Problem Statement :

Assume that a restaurant owner has aggregated 10,000 reviews for their restaurant from your website over the past year and are having a hard time synthesizing and summarizing their reviews into actionable insights that could help them make better decisions to improve their restaurant.

If you are the restaurant owner -
● List the problem statements that you'd have to solve, and
● What solutions you can envision and what models you would build that helps this restaurant owner obtain actionable insights from reviews written by people.

Solution:

Platform Used - Jupyter Notebook
Language - Python
Data set used - https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants
Learning Method - Supervised Learning
Approach - Natural Language Processing

About Dataset :
As a restaurant owner, I am using the Zomato dataset specifically to get a fair idea about the factors affecting the establishment of different types of restaurants at different places in Bengaluru, the aggregate rating of each restaurant. This dataset contains 12,000 restaurants with restaurants serving dishes from all over the world.

Questions to ask as a restaurant owner

1. Percentage of positive and negative reviews given by the customers for the company
2. The cause behind the positive and negative reviews
3. Keep working on the positive aspect as usual and doing EDA over the negative reviews
4. Find the cause behind the negative review specified to my company. It may be some cuisine, behavior of the employees, time taken to place the order
5. Do some competitive analysis as to why other restaurants are doing better than us. Maybe add some new cuisines, locations, or poor marketing.
6. The amount of profit I can expect after spending on some key data points. Will it be beneficial to expand for me or I will face loss.

I have pointed out the questions as a restaurant owner in a business perspective. Now through this dataset, I want to analyze some key features such as:

• Location of the restaurant
• Approx Price of food
• Theme based restaurant or not
• Which locality of that city serves that cuisines with the maximum number of restaurants
• The needs of people who are striving to get the best cuisine of the neighborhood

• Is a particular neighborhood famous for its own kind of food.

For solving this problem I have used the Natural Language Processing branch of machine learning which can extensively find out outstanding insights as it can work on a huge amount of natural language data. The review which we will analyze is in the English language so it is important to specify a correct approach for this problem which can understand human language in a better way and find out useful insights from the data.

I am using a supervised learning algorithm as we are working on already collected data set and finding out a relation between the input and the output.

Procedure:

1. The process starts with loading the important python libraries which we will be using for the analysis and visualization such as pandas, numpy, matplotlib, seaborn, re for regular expression, nltk (natural language toolkit), strings, spacy (natural language processing).
2. Reading the file using pandas data-frame and printing the first few rows of the data to just get a grasp of how the data is stored, what all columns are there in the data.
3. Getting some information about the data present like how many null values are present and what are the data types of each column.
4. Doing some data cleaning by replacing the null values with the mean value in the rate column as they don't affect in the analysis
5. Dropping the rows where there are null values in other columns
6. I have done some feature extraction by dropping the non-important rows which will not affect the analysis
7. Now grouping the restaurants by their delivery type and printing a bar graph showing the rating for each type of service provided by the restaurant. This shows which delivery type we have to improve.
8. Plotted a bar graph which shows the average cost of 2 persons for each type of delivery service they take. The service which costs less will be preferred more by the customers so we have to expand more on that part.
9. Grouping and plotting the data to get an insight of how the location is affecting the reviews. People from which location are putting up reviews and my restaurant is located there or not.
10. Grouping the data by price and location. This gives us an idea of customers from which location is spending more on restaurants and whether my restaurant is located or not.
11. Now listing the count of dishes to find out which is the most liked dish and to check whether the liked dishes are present in my menu or I have to add them.
12. Plotting a bar graph to find out top 15 dishes most liked by the customers and to extensively add or improve them if required.
13. Finding the count of different cuisines from the review. Similarly plotting a bar graph to find out the top 15 most liked cuisines and to add or improve the quality if need be.
14. Applying NLP by first removing punctuations, numbers, and special characters so that they don't impact the analysis.
15. Now converting the whole review into a single case (lower) to get each and every word to work.

16. Doing some cleaning by removing the words with less than 3 such as pronouns as they are not that important.
17. Created stop words to treat each word as an entity.
18. Using WordCloud to determine the frequency and the importance of each word present to find out the most occurring word. This can give us an idea of which word is making an impact through the customer reviews and how are we relating to that word.
19. Now I have defined the best restaurant by taking the following parameters into account:
    No. of Votes
    Rating
    Average cost
20. This analysis helps us to get an idea of what is lacking in our restaurant and what are the factors responsible for which that restaurant is coming out the best and what we have to work on to improve.

I am attaching my GitHub repository so that you can visit and see the complete code. I have tried to keep it simple and understandable so that we can effectively draw some useful business insights and don't get confused about too much complex analysis part.

Github Repository of the project:

https://github.com/DebBiswas07/Restuarant-Review-Project.git

I have uploaded the required python notebook for the reference.

The insights drawn are as follows:

1. Top three restaurant types by rating are
   Drinks and nightlife
   Pubs and bars
   Buffet
2. Top three restaurant types by average cost for 2 persons are
   Drinks and nightlife
   Pubs and bars
   Buffet
   This shows that as people are spending more and more they are also giving more positive reviews. Maybe because they are enjoying the service available which is making them satisfied even after spending more so its value for money.
3. Top three locations by the number of votes
   Old Airport Road
   MG Road
   Indiranagar
   The locations may where people like to eat outdoor and spend outdoor. Also, they may care about giving reviews or maybe they are more socially active and the services are well provided.
4. Top three locations depending upon the average cost for 2 persons are

Church Street
MG Road
Brigade Road
These may be posh locality where restaurants are charging high. Maybe the cost of running a restaurant at these locations is high.
MG Road is common. Hence In spite of being costly people are visiting and spending their money. Hence they are experiencing satisfied service.

5. Top 3 liked dish
   Pasta
   Pizza
   Cocktails
6. Top 3 most liked cuisine
   North Indian
   Chinese
   South Indian
   Customers are craving for them the most.