

Data Visualization Final Report

By- Thushar Mohan Gaddigopula

University of Colorado Denver

(Using Zomato Bangalore date set to understand the variables affecting restaurant ratings and types in Bengaluru, including demographics, popular cuisines, and restaurant themes.)



Introduction:

The study utilizes data obtained from Zomato Bangalore, a food delivery platform operating in India, akin to the Uber Eats platform in the United States. With the entrance of domestic and foreign enterprises, the restaurant industry in India has undergone significant transformation, resulting in a dearth of skilled professionals. To attract a larger customer base, Indian restaurants have resorted to online platforms, while culinary schools have ramped up their offerings to meet the growing demand. In this context, new restaurants could benefit from investigating factors influencing their location, such as rising disposable incomes, an expanding young population, greater exposure to diverse cultures and cuisines, and an upswing in people's inclination to dine out, as the market for the food services sector continues to expand.

The present study employs data from the Zomato dataset to explore the factors affecting restaurant ratings and types in Bengaluru. Given the high demand for restaurants in this city, investigating the local demographics is essential for identifying prevalent cuisines and restaurant themes. Gathering data on pricing, location, and cuisine type can facilitate the identification of local food trends and demands.

Objectives:

- The objective of this study is to analyze the Zomato dataset, which contains information on over 50,000 restaurants in Bengaluru, India.
- The study aims to identify the factors that influence the aggregate rating of each restaurant and examine the local demographics, including popular cuisines and customer needs.
- The research methodology includes data cleaning, scraping, and analysis, culminating in a final project notebook and documentation.
- The study has five specific goals, including analyzing the factors affecting restaurant ratings, identifying prevalent cuisine types across different localities, exploring the relationship between food prices and ratings, identifying patterns or clusters among restaurants based on location and cuisine, and providing insights and recommendations to new restaurants on optimal location, cuisine type, and pricing strategies.
- The research methodology employs both quantitative and qualitative techniques to achieve the study's objectives and goals.
- The study aims to provide valuable insights into the restaurant industry in Bengaluru, India, and offer recommendations to new restaurants seeking to optimize their location, cuisine type, and pricing strategy.

Exploratory Data Analysis:

In this project, the Zomato dataset underwent a data cleaning process to prepare it for analysis. The cleaning involved removing redundant entries, addressing missing values, and eliminating unnecessary columns. The resulting data was then subjected to processing and analysis to identify patterns and insights related to variables such as location, estimated price range, cuisine type, and the popularity of specific dishes in various localities. These variables were investigated for their impact on the overall ranking of restaurants in Bengaluru. The rigorous cleaning process ensured the quality and reliability of the dataset, allowing for accurate and insightful analysis of the restaurant industry in Bengaluru.

This study utilized both Excel and R to conduct an exploratory data analysis of the Zomato Bangalore dataset, focusing on restaurant data. The analysis involved the visualization of various data types, such as average cost, price range, cost for two, votes, and aggregate rating. The visualizations were created using a range of graphs and charts, including scatter plots, line graphs, tree maps, bar graphs, and box plots. The visualizations provided insights into the distribution of data, trends over time, and relationships between different variables. This approach allowed for a comprehensive understanding of the data, which could be used to gain valuable insights into the restaurant industry in Bangalore.

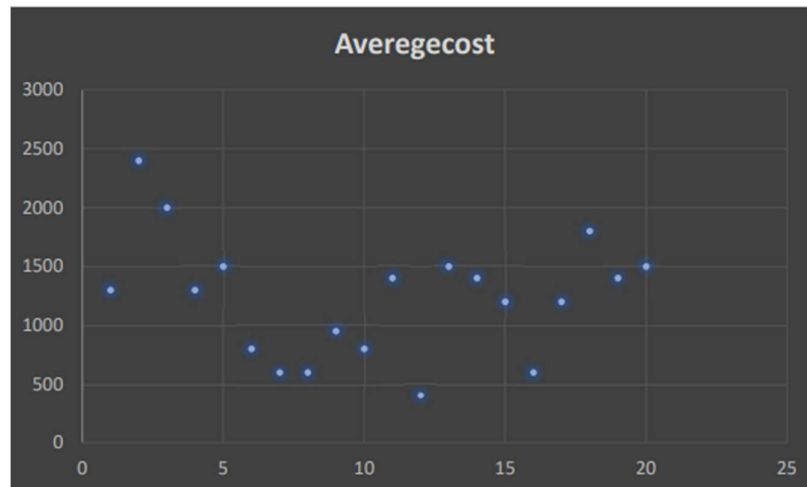
By utilizing Excel and R for data visualization and analysis, a more comprehensive understanding of the Zomato Bangalore dataset was achieved. The resulting visualizations enabled the identification of trends and patterns that would not have been easily observable through raw data analysis alone. These visualizations can inform business decisions or personal choices and serve as a starting point for further analysis and investigation. The approach used in this study allowed for a thorough exploration of the data and provided a foundation for future research and decision-making in the restaurant industry.

Dataset Used:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V			
1	1	Restaurant	Restaurant	Country	City	Address	Locality	Locality	Ve	Longitude	Latitude	Cuisines	Average	Cost	Currency	Has Table	Has Online	Is deliverin	Switch to	Price range	Aggregate	Rating	color	Rating text	Votes
2	728	50943	Sultans of	1	Bangalore	BluPetal H	BluPetal H	BluPetal H		77.61543	12.93328	North Indi	1300	Indian	Rup	Yes	Yes	No	No	3	4.1	Green	Very Good	2416	
3	729	58268	The Fatty I	1	Bangalore	610, 3rd Fl	Indiranaga	Indiranaga		77.6454	12.97022	Asian	2400	Indian	Rup	Yes	Yes	No	No	4	4.7	Dark Gree	Excellent	2369	
4	730	51705	Toit	1	Bangalore	298, Namr	Indiranaga	Indiranaga		77.64071	12.97917	Italian, Am	2000	Indian	Rup	No	No	No	No	4	4.8	Dark Gree	Excellent	10934	
5	731	18162866	Three Dot	1	Bangalore	840/1,100	Indiranaga	Indiranaga		77.64049	12.98041	European,	1300	Indian	Rup	No	No	No	No	3	3.9	Yellow	Good	1354	
6	732	18407918	Bombay Bi	1	Bangalore	2989/B, 12	Indiranaga	Indiranaga		77.64575	12.97032	Modern In	1500	Indian	Rup	No	Yes	No	No	3	4.2	Green	Very Good	231	
7	733	56464	Glen's Bak	1	Bangalore	297, 100 F	Indiranaga	Indiranaga		77.64062	12.9791	Bakery, De	800	Indian	Rup	No	No	No	No	2	4	Green	Very Good	3533	
8	734	18221572	Onesta	1	Bangalore	501, Binna	Indiranaga	Indiranaga		77.64368	12.97845	Pizza, Cafe	600	Indian	Rup	No	Yes	No	No	2	4.3	Green	Very Good	1413	
9	735	18359919	Onesta	1	Bangalore	Site 15, 15	JP Nagar	JP Nagar, E		77.59679	12.90623	Pizza, Cafe	600	Indian	Rup	No	No	No	No	2	4.6	Dark Gree	Excellent	781	
10	736	18439634	ECHOES Ki	1	Bangalore	44, 4th B	C Koramang	Koramang		77.6158	12.93418	Continenti	950	Indian	Rup	No	No	No	No	2	4.7	Dark Gree	Excellent	276	
11	737	51040	Truffles	1	Bangalore	28, 4th 'B'	Koramang	Koramang		77.61429	12.9333	American,	800	Indian	Rup	No	Yes	No	No	2	4.7	Dark Gree	Excellent	9667	
12	738	54162	The Black	1	Bangalore	105, 1st A	Koramang	Koramang		77.61616	12.93436	North Indi	1400	Indian	Rup	Yes	No	No	No	3	4.1	Green	Very Good	5385	
13	739	18305628	Eat Street	1	Bangalore	11, 80 Fee	Koramang	Koramang		77.626	12.9395	North Indi	400	Indian	Rup	No	Yes	No	No	1	4.3	Green	Very Good	753	
14	740	18385443	Koramang	1	Bangalore	118, Koran	Koramang	Koramang		77.61413	12.93566	Continenti	1500	Indian	Rup	No	No	No	No	3	4.5	Dark Gree	Excellent	1288	
15	741	56618	AB's - Absc	1	Bangalore	90/4, 3rd F	Marathahi	Marathahi		77.69939	12.94993	European,	1400	Indian	Rup	No	No	No	No	3	4.6	Dark Gree	Excellent	6907	
16	742	18353121	Flechazo	1	Bangalore	9/1, 1st Fl	K Marathahi	Marathahi		77.69666	12.97538	Asian, Mec	1200	Indian	Rup	Yes	No	No	No	3	4.4	Green	Very Good	983	
17	743	18366652	Onesta	1	Bangalore	215, 216 &	New BEL F	New BEL F		77.571	13.0292	Pizza, Cafe	600	Indian	Rup	No	No	No	No	2	4.6	Dark Gree	Excellent	627	
18	744	18430785	Communit	1	Bangalore	67 & 68, B	Residency	Residency		77.60818	12.97253	Continenti	1200	Indian	Rup	No	No	No	No	3	4.2	Green	Very Good	334	
19	745	58882	Big Brewsk	1	Bangalore	Behind MK	Sarjapur R	Sarjapur R		77.68324	12.91304	Finger Foo	1800	Indian	Rup	Yes	Yes	No	No	3	4.5	Dark Gree	Excellent	5705	
20	746	18422898	Hoot	1	Bangalore	BBMP 203	Sarjapur R	Sarjapur R		77.6784	12.91426	Continenti	1400	Indian	Rup	No	No	No	No	3	3.9	Yellow	Good	405	
21	747	18339874	Farzi Cafe	1	Bangalore	202, Level	UB City	UB City, Bi		77.59601	12.97216	Modern In	1500	Indian	Rup	No	No	No	No	3	4.4	Green	Very Good	754	

Graphs plotted for EDA:

1) Scatter Plot:

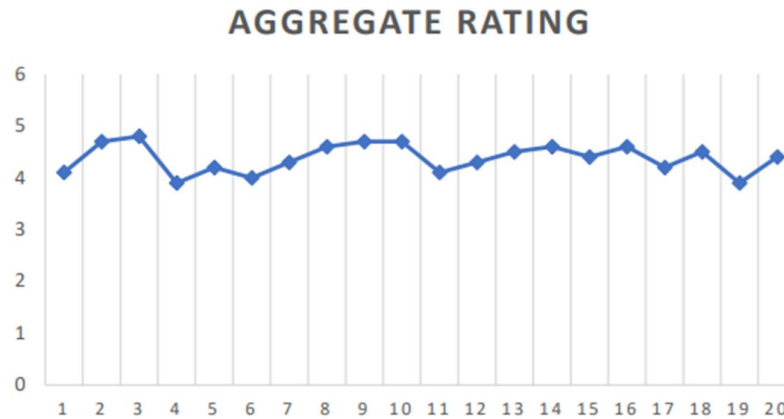


The scatter plot is a graphical representation that shows the relationship between two variables. The x-axis represents one variable, while the y-axis represents the other. In this case, the scatter plot displays the relationship between average cost and another variable. Each point on the plot represents a specific level of output and the corresponding average cost. The shape and trend of the plot can reveal important information about the relationship between the two variables. For example, a downward-sloping line indicates that as output increases, average cost decreases, while a flat or upward-sloping line may indicate that as output increases, average cost remains the same or increases, respectively.

2) Line Graph:

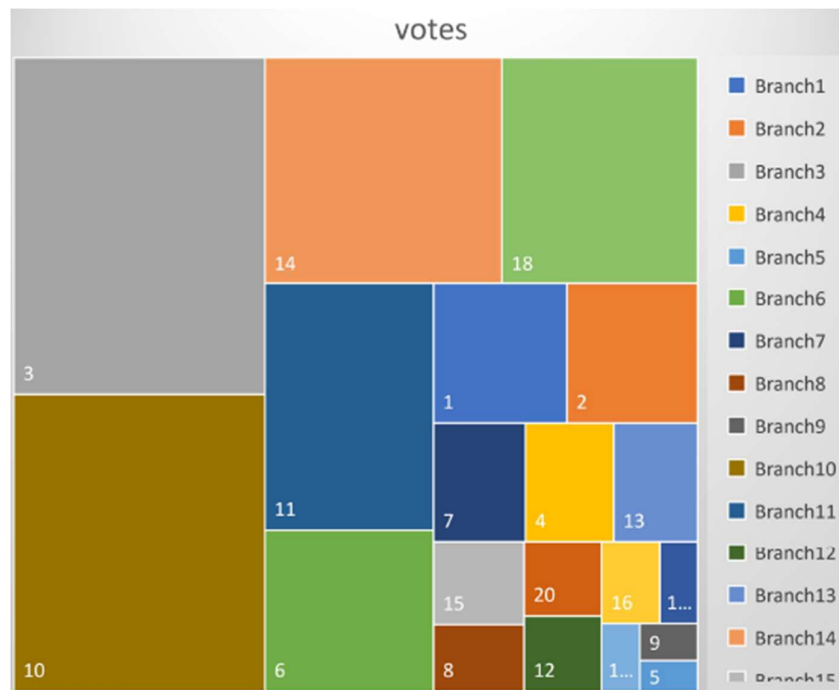


The graph displays the price trend of a product or service over time, revealing pricing strategies. A downward slope indicates a decrease in price range, while an upward slope suggests an increase. A flat line shows no significant change. Price range data can identify trends in pricing over time and compare pricing across products or services.



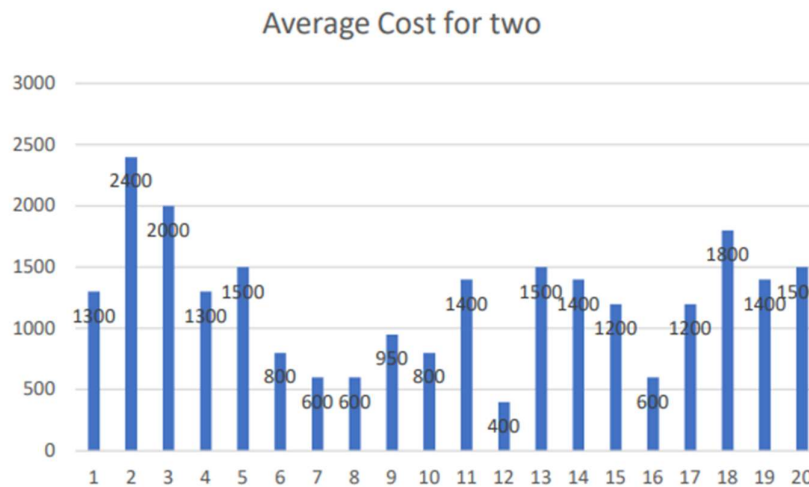
The graph shows trends in sales or revenue for a product over time. Each point on the line represents a specific time period and the corresponding aggregate rate. An upward slope indicates increasing revenue, a downward slope indicates decreasing revenue, and a flat line indicates no significant change. Aggregate rate data can be used to identify sales or revenue trends over time and compare performance across products or services.

3) Tree Map:



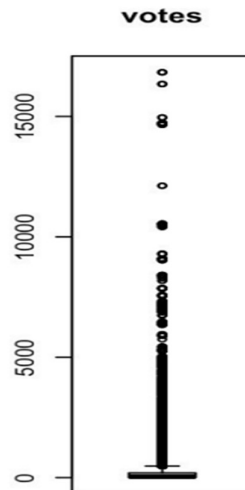
Utilized a tree map to analyze voting data and determine the most popular restaurant categories and cuisines among their users in Bangalore. The rectangles in the tree map represent categories and subcategories, with size corresponding to their value or frequency. This hierarchical arrangement of data allows for identification of trends and preferences, which can inform decision-making regarding restaurant promotion and user engagement.

4) Bar Graph:



Zomato's graph displays the average price range for a meal for two people at various restaurants in Bangalore. Each bar represents a restaurant or group of restaurants, with the height indicating the average cost for two people to dine there. The graph helps understand the price points of different restaurants and their position in the market.

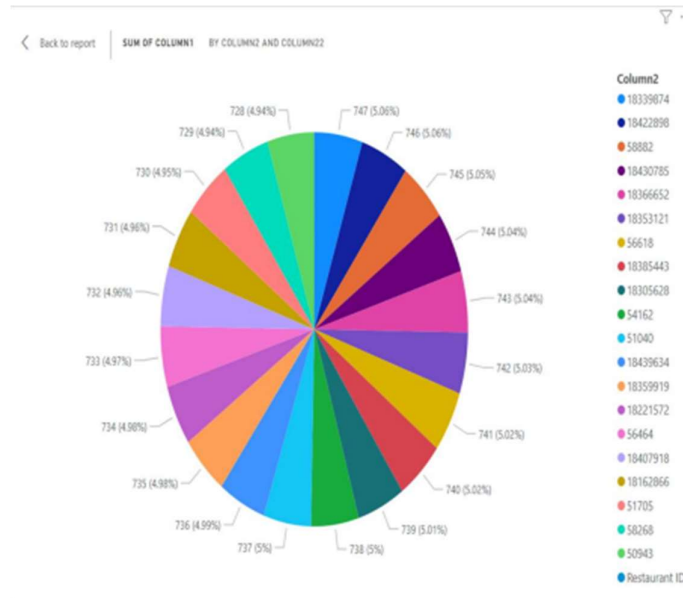
5) Box Plot:



This compares the vote distribution of different restaurants in Zomato Bangalore, identifying higher and lower vote counts and their distribution among restaurants. It helps to recognize highly popular restaurants and those with a wide range of votes, providing insight into their popularity and how they compare to competitors.

Visualization of Zomato Dataset using PowerBI, Tableau, and R to uncover relationships between parameters for improved decision-making:

Pie Chart:



The pie chart is an effective way to represent the distribution of votes among the different restaurants in your dataset. Each slice of the pie corresponds to a unique restaurant ID, with the size of the slice representing the number of votes received by that restaurant. This provides a clear visualization of the relative popularity of each restaurant based on the number of votes they have received.

This information can be useful for making informed business decisions. For example, you can identify the most popular restaurants and consider promoting them to attract more customers. On the other hand, you can investigate why some restaurants are not receiving as many votes and take corrective measures to improve their popularity.

It's important to note that pie charts are most effective when you want to compare parts of a whole. In this case, the "whole" is the total number of votes, while the "parts" are the votes received by each restaurant. The pie chart allows you to easily see how the votes are distributed among the different restaurants, making it an ideal visualization for this type of data.

Funnel Chart:



A Funnel chart is a powerful visualization tool that showcases the number of votes each cuisine received, arranged in descending order. The X-axis of the chart represents the number of votes, while the Y-axis shows the different cuisines in the dataset. This chart provides a clear picture of the relative popularity of each cuisine based on the number of votes it received, with the most popular cuisine at the top of the chart.

The insights obtained from a Funnel chart can be used to make data-driven decisions for your business. For instance, by promoting the most popular cuisines, you can attract more customers and increase your revenue. Alternatively, if certain cuisines are not receiving many votes, you may want to investigate the underlying issues affecting their popularity and take corrective measures.

Funnel charts are particularly useful for visualizing the stages of a process or journey and how many items pass through each stage. In this case, the chart shows the number of votes received by each cuisine, starting with the most popular and ending with the least popular. This information can help you to identify areas that need improvement and make data-driven decisions that will lead to better business outcomes.

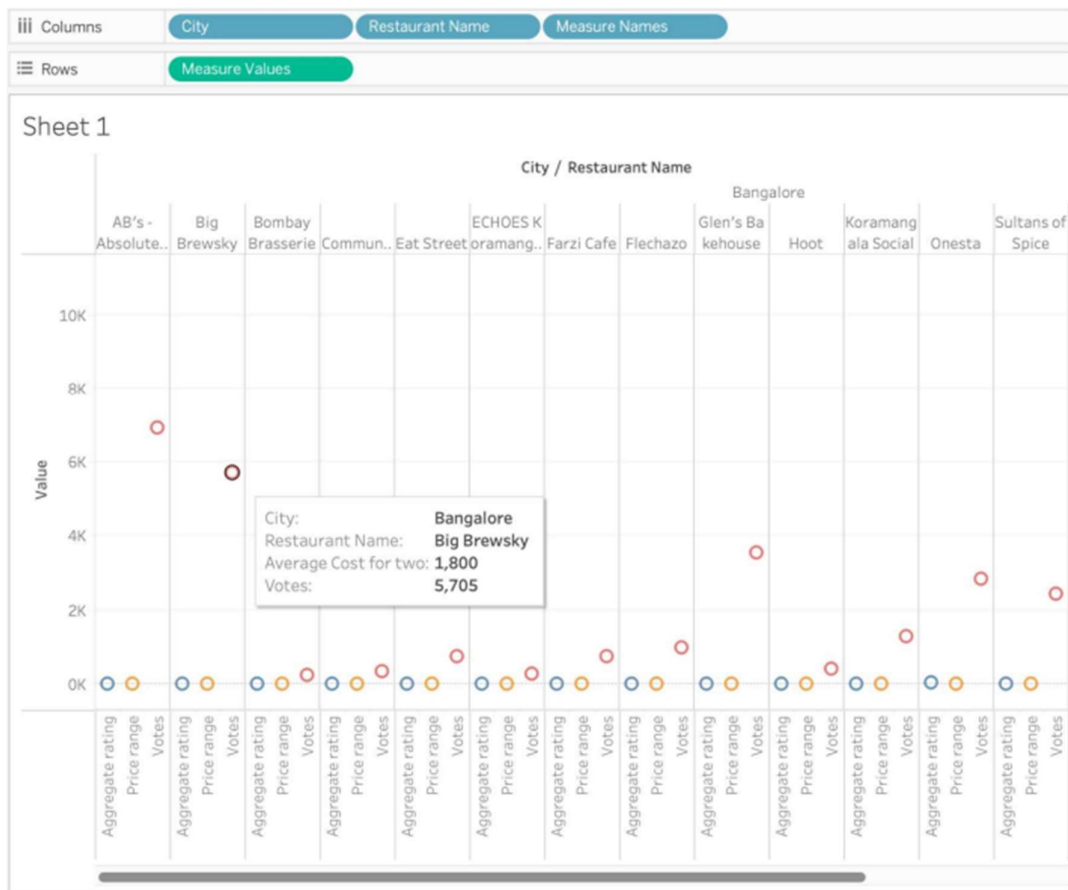
Waterfall Chart:



A waterfall chart is an effective visualization tool that displays the total number of restaurants for each cuisine in a descending order. With the X-axis representing different cuisines and the Y-axis indicating the number of restaurants, the chart provides a clear picture of how cuisines are ranked based on the number of restaurants they have.

Using this chart, you can make informed decisions about your business. For example, you may want to promote the most popular cuisines to attract more customers or investigate why some cuisines have fewer restaurants and take steps to increase their popularity. Overall, the waterfall chart helps you to identify the popular and less popular cuisines, which can guide your decision-making processes and ultimately improve your business outcomes.

Side-by-Side Circle Chart:

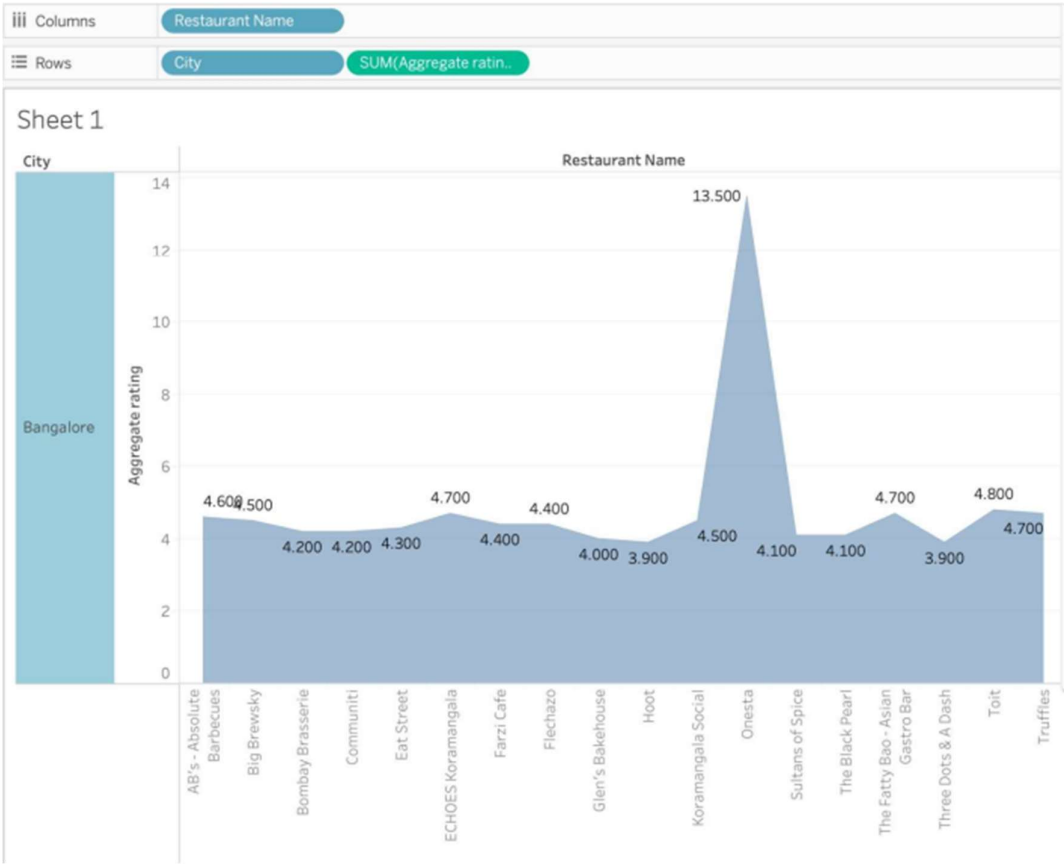


I chose to utilize a side-by-side circle chart as it is an effective way to represent the relative performance of various restaurants across different cities. The use of different colors for each restaurant enhances the chart's visual appeal and enables the reader to quickly identify each restaurant.

By incorporating measure names and values, I was able to display multiple metrics for each restaurant in the same chart, simplifying the comparison of their performance on different measures. Furthermore, by incorporating city as a dimension, I was able to group the restaurants by location, simplifying the assessment of their performance across different markets.

Overall, the side-by-side circle chart utilizing city, restaurant names, measure names, and measure values was an efficient method for visualizing the performance of multiple restaurants across different cities, allowing the reader to quickly comprehend the data and extract valuable insights.

Stacked Area Chart:



I selected an area chart for its ability to convey changes over time while also showing the relative performance of different restaurants. The use of distinctive colors for each restaurant and city enhances the chart's visual appeal and assists readers in identifying each data point quickly.

By utilizing restaurant name, city, and SUM(Aggregate rating), I was able to present the performance of each restaurant in every city over time, making it simple to compare their performances against each other. The summation of ratings gives a comprehensive view of the overall performance of each restaurant, and the city dimension allows for effortless comparison between different markets.

Overall, the area chart using restaurant name, city, and SUM(Aggregate rating) was an effective method for visualizing the performance of various restaurants in various cities over time. The chart was easy to interpret and provided a clear depiction of how each restaurant's aggregate rating was evolving over time in each city.

Conclusion:

The purpose of this research project is to conduct an in-depth analysis of the Zomato dataset with a focus on identifying the key factors that influence restaurant ratings and types in Bengaluru, India. The project aims to explore various independent variables such as the geographical location, cuisine type, pricing, and popularity of specific dishes to provide insights and recommendations to new restaurants in the region.

The study will employ a comprehensive data cleaning and processing approach to ensure the accuracy and validity of the results. Furthermore, we will use powerful data visualization tools such as Tableau or Power BI and Excel to analyze the dataset and identify trends and patterns in the data.

To achieve our objectives, we will employ a rigorous methodology that involves collecting, cleaning, and transforming the data to extract valuable insights and trends. We will utilize both descriptive and inferential statistical techniques to analyze the data, including measures of central tendency, dispersion, correlation, and regression analysis.

Our analysis will focus on identifying the key factors that drive restaurant ratings and popularity in the Bengaluru region. We will explore the impact of variables such as location, cuisine type, pricing, and dish popularity on restaurant ratings. Our analysis will also provide recommendations on optimal location, cuisine type, and pricing strategy for new and existing restaurants based on the local demographics and food trends.

Overall, this project is expected to provide valuable insights and recommendations to help new and existing restaurants in Bengaluru make informed decisions about their business strategy, thereby improving their competitiveness and success in the market.

References and Source:

Source:

<https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalorerestaurants/versions/1?resource=download>.

References:

- Zomato offers a public API for developers to access and extract restaurant data from their database. This API can be used to obtain the latest data and update analysis.
- Kaggle is a platform for data science competitions and has several datasets related to restaurant ratings, reviews, and pricing, including the Zomato dataset.
- Google Analytics course focused on plotting and visuals.

