Cognifyz: Data Analyst Internship

Dataset :

- **Columns :**
  - **Restaurant ID :** this is the ID of the Restaurant
  - **Restaurant Name :** Name of the restaurant
  - **Country Code :** code of the country
  - **City :** the city customer live in it
  - **Address :** The Full Address of the Customer
  - **Locality :** the address of the City
  - **Locality Verbose** : the long range of the address of City
  - **Longitude , Latitude :** coordinates
  - **Cuisines :** the kind of the Kitchen(French – german – Japan- desserts)
  - **Average Cost for two :** the average cost of 2 kinds of Kitchens
  - **Currency :** the kind of coin
  - **Has Table booking :** have a booking in the sceduale
  - **Has Online delivery :** have a online delivery ?
  - **Is delivering now :** are the delivering now ?
  - **Price range & Price Range Levels :** the pricing options or categories available at each restaurant. It helps users understand the cost range they can expect when dining at a particular restaurant. The column likely contains values such as "poor" , "Low," "Medium," , "High" to indicate the price range options.
  - **Aggregate rating :** the rate of the restaurant
  - **Rating color** : the colors that customer rating it
  - **Rating text :** here customer rate by using word
  - **Votes :** how many people votes for each restaurant ?

Data prepare :

1. Search about Duplicate Values : Done
2. Search about NullValues & remove it : Done
3. DataType of each column : Done
4. Remove useless Columns : Done
5. remove Null Values or [column numeric : replace with mode or median , Categorical : common features or Not Defined] or fill it with 0
6. Rename Columns : Done
7. Check each row is organized and categorized : Done
8. Remove unnecessary Formats : Done

Business Questions :

- **Level 01 :**
  - **Task01: Top Cuisines :**
    - **(Q1) Determine the top three most common cuisines in the dataset ?**
      - Here we find the best 3 common cuisines in all Dataset this depend on different Features like [Voting – rating – rate Name – count customer] so after some analysis I found the top 3 common cuisines are(North Indian, Mughlai - North Indian - North Indian, Chinese) BUT if we use other features we will find the best common cuisines are (North indian - North Indian ,Chinese - Chinese - North Indian , Mughlai - Café – Italian)

    - **(Q2) Calculate the percentage of restaurants that serve each of the top cuisines ?**
      - It was [barbeque Nation – pind Balluchi – Punjab Grill ……] these restaurant serva the most or

the top Cuisines you can find more restaurants in Dashboard Level 1

- Task02: City Analysis
  - **(Q3) Identify the city with the highest number of restaurants in the dataset ?**
    - The city that have most number of restaurants where number restaurant reach to more than 5000 restaurant was [New Delhi] and there are more cities have more than 1000 restaurants like [Gurgoan – Noida]
  - **(Q4) Calculate the average rating for restaurants in each city. ?**
    - The average rating for restaurant was between [1.8 – 4.9] and I found there are restaurant have high rating like (inner city – queozencity – makati city – passig city – mandaluyoung city – beechworth city – London – Taguig city - lincoin – Secunderabad – tagytay) these city have rating between [4.5 – 4.9] so I think they are the best restaurants
  - **(Q5) Determine the city with the highest average rating…?**
    - I found there are restaurant have high rating like (inner city – queozencity – makati city – passig city – mandaluyoung city – beechworth city – London – Taguig city - lincoin – Secunderabad – tagytay) these city have rating between [4.5 – 4.9] so I think they are the best restaurants
- Task03: Price Range Distribution
  - **(Q6) Create a histogram or bar chart to visualize the distribution of price ranges among the restaurants .?**
    - We have 2 ranges or intervals ([1 – 2] , [3 – 4]) and after make analysis on it I found the interval [1 – 2] is

most common more than the other interval where values were (7600k , 200K) that mean the most resturants of price in range [1 – 2] price range

- **(Q7) Calculate the percentage of restaurants in each price range category ?**
  - I tried to divide the ranges in these comparssion [poor price (1) – lower price (2) – medium price (3) – high price (4)] , so I found the most common kind of price range was poor price with rate 46.53% and come after it lower price by rate 14.75 % and after it high price 6.14% , so we need to focus on the poor price category as it's more common where the customer prefer it

- Task04: Online Delivery
  - **(Q8) Determine the percentage of restaurants that offer online delivery..?**

    *   After conducting an analysis of customer preferences regarding online delivery, it was observed that a majority of customers, approximately 74.34%, do not opt for online delivery. This indicates that not having online delivery is the more common choice among customers.However, it is worth noting that the remaining customers, accounting for the remaining 25.66%, do prefer online delivery. This suggests that there is still a significant portion of the customer base that values the convenience and accessibility offered by online delivery services.These findings shed light on the importance of providing online delivery options for restaurants. While a majority of customers may not currently utilize this service, there is a substantial market segment that does. By offering online delivery, restaurants have the opportunity to cater to the needs and preferences of these customers, potentially increasing their customer base and overall satisfaction.These insights are

valuable for businesses in the food industry as they make decisions regarding the implementation of online delivery services.

- **(Q9) Compare the average ratings of restaurants with and without online delivery.?**
  - **According to the data analysis, it was observed that the average rating of restaurants offering online delivery is more prevalent compared to those that do not provide this service. The results indicate that approximately 56.86% of the restaurants with online delivery have higher ratings, while 43.14% of the restaurants without online delivery have lower ratings.This finding suggests a positive correlation between online delivery services and customer satisfaction, as indicated by the higher average ratings. It is important to note that the availability of online delivery may contribute to increased convenience and accessibility for customers, which could positively impact their overall dining experience.These results provide valuable insights into the preferences and expectations of customers when it comes to restaurant services.**

- **Level 02 :**
  - **Task01: Restaurant Ratings:**
    - **(Q1) Analyze the distribution of aggregate ratings and determine the most common rating range?**
      - **During the analysis process, the aggregating rating was divided into intervals to make the distribution more understandable. It was found that the majority of customers, around 21,000, rated restaurants within the range of 2 to 4. This indicates that the most common rating for restaurants falls within this interval.However, when looking at the higher ratings between 4 and 5, there were**

only 5,000 customers who rated restaurants within this range. This suggests that fewer customers gave ratings in the higher range.On the other hand, the lowest ratings between 0 and 2 had even fewer customers. This implies that there were a lower number of people who rated restaurants within this range.In summary, the analysis reveals that the most common rating for restaurants falls within the range of 2 to 4, while the higher ratings (4 to 5) and lower ratings (0 to 2) had relatively fewer customers.

- **(Q2) Calculate the average number of votes received by restaurants?**
  - After conducting the analysis, it was found that the following ten restaurants had the highest number of votes, arranged in descending order: Toit, Hauz Khas Social, Peter cat, Big Brewsky, The Black Pearl, BarBQ, Warehous Café, Truffles, Mocambo, and Gulati. These restaurants are more commonly preferred by customers compared to others.This information indicates that these ten restaurants have managed to attract a larger customer base and receive a higher number of votes. It suggests that customers have a strong preference for these establishments, possibly due to factors such as food quality, ambiance, service, or overall dining experience.Identifying these popular restaurants can be valuable for businesses in the food industry as they can gain insights into the factors that contribute to customer preference and satisfaction. By understanding the reasons behind the popularity of these establishments, other restaurants can potentially improve their offerings and strive to meet customer expectations.

- **Task02: Cuisine Combination**

- **(Q3) Identify the most common combinations of cuisines in the dataset ?**
  - According to the analysis, the most prevalent combinations of cuisines, ranked in descending order, are as follows:
    - North Indian and Mughlai cuisine, accounting for 27.24% of the occurrences.
    - North Indian and Chinese cuisine, representing 24.82% of the combinations.
    - North Indian cuisine alone, comprising 21.29% of the total.
    - Chinese cuisine, contributing to 15.54% of the combinations.
    - Café cuisine, constituting 11.11% of the occurrences.

  These findings indicate that the combination of North Indian and Mughlai cuisines is the most frequently observed, followed by North Indian and Chinese cuisines. It is important to note that these results are specific to the analyzed dataset and may vary in different contexts or datasets.

- **(Q4) Determine if certain cuisine combinations tend to have higher ratings.. ?**
  - Based on a comprehensive analysis, specific cuisine combinations have emerged with consistently higher ratings than others. Notably, the average ratings indicate that combinations such as "North Indian - North Indian," "Chinese - North Indian," and "Mughlai - Café - Chinese" exhibit the highest levels of satisfaction among consumers. These findings underscore the preference for these particular culinary fusions, reflecting a noteworthy trend in elevated customer satisfaction. The discernment of these high-performing combinations provides valuable insights for culinary establishments seeking to optimize their menu offerings and enhance overall customer experience.

- o **Task03: Geographic Analysis**
  - ▪ **(Q5) Plot the locations of restaurants on a map using longitude and latitude coordinates. .?**
    - *** I make it and it show me a lot of restaurants in different countries**
  - ▪ **(Q6) Identify any patterns or clusters of restaurants in specific areas. ?**
    - • **The spatial analysis of restaurant distributions reveals discernible clusters within specific localities, each characterized by a distinct culinary profile. These clusters are defined as follows:**
    - • **Waikiki:Culinary Profile: American, Asian, Burger**
    - • **Golf Course Road :Culinary Profile: American, Continental, Burger**
    - • **Athens:Culinary Profile: American, Burger, Sandwich**
    - • **Hyde Park:Culinary Profile: American, Desserts, Steak, Burger**
    - • **Windermere:Culinary Profile: [Specific cuisine profiles not provided]**
    - • **The Dubai Mall, Downtown Dubai:Culinary Profile: American, Desserts**
    - • **Restaurant Row:Culinary Profile: American**
    - • **Golf Course Road**
    - • **Culinary Profile : American, Continental, Burger**
      This categorization based on localities and their associated culinary profiles offers a strategic perspective for stakeholders aiming to understand the gastronomic landscape within specific regions. Such insights can inform decision-making processes related to marketing strategies, menu planning, and the establishment of new culinary ventures, fostering a nuanced understanding of consumer preferences in diverse locales.

- Task04: Restaurant Chains
  - (Q7) Identify if there are any restaurant chains present in the dataset...?
    - In the comprehensive evaluation of the culinary landscape, prominent restaurant chains have emerged as significant contributors to the gastronomic milieu. Among the multitude of chains, a discerning analysis reveals the prevalence of certain establishments that stand out for their widespread presence. The ten most frequently encountered restaurant chains, indicative of their ubiquity and consumer appeal, are as follows:
      - Café Coffee Day
      - Domino's Pizza
      - Subway
      - Green Chick Chop
      - McDonald's
      - Pizza Hut
      - Giani
      - Baskin Robbins
      - Barbeque Nation

      This identification of the most common restaurant chains not only underscores their market prevalence but also serves as valuable intelligence for stakeholders seeking insights into the competitive landscape and consumer preferences. This knowledge is pivotal for strategic decision-making, marketing endeavors, and understanding the dynamics of the contemporary culinary market.

  - (Q8) Analyze the ratings and popularity of different restaurant chains..?

In the culmination of a meticulous analysis, certain restaurants have emerged as exemplars of high popularity, gauged by a confluence of both elevated ratings and a substantial volume of votes. The foremost establishments, distinguished by their commendable reputation and significant patron engagement, are as follows:

- Toi
    - Total Votes: 10,934
- Hauz Khas Social
    - Total Votes: 7,931
- Peter Cat
    - Total Votes: 7,574
- Truffles
    - Total Votes: 4,841
- Joey's
    - Total Votes: 3,903

These establishments not only command noteworthy ratings but also have garnered a substantial number of votes, indicative of their widespread appeal and active patronage. Such insights are pivotal for industry stakeholders seeking to understand the dynamics of consumer engagement and make informed decisions in the competitive landscape of the culinary domain.

- **Level 03 :**
    - **Task01: Restaurant Reviews:**
        - **(Q1) Analyze the text reviews to identify the most common positive and negative keywords?**
            - **In the analysis of textual ratings provided by patrons, a discernible pattern emerges, showcasing the prevalence of specific descriptors that encapsulate the overall sentiment toward various establishments. The most commonly**

encountered textual ratings, indicative of the collective sentiment expressed by diners, are:

- Average
- Not Rated
- Good
- Very Good

These descriptors, recurrent in the assessments of diverse dining experiences, serve as key indicators of the nuanced feedback provided by patrons. Understanding the prominence of these specific textual ratings affords stakeholders valuable insights into the perceived quality of culinary offerings and aids in formulating strategies to enhance customer satisfaction.

- **(Q2) Calculate the average length of reviews and explore if there is a relationship between review length and rating.?**
    - Average of length of reviews was 7 , but after we make a scatter plot between review length and rating I found there is no any relation between them

- **Task02: Votes Analysis**
    - **(Q3) Identify the restaurants with the highest and lowest number of votes?**

        For the highest-ranked restaurants:

        - **Barbeque Nation (28142 votes):**
            - Example: Barbeque Nation's substantial vote count suggests it enjoys a high level of patronage.
            - Use case: The restaurant's popularity may be attributed to its unique dining experience, diverse menu, or exceptional customer service.
        - **AB's (13400 votes):**
            - Example: AB's has garnered a considerable number of votes, reflecting positive customer opinions.

- Use case: Investigate what aspects of AB's, such as its cuisine, ambiance, or location, contribute to its popularity.
- **Toit (10934 votes):**
  - Example: Toit's high vote count implies a strong customer base and satisfaction.
  - Use case: Explore Toit's menu, promotions, or any specific features that attract customers and enhance their dining experience.
- **Big Chill (10853 votes):**
  - Example: Big Chill's substantial votes suggest it is a favored choice among customers.
  - Use case: Look into the unique selling points of Big Chill, such as its menu variety or nostalgic ambiance, to understand its appeal.
- **Farzi Café (10098 votes):**
  - Example: Farzi Café's four-digit vote count highlights its popularity among diners.
  - Use case: Analyze what distinguishes Farzi Café from competitors and investigate factors contributing to customer satisfaction.
- **Truffles (9682 votes):**
  - Example: Truffles' significant votes indicate a positive perception among the voting audience.
  - Use case: Research Truffles' offerings, customer service, or marketing strategies that contribute to its success.
- **For the lowest-ranked restaurants:**
  - #Hastage, 13 Café, 44 Grills, 6 Packs Moms, 7 Colours Xpress Kitchen:

- **Example: These establishments received comparatively lower votes, suggesting areas for improvement.**
- **Use case: Investigate customer reviews or feedback to identify specific issues and devise strategies for enhancement.**

- **(Q4) Analyze if there is a correlation between the number of votes and the rating of a restaurant.?**
  - **There are a strong relationship between the number of votes and rating of restaurants so when the number of voting increase so the aggregating will increase**

- **Task03: Price Range vs. Online Delivery and Table Booking**
  - **(Q5) Analyze if there is a relationship between the price range and the availability of online delivery and table booking.. .?**

    * there is no any relationship between them

  - **(Q6) Determine if higher-priced restaurants are more likely to offer these services. ?**
    - **It seems like you've conducted an analysis and found that restaurants within the price range of [3 - 4] have a higher rating of 56.11%, suggesting that they may be more likely to offer services compared to those in the [1 - 2] price range with a rating of 43.89%.**
    In a professional context, this information can be valuable for decision-making and strategic planning. Here's how you might convey this in a more structured manner:

    Analysis Summary:

    1- Price Range [3 - 4]:
      - **Rating: 56.11%**
      - **Observation: Restaurants within the [3 - 4] price range exhibit a higher rating, indicating a positive**

correlation between pricing and customer satisfaction.

- **Implication: This suggests that establishments in this price range may be more inclined or better equipped to offer services that contribute to higher customer satisfaction.**

2- **Price Range [1 - 2]:**

- **Rating: 43.89%**
- **Observation: Restaurants in the [1 - 2] price range have a lower rating compared to the [3 - 4] range.**
- **Implication: There may be opportunities for improvement or strategic adjustments in service offerings for restaurants in this price range to enhance customer satisfaction.**

3- **Recommendations:**

- **Consider investigating the specific services that contribute to higher ratings in the [3 - 4] price range. Identify key factors driving customer satisfaction.**
- **Explore potential service enhancements or adjustments for restaurants in the [1 - 2] price range to align with customer expectations and improve ratings.**

- **Continuously monitor customer feedback and adapt service strategies accordingly to maintain or improve overall satisfaction levels.**

Thanks You ☺