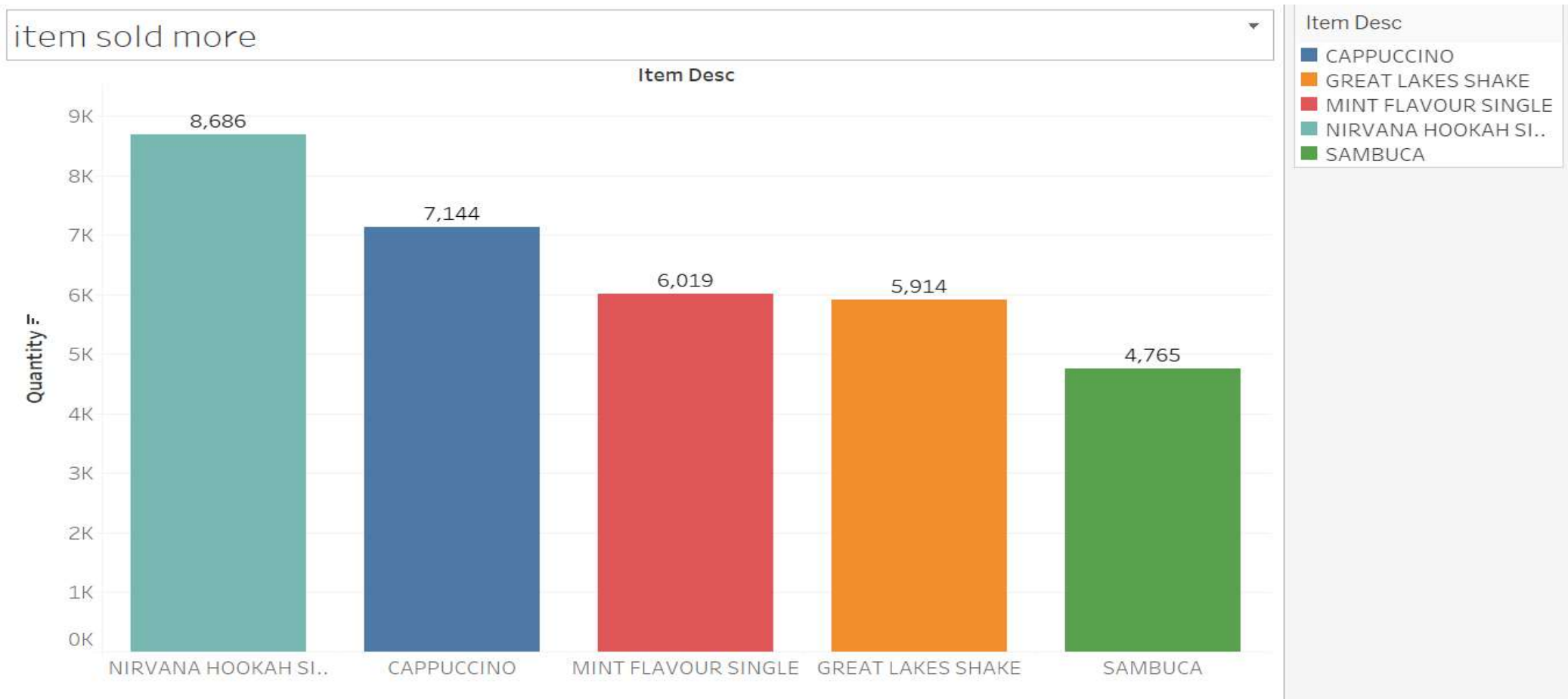- The minimum sales has been done in April and May month

Recommendation:

- Great Lakes shakes need to be combined with recommended items Nirvana Hookah Single and vice versa

- Great Lakes shakes need to be combined with recommended items Cappuccino and vice versa

- Basket Item i.e. Mint Flavors single with recommended to Cappuccino
- Noticed that during the month of May , Tobacco has a brighter consumption pattern than most other categories . Hence , the promotion of Tobacco can be increased to tap this trend

- Inclusion of combo meals will help in boosting the revenue during the lean period and overall since Food is a versatile category here

MARKET BASKET ANALYSIS

Shopping basket → Shopping basket recommended

- Market Basket Analysis Meaning
- MRA KNIME WorkFlow & Qutput Table

MARKET BASKET ANALYSIS MEANING

- Market basket analysis is a data mining technique used by retailers to increase sales by better understanding customer purchasing patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings, as well as products that are likely to be purchased together.

MRA KNIME WORKFLOW
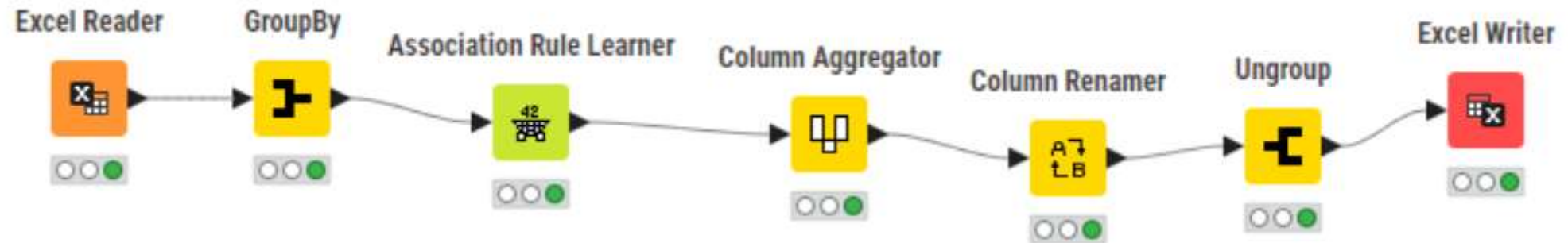
OUTPUT TABLE

Support	Confidenc	Lift	recommended Items	implies	Basket Items
0.00513	0.073701	0.614826	NIRVANA HOOKAH SINGLE	<---	GREAT LAKES SHAKE
0.00513	0.042794	0.614826	GREAT LAKES SHAKE	<---	NIRVANA HOOKAH SINGLE
0.005473	0.045655	0.960045	QUA MINERAL WATER(1000ML)	<---	NIRVANA HOOKAH SINGLE
0.005473	0.115084	0.960045	NIRVANA HOOKAH SINGLE	<---	QUA MINERAL WATER(1000ML)
0.005544	0.079655	1.019463	CAPPUCCINO	<---	GREAT LAKES SHAKE
0.005544	0.070958	1.019463	GREAT LAKES SHAKE	<---	CAPPUCCINO
0.005987	0.072092	0.92267	CAPPUCCINO	<---	MINT FLAVOUR SINGLE
0.005987	0.076628	0.92267	MINT FLAVOUR SINGLE	<---	CAPPUCCINO
0.00633	0.12976	1.082471	NIRVANA HOOKAH SINGLE	<---	POUTINE WITH FRIES
0.00633	0.052807	1.082471	POUTINE WITH FRIES	<---	NIRVANA HOOKAH SINGLE

Threshold Values of Support and Confidence

- Support of Minimum: 0.005

- Maximum Item Set Length : 10

- Minimum Confidence Level: 0.04

Market basket analysis, support, confidence, and lift values

- In market basket analysis, support, confidence, and lift values are used to measure the strength of association between items in a transaction dataset.

- **SUPPORT**: Support is a measure that gives an idea of how frequent an itemset is in all the transactions. It measures how frequent the itemset occurs in the dataset. High support indicates that the itemset is popular and should be considered for promotion or placement together.

- **LIFT:** Lift controls for the support (frequency) while calculating the conditional probability of occurrence of {Y} given {X}. It measures the strength of the association between two items. High confidence indicates that the items are likely to be bought together and can be used to recommend or suggest items to customers.

- **CONFIDENCE:** Confidence measures the likelihood of items given that the shopping cart already has other items. It is calculated as the support of the itemset divided by the product of the individual supports of the items. A lift value of 1 indicates that the items are independent, while a value greater than 1 indicates a positive association between the items. A lift value less than 1 indicates a negative association between the items. High lift indicates that the items have a strong association and can be used for cross-selling or bundling.

Association Rules

- Association rules are a technique used to find relationships or associations between items in a large dataset. These rules are based on the concept of frequent item sets, which are sets of items that appear together frequently in a transactional dataset.

- 10 rules have been found with the dataset and set parameters.

Support	Confidenc	Lift	recommended Items	implies	Basket Items
0.00513	0.073701	0.614826	NIRVANA HOOKAH SINGLE	<---	GREAT LAKES SHAKE
0.00513	0.042794	0.614826	GREAT LAKES SHAKE	<---	NIRVANA HOOKAH SINGLE
0.005473	0.045655	0.960045	QUA MINERAL WATER(1000ML)	<---	NIRVANA HOOKAH SINGLE
0.005473	0.115084	0.960045	NIRVANA HOOKAH SINGLE	<---	QUA MINERAL WATER(1000ML)
0.005544	0.079655	1.019463	CAPPUCCINO	<---	GREAT LAKES SHAKE
0.005544	0.070958	1.019463	GREAT LAKES SHAKE	<---	CAPPUCCINO
0.005987	0.072092	0.92267	CAPPUCCINO	<---	MINT FLAVOUR SINGLE
0.005987	0.076628	0.92267	MINT FLAVOUR SINGLE	<---	CAPPUCCINO
0.00633	0.12976	1.082471	NIRVANA HOOKAH SINGLE	<---	POUTINE WITH FRIES
0.00633	0.052807	1.082471	POUTINE WITH FRIES	<---	NIRVANA HOOKAH SINGLE

SUMMARY

- Food Is more likely to be get order followed by beverage and Tabacco.

- Least ordered is the category for liquor and Tabacco with a count of only 54 times only

- Tabacco Is more likely to be get sales followed by food

- Least ordered is the category for liquor and Tabacco with a count sales of only 42K

- Q2 performed low as compared to all other quarters.

- Q4 performed very well followed by Q1.

- Q3 the sales is nearby around 80 lakh.

- Total sales is highest in December month.

- 2nd highest sales has been done in January month

- The minimum sales has been done in April and May month