**Designing a Marketing Campaign for a Restaurant Chain**

**Using Exploratory Data Analysis**



**Objective**

*To utilize exploratory data analysis (EDA) skills to understand customer preferences,*

*dining trends, and competitive landscape in various regions of India, and to design*

*an effective marketing campaign for a restaurant chain.*

**Data Cleaning and Preparation**

***● Identify and handle missing values.***

A graph with numbers and letters

Description automatically generated with medium confidence

This analysis is done by using the missingno library of panda and the function is:



The above bar chart shows the missing values in each columns. Where the bars are 100 % or 1 which means there are no missing values. For instance, in zip code there are more than 50 values are missing which we later exclude from our columns. Also, in other columns there are missing values too, like in the column of timing, cuisines, and address.

A black and white image of a bar code

Description automatically generated

Matrix function is used to determine the relationship between columns missing values (indicated as white line) which gives the idea whether the missing values follows any pattern. This pattern helps to understand if a particular value is missing then another is missing too which help in the imputation strategies to fill the data. We can see there is a connection between cuisines and timings missing values. And other columns have no relationship. Its function is:



A diagram of a graph

Description automatically generated with medium confidence

The correlation functions show the correlation between two variables, it shows the relation of missing values in graphical representation format. Correlation means either the variables are positively correlated or negatively correlated.

A black text on a white background

Description automatically generated

If correlation is positive and 1 which means if one variable is increase, then other is increasing too or vice versa. 1 means the correlation is very strong and if correlation between 0.6-0.79 which means moderate relation, and from 0 to 0.39 weak. The correlation only considers the numerical variables (not categorical variables). In above, we can see that there is weak correlation between all variables which have colors Like Cuisine and Timings has 0.1 correlation, whereas address and cuisines has 0.3 correlation. However, the correlation is weak which can be neglected. Furthermore, there are white boxes too which do not show which means there is 0 correlation.

***Errors (Wrong Entry)***

A screenshot of a computer

Description automatically generated

If the missing value is in the column, then it must be NaN or Empty space. But if it is written nan or “NaN” then it is considered as the wrong entry which needs to be imputed or removed. Similarly, the above code shows that there are incorrect entries like nan. This we must treated here we can’t take mean, mean r mode because it’s a zip code. So, for changing nan with NaN, we have used **np.nan.** This thing changed from nan to NaN (missing values). After doing this we can see in the picture there is NaN in the column which means now it is replaced and considered as the missing value.

A screenshot of a computer

Description automatically generated

In this example we can see that we have followed the same steps and then fill NaN with mean value. After finding unique there is no nan there.

***Mislabeled Categories***

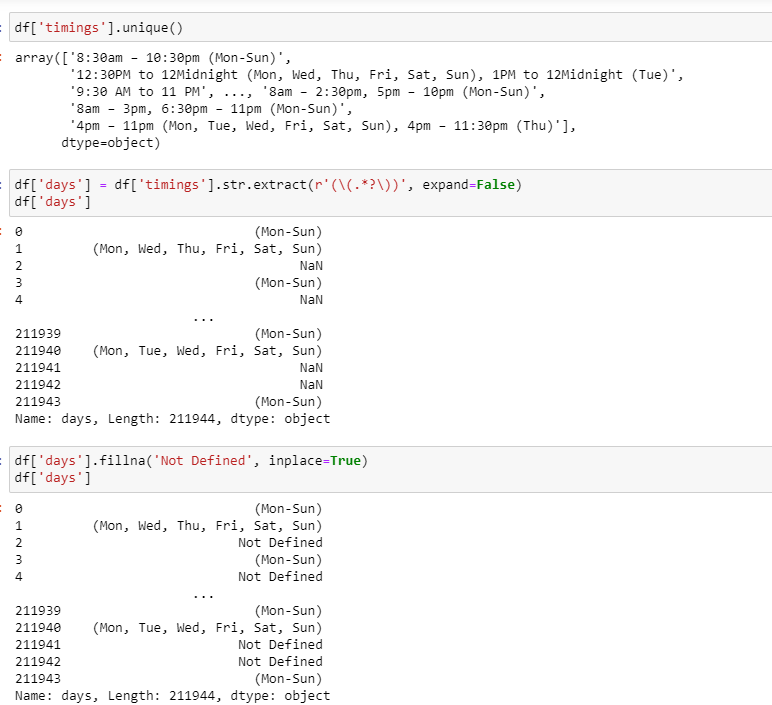
**A screenshot of a computer screen

Description automatically generated**

In the above picture we can see there are words which are in different languages which can’t be understood while doing analysis. Therefore, we must translate them into English:

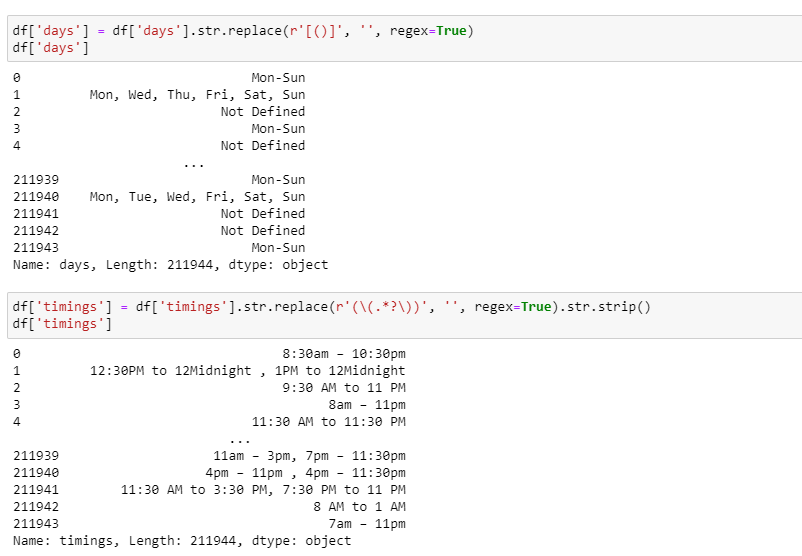


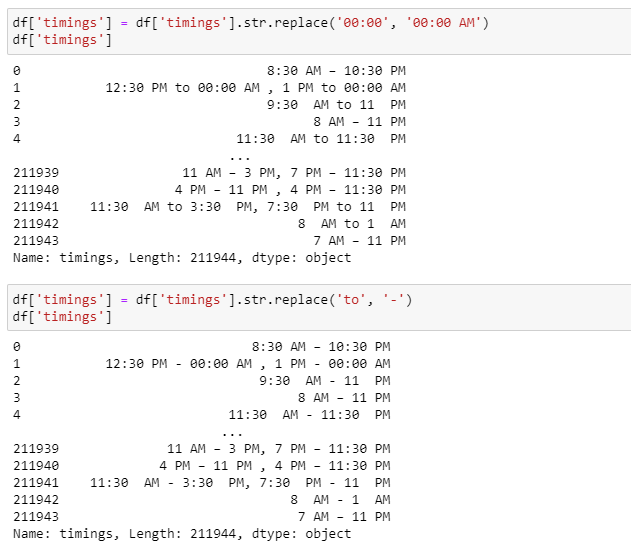
**Extract Information from existing column and make a new column**



In the above image we can see the column timing has combined information of timings and days. Also inconsistent format like Pm PM or pm etc. So after finding the issues in the next code of extract we extract days into another columns and where days are not specified we fill it with Not Defined.

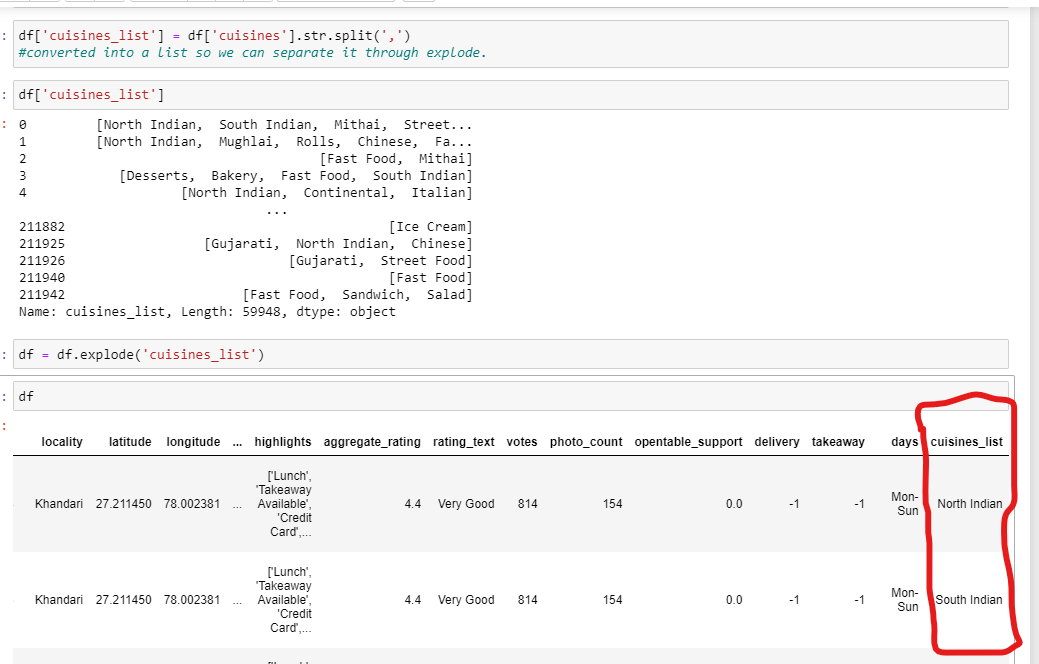
More cleaning has been done like removing brackets:





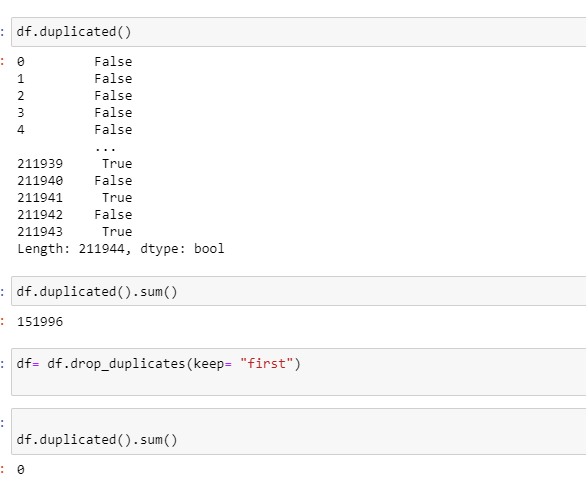
***Explode function***

Explode function is used to convert each element of the specified columns into a row. After this, each cuisine is shown in a separate row. Here we must consider one thing explode increase the number of rows as the same row add up with different cuisine (Highlighted into red). This will be helpful for each cuisine visualization.



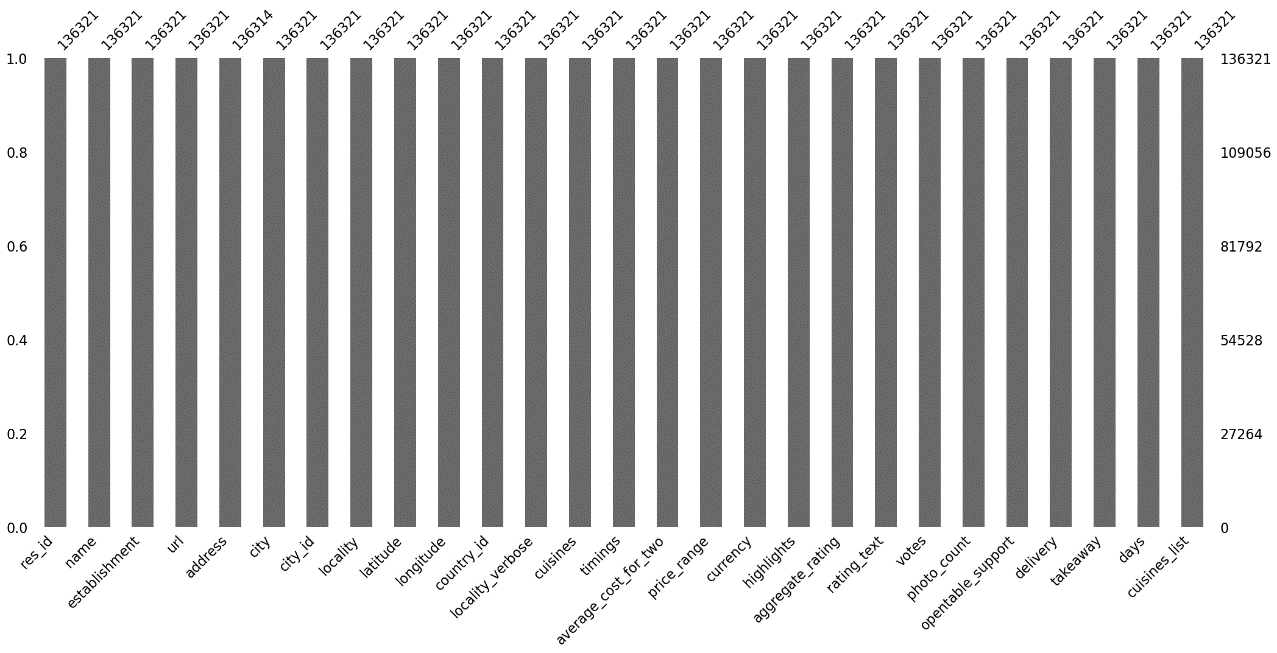
***Duplicated Values:***

We also have to remove duplicated values but after checking the context like if a column has age then two or more people have same ages. In row context, if the full row has duplicates then we keep only first one and remove others.



***After Cleaning***

After Cleaning our bar chart shows no missing values.



A black and white striped background

Description automatically generated with medium confidence

**Exploratory Data Analysis**

**Descriptive Statistics**

**A screenshot of a computer

Description automatically generated**

Description shows the statistical information of the numerical columns. In the first table we can see each column’s total values, column average value, standard deviation which means how far the values are from mean, the minimum and maximum values. For instance, in votes there are 136321 values, and its meaning in 355.64. However, its standard deviation is 901.94 which means the it values are spread widely across the mean. Higher standard deviation against means there is possibility of outliers or a skewed distribution.

***A screenshot of a computer

Description automatically generated***

This is also showing the analysis but for object columns (categorical columns). For instance, the column means, where count means number of values, unique represents the unique number of values in the column, top means the name which appeared the most, whereas frequency means how many times the top value appears.

Furthermore, “. Shape” shows the rows and columns. After exploding the data size increased here from 59948 to 136321. There are additional columns too name: cuisine list and days.

***Distribution Analysis***

A graph with blue squares

Description automatically generated

This analysis shows the aggregate rating frequency. The distribution of data is nearly normal distribution. Here we are not removing the 0 because it shows the data which has not rated by the customers, and it is not considered as the outliers.

A green bar graph with white text

Description automatically generated

This bar chart shows which rating is given by customers to any restaurant or its cuisine.

A graph with blue bars

Description automatically generated

It shows the distribution of each Price. The distribution is positively skewed. so, in featuring engineering we will do normalization. If the histogram is bell shaped then we will used z-score for feature engineering.

A graph with a number of blue and black lines

Description automatically generated

*Cuisine List Distribution chart*

Here we can see the distribution is positively skewed in cuisine list distribution. As it is a categorical column and its distribution, we will use nominal. Coding method i.e. one hot coding for feature engineering so our model will understand it.

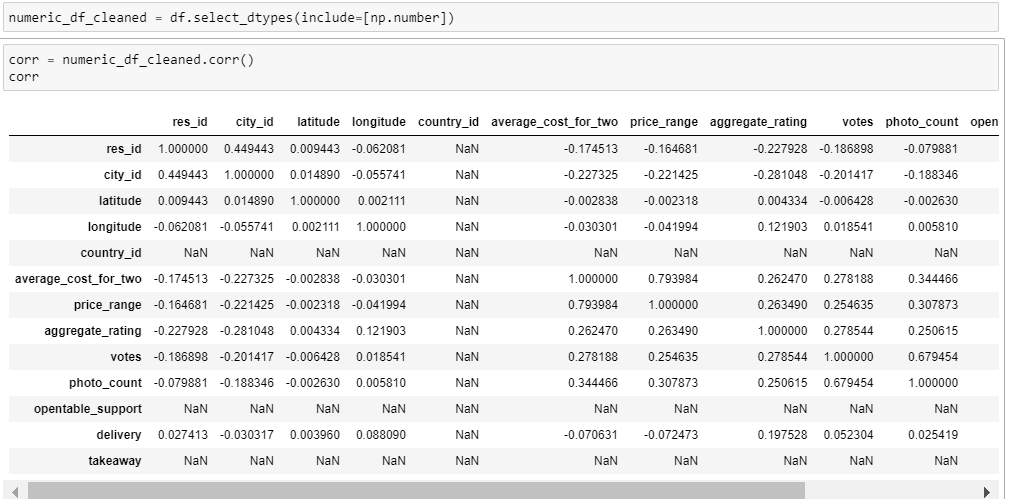
A graph of different types of food

Description automatically generated

This bar shows North Indian is one of the most famous cuisine in India across all india followed by Chinese, Fast Food, Desserts. Here one more thing we can see North Indian is repeating again with North Indian. We should use fuzzy fuzzy library for this to make our words, cases same.

***Correlation Analysis: Examine the Relationships between Different Variables.***

First, we calculate the correlation:



Then we will make the visualization:

A screenshot of a computer

Description automatically generated

Here in this heatmap we can see the correlation between variables. Here we can see there is correlation between variables but a weak correlation. However, in the two columns there is a strong correlation which **is Price Range and Average Cost of two.** The other one is photo and photocount. If we want to perform a linear regression for these two, it should have low correlation The reason is that in liner regression it must be necessary, that each feature is independent and directly contributed to the target variable which is dependent variable. For this we must treat them before doing linear regression in Machine Learning.

**Regional Analysis**

*Compare the restaurant trends and customer preferences across different cities or regions in India.*

**Average Cost for Two by Cities and Top 10**

***A screenshot of a computer

Description automatically generated***

A graph of blue lines

Description automatically generated

The above graph shows the average cost for two by city across India and it is highest in Mumbai **which means the average cost for two for cuisines are highest in Mumbai.** Its another meaning is that the cuisines prices are higher in Mumbai as compared to others.

A screenshot of a computer code

Description automatically generated

A graph of blue and white bars

Description automatically generated

This graph only shows the top 10 cities with average cost where we can see the sequence wise other cities.

***Average Rating by City and Top 10***



A graph of blue lines

Description automatically generated

The above graph shows the average rating mean by city across India.

A screenshot of a computer code

Description automatically generated

A graph with blue and white vertical lines

Description automatically generated

Here we can see the average rating for two in Top 10 cities of India. We are seeing the people are giving good rating to restaurants or cuisines followed by Mumbai and New Dehli. However, there are those cities as well which gives low rating too like Alappuzha and Pushkar. Their low rating is below 1. Maybe they don’t like the cuisines or quality of the cuisine.

**Customer Preference Analysis**

*Analyze the types of cuisines that are popular in different regions*

***Total Number of Cuisine By City and Top 10***

**A screenshot of a computer program

Description automatically generated**

A graph with blue lines

Description automatically generated

**A screenshot of a computer program

Description automatically generated**

A graph of blue bars

Description automatically generated

Highest number of cuisines are in Mumbai followed by Chennai Bangalore. It means the more variety is in these cities. There are other cities which have low cuisines which are Graeter Noida and others.

A screenshot of a computer code

Description automatically generated

A graph of different countries/regions

Description automatically generated with medium confidence

A screenshot of a computer code

Description automatically generated

A graph of different types of cuisines

Description automatically generated

We saw most popular cuisine across India is North Indian where particularly in Mumbai it is Fast Food and Chinese. It shows which cuisine the people of Mumbai likes the most.

***Examine the relationship between restaurant ratings, price range, and popularity.***

**A screenshot of a computer program

Description automatically generated**

A red and blue squares

Description automatically generated

The above correlation matrix shows the relationship between three variables i,e. price, aggregate rating and vote. From the analysis there is a correlation between variables which means a liner relationship, but it is week correlation which is best for correlation. All variables tend to go up in response to one another, the relationship is not very strong. Between price and rating, if price increases then there is a low tendency for the rating to increase as the relationship is not strong. Similarly for price and votes. It can be said that more expensive restaurants tend to attract more reviews or votes, but the relation is weak if its strong then it will be beneficial for the restaurant. The same goes to rating and votes.

**Competitive Analysis**

*Identify major competitors in each region based on cuisine, pricing, and ratings.*

*Major Competitors Based on Cuisine*

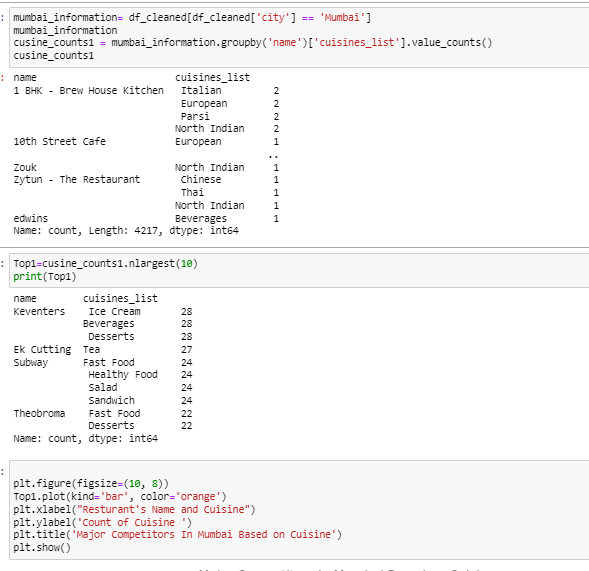
A screenshot of a computer

Description automatically generated

A graph of a number of blue bars

Description automatically generated with medium confidence

The major competitors across the India with cuisine are Domino’s Piza (Piza and Fast food), and Café Coffee Day stands 2nd which cuisine sandwich, Fast Food, Café, and Beverages.



A graph of a number of different types of food

Description automatically generated

The major competitors in Mumbai are Keventers (cuisine = Ice cream, Beverages, Deserts) followed by Ek Cutting, Subway. It shows people in Mumbai consume more desserts and then anything.

A screenshot of a computer program

Description automatically generated

A bar graph with blue lines

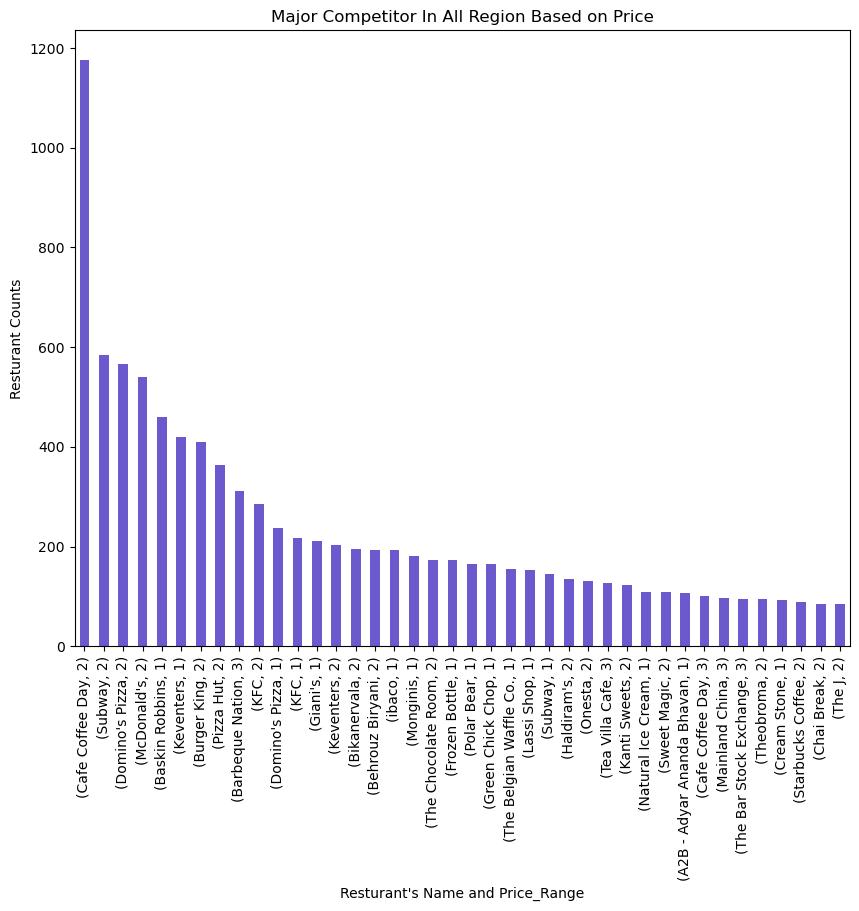
Description automatically generated

The most number of cusinies are offered by Café Coffee day which means they have a lot of variety for its customers.

***Major Competitors on Prices:***

A screenshot of a computer

Description automatically generated



Major Competitor based on price are Café Coffee day with price 2, followed by subway etc. Café Coffee Day also offered a number of cuisines too so customer may attract towards it too.

A screenshot of a computer program

Description automatically generated

A graph of a bar chart

Description automatically generated

Major Competitor based on price in particularly Mumbai is Subway and they sell out their products at price 1.

***Major Competitors Based on Rating***

A screenshot of a computer code

Description automatically generated

A graph of a bar chart

Description automatically generated

McDonald’s and Dimino’s Piza customer provide more rating as compared to others which means they are engaged with company.

A screenshot of a computer program

Description automatically generated

A graph of a number of blue bars

Description automatically generated

Particularly in Mumbai the most number of rating giving to Colaba Social followed then Ek cutting, and Candy and Green. It shows the engagement level of customers are high.

**Market Gap Analysis**

*Identify any gaps in the market that the restaurant chain can capitalize on (e.g., underrepresented cuisines, price ranges).*

A screenshot of a computer

Description automatically generated

A graph with numbers and text

Description automatically generated

The Street Food, biryani, mithai are the less popular cusinins in india in Zomato Restaurant.

A screenshot of a computer code

Description automatically generated

A graph of a number of different types of food

Description automatically generated

It shows that underrepresented cuisine with price are beverages, continental, mithai, and so on. The reason might be any that people buy less drinks while eating. Also, Mithai or Desserts only served at the end and people consumed it less.

**Designing the Marketing Campaign**

From the above analysis, it is suggested that marketing campaign should capitalize on the discovered correlation between price range, restaurants ratings and voting. They have weak positive correlation which means customers are slightly influenced by price and restaurant. However, it is not an ideal parameter which can be used for them to attract.

**Branding Strategy:** This strategy will emphasis the quality of the food so those restaurants where customers don’t prefer to go, they will attract and enhance their experience. Not only this, but good quality food also become the reason of word-of-mouth advertisement and people attract more.

Their rating will increase credibility and more customers will be attracted. Also introduce loyalty cards or discount cards in those restaurants where customer engagement is less. This thing will attract them.

**Cost Sensitive Segments:** offer foods at low prices too at special hours without compromising quality.

**Regional Targeting:** Based on the city, the marketing team should plan tailor promotions which promote their preferences and cuisines with local tastes. Host local event to boost community engagement where they also taste the food and will visit the restaurant again.

**Promotional Tactics:** Use social media platforms to target the targeted audience where popularity (votes) is linked to higher ratings. Use direct marketing through TV channels to create awareness about new products and do celebrity endorsements.