MSCI 719: Operations Analytics

Assignment 2: HAPPY COW ICE CREAM: DATA-DRIVEN SALES

FORECASTING

Student: Advait Shah

Faculty: Engineering

Department: Management Sciences

Instructor: Prof. Hossein Abouee Mehrizi

Contents

1. Effect of Flavour/ Customer Types

Mary and Prem, the sales assistant, believe that different groups of customers have different behaviour regarding the different groups of flavours. Based on the data and your intuition regarding ice cream, create 4 groups of flavours and address the following questions.

1.1. Create a dynamic dashboard that visualizes the sales of flavours in different time units, i.e., daily, weekly, and monthly.

Here is the head of our dataset:

Product	identity	Sat 1st Apr 2017	Sun 2nd Apr 2017	Mon 3rd Apr 2017	Tue 4th Apr 2017	Wed 5th Apr 2017	Thu
-$5 LUSH COUPON ON $40	event	0	0	0	0	0	
Mango	tourism	0	0	40	0	0	
S. Caramel Staff	Staff	0	0	0	0	60	
S. Caramel St	student	390	0	806	0	650	
Chocolate	tourism	40	0	0	0	120	
Mango Staff	Staff	30	0	60	0	90	
Strawberry	tourism	0	0	40	0	0	
Mint Choco St	student	208	0	338	0	416	

Basically, we have sales and revenue data for each product (Ice-cream) category along with type of customer who bought it.

This data also includes some of the products entries consisting of sales and revenue values in negative and zero (tax row). So, using Tableau Prep (ETL Tool), we have created a workflow, as shown below, to load this Microsoft Excel file data, and cleaned it as mentioned above. Moreover, we have grouped products into four flavours in the 3rd transformation step.
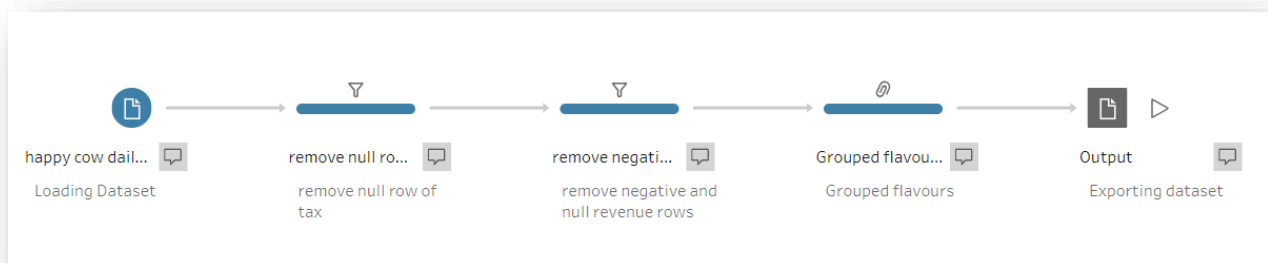

Fig: Data Transformation: ProcessFlow1 steps

Fig: Products Grouped as per Flavours

Furthermore, we need to convert transaction date columns into single column with different date wise sales figures, which will allow us to group data as per time granularity requirements. We have achieved this below step in Tableau Prep Software, and created Dataset: Processed2.xlsx for further data analysis:

Fig: Data Transformation: ProcessFlow2 steps

Now, we have imported our data into Tableau Desktop for visualization and further data massaging, filtering, and creating business insights. As a first step, we analysed day-wise sales, and removed days with zero total sales, as we assume that shop would be closed on those days, because they were falling on Sundays. These is highlighted in below figure:
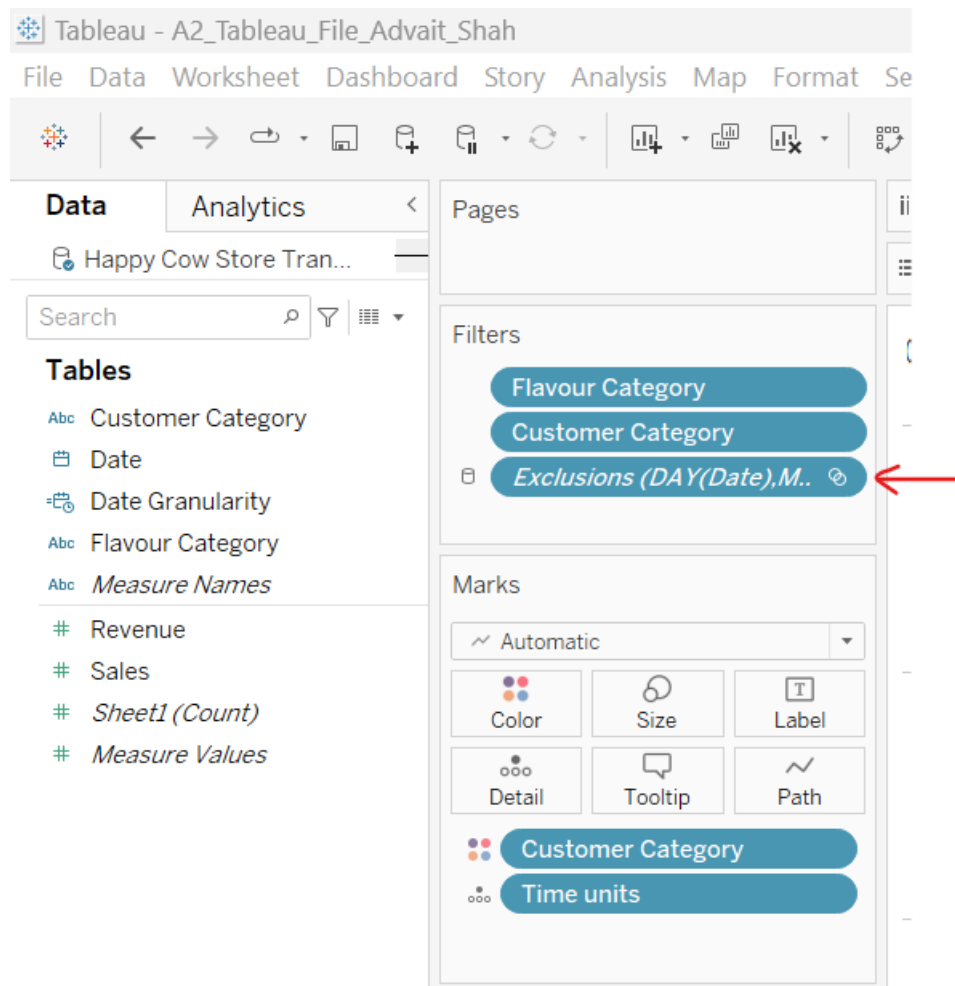

Fig: Data Filtered to remove Zero sales on some of the Sundays

Then, I created parameter to group days into Daily, Weekly and Monthly timeframes, and named it as "Time unit". And with this we are ready to prepare required charts to visualize flavour-wise and

customer group based sales in this given time period. Thus, next I created following Tableau Dynamic Dashboard for this purpose:
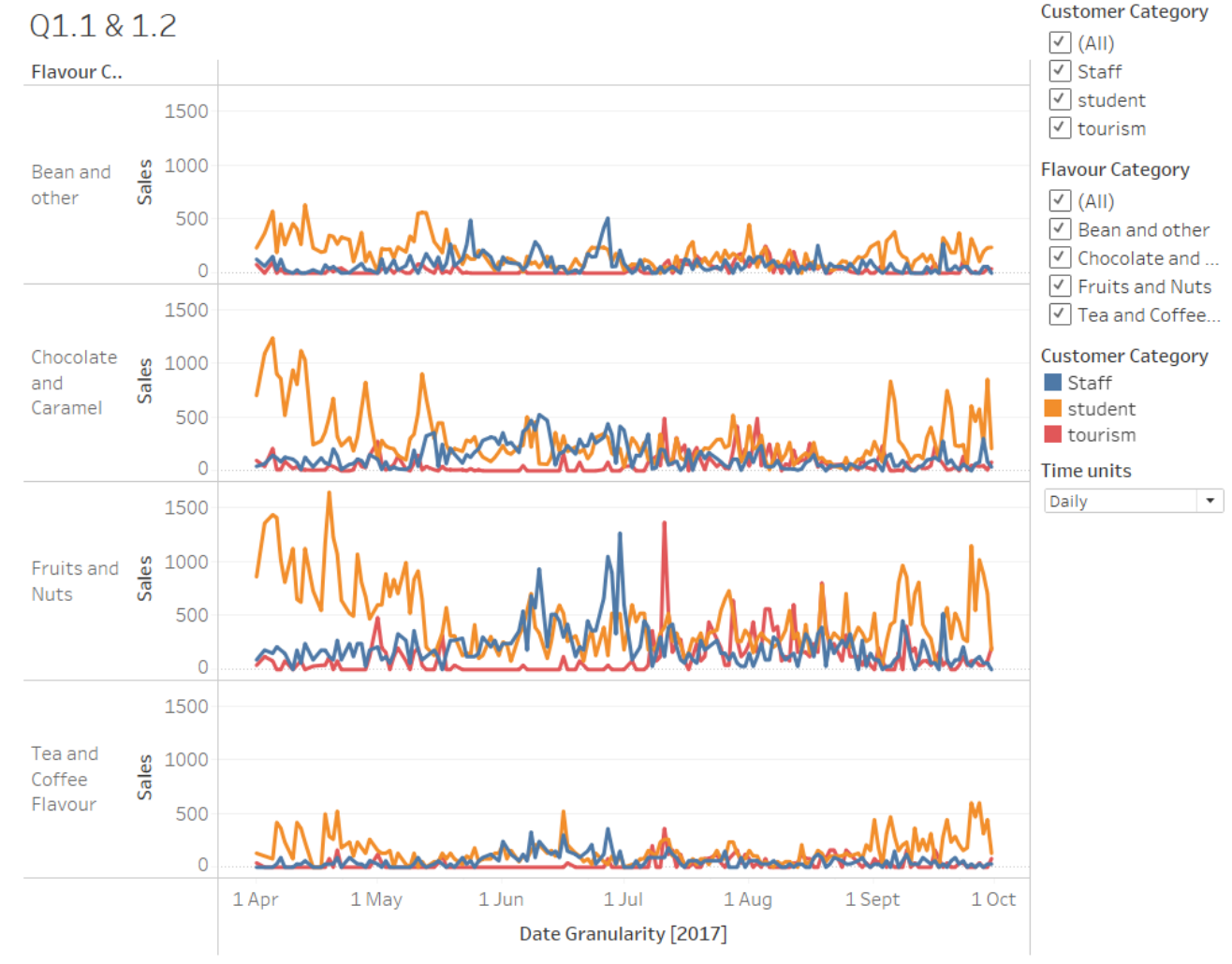

Fig: Dynamic Dashboard to visualize the sales of flavours, and with respect to different time units

Then, I have shown below the visualization when we use weekly and monthly time unit parameter:
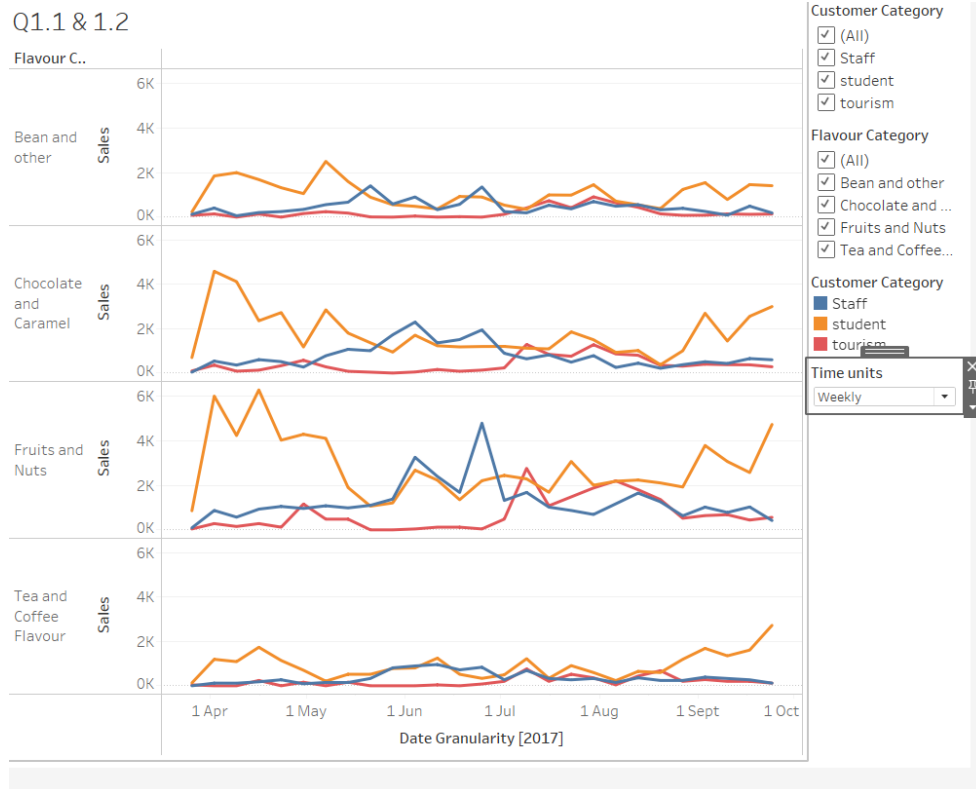
Fig: Dynamic Dashboard to visualize the sales of flavours, weekly time units activated
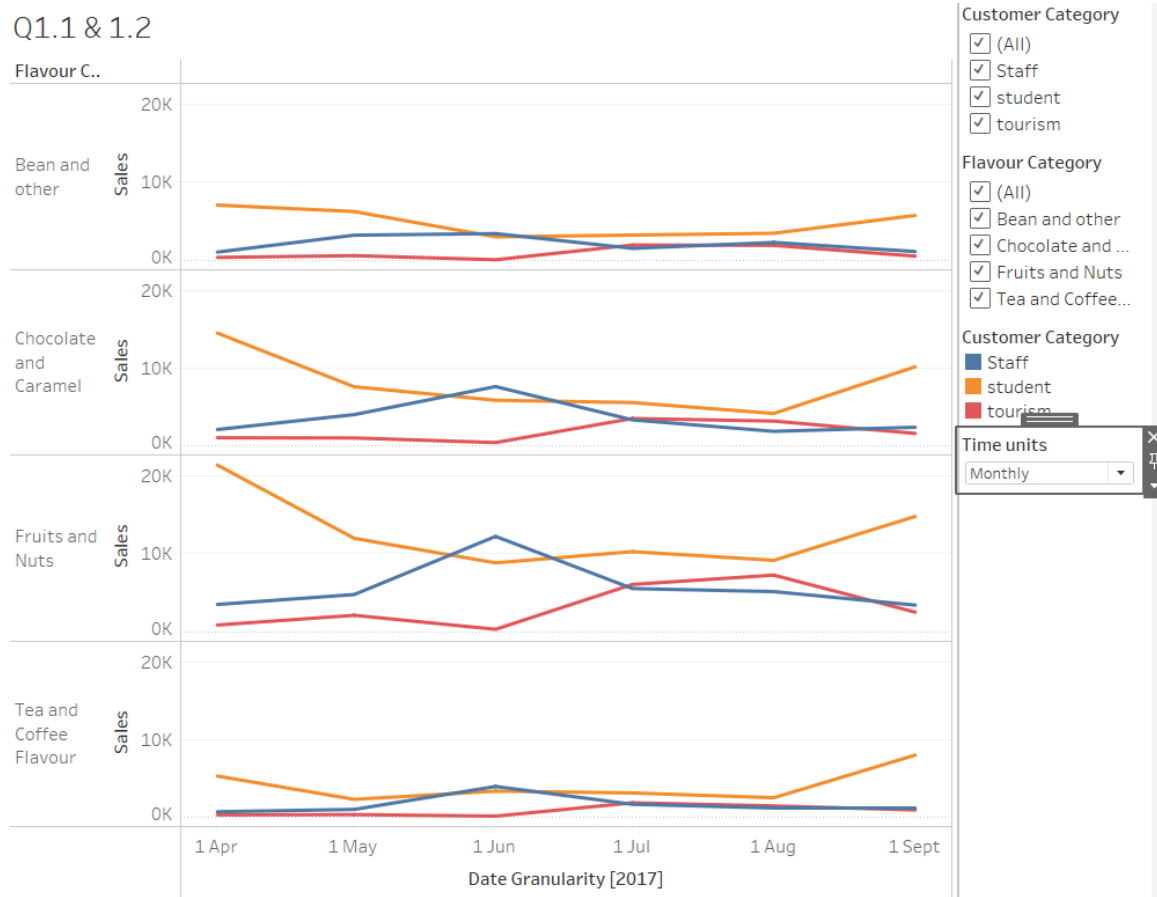

Fig: Dynamic Dashboard to visualize the sales of flavours, Monthly time units activated

1.2. Examine if there is a difference in the behaviour of customer groups for different flavour groups:

Sales of icecream at happy cow varies significantly for different customer groups over the year. From April, sales accounted to "students" starts to dip, whereas sales from "staff" and "tourism" starts picking up from month of May. And until, August all 3 groups accounts for almost same sales figures. But, again from September most of the sales begins to come through students customer category.

Now, looking at the preference of flavours for this customer groups, we can observe that "Fruits and Nuts" and "Chocolate and caramel" flavours are consistently liked by all 3 customer groups, namely, "Staff", "Students" and "Tourism". Whereas "Tea and Coffee" and "Bean and other" flavours are more seasonal, and their total daily sales are comparatively lesser than the other 2 flavour groups. Interestingly, "tea and coffee" is least preferred by "students" group. Also, we can see a spike in the sales of "Tea and Coffee" and "Fruits and nuts" flavour ice-creams accounted to the "tourism" group.

We can also test our hypothesis on following total sales data:

Null hypothesis, H0: Homogeneous total sales for all 3 Customer Groups and Flavour categories

Alternate hypothesis, H1: total sales differ significantly across these flavour categories

As this will be a paired dataset case, we will test this hypothesis using 2-way ANOVA method, as shown below, in Microsoft Excel after creating data summary in Tableau and/ or pivot table:

Sum of Sales	Flavour Groups				
Customer Groups	Bean and other	Chocolate and Caramel	Fruits and Nuts	Tea and Coffee Flavour	Grand Total
Staff	12619.97	21169.59	34680.71	9618.73	78089.00
student	28791.88	47815.64	76551.81	24558.67	177718.00
tourism	5417.91	10560.26	19190.83	4920.00	40089.00
Grand Total	46829.76	79545.49	130423.35	39097.39	295896.00

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
Staff	4	78088.99988	19522.24997	126071313.6
student	4	177718	44429.49999	560897959.8
tourism	4	40089	10022.25	43861772.72
Bean and other	3	46829.75628	15609.91876	143290526.9
Chocolate and Caram	3	79545.49473	26515.16491	368422240.2
Fruits and Nuts	3	130423.3549	43474.45164	880567788.8
Tea and Coffee Flavo	3	39097.39391	13032.46464	105159509.3

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	2525973274	2	1262986637	16.16082109	0.003838145	5.143253
Columns	1723586282	3	574528760.6	7.351508115	0.019601946	4.757063
Error	468906856.8	6	78151142.81			
Total	4718466412	11				

As we can see, p-value is less than 0.05 for both rows and columns groups, and therefore we reject null hypothesis, and can conclude that these flavour groups as well as customer groups do have different behaviours for at least one of their underlying groups compared to the rest of the groups.

1.3. Is there any difference in the revenue generated by each group? Address this question both visually and statistically. What is your suggestion to increase the revenue based on the data?

To answer this, I have created a bar chart showing average revenue for each flavour groups.

Average revenue for a flavour group is calculated as, (Total revenue from a flavour group) / (no. of flavours in that group)
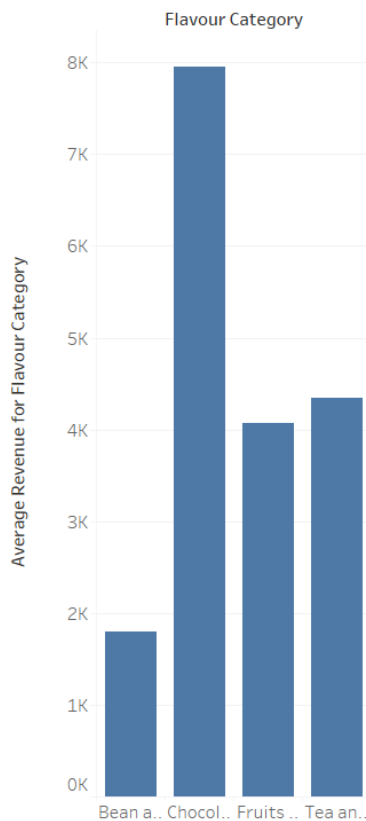
Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Bean and other	26	46829.75628	1801.144472	9902240.053
Chocolate and Caramel	10	79545.49473	7954.549473	52716984.83
Fruits and Nuts	32	130423.3549	4075.729842	14654658.97
Tea and Coffee Flavour	9	39097.39391	4344.154879	10940117.3

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	281440539.3	3	93813513.1	5.418780782	0.002019243	2.730018714
Within Groups	1263824231	73	17312660.7			
Total	1545264771	76				

Fig: Anova Test results for flavour groups revenues

Fig: Average revenue from each Flavour Groups

Also, we can see the total revenue and difference in revenues between the groups from the Anova table shown above. As expected, p-value for this groups comparison is less than 0.05, which means null hypothesis of equal revenue from each flavour group should be rejected, and it suggests that there is a significant difference in the average revenue values between these flavour groups.

Moreover, when we look at the total revenue values generated by each flavour groups, we find that "Fruits and Nuts" has generated highest revenue, however, average revenue value is the highest for "Chocolate and Caramel". That means, although there are less flavours in the "Chocolate and Caramel" category, each flavour in that group generates high amount of revenue. On the other hand, " fruits and nuts" has more number of flavours falling under its category and thus generates high revenue in total, but some of the flavours are not running that quite well and therefore it reduces average revenue generated by this flavour group. Then, looking at "Tea and coffee" flavour group, its average revenue is comparable to "Fruits and Nuts" group, but its total revenue is far less because it has very less flavours under that category. And lastly, "Bean and other" Category has comparatively very less average revenue generated, although its total revenue is at par with "tea and coffee" group because it consists of a greater number of flavours.

We can make following suggestion to "Happy Cow" stores based on this revenue analysis:

- It should focus more on adding some flavours under "tea and coffee" flavour groups, as it has potential to consistently produce good amount of revenue.
- The company can also think of discontinuing some of the flavours under "Bean and other" flavour group which are not contributing that much to the revenue.
- Also, I would suggest store to continue its focus on "chocolate and caramel" flavour group because it is generating very high average revenues for the store.
- And store can also think of reducing some of the worst performing flavour offerings under "fruits and nuts" group and thereby further increase average revenue from this flavour group.

2. Outliers Detection

2.1. Visualize the total daily sales corresponding to each customer group and determine the outliers in the data based on the one discussed in Lecture 3.

We can prepare following visualization for daily sales from accounted to each customer groups, after excluding zero total sales days, using the tableau desktop:
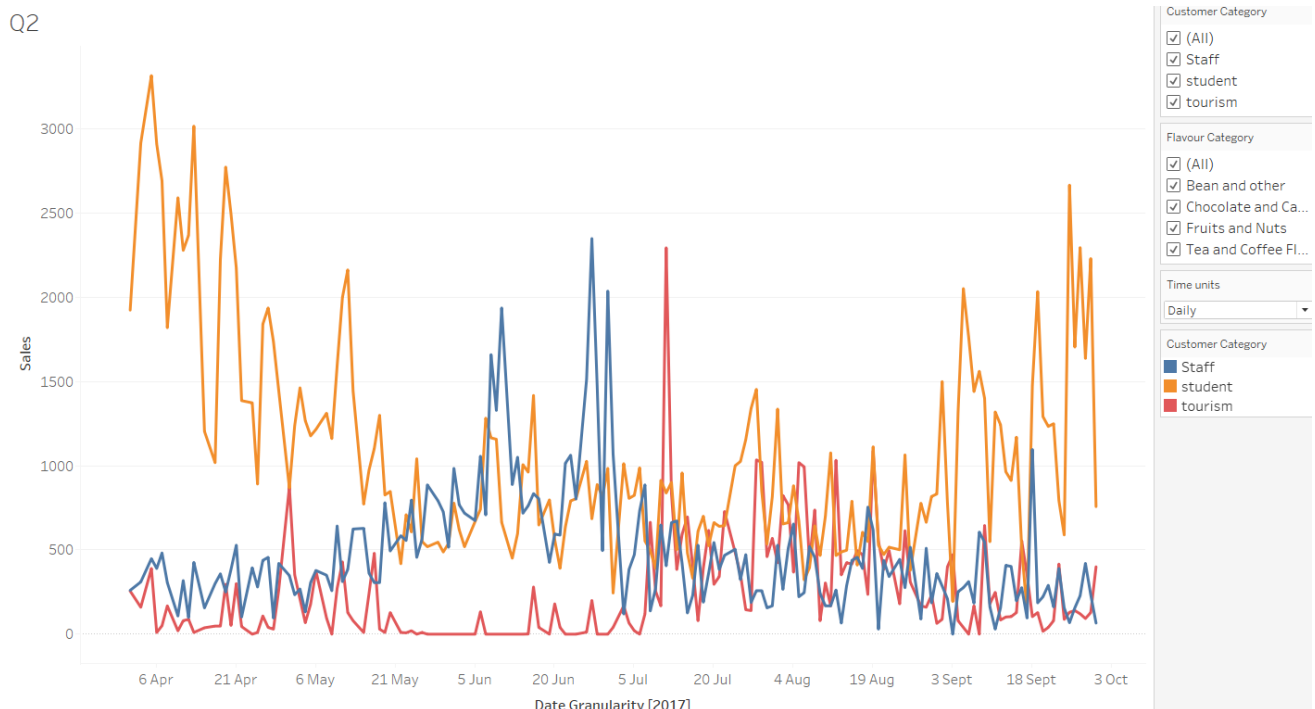

Fig: total daily sales corresponding to each customer group

Now, we can find the outliers in the data using the following method: if the data value is outside the (Q1-1.5IQR, Q3+1.5IQR) range, we will term our data point as an outlier. And, below are the ranges for identifying outliers in each customer group sales:

	Staff	student	tourism
Q1	249	617.75	20
Q3	600	1327.5	384
IQR	351	709.75	364
Lower Bound	-277.5	-446.875	-526
Upper Bound	1126.5	2392.125	930

So, based on above calculation in MS Excel, we identified and highlighted outlier points for each customer group sales, in red colour circles, in the following chart:
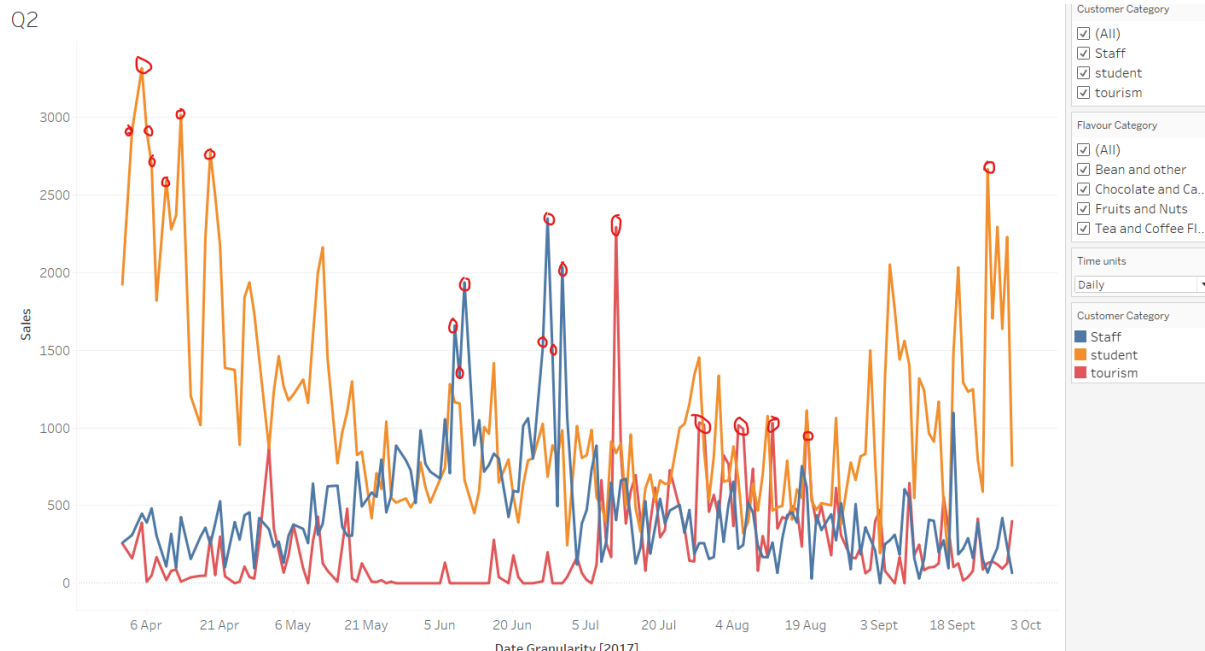
With this, we end our discussion on outliers here, but further investigation on outliers can provide deeper insights on "what" and "why" related to these outliers and they can be treated accordingly in the further data analysis of this case.

Attachments:

"A2_Tableau_File_Advait_Shah.twb"

"HK1227-XLS-ENG (1).xlsx"

"ProcessFlow1.tfl"

"ProcessFlow2.tfl"

"Processed1.xlsx"

"Processed2.xlsx"

References:

[1] H A Mehrizi, eBook: MSCI 719 Winter 2023 Cases Multiple (ID: 9723713) Accessed: Jan. 22, 2023. [Online].

Available: https://www.campusebookstore.com/integration/AccessCodes/default.aspx?permalinkId=ee044bf2-fe82-4db0-ad22-088e81954eef&frame=YES&t=permalink&sid=4u2faw45zyslbp45bbqlpc55