Project Title: Coffee Metrics

Authors: Aishwarya Gupta, Rahul Bhogale

A. Introduction:

Coffee is one of the most consumed beverages in the world, and the coffee chain industry has grown rapidly in recent years. The Coffee Chains dataset, available on Kaggle (Coffee Chains Dataset | Kaggle), presents a wealth of data that provides a comprehensive overview of the performance and market trends of a sample of coffee chains across the United States. This dataset is a valuable resource for coffee chain operators and investors who seek to gain insights into market dynamics, identify areas for improvement, and make informed decisions to grow their business.

The dataset contains over 4,000 rows and 20 columns of information, covering a range of key variables including Area Code, Date, Market, Market Size, Product, Product Line, Product Type, State, Type, Budget COGS, Budget Margin, Budget Profit, Budget Sales, COGS, Inventory, Margin, Marketing, Profit, Sales, and Total Expenses. This level of granularity allows for thorough analysis of distinct aspects of the business and enables users to gain a deep understanding of how each coffee chain is performing.

By analyzing the Market and Market Size variables, coffee chain operators and investors can gain insights into market trends, such as which markets are growing, and which ones are not performing as well. They can also evaluate the size of the market, which provides a basis for assessing potential revenue opportunities.

Furthermore, the dataset provides information on the performance of assorted products, including Product, Product Line, and Product Type. This information can be used to identify top-performing products and product lines and to develop effective marketing and promotional strategies to drive sales of these products.

The dataset also offers valuable financial information on coffee chains. Variables such as Budget COGS, Budget Margin, Budget Profit, Budget Sales, COGS, Inventory, Margin, Marketing, Profit, Sales, and Total Expenses can be used to evaluate the financial performance of coffee chains. By analyzing this information, users can identify areas of the business that are not performing well and develop strategies to improve profitability.

In conclusion, the Coffee Chains dataset is an invaluable resource for coffee chain operators and investors looking to gain insights into market trends, product performance, and financial performance. The detailed information included in the dataset allows for informed decision-making and the development of effective strategies to grow the business. Follow the link to access the full dataset and begin exploring the trends and patterns in the coffee chain industry.

B. Data Description:

The Coffee Chain Dataset has over 4,000 rows and 20 columns with their field name and data description mentioned as below.

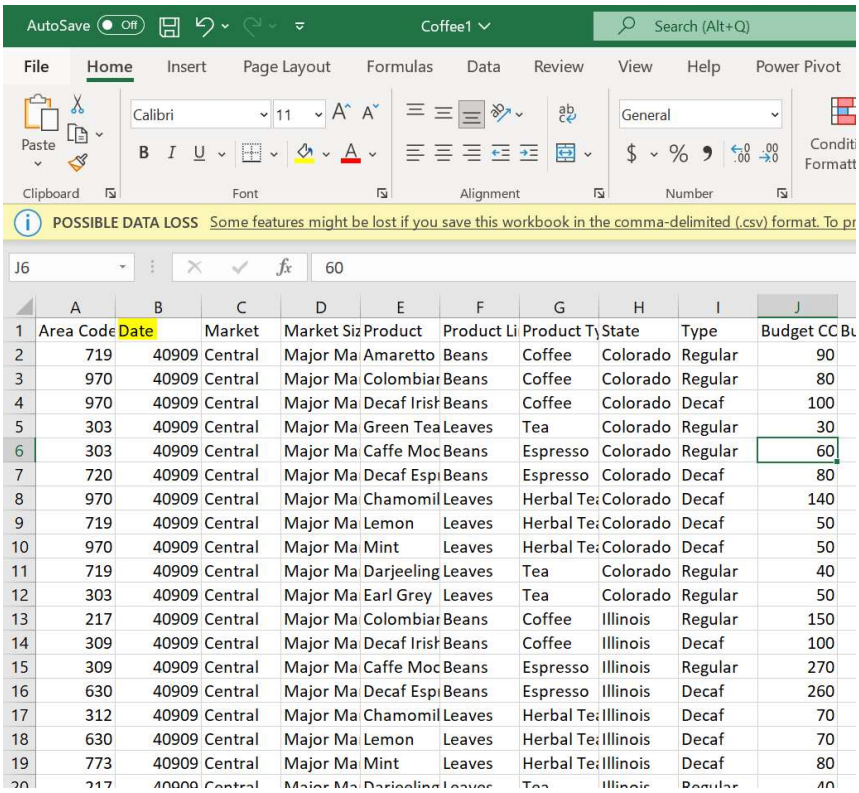
Field Name	Data Description
1. Area Code	A unique identifier for a specific geographic area.
2. Date	The date when the sales data was recorded.
3. Market	The geographic market in which the coffee chain operates.
4. Market Size	The size of the market in which the coffee chain operates.
5. Product	The name of the product sold by the coffee chain.
6. Product Line	The specific line of products to which the product belongs.
7. Product Type	The category of the product, such as hot or cold beverages, snacks, etc.
8. State	The state in which the coffee chain operates.
9. Type	The type of coffee chain, such as franchise or company owned.
10. Budget COGS	The estimated cost of goods sold for a given period.
11. Budget Margin	The estimated margin for a given period.

12. Budget Profit	The estimated profit for a given period.
13. Budget Sales	The estimated sales for a given period.
14. COGS	The actual cost of goods sold for a given period.
15. Inventory	The inventory level for a given period.
16. Margin	The actual margin for a given period.
17. Marketing	The marketing expenses for a given period.
18. Profit	The actual profit for a given period.
19. Sales	The actual sales for a given period.
20. Total Expenses	The total expenses for a given period.

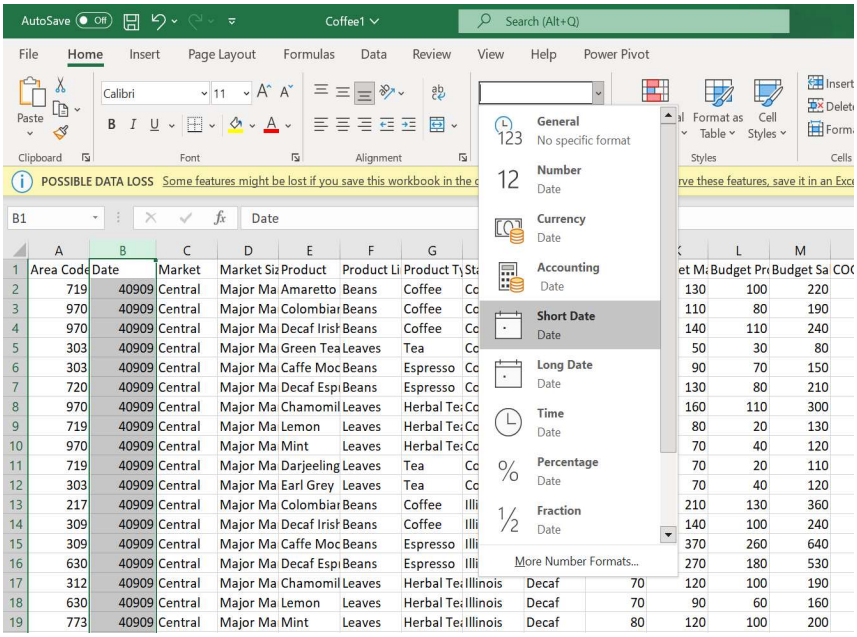
C. Data Cleaning:

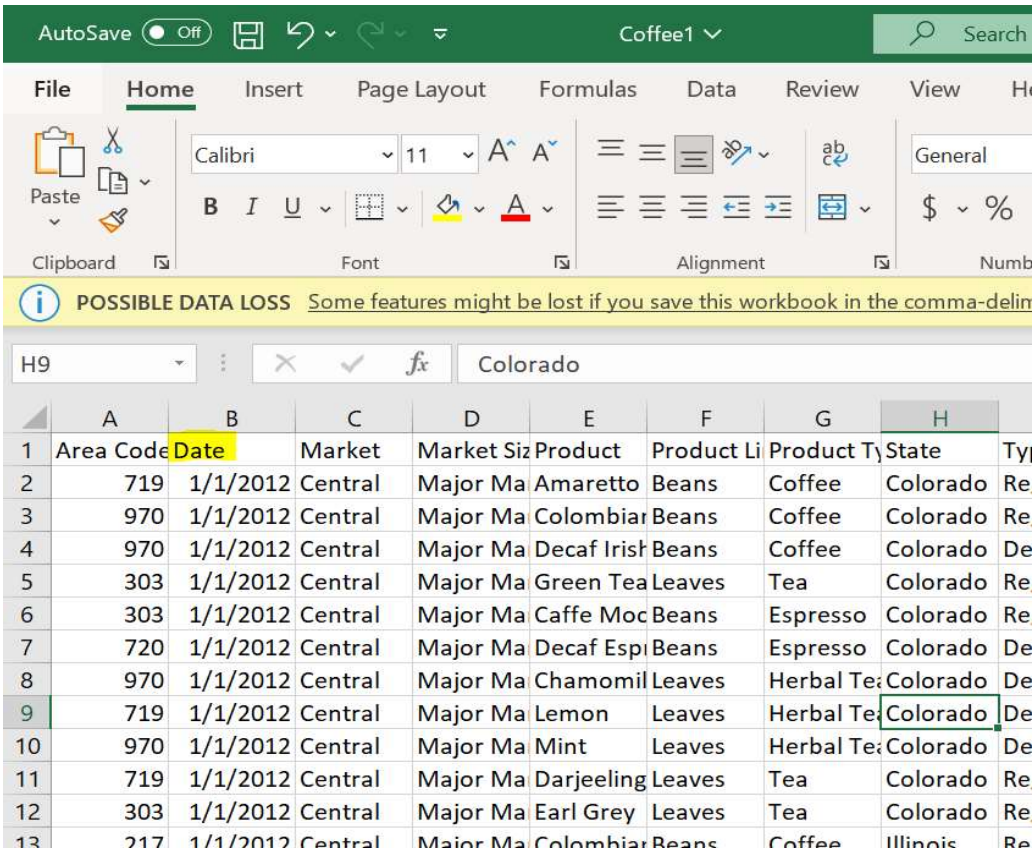
1. Data type Formatting:

Converting the Date column into date as datatype from number value



	A	B	C	D	E	F	G	H	I	J
1	Area Code	Date	Market	Market Siz	Product	Product Li	Product Ty	State	Type	Budget CC
2	719	40909	Central	Major Ma	Amaretto	Beans	Coffee	Colorado	Regular	90
3	970	40909	Central	Major Ma	Colombiar	Beans	Coffee	Colorado	Regular	80
4	970	40909	Central	Major Ma	Decaf Irish	Beans	Coffee	Colorado	Decaf	100
5	303	40909	Central	Major Ma	Green Tea	Leaves	Tea	Colorado	Regular	30
6	303	40909	Central	Major Ma	Caffe Moc	Beans	Espresso	Colorado	Regular	60
7	720	40909	Central	Major Ma	Decaf Espi	Beans	Espresso	Colorado	Decaf	80
8	970	40909	Central	Major Ma	Chamomil	Leaves	Herbal Tea	Colorado	Decaf	140
9	719	40909	Central	Major Ma	Lemon	Leaves	Herbal Tea	Colorado	Decaf	50
10	970	40909	Central	Major Ma	Mint	Leaves	Herbal Tea	Colorado	Decaf	50
11	719	40909	Central	Major Ma	Darjeeling	Leaves	Tea	Colorado	Regular	40
12	303	40909	Central	Major Ma	Earl Grey	Leaves	Tea	Colorado	Regular	50
13	217	40909	Central	Major Ma	Colombiar	Beans	Coffee	Illinois	Regular	150
14	309	40909	Central	Major Ma	Decaf Irish	Beans	Coffee	Illinois	Decaf	100
15	309	40909	Central	Major Ma	Caffe Moc	Beans	Espresso	Illinois	Regular	270
16	630	40909	Central	Major Ma	Decaf Espi	Beans	Espresso	Illinois	Decaf	260
17	312	40909	Central	Major Ma	Chamomil	Leaves	Herbal Tea	Illinois	Decaf	70
18	630	40909	Central	Major Ma	Lemon	Leaves	Herbal Tea	Illinois	Decaf	70
19	773	40909	Central	Major Ma	Mint	Leaves	Herbal Tea	Illinois	Decaf	80
20	217	40909	Central	Major Ma	Darjeeling	Leaves	Tea	Illinois	Regular	40





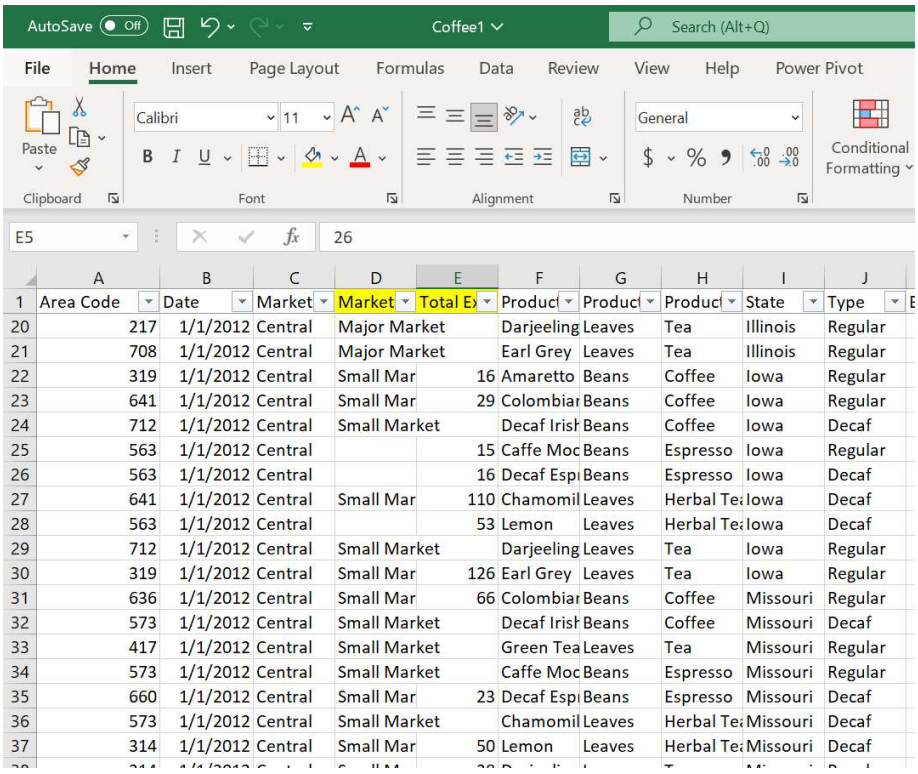
	A	B	C	D	E	F	G	H	
1	Area Code	Date	Market	Market Siz	Product	Product Li	Product Ty	State	Ty
2	719	1/1/2012	Central	Major Ma	Amaretto	Beans	Coffee	Colorado	Re
3	970	1/1/2012	Central	Major Ma	Colombiar	Beans	Coffee	Colorado	Re
4	970	1/1/2012	Central	Major Ma	Decaf Irish	Beans	Coffee	Colorado	De
5	303	1/1/2012	Central	Major Ma	Green Tea	Leaves	Tea	Colorado	Re
6	303	1/1/2012	Central	Major Ma	Caffe Moc	Beans	Espresso	Colorado	Re
7	720	1/1/2012	Central	Major Ma	Decaf Espi	Beans	Espresso	Colorado	De
8	970	1/1/2012	Central	Major Ma	Chamomil	Leaves	Herbal Te	Colorado	De
9	719	1/1/2012	Central	Major Ma	Lemon	Leaves	Herbal Te	Colorado	De
10	970	1/1/2012	Central	Major Ma	Mint	Leaves	Herbal Te	Colorado	De
11	719	1/1/2012	Central	Major Ma	Darjeeling	Leaves	Tea	Colorado	Re
12	303	1/1/2012	Central	Major Ma	Earl Grey	Leaves	Tea	Colorado	Re
13	217	1/1/2012	Central	Major Ma	Colombiar	Beans	Coffee	Illinois	Re

Date as date datatype is used for Yearly, Quarterly, Monthly, Daily analysis for further visualization on Tableau.

2. Missing Values:

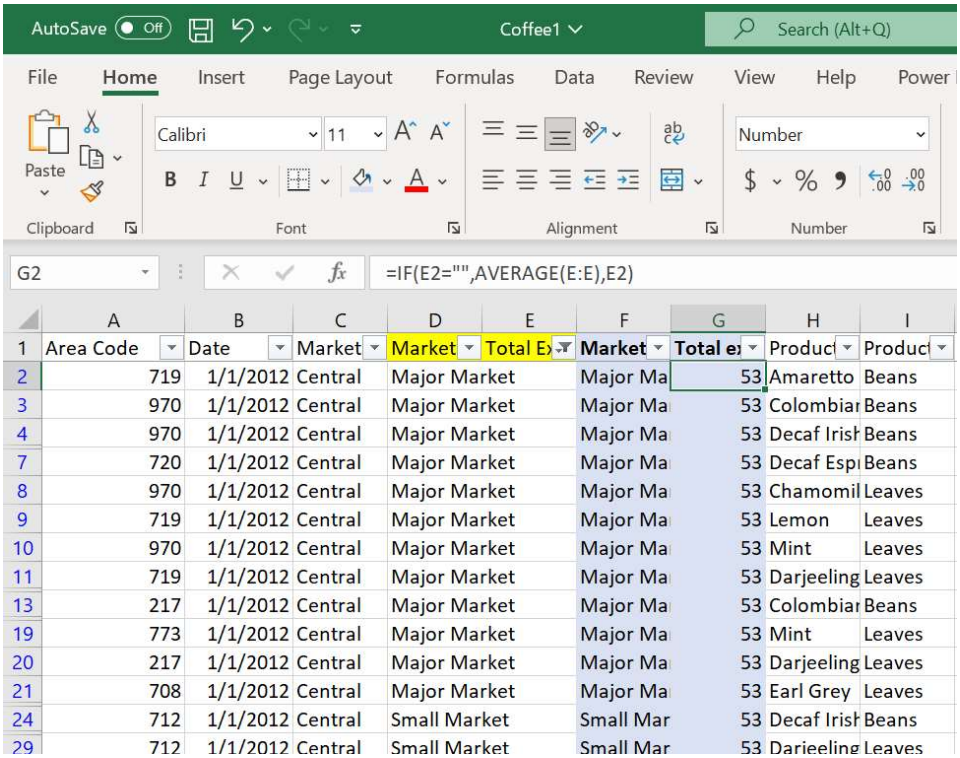
If the data type of a ***column is measure***, replace missing null values by ***Average function***.

If the data type of a ***column is dimension***, replace missing null values by ***Random function.***



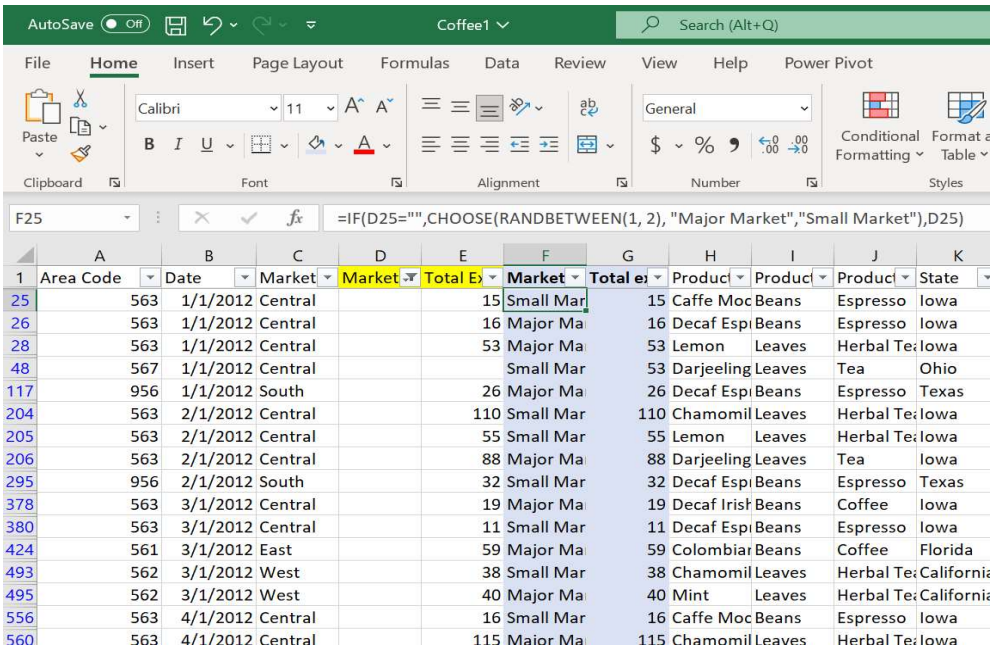
	A	B	C	D	E	F	G	H	I	J
1	Area Code	Date	Market	Market	Total Ex	Produc	Produc	Produc	State	Type
20	217	1/1/2012	Central	Major Market		Darjeeling	Leaves	Tea	Illinois	Regular
21	708	1/1/2012	Central	Major Market		Earl Grey	Leaves	Tea	Illinois	Regular
22	319	1/1/2012	Central	Small Mar	16	Amaretto	Beans	Coffee	Iowa	Regular
23	641	1/1/2012	Central	Small Mar	29	Colombiar	Beans	Coffee	Iowa	Regular
24	712	1/1/2012	Central	Small Market		Decaf Irish	Beans	Coffee	Iowa	Decaf
25	563	1/1/2012	Central		15	Caffe Moc	Beans	Espresso	Iowa	Regular
26	563	1/1/2012	Central		16	Decaf Espı	Beans	Espresso	Iowa	Decaf
27	641	1/1/2012	Central	Small Mar	110	Chamomil	Leaves	Herbal Te	Iowa	Decaf
28	563	1/1/2012	Central		53	Lemon	Leaves	Herbal Te	Iowa	Decaf
29	712	1/1/2012	Central	Small Market		Darjeeling	Leaves	Tea	Iowa	Regular
30	319	1/1/2012	Central	Small Mar	126	Earl Grey	Leaves	Tea	Iowa	Regular
31	636	1/1/2012	Central	Small Mar	66	Colombiar	Beans	Coffee	Missouri	Regular
32	573	1/1/2012	Central	Small Market		Decaf Irish	Beans	Coffee	Missouri	Decaf
33	417	1/1/2012	Central	Small Market		Green Tea	Leaves	Tea	Missouri	Regular
34	573	1/1/2012	Central	Small Market		Caffe Moc	Beans	Espresso	Missouri	Regular
35	660	1/1/2012	Central	Small Mar	23	Decaf Espı	Beans	Espresso	Missouri	Decaf
36	573	1/1/2012	Central	Small Market		Chamomil	Leaves	Herbal Te	Missouri	Decaf
37	314	1/1/2012	Central	Small Mar	50	Lemon	Leaves	Herbal Te	Missouri	Decaf

Filter the blank values from Data Tab >> Filter >> Select Total Expense column >> filter blank values >> Use Average function



	A	B	C	D	E	F	G	H	I
1	Area Code	Date	Market	Market	Total E	Market	Total e	Produc	Produc
2	719	1/1/2012	Central	Major Market		Major Ma	53	Amaretto	Beans
3	970	1/1/2012	Central	Major Market		Major Ma	53	Colombiar	Beans
4	970	1/1/2012	Central	Major Market		Major Ma	53	Decaf Irish	Beans
7	720	1/1/2012	Central	Major Market		Major Ma	53	Decaf Espı	Beans
8	970	1/1/2012	Central	Major Market		Major Ma	53	Chamomil	Leaves
9	719	1/1/2012	Central	Major Market		Major Ma	53	Lemon	Leaves
10	970	1/1/2012	Central	Major Market		Major Ma	53	Mint	Leaves
11	719	1/1/2012	Central	Major Market		Major Ma	53	Darjeeling	Leaves
13	217	1/1/2012	Central	Major Market		Major Ma	53	Colombiar	Beans
19	773	1/1/2012	Central	Major Market		Major Ma	53	Mint	Leaves
20	217	1/1/2012	Central	Major Market		Major Ma	53	Darjeeling	Leaves
21	708	1/1/2012	Central	Major Market		Major Ma	53	Earl Grey	Leaves
24	712	1/1/2012	Central	Small Market		Small Mar	53	Decaf Irish	Beans
29	712	1/1/2012	Central	Small Market		Small Mar	53	Darjeeling	Leaves

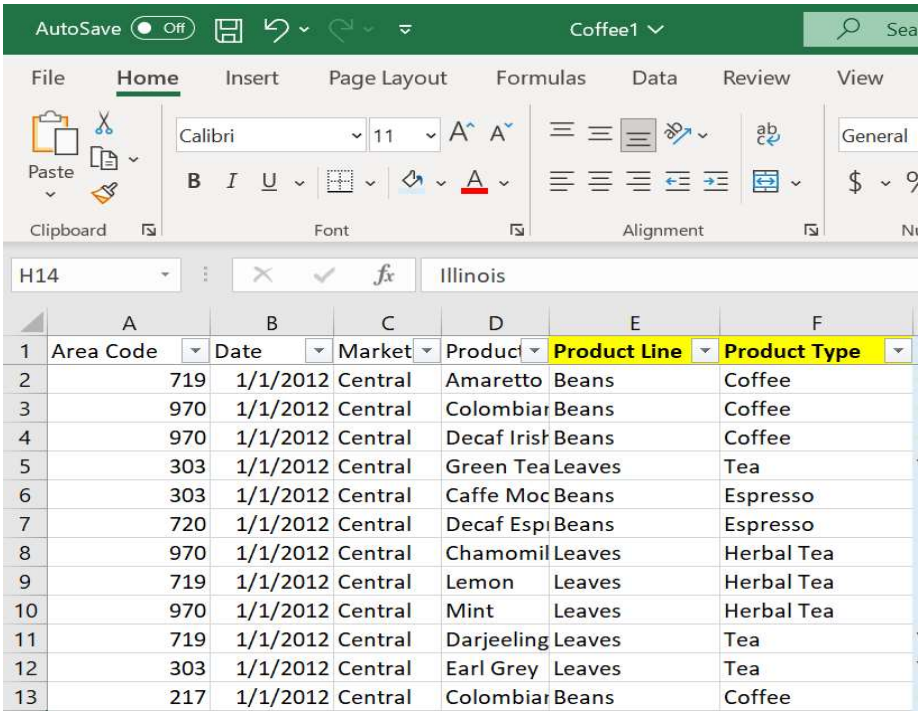
Filter the blank values from Data Tab >> Filter >> Select Market column >> filter blank values >> Use Average function



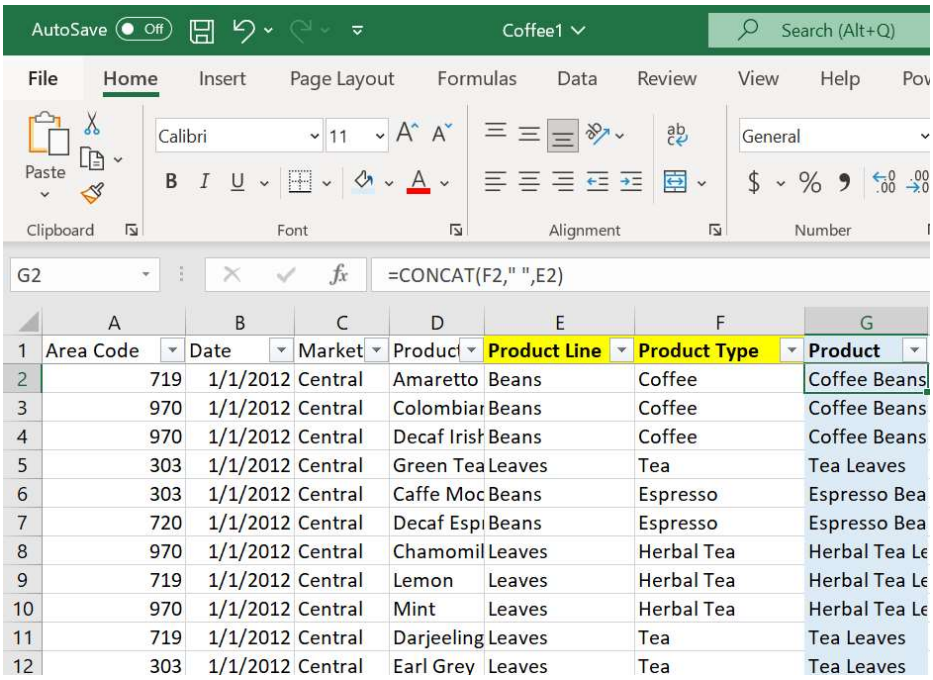
	A	B	C	D	E	F	G	H	I	J	K
1	Area Code	Date	Market	Market	Total E	Market	Total e	Produc	Produc	Produc	State
25	563	1/1/2012	Central		15	Small Mar	15	Caffe Moc	Beans	Espresso	Iowa
26	563	1/1/2012	Central		16	Major Ma	16	Decaf Espı	Beans	Espresso	Iowa
28	563	1/1/2012	Central		53	Major Ma	53	Lemon	Leaves	Herbal Tea	Iowa
48	567	1/1/2012	Central			Small Mar	53	Darjeeling	Leaves	Tea	Ohio
117	956	1/1/2012	South		26	Major Ma	26	Decaf Espı	Beans	Espresso	Texas
204	563	2/1/2012	Central		110	Small Mar	110	Chamomil	Leaves	Herbal Tea	Iowa
205	563	2/1/2012	Central		55	Small Mar	55	Lemon	Leaves	Herbal Tea	Iowa
206	563	2/1/2012	Central		88	Major Ma	88	Darjeeling	Leaves	Tea	Iowa
295	956	2/1/2012	South		32	Small Mar	32	Decaf Espı	Beans	Espresso	Texas
378	563	3/1/2012	Central		19	Major Ma	19	Decaf Irish	Beans	Coffee	Iowa
380	563	3/1/2012	Central		11	Small Mar	11	Decaf Espı	Beans	Espresso	Iowa
424	561	3/1/2012	East		59	Major Ma	59	Colombiar	Beans	Coffee	Florida
493	562	3/1/2012	West		38	Small Mar	38	Chamomil	Leaves	Herbal Tea	California
495	562	3/1/2012	West		40	Major Ma	40	Mint	Leaves	Herbal Tea	California
556	563	4/1/2012	Central		16	Small Mar	16	Caffe Moc	Beans	Espresso	Iowa
560	563	4/1/2012	Central		115	Major Ma	115	Chamomil	Leaves	Herbal Tea	Iowa

From the above snapshots, Random values (dimensions) and Average (measures) are replacing the missing values to perform unbiased analysis and visualization.

3. Combining Columns:



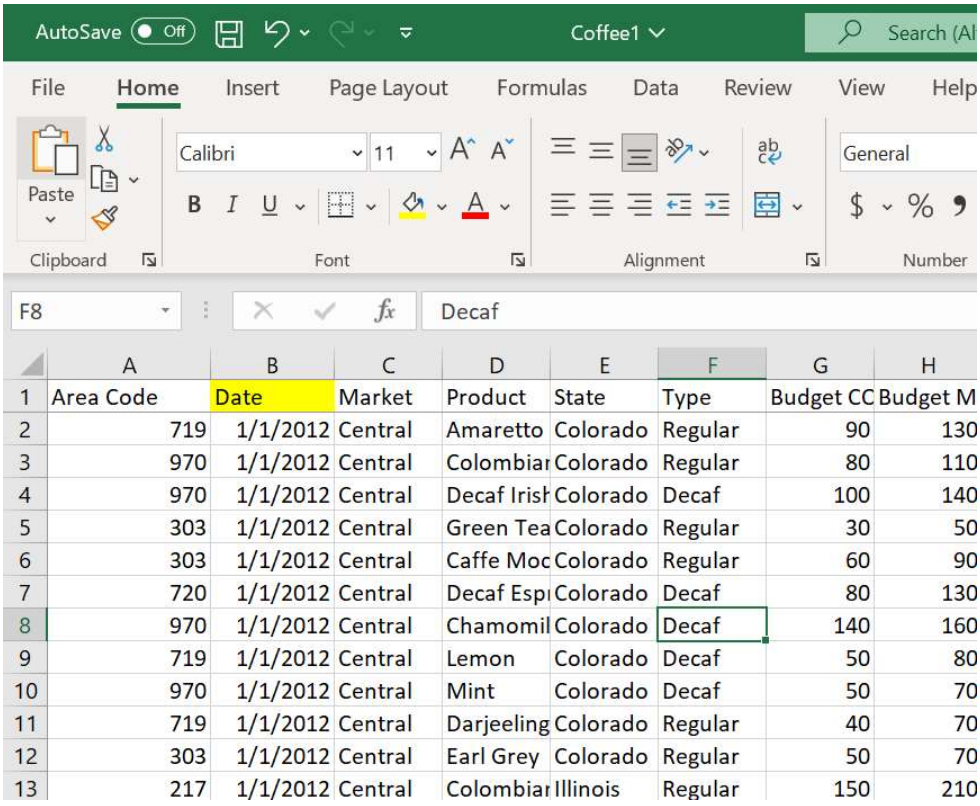
	A	B	C	D	E	F
1	Area Code	Date	Market	Product	Product Line	Product Type
2	719	1/1/2012	Central	Amaretto	Beans	Coffee
3	970	1/1/2012	Central	Colombian	Beans	Coffee
4	970	1/1/2012	Central	Decaf Irish	Beans	Coffee
5	303	1/1/2012	Central	Green Tea	Leaves	Tea
6	303	1/1/2012	Central	Caffe Moc	Beans	Espresso
7	720	1/1/2012	Central	Decaf Espr	Beans	Espresso
8	970	1/1/2012	Central	Chamomil	Leaves	Herbal Tea
9	719	1/1/2012	Central	Lemon	Leaves	Herbal Tea
10	970	1/1/2012	Central	Mint	Leaves	Herbal Tea
11	719	1/1/2012	Central	Darjeeling	Leaves	Tea
12	303	1/1/2012	Central	Earl Grey	Leaves	Tea
13	217	1/1/2012	Central	Colombian	Beans	Coffee



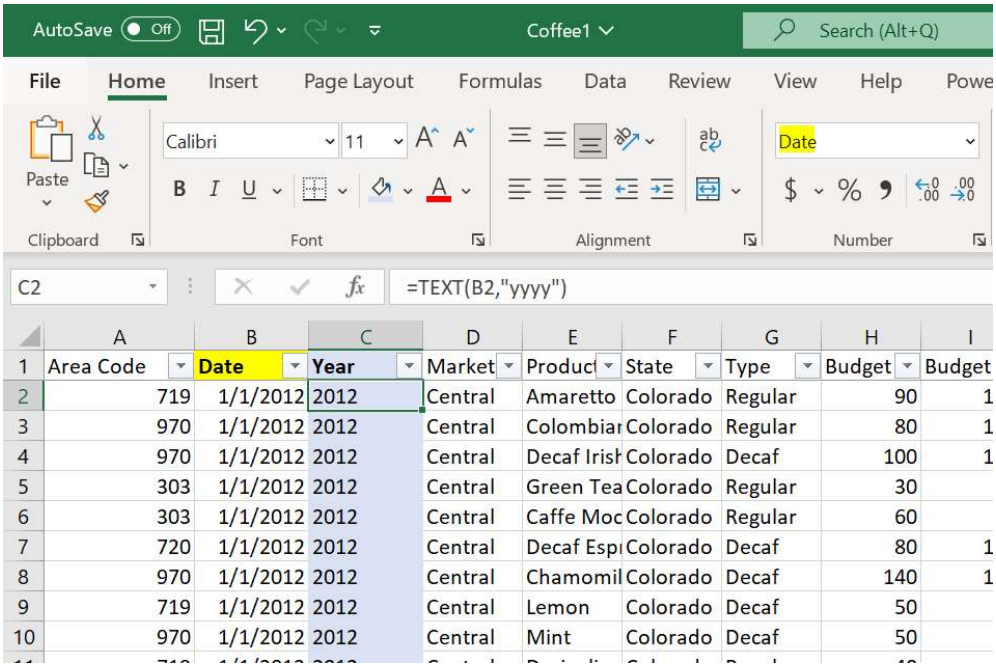
	A	B	C	D	E	F	G
1	Area Code	Date	Market	Product	Product Line	Product Type	Product
2	719	1/1/2012	Central	Amaretto	Beans	Coffee	Coffee Beans
3	970	1/1/2012	Central	Colombian	Beans	Coffee	Coffee Beans
4	970	1/1/2012	Central	Decaf Irish	Beans	Coffee	Coffee Beans
5	303	1/1/2012	Central	Green Tea	Leaves	Tea	Tea Leaves
6	303	1/1/2012	Central	Caffe Moc	Beans	Espresso	Espresso Bea
7	720	1/1/2012	Central	Decaf Espr	Beans	Espresso	Espresso Bea
8	970	1/1/2012	Central	Chamomil	Leaves	Herbal Tea	Herbal Tea Le
9	719	1/1/2012	Central	Lemon	Leaves	Herbal Tea	Herbal Tea Le
10	970	1/1/2012	Central	Mint	Leaves	Herbal Tea	Herbal Tea Le
11	719	1/1/2012	Central	Darjeeling	Leaves	Tea	Tea Leaves
12	303	1/1/2012	Central	Earl Grey	Leaves	Tea	Tea Leaves

Using concatenation, we can combine the above two columns Product line and Product type as shown using the formula.

4. Splitting Columns:



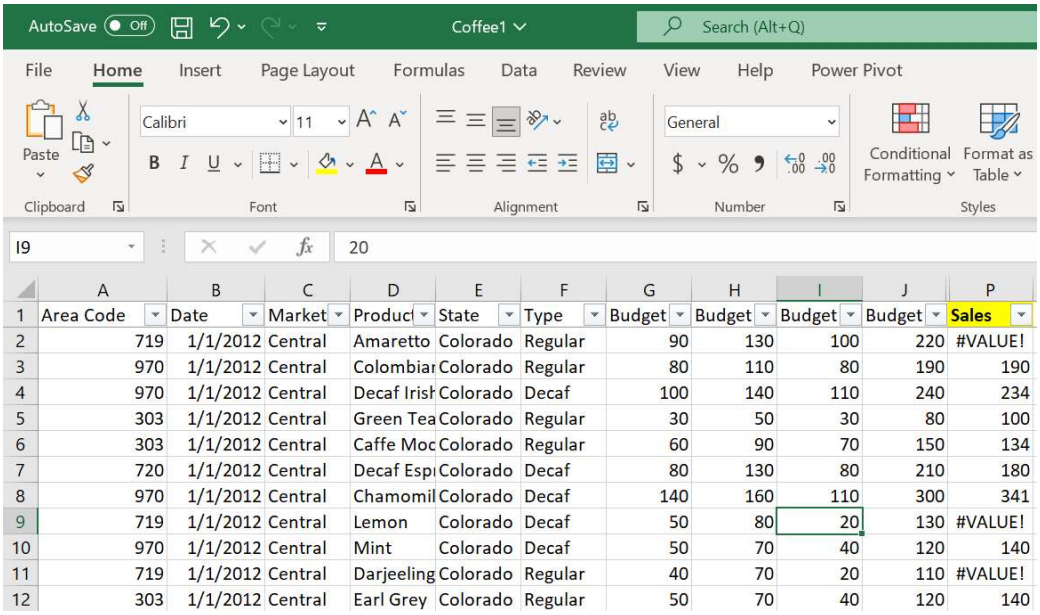
	A	B	C	D	E	F	G	H
1	Area Code	Date	Market	Product	State	Type	Budget CC	Budget M
2	719	1/1/2012	Central	Amaretto	Colorado	Regular	90	130
3	970	1/1/2012	Central	Colombiar	Colorado	Regular	80	110
4	970	1/1/2012	Central	Decaf Irish	Colorado	Decaf	100	140
5	303	1/1/2012	Central	Green Tea	Colorado	Regular	30	50
6	303	1/1/2012	Central	Caffe Moc	Colorado	Regular	60	90
7	720	1/1/2012	Central	Decaf Espı	Colorado	Decaf	80	130
8	970	1/1/2012	Central	Chamomil	Colorado	Decaf	140	160
9	719	1/1/2012	Central	Lemon	Colorado	Decaf	50	80
10	970	1/1/2012	Central	Mint	Colorado	Decaf	50	70
11	719	1/1/2012	Central	Darjeeling	Colorado	Regular	40	70
12	303	1/1/2012	Central	Earl Grey	Colorado	Regular	50	70
13	217	1/1/2012	Central	Colombiar	Illinois	Regular	150	210



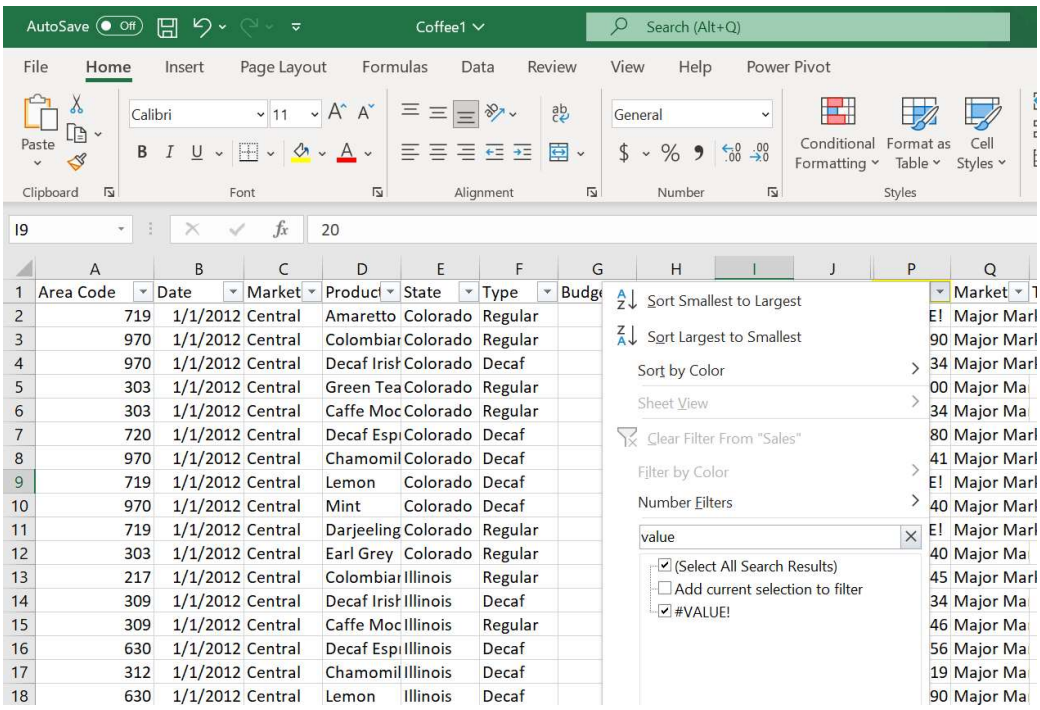
	A	B	C	D	E	F	G	H	I
1	Area Code	Date	Year	Market	Product	State	Type	Budget	Budget
2	719	1/1/2012	2012	Central	Amaretto	Colorado	Regular	90	1
3	970	1/1/2012	2012	Central	Colombiar	Colorado	Regular	80	1
4	970	1/1/2012	2012	Central	Decaf Irish	Colorado	Decaf	100	1
5	303	1/1/2012	2012	Central	Green Tea	Colorado	Regular	30	
6	303	1/1/2012	2012	Central	Caffe Moc	Colorado	Regular	60	
7	720	1/1/2012	2012	Central	Decaf Espı	Colorado	Decaf	80	1
8	970	1/1/2012	2012	Central	Chamomil	Colorado	Decaf	140	1
9	719	1/1/2012	2012	Central	Lemon	Colorado	Decaf	50	
10	970	1/1/2012	2012	Central	Mint	Colorado	Decaf	50	

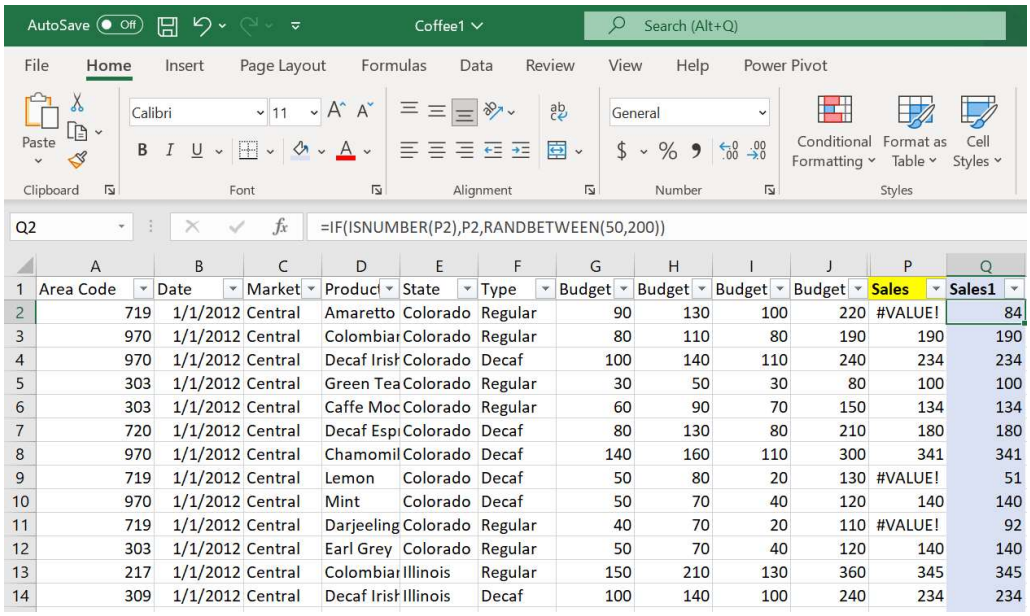
Extract Year from Date column as shown above using Text formatting and later converting into date format for the data analysis.

5. Invalid Data:



	A	B	C	D	E	F	G	H	I	J	P
1	Area Code	Date	Market	Product	State	Type	Budget	Budget	Budget	Budget	Sales
2	719	1/1/2012	Central	Amaretto	Colorado	Regular	90	130	100	220	#VALUE!
3	970	1/1/2012	Central	Colombian	Colorado	Regular	80	110	80	190	190
4	970	1/1/2012	Central	Decaf Irish	Colorado	Decaf	100	140	110	240	234
5	303	1/1/2012	Central	Green Tea	Colorado	Regular	30	50	30	80	100
6	303	1/1/2012	Central	Caffe Moc	Colorado	Regular	60	90	70	150	134
7	720	1/1/2012	Central	Decaf Espi	Colorado	Decaf	80	130	80	210	180
8	970	1/1/2012	Central	Chamomil	Colorado	Decaf	140	160	110	300	341
9	719	1/1/2012	Central	Lemon	Colorado	Decaf	50	80	20	130	#VALUE!
10	970	1/1/2012	Central	Mint	Colorado	Decaf	50	70	40	120	140
11	719	1/1/2012	Central	Darjeeling	Colorado	Regular	40	70	20	110	#VALUE!
12	303	1/1/2012	Central	Earl Grey	Colorado	Regular	50	70	40	120	140





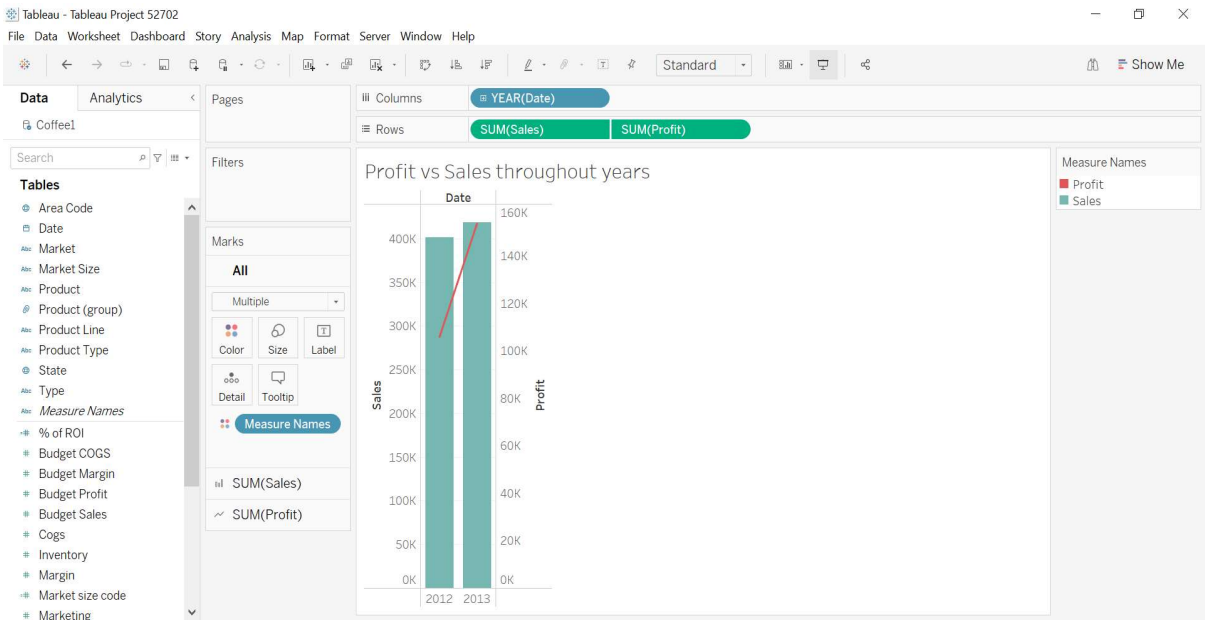
	A	B	C	D	E	F	G	H	I	J	P	Q
1	Area Code	Date	Market	Produc	State	Type	Budget	Budget	Budget	Budget	Sales	Sales1
2	719	1/1/2012	Central	Amaretto	Colorado	Regular	90	130	100	220	#VALUE!	84
3	970	1/1/2012	Central	Colombiar	Colorado	Regular	80	110	80	190	190	190
4	970	1/1/2012	Central	Decaf Irish	Colorado	Decaf	100	140	110	240	234	234
5	303	1/1/2012	Central	Green Tea	Colorado	Regular	30	50	30	80	100	100
6	303	1/1/2012	Central	Caffe Moc	Colorado	Regular	60	90	70	150	134	134
7	720	1/1/2012	Central	Decaf Espi	Colorado	Decaf	80	130	80	210	180	180
8	970	1/1/2012	Central	Chamomil	Colorado	Decaf	140	160	110	300	341	341
9	719	1/1/2012	Central	Lemon	Colorado	Decaf	50	80	20	130	#VALUE!	51
10	970	1/1/2012	Central	Mint	Colorado	Decaf	50	70	40	120	140	140
11	719	1/1/2012	Central	Darjeeling	Colorado	Regular	40	70	20	110	#VALUE!	92
12	303	1/1/2012	Central	Earl Grey	Colorado	Regular	50	70	40	120	140	140
13	217	1/1/2012	Central	Colombiar	Illinois	Regular	150	210	130	360	345	345
14	309	1/1/2012	Central	Decaf Irish	Illinois	Decaf	100	140	100	240	234	234

Invalid Data could be a number value in date or text field, special characters, or any error values such as **#VALUE!, #NAME?** as shown above etc. could be replaced using ***Is Number function paired with Random*** no's generated between 50 and 200

D. Data Visualization:

1. How does the trend of coffee sales and profit vary over the years?

Bar & line graph, dual axis



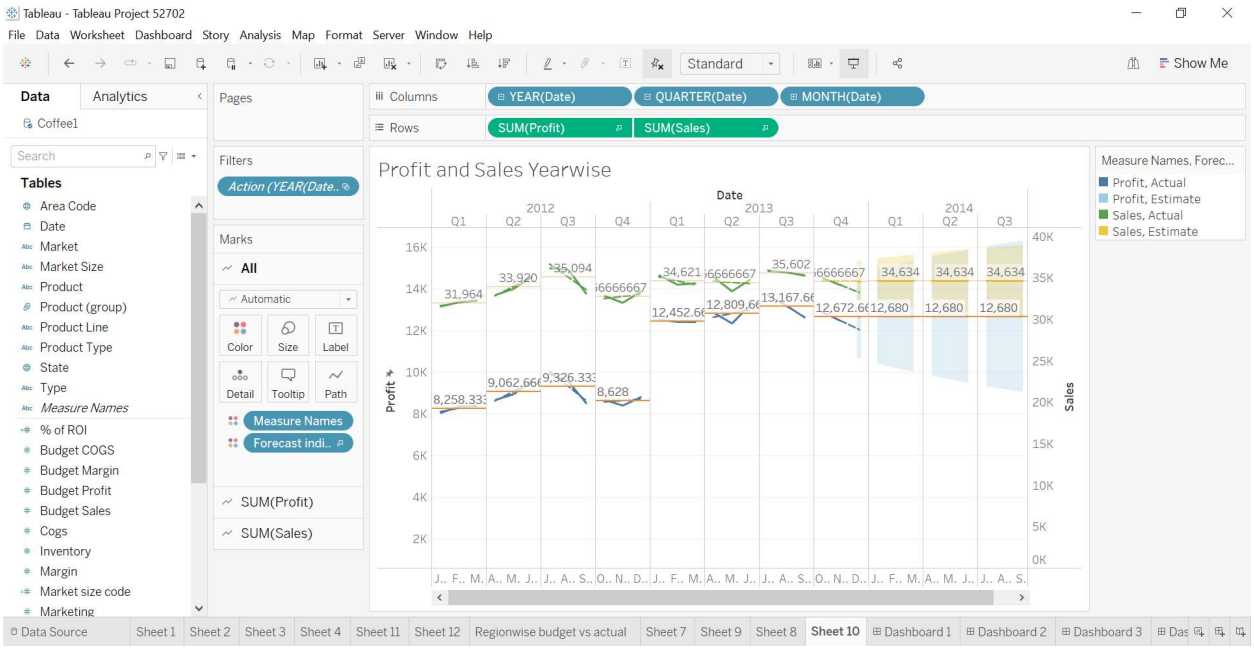
Based on the dual axis bar and line graph provided, we can observe the trend of profit and sales analysis over the period of 2012 and 2013. The green bar graph represents the total sales, while the red line graph represents the total profit. We can see that both the sales and profit have gradually increased over these two years.

From the graph, there is a positive correlation between sales and profit, as they are linearly proportional to each other. This suggests that an increase in sales has resulted in an increase in profit.

In conclusion, this graph provides valuable insights into the performance of the company over the two-year period and suggests a positive trend in both sales and profit.

2. **What will be the forecasted sales and profit coffee trend for the following year?**

Line graph, dual axis, Forecast, Trend lines, Reference lines, Dates



From the above chart, it is quite evident that we are forecasting the overall sales and profit over the years 2012 and 2013 and projected the respective values for 2014 in quarterly manner.

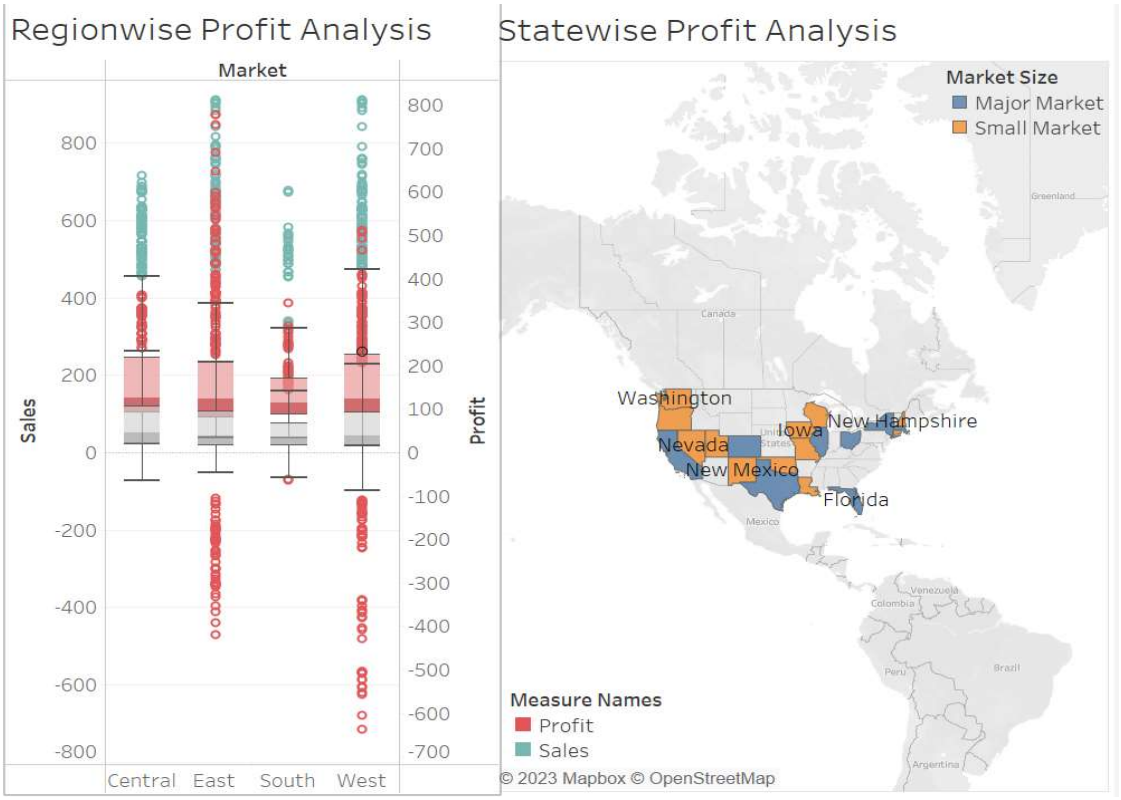
Here, we have used multiple concepts such as dual axis for two different measures (sales and profit), line graph, created trend lines and reference lines which are highlighted with dotted lines and average values as follow.

From the above data visualization, we interpret the results as follows:

1. The total forecasted profit value would subsequently increase for quarters of 2014 and would range from 8,000 to 16,000 shown in blue.
2. The overall forecasted sales value would range between 30,000 to 40,000 for 2014 quarters is shown in yellow.

3. **How Sales and Profit varies across different region and states?**

Box & Whiskers plot, Geographical plot, Dual axis



To gain a comprehensive understanding of the performance of different regions and states, we have analyzed profit and sales using a combination of Box and Whiskers plots with a dual axis and a Geographical map.

Our analysis of the Box plot has yielded the following conclusions:

a. The red box plot representing profit clearly indicates that the South region has performed significantly lower compared to other regions. An outlier is displayed in red color for easy identification.

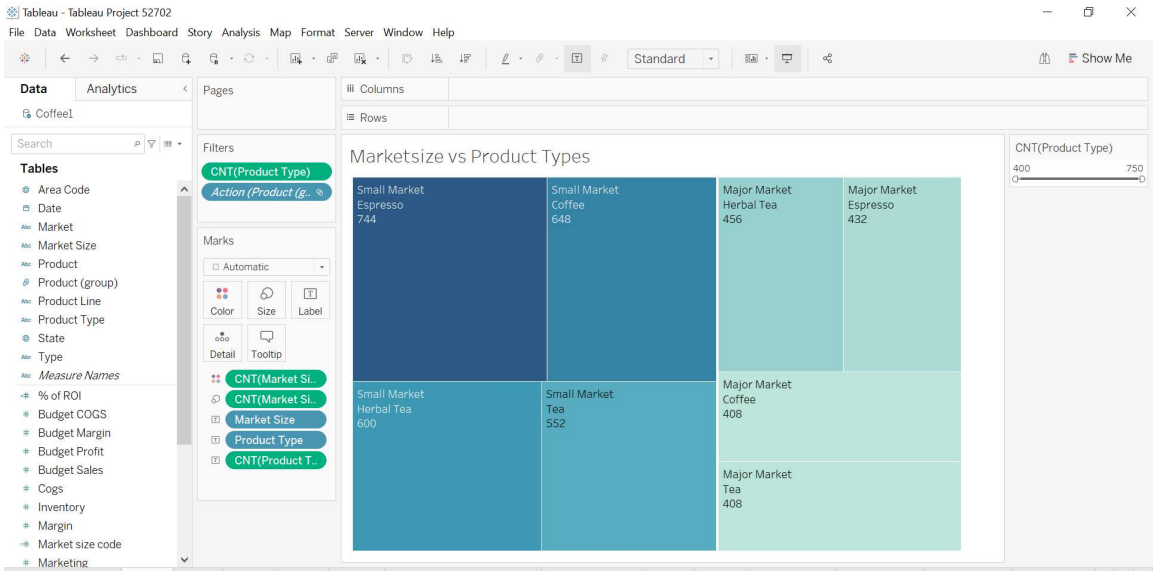
b. The Grey box plot, which depicts sales over different regions, shows that the South region has the least contribution to sales when compared to other regions. An outlier is displayed in green color for easy identification.

Moving on to our analysis of the Geographical map, we have concluded that the market size varies significantly across different states. Our observations indicate that the major market covers fewer states when compared to the smaller markets, which are displayed in blue and yellow for easy reference.

Overall, our analysis using Box and Whiskers plots with a dual axis and a Geographical map provides valuable insights into the performance of different regions and states, helping businesses to make informed decisions based on data-driven evidence.

4. **What is the distribution of product types sold within each market size?**

Heat Map graph

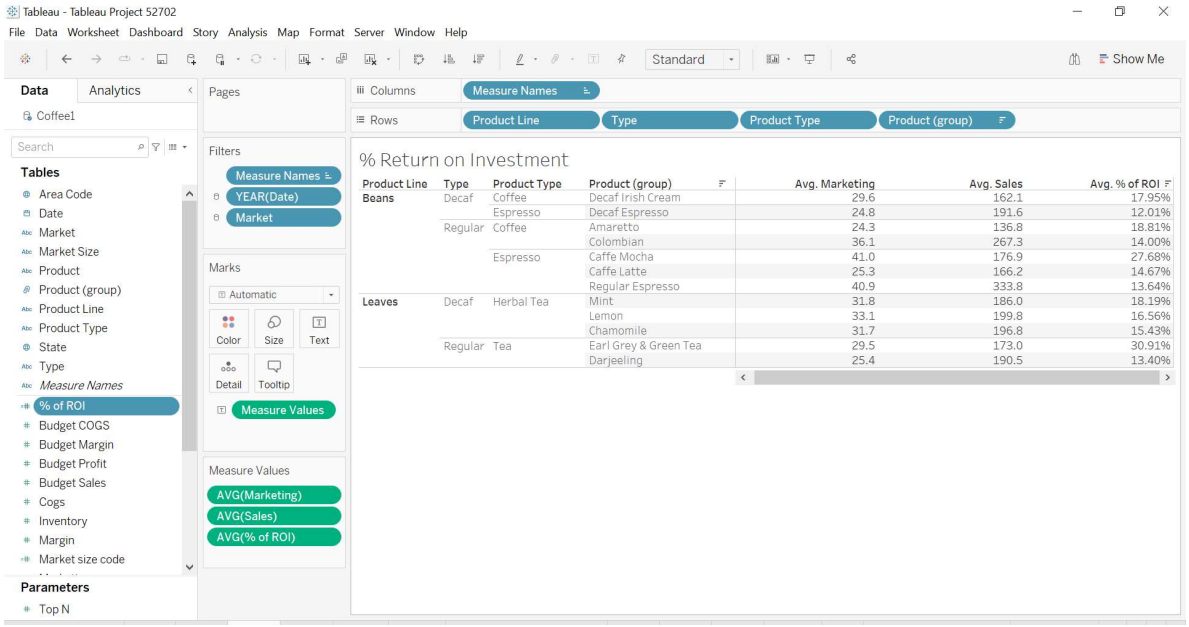


In the provided heat map chart, we have conducted an analysis of the market size distribution across various product types. The size of the box and the intensity of the color gradient, depicted in blue, indicate the relative value of each product type.

Based on our observations, we can infer that the Small market segment has shown the highest sales figures for Espresso, while Tea has been the least preferred product. On the other hand, the Major market segment has witnessed the highest sales for Herbal Tea, with Regular Tea being the least popular among the goods sold.

Overall, this analysis highlights the variations in consumer preferences across different product categories and market segments. Such insights can be valuable for businesses to tailor their marketing strategies and target their products to the right audience.

5. What is the return on investment % over different product lines?

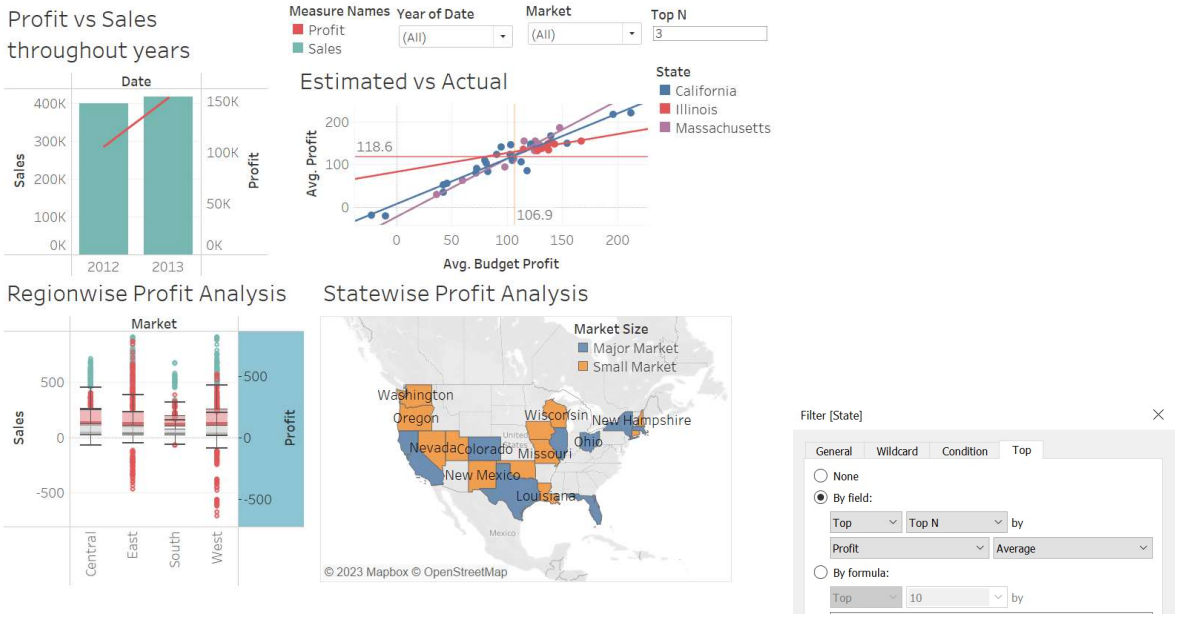


The tabular chart provided demonstrates the analysis of the % Return on Investment metric by comparing the average amount spent on marketing against the average amount of sales made on different coffee products across various product lines and types. This analysis was performed using advanced data manipulation techniques, including Grouping, Calculated fields, and Ranking.

In particular, the "Product" column was used as the basis for grouping, with Early Grey and Green Tea products being combined into a particular group. The *calculated field, "Avg % ROI*," was derived from the average value of the marketing column divided by the average value of the sales column.

The resulting data was then *ranked based on the sorted values of the Avg % ROI field*, from highest to lowest. This approach provides a professional and rigorous method for analyzing the performance of different coffee products in terms of return on investment.

E. Dashboard:



The above dashboard incorporates various analytical concepts such as Dual axis, bar, line, Scatter plot, box & whiskers, geo map, Trendline, Reference line, Rank & Top N parameter. The new graph depicts estimated vs actual values of profit using a Scatter plot, which enables us to analyze the correlation between the two.

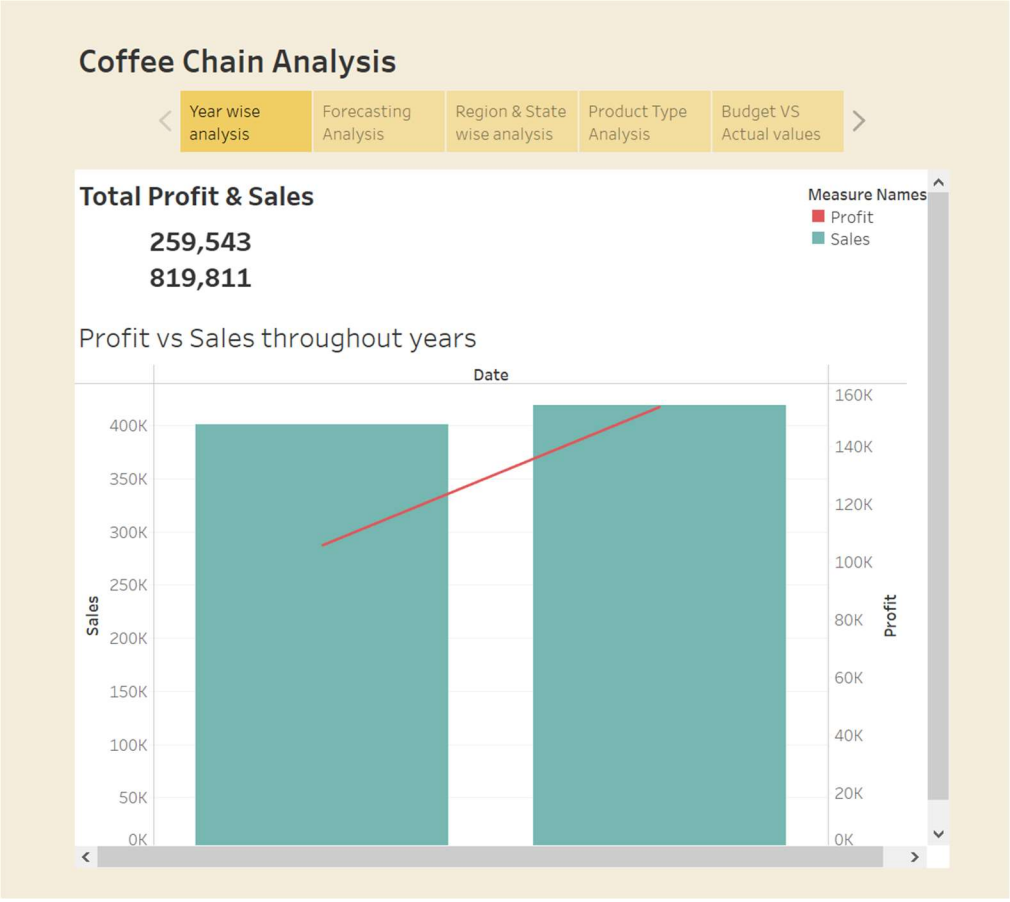
By utilizing the Trendline and Reference line, we can effectively identify how the budget/estimated vs actual values are varying across each data point. In the displayed figure, Massachusetts and Illinois exhibit a high correlation, as the data points fall across the linear regression line, followed by California, where the data points are dispersed for estimated vs actual profit values.

The Top N parameter is utilized to select the top three states with the highest average profit values across the entire USA. Overall, this dashboard and its components provide a professional and insightful visualization of the analyzed data.

F. Story telling:

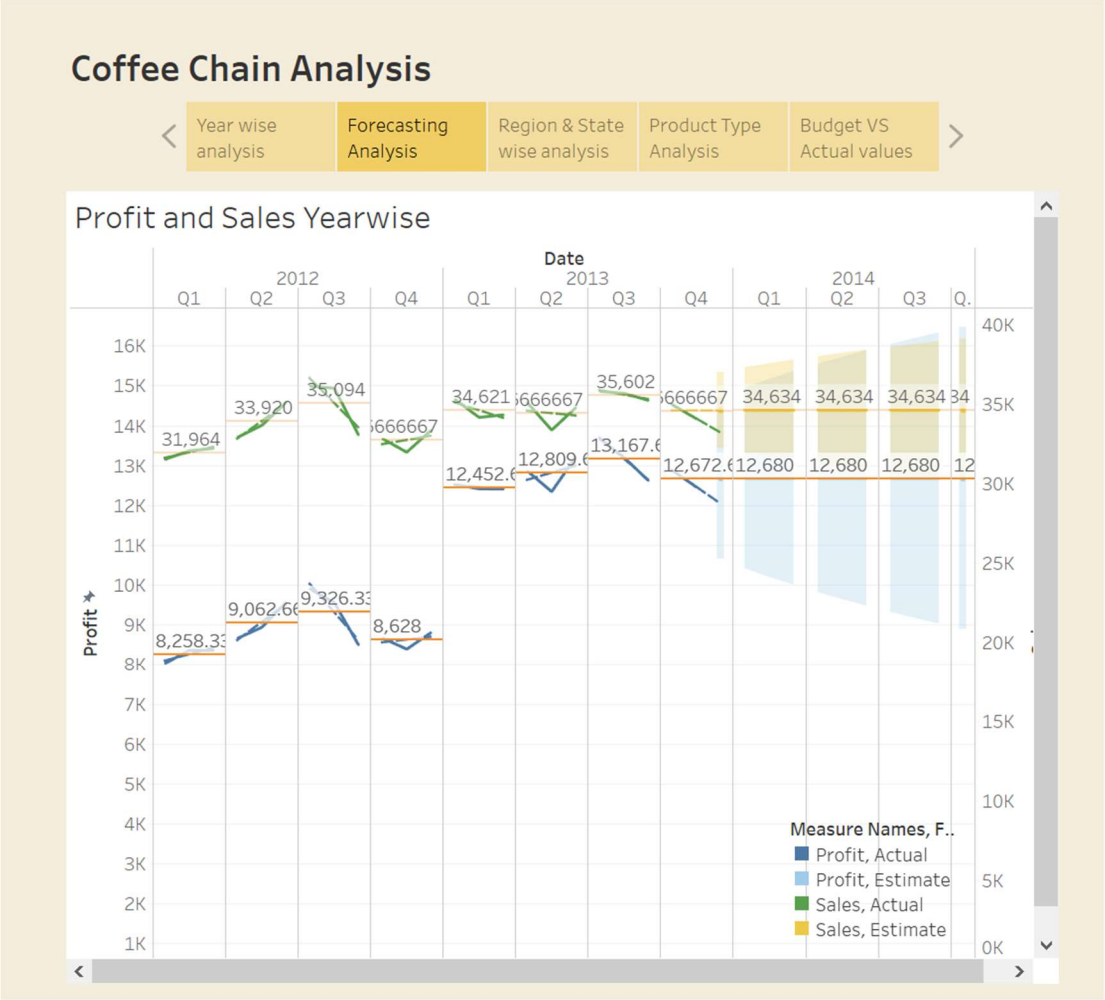
Coffee is cultivated and harvested in over fifty developing nations across Latin America, Africa, and Asia. This commodity serves as a significant source of revenue for 20-25 million families worldwide, making it a vital economic contributor to these regions. [1]

As we delve into the world of data, we are faced with an overwhelming amount of information. In this data story dashboard, we will be analyzing the trends of sales and profit across different regions, states, and product types for the years 2012 to 2014. We will be using various data visualization techniques such as Dual axis bar and Line graphs, Box and Whisker plots, Heat maps, Tabular charts and Scatter Plot to make sense of this data.



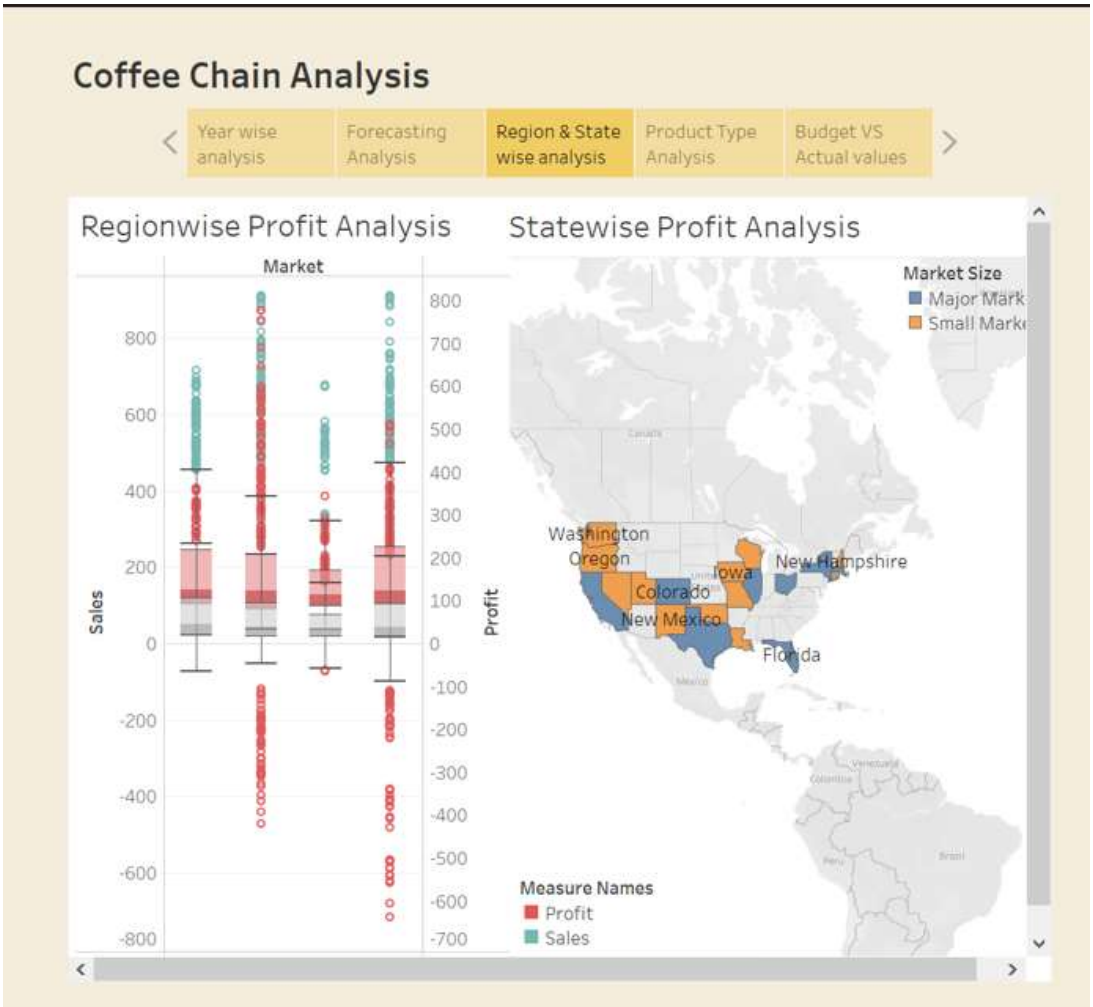
(1)

First, let us look at the overall Profit & Sales values and the dual axis bar and line graph which shows the trend of profit vs sales analysis over the period of 2012 and 2013. The total sales are represented as a bar graph in green color and the total profit is shown as a line graph in red color. We can see that both sales and profit are gradually increasing over these years, indicating a positive correlation between the two. (1)



(2)

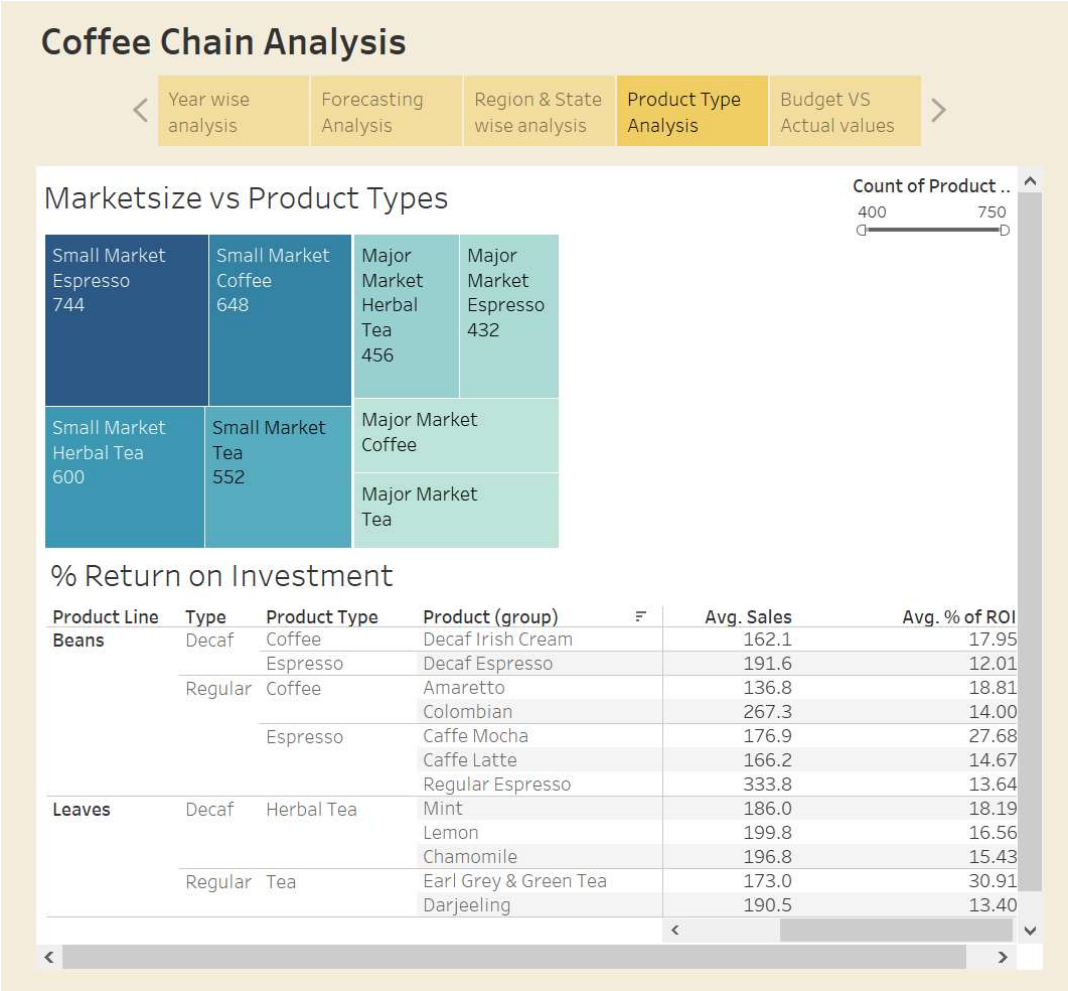
The projected values for 2014 are also shown in quarterly manner, with the forecasted profit ranging from 8,000 to 16,000 in blue and the overall forecasted sales ranging between 30,000 to 40,000 in yellow. (2)



(3)

Moving on, we analyze the performance of different regions and states through the combination of box and whiskers plots and geographical maps. The box plot shows the distribution of profit and sales across different regions. The red box plot represents profit and depicts that the South region has significantly performed the lowest compared to other regions. The outlier is shown in red color. On the other hand, the grey box plot shows the

sales over different regions where the South has the least sales contribution compared to other regions, and the outlier is shown in green color. The geographical map shows us how the market size varies across different states, with the major market covering fewer states compared to small markets and shown in blue and yellow. (3)[3]



Product Line	Type	Product Type	Product (group)	Avg. Sales	Avg. % of ROI
Beans	Decaf	Coffee	Decaf Irish Cream	162.1	17.95
		Espresso	Decaf Espresso	191.6	12.01
	Regular	Coffee	Amaretto	136.8	18.81
			Colombian	267.3	14.00
		Espresso	Caffe Mocha	176.9	27.68
			Caffe Latte	166.2	14.67
			Regular Espresso	333.8	13.64
Leaves	Decaf	Herbal Tea	Mint	186.0	18.19
			Lemon	199.8	16.56
			Chamomile	196.8	15.43
	Regular	Tea	Earl Grey & Green Tea	173.0	30.91
			Darjeeling	190.5	13.40

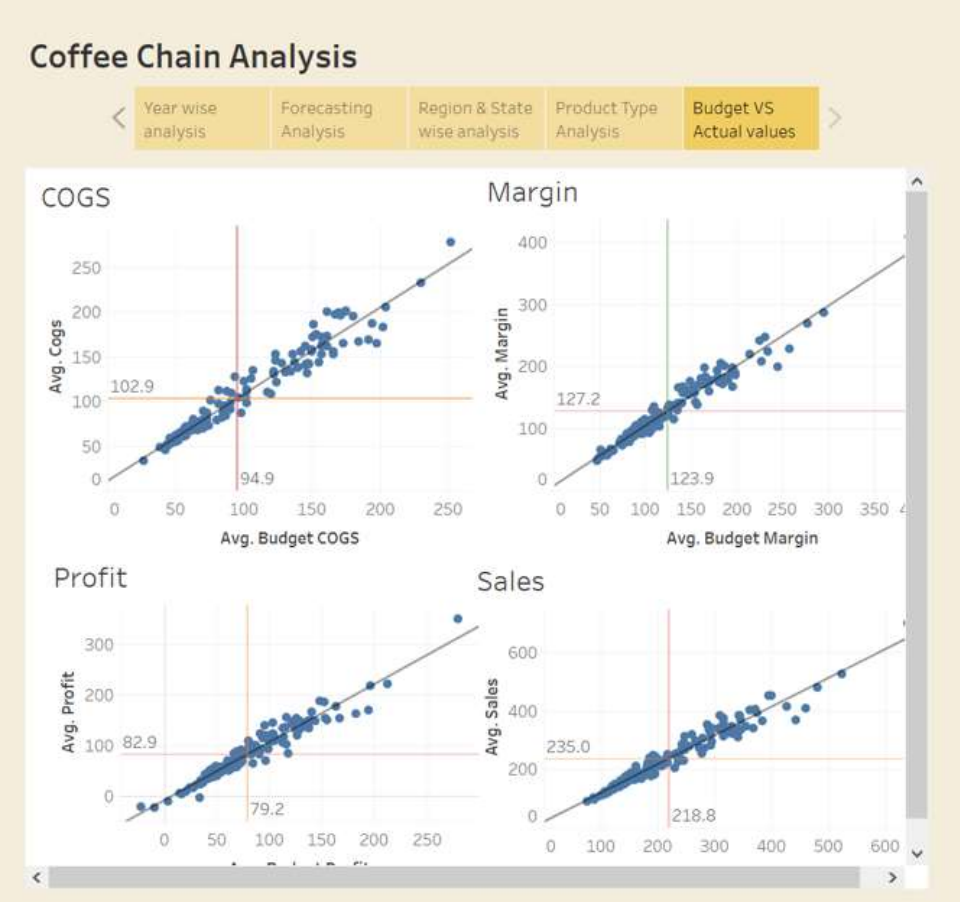
(4)

Moving on to the heat map chart, we analyze the distribution of market size across different product types. The bigger the size of the box and the deeper the color gradient as shown in blue, the greater the value of the product type. As shown from the chart, we observe that small markets sold Espresso the most followed by tea being sold the least [2]. With major markets,

Herbal Tea tops the charts, followed by Regular Tea at the bottom in terms of goods being sold. (4)

Lastly, we have a tabular chart that explains the % Return on Investment metric through analyzing the average amount spent on marketing vs. the average amount of sales made on coffee products over different product lines and types. The grouping function is performed on the "Product" column where we have grouped Early Grey and Green Tea.

The calculated field is Avg % ROI, which is based on avg (Marketing) / avg (Sales). The ranking is based on sorting Avg % ROI from highest to lowest. (4)



(5)

To further delve into our analysis, we sought to determine the correlation between the estimated or budgeted amount and the actual values for various metrics, including COGS, Margin, Profit, and Sales, using scatter plots, linear trend lines, and average value reference lines. Upon analyzing these graphs, we have arrived at the conclusion that the COGS graph exhibits the weakest correlation between planned and actual values. This is due to the highly scattered nature of the data points, in contrast to the positive correlation evident in the other graphs. (5)

In conclusion, by analyzing the trends of sales and profit across different regions, states, and product types, we can gain valuable insights into the performance of the business. By using various data visualization techniques, we can present this information in a clear and concise manner, making it easy to understand for all stakeholders. This data story is just the beginning, and we can use this information to make informed decisions that will drive the business forward.

Citations:

[1] Petit, N. (2007) Ethiopia's Coffee Sector: A Bitter or a Better Future. Journal of Agrarian Change.

[2] Todd Smith (2018), Ivory Research: Starbucks: Reasons for Success.

[3] Abdul Momin: Dunkin' Donuts SWOT Analysis 2022.