MRA PROJECT

Submitted
by
Lakshana Bhat

Exploratory Analysis (45 Marks)

Exploratory Analysis of data & an executive summary (in PPT) of your top findings, supported by graphs.

Top five rows of the dataset:

	Date	Bill Number	Item Desc	Time	Quantity	Rate	Tax	Discount	Total	Category
0	2010-04-01	G0470115	QUA MINERAL WATER(1000ML)	13:15:11	1	50.0	11.88	0.0	61.88	BEVERAGE
1	2010-04-01	G0470115	MONSOON MALABAR (AULAIT)	13:15:11	1	100.0	23.75	0.0	123.75	BEVERAGE
2	2010-04-01	G0470116	MASALA CHAI CUTTING	13:17:35	1	40.0	9.50	0.0	49.50	BEVERAGE
3	2010-04-01	G0470117	QUA MINERAL WATER(1000ML)	13:19:55	1	50.0	11.88	0.0	61.88	BEVERAGE
4	2010-04-01	G0470283	MOROCCAN MINT TEA	01:20:18	1	45.0	10.69	0.0	55.69	BEVERAGE

Setting the Data variable as the index for our DataFrame.

The Variables in our dataset:

```
<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  datetime64[ns]
 1   Bill Number   145830 non-null  object
 2   Item Desc     145830 non-null  object
 3   Time          145830 non-null  datetime64[ns]
 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: datetime64[ns](2), float64(4), int64(1), object(3)
memory usage: 11.1+ MB
```

Cleaning the object variables:

Cleaning the inconsistent spellings for 'LIQUOR' the 'Category' Variable:

```
FOOD                56658
BEVERAGE            43513
TOBACCO             36294
LIQUOR               6199
MISC                 1160
WINES                 805
MERCHANDISE           473
LIQUOR & TOBACCO       47
LIQUOR                  1
```

```
FOOD                56658
BEVERAGE            43513
TOBACCO             36294
LIQUOR               6200
MISC                 1160
WINES                 805
MERCHANDISE           473
LIQUOR & TOBACCO       47
```

Four point Summary of the data:

	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	2010-12-31 00:00:00	G0490530	NIRVANA HOOKAH SINGLE	22:25:36	NaN	NaN	NaN	NaN	NaN	FOOD
freq	834	23	8553	33	NaN	NaN	NaN	NaN	NaN	57023
first	2010-04-01 00:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
last	2011-03-31 00:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	NaN	NaN	NaN	1.121299	161.782259	48.929061	0.095079	224.959852	NaN
std	NaN	NaN	NaN	NaN	0.477237	102.244631	40.272851	3.720735	164.960776	NaN
min	NaN	NaN	NaN	NaN	1.000000	0.010000	0.000000	0.000000	0.010000	NaN
25%	NaN	NaN	NaN	NaN	1.000000	95.000000	22.560000	0.000000	117.560000	NaN
50%	NaN	NaN	NaN	NaN	1.000000	125.000000	32.060000	0.000000	167.060000	NaN
75%	NaN	NaN	NaN	NaN	1.000000	225.000000	72.000000	0.000000	315.000000	NaN
max	NaN	NaN	NaN	NaN	30.000000	2100.000000	2731.250000	825.000000	14231.250000	NaN

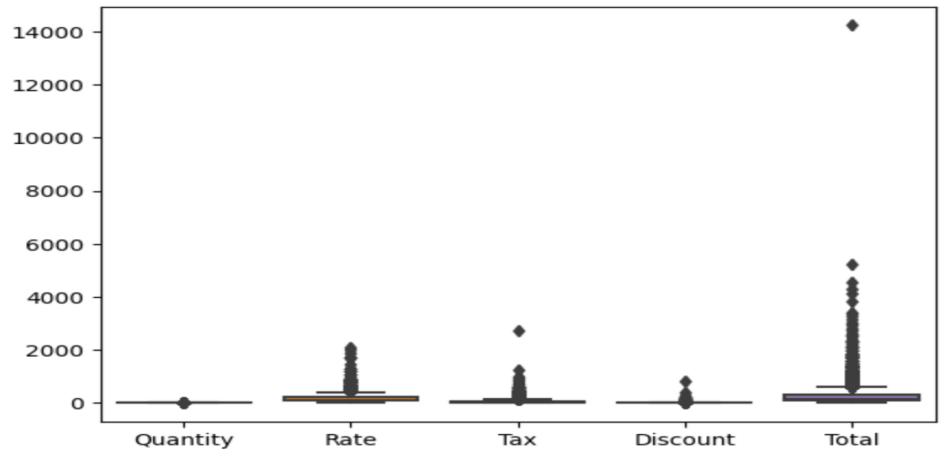
Checking the Number of Duplicates in the data:

There are 680 Duplicate rows in the data. Since the duplicates ratio is very low, when compared with the total number of rows, we drop the duplicates.

Checking the number of Nulls in the data:

```
Date           0
Bill Number    0
Item Desc      0
Time           0
Quantity       0
Rate           0
Tax            0
Discount       0
Total          0
Category       0
dtype: int64
```

Checking the Number of Outliers in the data using boxplots:



It is evident from the boxplot that our data has many outliers. Since we are trying to understand our data, and not make any prediction. We choose to not treat the outliers.

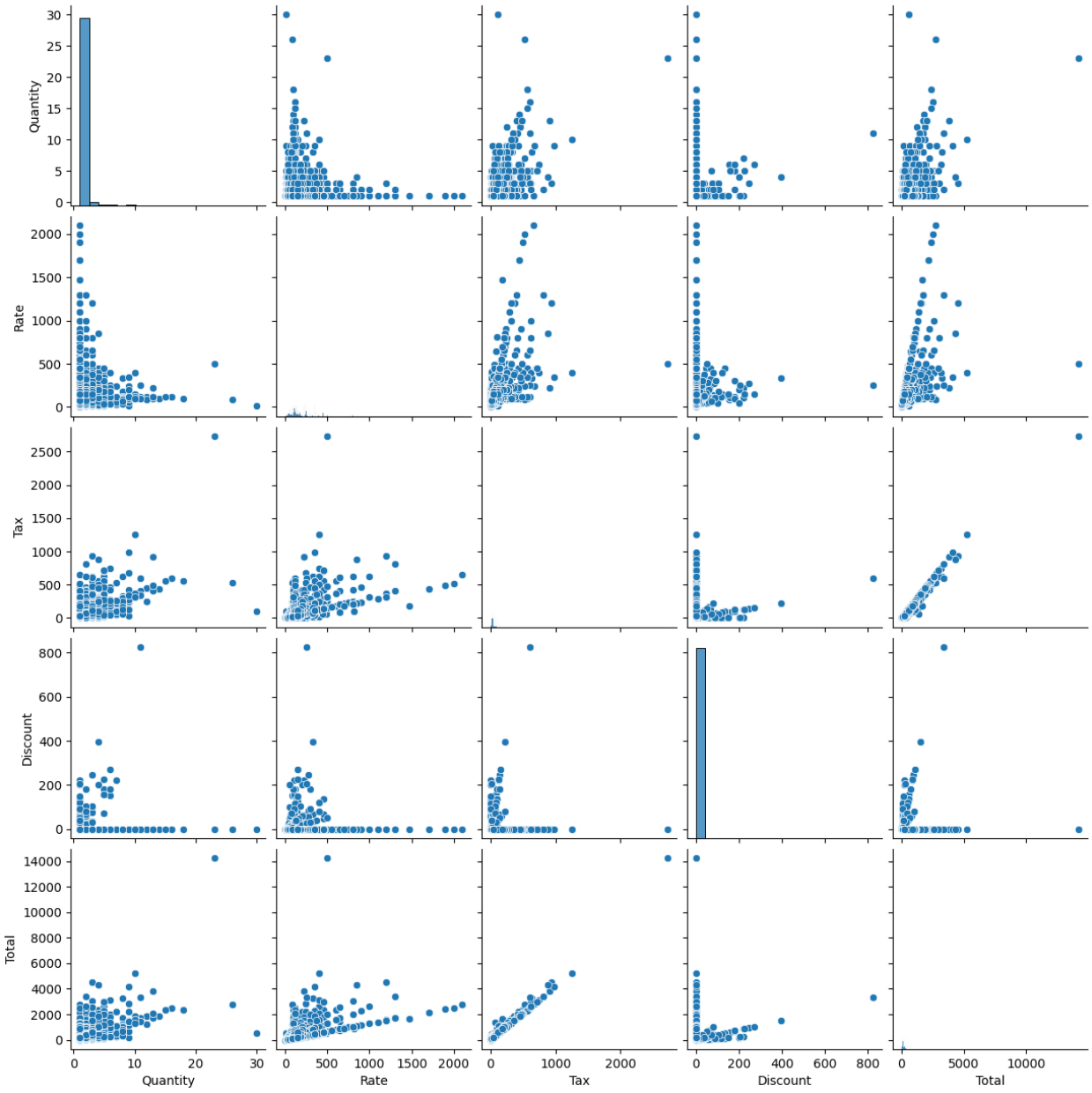
Understanding the Correlation between the Variables:

Inference:

As expected there is strong correlation between all tax, rate and total.

Understanding the Covariance between the variables:

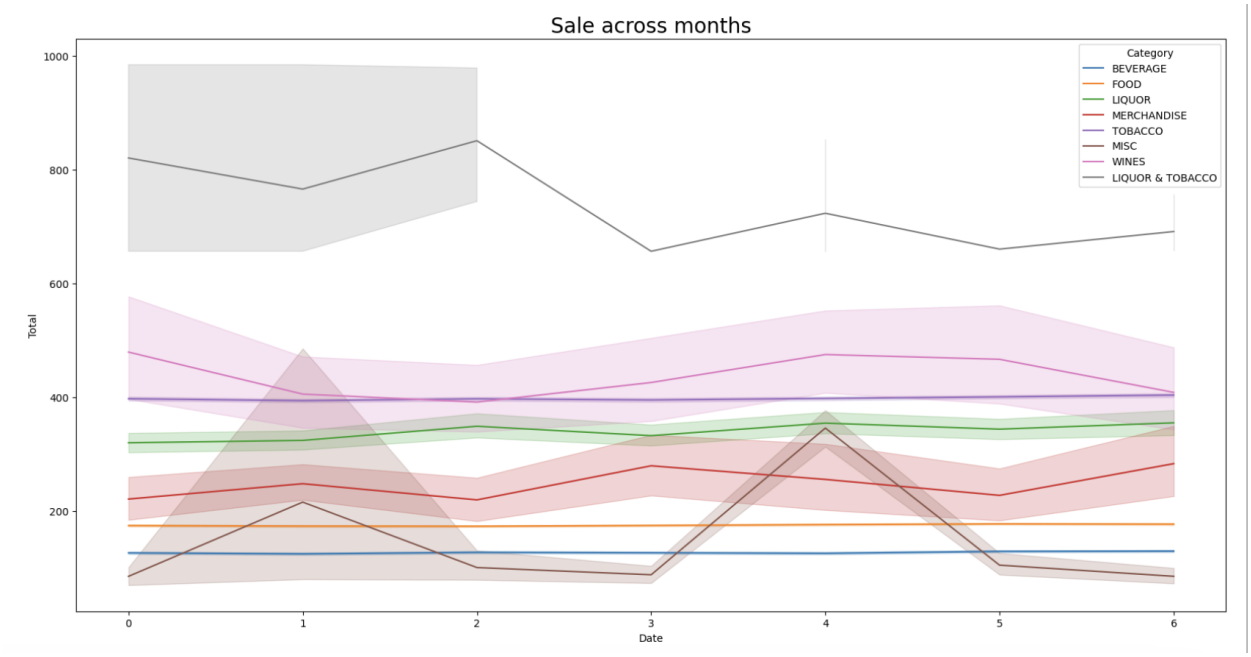
Pairplot for the data:



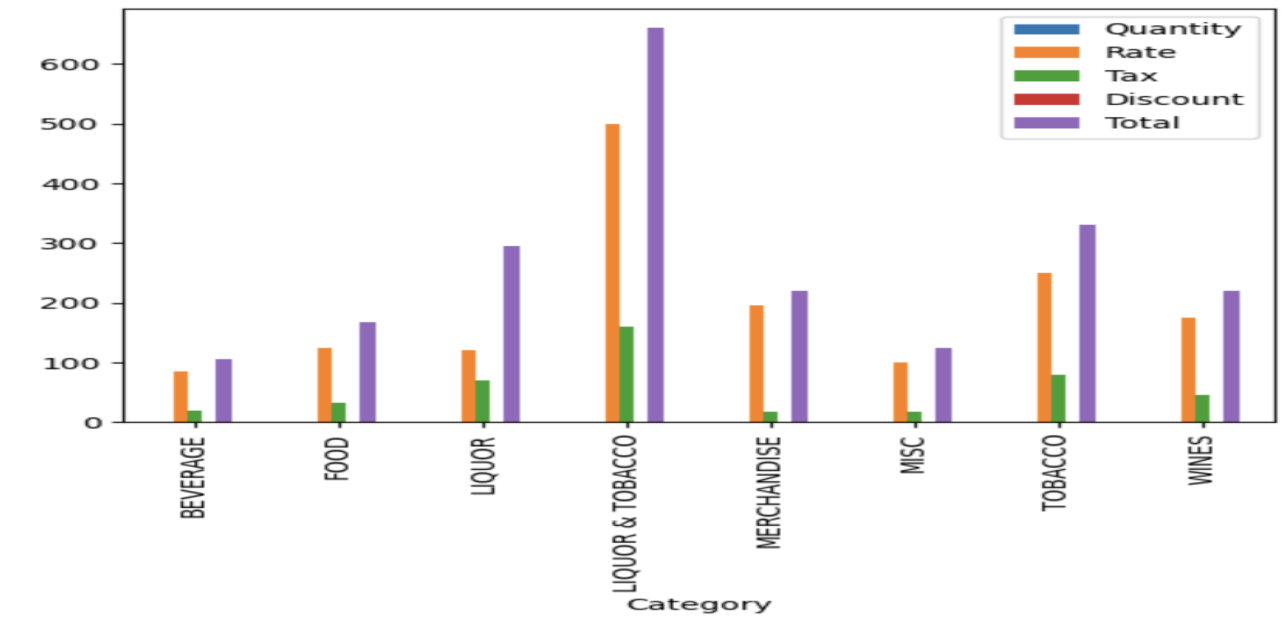
Inference:

Some of the variables have a liner relationship between them, like tax and total, rate and tax etc.

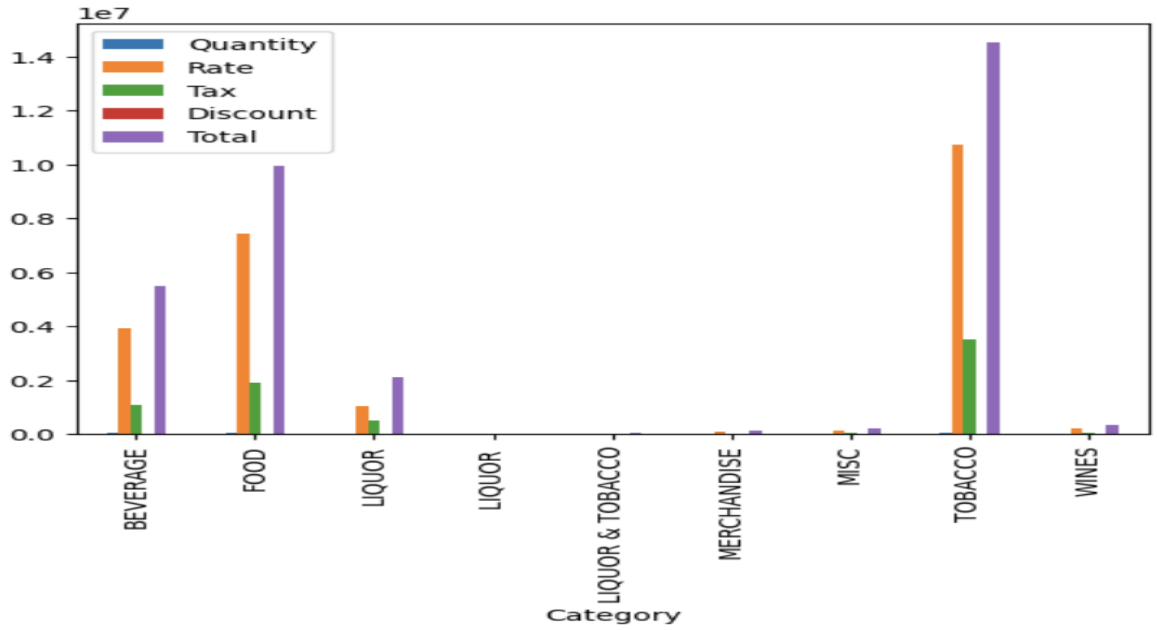
What kind of trends do you notice in terms of consumer behaviour over different times of the day and different days of the week? Can you give concrete recommendations based on the same?



Plotting the Total value for all the numeric variables across the different Categories:



Plotting the Median for all the numeric variables across all categories:



Inference:

1. Liquor and Tobacco sales peak during the beginning of the week and show a gradual decline towards the end.
2. Wine is more popular todawards the end of the week
3. Misc item show peak sales during Mondays and Thursdays
4. Merchandise sales are also hughes during Saturdays
5. Irrespective of the week Liquor and Tobacco contribute the most to the overall Sale

Items that can be taken off the menu

Analysing each menu category and the least frequently ordered items. Displaying the least ordered items for each menu category.

Since the following items within these categories have been least ordered throughout the year, they can be taken off the menu.

Tobacco:

Items that can be taken off the tobacco menu category

	ItemDesc
BENSON & HEDGES SPL	9
SPICE SHEESHA	6
CLASSIC MENTHOL	4
CLASSIC ULTRA MILD	4
BENSON & HEGDES GOLD BLUE	3
GOLD FLAKE LIGHTS-BIG	3
MINT FLAVOUR DOUBLE	3
GREAT LAKES HOOKAH SINGLE	2
ICE SPICE SHEESHA	2
CLASSIC MENTHOL RUSH	2
APPLE FLAVOUR DOUBLE	2
GOLD FLAKE ULTRA LIGHTS(20)	1
CLASSIC REGULAR	1
AL SIKANDARI HOOKAH DOUBLE	1
INDIA KINGS OCEAN BLUE	1

Food:

Items that can be taken off of the Food menu category.

	ItemDesc
MAGGI NDL BURMESE CURRY	9
CAFE ZABAGLIONE	9
ALIO AGLIO	9
POLENTA & CORN CAKES	8
3COURSE NON-VEG MEAL	7
RED VELVET CHEESE CAKE	7
MUSHROOM & CORN	7
TOAST CIABATA	6
3COURSE VEG MEAL	6
SCRAMBLED EGGS + BEVERAGE	6
MOTHERS DAY SPL	5
CALAMARI FRITO	5
BEANS NACHO CHILLI W MEAT	5
CHICKEN HAM	5
NONVEG PASTA PESTO	5
CHEESE FINGERS	4
J.PCHENET SPARKLING ROSE (BTL)	3
2 MUFFINS + BEVERAGE	3
WAFFLES + BEVERAGE	3
VEGETABLE PASTA	3
CARROT CAKE	3
NIRVANA HOOKAH DOUBLE	2
STRAWBERRY MERINGUE	2
SUNNY SIDEUP + BEVERAGE	1
CAPONATA	1
ADD BUTTERED TOAST	1
SCHNEIDER 2+1	1

Beverage:

Items that can be taken off of the Beverage menu category

	ItemDesc
JAVA ESTATES (AULAIT)	15
N R G HOOKAH	15
NEW ORLEANS BLUE (REG)	15
VARLHONA HOT CHOCOLATE	11
BOTTLED WATER (1LITRE)	10
NEW ORLEANS BLUE (AULAIT)	10
HOUSE BLEND DE CAFFE (AU LAIT)	8
WHAT A MELON	4
PEACH BULL	1
MOCAFE HOT CHOCOLATE(SF)	1
MIXED FLAVOUR SINGLE	1
DECAFFINATE COFFEE FRAPPE	1
2 AXE TWIST	1

Liquor:

Items that can be taken off of the Liquor menu category

	ItemDesc
VODKA (SM)	14
TUBORG 2+1	13
STELLA ARTOIS MUG (1 LTR)	13
CARLSBERG 2+1	12
BEER TANK 3.5 LITRE	10
ZINZI WHITE (GLS)	9
ZINZI RED (GLS)	9
1+1 KF 2 LITER	7
WHITE SANGRIA (CARAFE)áááááááá	6
TEQUILA	5
BROOKLYN BUCKET - 4	4
SCHNEIDER WEISSE	4
STELLA 1LTR 2+1	4
BROOKLYN	4
SCHNEIDER BUCKET - 6	3
WHITE RUM (SM)	3
UNLIMITED BEER	3
STELLA ARTOIS	2
WHISKEY (SM)	1
ZINZI WHITE (BTL)	1

Merchandise:

Items that can be taken off of the Merchandise menu categroy

	ItemDesc
CH TSHIRTS	1
ROCK THE BOAT TEA LIGHT HOLDER	1
COUNTRY LEMONADE GLASS(HANSA)	1
SHAKE GLASS	1
AVALANCHE BOWL	1
FLAVOR 1000 GMS	1
CH CRICKET NOTEBOOK	1
KITSCH PINK	1
GREAT LAKES MUG SINGLE	1
MUGS - PLAIN COLOUR	1
SILVER STYLER 1.0	1
BEACH GREEN	1
ZEN ROCK SQUARE VASE	1
ETCHED LEAF TLIGHT HOLDER	1
MODEL-P (IRON ASHTRAY)	1
BENARAS BLUE	1
GOLDEN DELIGHT 1.1	1
CH TIN SMALL	1
ASH TRAYS	1
CUTTING GLASS	1
UDAIPUR LILY	1
GUERILLA COOL	1
KITSCH BLUE	1
DHARMATEA LIGHT HOLDER	1
KONKAN STRIPE	1
DIP BOWL	1
CHAIRMAN COOL	1

Misc:

Items that can be taken off of Misc Menu category

	ItemDesc
QUESO CROQUETAS FRITOS	2
FISH FINGER	2
PHILADELPHIA CREAM CHEESECAKE	2
GIN (SM)	2
POLLO CON AIOLI	2
ORANGE ARRABIATA	2
NUTELLA CREPES	2
HOTDOG WRAP	2
DARK RUM (SM)	2
ROAST CHICKEN SALAD	2
GREEK GYROS CRUSTINI	2
RASPBERRY LITE SHAKE	1
POLLO PELOTA CON TAMATE SALSA	1
RED BULL 2+1	1
RED SANGRIA (CARAFE) ááááááááá	1
ROMA TOMATO & JALAPENO CROQUE	1
PARTY CHARGES @ 500/-	1
HOEGAARDEN LTR MUGS (2+1)	1
GNOCCHI CON POMMODORO	1
GRAPPO SHEESHA	1
MEXICAN CHILLY CREPES	1
MEDITER RANEAN PANINO	1
M & M SHAKE	1
HOEGAARDEN GLS (2+1)	1
JAPANESE YAKITORI WRAP	1
ITALIAN CAPONATA PANINO	1
MANGO FLAVOUR SINGLE	1

Wines:

Items that can be taken off of Wines menu category

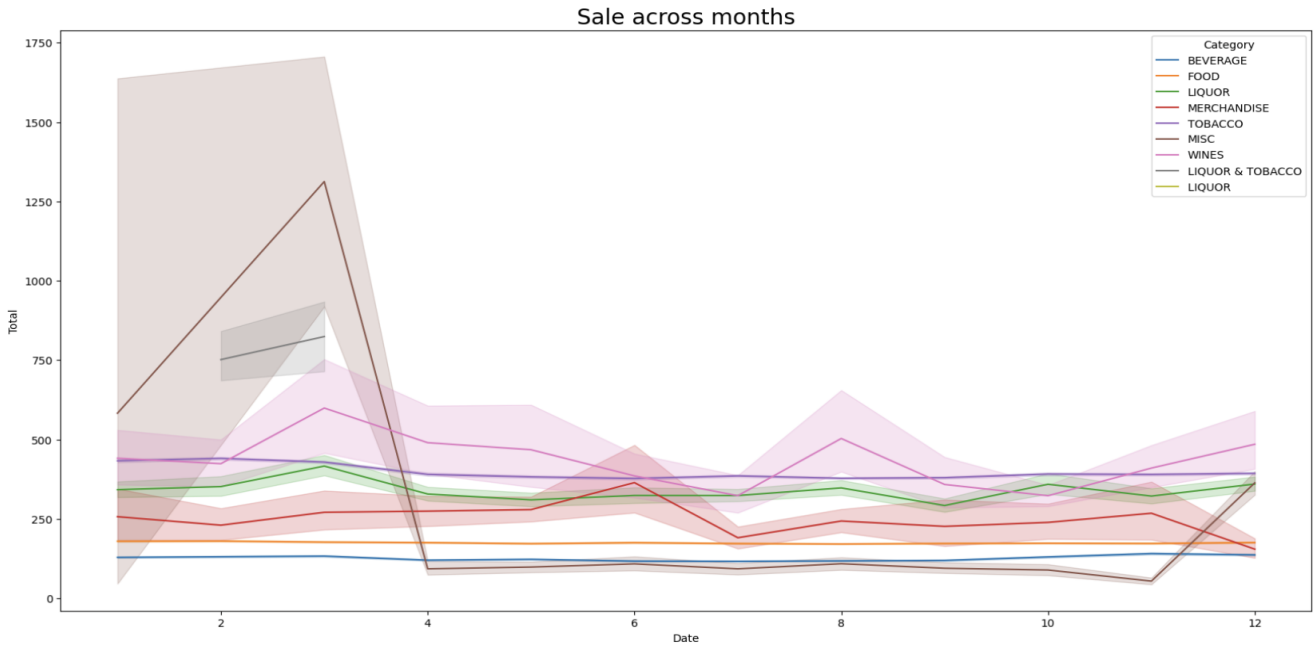
	ItemDesc
VLN CAB SAUV CLASIQ (BTL)	8
VLN SAUV BLANC (BTL)	7
VLN CAB SAUV (BTL)	6
4 SEASONS CLAS SYRAH(GLS)	5
SULA BRUT (BTL)	5
DIA SPARKLING WINE(BTL)	4
B1G1 ZINZI WHITE (BTL)	3
4 SEASONS CLAS SAUV(GLS	3
MAISON PIERRE SAUV MARSAN	3
B1G1 ZINZI RED (BTL)	3
SULA SATORI MERLOT (BTL)	2
1+1 BTL4 SEASON WHITE	2
B1G1 4SEASON CLAS SYRAH(GLS)	2
1+1 VLN CAB SAUV CLASIQ (BTL)	2
4 SEASONS CLAS SAUV(BTL)	2
1+1 WINE BOTTLE	2
B1G1 4SEASON CLAS SAUV(BTL)	1
4 SEASONS CLAS SYRAH(BTL)	1
GOSSIPS CHARD AUS (BTL)	1
MANDALA VALLEY CHENIN BLANC(GL	1
MANDALA VALLEY RED ZINFANDEL(G	1
MATEUS ROSE PORTUGAL(BTL)	1
2 OCEAN PINOTAGE (BTL)	1
1+1 VLN CAB SAUV (BTL)	1
SULA CHENIN BLANC (BTL)	1
1+1 VLN SAUV BLANC (BTL)	1
B1G1 4SEASON CLAS SAUV(GLS)	1

Trends across months

Plotting the Total variable, across the different months for all the categories.

Inference:

1. There is a peak in the sale of 'Misc' Category between Decenber and January to April, the highest being in March.
2. Beverages are the least popular Category throught the year.
3. Customers Enjoy wines the most during March and August
4. There is a noticeable peak in the sale of merchandise during June.
5. There is a noticeable peak in liquor sale during March and October.



Menu Analysis

Identify the most popular combos that can be suggested to the restaurant chain after a thorough analysis of the most commonly occurring sets of menu items in the customer orders. The restaurant doesn't have any combo meals. Can you suggest the best combo meals?

Finding the most commonly ordered items:

The below list contains the **15 most commonly ordered items** on the menu.

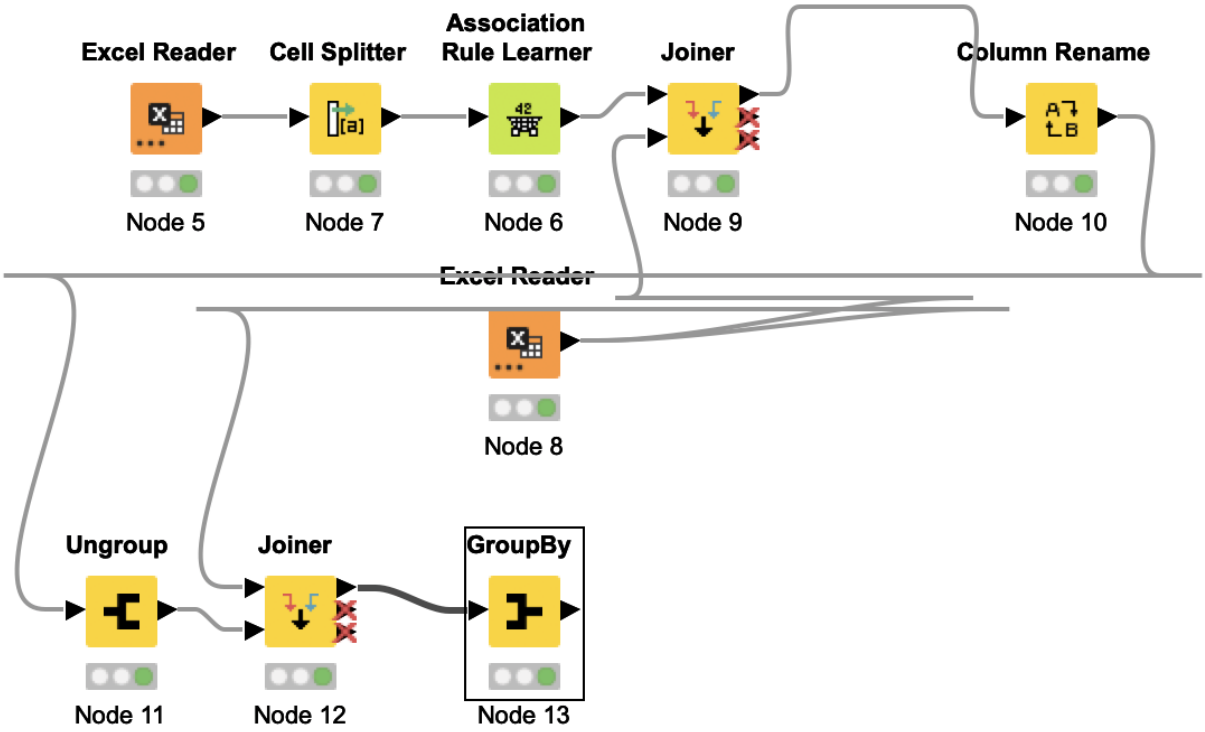
	ItemDesc
NIRVANA HOOKAH SINGLE	8392
MINT FLAVOUR SINGLE	5812
CAPPUCCINO	5474
GREAT LAKES SHAKE	4875
SAMBUCA	4423
POUTINE WITH FRIES	3432
QUA MINERAL WATER(1000ML)	3330
JR.CHL AVALANCHE	3307
CALCUTTA MINT	3307
B.M.T. PANINI	2621
GREEN APPLE FLAVOUR SINGLE	2528
MOROCCAN MINT TEA	2256
N R G HOOKAH	2236
MASALA CHAI CUTTING	2234
LEMON ICED TEA	2193

Use of Market Basket Analysis (Association Rules)

The association rule and its relevance in this case

- The Association rule helps us identify and establish the relationship present within variables in a dataset. In the case of Menu analysis, the usage of Assocaition use and Market Basket Analysis will identify which items in the menuare more frequently ordered together.
- It is very useful in understanding customer behaviour in terms of the product consumption.
- By Implimenting Market Basket Analysis, we will be able to make meaningful suggestions to the customer when orderding the the restaurant based on the products chosen.
- This can also let us create meaningful combos in the menu with enticing offers to boost sales.

KNIME workflow



Support

The Support value of a particular product gives you an understanding of the popularity of the product. It is the ratio between the number of times the product is purchased vs the total number of purchases of all the product.

Choosing a support value of 0.0009 after multiple iterations

Confidence

The Confidence between two products is the conditional probability that someone buys a product provided they purchase the other product.

Choosing a Confidence value of 0.3 after multiple iterations.

Associations Identified

Support	Confid...	Lift	Consequent	...	Items
0.001	0.36	9.671	B.M.T.PANINI	<...	[MAGGINDLARRABIATA,SAMBUCA]
0.001	0.47	7.439	SAMBUCA	<...	[MAGGINDLARRABIATA,B.M.T.PANINI]
0.001	0.465	79.364	LEMONINFUSEDCHARGRILLEDVEG	<...	[ADDHERBROASTCHICKEN]
0.001	0.412	5.924	GREATLAKESSHAKE	<...	[VANILLAICECREAM]
0.001	0.464	5.936	CAPPUCCINO	<...	[ADDHAZELNUTFLAVOUR]
0.002	0.546	18.235	CAFFELATTE	<...	[ADDHAZELNUTFLAVOUR]

Possible Combos with Lucrative Offers

From the Above table some of the Possibel combos are:

1. Cappuccino + Add Hazelnul Flavour
2. Caffe Latte + Add Hazelnul Flavour
3. Maggie Arrabiata + Sambuca + B.M.T.Panini
4. Lemon infused Char Grilled Veg + Herb roasted Chicken
5. Great Lakes Shake + Vanilla Icecream

Recommendations

1. The Restaurant should offer a discount on Hazelnul flavour for Cuppuccino and Caffe Latte.
2. The restaurant should offer B.M.T Panini for free when purchasing Maggindlarrabiata and Sambuca
3. The Restaurant should offer 50% off on Greatlakes shake when customer purchases Vanilla Icecream
4. The Restaurant should offer sale on Liquor and Tobacco items during the end of the month.
5. The Restaurant should offer sale of Misc items from march to october.