Marketing and Retail Analytics

Project - Milestone 2

Agenda & Executive Summary of the data

- **Contents**
  * Data set and problem statement
  * Exploratory Data Analysis and Inferences
  * Market Basket Analysis
  * Inferences from MBA analysis and Recommendations

- **Problem statement**

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

About data set

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145830 entries, 0 to 145829
Data columns (total 10 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   Date         145830 non-null  object
 1   Bill Number  145830 non-null  object
 2   Item Desc    145830 non-null  object
 3   Time         145830 non-null  object
 4   Quantity     145830 non-null  int64
 5   Rate         145830 non-null  float64
 6   Tax          145830 non-null  float64
 7   Discount     145830 non-null  float64
 8   Total        145830 non-null  float64
 9   Category     145830 non-null  object
dtypes: float64(4), int64(1), object(5)
memory usage: 11.1+ MB
```

- The given dataset contains 145830 rows and 10 columns of data.
- It has four float data type, one integer data type and 5 object data type of columns.
- The dataset has both no null values and no duplicate rows of data
- There is presence of outliers in variables such as Quantity,Rate,Tax and Total.
- Variable 'Discount' has highest positive skewness(106.09) and Variable 'Rate' has lowest positive skewness(1.566)

Data Head

	Date	Bill Number	Item Desc	Time	Quantity	Rate	Tax	Discount	Total	Category
0	01-04-2010	G0470115	QUA MINERAL WATER(1000ML)	13:15:11	1	50.0	11.88	0.0	61.88	BEVERAGE
1	01-04-2010	G0470115	MONSOON MALABAR (AULAIT)	13:15:11	1	100.0	23.75	0.0	123.75	BEVERAGE
2	01-04-2010	G0470116	MASALA CHAI CUTTING	13:17:35	1	40.0	9.50	0.0	49.50	BEVERAGE
3	01-04-2010	G0470117	QUA MINERAL WATER(1000ML)	13:19:55	1	50.0	11.88	0.0	61.88	BEVERAGE
4	01-04-2010	G0470283	MOROCCAN MINT TEA	01:20:18	1	45.0	10.69	0.0	55.69	BEVERAGE
5	01-04-2010	G0470283	QUA MINERAL WATER(1000ML)	01:20:18	1	50.0	11.88	0.0	61.88	BEVERAGE
6	01-04-2010	G0470118	MASALA CHAI CUTTING	13:21:34	1	40.0	9.50	0.0	49.50	BEVERAGE
7	01-04-2010	G0470118	MOROCCAN MINT TEA	13:21:34	1	45.0	10.69	0.0	55.69	BEVERAGE
8	01-04-2010	G0470284	CAPPUCCINO	01:29:38	1	60.0	14.25	0.0	74.25	BEVERAGE
9	01-04-2010	G0470285	MOROCCAN MINT TEA	01:29:46	1	45.0	10.69	0.0	55.69	BEVERAGE

	Date	Bill Number	Item Desc	Time	Quantity	Rate	Tax	Discount	Total	Category
count	145830	145830	145830	145830	145830.000000	145830.000000	145830.000000	145830.000000	145830.000000	145830
unique	365	69982	580	36200	NaN	NaN	NaN	NaN	NaN	9
top	31-12-2010	G0490530	NIRVANA HOOKAH SINGLE	22:25:36	NaN	NaN	NaN	NaN	NaN	FOOD
freq	834	23	8553	33	NaN	NaN	NaN	NaN	NaN	57023
mean	NaN	NaN		NaN	1.121299	161.782259	48.929061	0.095079	224.959852	NaN
std	NaN	NaN		NaN	0.477237	102.244631	40.272851	3.720735	164.960776	NaN
min	NaN	NaN		NaN	1.000000	0.010000	0.000000	0.000000	0.010000	NaN
25%	NaN	NaN		NaN	1.000000	95.000000	22.560000	0.000000	117.560000	NaN
50%	NaN	NaN		NaN	1.000000	125.000000	32.060000	0.000000	167.060000	NaN
75%	NaN	NaN		NaN	1.000000	225.000000	72.000000	0.000000	315.000000	NaN
max	NaN	NaN		NaN	30.000000	2100.000000	2731.250000	825.000000	14231.250000	NaN

- **Maximum** and **Minimum** values for respective numerical variables are highlighted in **Green** and **Red** colors
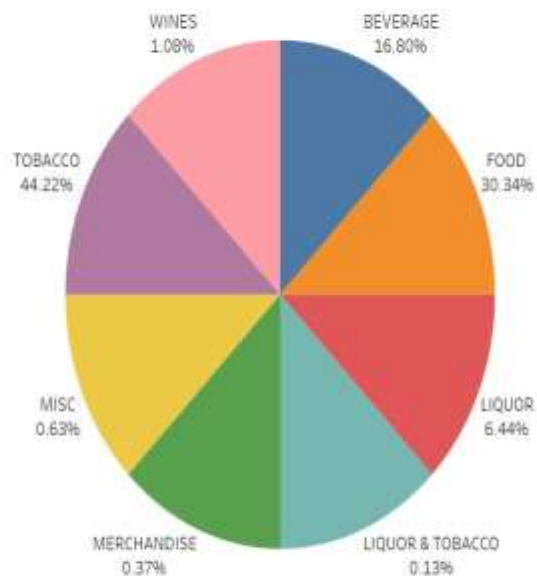- **Top** and **Freq** for respective string variables are highlighted in **Orange** color

Exploratory Analysis and Inferences
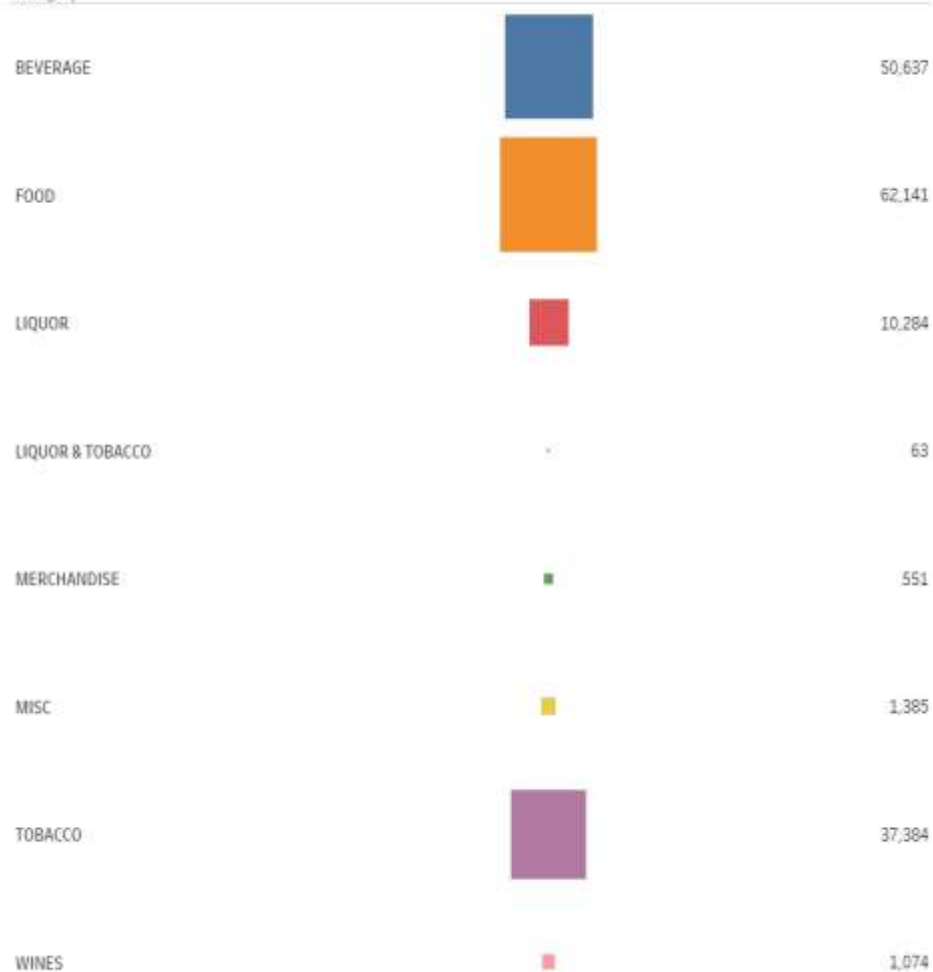
Category vs Sales Amt

Category
- BEVERAGE
- FOOD
- LIQUOR
- LIQUOR & TOBACCO
- MERCHANDISE
- MISC
- TOBACCO
- WINES

WINES 1.08%
BEVERAGE 16.80%
FOOD 30.34%
LIQUOR 6.44%
LIQUOR & TOBACCO 0.13%
MERCHANDISE 0.37%
MISC 0.63%
TOBACCO 44.22%

Category vs Sales Qty

Category

BEVERAGE	50,637
FOOD	62,141
LIQUOR	10,284
LIQUOR & TOBACCO	63
MERCHANDISE	551
MISC	1,385
TOBACCO	37,384
WINES	1,074

Category 'TOBACCO' has highest Sales amount contribution (44.22%) and category 'LIQUOR & TOBACCO' has lowest Sales amount contribution(0.13%)
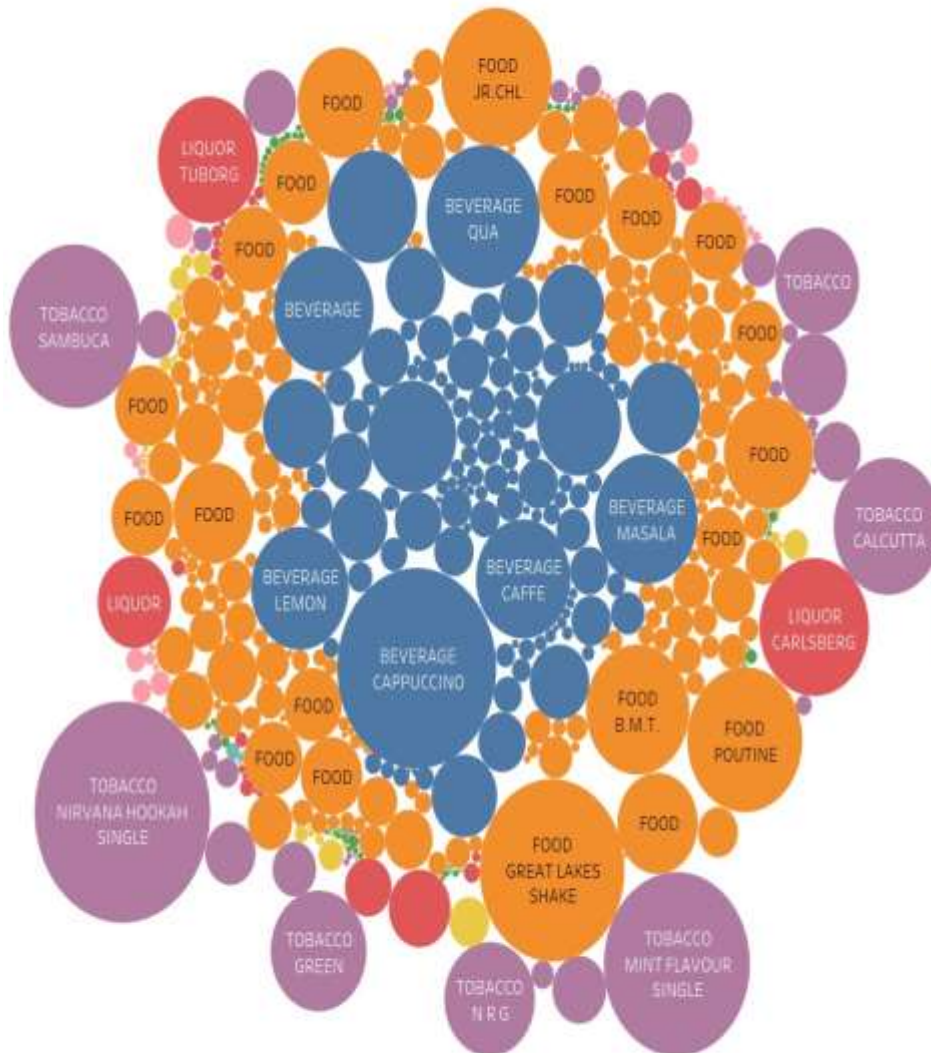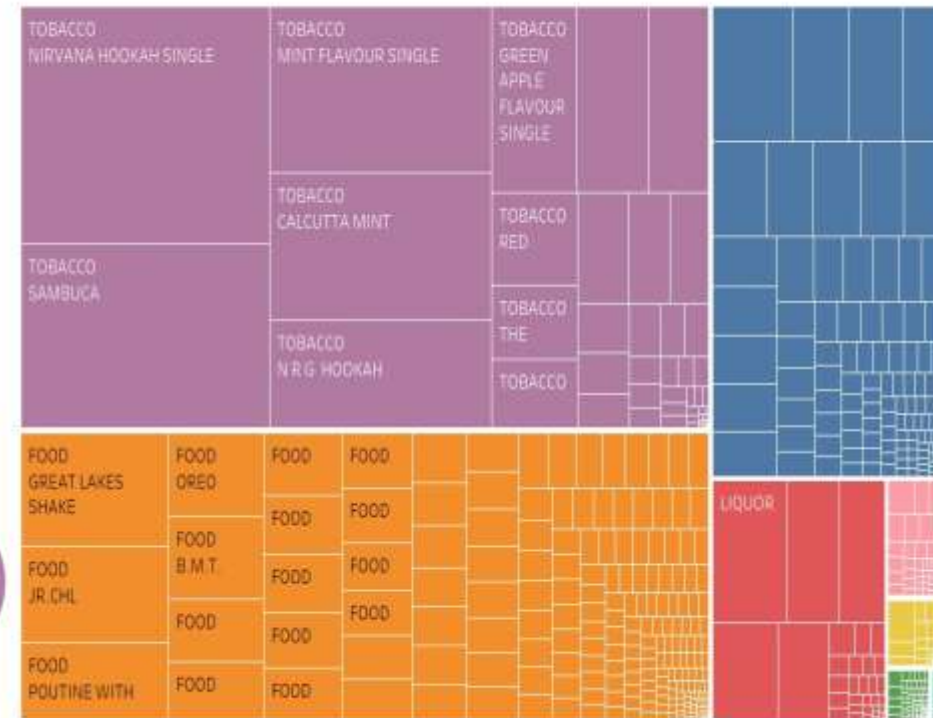
Category 'FOOD' has highest Sales Qty (62141 Nos) and category 'LIQUOR & TOBACCO' has lowest Sales Qty contribution(63 Nos)

Item vs Sales Qty

Item vs Sales Amount

Item 'NIRVANA HOOKAH SINGLE' has highest Sales Qty (8686) and Item '1+1 VLN CAB SAUV' and many items have lowest Sales Qty of value One.

Item 'NIRVANA HOOKAH SINGLE' has highest Sales Amount (2953135) and Item 'MOTHERS DAY SPL' has lowest Sales Amount of value zero.

Most of the Add On items such as ADD ON S,ADD BUTTERED TOAST ,CHEDDAR CHEESE ,MUSHROOM,ADD HAZELNUT FLAVOUR etc., are contributing very less Sales amount. We can think to keep or remove these items based on contribution attributed by MBA analysis
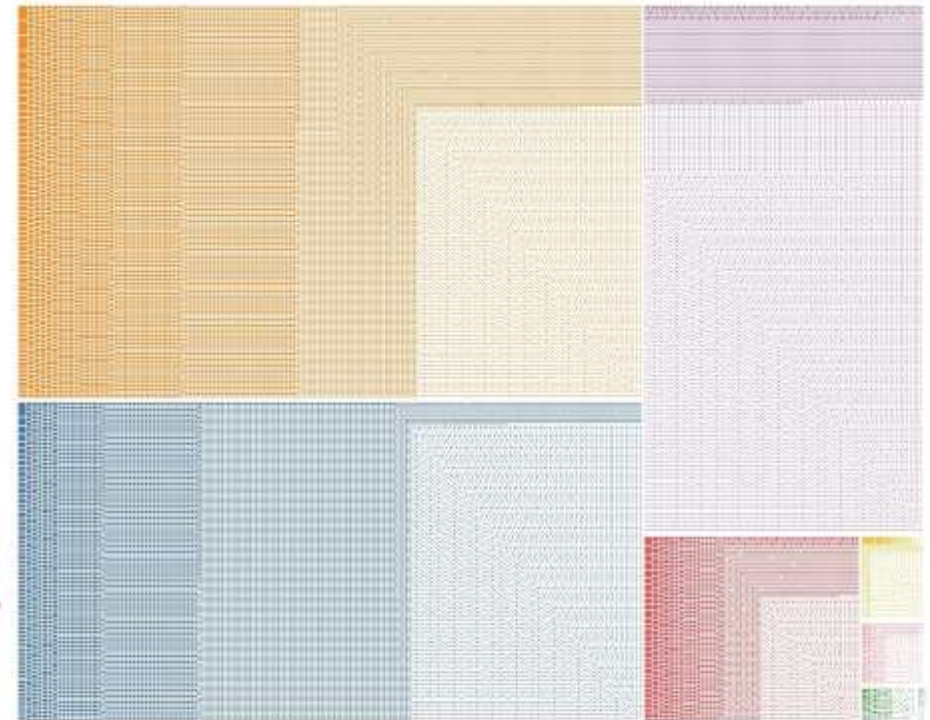
Category
■ BEVERAGE   ■ FOOD   ■ LIQUOR   ■ LIQUOR & T..   ■ MERCHAND..   ■ MISC   ■ TOBACCO   ■ WINES

Bill Number vs Sales Amount
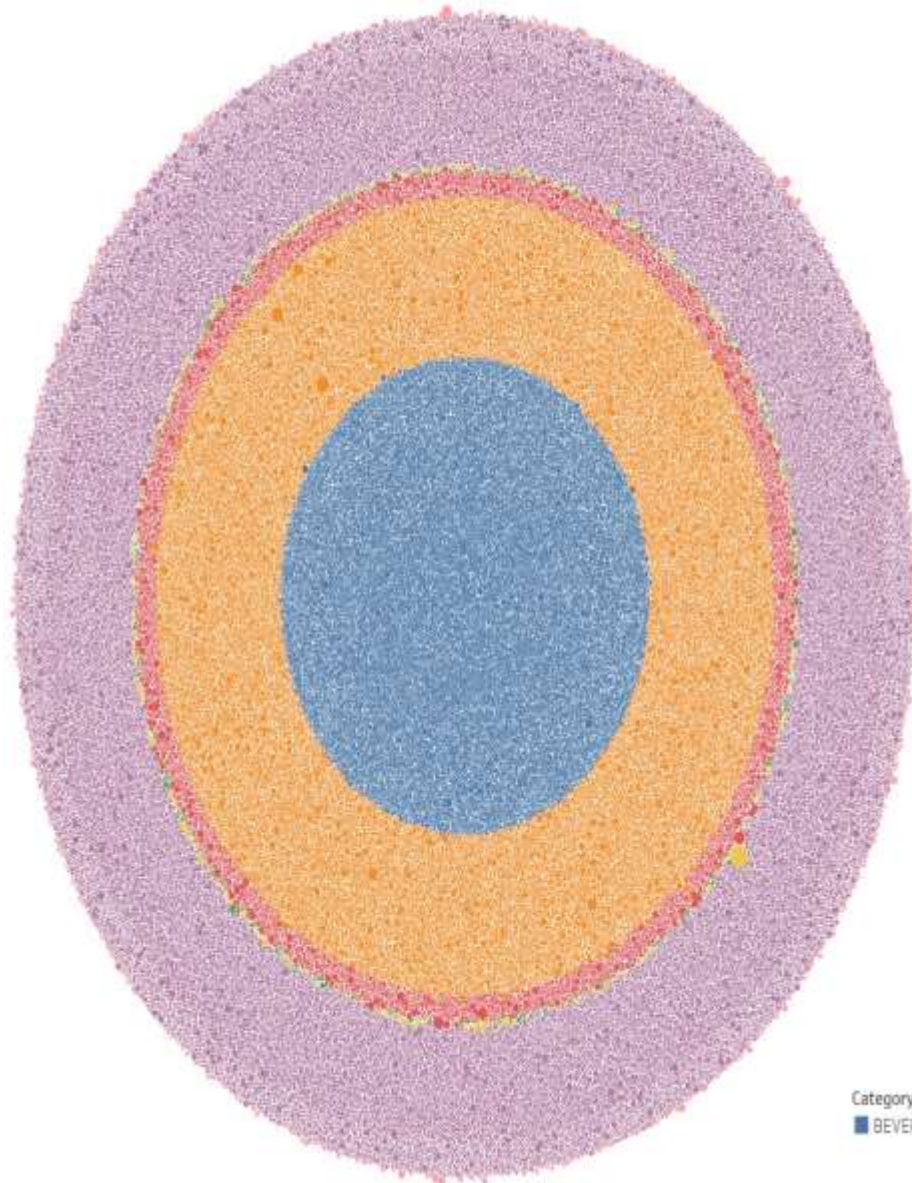
Bill Number vs Qty

Bill Number 'G0481424' has highest Sales Qty (64) and Bill Number 'G0470109' and many items have lowest Sales Qty of value One.

Bill Number 'G0522713' has highest Sales Amount (14231) and Bill Number 'N0030448' has lowest Sales Amount of value One
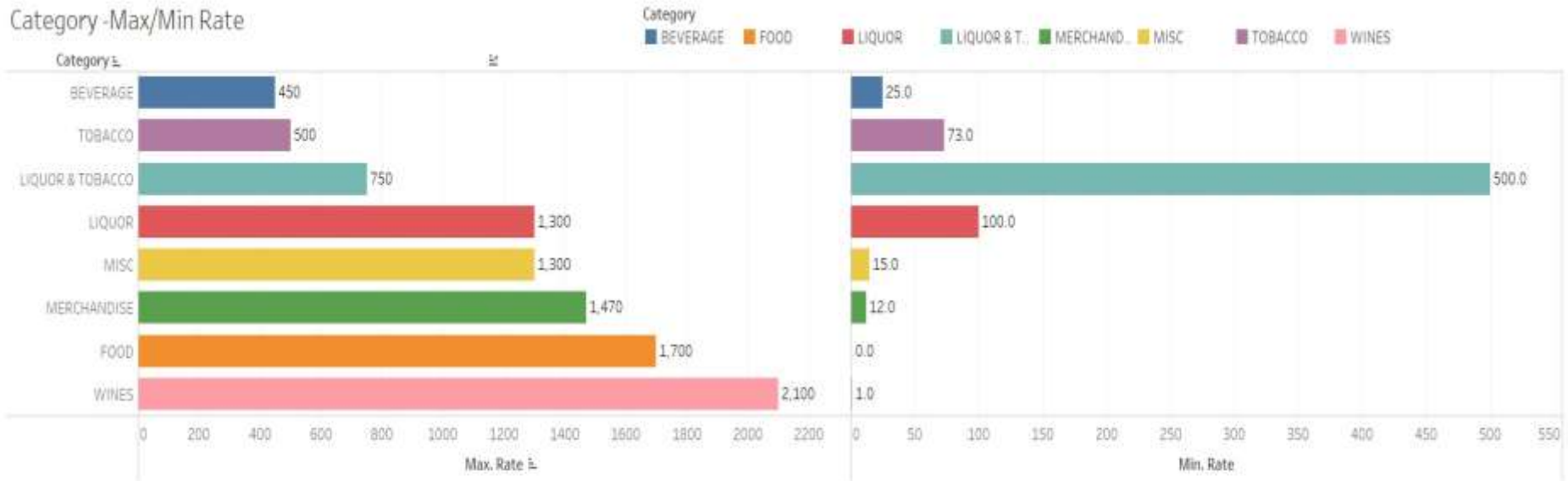
Category
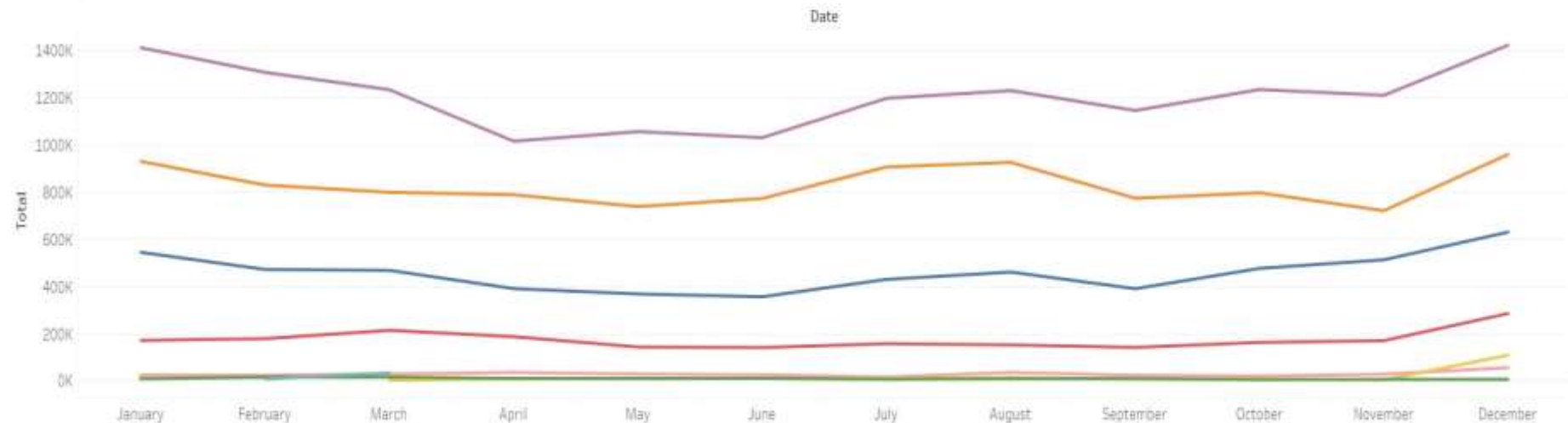■ BEVERAGE ■ FOOD ■ LIQUOR ■ LIQUOR & T... ■ MERCHAND.. ■ MISC ■ TOBACCO ■ WINES

Category -Max/Min Rate

'WINES' has **highest** Rate of value 2100 and 'FOOD' has **lowest** Rate of value zero
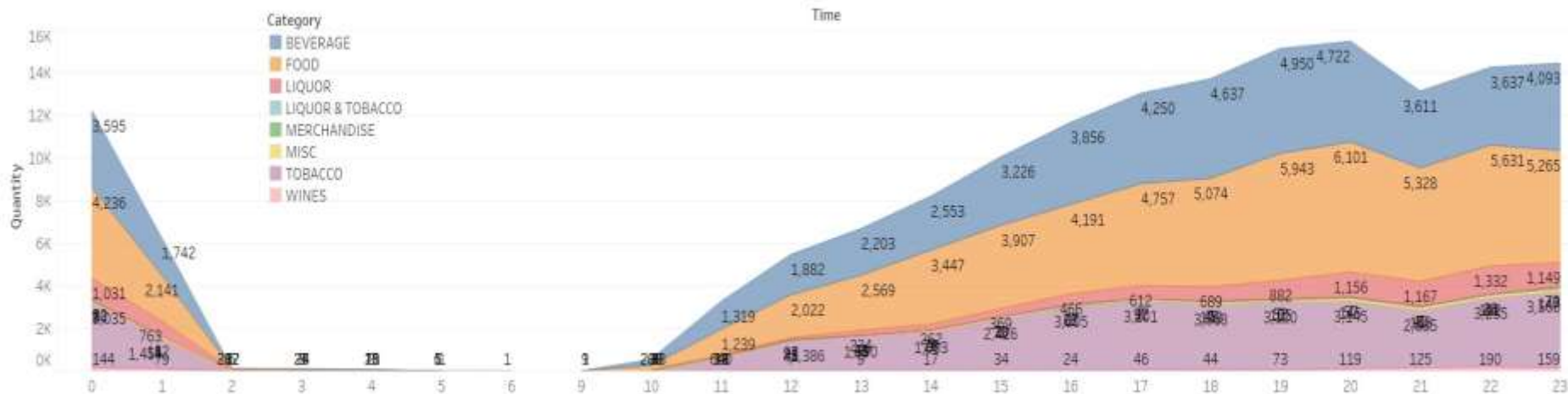
Monthly Sales Amount

Period of Months 'November-February' have **highest** Overall Sales amount and Period of Months 'April-June' have lowest Overall Sales amount.

'TOBACCO' has **highest** Sales Amount along all months and 'MERCHANDISE' has **lowest** Sales Amount along most of the months

Category vs Time

Hours '18-21' have highest Overall Sales Qty and Hours '2-10' have lowest Overall Sales Qty
'BEVERAGE' & 'FOOD' have highest Sales Qty along all Hours and 'WINES' have lowest Sales Qty along most of the Hours

Category vs Day
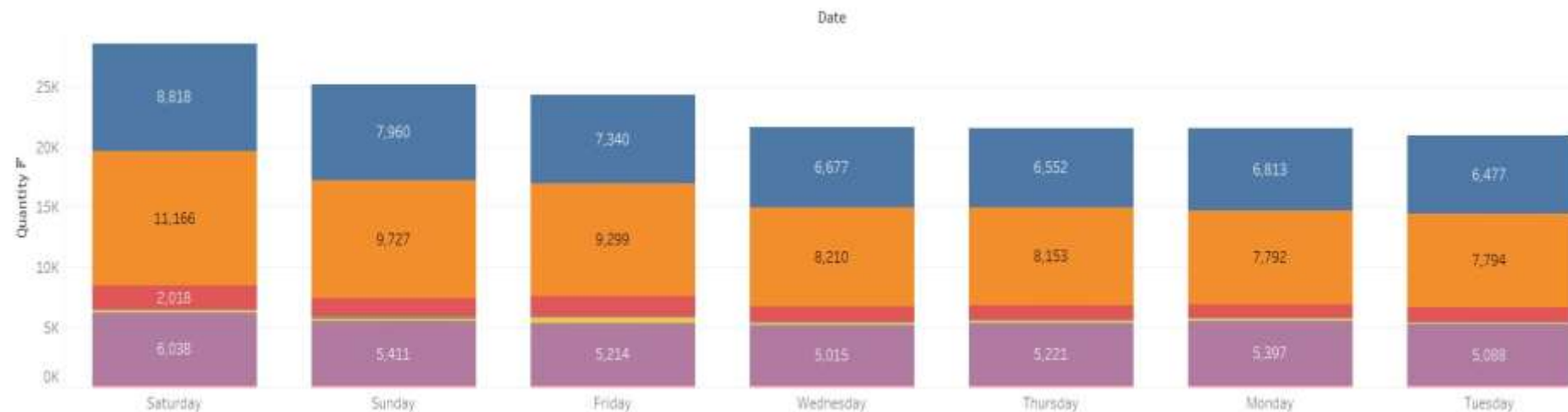
Days 'Saturday-Sunday' have highest Overall Sales Qty and Days 'Monday-Tuesday' have lowest Overall Sales Qty
'BEVERAGE' & 'FOOD' have highest Sales Qty along all days and 'MERCHANDISE' & 'MISC' have lowest Sales Qty along most of the days

Quarterly Sales amount

	Total
	18,397 3,955,915

	Date			
Category	Q1	Q2	Q3	Q4
BEVERAGE	1,487,198	1,117,464	1,283,578	1,621,999
FOOD	2,561,433	2,302,713	2,608,645	2,479,937
LIQUOR	565,935	473,614	453,568	620,330
LIQUOR & TOBACCO	42,562			
MERCHANDISE	41,505	32,935	26,774	18,859
MISC	18,397	31,467	33,787	123,207
TOBACCO	3,955,915	3,105,128	3,575,952	3,870,081
WINES	78,813	91,933	76,548	105,617

'TOBACCO' have **highest** Sales amount along all Quarters and **'MERCHANDISE' & 'MISC'** have **lowest** Sales Qty along most of the Quarters

Quantity vs Rate
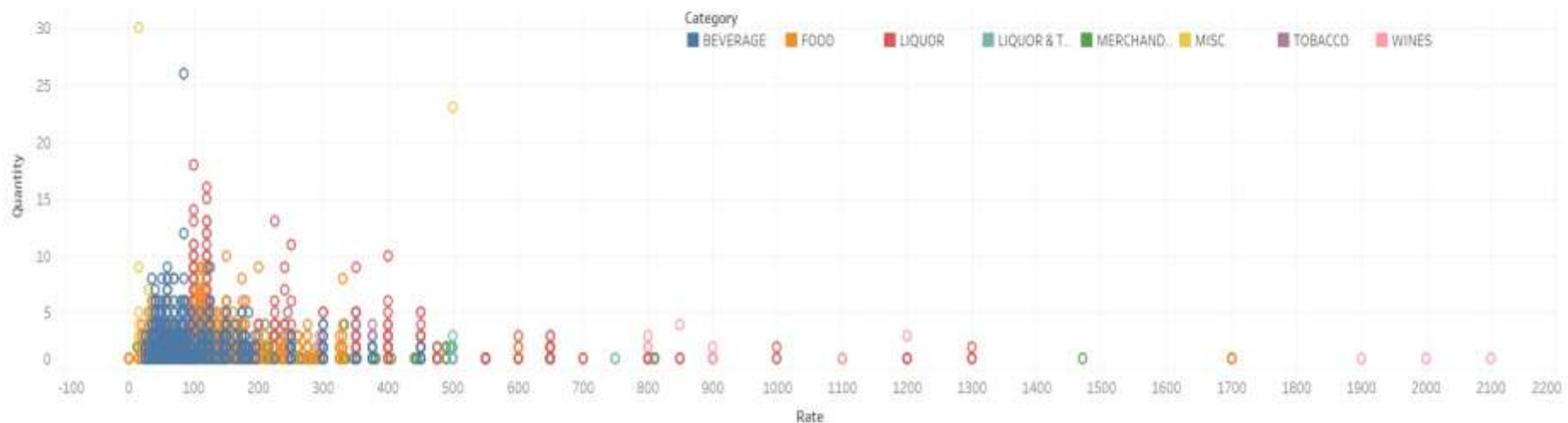
Majority of the Items are in range of **0-10** for **Quantity** and **0-100** for **Rate**

Distplot and Boxplot distribution
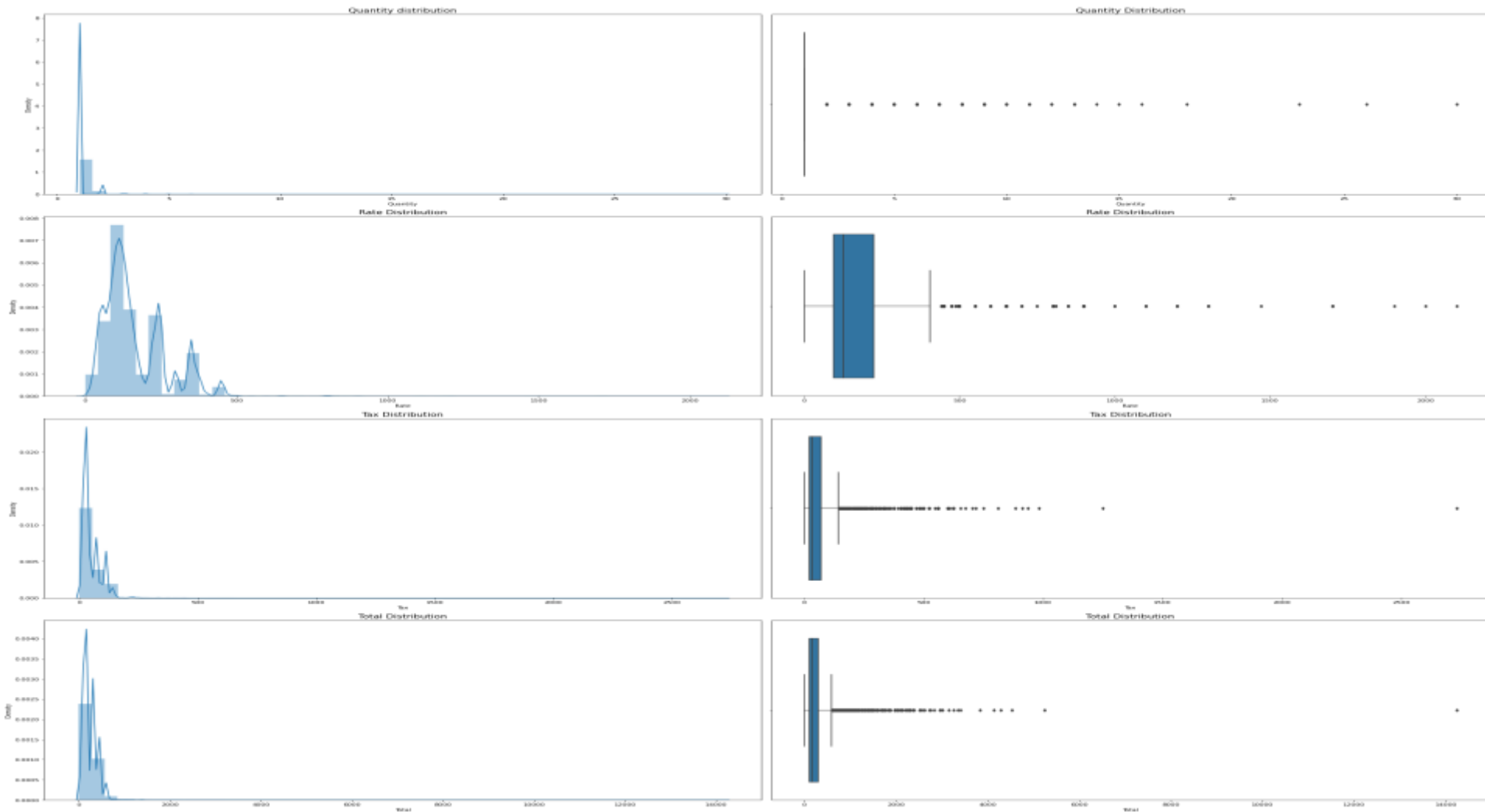
- There is presence of **outliers** in variables such as **Quantity, Rate and Total**.
- Variable **'Discount'** has highest **positive** skewness(106.09) and Variable '**Rate'** has lowest **positive** skewness(1.566)

Heat Map

- Variables **'Total'** and **'Tax'** have **highest positive** Correlation(0.988) and Variables 'Quantity' and 'Rate' have highest **negative** Correlation(-0.11)

Pair Plot
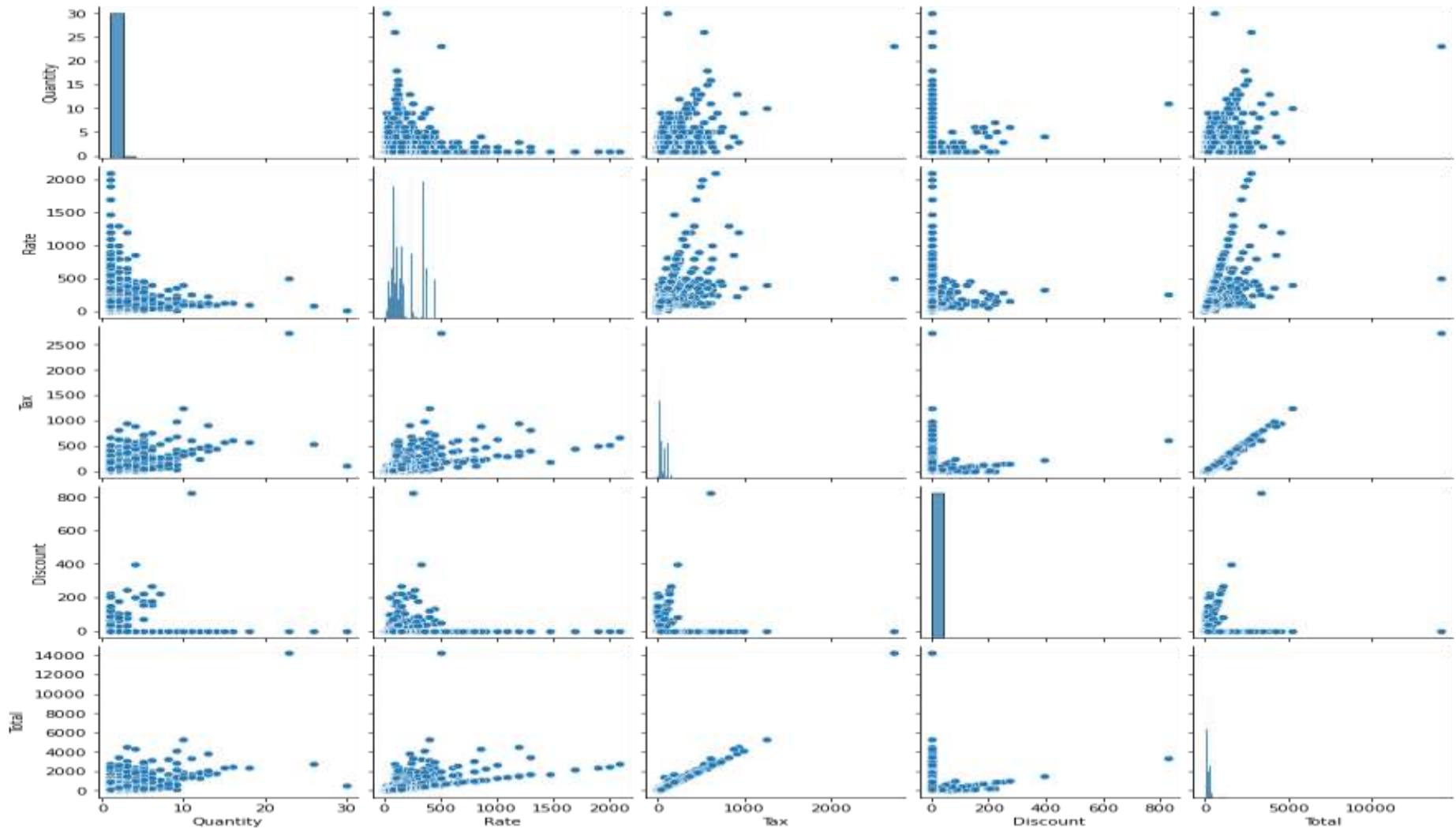
- From the plot , we can see that none of the variables are symmetric
- Variables **'Total'** and **'Tax'** have almost linear realtionship between them.

Menu analysis

Association rules helps us finding interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occurs in a transaction.

A typical example is Market Basket Analysis.

- Market Basket Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.

- Given a set of transactions, we can find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction.

- This association rule has 2 parts : an antecedent (if) and a consequent (then). An antecedent is something that's found in data, and a consequent is an item that is found in combination with the antecedent

Association rules are created by thoroughly analyzing data and looking for frequent if/then patterns. Then, depending on the following two parameters, the important relationships are observed:
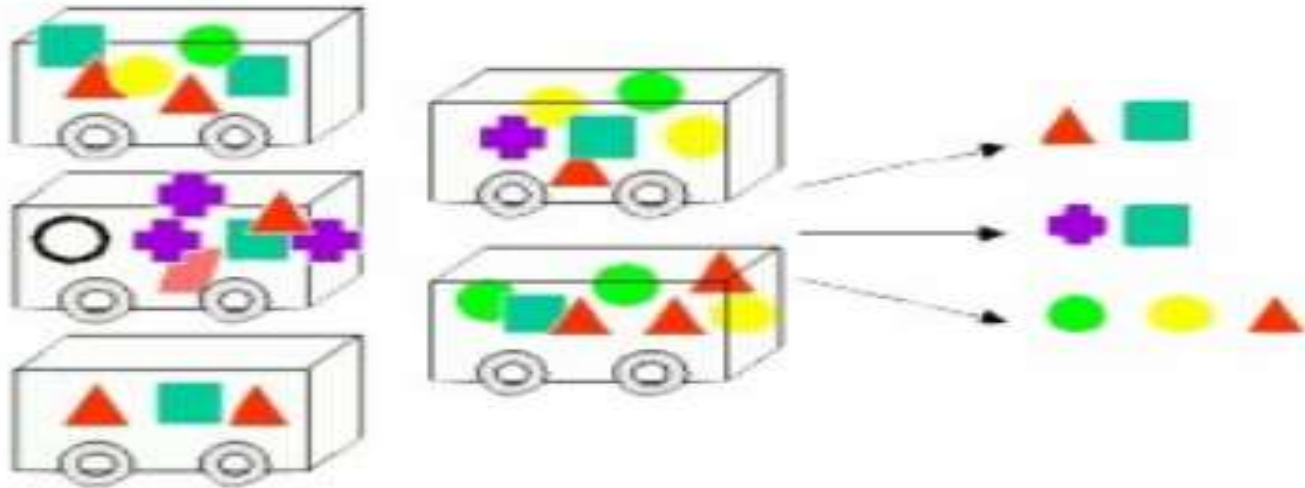
Support: Support indicates how frequently the if/then relationship appears in the database.
Confidence: Confidence tells about the number of times these relationships have been found to be true.
So, in a given transaction with multiple items, Association Rule Mining primarily tries to find the rules that govern how or why such products/items are often bought together.

KNIME WorkFlow Image and Threshold
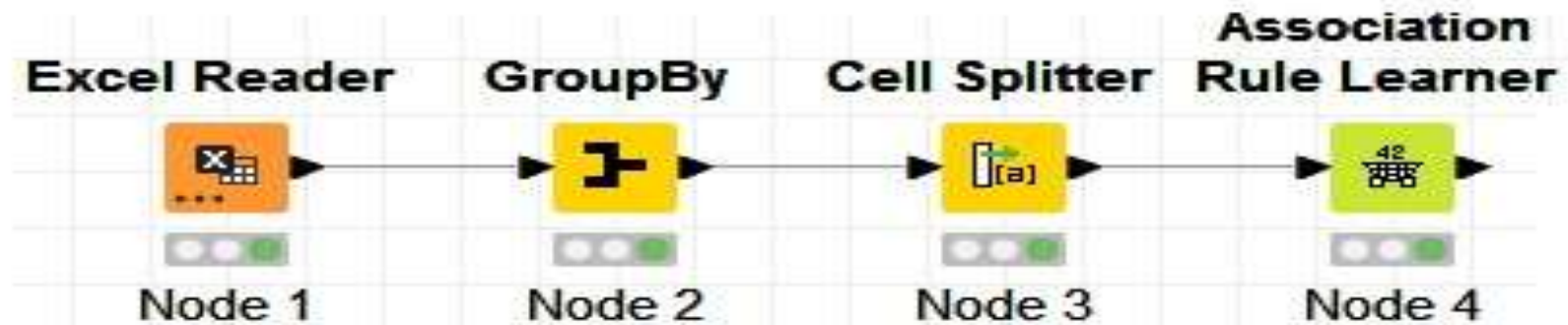
KNIME workflow Image for Market Basket Analysis

We have taken following threshold values for calculation of frequent itemsets through association rule learner

Minimum Support – **0.0006** (Probability of having particular item out of all items)

Minimum Confidence - **0.1** (Probability of having certain item when there are particular items already in the basket)

Association Table

Support	Confidence	Lift	Consequent	Consequent Item Category	implies	Items	Items Category
0.00069	0.462	101.89	KHEEMA GHOTALA	FOOD	<---	[BUTTERED TOASTS]	FOOD
0.00069	0.151	101.89	BUTTERED TOASTS	FOOD	<---	[KHEEMA GHOTALA]	FOOD
0.00104	0.178	79.36	ADD HERB ROAST CHICKEN	MISC	<---	[LEMON INFUSED CHAR GRILLED VEG]	FOOD
0.00104	0.465	79.36	LEMON INFUSED CHAR GRILLED VEG	FOOD	<---	[ADD HERB ROAST CHICKEN]	MISC
0.00073	0.325	45.93	ORANGE ARRABIATA	FOOD	<---	[ADD HERB ROAST CHICKEN]	MISC
0.00073	0.103	45.93	ADD HERB ROAST CHICKEN	FOOD	<---	[ORANGE ARRABIATA]	FOOD
0.00063	0.638	21.30	CAFFE LATTE	BEVERAGE	<---	[ADD CARAMEL FLAVOUR]	BEVERAGE
0.00084	0.223	20.80	ADD FRIES	FOOD	<---	[TRADITIONAL ITALIAN CRUSTINI]	FOOD
0.00161	0.546	18.24	CAFFE LATTE	BEVERAGE	<---	[ADD HAZELNUT FLAVOUR]	BEVERAGE
0.00090	0.192	15.26	MAGGI NDL ARRABIATA	FOOD	<---	[SAMBUCA, B.M.T. PANINI]	TOBBACCO , FOOD
0.00066	0.130	12.09	ADD FRIES	FOOD	<---	[CHICKEN SLOUVLAKI WRAP]	FOOD
0.00064	0.119	11.11	ADD FRIES	FOOD	<---	[MEDITER RANEAN PANINO]	FOOD
0.00074	0.195	11.04	GARDEN FRESH PANINI	FOOD	<---	[FRENCH FRIES]	FOOD
0.00090	0.360	9.67	B.M.T. PANINI	FOOD	<---	[MAGGI NDL ARRABIATA, SAMBUCA]	FOOD , TOBACCO
0.00067	0.176	9.32	COUNTRY ROAST CHICKEN PANINI	FOOD	<---	[FRENCH FRIES]	FOOD
0.00066	0.535	8.46	SAMBUCA	TOBACCO	<---	[QUA MINERAL WATER(1000ML), RED BULL 2+1]	BEVERAGE , BEVERAGE
0.00066	0.140	8.41	RED BULL 2+1	BEVERAGE	<---	[QUA MINERAL WATER(1000ML), SAMBUCA]	BEVERAGE, TOBACCO
0.00113	0.105	7.48	SATAY CHICKEN PANINI	FOOD	<---	[ADD FRIES]	FOOD
0.00090	0.470	7.44	SAMBUCA	TOBACCO	<---	[MAGGI NDL ARRABIATA, B.M.T. PANINI]	FOOD,FOOD
0.00101	0.266	7.14	B.M.T. PANINI	FOOD	<---	[FRENCH FRIES]	FOOD
0.00124	0.116	6.12	COUNTRY ROAST CHICKEN PANINI	FOOD	<---	[ADD FRIES]	FOOD
0.00137	0.464	5.94	CAPPUCCINO	BEVERAGE	<---	[ADD HAZELNUT FLAVOUR]	BEVERAGE
0.00124	0.412	5.92	GREAT LAKES SHAKE	FOOD	<---	[VANILLA ICECREAM]	FOOD

•From above table, for row 1, **Support(0.00069)** indicates how frequently the given item appears with respect to all items.

•**Confidence(0.462)** indicates the chance of having the consequent item (KHEEMA GHOTALA) when BUTTERED TOASTS is ordered.

•**Lift** is the ratio of confidence to support

Combos and offers

	Combos		
Items	**Consequent**	**Items**	**Consequent**
[ADD FRIES]	SATAY CHICKEN PANINI	[LINDT CHOCOLATE SHAKE]	OREO COOKIE SHAKE
	COUNTRY ROAST CHICKEN PANINI		GREAT LAKES SHAKE
	B.M.T. PANINI	[MAGGI NDL ARRABIATA]	B.M.T. PANINI
	PHILLYCREAM CHEESE &CHILLY PAN		SAMBUCA
[ADD HAZELNUT FLAVOUR]	CAFFE LATTE	[MAGGI NDLCREAM/ CHEE/GARLIC]	GREAT LAKES SHAKE
	CAPPUCCINO		POUTINE WITH FRIES
[ADD HERB ROAST CHICKEN]	LEMON INFUSED CHAR GRILLED VEG		GREAT LAKES SHAKE
	ORANGE ARRABIATA		NIRVANA HOOKAH SINGLE
[ADD ON S]	B.M.T. PANINI	[MEDITER RANEAN PANINO]	ADD FRIES
	POUTINE WITH FRIES		GREAT LAKES SHAKE
[B.M.T. PANINI]	POUTINE WITH FRIES	[MEZE PLATTER]	QUA MINERAL WATER(1000ML)
	SAMBUCA		NIRVANA HOOKAH SINGLE
	GREAT LAKES SHAKE	[ORANGE ARRABIATA]	ADD HERB ROAST CHICKEN
	NIRVANA HOOKAH SINGLE		SAMBUCA
[CHAI LATTE]	QUA MINERAL WATER(1000ML)		GREAT LAKES SHAKE
	MASALA CHAI CUTTING		NIRVANA HOOKAH SINGLE
[COTTAGE CHEESE PANINI]	POUTINE WITH FRIES	[POUTINE WITH WEDGES]	QUA MINERAL WATER(1000ML)
	GREAT LAKES SHAKE		GREAT LAKES SHAKE
	NIRVANA HOOKAH SINGLE		NIRVANA HOOKAH SINGLE
[DOPPIO]	CAFFE LATTE	[QUA MINERAL WATER(1000ML), SAMBUCA]	RED BULL 2+1
	QUA MINERAL WATER(1000ML)		B.M.T. PANINI
	CAPPUCCINO		GREAT LAKES SHAKE
[ESPRESSO]	QUA MINERAL WATER(1000ML)	[QUA MINERAL WATER(500ML)]	SAMBUCA
	CAPPUCCINO		NIRVANA HOOKAH SINGLE
[FRENCH FRIES]	GARDEN FRESH PANINI	[RED BULL 2+1]	SAMBUCA
	COUNTRY ROAST CHICKEN PANINI		CALCUTTA MINT
	B.M.T. PANINI		NIRVANA HOOKAH SINGLE
[LEMON INFUSED CHAR GRILLED VEG]	ADD HERB ROAST CHICKEN	[RED BULL ENERGY DRINK]	SAMBUCA
	GREAT LAKES SHAKE		CALCUTTA MINT
	NIRVANA HOOKAH SINGLE		NIRVANA HOOKAH SINGLE
[SAMBUCA, GREAT LAKES SHAKE]	B.M.T. PANINI	[SAMBUCA, B.M.T. PANINI]	MAGGI NDL ARRABIATA
	QUA MINERAL WATER(1000ML)		QUA MINERAL WATER(1000ML)
[VEG. CLUB WRAP]	POUTINE WITH FRIES		POUTINE WITH FRIES
	GREAT LAKES SHAKE		GREAT LAKES SHAKE

• From above table, we can provide different combo offers on the above listed items and their respective consequents. There by we could increase sales of the restaurant

Recommendations

Items	Consequent
[3 RED BULL]	NIRVANA HOOKAH SINGLE
[ADD CARAMEL FLAVOUR]	CAFFE LATTE
[B.M.T. PANINI, GREAT LAKES SHAKE]	SAMBUCA
[BLACK CURRANT ICED TEA]	LEMON ICED TEA
[BLUEBERRY BRAIN FREEZER SHAKE]	GREAT LAKES SHAKE
[BUN MASKA & CHAI]	MASALA CHAI CUTTING
[BUTTERED TOASTS]	KHEEMA GHOTALA
[CAFFE LATTE]	CAPPUCCINO
[CHICKEN SALAMI PANINI]	GREAT LAKES SHAKE
	NIRVANA HOOKAH SINGLE
[CHICKEN SLOUVLAKI WRAP]	ADD FRIES
[CHUNKY CHOCO CHIP COOKIES]	CAPPUCCINO
[COOL CALIFORNICA]	BERRY BLAST
[COUNTRY ROAST CHICKEN PANINI]	POUTINE WITH FRIES
[CURRANT COOLER]	BERRY BLAST
[GARDEN FRESH PANINI]	GREAT LAKES SHAKE
[GREAT LAKES FLOATS W CHOC]	CAPPUCCINO
[IRISH COFFEE]	CAPPUCCINO
[KHEEMA GHOTALA]	BUTTERED TOASTS
[KIT KAT SHAKE]	NIRVANA HOOKAH SINGLE
[MAGGI NDL ARRABIATA, B.M.T. PANINI]	SAMBUCA
[MAGGI NDL ARRABIATA, SAMBUCA]	B.M.T. PANINI
[MASALA CHAI CUTTING]	CAPPUCCINO
[MOROCCAN MINT TEA]	NIRVANA HOOKAH SINGLE
[PHILLYCREAM CHEESE &CHILLY PAN]	POUTINE WITH FRIES
[POUTINE WITH FRIES, B.M.T. PANINI]	SAMBUCA
[POUTINE WITH FRIES, SAMBUCA]	B.M.T. PANINI
[POUTINE WITH FRIES]	NIRVANA HOOKAH SINGLE
[QUA MINERAL WATER(1000ML), B.M.T. PANINI]	SAMBUCA
[QUA MINERAL WATER(1000ML), GREAT LAKES SHAKE]	SAMBUCA
[QUA MINERAL WATER(1000ML), RED BULL 2+1]	SAMBUCA
[QUA MINERAL WATER(1000ML)]	NIRVANA HOOKAH SINGLE
[RED BULL 2+1, SAMBUCA]	QUA MINERAL WATER(1000ML)
[SCRAMBLED EGGS]	QUA MINERAL WATER(1000ML)
[SPANISH OMELETTE BREAKFAST]	QUA MINERAL WATER(1000ML)
[STRAWBERRY ICED TEA]	LEMON ICED TEA
[THE CHOCO LATTE]	CAPPUCCINO
[THE FERROR ROCHER SHAKE]	KIT KAT SHAKE
[TRADITIONAL ITALIAN CRUSTINI]	ADD FRIES
[TROPICAL SMOOTHIE]	BERRY BLAST
[ULTIMATE HOT CHOCOLATE]	CAPPUCCINO
[VANILLA ICECREAM]	GREAT LAKES SHAKE

• From above table, we can make buy one get one or 2+1 offers by joining the respective Items and Consequent. This will help in attracting more customers towards the restaurant.

• The Restaurant can prepare catalog to place the optimally concerning category items together on the menu.There would be higher chances of order increase and resulting in the increase of revenue of the restaurant.

• Providing more offers during weekdays and festival months would be good option to make best use of peak rush.