

# **MARKETING & RETAIL ANALYTICS**

**Ashesh Sahoo**



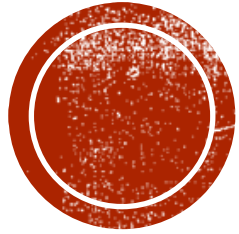
# PROBLEM STATEMENT

A Café chain shared data of one of its restaurants. The key objective here is to conduct thorough data analysis and come up with recommendations that can help the café chain to increase its revenue. Data provided is POS (Point of sale) for one of its chains. Goal is to identify the most popular combos that can be suggested to the Cafe chain after a thorough analysis of the most commonly occurring sets of menu items in the customer orders. The Store doesn't have any combo meals. Using the market basket analysis, recommend combo plans that café chain can offer to its customers

## **Steps to be followed :**

- Exploratory Analysis of data & an executive summary
- Trends across Products, Dates, Categories
- Importance of the association rules and their relevance in the current use case
- Add KNIME workflow Image
- The threshold values of Support and Confidence
- Associations in a tabular format
- Support, confidence & lift values are calculated
- Recommendations through analysis





# **EXPLORATORY DATA ANALYSIS**

# DATA SUMMARY

- There are no missing values in the data
- There are 145,830 rows and 10 Columns in the original dataset provided
- There are no duplicate rows in the data
- There are 69982 Unique Transactions and 580 Unique products
- The data is for the year April,2010 – Mar,2011
- There are 9 unique categories in the data
- Additional columns like Month, Weekday of Billed date, Hour, Minute of Billed time are included for further analysis



# DATA ANALYSIS

Below image shows the data summary at high level

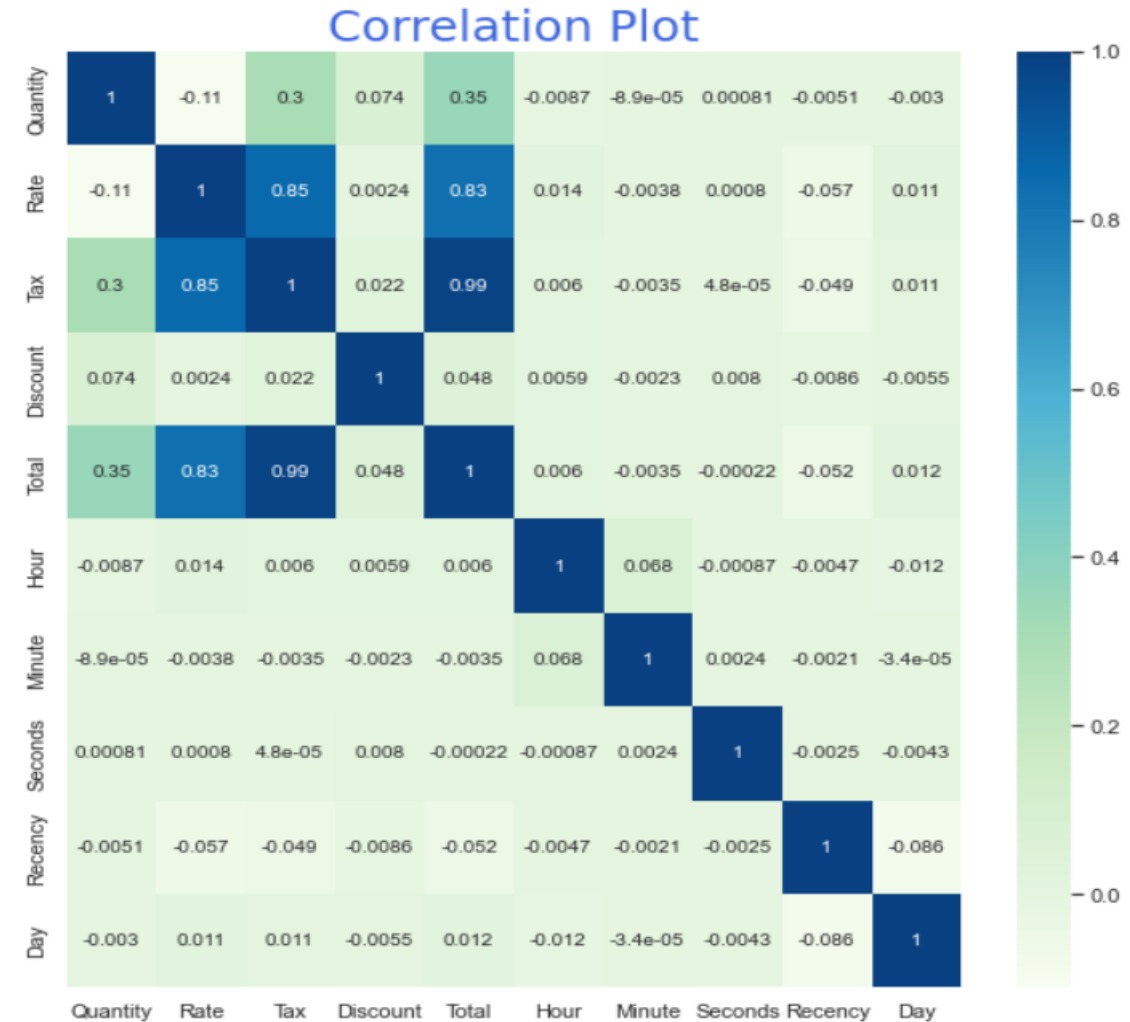
	count	mean	std	min	25%	50%	75%	max
Quantity	145830.0	1.121299	0.477237	1.00	1.00	1.00	1.0	30.00
Rate	145830.0	161.782259	102.244631	0.01	95.00	125.00	225.0	2100.00
Tax	145830.0	48.929061	40.272851	0.00	22.56	32.06	72.0	2731.25
Discount	145830.0	0.095079	3.720735	0.00	0.00	0.00	0.0	825.00
Total	145830.0	224.959852	164.960776	0.01	117.56	167.06	315.0	14231.25
Hour	145830.0	16.325070	5.634004	1.00	13.00	17.00	21.0	24.00
Minute	145830.0	29.366063	17.339664	0.00	14.00	29.00	45.0	59.00
Seconds	145830.0	28.476462	18.953535	0.00	14.00	30.00	45.0	59.00
Recency	145830.0	177.338367	103.828389	0.00	89.00	175.00	264.0	364.00
Day	145830.0	15.872811	8.881345	1.00	8.00	16.00	24.0	31.00

	Bill Number	Item Desc	Time	Category
count	145830	145830	145830	145830
unique	69982	580	36200	9
top	G0490530	NIRVANA HOOKAH SINGLE	22:25:36	FOOD
freq	23	8553	33	57023



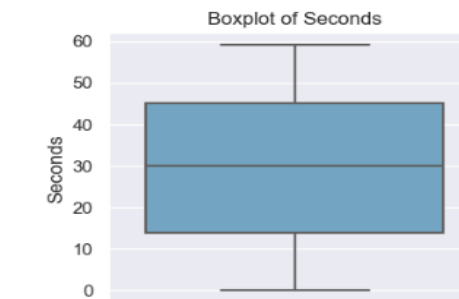
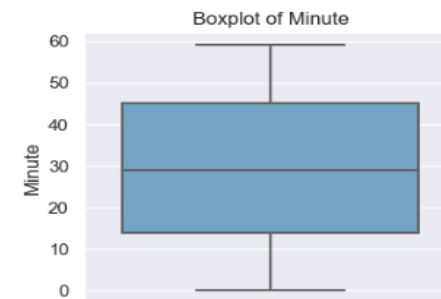
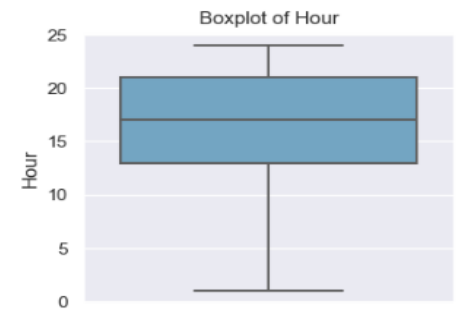
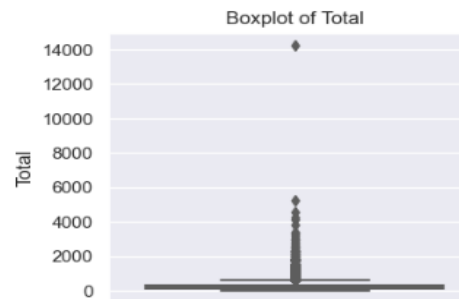
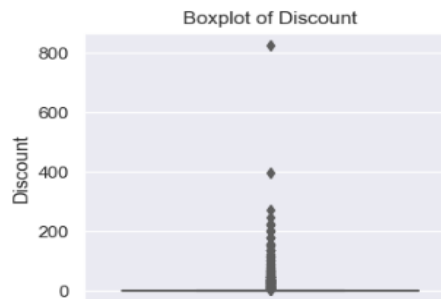
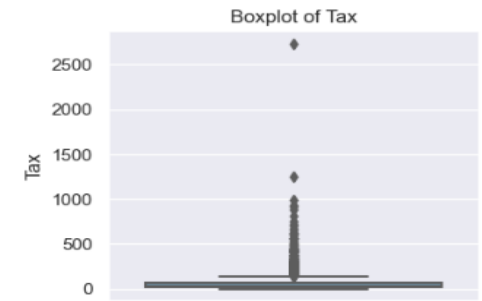
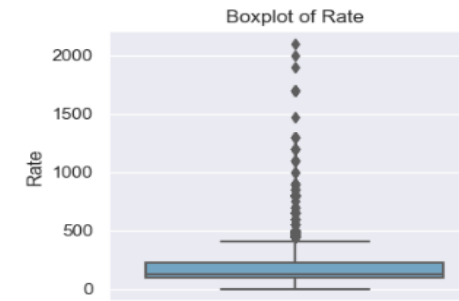
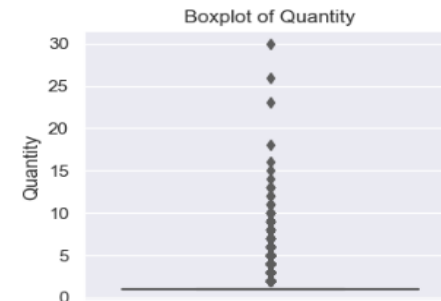
# CORRELATION CHART

- High correlations are noted between Total, Tax, and Rate.
- Some correlation can also be seen between Total, Quantity and Tax but it can be ignored as the value is low.



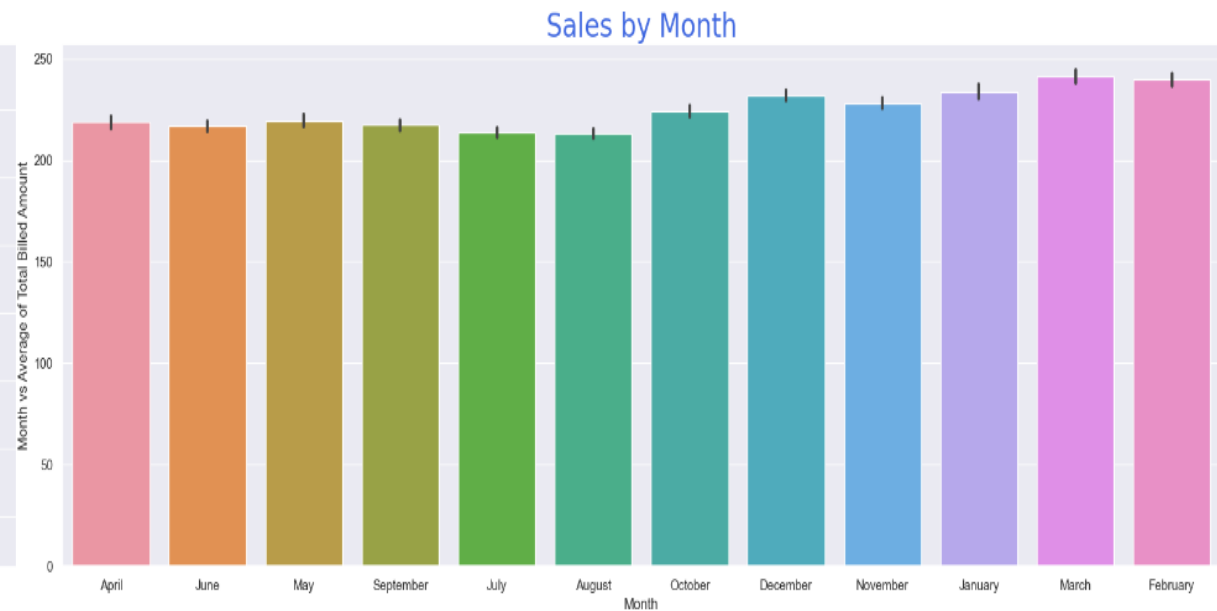
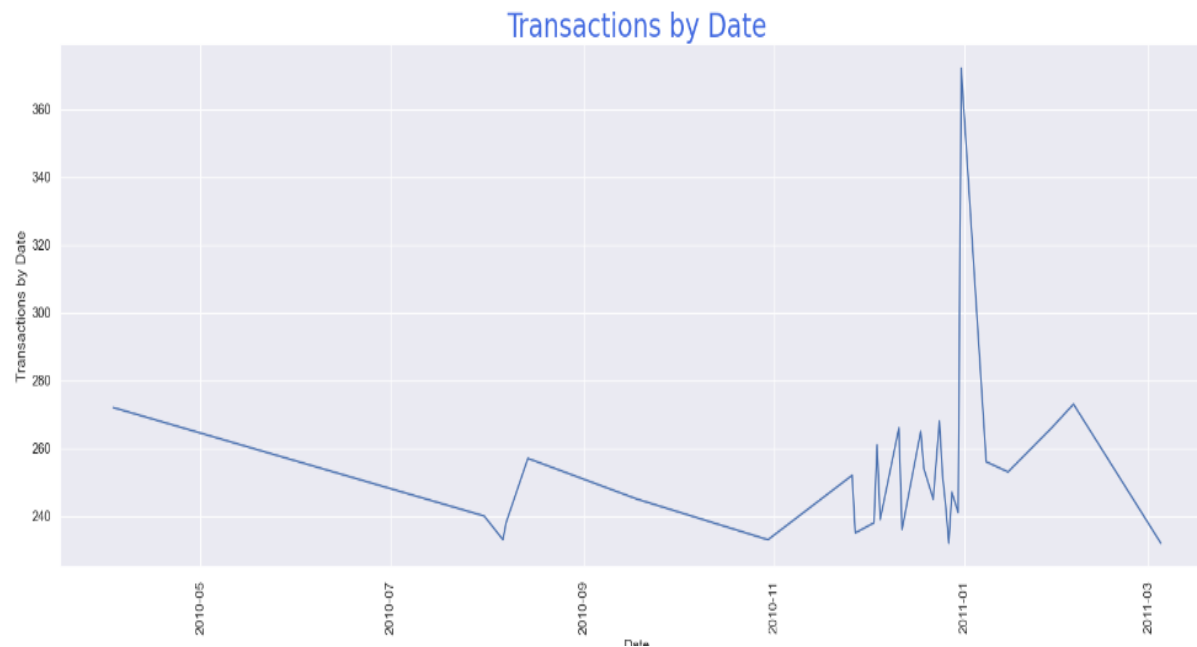
# BOXPLOTS

- Significant outliers are noted in Quantity, Rate, Tax, Discount, and Total
- Hours are skewed indicating more transaction happen in the during 2<sup>nd</sup> half of the day



# ORDER COUNTS TRENDS

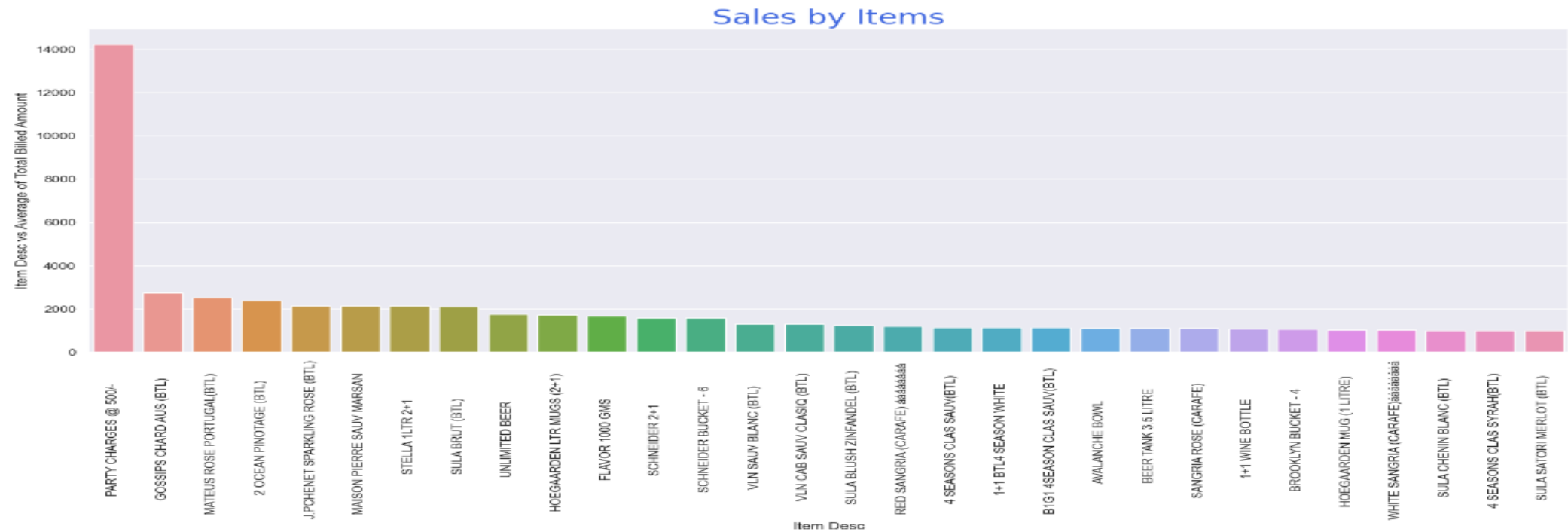
There are good number of transactions seen in the month of December and January. It could be due to Christmas and new year. Not enough data to check seasonality.





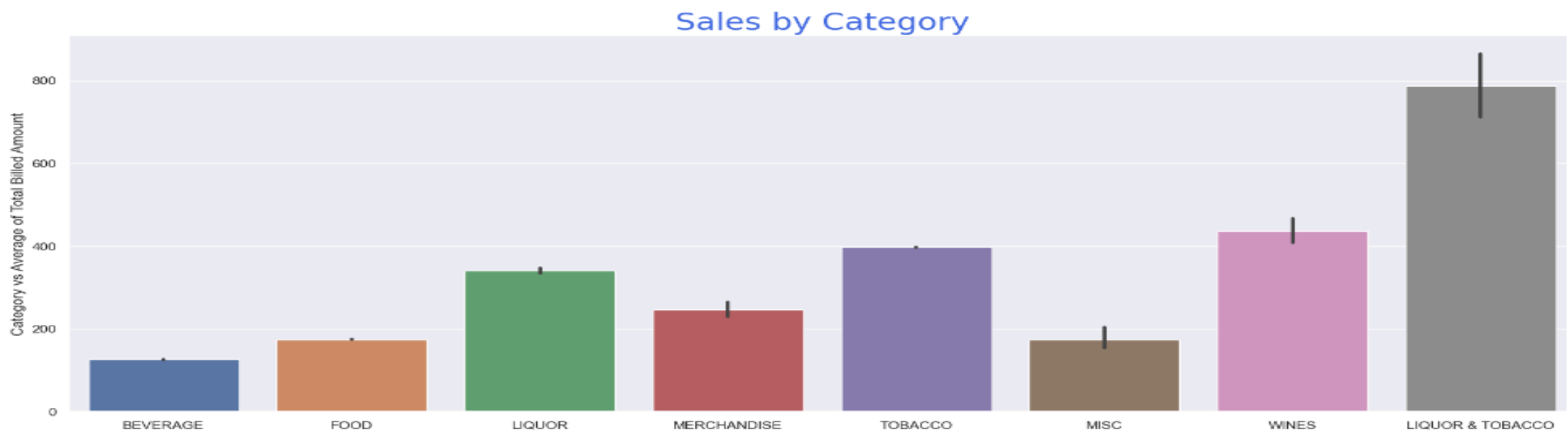
# SALES BY ITEM

Top product sold is “Party Charges”. Average sales is approx. 14000 .Subsequent products are sold at an average of approx. 2000.



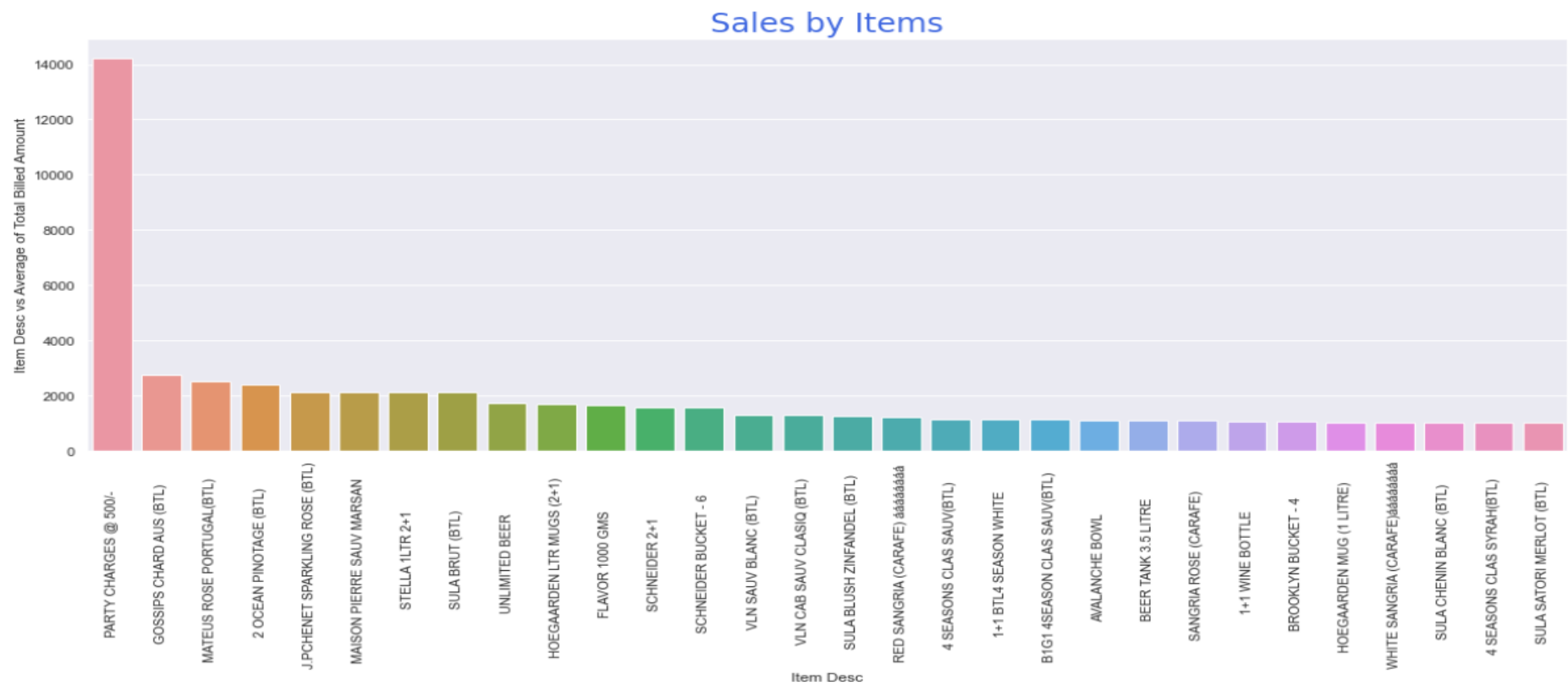
# SALES BY CATEGORY

Top product sold is “Liquor and Tobacco”. Average sales is approx. 800. Subsequent products are sold at an average of approx 400.



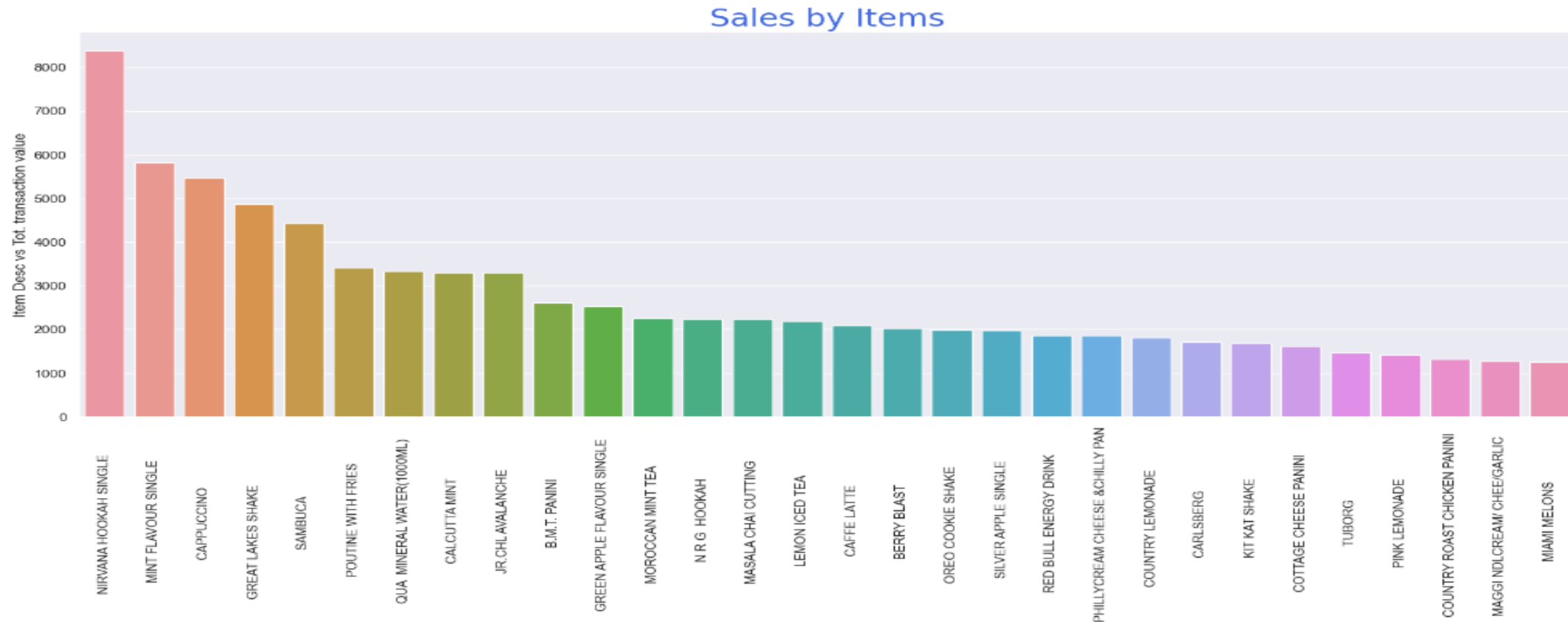
# ITEM WISE TOTAL

Top Item sold is “PARTY CHARGES @ 500/-”



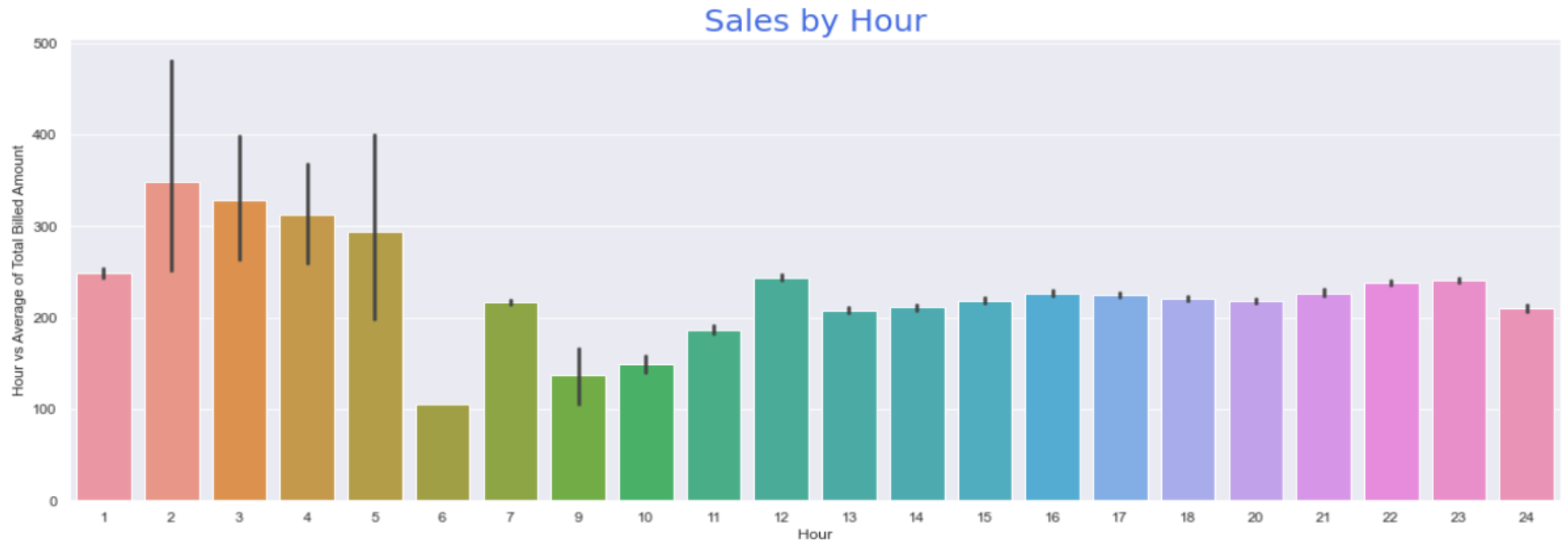
# ITEM WISE TRANSACTION VALUE

Top Item sold based on total transaction value is “NIRVANA HOOKAH SINGLE”



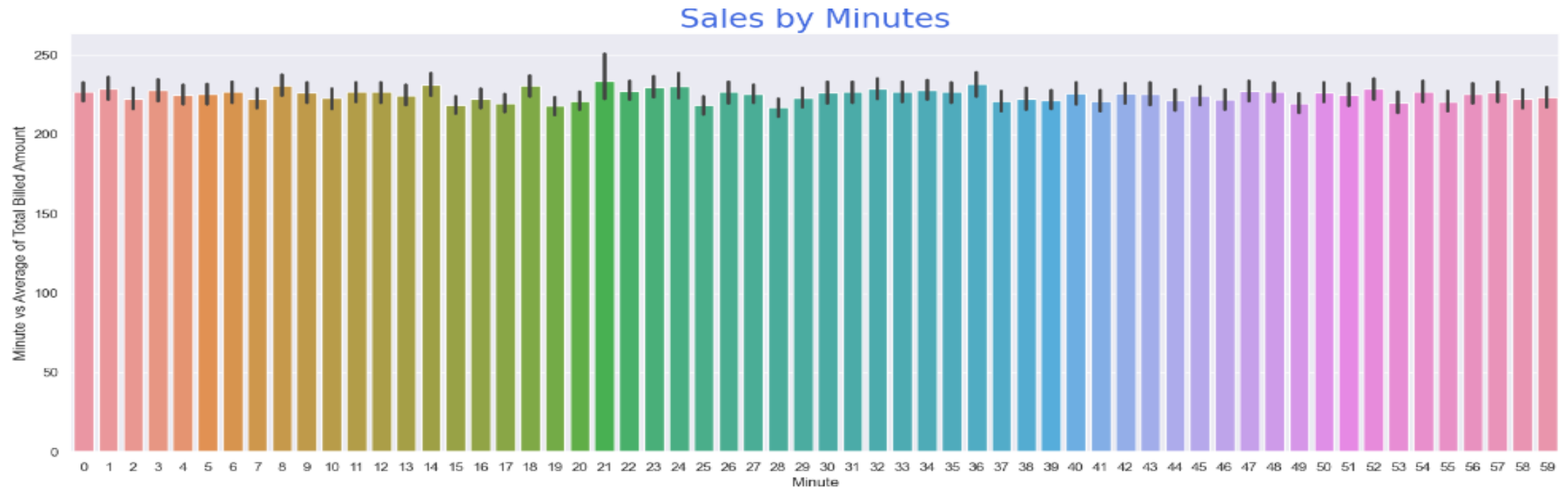
# HOURLY SALE TRANSACTION

Most sales transaction during 1<sup>st</sup> half of the day. Average sales cross approx. 300 in the initial 5 hours. Average sales is close to 200 for rest of the day



# MINUTE WISE SALE TRANSACTIONS

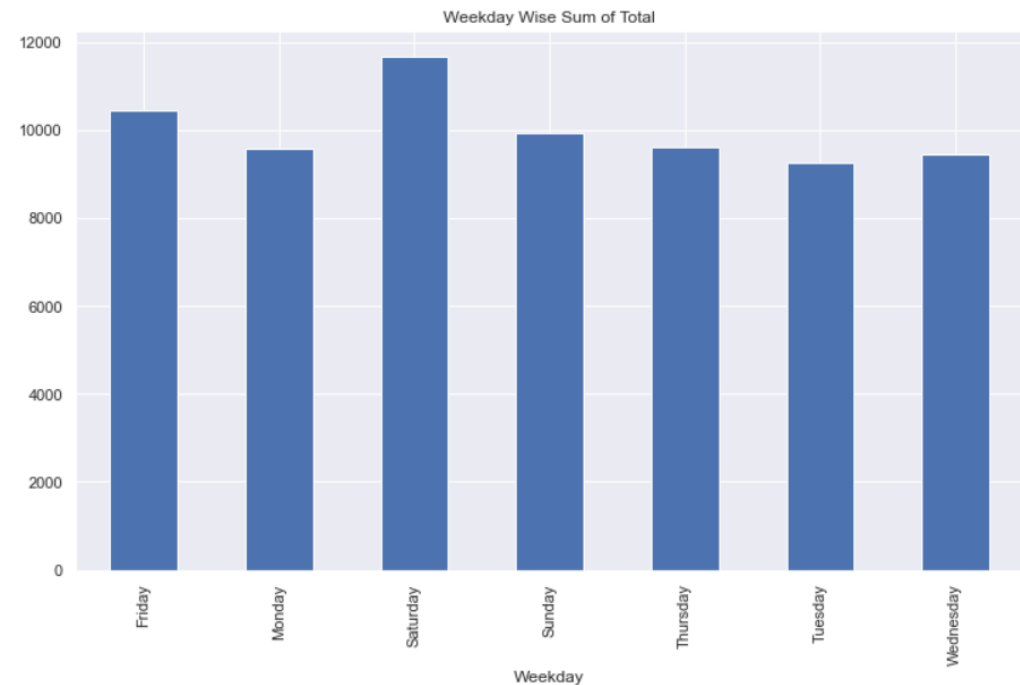
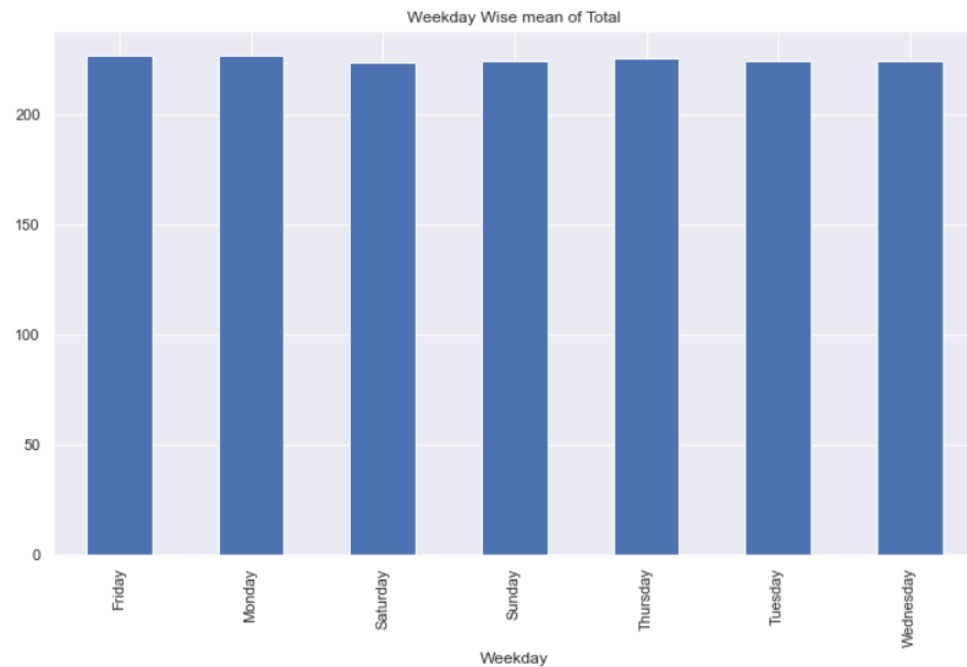
No correlation is found between sales and minute of Sales transaction



# TRANSACTIONS BY DAY OF WEEK

Weekend i.e. “Saturday” and “Sunday” witnesses more transactions compared to other days in the week

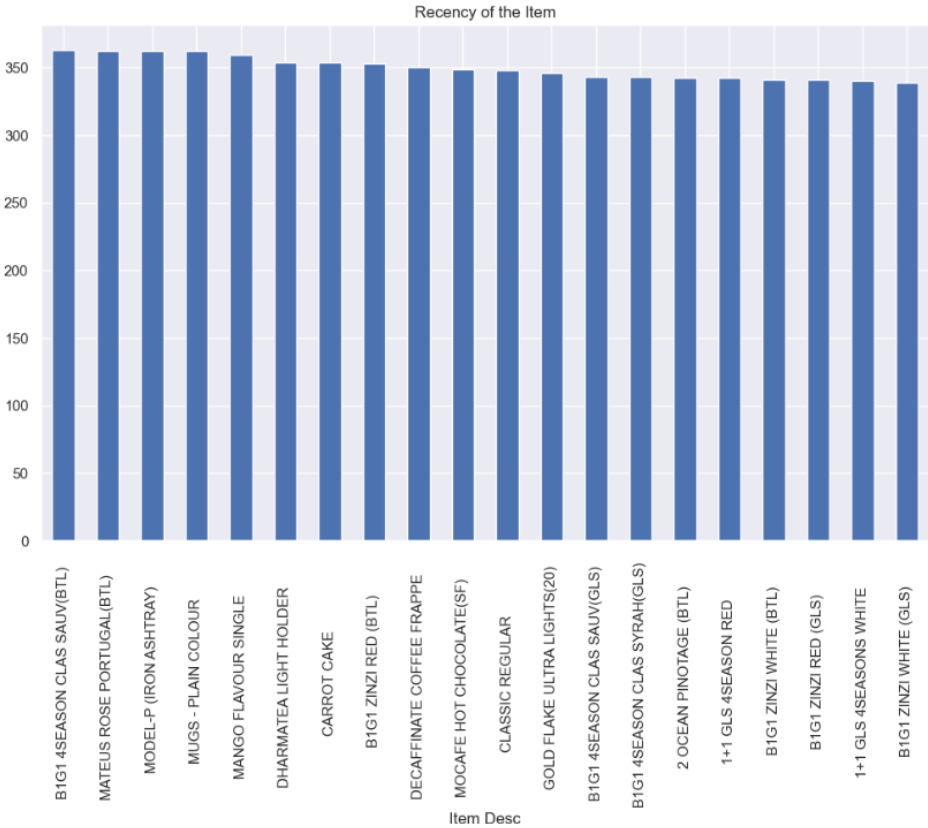
Average total is similar across all days of the week



# ITEMS SUGGESTED TO BE DROPPED FROM THE MENU

Items shown in the image were purchased an year ago and have no latest transactions. Recency is calculated from base date i.e. 01 Apr 2010 .

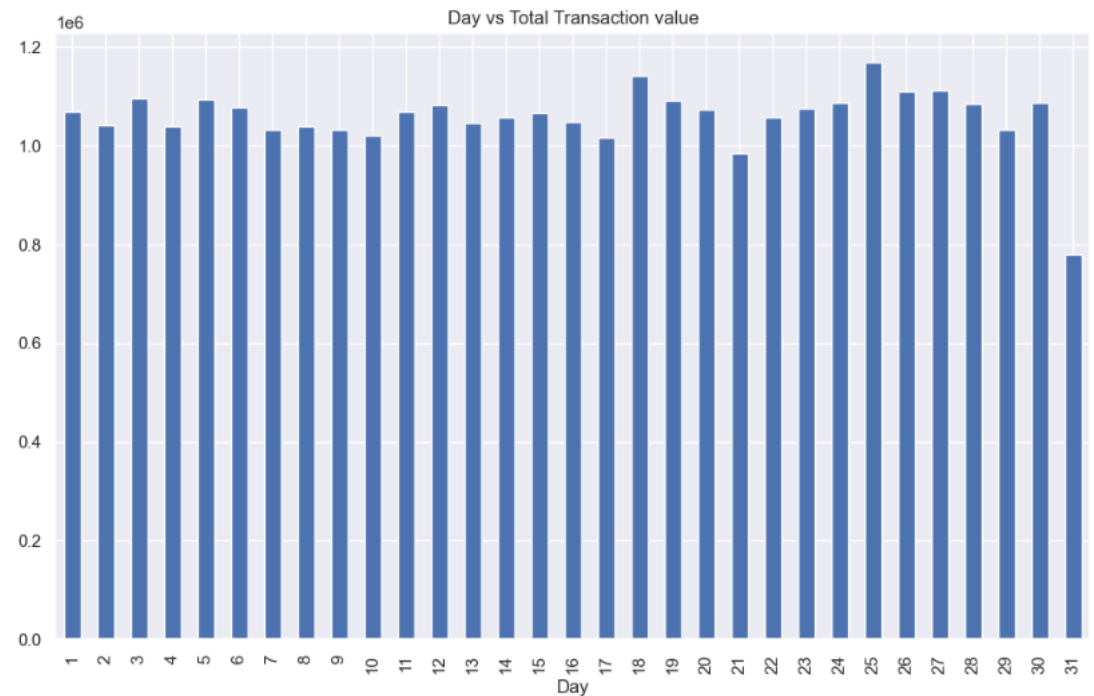
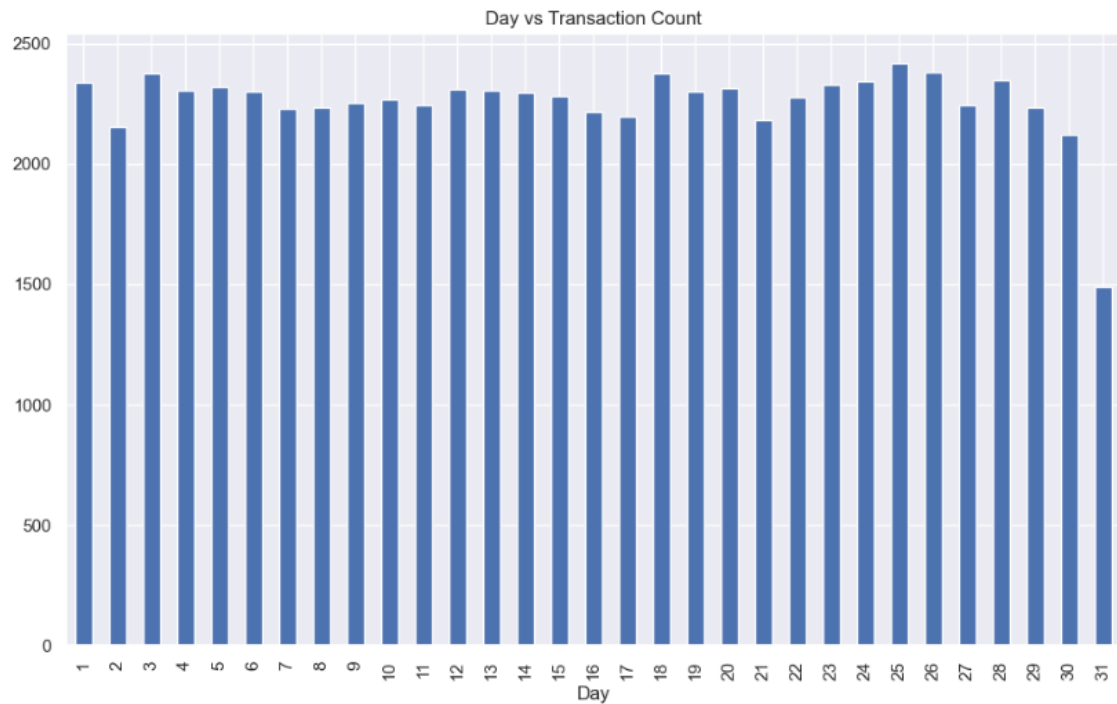
Items shown in the image are suggested to be dropped from the food menu.





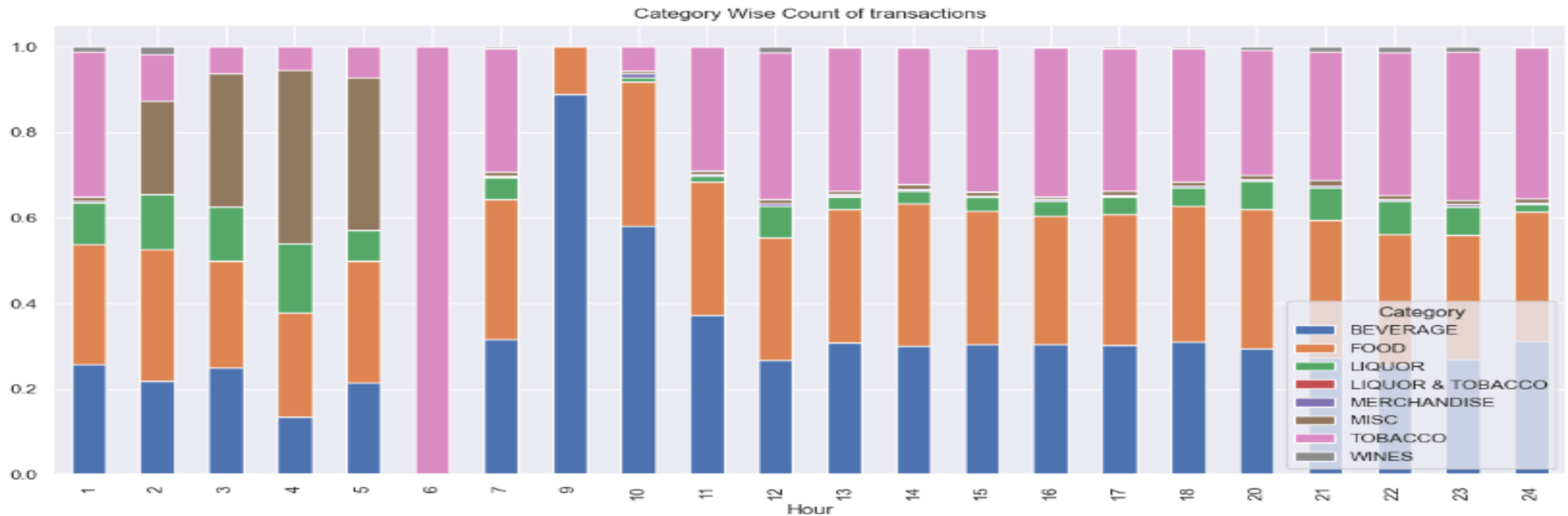
# TRANSACTIONS BY DAY OF MONTH

No trend is identified on the Day of month w.r.t Average of Total transaction value or Transaction count



## TRANSACTIONS BY CATEGORY DURING THE DAY (HOUR)

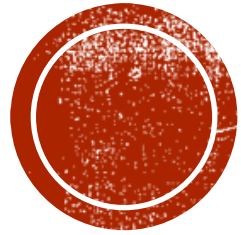
- Good number of transactions are recorded for Category “Misc” initial hours of the day
- Transactions for only 1 category i.e. “Tobacco” is seen during 6<sup>th</sup> hour of the day
- Very high transactions for “Merchandise” is seen during 9<sup>th</sup> hour of the day.



# KEY OBSERVATIONS

- Liquor& Tobacco is the most frequently ordered item followed by Wines and Tobacco
- Data set is for 9 months of 2010, 3 months of 2011
- 1<sup>st</sup> half of the day records the highest amount of Total transaction value.
- Initial few hours of the day have relatively high transactions related to Category “Misc” than other parts of the day
- There is no significant trend between the minutes of transactions compared to Total transaction value/Transaction count
- Saturday and Sunday have high transactions compared to other weekdays.
- Products like “B1G1 4SEASON CLAS SAUV(BTL)”, and “MATEUS ROSE PORTUGAL(BTL)” “seem to be obsolete items, that can be removed from the menu.
- Transactions for only 1 category i.e. “Tobacco” is seen during 6th hour of the day
- Very high transactions for “Merchandise” is seen during 9th hour of the day.





# **MARKET BASKET ANALYSIS**

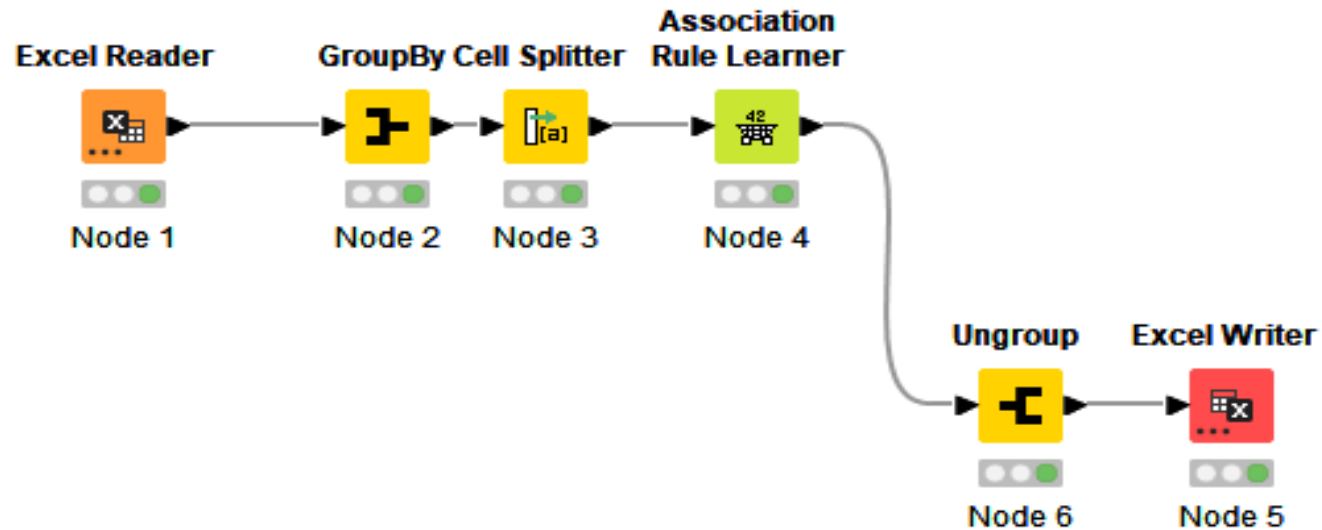
# WHY MARKET BASKET ANALYSIS ?

- Problem statement involves the business of a café chain restaurant where there is a lot of transactions across multiple items and Categories. Café chain owner wants to understand customer purchase pattern so that combo offers can be designed and presented to customer.
- The output of it can be used as changes to offerings to customer which will result in increased customer experience, customer retention, sales, etc. that'll help the business continue to grow revenue, profit
- Association rules will help to identify the items that are most likely to be purchased by the customer when they purchase a specific item
- key metrics using which purchase of potential items is evaluated are Support, Lift, and confidence.



# KNIME WORKFLOW

- Threshold value for Support is 0.001 and Confidence is 0.001
- 633 association rules are formed using above mentioned support and confidence values.



# OUTPUT OF ASSOCIATION RULES LEARNER

RowID	Support	Confidence	Lift	Consequent	implies	Items
rule58_1	0.001043125	0.17804878	79.36439335	ADD HERB ROAST CHICKEN	<---	LEMON INFUSED CHAR GRILLED VEG
rule59_1	0.001043125	0.464968153	79.36439335	LEMON INFUSED CHAR GRILLED VEG	<---	ADD HERB ROAST CHICKEN
rule366_1	0.001614701	0.54589372	18.23519537	CAFFE LATTE	<---	ADD HAZELNUT FLAVOUR
rule367_1	0.001614701	0.053937947	18.23519537	ADD HAZELNUT FLAVOUR	<---	CAFFE LATTE
rule116_1	0.001128862	0.105053191	7.478975022	SATAY CHICKEN PANINI	<---	ADD FRIES
rule117_1	0.001128862	0.080366226	7.478975022	ADD FRIES	<---	SATAY CHICKEN PANINI
rule27_1	0.001014547	0.027255278	7.143741149	FRENCH FRIES	<---	B.M.T. PANINI
rule26_1	0.001014547	0.265917603	7.143741149	B.M.T. PANINI	<---	FRENCH FRIES
rule195_1	0.001243177	0.06580938	6.124297888	ADD FRIES	<---	COUNTRY ROAST CHICKEN PANINI
rule194_1	0.001243177	0.115691489	6.124297888	COUNTRY ROAST CHICKEN PANINI	<---	ADD FRIES
rule258_1	0.001371781	0.463768116	5.935519439	CAPPUCCINO	<---	ADD HAZELNUT FLAVOUR
rule259_1	0.001371781	0.017556693	5.935519439	ADD HAZELNUT FLAVOUR	<---	CAPPUCCINO
rule196_1	0.001243177	0.412322275	5.923863158	GREAT LAKES SHAKE	<---	VANILLA ICECREAM
rule197_1	0.001243177	0.017860809	5.923863158	VANILLA ICECREAM	<---	GREAT LAKES SHAKE



# INFERENCE (HIGHEST LIFT)

RowID	Support	Confidence	Lift	Consequent	implied	Items
rule58_1	0.001043125	0.17804878	79.36439335	ADD HERB ROAST CHICKEN	<---	LEMON INFUSED CHAR GRILLED VEG
rule59_1	0.001043125	0.464968153	79.36439335	LEMON INFUSED CHAR GRILLED VEG	<---	ADD HERB ROAST CHICKEN
rule366_1	0.001614701	0.54589372	18.23519537	CAFFE LATTE	<---	ADD HAZELNUT FLAVOUR
rule367_1	0.001614701	0.053937947	18.23519537	ADD HAZELNUT FLAVOUR	<---	CAFFE LATTE
rule116_1	0.001128862	0.105053191	7.478975022	SATAY CHICKEN PANINI	<---	ADD FRIES

## Rule58\_1

- Support indicates that 0.10% of total transactions were of “Lemon Infused Char Grilled Veg”
- Confidence indicates that 17.8% of the total transactions where “Lemon Infused Char Grilled Veg” was purchased also had a purchase of “Add Herb Roast Chicken”
- lift is there is a 7936% increase in expectations that customers will buy “Add Herb Roast Chicken” when they purchase “Lemon Infused Char Grilled Veg”

## Rule366\_1

- Support indicates that 0.16% of the transactions were for “Add Hazelnut Flavor”
- Confidence says that 54.5% of the customers who purchased “Add Hazelnut Flavor” also purchased “Caffe Latte”
- lift is that there is a 1823% increase in expectations that customers will buy “Caffe Latte” when they purchase “Add Hazelnut Flavor”





## INTERPRETATION – MOST PURCHASED ITEM (NIRVANA HOOKAH SINGLE )

RowID	Support	Confidence	Lift	Consequent	implied	Items
rule579_1	0.003086508	0.025748003	1.544041791	RED BULL 2+1	<---	NIRVANA HOOKAH SINGLE
rule631_1	0.006330199	0.052807248	1.082471236	POUTINE WITH FRIES	<---	NIRVANA HOOKAH SINGLE

### rule579\_1

- Support indicates that 0.3% of the transactions were for “Nirvana Hookah Single”
- Confidence says that 2.5% of the total transaction where “Nirvana Hookah Single” was purchased also had a purchase of “Red Bull 2+1”
- lift is that there is a 54% increase in expectations that customers will buy “Red Bull 2+1” when we know that they purchase “Nirvana Hookah Single”



## SELECTING TOP RULES ( CONFIDENCE > 0.2 AND LIFT > 2)

RowID	Support	Confidence	Lift	Consequent	implied	Items
rule480_1	0.002157698	0.200797872	5.394332707	B.M.T. PANINI	<---	ADD FRIES 1
rule26_1	0.001014547	0.265917603	7.143741149	B.M.T. PANINI	<---	FRENCH FRIES
rule366_1	0.001614701	0.54589372	18.23519537	CAFFE LATTE	<---	ADD HAZELNUT FLAVOUR 2
rule258_1	0.001371781	0.463768116	5.935519439	CAPPUCCINO	<---	ADD HAZELNUT FLAVOUR
rule196_1	0.001243177	0.412322275	5.923863158	GREAT LAKES SHAKE	<---	VANILLA ICECREAM
rule59_1	0.001043125	0.464968153	79.36439335	LEMON INFUSED CHAR GRILLED VEG	<---	ADD HERB ROAST CHICKEN
rule604_1	0.004143923	0.248500428	3.931846481	SAMBUCA	<---	RED BULL 2+1

### rule480\_1

- Support indicates that 0.2% of total transactions were for “Add Fries”
- Confidence indicates that 20% of the transactions that involved a purchase of “Add Fries” also had a purchase of “B.M.T. PANINI”
- lift is that there is a 439% increase in expectations that customers will buy “B.M.T. PANINI” when they purchase “Add Fries”



# Recommendation

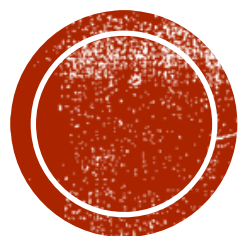
Using the rules, we can suggest the business that they can introduce a combo with B.M.T. PANINI, French Fries and Add Fries

With the data, we can see that 754 Bill Numbers have “Add Fries”, 2,606 Bill Numbers have “B.M.T. PANINI”, 268 B Numbers have “French Fries”.

Items	Sum of Total	No of Bills
Add Fries	40,422	752
B.M.T. PANINI	391,067	2,605
French Fries	12,560	267

Using this data, we can suggest a combo of Add Fries, B.M.T PANINI, French Fries as Platter with suitable price.





**THANK YOU**

