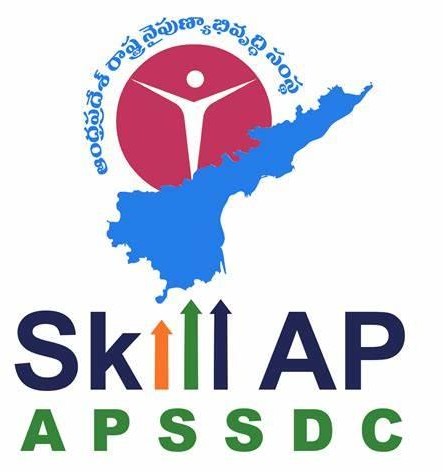
**FOOD DELIVERY ANALYSIS**

Prepared in the partial fulfillment of the Summer Internship Program on Data Analysis AT



*Under the guidance of*

**Ms. Narmada Mani, APSSDC Mrs. Meenakshi, APSSDC**

*Submitted by*

**Madhava Rao Sakaray, AP21110011586**

**Likhith Sai Thimmannagari, AP21110011584**

**K.Revanth Kumar Goud, AP21110011566**

**SRM University, AP**

(June-Aug, 2023)

# ACKNOWLEDGEMENT

We would like to extend our sincere appreciation to Meenakshi Ma’am and Narmada Ma’am, for their invaluable guidance and consistent support during our summer internship program, which was made possible through the efforts of APSSDC. Your expertise not only improved our technical skills but also broadened our industry perspective.

We also want to express our gratitude to APSSDC that actively promotes skill development and empowers students like us to gain real-world experience. The resources, opportunities, and exposure provided by APSSDC were invaluable in shaping my understanding of the industry and preparing me for future challenges.

Once again, thank you, Narmada Ma’am, Meenakshi Ma’am, and APSSDC, for your significant contribution to our professional growth. We look forward to carrying forward the knowledge and insights gained during this internship as we embark on the next phase of my academic and professional journey. Your guidance will always be a beacon of inspiration for us.

Thank you.

Sincerely,

Madhava Rao Sakaray, AP21110011586

Likhith Sai Thimmannagari, AP21110011584

K.Revanth Kumar Goud, AP21110011566

# ABSTRACT

The "Food Delivery Dataset Analysis" project aims to explore and analyze a comprehensive dataset from a prominent food delivery platform to uncover valuable insights and trends within the industry.

With the rapid growth of the food delivery market, understanding customer preferences, delivery patterns, and other critical aspects becomes pivotal for improving service efficiency and customer satisfaction.

This research utilizes a diverse dataset encompassing customer information, order details, delivery routes, restaurant data, and feedback ratings.

Leveraging advanced data analytics techniques, we analyze the dataset to identify key factors that influence user behavior, popular cuisines, peak ordering hours, delivery times, and customer satisfaction levels.

Data visualization provides a quick and effective way to communicate information in a universal manner using visual information. Starting from a clean dataset allows you to focus on creating an effective visualization rather than trying to diagnose and fix issues while creating visualizations.

# INTRODUCTION

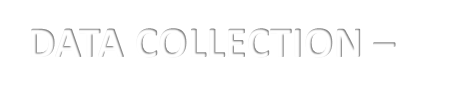
The rise of online platforms have revolutionized the way people order and enjoy food. Food delivery services have emerged as a dominant force in the food industry, offering convenience and a diverse range of culinary options to consumers.

The "Food Delivery Dataset Analysis" project aims to leverage this wealth of data to gain deeper insights into the dynamics of the food delivery market.

By exploring and analyzing a comprehensive dataset from a prominent food delivery platform, we seek to uncover hidden patterns, trends, and correlations that can lead to informed decision-making and improvements in service quality.

With the ever-increasing popularity of food delivery platforms, an immense amount of data is generated daily, encompassing crucial information about customer behavior, restaurant performance, delivery logistics, and overall service efficiency.

# OVERVIEW/METHOLODOLOGY



**1. DATA COLLECTION –**

* This step involves collecting dataset from a reliable website (like Kaggle ), as it directly affects the quality of your analysis.
* The dataset used is **Zomato.csv**. It has multiple columns such as url , address, name, online order, book table, rate, votes, phone, location, rest type, dish liked, approx cost(for two people), reviews list, menu item ,listed in(type), listed in(city).
* Import this csv file to Jupyter Notebook using **read\_csv()**

method in pandas.

## DATA PREPROCESSING –

* This step involves several tasks aimed at cleaning, transforming, and organizing the data to improve its quality, relevance, and suitability for the specific analysis.
* Converting Rating into a float value by removing “/5” and dropping duplicates and unnecessary and insignificant columns like url, listed\_in(city) and also changing the names locations, rest\_types and cuisines which have less than considerable values and changing their name to “others” for easy evaluation and analysis of data.
* **Isnull()** method is used to display all the null values in a column of the dataset. Null values are needed to handle as null values can lead to inaccurate results, predictions, negatively impact the analysis, and can cause inconsistence in the analysis of the dataset.
* Later using **dropna()** method , drop the rows that contains null value and keeping inplace as True.

## QUERIES RESOLUTION-

* Query resolution is critical for data analysis because it ensures that the inquiries or questions on the data are addressed accurately and efficiently. This allows focusing on deriving valuable insights from the data.
* **groupby()** method is used to group data based on one or more categorical variables, allowing one to apply aggregate functions to each group independently. For example, groupby() can be used for Category to find out all details based on each category.
* **sort\_values()** method is used to sort the rows in a dataset,

which allows one to better understand data’s distribution, find patterns, and present data in more organized manner. For example, sort\_values() can be used to find out Locations and cuisines and restaurant types which have High demand in descending /ascending order.

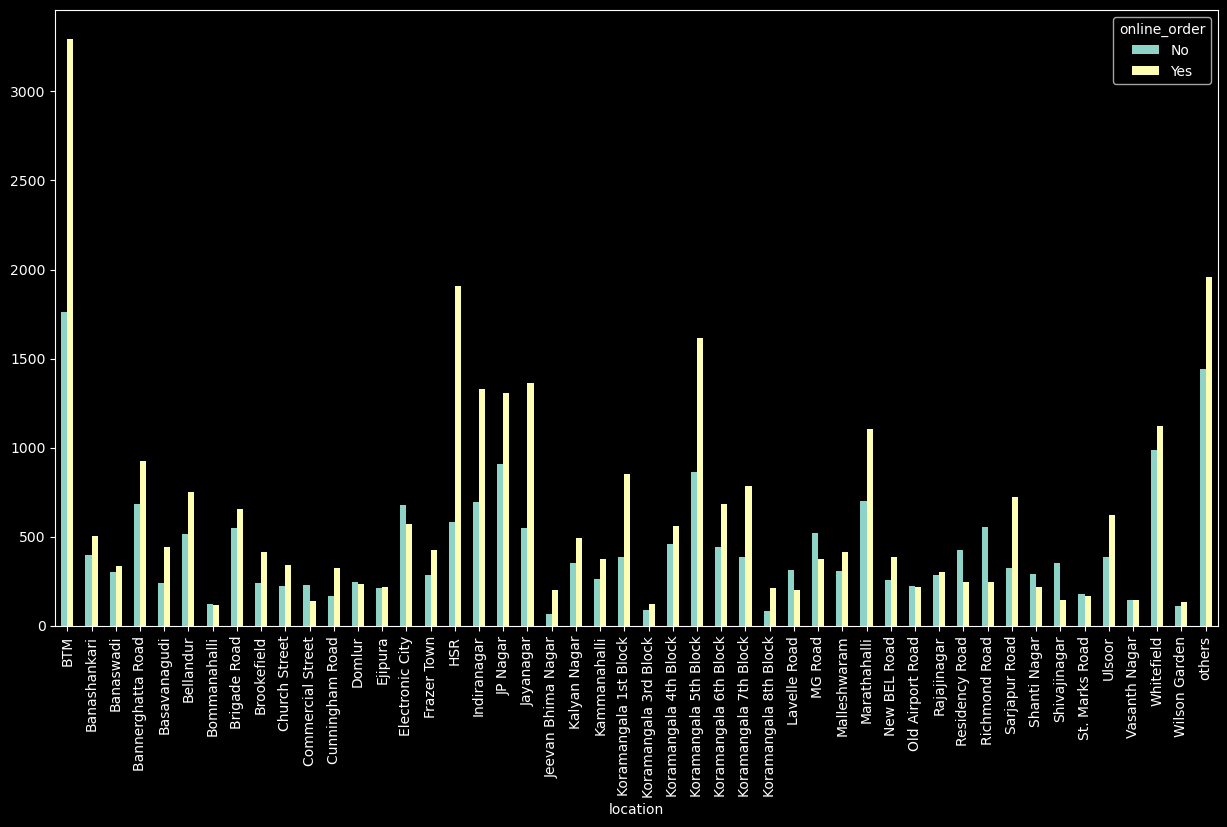
* **value\_counts()** method helps to count the occurrences of each unique value and provide summary of the data’s frequencies. For example, value\_counts() can be used to find the number of Restaurants in particular locations,no. of restaurants of types Votes for restaurants and Number of online orders .

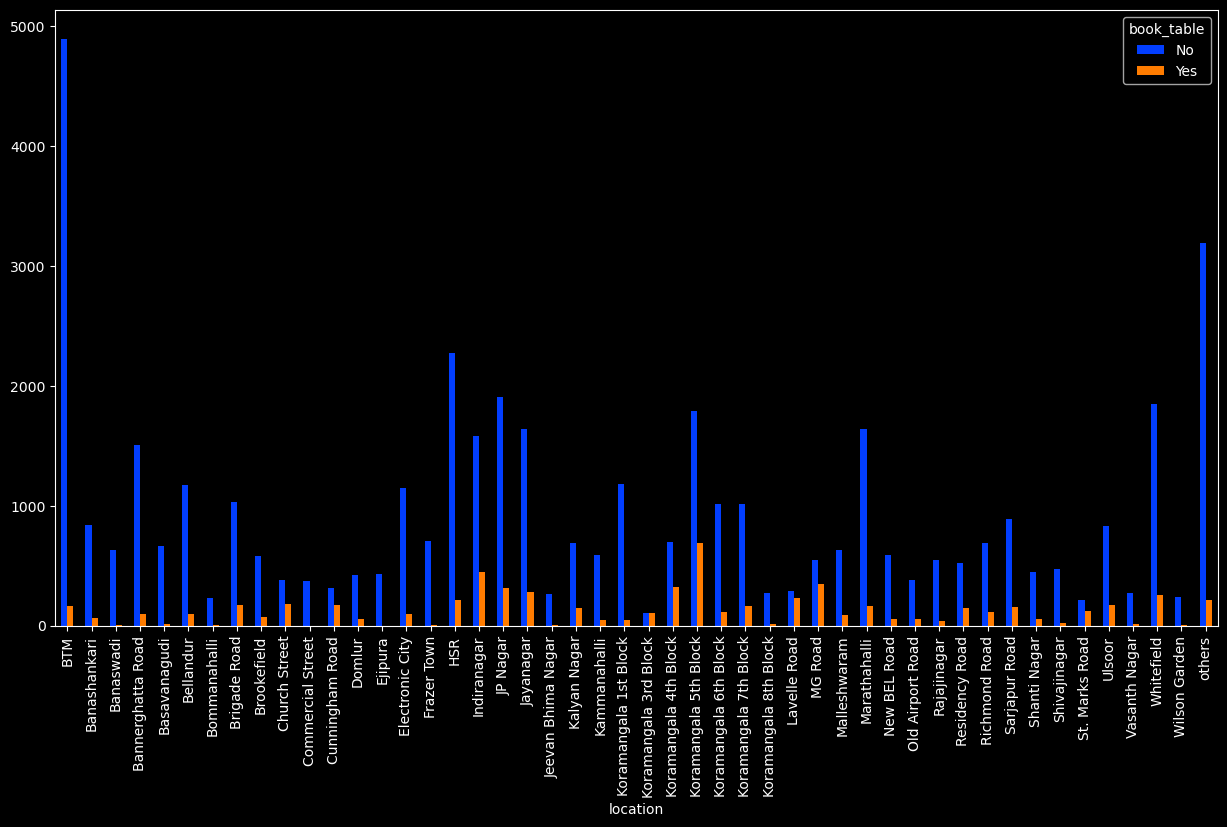
## Few Queries that are resolved are-

* + What type of dishes are most popular and liked most by customers?
  + Which locations are most suitable for establishing a new restaurant?
  + Which facilities of restaurants are attractive and improves the restaurant rating?
  + How to change menu to manage large volumes of orders in particular times?

## DATA VISUALIZATION –

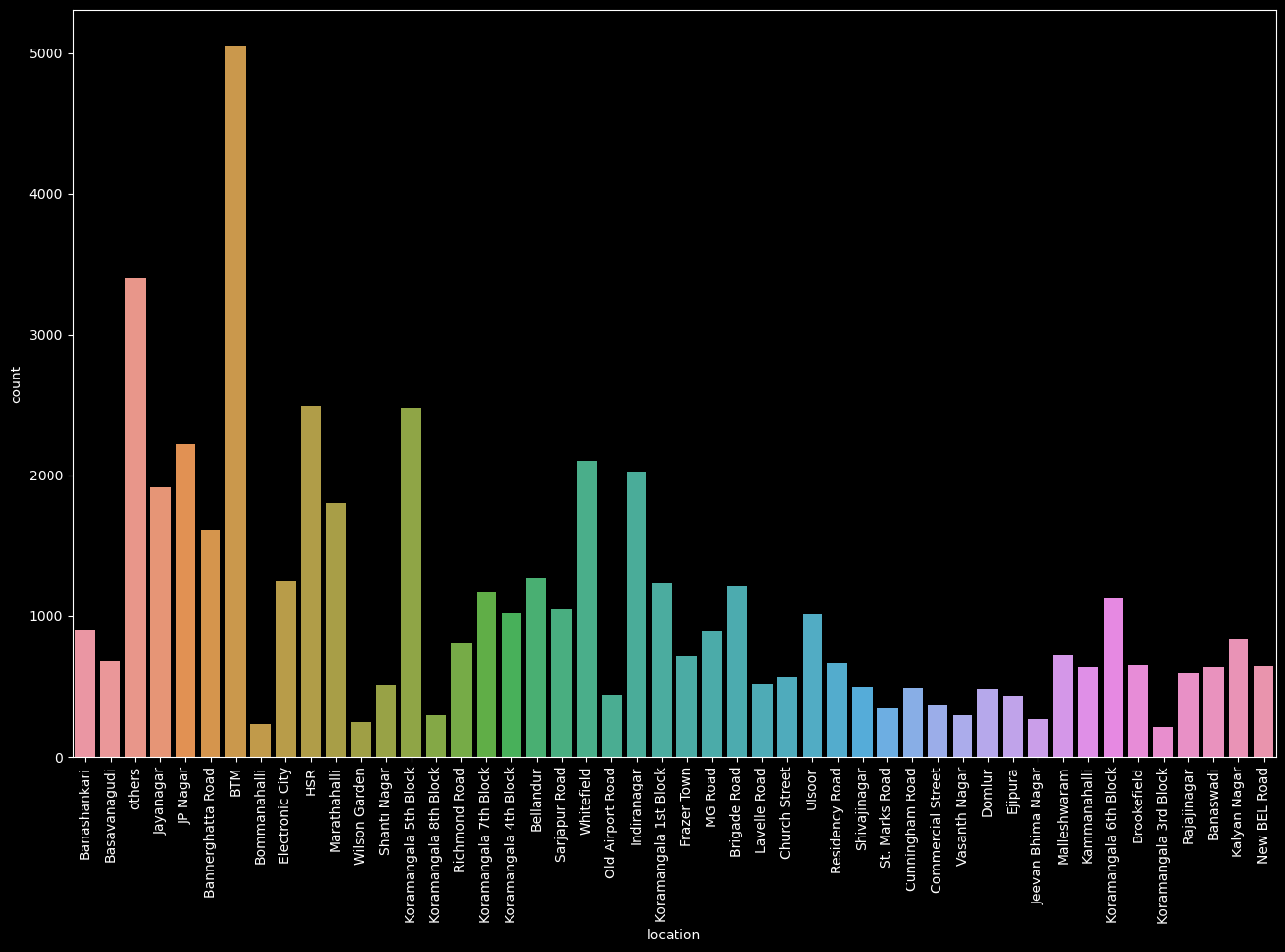
* It involves creating graphical representations of data to visually present number restaurants in locations,votes and rating that might be difficult to understand from the raw data alone.
* Matplotlib and Seaborn are extensive used to plot graphs.
* **Bar graphs** are represents frequency or count of a particular category as rectangular bars. For example- plot showing number of restaurants accepting online orders and restaurants not accepting online orders, also restaurants having book table facility located on respective location on X axis.



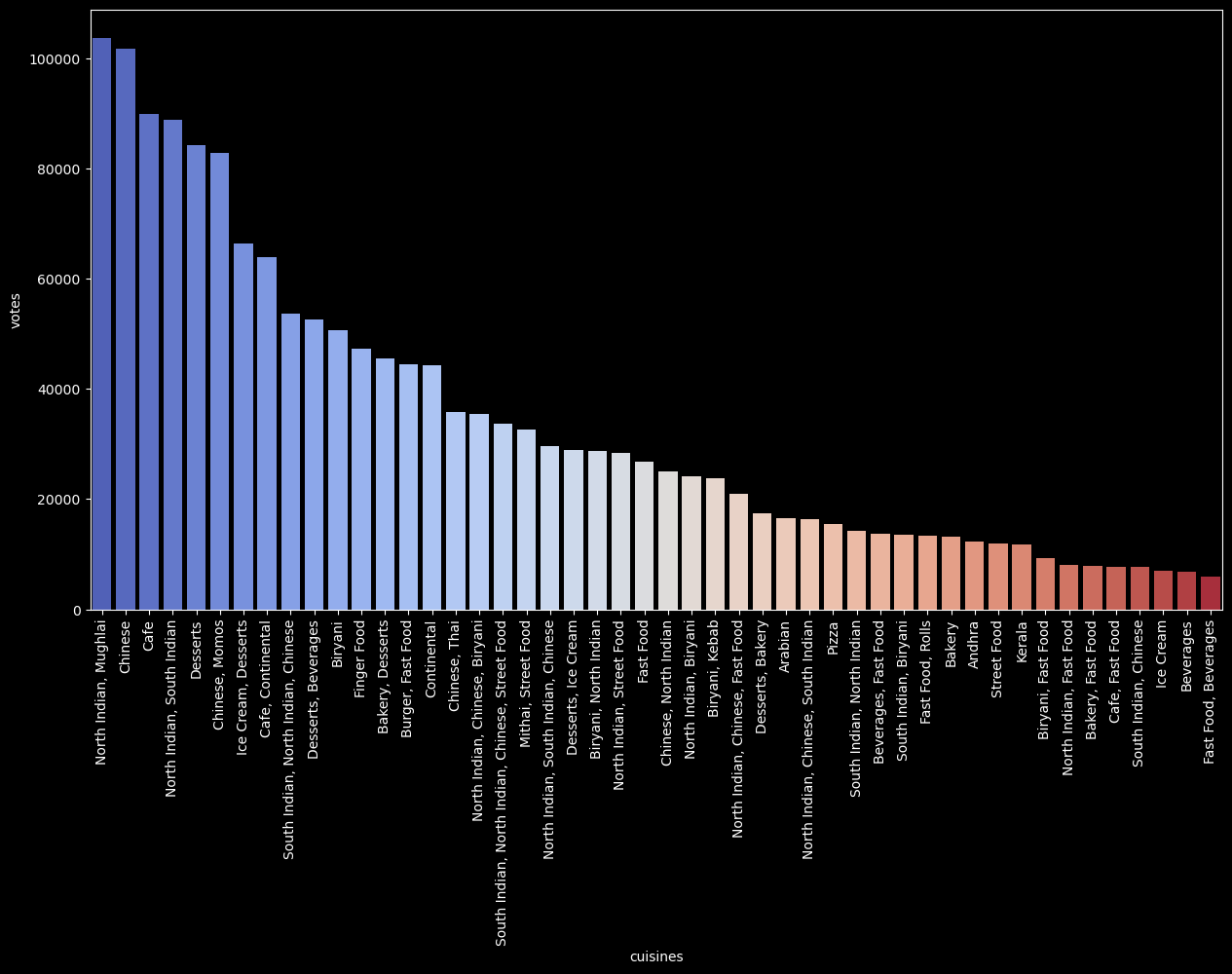


* **Count plot** is type of bar plot that displays the count of observations in each category of a categorical variable. For example- Plot showing number of restaurants present in particular locations in the city, and the count plot showing votes for dishes types showing in descending order.

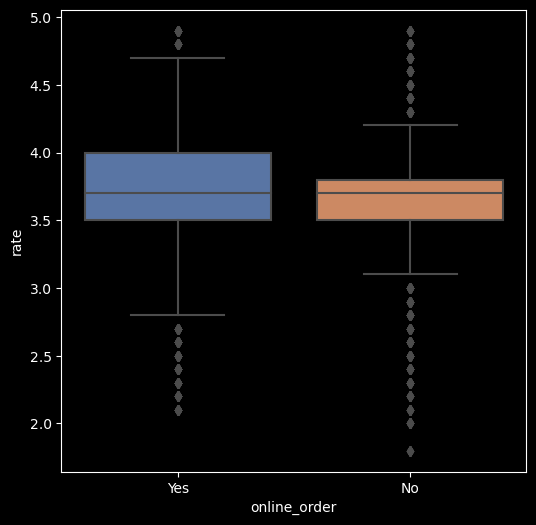
No.of restaurants at particular locations



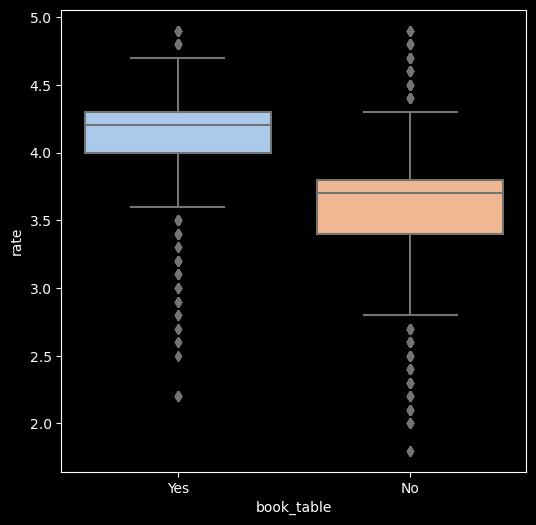
Votes for different types of dishes



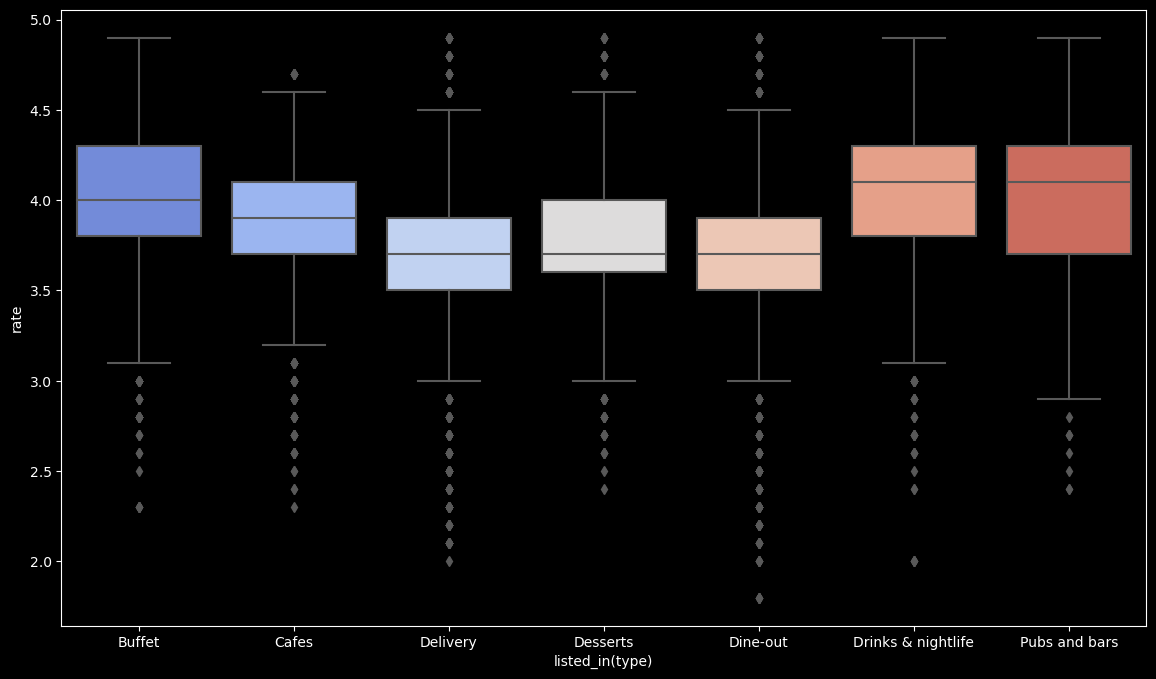
* **Box plot** is a graph that gives a visual indication of how a data set's mean, median, mode, minimum, maximum and outer values are spread out and compare to each other.For example box plots showing Ratings for restaurants based on online order facility and book table facility and also on the type of the restaurant.



Box plot showing rating considering book table facility



Box plot for restaurant types and their Rating.



## BEHAVIOUR ANALYSIS-

* Behaviour analysis provides critical insights that can drive decision-making, improve user experience, optimize marketing strategies, and enhance overall business performance.

## Important remarks that can be made by behaviour analysis-

* + North Indian, Mughlai dishes are the most popular among customers in the city.
  + Book Table facility is responsible for having a good rating of a restaurant.
  + Restaurants offering online order facility have the capability of achieving maximum rating.
  + Buffet, Bar and Night club type restaurants have the highest average ratings and also the maximum rating.
  + For an individual who is starting a new restaurant it good if they choose location with less restaurants to avoid heavy competition, also provide online order and book table facility,for getting good ratings.

# REQUIREMENTS

## Operating System Used –

* Windows OS

## Software requirements -

* **Data Analytic Tools**: Pandas and Numpy - These popular libraries provide essential tools for data processing and analysis.
* **Data Visualisation Tools**: Matplotlib and Seaborn - Libraries in Python for creating data visualizations.
* **IDE** – Jupyter Notebook

## Hardware requirements-

* Multicore processor with atleast 2.5 GHz clock speed (recommended 4 core or more)
* Minimum 16 GB RAM for handling large dataset and memory- intensive operations.
* Adequate storage space for storing dataset and analysis results.

# CONCLUSION

In conclusion, the food delivery dataset analysis has provided valuable insights that can guide strategic decision-making, improve customer satisfaction, and drive sustainable growth for the food delivery platform. By capitalizing on these findings and continuing to leverage data-driven approaches, the platform can maintain a competitive edge in the dynamic and evolving food delivery market.

The analysis also highlighted promising market opportunities. Untapped areas with high demand for food delivery services were discovered, presenting the platform with opportunities for strategic expansion and increased market share.

**REFERENCES**

* 1. Dataset - <https://www.kaggle.com/datasets>
  2. Documentations –
     + <https://matplotlib.org/>
     + [https://seaborn.pydata.org/](https://scikit-learn.org/stable/modules/)
     + <https://pandas.pydata.org/docs/user_guide/>
  3. AP Skill Development Corporation (2023) – SRM Data Analysis Summer Internship – [https://github.com/AP-Skill-](https://github.com/AP-Skill-Development-Corporation/SRM-Data-analysis-summer-internship-2023/) [Development-Corporation/SRM-Data-analysis-summer-](https://github.com/AP-Skill-Development-Corporation/SRM-Data-analysis-summer-internship-2023/) [internship-2023/](https://github.com/AP-Skill-Development-Corporation/SRM-Data-analysis-summer-internship-2023/)