**RESTAURANT DATA ANALYSIS – USING PYTHON**

**Introduction:**

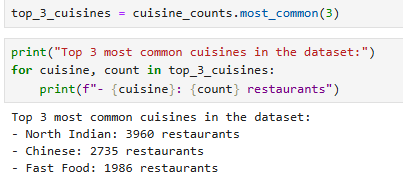
Presenting a comprehensive data analysis of the restaurant industry. This project was undertaken as part of my data analyst internship with **Cognifyz Technologies.**

Our analysis has been conducted using Python with key libraries such as pandas for data manipulation, and matplotlib, seaborn, and statsmodels for visualization and statistical modeling. Through this exploration, we seek to answer critical questions about the current restaurant landscape, identify emerging patterns, and ultimately offer data-driven recommendations.

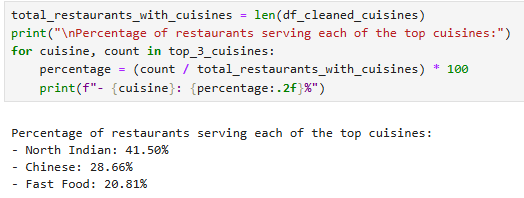
**LEVEL 1**

✔️ **Task 1**

* Determine the top three most common cuisines in the dataset.

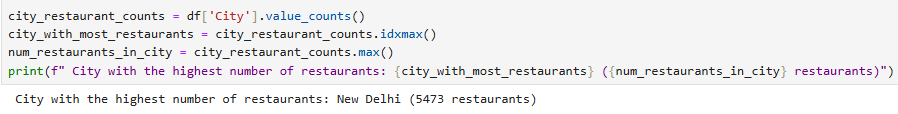


* Calculate the percentage of restaurants that serve each of the top cuisines.

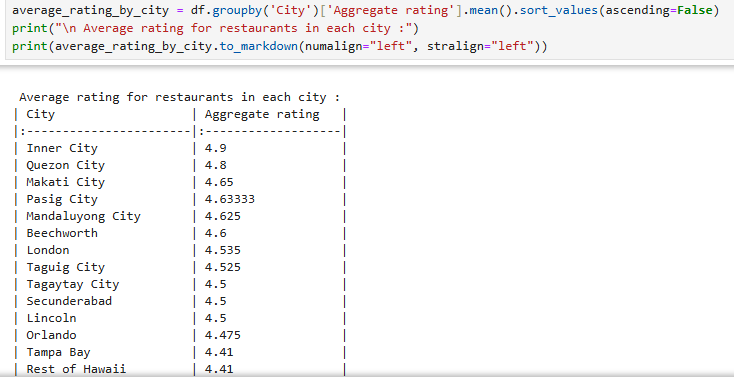


✔️ **Task 2**

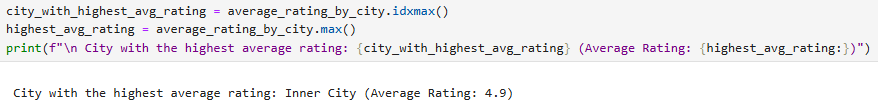
* Identify the city with the highest number of restaurants in the dataset.



* Calculate the average rating for restaurants in each city.

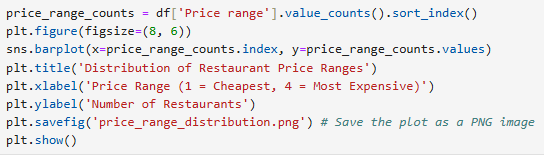


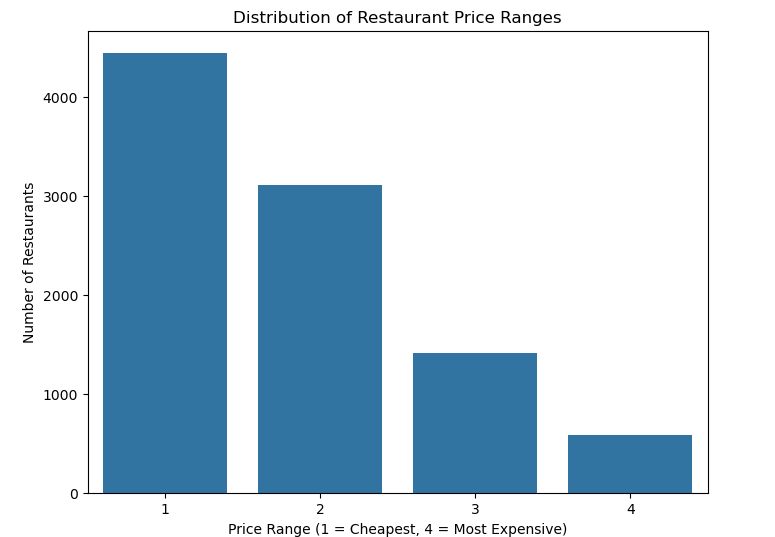
* Determine the city with the highest average rating.



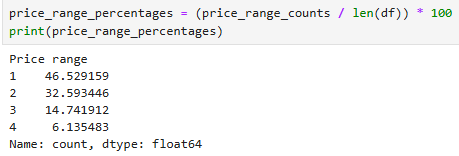
✔️ **Task 3**

* Create a histogram or bar chart to visualize the distribution of price ranges among the restaurants.



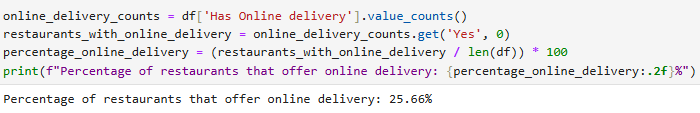


* Calculate the percentage of restaurants in each price range category.

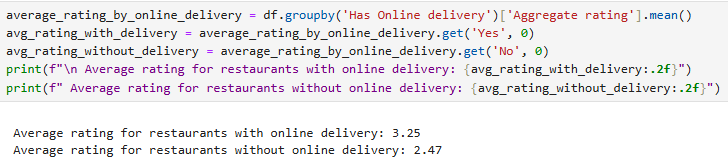


✔️ **Task 4**

* Determine the percentage of restaurants that offer online delivery.



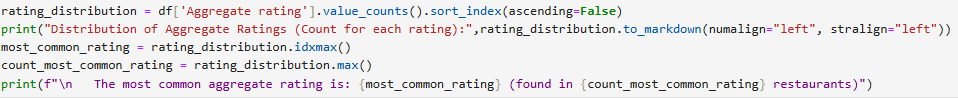
* Compare the average ratings of restaurants with and without online delivery.

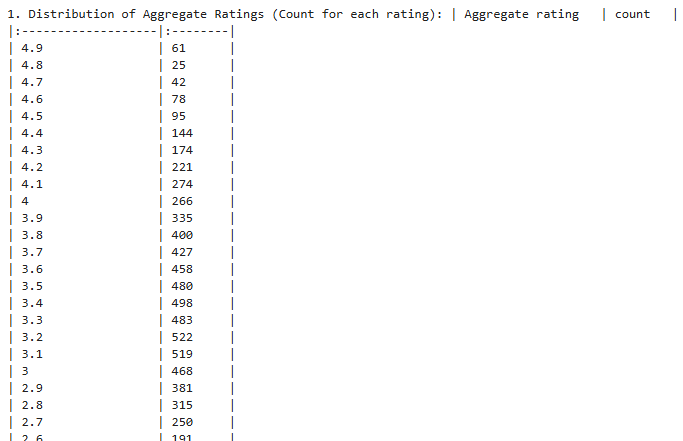


**LEVEL 2**

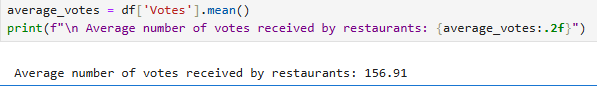
✔️ **Task 1**

* Analyze the distribution of aggregate ratings and determine the most common rating range.



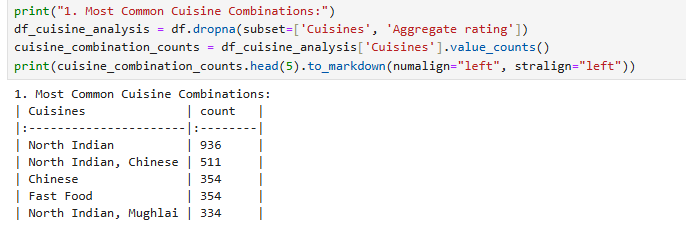


* Calculate the average number of votes received by restaurants.

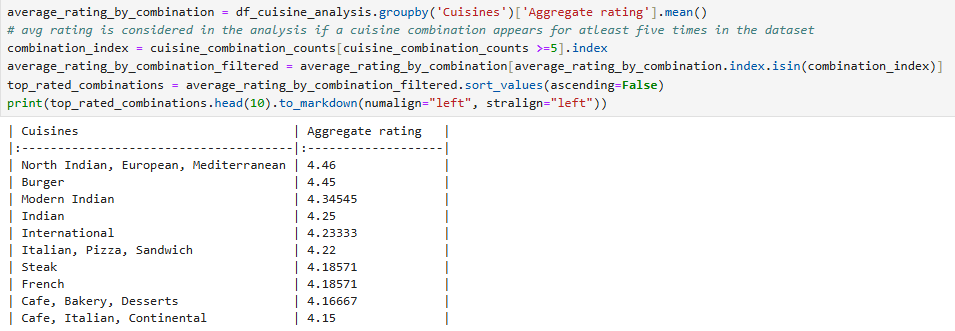


✔️ **Task 2**

* Identify the most common combinations of cuisines in the dataset.

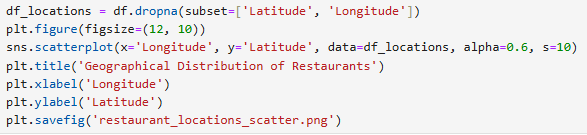


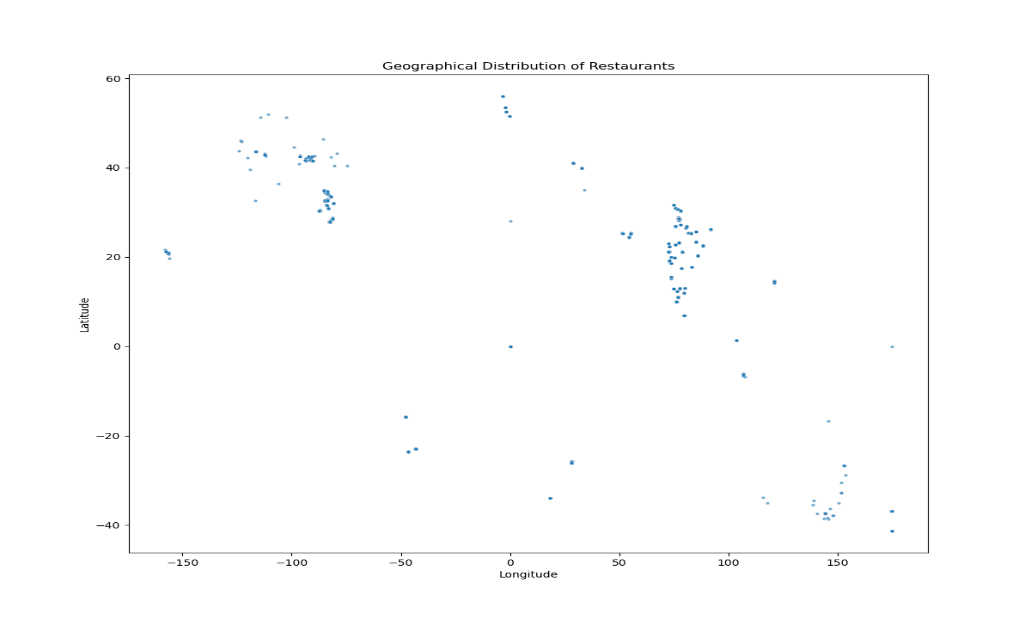
* Determine if certain cuisine combinations tend to have higher ratings.



✔️ **Task 3**

* Plot the locations of restaurants on a map using longitude and latitude coordinates.



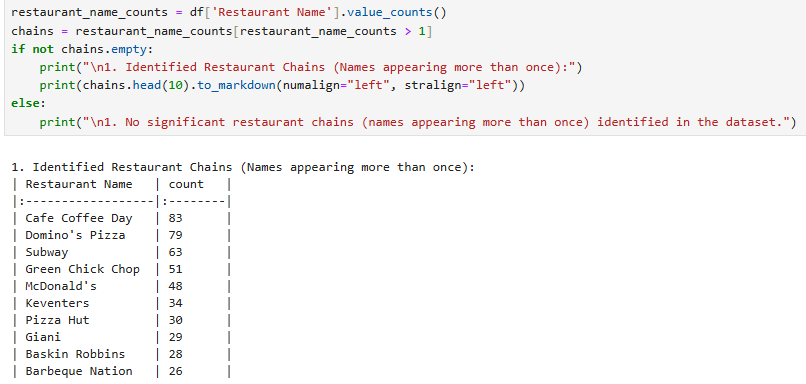


* Identify any patterns or clusters of restaurants in specific areas

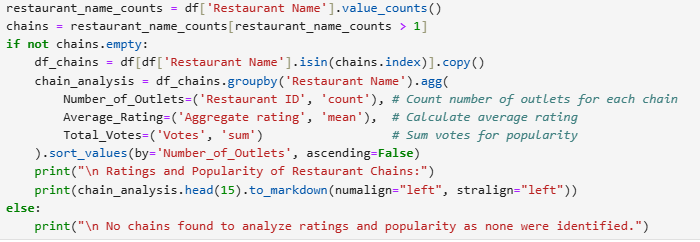


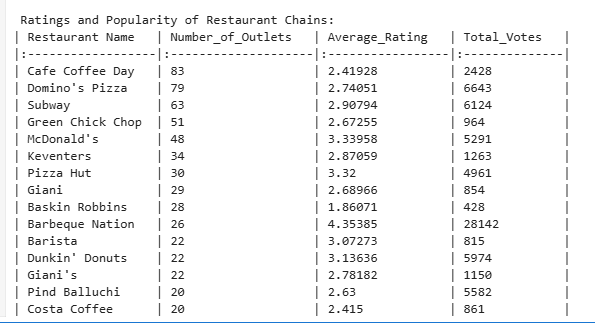
✔️ **Task 4**

* Identify if there are any restaurant chains present in the dataset.



* Analyze the ratings and popularity of different restaurant chains.

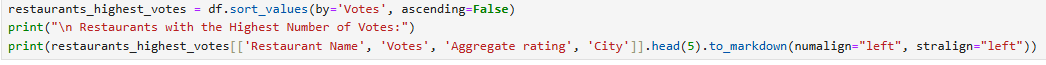


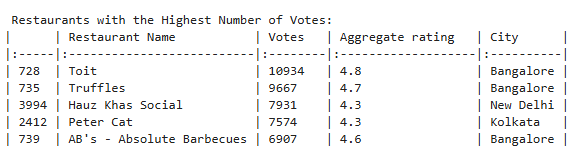


**LEVEL 3**

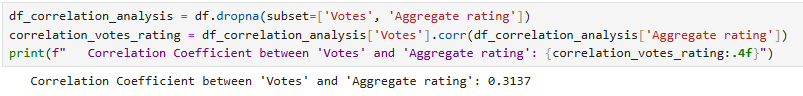
✔️ **Task 2**

* Identify the restaurants with the highest and lowest number of votes.





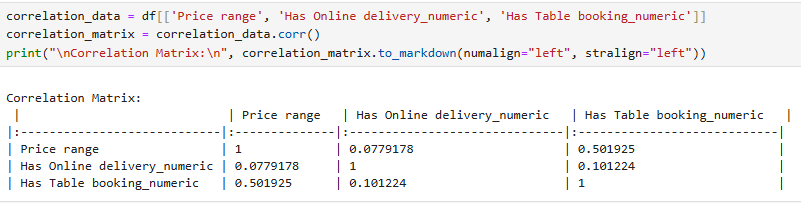
* Analyze if there is a correlation between the number of votes and the rating of a restaurant.



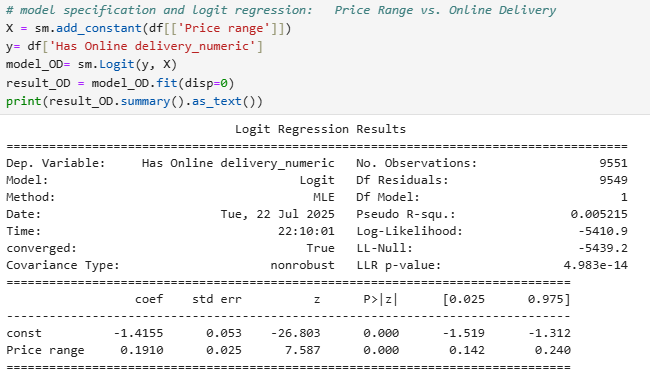
* **weak to moderate positive linear relationship between Votes and Aggregate Rating**

✔️ **Task 3**

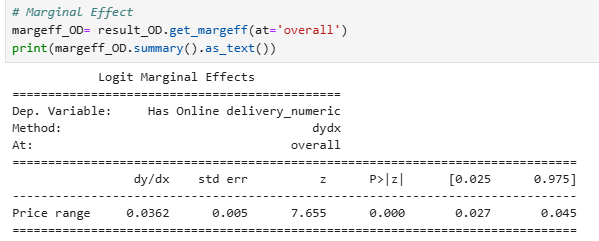
* Analyze if there is a relationship between the price range and the availability of online delivery and table booking.



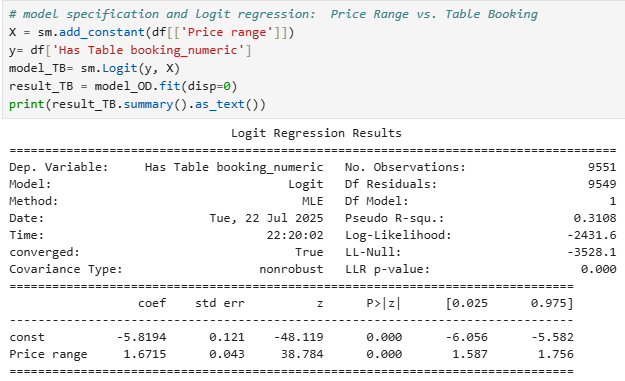
* **Price Range and Has Online Delivery: very weak positive correlation**.
* **Price Range and Has Table Booking: moderately strong positive correlation.**
* Determine if higher-priced restaurants are more likely to offer these services.



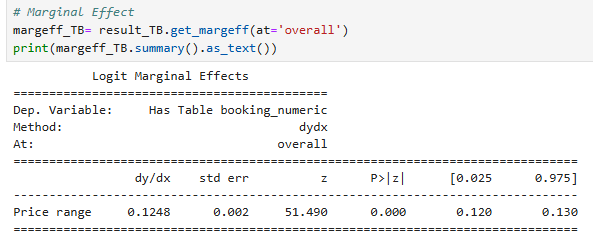
* Examining the influence of 'Price range' on the likelihood of offering Online Delivery.
* Overall Model Significance (LLR p-value): A very small p-value (< 0.05) suggests 'Price range' has a significantly better explanatory capacity in this regard.
* Coefficient for Price Range (**0.1910**): In a logit model, a positive coefficient means that as price range increases, the log-odds of having online delivery increase. The p-value associated with the 'Price range' coefficient suggests that it is highly significant.



* On average, for every one-unit increase in the 'Price range', the probability of a restaurant offering online delivery increases by approximately **0.0362, or 3.62 percentage points**.



* Examining the influence of 'Price range' on the likelihood of offering Table Booking.
* Overall Model Significance (LLR p-value): A very small p-value (0.000) which is (< 0.05) suggests 'Price range' has a significantly better explanatory capacity in this regard.
* Coefficient for Price Range (**1.6715**): The positive coefficient means that as price range increases, the log-odds of having table booking increase. The p-value associated with the 'Price range' coefficient suggests that it is highly significant.



* On average, for every one-unit increase in the 'Price range', the probability of a restaurant offering online delivery increases by approximately **0.1248, or 12.48 percentage points**.
* The marginal effect for table booking (12.48 percentage points) is substantially larger than for online delivery (3.62 percentage points), indicating that price range has a much stronger positive association with the availability of table booking.
* In summary, higher-priced restaurants are indeed more likely to offer both online delivery and, more notably, table booking services. This suggests that these services might be considered standard offerings or a strategic decision for establishments in higher price categories.

**💡 Actionable Recommendations:**

* **For New Restaurants**: Consider focusing on North Indian, Chinese, or Fast Food if targeting a broad market, but also investigate niche cuisine combinations that tend to have higher ratings for potentially better customer satisfaction.
* **For Existing Restaurants:** If your restaurant is in a higher price range, offering table booking is likely an expected service by customers. Online delivery can also positively impact ratings.
* **For Market Analysis:** Cities like New Delhi are saturated with restaurants, while areas like Inner City show a strong preference for highly-rated establishments.
* **For Investment:** Restaurant chains like Barbeque Nation show high popularity and good average ratings, indicating strong brand recognition and customer satisfaction.
* Restaurants aiming for higher ratings should focus on strategies to encourage customer engagement and reviews, as there's a positive relationship between votes and aggregate rating.