

MRA Extended Project Report

Submitted to



by

Subhadeep Seal

in Partial Fulfillment of PGP DSBA



Part 1 - Exploratory Analysis

1.1

Exploratory Data Analysis

We are exploring a dataset of transactions at a particular cafe - for a period of 1 year, from 1 April 2010 to 31 March 2011

Each record has 9 features: which includes a Bill Number, the Item that was ordered, the Category under which the said item is classified, date and time stamps of the order, along with the Price, Quantity, Tax and the Total amount charged for the item.

The dataset has 1,45,830 rows, accounting for 69,982 orders (identified by Unique Bill Numbers)

The Total Revenue from sales at this cafe, for the 1 year period, is a little over INR 3.28 Crore (INR 3,28,05,895.27.) We assume the currency is Indian Rupees or INR for this exercise.

There are 680 duplicate records in the dataset. We will, however, not delete these, and assume that they are legitimate entries for the purpose of this exercise.

There are no missing values in the dataset.



1.1

Orders at the cafe

By grouping the data on the basis on Unique Bill Numbers, we can study the orders placed at the cafe.

Number of orders placed at the cafe during the year: 69,982

Median Order Value (Bill Total): INR 391

Mean Order Value (Bill Total): INR 468.78

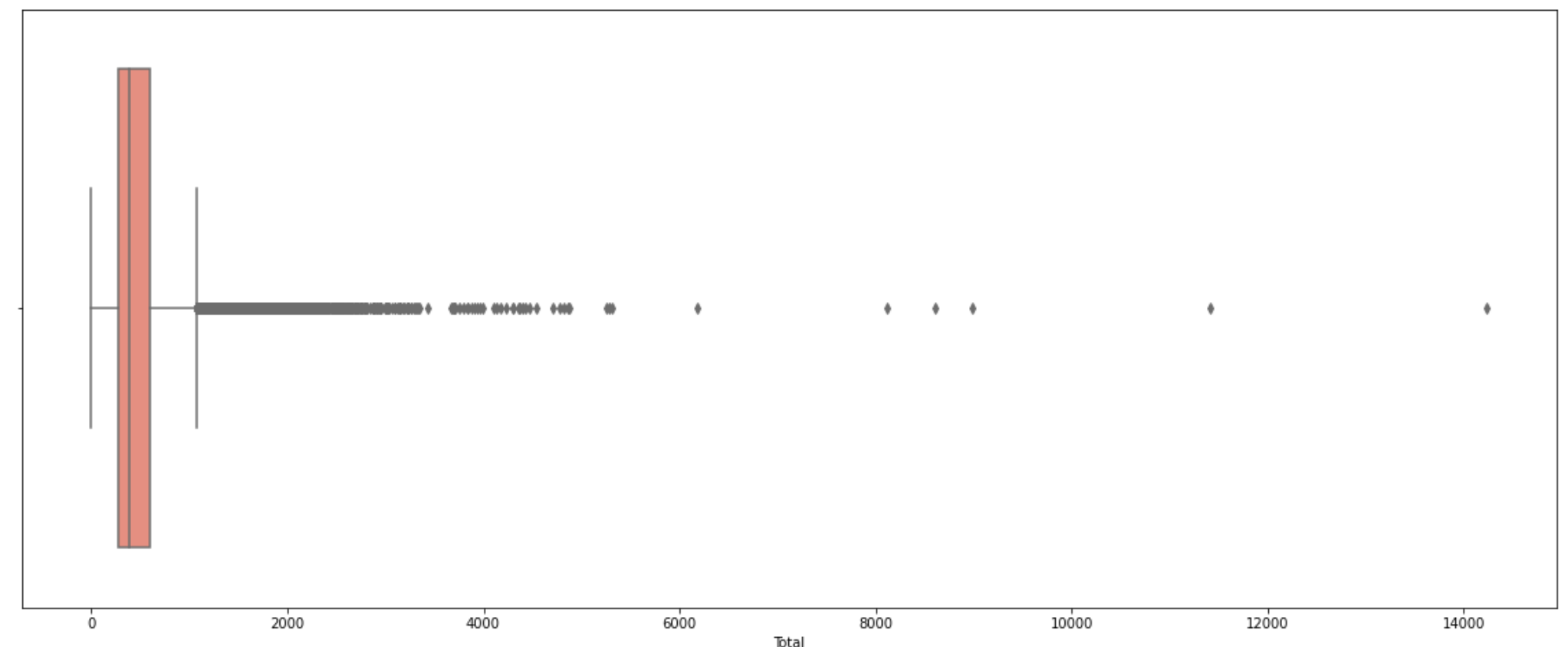
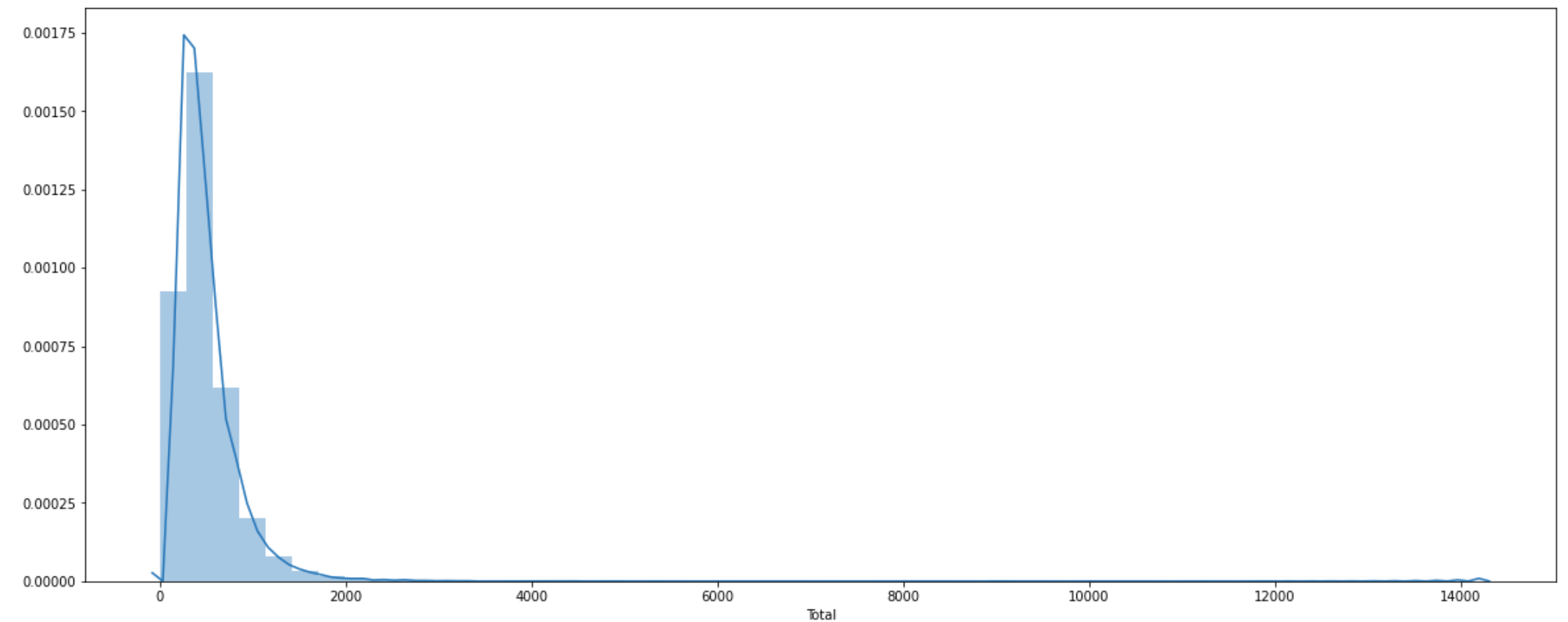
The notable difference in the mean and median values suggests a skewed distribution.

The featured histogram and box plot for order values (Total per Bill number), both corroborate the right skew. Typically order values are concentrated between INR 250 to INR 600. But there is a significant number of high value orders, which constitute outliers.

One can see the large range of values for orders placed, from INR 1 to INR 14,231.

It is also observed that the typical order consists of 1 to 3 units, 2 being the median quantity per order.

In relatively rare instances do quantity of items in an order cross 5 units. Although in exceptional cases, outlier orders that have over 50/60 units of Items.



1.1

Outlier Orders at the cafe

We take a broad look at the high value orders or outliers

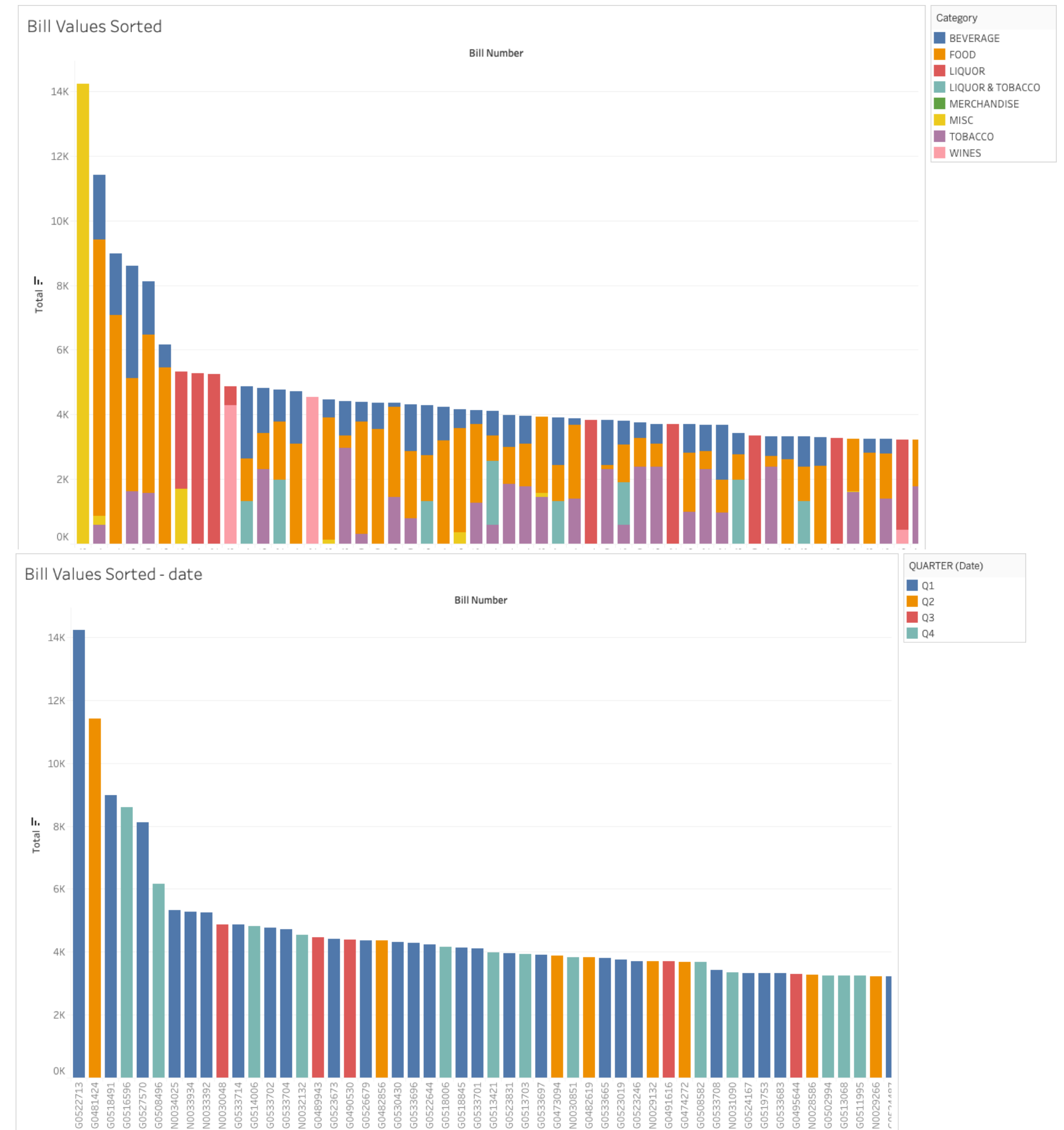
On the right are bar plots of Bills, arranged in descending order of value (Total Amount)

The top plot features the constituent Categories of each Bill.
Food, along with non-alcoholic beverage, appears to be the predominant combination of the outlier orders.

The bottom plot colour codes Bills based on the quarter in which the order was placed.

Q1 (Jan 2011 - Mar 2011) appears to have the orders with the highest values, followed by Q4 (Oct 2010 - Dec 2010)

In the slides ahead, we will compare these to the general pattern.



1.1

Menu Items

There are 580 unique items in the dataset, that form the cafe’s menu

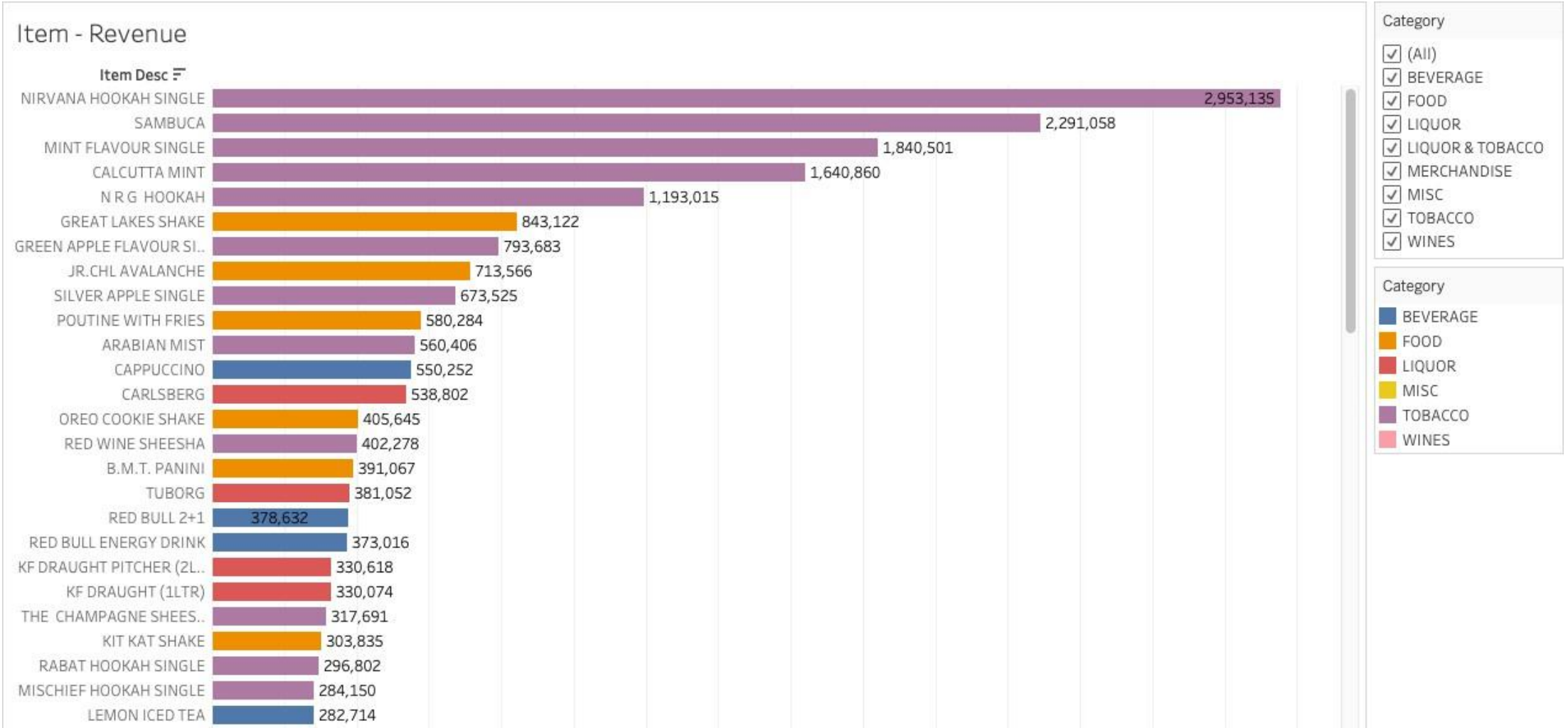
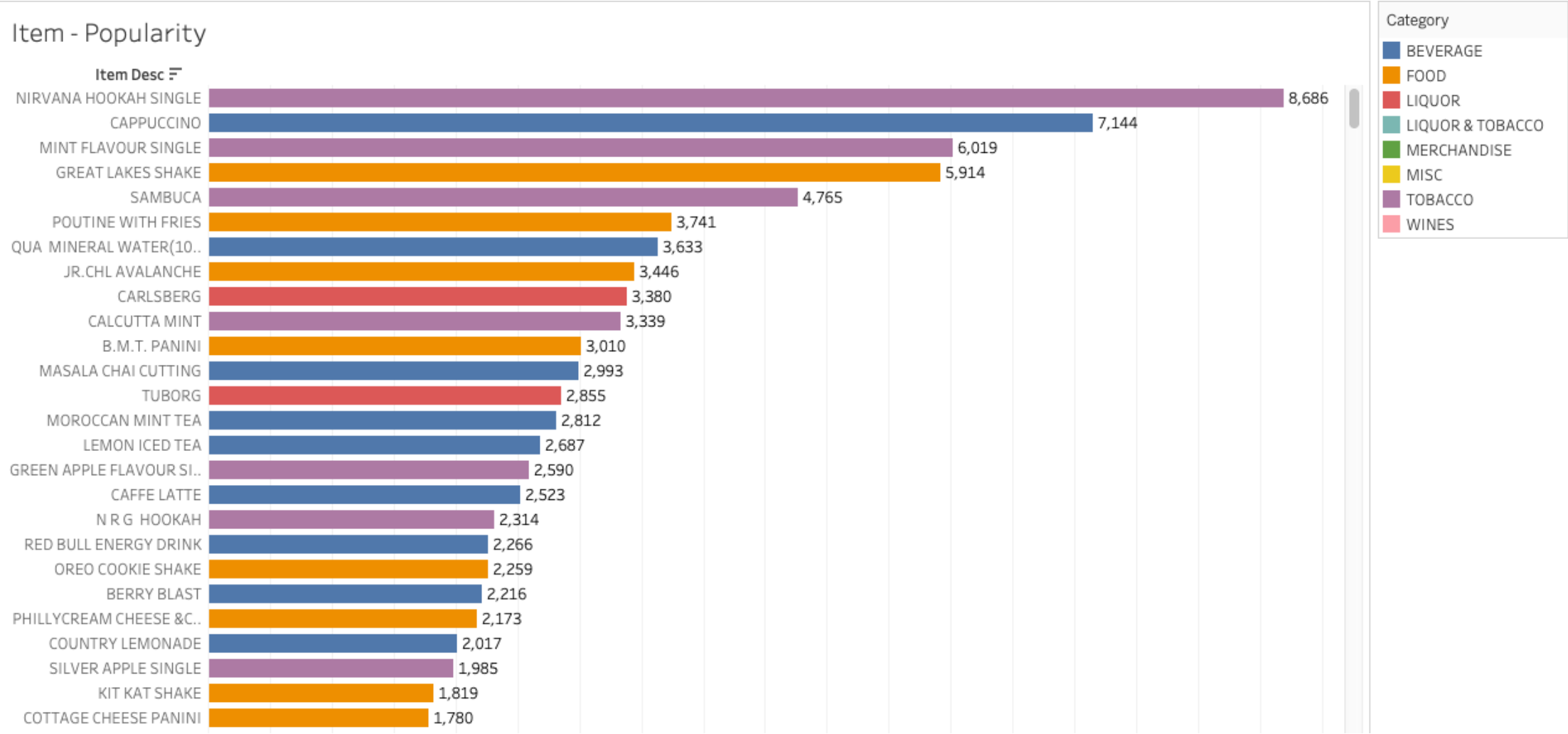
On the right are the top ranked items, sorted by units sold (popularity), and by sales proceeds (revenue):

The Nirvana Hookah Single is the most popular item in the menu, with 8,686 units sold over the year

The Nirvana Hookah Single is also highest in the revenue rankings, bringing almost INR 30 lakhs in earnings. That’s almost 10% of the cafe’s annual earnings coming from one product item.

What’s also noticeable is that the top 5 items bringing the highest revenues, belong to the Tobacco category

Another interesting observation is that earnings from the top 10% of the products in the menu (58 items) constitute over 77% of the cafe’s annual revenue.



1.1

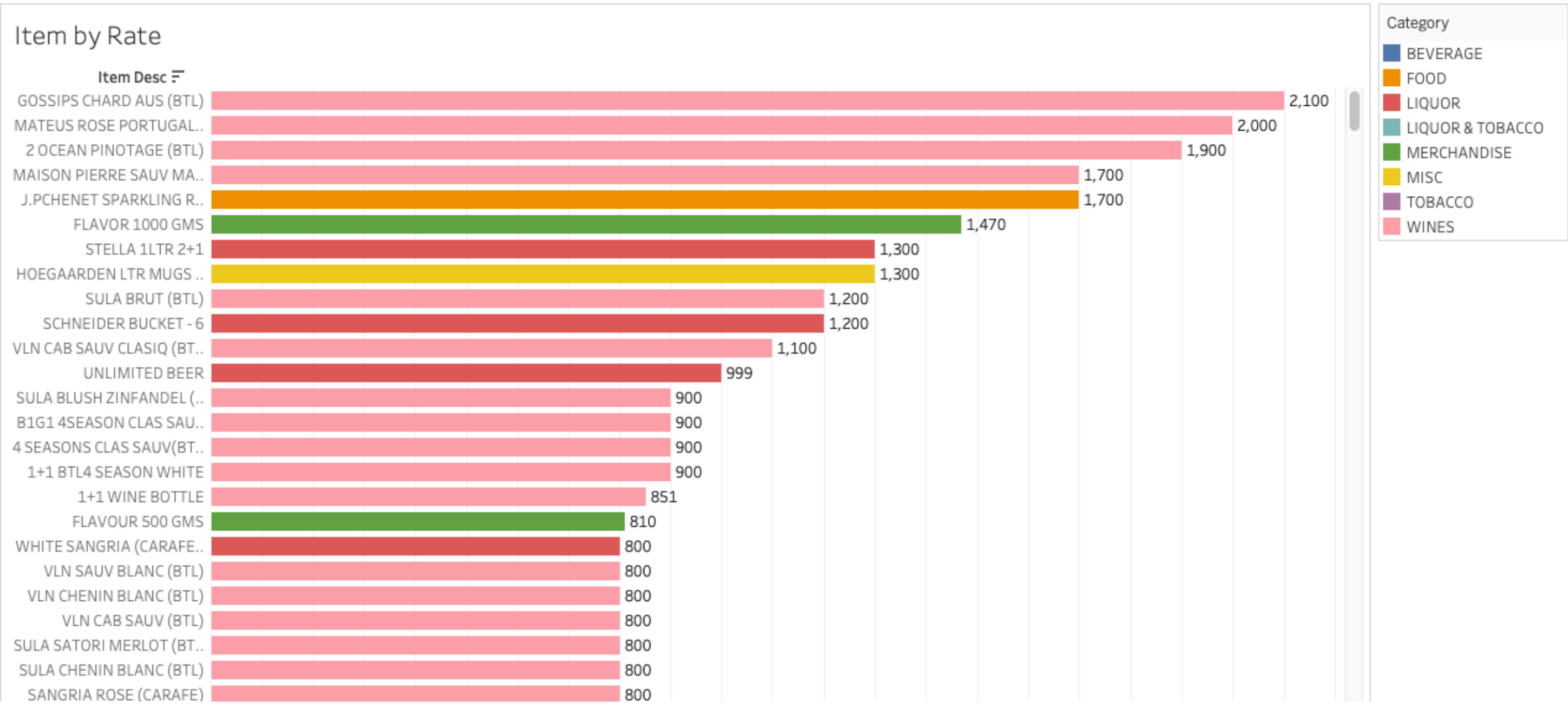
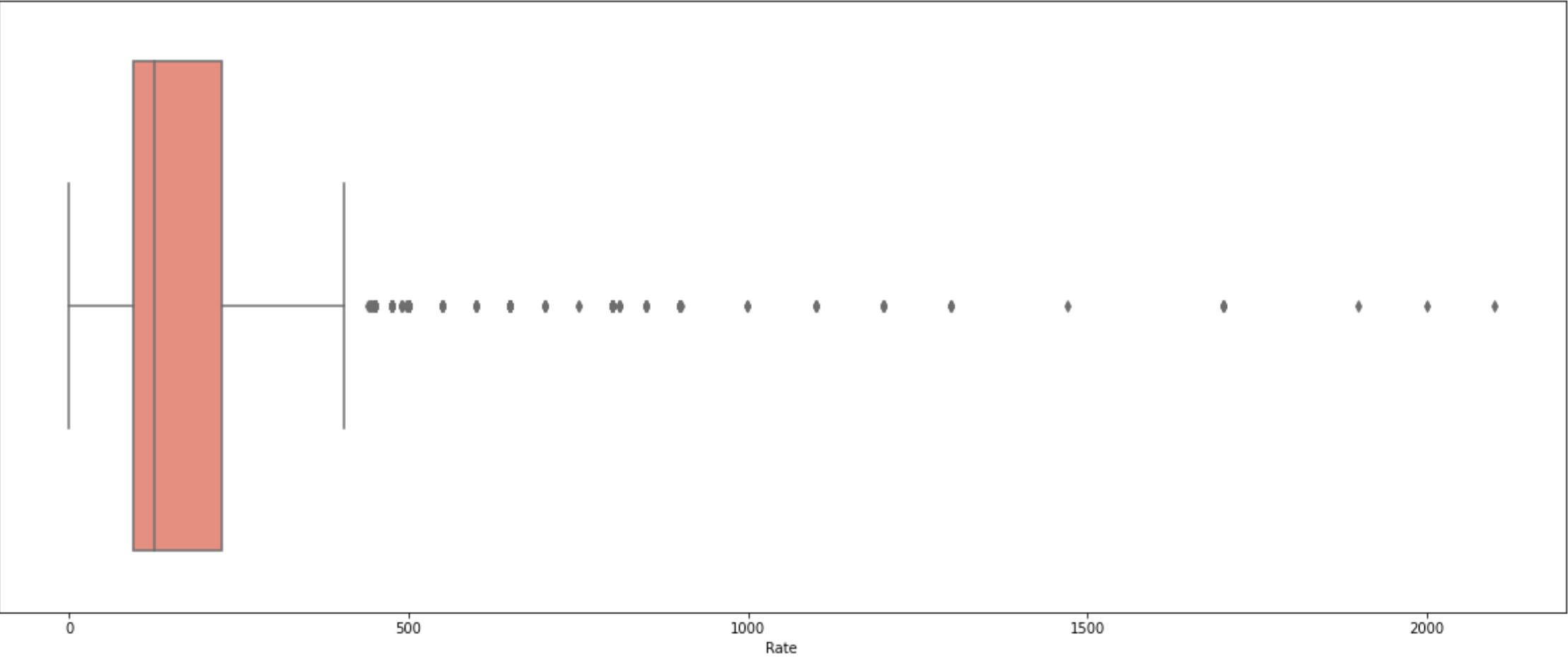
Menu Items - Rate

Of the Rate of Items listed in the menu, one can see the range of price points from INR 12 to INR 2100.

Median Price of items (Rate): INR 125
Mean Price of items (Rate): INR 161.78

The featured box plot for the column Rate illustrates the right skew of the distribution. Typically the price points are concentrated between INR 95 to INR 225. But there are a number of products over ~ INR 500, which constitute outliers.

The adjacent bar plot lists the most expensive products on the menu, which is markedly dominated by Wines.



1.1

Product Categories

The items on the cafe menu are grouped in 7 broad categories: Food, Beverage (non-alcoholic drinks), Wine, Liquor (other alcoholic beverages), Tobacco, Merchandise and Miscellaneous.

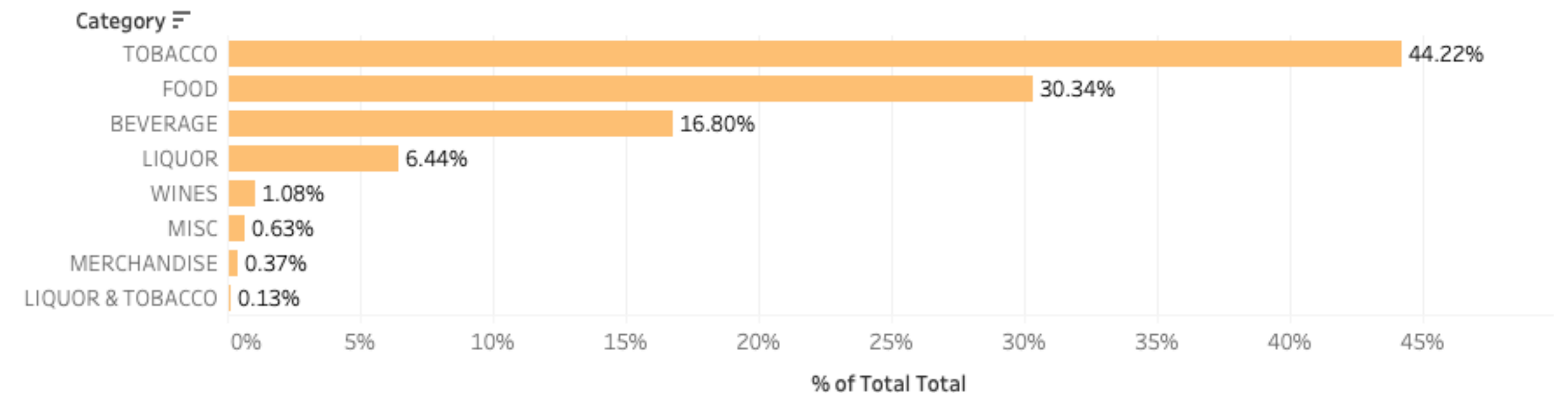
One should note that Miscellaneous appears to comprise mostly misclassified products, which should ideally be listed under any of the other 6 categories.

The adjacent bar charts provide category wise split for quantity sold and revenue.

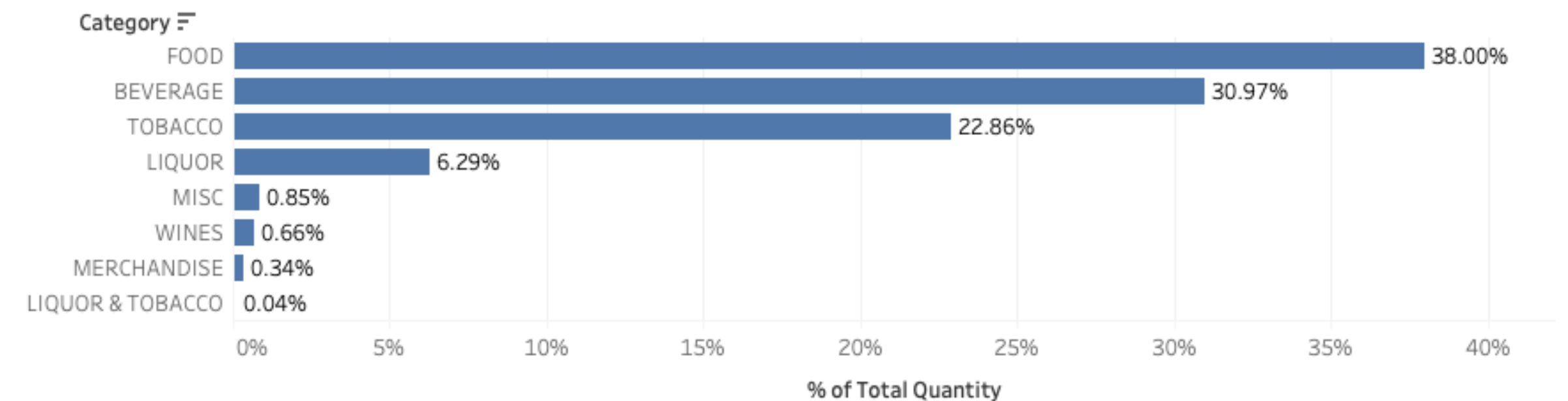
Tobacco as a category accounts for over 44% of the cafe's annual revenues, even as it accounts for only 22% of the orders.

Food is the most popular category, and accounts for the category from which items are most ordered. Beverage accounts for the second most popular category. Often ordered together, they account for over 68% of the ordered items.

Category and Revenue



Category and Quantity Sold



1.1

Sales over Time

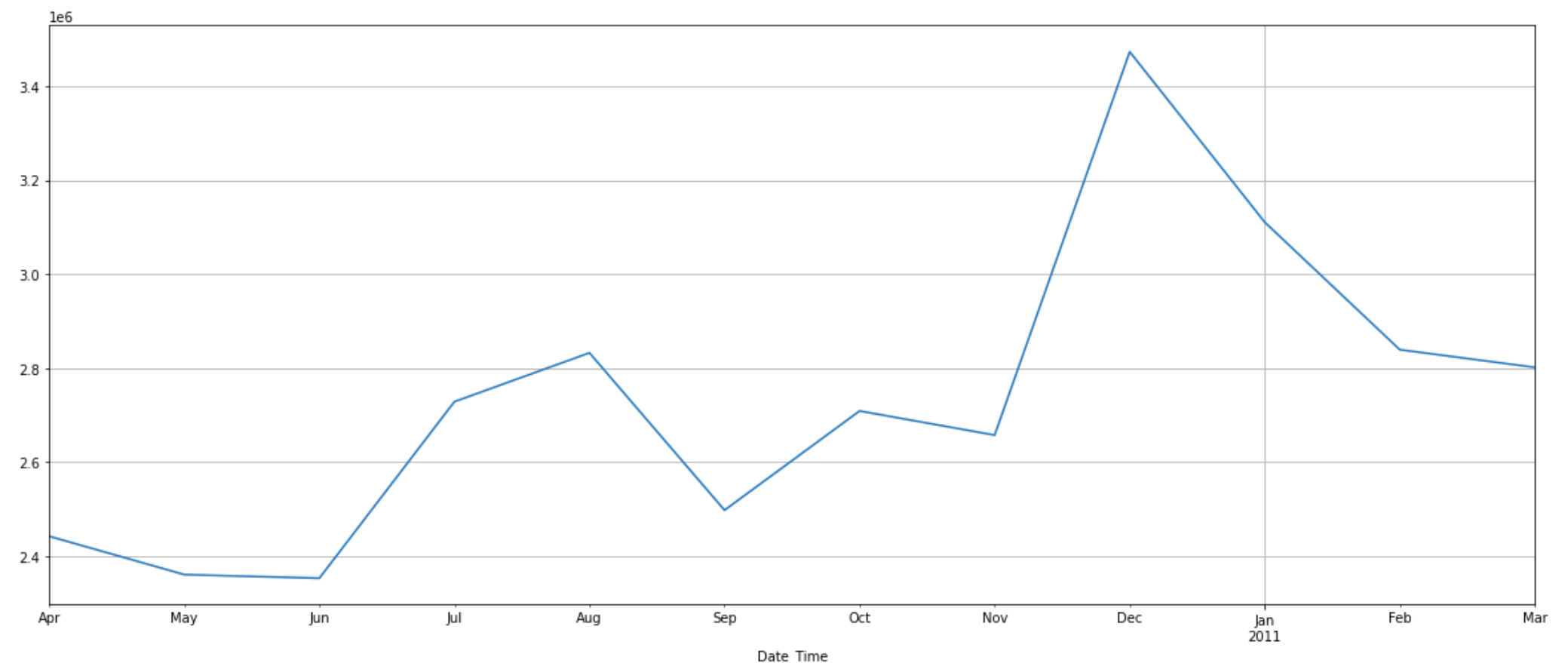
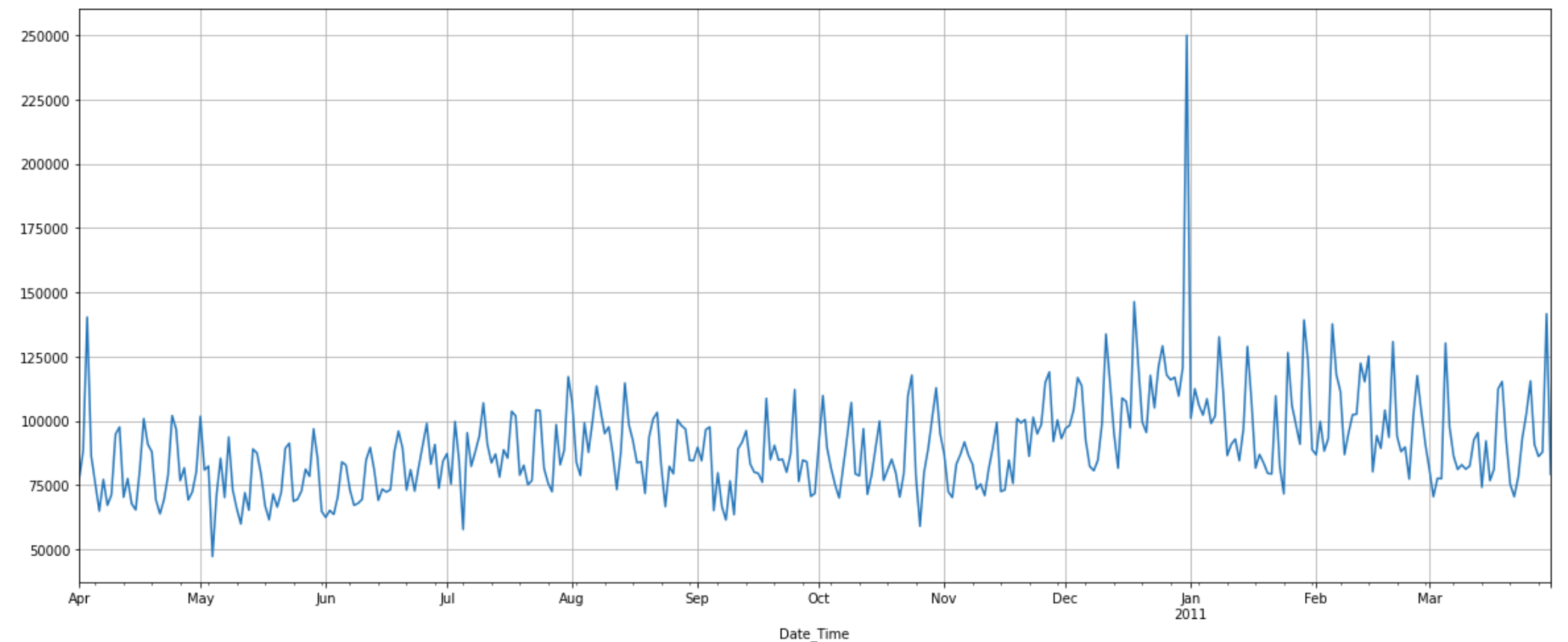
The adjacent plots visualise Revenue generated from sales at the cafe, for a period of 1 year, from 1 April 2010 to 31 March 2011

The plot on the top right depicts sales at the cafe at a daily granular level. There is slight upward trend overall, but is marked by a gentle undulation. There does appear to be a weekly seasonal pattern that sustains across the entire time period.

The average sales at the cafe, per day, is close to INR 90,000, with the IQR ranging roughly from INR 78,000 to INR 1,00,000. The lowest daily sales was recorded on May 4, 2010 at INR 47,177. An exponential spike is seen on New Year's eve (Dec 31 2010) with sales amounting to INR 2.5 lakhs.

The plot below is the monthly revenue plot. December 2010 is the highest earning month by far with close to INR 35 lakhs in revenues, followed by January 2011 with INR 31 lakhs in sales revenue. August 2010, and February and March 2011 are also strong months, with sales around INR 28 lakhs.

The April-May-June (AMJ 2010) quarter was the weakest in terms of earnings. The Oct-Nov-Dec (OND 2010) quarter saw the most earnings, closely followed by Jan-Feb-Mar 2011.



1.1

Summary

We reviewed transaction data of a particular cafe for a period of 1 year, from 1 April 2010 to 31 March 2011

There were a net 69,982 orders placed, and total revenue earned during the period was around INR 3.28 Crore

The Median Order Value (Bill Total) is INR 391. And median quantity of item/s per order is 2. Improving on both of these counts can see significant growth in business. The Mean Order Value (Bill Total) is INR 468.78, suggesting a skewed distribution owing to a notable number of high value orders.

There are 580 unique items in the dataset. But over 77% of the cafe's annual revenue earnings come from 10% of the menu, ie from 58 items.

The Top 5 items bringing the highest revenues belong to the Tobacco category

Tobacco as a category accounts for over 44% of the cafe's annual revenues, even as it accounts for only 22% of the orders.

Food is the most popular category in terms of quantity sold. Often ordered together, Food and Beverage account for over 68% of the ordered items.

The average sales at the cafe, per day, is close to INR 90,000, with the IQR ranging roughly from INR 78,000 to INR 1,00,000.

The Oct-Nov-Dec (OND 2010) quarter saw the most earnings, closely followed by Jan-Feb-Mar 2011. The Apr-May-Jun 2010 quarter saw the lowest sales levels.

1.2

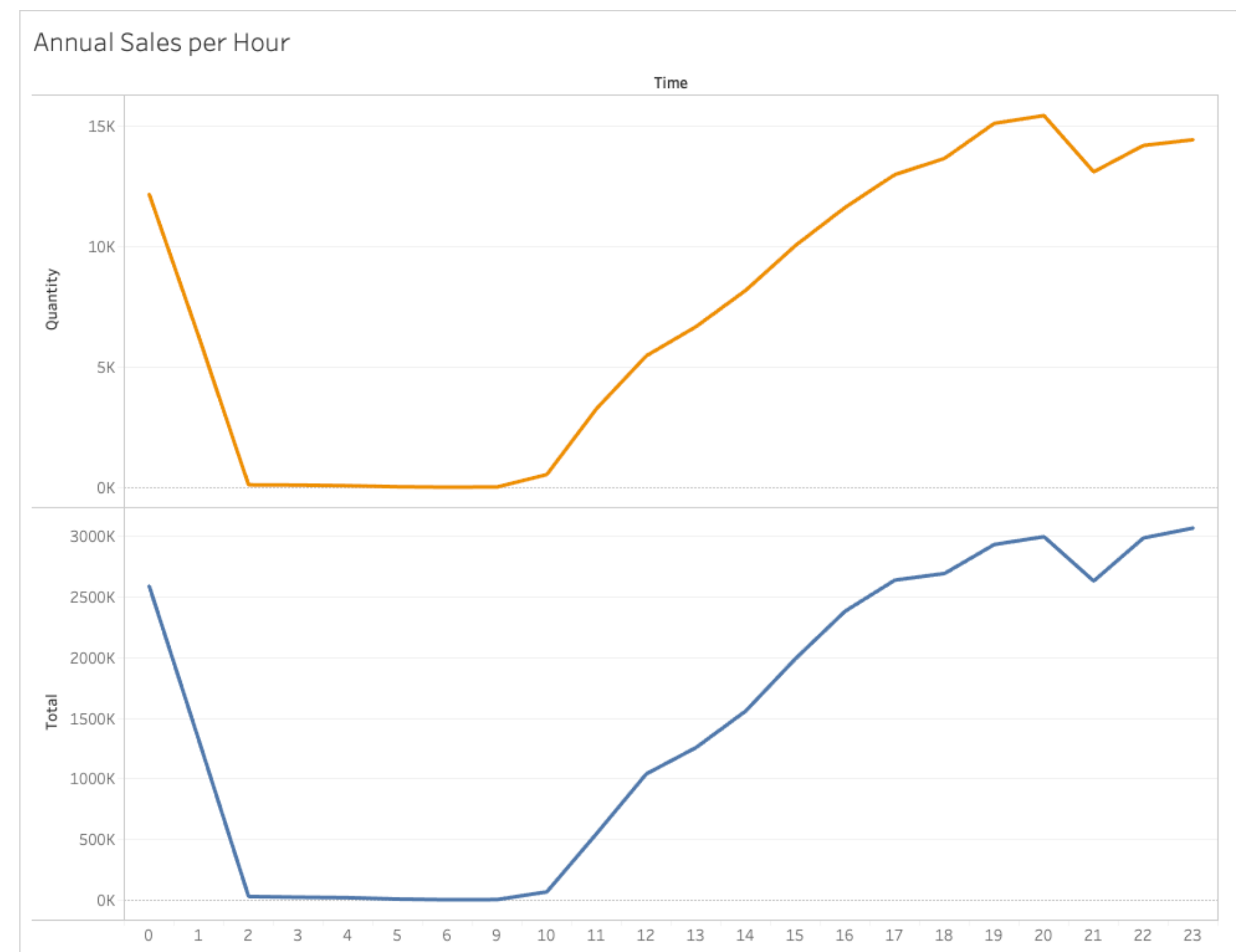
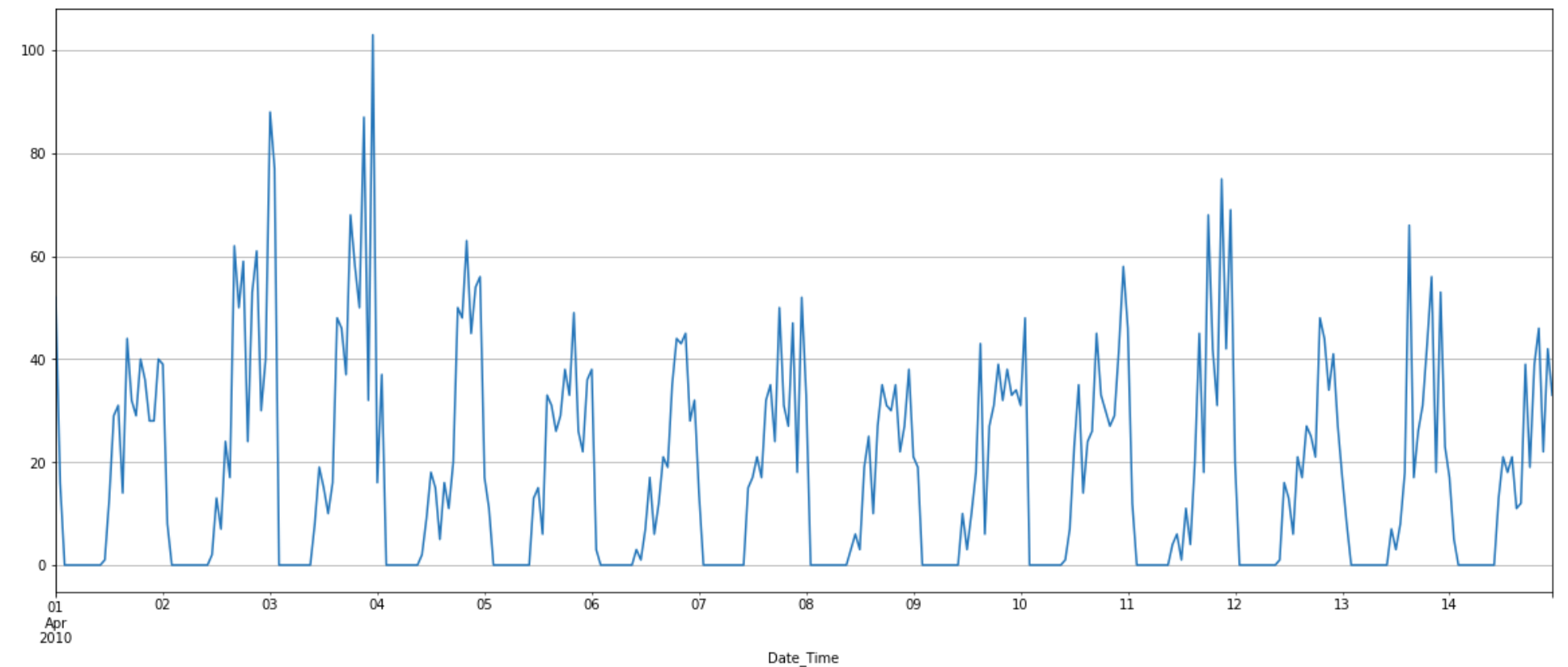
Trends at different times of day

The plot on top depicts hourly sales for a micro period of a fortnight, from April 1 2010 to April 15. A noticeable pattern in seen in traffic and sales at the cafe, at a day-to-day level.

If we are to generalise the pattern, by plotting the Annual Sales over the hour of each day, we get the plot at the bottom.

On a typical day, there is a steady rise in revenue from around 10 AM when the cafe appears to open, to around 8 PM at night, where it appears to peak. After 8 PM, sales appear to dip slightly before rising again until 11 PM, the second peak.

The plot suggests that the cafe is in operation from around 9 AM to 2 AM.



1.2

Trends at different times of day

The plots on the right depict Sales (in terms of units sold and revenue) across different hours of the day, split per Category.

In terms of quantities sold, Food is the most popular category all through the day, followed by Beverage

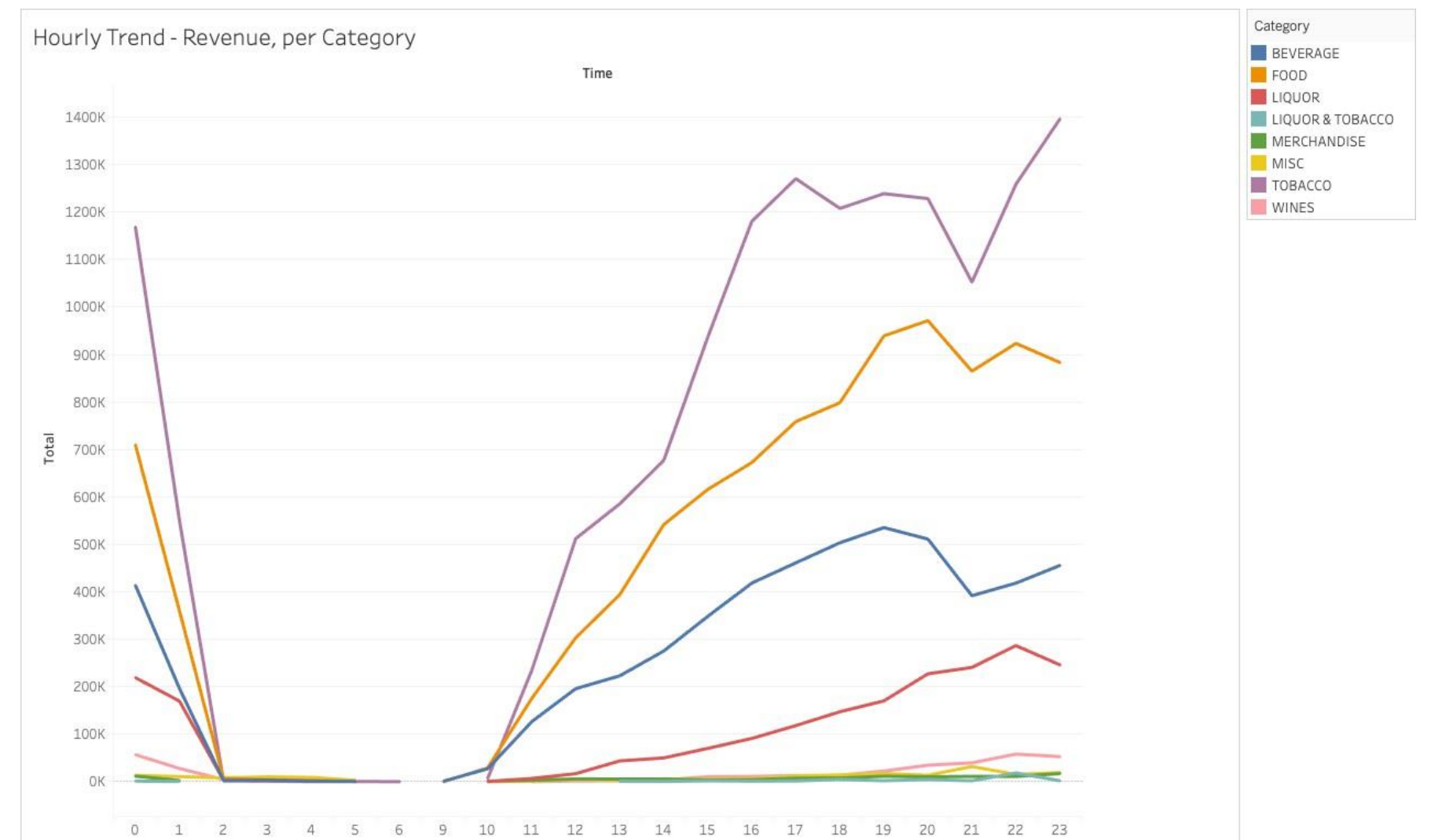
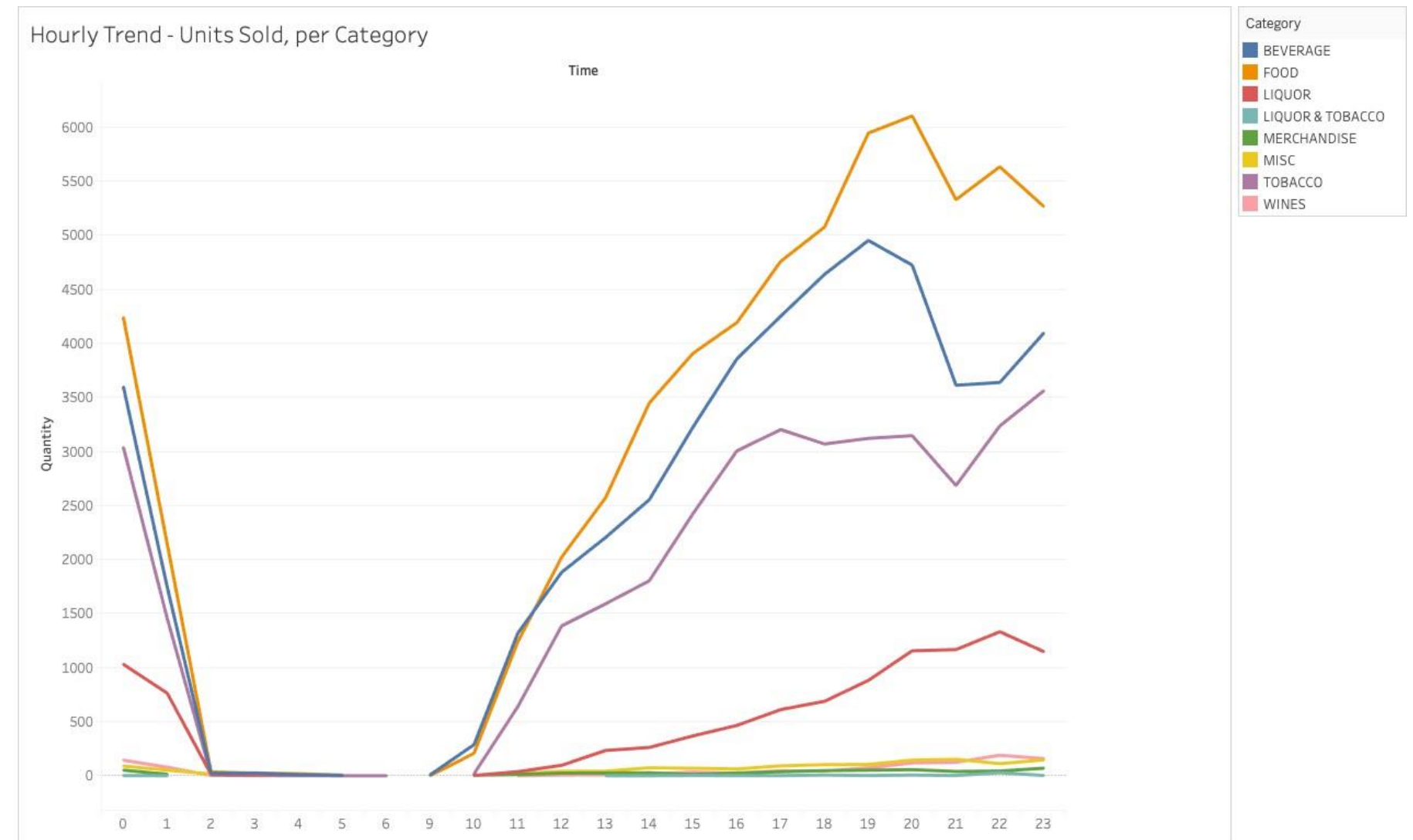
From 9 AM to 12 noon, Beverage and Food curves overlap and are equally popular indicating that they might be ordered together in breakfast hours

Orders across all categories steadily rise from morning through the afternoon. Food orders peak at 8 PM, whereas Beverage orders peak around 7 PM

From a revenue standpoint, Tobacco is the clear winner. It is the highest earning Category all through the day. Tobacco Sales peak at 11 PM, with another lower peak at 4 PM

Liquor Sales peak at 10 PM

From an overall Sales perspective, 8 PM and 11 PM are the most busy and lucrative time-slots respectively.



1.2

Trends over Weekdays

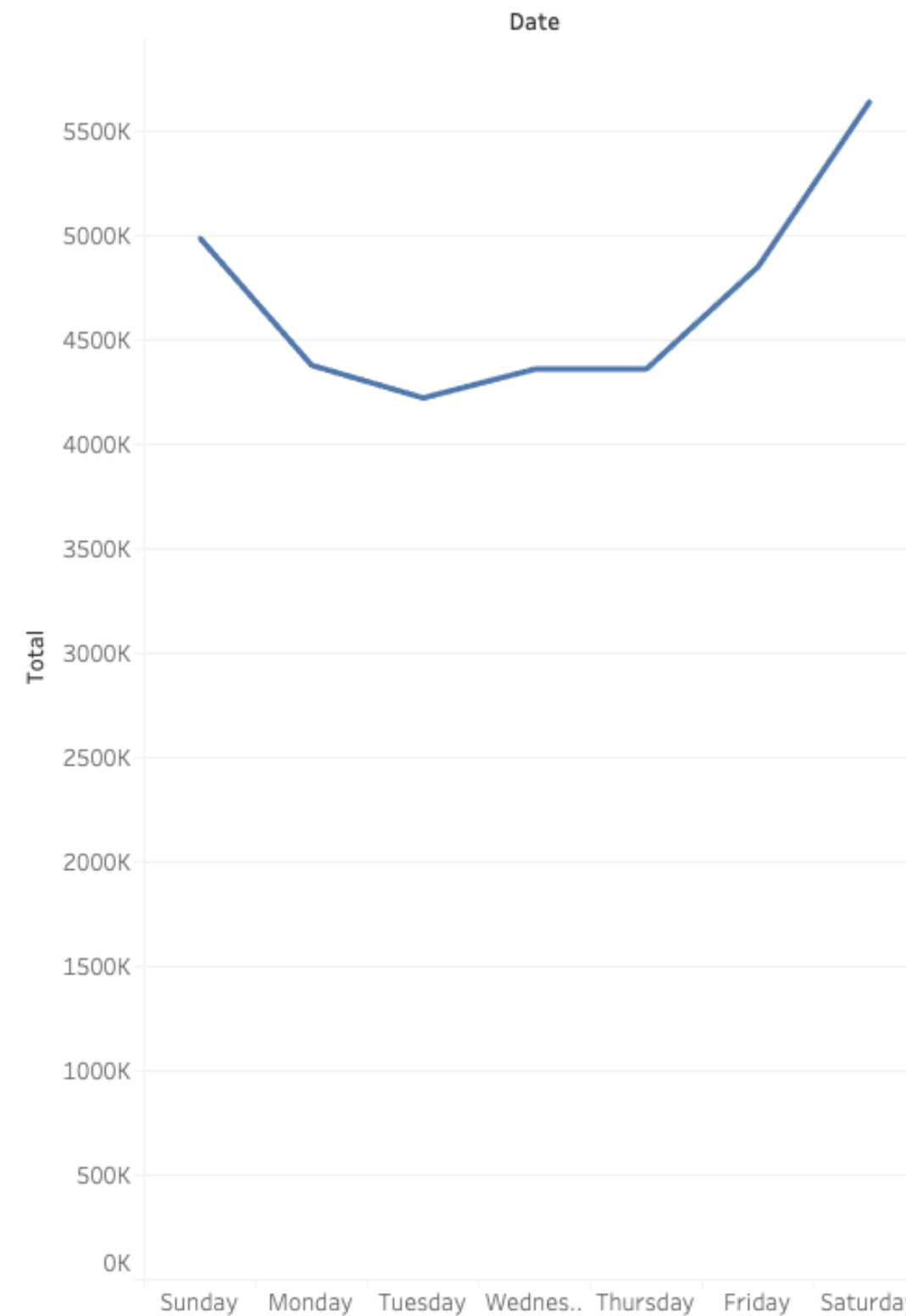
The adjacent plots depict Annual Sales, consolidated and per category, aggregated by weekdays.

Saturday is typically the busiest day, and most lucrative from a revenue standpoint.

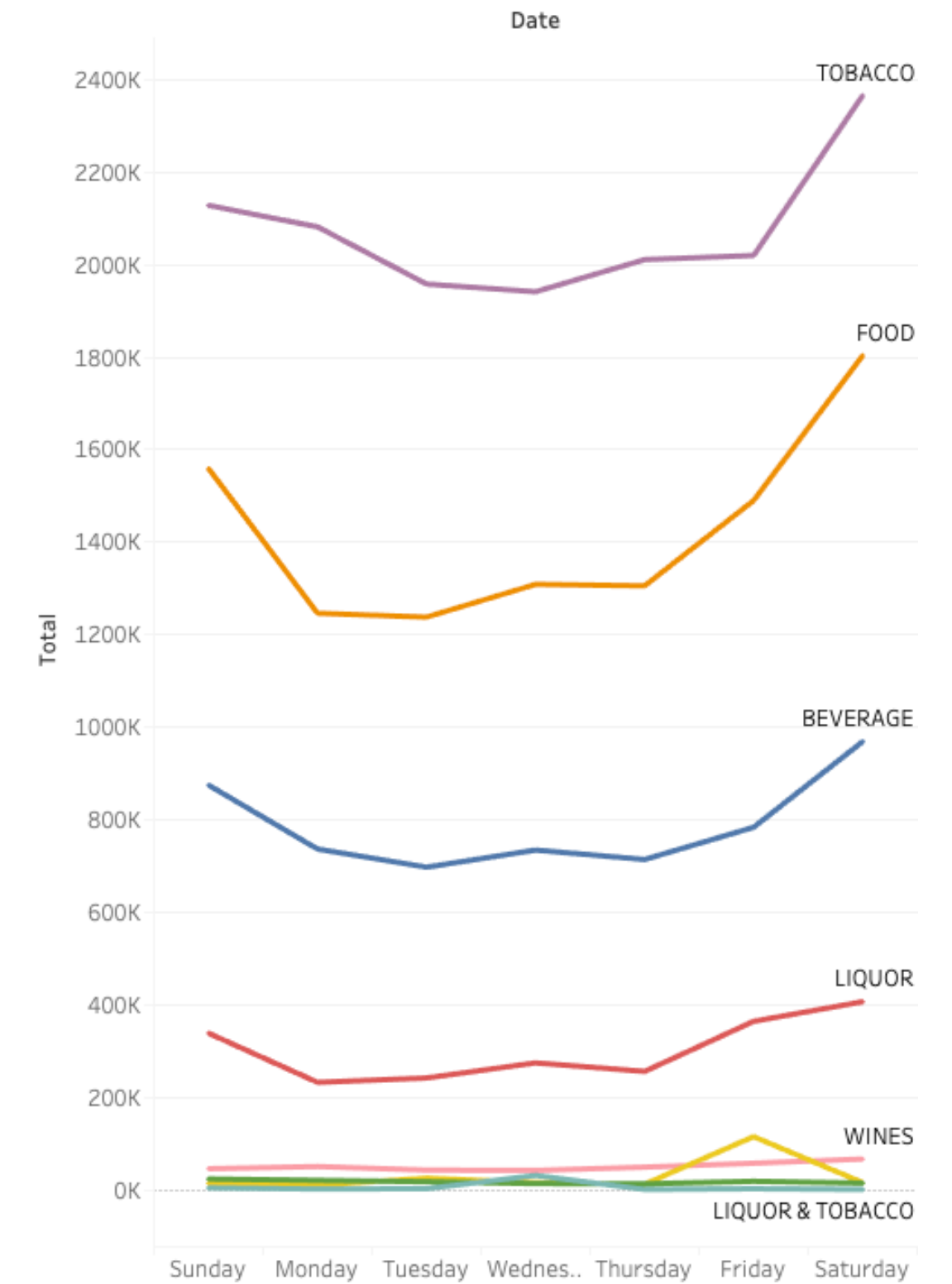
Sunday and Friday follow as the two next most lucrative days. Basically indicating that Friday through Sunday, the weekend, is when the cafe conducts most business

Tuesday shows the lowest earnings, though Monday through Thursday show a generally uniform pattern.

Annual Sales by weekdays



Category Annual Sales by weekdays



1.2

Recommendations

We see a general concentration of business over weekends, and in the evenings. It does appear that the cafe and its resources aren't optimally engaged over weekdays, and in the morning and afternoon slots.

In order to promote more traffic during this period, the cafe can be made more amenable to working individuals and teams. Freelancers, Remote workers, Entrepreneurs, Professionals and students can then work out from this space in the 'off hours'. The cafe's menu and operations can be more tailored to the requirements of the working individual who treats the cafe as more of a workspace in these hours.

Different monetisation means can be explored. Aside from charges for food, there can even be a periodic subscription to use the cafe's premises and infrastructure on a regular basis.

1.3

Least popular items on the menu

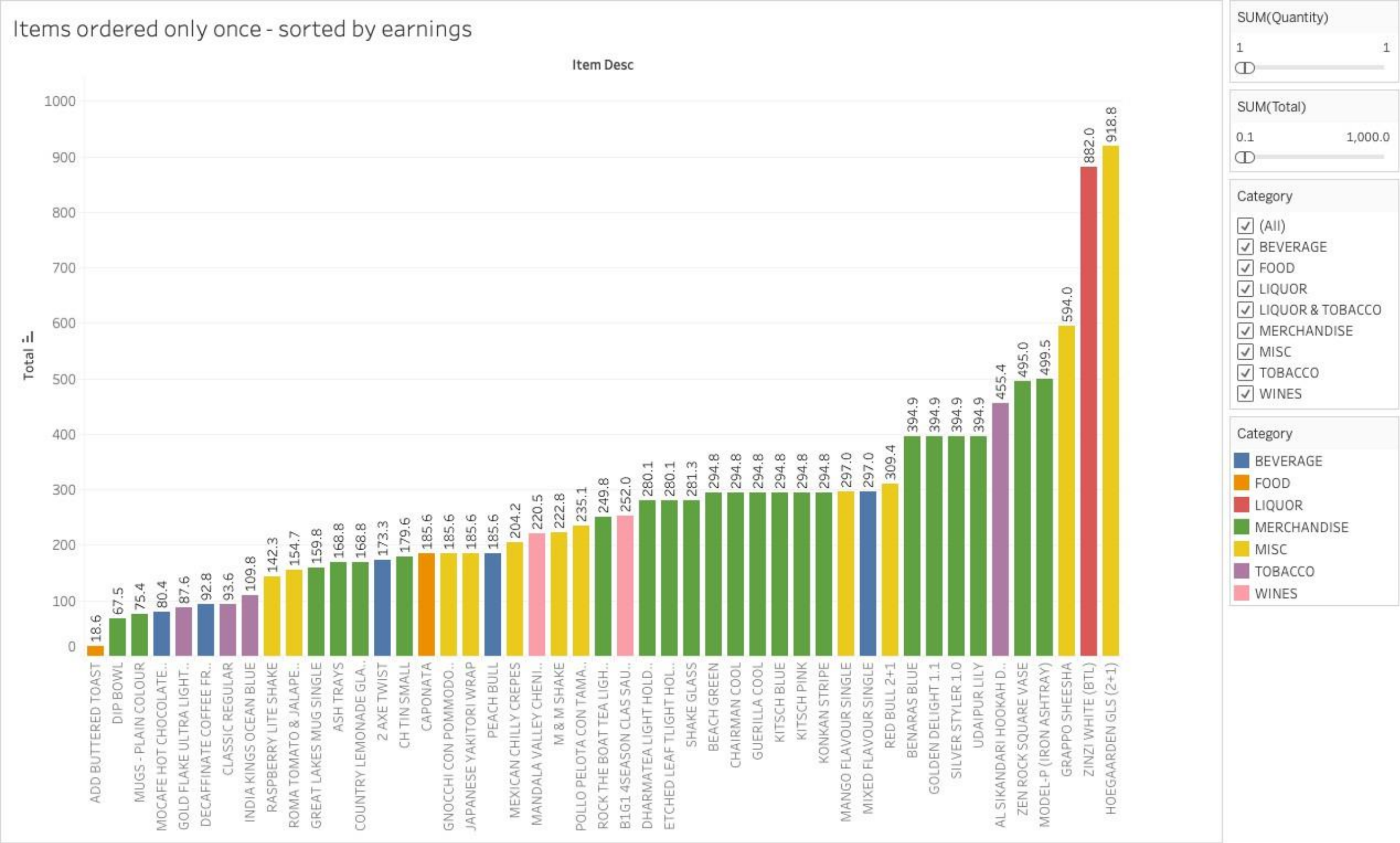
We have earlier noted that the highest earning 55 items contribute to over 75% of the cafe’s earnings.

In contrast, there are 54 items that have sold only one unit each, throughout the year.

The adjacent plot lists them in order of the revenue generated from the sales of these items.

It is seen that over 40% of these are from the Merchandise category.

And around 23% belong to Miscellaneous - which really is an assortment of items from all the other categories. So effectively, the 60% of the items can be uniformly attributed to the remaining categories.



1.3

Items that can be taken off the Menu

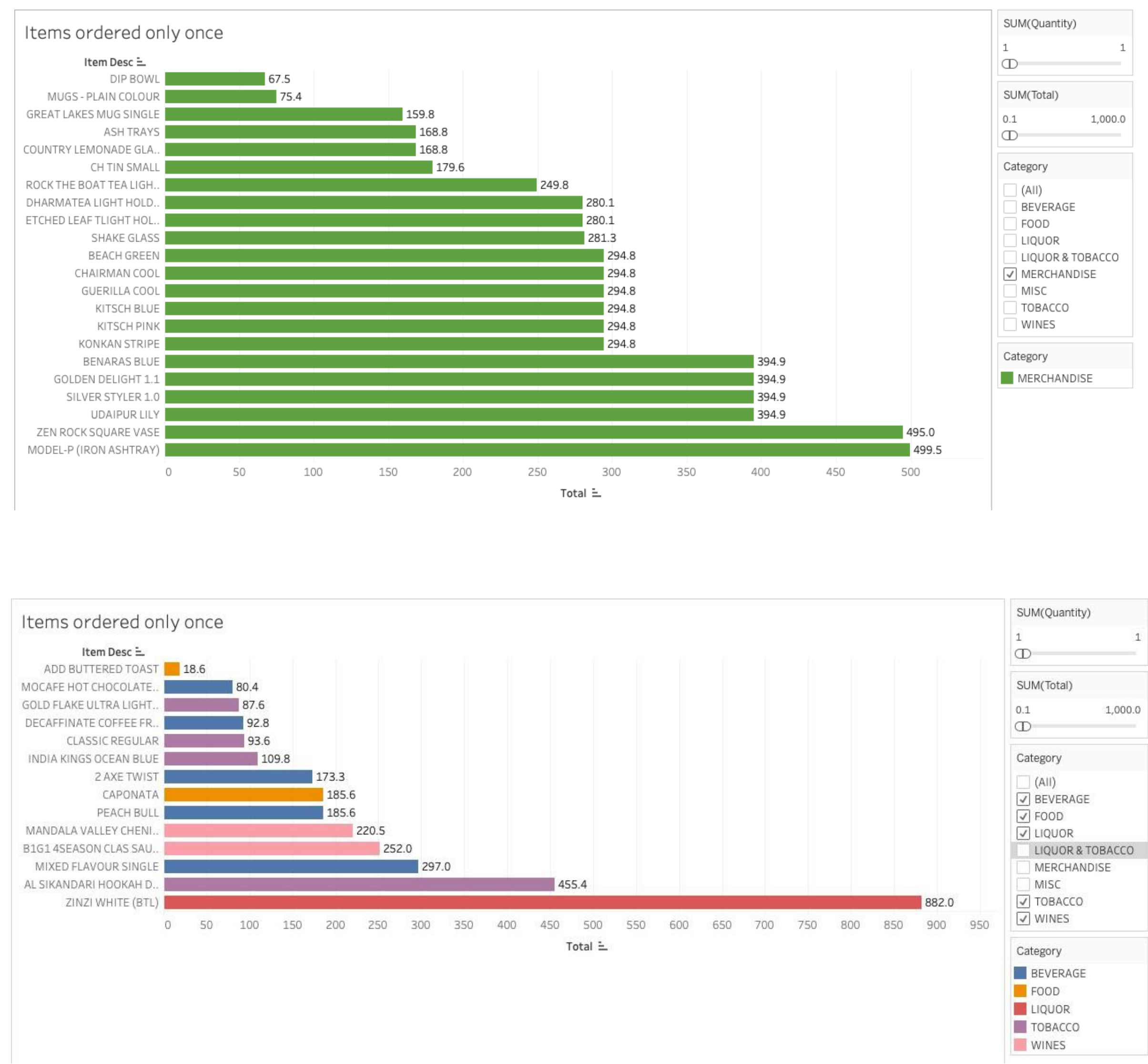
The plot on top depicts the least popular Merchandise Items, sorted based on the price. All of these have sold only 1 unit in the whole year.

Merchandise as a category accounts for 40 of the least popular 100 items as well. So it would be advisable to revisit the category as a whole. And one can start considering the listed items to be taken off the menu.

The plot below lists the least popular items across the other categories. The items look fairly evenly distributed across categories.

From among these, Perishable items within Food, and Tobacco can be considered for removal from the menu.

Miscellaneous (which constitutes ~20% of the set) covers Items from almost all categories. These items mostly appear to have been mis-classified. So we haven't included the category in the visualisation here.



1.3

Taking Items off the Menu

Approaching the issue from a revenue standpoint:

Listing the 50 items across categories that earned the lowest amount of revenue for the cafe.

Again, Merchandise has the bulk of the items in this set. And as suggested, this category may need to be reviewed.

Food also has a sizeable number, if items from Miscellaneous are also counted. They could try promoting these and also items from Beverage, and if they still don't show demand, maybe should consider dropping them.

Perishable items from Tobacco which show little demand should also be considered for dropping from the menu.

Table 1

Item Desc	Category	Quantity	Total
MOTHERS DAY SPL	FOOD	5	0.05
ADD BUTTERED TOAST	FOOD	1	18.56
CAPONATA	FOOD	1	185.63
CHICKEN HAM	FOOD	5	216.55
MUSHROOM & CORN	FOOD	7	216.58
SUNNY SIDEUP + BEVERAGE	FOOD	2	245.02
TOAST CIABATA	FOOD	6	297.00

Table 1-3

Item Desc	Category	Quantity	Total
GOLD FLAKE ULTRA LIGHTS(20)	TOBACCO	1	87.60
CLASSIC REGULAR	TOBACCO	1	93.60
INDIA KINGS OCEAN BLUE	TOBACCO	1	109.80
CLASSIC MENTHOL RUSH	TOBACCO	2	199.92
GOLD FLAKE LIGHTS-BIG	TOBACCO	3	262.80

Table 1-4

Item Desc	Category	Quantity	Total
MANDALA VALLEY CHENIN BLANC(GL	WINES	1	220.50
B1G1 4SEASON CLAS SAUV(GLS)	WINES	1	252.00

Table 1-1

Item Desc	Category	Quantity	Total
MOCAFE HOT CHOCOLATE(SF)	BEVERAGE	1	80.44
DECAFFINATE COFFEE FRAPPE	BEVERAGE	1	92.81
2 AXE TWIST	BEVERAGE	1	173.25
PEACH BULL	BEVERAGE	1	185.63
MIXED FLAVOUR SINGLE	BEVERAGE	1	297.00

Table 1-5

Item Desc	Category	Quantity	Total
ADD GROUND MEAT	MISC	3	111.39
RASPBERRY LITE SHAKE	MISC	1	142.31
ADD CHICKEN BACON	MISC	4	148.52
ROMA TOMATO & JALAPENO CROQUE	MISC	1	154.69
GNOCCHI CON POMMODORO	MISC	1	185.63
JAPANESE YAKITORI WRAP	MISC	1	185.63
MEXICAN CHILLY CREPES	MISC	1	204.19
M & M SHAKE	MISC	1	222.75
POLLO PELOTA CON TAMATE SALSA	MISC	1	235.13
MEDITER RANEAN PANINO	MISC	2	259.88
MANGO FLAVOUR SINGLE	MISC	1	297.00
RED BULL 2+1	MISC	1	309.38

Table 1-2

Item Desc	Category	Quantity	Total
CUTTING GLASS	MERCHANDISE	2	27.00
DIP BOWL	MERCHANDISE	1	67.50
MUGS - PLAIN COLOUR	MERCHANDISE	1	75.38
CH WRAPPING PAPER	MERCHANDISE	2	90.00
GREAT LAKES MUG SINGLE	MERCHANDISE	1	159.75
COUNTRY LEMONADE GLASS(HANSAN)	MERCHANDISE	1	168.75
ASH TRAYS	MERCHANDISE	1	168.75
CH TIN SMALL	MERCHANDISE	1	179.55
ROCK THE BOAT TEA LIGHT HOLDER	MERCHANDISE	1	249.75
DHARMATEA LIGHT HOLDER	MERCHANDISE	1	280.13
ETCHED LEAF TLIGHT HOLDER	MERCHANDISE	1	280.13
SHAKE GLASS	MERCHANDISE	1	281.25
CHAIRMAN COOL	MERCHANDISE	1	294.75
GUERRILLA COOL	MERCHANDISE	1	294.75
KITSCH PINK	MERCHANDISE	1	294.75
BEACH GREEN	MERCHANDISE	1	294.75
KONKAN STRIPE	MERCHANDISE	1	294.75
KITSCH BLUE	MERCHANDISE	1	294.75
CH CRICKET NOTEBOOK	MERCHANDISE	2	300.30

1.4

Trends across Months

The plots on the right depict Sales across months, from April 2010 to March 2011. The top plot depicts quantities sold and net proceeds from Sales for the period. The plot below splits Monthly Sales by Category.

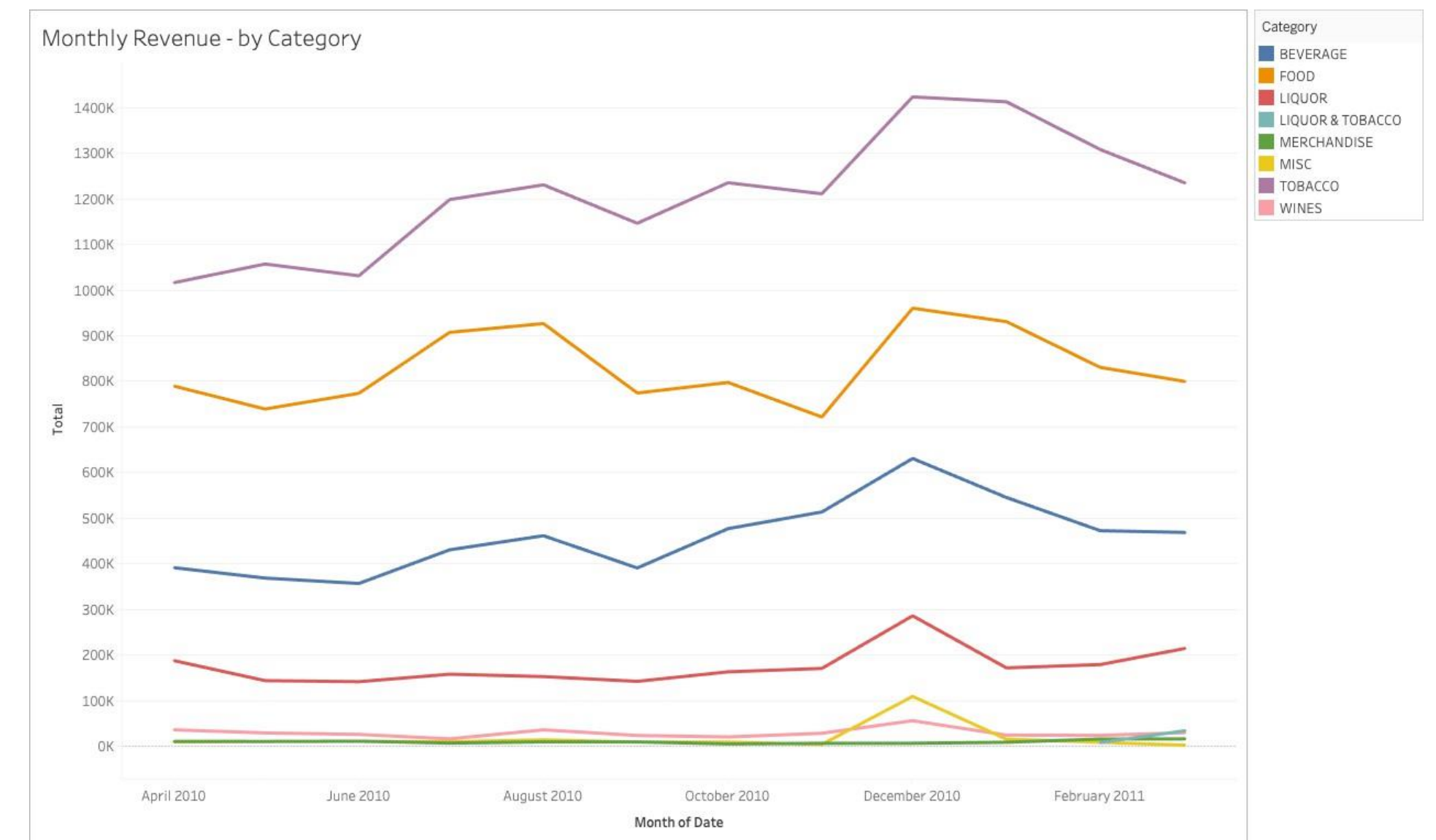
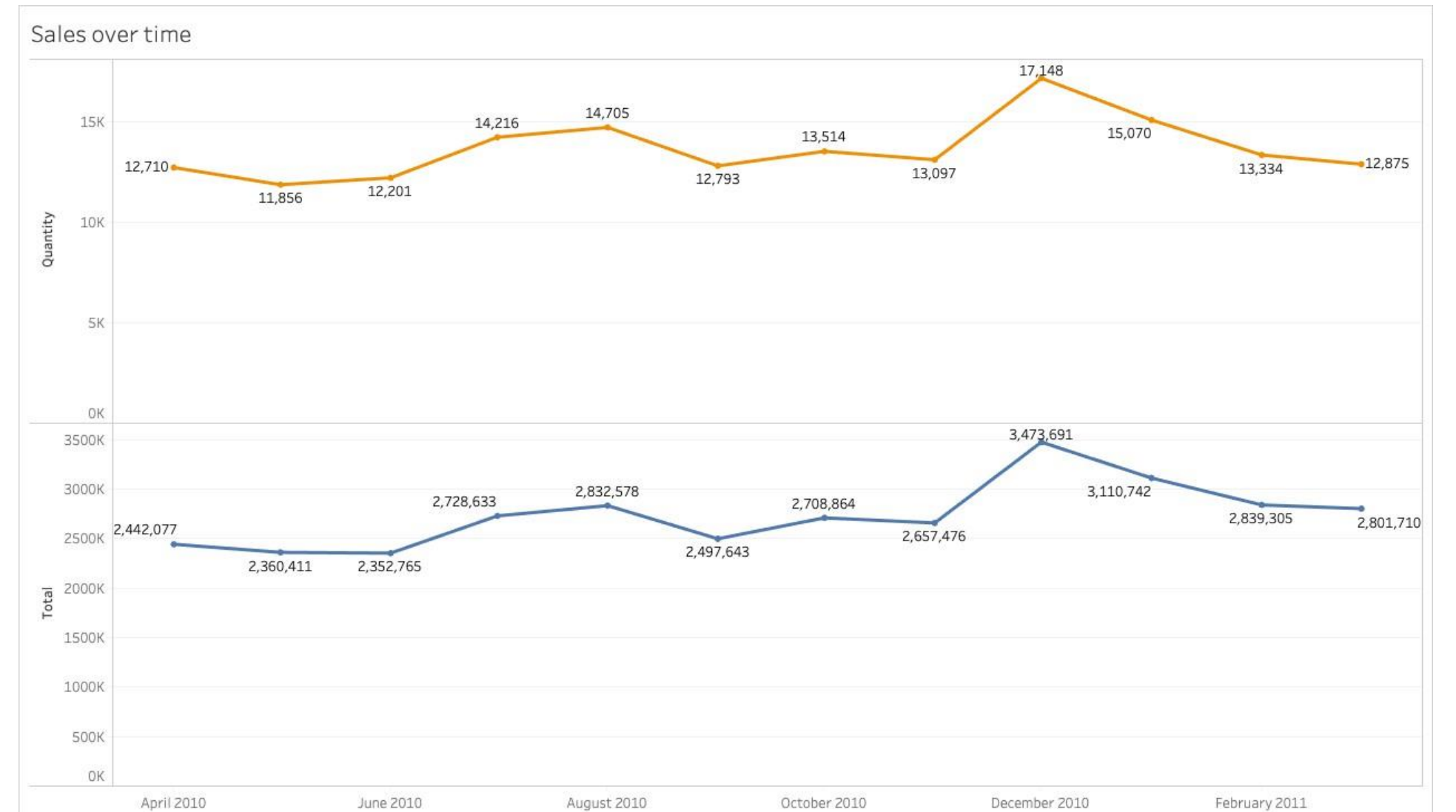
December 2010 is clearly the winner when it comes to Sales, with revenue nearing INR 35 lakhs. January 2011 is the next highest ranking month, with INR 31 lakhs in revenue.

June 2010 shows the lowest earnings at around INR 23 lakhs, and May 2010 has the lowest quantities sold.

At a Category level, a similar trend is seen. A couple of interesting observations:

Food records the lowest sales in the month of November 2011, where the general slump is seen in May and June 2010.

Liquor Sales are seen picking up in February and March 2011, where other categories show a gradual decline after December 2010.



1.4

Trends across Months

The 2 adjacent plots depict Monthly Sales, aggregated for Weekdays and Hours of the Day, respectively.

The Sales by Hours of the Day plot shows a fairly consistent pattern of monthly sales across all months, with Sales gradually building up with every hour since opening, with the peak at 11 PM.

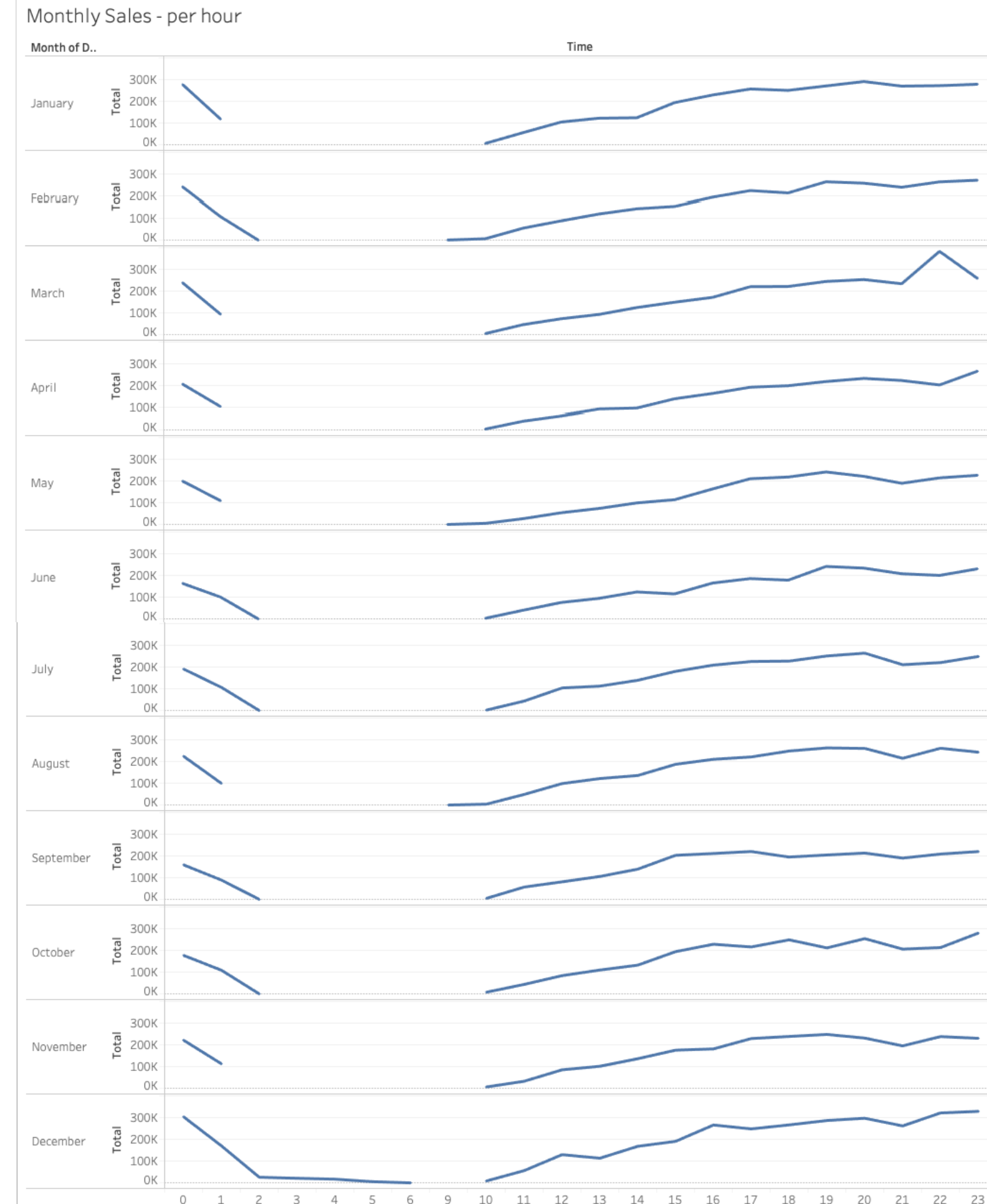
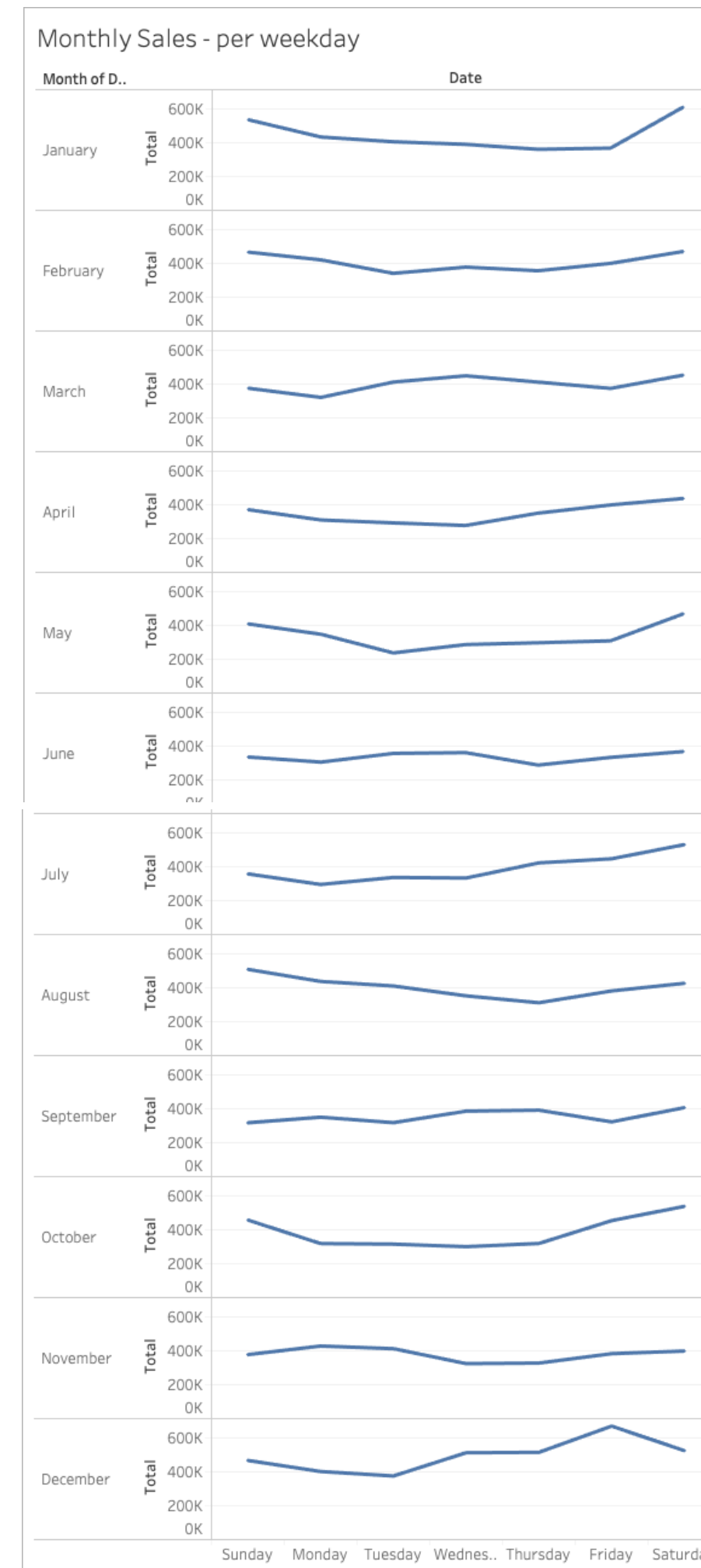
March however shows peak sales at 10 PM, which is distinct from any other month

Also noticeable is that for the month of February, May and August, business starts at 9 AM, which is earlier than usual.

December shows late hours business until 6 AM, which is probably accounted for by the New Year's Eve celebration

The Sales by Weekday, in comparison, shows much more distinct behaviour across months.

In general, weekends are still the busier than the rest of the week. However, November shows an unusually flat curve, with Monday and Tuesday accounting for most sales.



1.4

Trends across Quarters

To complete the monthly analysis, we aggregate monthly sales in quarters. The plots on this slide show Quarterly Sales for the period.

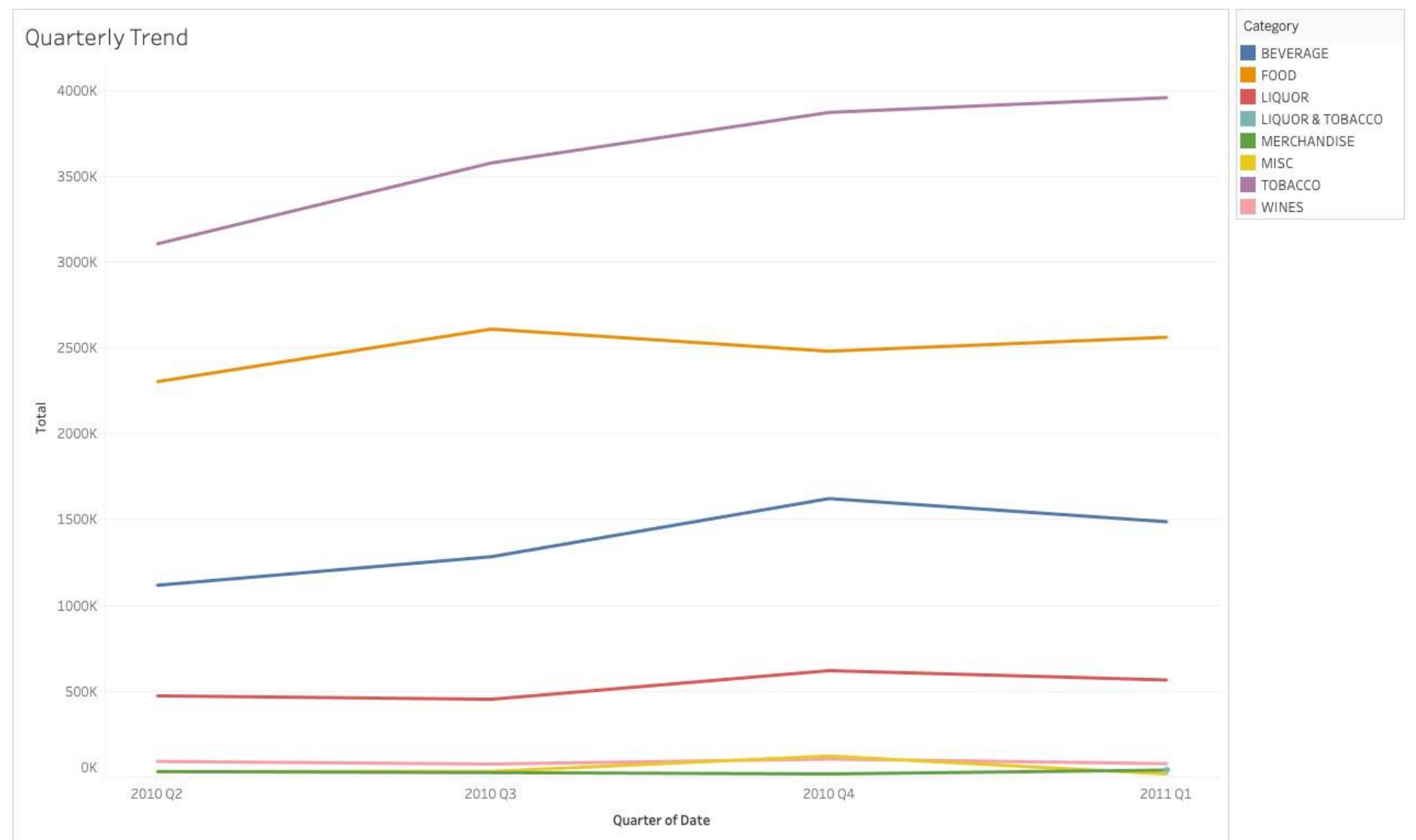
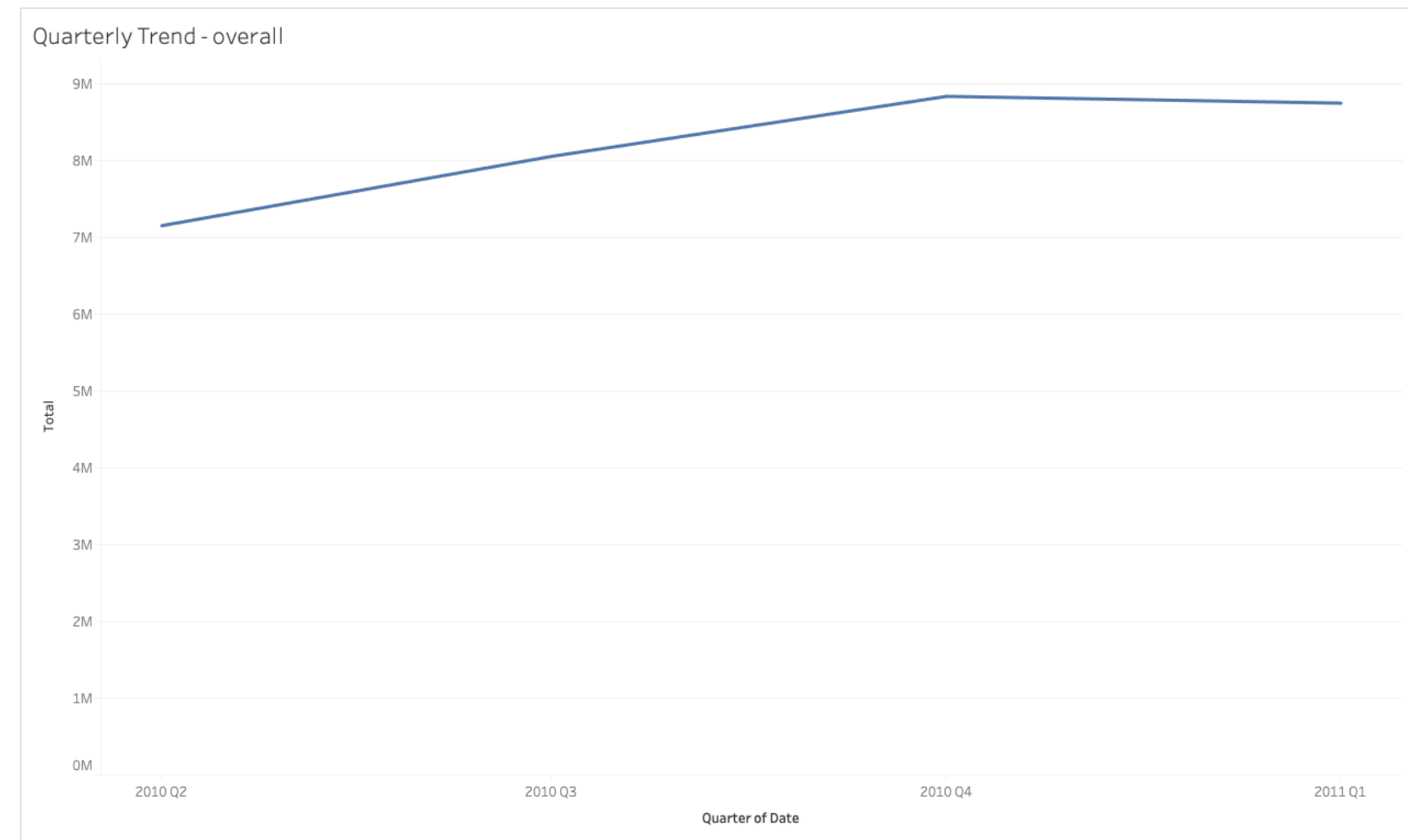
The Oct-Nov-Dec 2010 quarter saw the highest earnings, at INR 88.40 lakhs.

The Jan-Feb-Mar 2011 quarter comes a close second with INR 87.52 lakhs in revenues.

The July-Aug-Sep ranks third with INR 80.59 lakhs in earnings.

The April-May-June (AMJ 2010) quarter was the weakest in terms of earnings, with 71.55 lakhs.

Most categories follow the general trend. However Food and Tobacco tend to perform best in the Jan-Feb-March 2011 quarter.



Part 2 - Menu Analysis

2.1

Market Basket Analysis

We know from the EDA exercise that a typical Bill / Order at the cafe has 2 units of an item or items, and has a net order value of INR 391.50 - going by the median Quantity and median Total for Bills

We've also observed that outlier orders have many more items, and some combinations of categories such as Food and Beverage were apparent during the course of our EDA

In order to increase Revenues and Profits, it will be important to induce people to buy more per order.

So we conduct a Market Basket Analysis to build Association rules, which will indicate which items are likely to be bought together

If an association is established between Items, the customer can be induced to spend more per visit in various ways.

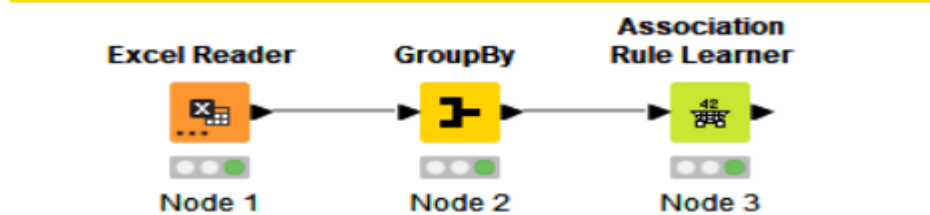
For the purpose of this exercise we will focus on building Combo offers. That is, the aim will be to incentivise purchase of more items by offering the right combos at the right price.



2.1

MBA – results

KNIME WORKFLOW



On the right are some of the Association Rules generated by the Market Basket Analysis conducted on the data.

Configuration specifics:
Minimum Support 0.001
Minimum Confidence 0.05

Results - 313 rules were generated using the above thresholds.

The rules here have been sorted in descending order of Lift. We will focus principally on rules and combinations that have a lift of greater than 4. There are 33 such instances in our analysis.

As a secondary measure, we will also look at Combinations that have relatively High support and Lift between 2 and 3.

For this exercise we will ignore all associations with lift less than 2 - for that implies little or no association.

Association Rules - in descending order of Lift						
RowID	Support	Confidence	Lift	Consequent	implies	Items
rule21	0.00104312537509645	0.178048780487805	79.3643933509399	ADD HERB ROAST CHICKEN	<---	[LEMON INFUSED CHAR GRILLED VEG]
rule22	0.00104312537509645	0.464968152866242	79.3643933509399	LEMON INFUSED CHAR GRILLED VEG	<---	[ADD HERB ROAST CHICKEN]
rule147	0.00161470092309451	0.545893719806763	18.2351953696978	CAFFE LATTE	<---	[ADD HAZELNUT FLAVOUR]
rule148	0.00161470092309451	0.0539379474940334	18.2351953696978	ADD HAZELNUT FLAVOUR	<---	[CAFFE LATTE]
rule44	0.00112886170729616	0.105053191489362	7.47897502218567	SATAY CHICKEN PANINI	<---	[ADD FRIES]
rule45	0.00112886170729616	0.0803662258392675	7.47897502218567	ADD FRIES	<---	[SATAY CHICKEN PANINI]
rule12	0.00101454659769655	0.265917602996255	7.14374114889977	B.M.T. PANINI	<---	[FRENCH FRIES]
rule75	0.00124317681689577	0.0658093797276853	6.12429788843467	ADD FRIES	<---	[COUNTRY ROAST CHICKEN PANINI]
rule74	0.00124317681689577	0.115691489361702	6.12429788843467	COUNTRY ROAST CHICKEN PANINI	<---	[ADD FRIES]
rule101	0.00137178131519534	0.463768115942029	5.93551943852507	CAPPUCCINO	<---	[ADD HAZELNUT FLAVOUR]
rule76	0.00124317681689577	0.412322274881517	5.92386315761821	GREAT LAKES SHAKE	<---	[VANILLA ICECREAM]
rule114	0.00142893886999514	0.186567164179104	5.84437926749422	MASALA CHAI CUTTING	<---	[BUN MASKA & CHAI]
rule9	0.00101454659769655	0.181122448979592	5.77989567920191	LEMON ICED TEA	<---	[BLACK CURRANT ICED TEA]
rule28	0.00107170415249636	0.0672645739910314	5.53150342777951	KF DRAUGHT (1LTR)	<---	[KF DRAUGHT (1/2LTR)]
rule29	0.00107170415249636	0.0881316098707403	5.5315034277795	KF DRAUGHT (1/2LTR)	<---	[KF DRAUGHT (1LTR)]
rule213	0.00215769769369266	0.200797872340426	5.39433270715073	B.M.T. PANINI	<---	[ADD FRIES]
rule214	0.00215769769369266	0.0579654510556622	5.39433270715073	ADD FRIES	<---	[B.M.T. PANINI]
rule41	0.00111457231859621	0.0618066561014263	4.93198792165338	COOL CALIFORNICA	<---	[MIAMI MELONS]
rule40	0.00111457231859621	0.088939566704675	4.93198792165338	MIAMI MELONS	<---	[COOL CALIFORNICA]
rule111	0.00141464948129519	0.082089552238806	4.92270012405837	RED BULL 2+1	<---	[ARABIAN MIST]
rule112	0.00141464948129519	0.0848329048843188	4.92270012405837	ARABIAN MIST	<---	[RED BULL 2+1]
rule165	0.00178617358749393	0.142531356898518	4.91119124493947	BERRY BLAST	<---	[COOL CALIFORNICA]
rule166	0.00178617358749393	0.0615460364352536	4.91119124493947	COOL CALIFORNICA	<---	[BERRY BLAST]
rule55	0.00117172987339602	0.0620272314674735	4.41585932101397	SATAY CHICKEN PANINI	<---	[COUNTRY ROAST CHICKEN PANINI]
rule56	0.00117172987339602	0.0834181078331638	4.41585932101397	COUNTRY ROAST CHICKEN PANINI	<---	[SATAY CHICKEN PANINI]
rule133	0.00151467520219485	0.0631328171530673	4.29364510204661	THE FERROR ROCHER SHAKE	<---	[KIT KAT SHAKE]
rule132	0.00151467520219485	0.103012633624879	4.2936451020466	KIT KAT SHAKE	<---	[THE FERROR ROCHER SHAKE]
rule19	0.00104312537509645	0.0970744680851064	4.2143085766327	COTTAGE CHEESE PANINI	<---	[ADD FRIES]
rule60	0.00117172987339602	0.109042553191489	4.10269675131549	PHILLYCREAM CHEESE &CHILLY PAN	<---	[ADD FRIES]
rule183	0.00191477808579349	0.152099886492622	4.08608608695842	B.M.T. PANINI	<---	[MAGGI NDL ARRABIATA]
rule184	0.00191477808579349	0.0514395393474088	4.08608608695842	MAGGI NDL ARRABIATA	<---	[B.M.T. PANINI]
rule5	0.00101454659769655	0.0501766784452297	4.00395018352801	COOL CALIFORNICA	<---	[PINK LEMONADE]
rule4	0.00101454659769655	0.080957810718358	4.00395018352801	PINK LEMONADE	<---	[COOL CALIFORNICA]
rule290	0.00414392272298591	0.0655663576757857	3.93184648060568	RED BULL 2+1	<---	[SAMBUCA]
rule289	0.00414392272298591	0.248500428449015	3.93184648060568	SAMBUCA	<---	[RED BULL 2+1]
rule175	0.00185762053099368	0.0883152173913044	3.60167572463768	CARLSBERG	<---	[TUBORG]
rule176	0.00185762053099368	0.0757575757575758	3.60167572463768	TUBORG	<---	[CARLSBERG]
rule80	0.00125746620559572	0.0808823529411765	3.37123813194128	KIT KAT SHAKE	<---	[TOBLERONE SHAKE]
rule81	0.00125746620559572	0.0524121500893389	3.37123813194128	TOBLERONE SHAKE	<---	[KIT KAT SHAKE]
rule161	0.00172901603269412	0.0668877833056938	3.30808540727849	PINK LEMONADE	<---	[COUNTRY LEMONADE]
rule162	0.00172901603269412	0.0855123674911661	3.30808540727849	COUNTRY LEMONADE	<---	[PINK LEMONADE]
rule243	0.0025006430224915	0.198637911464245	3.14290714901443	SAMBUCA	<---	[MAGGI NDL ARRABIATA]
rule231	0.00237203852419194	0.0501965527668582	3.01015865958035	RED BULL 2+1	<---	[CALCUTTA MINT]

2.2

Associations

On the right are 20 Associations derived from the 33 rules that we are focusing on. As mentioned previously, we are primarily considering only those associations that have a Lift of 4 or more.

All combinations here have a support of 0.001, except for BMT Panini and Fries, which is slightly higher at 0.002.

Arabian Mist and Red Bull 2 + 1 are the most expensive combination. This also stands out as the only mixed-category combination. The rest of the combinations are arguably items in the same category.

Cappuccino with an added Hazelnut flavour is least expensive of combinations, but given that Cappuccino ranks 2 in terms of popularity, it will make up in volumes.

Interestingly, Fries / French Fries is an accompaniment that features in many Food item combinations.

20 Associations based on LIFT						
Rule	Lift	Item 1	Price (Item 1)	Item 2	Price (Item 2)	Total Price
rule22	79.3643933509399	LEMON INFUSED CHAR GRILLED VEG	150	[ADD HERB ROAST CHICKEN]	30	180
rule147	18.2351953696978	CAFFE LATTE	70	[ADD HAZELNUT FLAVOUR]	25	95
rule44	7.47897502218567	SATAY CHICKEN PANINI	115	[ADD FRIES]	35	150
rule12	7.14374114889977	B.M.T. PANINI	105	[FRENCH FRIES]	35	140
rule74	6.12429788843467	COUNTRY ROAST CHICKEN PANINI	115	[ADD FRIES]	35	150
rule101	5.93551943852507	CAPPUCCINO	60	[ADD HAZELNUT FLAVOUR]	25	85
rule76	5.92386315761821	GREAT LAKES SHAKE	110	[VANILLA ICECREAM]	30	140
rule114	5.84437926749422	MASALA CHAI CUTTING	40	[BUN MASKA & CHAI]	60	100
rule9	5.77989567920191	LEMON ICED TEA	85	[BLACK CURRANT ICED TEA]	85	170
rule28	5.53150342777951	KF DRAUGHT (1LTR)	240	[KF DRAUGHT (1/2LTR)]	125	365
rule213	5.39433270715073	B.M.T. PANINI	105	[ADD FRIES]	35	140
rule41	4.93198792165338	COOL CALIFORNICA	85	[MIAMI MELONS]	85	170
rule112	4.92270012405837	ARABIAN MIST	350	[RED BULL 2+1]	250	600
rule165	4.91119124493947	BERRY BLAST	95	[COOL CALIFORNICA]	85	180
rule56	4.41585932101397	COUNTRY ROAST CHICKEN PANINI	115	[SATAY CHICKEN PANINI]	115	230
rule133	4.29364510204661	THE FERROR ROCHER SHAKE	175	[KIT KAT SHAKE]	135	310
rule19	4.2143085766327	COTTAGE CHEESE PANINI	105	[ADD FRIES]	35	140
rule60	4.10269675131549	PHILLYCREAM CHEESE &CHILLY PAN	105	[ADD FRIES]	35	140
rule183	4.08608608695842	B.M.T. PANINI	105	[MAGGI NDL ARRABIATA]	125	230
rule5	4.00395018352801	COOL CALIFORNICA	85	[PINK LEMONADE]	85	170

2.2

Associations

On the right are 4 additional associations combining Lift and Support.

We include combinations with Lift value 2 or 3, when the corresponding support is 0.003 or more.

Here we find the combination of Sambuca and Red Bull appearing in 2 sets.

Also, we have more of cross-category combinations in this case, unlike in the previous selection based on Lift alone.

So combining both approaches, we have shortlisted 24 associations which we will develop into combos presently.

4 Associations based on LIFT & SUPPORT in combination							
Rule	Support	Lift	Item 1	Price (Item 1)	Item 2	Price (Item 2)	Total Price
rule289	0.00414392272298591	3.93184648060568	SAMBUCA	350	[RED BULL 2+1]	250	600
rule287	0.00404389700208625	2.40737031985783	SAMBUCA	350	[RED BULL ENERGY DRINK]	125	475
rule277	0.00367237289588751	2.0745715986269	B.M.T. PANINI	105	[QUA MINERAL WATER(1000ML)]	50	155
rule280	0.00380097739418708	2.09313260178536	POUTINE WITH FRIES	125	[B.M.T. PANINI]	105	230

2.3

Possible Combos and Offers

If we review the 24 shortlisted associations, we can group the combos in a few ways:

Combo 1. 'Add Fries to it!' : Paninis and French Fries

It can be noticed that out of the 20 associations shortlisted, 6 of them were paired with French Fries / Add Fries (which are essentially the same item).

The items are:

SATAY CHICKEN PANINI

B.M.T. PANINI

COUNTRY ROAST CHICKEN PANINI

COTTAGE CHEESE PANINI

PHILLYCREAM CHEESE &CHILLY PAN

All these Food Items listed are essentially Paninis, price points between INR 105 to INR 115. Now this indicates Paninis and Fries have a strong fit, but the Confidence indicators show that only 6% to 12% of people opt for Fries. In the case of BMT Panini it is 20%.

We could provide Fries at an aggressively discounted price for more people to opt for it. If an additional INR 35 dissuades customers from buying Fries, perhaps a substantial discount would induce a larger number of people to opt for Fries with their order.

Fries / French Fries could be priced at INR 15 / 20 when purchased with a Panini.

2.3

Possible Combos and Offers

Combo 2. 'Twinning!'

Applied to Cold Beverages such as Ice Teas, Non-Alcoholic Drinks and Shakes - to induce more people to opt for the more expensive non-alcoholic beverages, instead of the cheaper coffees and drinks

The combinations featured here are:

LEMON ICED TEA and BLACK CURRANT ICED TEA (both priced at INR 85)

COOL CALIFORNICA paired with MIAMI MELONS, PINK LEMONADE, BERRY BLAST (INR 85 / 95)

THE FERROR ROCHER SHAKE (INR 175) and KIT KAT SHAKE (INR 135)

The confidence measure for all these combinations vary. The Iced Teas show a 20% confidence but other drinks range from 6 to 14%. Nonetheless, there is scope to increase consumption of these drinks using combo offers.

One could offer a 25% to 50% off on the second drink in the cold beverage category or sub-category.

2.3

Possible Combos and Offers

Combo 3. KF Draught 1 + 1/2

This is applied specifically on KF Draught 1 ltr and 1/2 ltr products. And it is likely aimed at the 1 ltr consumer, who may not want to opt for another 1 ltr unit, but may yet prefer an additional smaller unit.

The combinations featured here are:

KF DRAUGHT (1LTR) priced at INR 240, and KF DRAUGHT (1/2LTR) priced at INR 125

One could offer 1 + 1/2 ltr for INR 340 (conservative) / INR 300 (aggressive), effectively offering a 20% or ~50% on the additional 1/2 ltr.

Combo 4: Premium Tobacco and Red Bull

Strong associations seen in the following:

ARABIAN MIST (INR 350) and RED BULL 2+1 (INR 250)

SAMBUCA (INR 350) and RED BULL 2+1 (INR 250)

SAMBUCA (INR 350) and RED BULL ENERGY DRINK (INR 125)

Sambuca / AM Combo with Red Bull 2 + 1 could be priced at INR 500, and with Red Bull Energy Drink at INR 450.

Thank you!

Appendix:

The supporting files are:

1. **MRA Project Extended project.ipynb** (with most of the EDA code and workings, attached)
2. Tableau Public link for visualisations: https://public.tableau.com/app/profile/subhadeep.seal/viz/MRA_Extendedproject_visualizations/BillValuesSorted
3. KNIME file/s for the Market Basket Analysis component (attached)