The background is a blue gradient. In the corners, there are decorative white line art elements resembling circuit boards or neural network connections, with small circles at the end of the lines.

Market Basket Analysis

Extended MRA PROJECT CAFÉ CHAIN DATASET

ANIL KUMAR PGBPDSBA AUG'23

CAFÉ CHAIN DATA ANALYSIS

- ▶ Case Study: The dataset provided is the dataset of a Café Chain of one of the restaurants. This data is of the years 2010 and 2011.
- ▶ Problem Statement: To Conduct EDA and derive most commonly occurring sets of items using Association rules. Recommend customers top selling products which result in increase of Revenue of the Café Chain.
- ▶ Need of Study:
- ▶ Objective: To check the Sales Data of the restaurant to analyse trends and patterns of Consumer Behaviour, to find out top selling products and items. Conducting Exploratory Data Analysis and Menu Analysis to identify most commonly occurring sets of items which are most bought by the customers.
- ▶ Top Categories of Products sold at the Restaurant are Tobacco products, Food products, Beverages, Liquor, Merchandise products, Liquor & Tobacco products, Merchandise and Wines.

EXPLORATORY DATA ANALYSIS

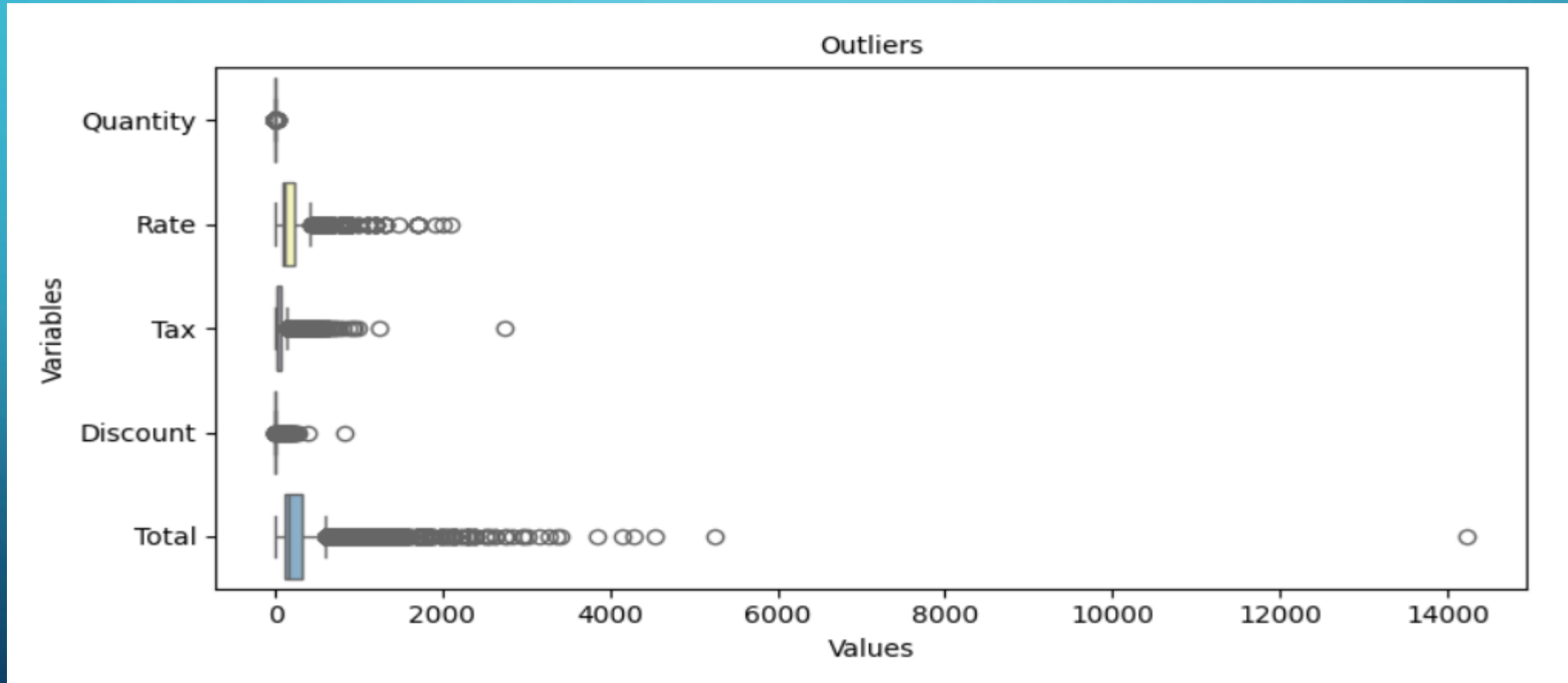
- ▶ **1. Data Overview:** Sample of the dataset:

	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

- ▶ **Shape of the Data:** There are 10 columns in the Data 10 and 145830 rows in each.
- ▶ **Data Types:** Date and Time are datetime data types.
Bill Number, Item desc and Category are object Data types.
Quantity, Rate, Tax, Discount and Total or Sales are Float or numeric type Data types.
- ▶ **Data Quality:** There are 0 null values.
There are 680 duplicate records, which have been dropped to get a clean dataset.
There are no junk values in the dataset.

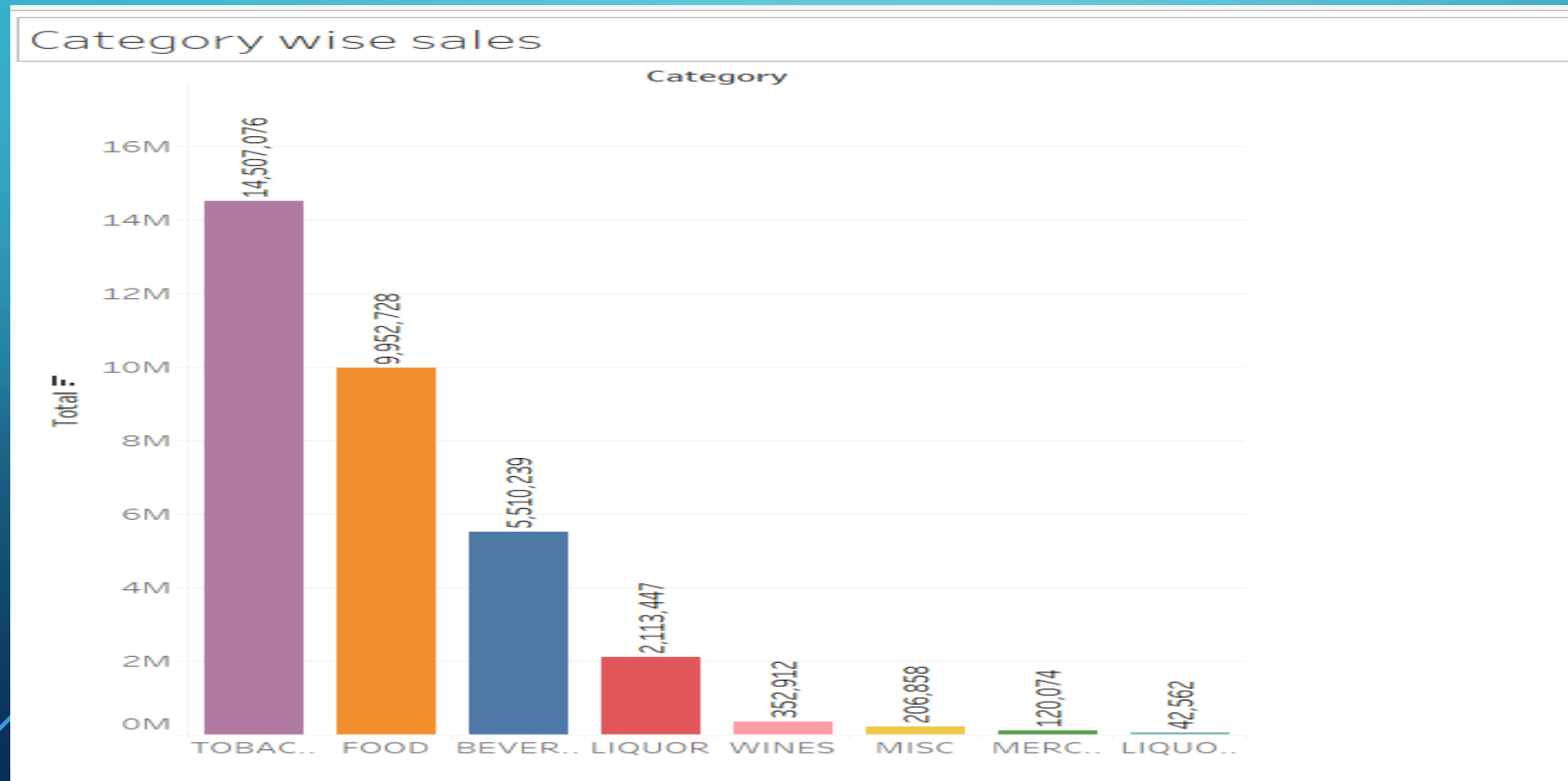
EXPLORATORY DATA ANALYSIS - OUTLIERS

- ▶ Outliers in the dataset are observed and retained for data exploration and insights.



UNIVARIATE ANALYSIS

- Tobacco is the category with the highest Sales Revenue. Followed by,
- Food, Beverage and Liquor.
- Liquor & Tobacco, Merchandise, Miscellaneous and Wines are underperforming Categories.



TOP PERFORMING OR MOST LOVED ITEMS

- Nirvana Hookah Single, Sambuca, Mint Flavour are some of the top selling Tobacco Products.
- Poutine with Fries, Great Lakes Shake, BMT Panini are some of the top selling items under Food Category.
- Cappuccino, Red Bull 2+1 and KF Draught Pitcher (2Ltr) are top selling Beverage Products.
- along with TUBORG are top selling Liquor Products.

WorldCloud for Most Loved Items

CARLSBERG RED BULL 2+1
POUTINE WITH FRIES CAPPUCINO
GREAT LAKES SHAKE CALCUTTA MINT
N R G HOOKAH GREEN APPLE FLAVOUR SINGLE
NIRVANA HOOKAH SINGLE TUBORG
SAMBUCA ARABIAN MIST MINT FLAVOUR SINGLE
JR. CHL AVALANCHE KF DRAUGHT PITCHER (2LTR) SILVER APPLE SINGLE
OREO COOKIE SHAKE
RED BULL ENERGY DRINK RED WINE SHEESHA B.M.T. PANINI

LEAST PERFORMING PRODUCTS

- Merchandise Items LIKE Ch Tin Small, Country Lemonade Glass, Caponata Food Item, Class Menthol Rush etc are poorly performing items along with others in the WordCloud.
- Ground Meat, Chicken Bacon are least preferred items under miscellaneous which are not usual toppings generally preferred with their meals.

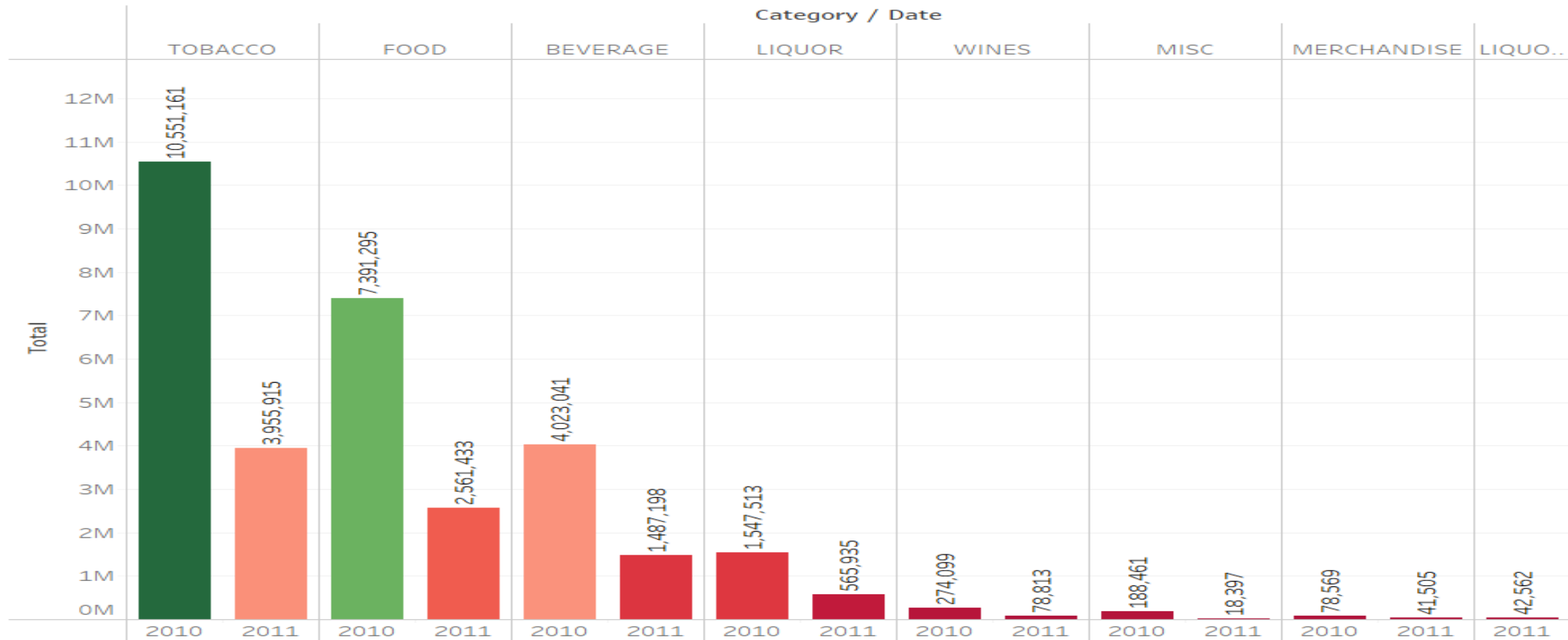
Least Performing Items

ASH TRAYS
CH WRAPPING PAPER CUTTING GLASS
CH TIN SMALL CAPONATA
INDIA KINGS OCEAN BLUE DIP BOWL MOCAFE HOT CHOCOLATE(SF)
DECAFFINATE COFFEE FRAPPE CLASSIC MENTHOL RUSH
COUNTRY LEMONADE GLASS(HANSA) MOTHERS DAY SPL
ADD BUTTERED TOAST
ADD GROUND MEAT GREAT LAKES MUG SINGLE
GOLD FLAKE ULTRA LIGHTS(20) ADD CHICKEN BACON
2 AXE TWIST CLASSIC REGULAR
MUGS - PLAIN COLOUR

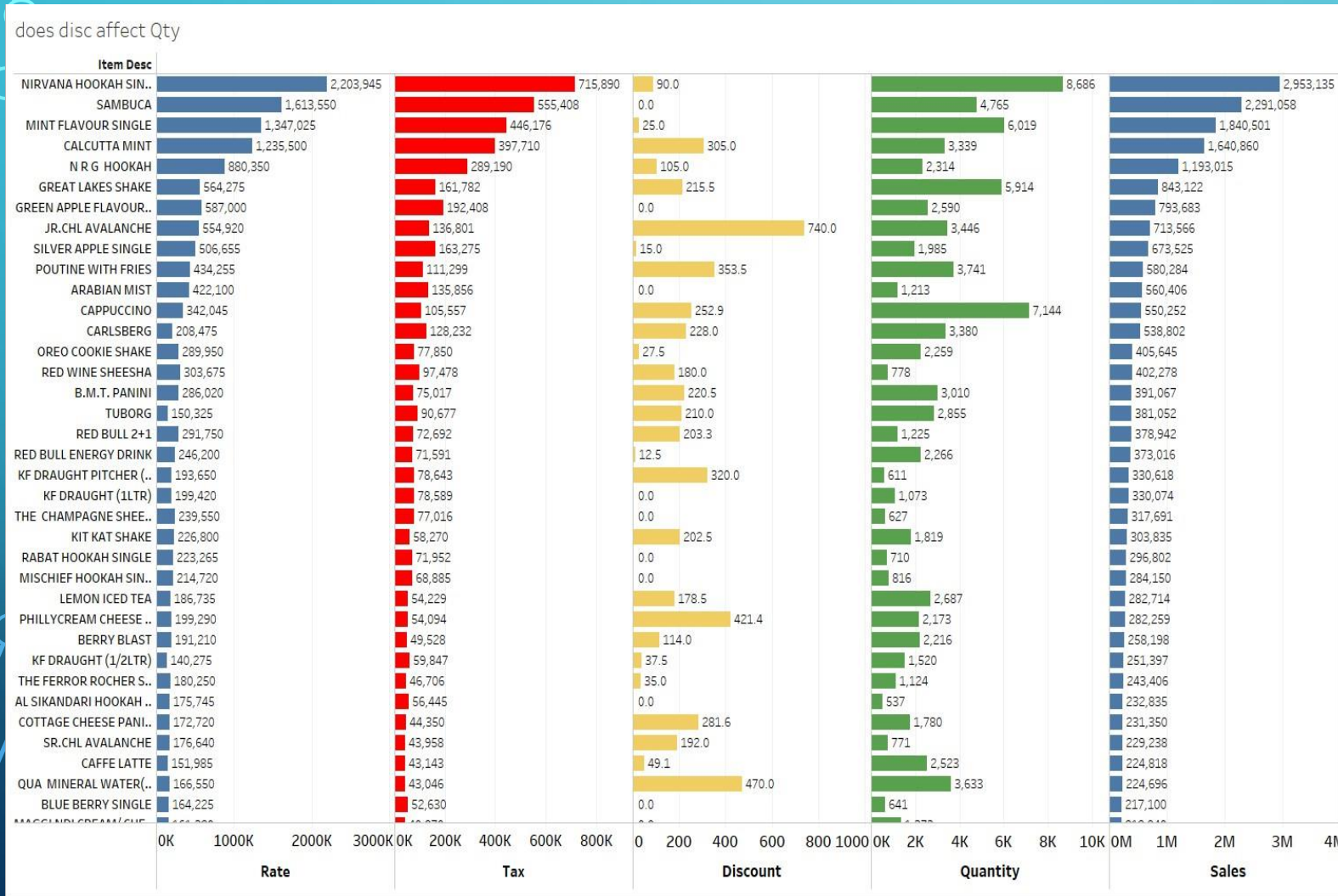
YEARLY SALES COMPARISON

- Tobacco, Food and Beverage Categories have performed very well in the year 2010.
- There is a sharp decrease in the Sales of almost all categories in the year 2011. However, Food and Tobacco seem to be having approx. equal Sales in the year 2011.
- Liquor & Tobacco seems to be a new category introduced in the year 2011 as data for 2010 is not present.
- Merchandise, Wines and Misc have been consistently low both years.

Yearly Sales Comprison



CONSUMER BEHAVIOUR PATTERNS



Most of the Tobacco and Food items are levied high taxes on them.

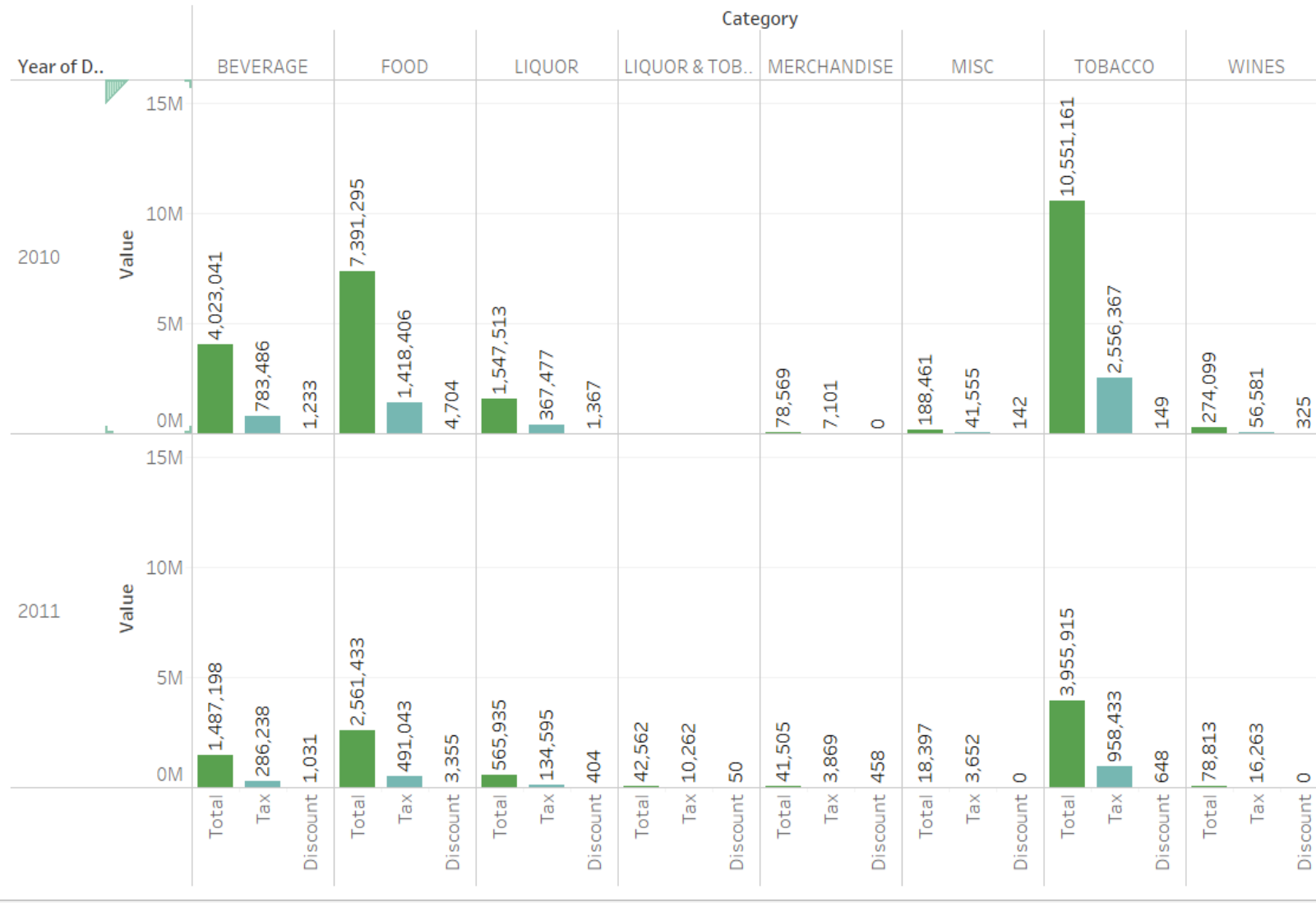
Discounts on items does not seem to increase the quantity as customers purchase items anyway.

Huge Discounts are observed for Jr. CHL Avalanche item, however, order quantities have not increased as it has for non discounted items such as Nirvana Hookah Single.

Cappucino and QUA Mineral Water, Great Lakes Shake, BMT Panini are the only items whose quantities have increased as discounts have increased.

PRICE LEVEL COMPARISON YEAR WISE

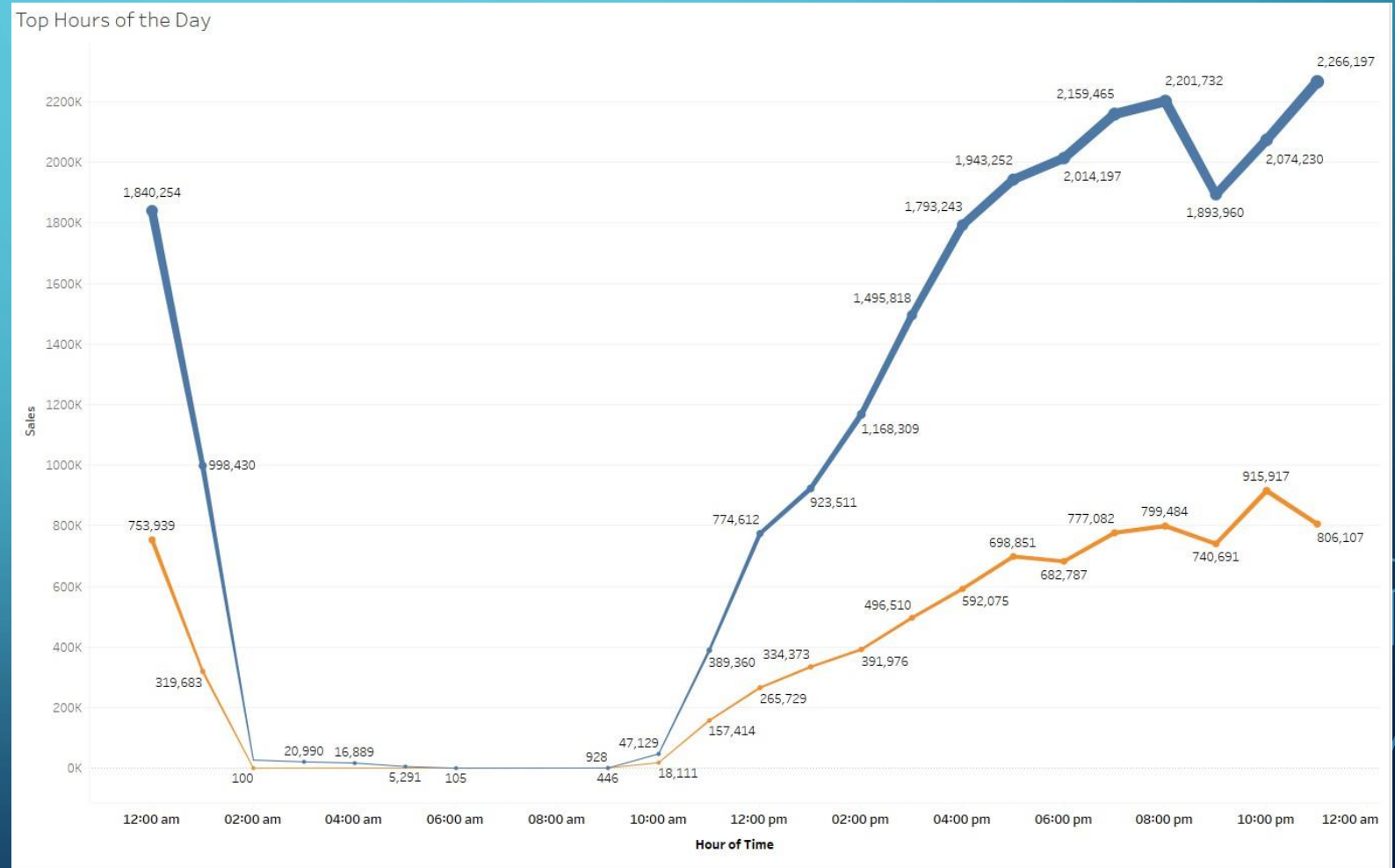
Sales/Tax/disc comparison Sales



- Tax Values remained the same for Beverages
- As observed, Tobacco, Food and Beverage made the highest Sales Revenue in the year 2019, but a sudden drop is observed in the year 2011.

CUSTOMER TRENDS DURING DIFFERENT TIMES OF THE DAY

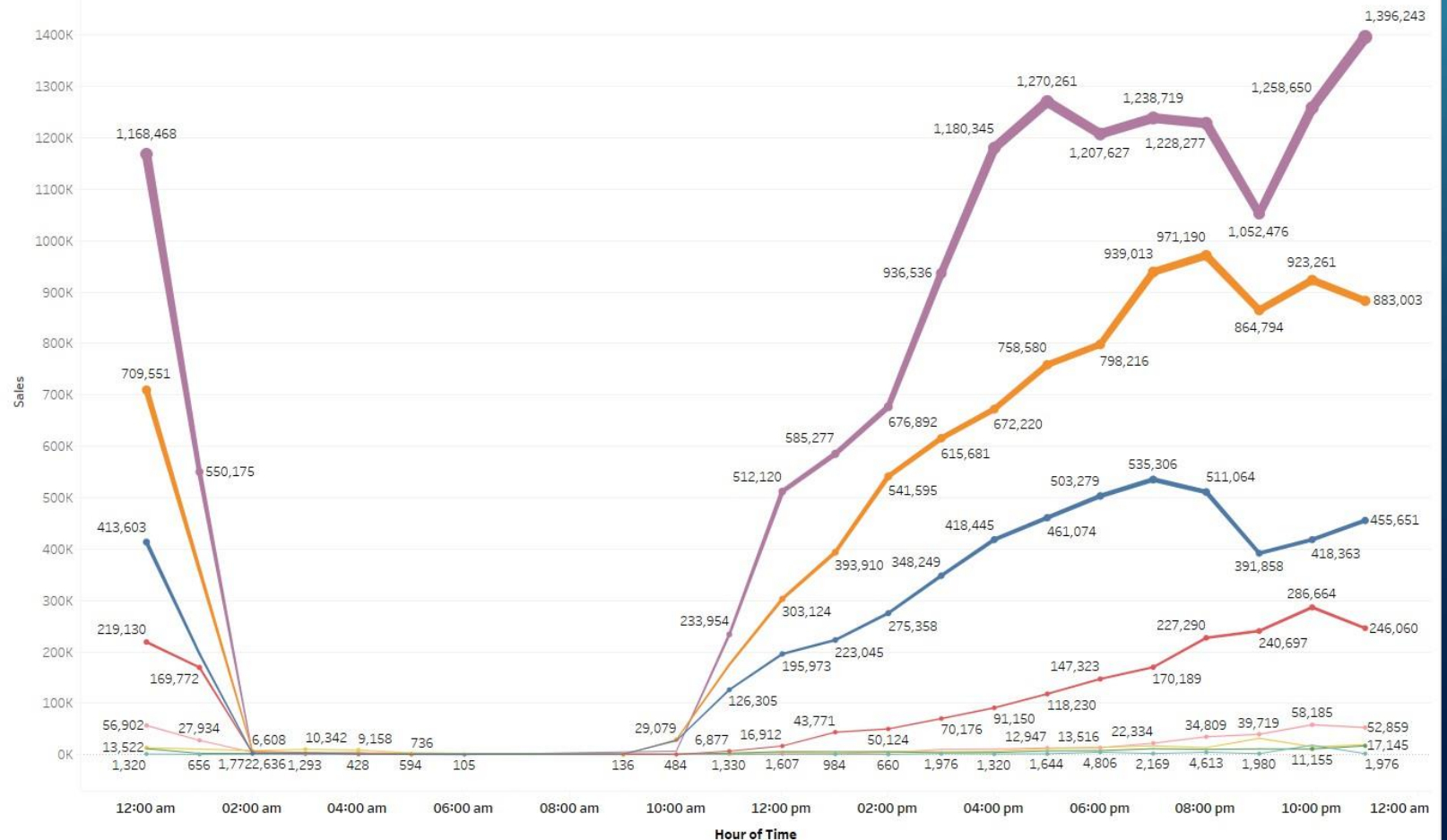
- Highest Footfall at the Café Chain is observed at night 8:00pm.
- The 'customer-rush' continues from morning 10:00AM until midnight 12:00 AM.
- This is observed during both years of 2010(in blue) and 2011(in orange)
- The Café does not shut down until 2:00am, and still continues to get orders during the wee hours.
- A sharp decrease in the count of customers ins observed in the year 2011, even though the Café operates during the same hours, and also earns the most profits at 8pm and then again at 12:00AM.



TOP CATEGORIES- HOURS OF THE DAY

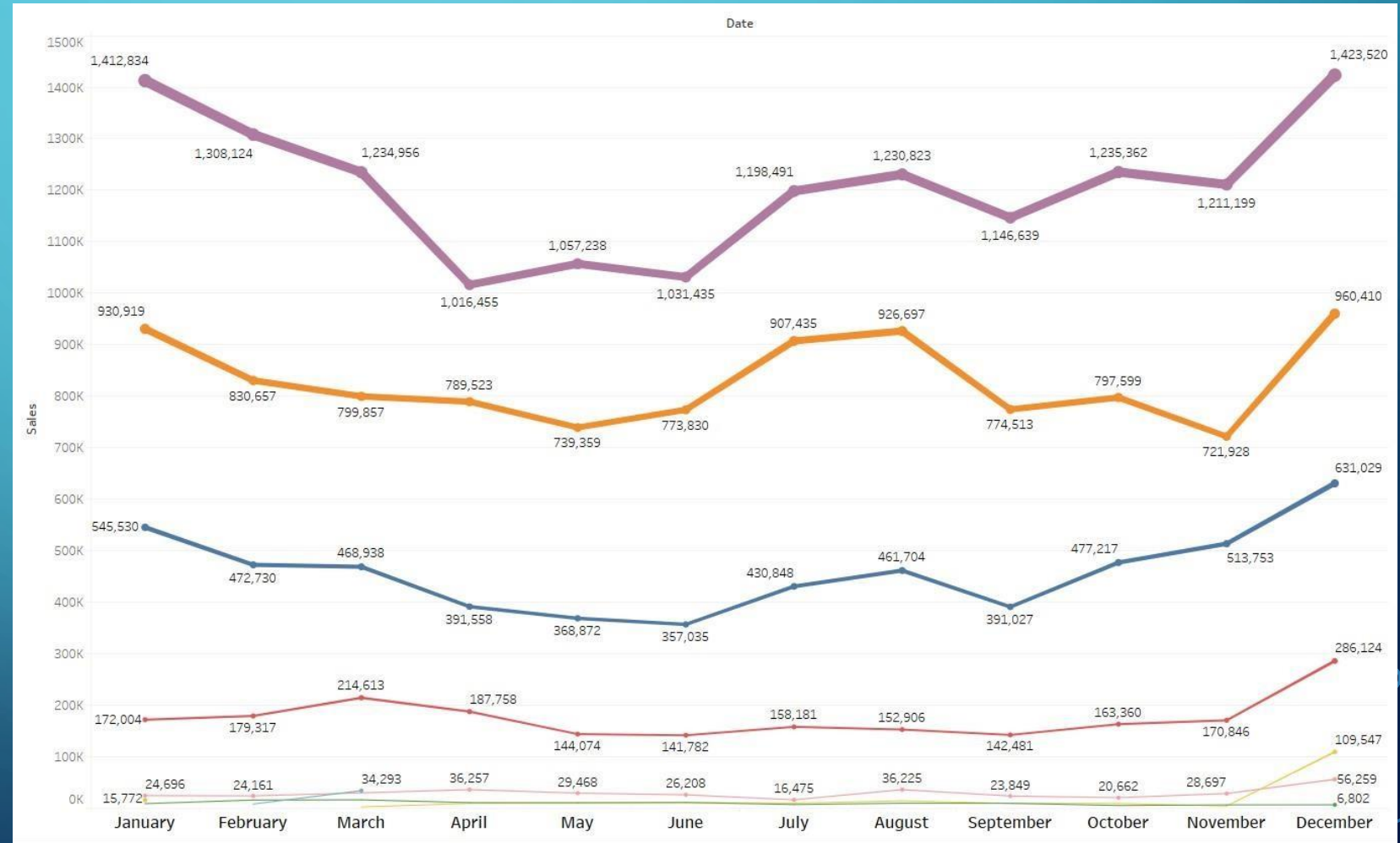
- Each line denotes the different categories which are top selling in the given hours.
- Tobacco is the top selling Category at all hours.
- Food and Beverage are sold much less
- Liquor continues to see a rise post 10:00pm.
- Merchandise, Miscellaneous and Wines seem to be at an all time low .
- Highest Footfall is seen between 4:00am and 12:00AM

Top Hours of the Day for each Category



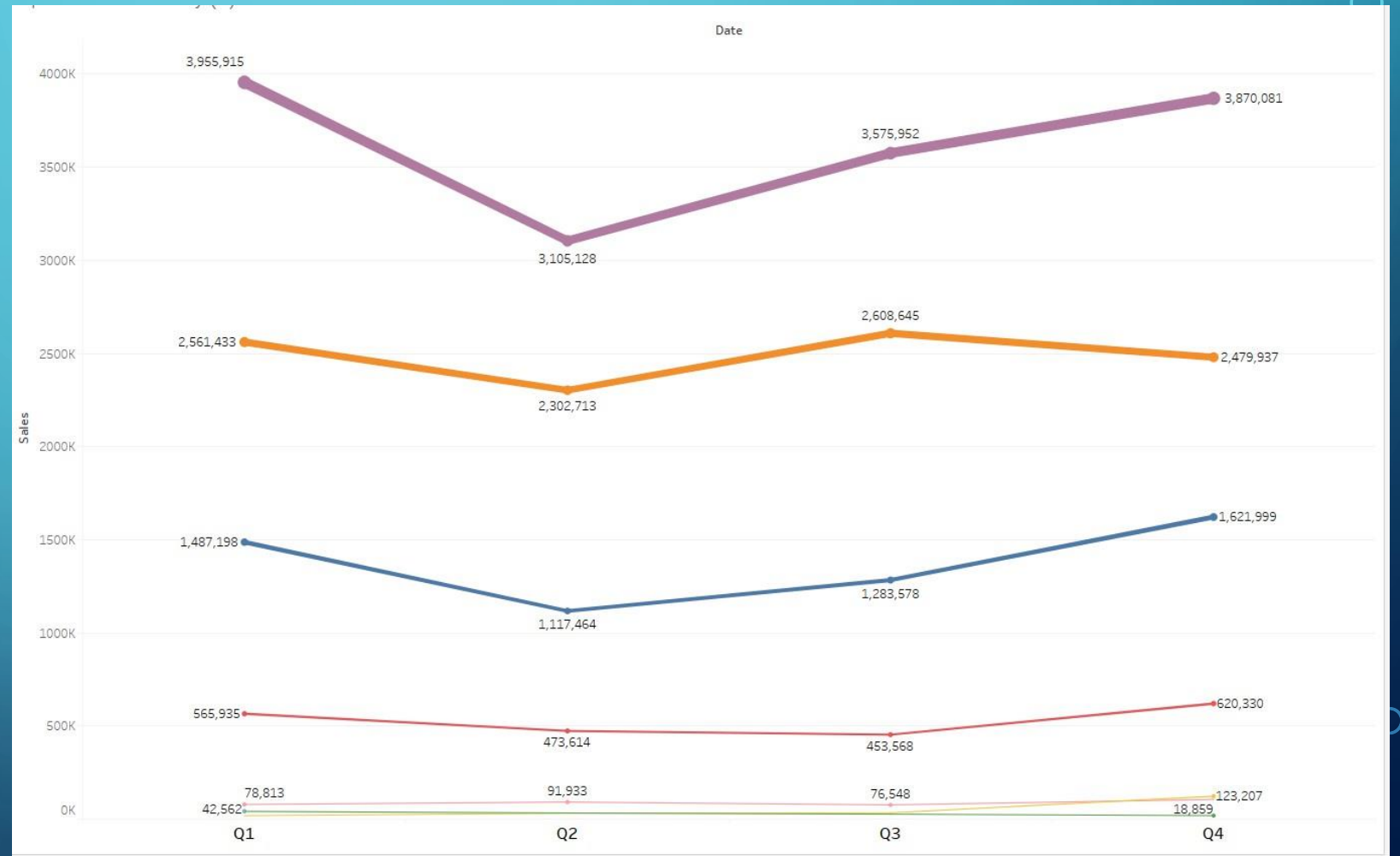
MONTHLY TRENDS FOR CATEGORIES

- Monthly Trends show that Tobacco, Food and Beverage are mostly seen with an upward trend after September, and continue to rise until March.
- This could be due to Holiday Season, therefore increase in footfall.
- March to July the trends are almost stable for tobacco but continues for Food and other categories.
- Tobacco, Food, beverage, all Categories seem to show an upward trend during the month of December, including Miscellaneous (Yellow).
- However, Merchandise and Wines seem to be low.



QUARTERLY TRENDS

- Q1, Q3, Q4 continue to rise, an upward trend is observed.
- In Q2 there is a dip as it is a non-holiday season.



INFERENCES

- ▶ Tobacco, Food and Beverages are the top Performing Categories and it is highly likely that customers buy them in a combination.
- ▶ Liquor and Tobacco is a new Category, and only 1 product is performing well. The "Beer Hookah". Other items are poorly selling.
- ▶ Tax or Discounts do not seem to affect the customer's purchase when buying Tobacco or Liquor or other items.
- ▶ Discounts are often provided more incase of Food and they are purchased more.
- ▶ Customers Footfall increases in the evening and is the highest during 8pm in the evening and continues at 12:00AM.
- ▶ There are not as many buyers for Liquor, as there are for Tobacco, Food and Beverage.
- ▶ Wines is also a Category which is not sold well, removing this Category could result in saving major Costs.

MENU ANALYSIS

THE MOST COMMONLY OCCURRING FOOD ITEMS

Most Occurring Sets	Confidence	Lift	Support	Recommended Product
RED BULL 2+1, PHILLYCREAM CHEESE &CHILLY PAN, GREAT LAKES SHAKE	1	15.82	0.000128	SAMBUCA
RED BULL 2+1, GREAT LAKES SHAKE, MAGGI NDL ARRABIATA	1	15.82	0.000128	SAMBUCA
QUA MINERAL WATER(500ML), TOMATOLINO	1	8.34	0.000126	NIRVANA HOOKAH SINGLE
N R G HOOKAH	1	3332.4	0.000214	2 RED BULL
QUA MINERAL WATER(1000ML), RED BULL 2+1, GREAT LAKES SHAKE	0.93	14.76	0.0002	SAMBUCA
RED BULL ENERGY DRINK, MASALA CHAI CUTTING, GREAT LAKES SHAKE	0.8	14.064	0.000114	SAMBUCA
B.M.T. PANINI, RED BULL 2+, GREAT LAKES SHAKE	0.8	13.56	0.000171	SAMBUCA
QUA MINERAL WATER(1000ML), B.M.T. PANINI, MAGGI NDL ARRABIATA	0.8	12.65	0.000285	SAMBUCA
SAMBUCA, GREAT LAKES SHAKE, GOOEY CHOCOLATE FUDGE	0.7	16.35	0.0001	QUA MINERAL WATER(1000ML)
QUA MINERAL WATER(1000ML), RED BULL ENERGY DRINK, B.M.T. PANINI, GREAT LAKES SHAKE	0.7	12.3	0.000100026	SAMBUCA
QUA MINERAL WATER(1000ML), RED BULL 2+1, POUTINE WITH FRIES	0.75	11.86	0.000128604	SAMBUCA

FOOD COMBINATIONS

Items with High Lift Values are guarantee products which are bought together as a Combo Meal.

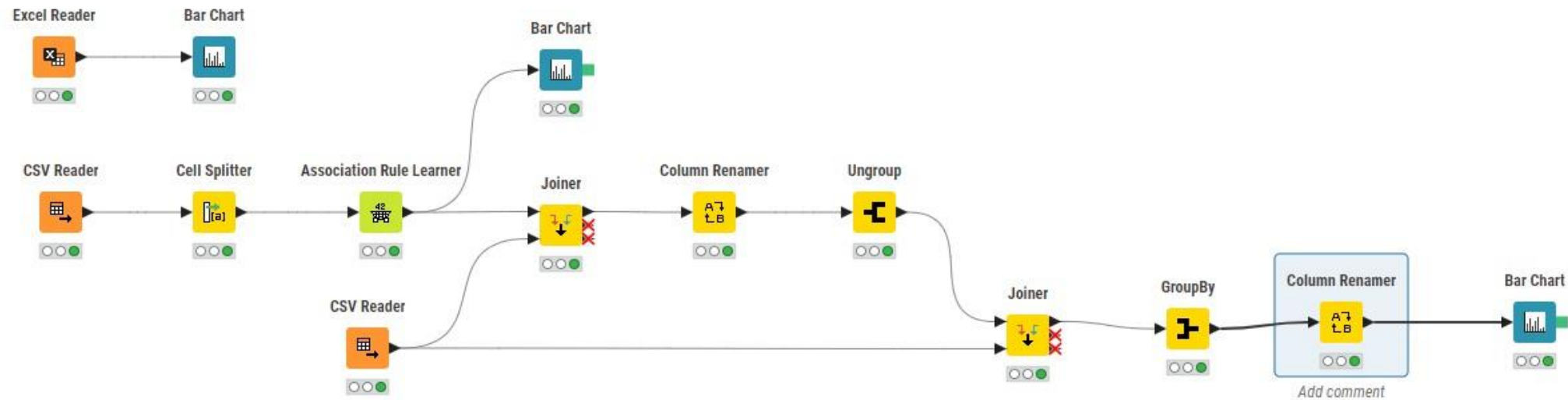
Also Items with high Confidence can also be sold as a great Combo Meal

- ▶ *PASTA ALFREDO VEG, 3 RED BULL with BEER HOOKAH*
- ▶ *N R G HOOKAH with 2 RED BULL*
- ▶ *NIRVANA HOOKAH SINGLE, ADD HERB ROAST CHICKEN with Lemon Infused Char Grilled Veg*
- ▶ *QUA MINERAL WATER(1000ML), COTTAGE CHEESE PANINI, POUTINE WITH FRIES with BMT Panini*

MARKET BASKET ASSOCIATION

- ▶ Market Basket Analysis (MBA), or Association Rule Mining, is a data mining technique used to discover interesting patterns, relationships, or associations among a set of items in large datasets.
- ▶ The goal is to uncover rules that indicate the likelihood of certain items being purchased together.
- ▶ The primary concept behind Market Basket Analysis is the identification of associations or relationships between different products based on historical transaction data. The most common representation of these associations is through association rules.
- ▶ This association is required in our Dataset to check the Buying Combination of items that consumers purchase. If we club the top selling Items together, this would increase in Recency, Frequency of a Customer, as well as help in increasing revenues for the company.

KNIME WORKFLOW TO ARRIVE AT ASSOCIATIONS



Threshold Values: Confidence: 0.50%, Support: 0.0001. Lift Values: 1 to 3,332.476

ASSOCIATIONS

► 1: Output Table  Flow Variables

Rows: 94 | Columns: 13

Table Statistics

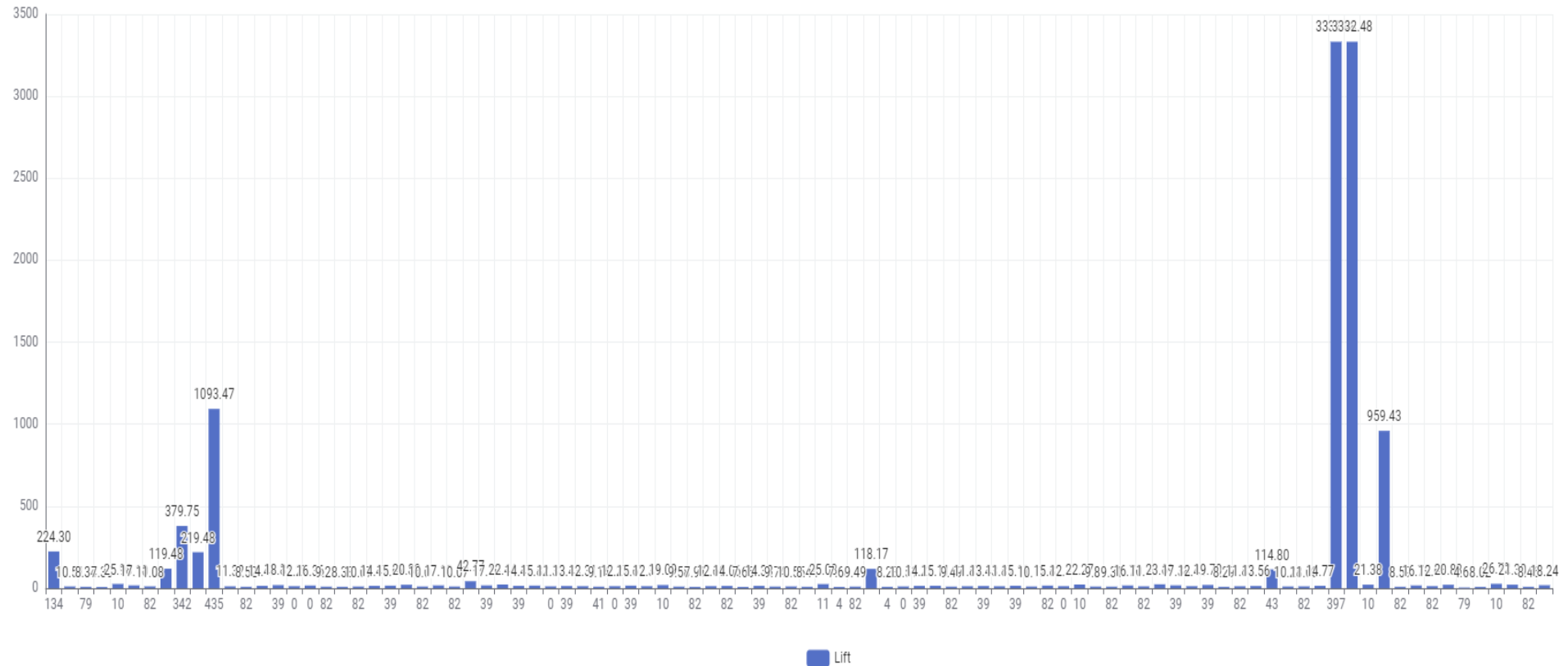
Items Set	Support Number (double)	Confidence Number (double)	Lift Number (double)	implies String	Basket_Item_... String	Rec_Prds_Na... String	Rec_Disc Number (double)	Rec_Price Number (double)	Basket_Product_Name String	
[513,60,41]	0	1	15.822	<---	513, 60, 41	SAMBUCA	0	462	RED BULL 2+1, PHILLYCREAM CHEESE &CHILLY PAN, GREAT LAKES SHAKE	0
[0,513,41]	0	0.933	14.767	<---	0, 513, 41	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), RED BULL 2+1, GREAT LAKES SHAKE	0
[11,2,41]	0	0.889	14.064	<---	11, 2, 41	SAMBUCA	0	462	RED BULL ENERGY DRINK, MASALA CHAI CUTTING, GREAT LAKES SHAKE	0
[9,362]	0	0.875	7.299	<---	9, 362	NIRVANA HOOKAH...	0	323.4	MIAMI MELONS, 3 RED BULL	0
[39,513,41]	0	0.857	13.562	<---	39, 513, 41	SAMBUCA	0	462	B.M.T. PANINI, RED BULL 2+1, GREAT LAKES SHAKE	0
[104]	0.001	0.8	26.723	<---	104	CAFFE LATTE	0	86.63	ADD VANILLA FLAVOUR	0
[0,39,52]	0	0.8	12.658	<---	0, 39, 52	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), B.M.T. PANINI, MAGGI NDL ARRABIATA	0
[0,513,60]	0	0.8	12.658	<---	0, 513, 60	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), RED BULL 2+1, PHILLYCREAM CHEESE &CHILLY PAN	0
[11,39,52]	0	0.786	12.432	<---	11, 39, 52	SAMBUCA	0	462	RED BULL ENERGY DRINK, B.M.T. PANINI, MAGGI NDL ARRABIATA	0
[82,41,32]	0	0.778	16.355	<---	82, 41, 32	QUA MINERAL WA...	0	61.88	SAMBUCA, GREAT LAKES SHAKE, GOOEY CHOCOLATE FUDGE	0
[26,39]	0	0.778	25.981	<---	26, 39	CAFFE LATTE	0	86.63	ADD HAZELNUT FLAVOUR, B.M.T. PANINI	0
[342,435]	0	0.778	219.478	<---	342, 435	3 RED BULL	0	371.25	PASTA ALFREDO VEG, BEER HOOKAH	0
[82,83,52]	0	0.778	20.895	<---	82, 83, 52	B.M.T. PANINI	0	259.88	SAMBUCA, N R G HOOKAH, MAGGI NDL ARRABIATA	0
[0,11,39,...]	0	0.778	12.306	<---	0, 11, 39, 41	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), RED BULL ENERGY DRINK, B.M.T. PANINI, GREAT LAKES SHAKE	0
[0,39,41,...]	0	0.778	12.306	<---	0, 39, 41, 52	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), B.M.T. PANINI, GREAT LAKES SHAKE, MAGGI NDL ARRABIATA	0
[0,513,54]	0	0.75	11.867	<---	0, 513, 54	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), RED BULL 2+1, POUTINE WITH FRIES	0
[11,39,7]	0	0.75	11.867	<---	11, 39, 7	SAMBUCA	0	462	RED BULL ENERGY DRINK, B.M.T. PANINI, LEMON ICED TEA	0
[0,41,52]	0	0.75	11.867	<---	0, 41, 52	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), GREAT LAKES SHAKE, MAGGI NDL ARRABIATA	0
[82,52,54]	0	0.733	19.701	<---	82, 52, 54	B.M.T. PANINI	0	259.88	SAMBUCA, MAGGI NDL ARRABIATA, POUTINE WITH FRIES	0
[397]	0	0.714	3,332.476	<---	397	N R G HOOKAH	0	594	2 RED BULL	0
[0,39,513]	0	0.714	11.302	<---	0, 39, 513	SAMBUCA	0	462	QUA MINERAL WATER(1000ML), B.M.T. PANINI, RED BULL 2+1	0
[435,362]	0	0.7	379.747	<---	435, 362	PASTA ALFREDO V...	0	247.5	BEER HOOKAH, 3 RED BULL	0
[0,66,54]	0	0.7	18.805	<---	0, 66, 54	B.M.T. PANINI	0	259.88	QUA MINERAL WATER(1000ML), COTTAGE CHEESE PANINI, POUTINE WITH FRIES	0

THRESHOLD VALUES: CONFIDENCE, SUPPORT, LIFT

- ▶ Support Value is chosen as 0.0001 to include maximum number of customers, and increase the probability of finding associations and Sets.
- ▶ Confidence value is chosen as 0.50. Keeping Low Confidence values ensures that we get a wide variety of sets and associations that people purchase. Keeping really high confidence limits the scope of diverse purchasing that customers do.
- ▶ Therefore keeping a low confidence value includes sets and rules of customers who buy anywhere ranging 2 items to 4 or 6 items in a single purchase.
- ▶ Lift values range from 1 to 3332. Top 2 items having Lift values 3332 are NRG Hookah and Nirvana Hookah Single.
- ▶ Other products have a Lift of 1093 which is Pasta alfredo and 3 Red Bull. These products are definitely bought together, hence these are having high Lift values.
- ▶ In the event when a customer is purchasing products, the item having high Lift value is recommended for purchase. Therefore “Lemon Infused Char Grilled Veg” is definitely bought.
- ▶ Other products such as Poutine with Fries, BMT Panini, Sambuca are recommended due to high Confidence values.

SUPPORT AND CONFIDENCE AND LIFT VALUES

Bar Chart



POSSIBLE COMBINATIONS

▶ *Lucrative offers:*

Combo Meal of Poutine with Fries, Café LATTE, QUA Mineral Water (1000ml) Free

Saigon Noodles with Chicken, Café Latte free

Sambuca Free with a purchase of BMT Panini, SAMBUCA, GREAT LAKES SHAKE, MAGGI NDL ARRABIATA

Coffee Mug Free with a purchase of Spinach Tempura, Stuffed Mushroom.

▶ *Discounts*

Discounts on purchase of Cappuccino, Café Latte and 3 Red Bull.

RECOMMENDATIONS

- ▶ The Liquor & Tobacco Category seems to be a mix of Liquor and Tobacco flavours.
- ▶ This would work much better if existing products in the Liquor and Tobacco category are added into this category, and old poor performing products are scrapped.
- ▶ Top Performing Categories are Food, Beverage and Tobacco hence Combos should be made using these categories for the customers.
- ▶ Discounts are offered very less in case of Beverages, these should be
 - increased to push sales in this Category.
- ▶ Discounts on QUA Mineral water should be scrapped, it should be provided as a part of Combo Meal when 2 or more Food and Beverage items are purchased.

LINK TO TABLEAU DASHBOARD

 [Cafe Data MRA Anil Kumar | Tableau Public](#)