```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Reading csv file

```
In [2]: def read_data_from_csv():
    hotels=pd.read_csv(r"C:\Users\Ramakrishna\Desktop\zomato.csv")
    return hotels
    read_data_from_csv()
```

Out[2]:

	address	dress name online_order book_table rate votes		votes	phone			
0	942, 21st Main Road, 2nd Stage, Banashankari, 	Jalsa	Yes	Yes	4.1/5	775	080 42297555\r\n+91 9743772233	Bar
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th	Spice Elephant	Yes	No	4.1/5	787	080 41714161	Bar
2	1112, Next to KIMS Medical College, 17th Cross	San Churro Cafe	Yes	No	3.8/5	918	+91 9663487993	Bar
3	1st Floor, Annakuteera, 3rd Stage, Banashankar	Addhuri Udupi Bhojana	No	No	3.7/5	88	+91 9620009302	Bar
4	10, 3rd Floor, Lakshmi Associates, Gandhi Baza	Grand Village	No	No	3.8/5	166	+91 8026612447\r\n+91 9901210005	Bas
56247	Four Points by Sheraton Bengaluru, 43/3, White	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6 /5	27	080 40301477	
56248	Number 10, Garudachar Palya, Mahadevapura, Whi	Vinod Bar And Restaurant	No	No	NaN	0	+91 8197675843	
56249	Sheraton Grand Bengaluru Whitefield Hotel & Co	Plunge - Sheraton Grand Bengaluru Whitefield H	No	No	NaN	0	NaN	
56250	Sheraton Grand Bengaluru Whitefield Hotel & Co	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3 /5	236	080 49652769	
56251	ITPL Main Road, KIADB Export Promotion Industr	The Nest - The Den Bengaluru	No	No	3.4 /5	13	+91 8071117272	

```
def read data from csv():
   hotels=pd.read csv('zomato.csv')
    return hotels
def remove_unwanted_columns():
   #DO NOT REMOVE FOLLOWING LINE
   #call read_data_from_csv() function to get dataframe
   hotels=read_data_from_csv()
    return hotels
def rename_columns():
   #DO NOT REMOVE FOLLOWING LINE
   #call remove_unwanted_columns() function to get dataframe
   hotels = remove_unwanted_columns()
   #task2: rename columns, only these columns are allowed in the dataset
   # 1.
           Ιd
   # 2.
           Name
   # 3.
           online_order
   # 4.
           book table
   # 5.
           rating
           votes
   # 6.
           location
   # 7.
   # 8.
           rest type
   # 9.
           dish_liked
   # 10. cuisines
   # 11. approx cost
   # 12.
           type
    return hotels
#task3: handle null values of each column
def null_value_check():
   #DO NOT REMOVE FOLLOWING LINE
   #call rename columns() function to get dataframe
   hotels=rename_columns()
   #deleting null values of name column
   #handling null values of online order
   #handling null values of book_table
   #handling null values of rating
   #handling null values of votes
   #handling null values of location
   #handling null values of rest_type
```

```
#handling null values of dishliked
   #handling null values of cuisines
   #handling null values of approxcost
   #handling null values of type
    return hotels
#task4 #find duplicates in the dataset
def find duplicates():
   #DO NOT REMOVE FOLLOWING LINE
   #call null_value_check() function to get dataframe
   hotels=null value check()
   #droping the duplicates value keeping the first
    return hotels
#task5 removing irrelevant text from all the columns
def removing irrelevant text():
   #DO NOT REMOVE FOLLOWING LINE
   #call find_duplicates() function to get dataframe
   hotels= find duplicates()
    return hotels
#task6: check for unique values in each column and handle the irrelevant
values
def check_for_unique_values():
   #DO NOT REMOVE FOLLOWING LINE
   #call removing irrelevant text() function to get dataframe
   hotels=removing irrelevant text()
    return hotels
#task7: remove the unknown character from the dataset and export it to
"zomatocleaned.csv"
def remove the unknown character():
   #DO NOT REMOVE FOLLOWING LINE
   #call check_for_unique_values() function to get dataframe
   dataframe=check for unique values()
   #remove unknown character from dataset
   #export cleaned Dataset to newcsv file named "zomatocleaned.csv"
   dataframe.to_csv('zomatocleaned.csv')
    return dataframe
#check if mysql table is created using "zomatocleaned.csv"
```

```
#Use this final dataset and upload it on the provided database for performing
analysis in MySQL
#To Run this task first Run the appliation for Terminal to create table named
'Zomato' and then run test.
def start():
    remove_the_unknown_character()

def task_runner():
    start()
```

Delete Unwanted data

```
In [3]: def remove_unwanted_columns():
    #DO NOT REMOVE FOLLOWING LINE
    #call read_data_from_csv() function to get dataframe
    hotels=read_data_from_csv()
    unwanted_columns= hotels.drop(columns=["address","phone"],inplace=True)
    return hotels
```

In [4]: remove_unwanted_columns()

Out[4]:

	name	online_order	book_table	rate	votes	location	rest_type	dish_liked	
0	Jalsa	Yes	Yes	4.1/5	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja	
1	Spice Elephant	Yes	No	4.1/5	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G	
2	San Churro Cafe	Yes	No	3.8/5	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc	
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	Banashankari	Quick Bites	Masala Dosa	
4	Grand Village	No	No	3.8/5	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe	F
	•••							•••	
56247	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6 /5	27	Whitefield	Bar	NaN	С
56248	Vinod Bar And Restaurant	No	No	NaN	0	Whitefield	Bar	NaN	
56249	Plunge - Sheraton Grand Bengaluru Whitefield H	No	No	NaN	0	Whitefield	Bar	NaN	
56250	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3 /5	236	ITPL Main Road, Whitefield	Bar	Cocktails, Pizza, Buttermilk	
56251	The Nest - The Den Bengaluru	No	No	3.4 /5	13	ITPL Main Road, Whitefield	Bar, Casual Dining	NaN	С

```
In [5]: remove_unwanted_columns().columns
Out[5]: Index(['name', 'online_order', 'book_table', 'rate', 'votes', 'location',
                'rest_type', 'dish_liked', 'cuisines', 'approx_cost(for two people)',
                'listed_in(type)'],
              dtype='object')
In [6]: def rename_columns():
            #DO NOT REMOVE FOLLOWING LINE
            #call remove_unwanted_columns() function to get dataframe
            hotels = remove_unwanted_columns()
            column_names={
                 "name": "name",
                 "online_order":"online_order",
                 'book_table':'book_table',
                 "rate": 'rating',
                 'votes':'votes',
                 'location':'location',
                 'rest_type':'rest_type',
                 'dish_liked':'dish_liked',
                 'cuisines':'cuisines',
                 'approx_cost(for two people)':'approx_cost',
                 'listed in(type)':'type'
            hotels=hotels.rename(columns=column_names)
            return hotels
```

In [7]: rename_columns()

Out[7]:	n	ame online	o order	hook table	rating	votes	

	name	online_order	book_table	rating	votes	location	rest_type	dish_liked
0	Jalsa	Yes	Yes	4.1/5	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja
1	Spice Elephant	Yes	No	4.1/5	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G
2	San Churro Cafe	Yes	No	3.8/5	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	Banashankari	Quick Bites	Masala Dosa
4	Grand Village	No	No	3.8/5	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe
56247	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6 /5	27	Whitefield	Bar	NaN
56248	Vinod Bar And Restaurant	No	No	NaN	0	Whitefield	Bar	NaN
56249	Plunge - Sheraton Grand Bengaluru Whitefield H	No	No	NaN	0	Whitefield	Bar	NaN
56250	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3 /5	236	ITPL Main Road, Whitefield	Bar	Cocktails, Pizza, Buttermilk
56251	The Nest - The Den Bengaluru	No	No	3.4 /5	13	ITPL Main Road, Whitefield	Bar, Casual Dining	NaN

```
In [8]: def null value check():
            #DO NOT REMOVE FOLLOWING LINE
            #call rename_columns() function to get dataframe
            hotels=rename columns()
            #deleting null values of name column
            hotels=hotels.dropna(subset=["name"])
            #handling null values of online_order
            hotels["online order"]=hotels["online order"].fillna('NA')
            #handling null values of book table
            hotels['book_table']=hotels['book_table'].fillna('NA')
            #handling null values of rating
            hotels["rating"]=hotels["rating"].fillna(0)
            #handling null values of votes
            hotels["votes"]=hotels["votes"].fillna(0)
            #handling null values of location
            hotels["location"]=hotels["location"].fillna('NA')
            #handling null values of rest_type
            hotels["rest_type"]=hotels["rest_type"].fillna('NA')
            #handling null values of dishliked
            hotels["dish liked"]=hotels["dish liked"].fillna('NA')
            #handling null values of cuisines
            hotels["cuisines"]=hotels["cuisines"].fillna('NA')
            #handling null values of approxcost
            hotels["approx_cost"]=hotels["approx_cost"].fillna(0)
            #handling null values of type
            hotels["type"]=hotels["type"].fillna('NA')
            return hotels
```

In [9]: null_value_check()

Out[9]:	name		online_order	book_table	rating	votes	location	rest_type	dish_liked
	0	Jalsa	Yes	Yes	4.1/5	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja
	1	Spice Elephant	Yes	No	4.1/5	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G
	2	San Churro Cafe	Yes	No	3.8/5	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc
	3	Addhuri Udupi Bhojana	No	No	3.7/5	88	Banashankari	Quick Bites	Masala Dosa
	4	Grand Village	No	No	3.8/5	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe
				•••					
	56247	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6 /5	27	Whitefield	Bar	NA
	56248	Vinod Bar And Restaurant	No	No	0	0	Whitefield	Bar	NA
	56249	Plunge - Sheraton Grand Bengaluru Whitefield H	No	No	0	0	Whitefield	Bar	NA
	56250	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3 /5	236	ITPL Main Road, Whitefield	Bar	Cocktails, Pizza, Buttermilk
	56251	The Nest - The Den Bengaluru	No	No	3.4 /5	13	ITPL Main Road, Whitefield	Bar, Casual Dining	NA

In [11]: find_duplicates()

٦.	.4-1	T 1 1	լլ։	
JU	ıuı	1 1		

	name	online_order	book_table	rating	votes	location	rest_type	dish_liked
0	Jalsa	Yes	Yes	4.1/5	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja
1	Spice Elephant	Yes	No	4.1/5	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G
2	San Churro Cafe	Yes	No	3.8/5	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	Banashankari	Quick Bites	Masa l a Dosa
4	Grand Village	No	No	3.8/5	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe
56247	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6 /5	27	Whitefield	Bar	NA
56248	Vinod Bar And Restaurant	No	No	0	0	Whitefield	Bar	NA
56249	Plunge - Sheraton Grand Bengaluru Whitefield H	No	No	0	0	Whitefield	Bar	NA
56250	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3 /5	236	ITPL Main Road, Whitefield	Bar	Cocktails, Pizza, Buttermilk
56251	The Nest - The Den Bengaluru	No	No	3.4 /5	13	ITPL Main Road, Whitefield	Bar, Casual Dining	NA

38413 rows × 11 columns

4

```
In [12]: find duplicates().columns
Out[12]: Index(['name', 'online_order', 'book_table', 'rating', 'votes', 'location',
                 'rest_type', 'dish_liked', 'cuisines', 'approx_cost', 'type'],
               dtype='object')
In [13]: | def removing_irrelevant_text():
             #DO NOT REMOVE FOLLOWING LINE
             #call find duplicates() function to get dataframe
             hotels= find duplicates()
             hotels=hotels[hotels["name"].str.contains("RATED Rated")==False]
             hotels=hotels[hotels["online_order"].str.contains("RATED|Rated")==False]
             hotels=hotels[hotels["book_table"].str.contains("RATED Rated")==False]
             hotels=hotels[hotels["rating"].str.contains("RATED Rated")==False]
             hotels=hotels[hotels["votes"].str.contains("RATED Rated")==False]
             hotels=hotels[hotels["location"].str.contains("RATED|Rated")==False]
             hotels=hotels[hotels["rest_type"].str.contains("RATED|Rated")==False]
             hotels=hotels[hotels["dish_liked"].str.contains("RATED|Rated")==False]
             hotels=hotels[hotels["cuisines"].str.contains("RATED Rated")==False]
             hotels=hotels[hotels["approx_cost"].str.contains("RATED Rated")==False]
             hotels=hotels[hotels["type"].str.contains("RATED Rated")==False]
             return hotels
```

In [14]:	removing_irrelevant_text()								
Out[14]:		name	online_order	book_table	rating	votes	location	rest_type	dish_liked
	0	Jalsa	Yes	Yes	4.1/5	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja
	1	Spice Elephant	Yes	No	4.1/5	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G
	2	San Churro Cafe	Yes	No	3.8/5	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc
	3	Addhuri Udupi Bhojana	No	No	3.7/5	88	Banashankari	Quick Bites	Masala Dosa
	4	Grand Village	No	No	3.8/5	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe
	56244	The Farm House Bar n Grill	No	No	3.7 /5	34	Whitefield	Casual Dining, Bar	NA
	56246	Bhagini	No	No	2.5 /5	81	Whitefield	Casual Dining, Bar	Biryani, Andhra Meal
	56247	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6 /5	27	Whitefield	Bar	NA
	56250	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3 /5	236	ITPL Main Road, Whitefield	Bar	Cocktails, Pizza, Buttermilk

30340 rows × 11 columns

The Nest -

Bengaluru

The Den

No

No 3.4 /5

13

localhost:8888/notebooks/Desktop/Zomato Analysis/Zomato Data Analysis.ipynb#

56251

NA

ITPL Main

Whitefield

Road,

Bar,

Casual

Dining

```
In [17]: def check_for_unique_values():
    #DO NOT REMOVE FOLLOWING LINE
    #call removing_irrelevant_text() function to get dataframe
    hotels=removing_irrelevant_text()
    index=hotels[~hotels["online_order"].isin(['Yes','No'])].index
    hotels.drop(index,inplace=True)
    hotels["rating"]= hotels["rating"].replace({'/5':''},regex=True).replace('Interval to the content of the cont
```

In [18]: check_for_unique_values()

[
Out[18]:		name	online_order	book_table	rating	votes	location	rest_type	dish_liked
	0	Jalsa	Yes	Yes	4.1	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja
	1	Spice Elephant	Yes	No	4.1	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G
	2	San Churro Cafe	Yes	No	3.8	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc
	3	Addhuri Udupi Bhojana	No	No	3.7	88	Banashankari	Quick Bites	Masala Dosa
	4	Grand Village	No	No	3.8	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe
	56244	The Farm House Bar n Grill	No	No	3.7	34	Whitefield	Casual Dining, Bar	NA
	56246	Bhagini	No	No	2.5	81	Whitefield	Casual Dining, Bar	Biryani, Andhra Meal
	56247	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6	27	Whitefield	Bar	NA
	56250	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3	236	ITPL Main Road, Whitefield	Bar	Cocktails, Pizza, Buttermilk
	56251	The Nest - The Den Bengaluru	No	No	3.4	13	ITPL Main Road, Whitefield	Bar, Casual Dining	NA

30336 rows × 11 columns

,

```
def read_data_from_csv():
   hotels=pd.read_csv('zomato.csv')
    return hotels
def remove_unwanted_columns():
   #DO NOT REMOVE FOLLOWING LINE
   #call read_data_from_csv() function to get dataframe
   hotels=read_data_from_csv()
    return hotels
def rename_columns():
   #DO NOT REMOVE FOLLOWING LINE
   #call remove_unwanted_columns() function to get dataframe
   hotels = remove_unwanted_columns()
   #task2: rename columns, only these columns are allowed in the dataset
   # 1.
   # 2.
           Name
           online_order
   # 3.
   # 4.
           book table
   # 5.
           rating
           votes
   # 6.
   # 7.
           location
          rest type
   # 8.
   # 9.
           dish_liked
   # 10. cuisines
   # 11. approx_cost
   # 12. type
    return hotels
#task3: handle null values of each column
def null value check():
   #DO NOT REMOVE FOLLOWING LINE
   #call rename columns() function to get dataframe
   hotels=rename columns()
   #deleting null values of name column
   #handling null values of online_order
   #handling null values of book table
   #handling null values of rating
   #handling null values of votes
   #handling null values of location
   #handling null values of rest_type
   #handling null values of dishliked
```

```
#handling null values of cuisines
   #handling null values of approxcost
   #handling null values of type
    return hotels
#task4 #find duplicates in the dataset
def find duplicates():
   #DO NOT REMOVE FOLLOWING LINE
   #call null value check() function to get dataframe
   hotels=null_value_check()
   #droping the duplicates value keeping the first
    return hotels
#task5 removing irrelevant text from all the columns
def removing_irrelevant_text():
   #DO NOT REMOVE FOLLOWING LINE
   #call find duplicates() function to get dataframe
   hotels= find duplicates()
    return hotels
#task6: check for unique values in each column and handle the irrelevant
values
def check_for_unique_values():
   #DO NOT REMOVE FOLLOWING LINE
   #call removing_irrelevant_text() function to get dataframe
   hotels=removing irrelevant text()
    return hotels
#task7: remove the unknown character from the dataset and export it to
"zomatocleaned.csv"
def remove the_unknown_character():
   #DO NOT REMOVE FOLLOWING LINE
   #call check_for_unique_values() function to get dataframe
   dataframe=check for unique values()
   #remove unknown character from dataset
   #export cleaned Dataset to newcsv file named "zomatocleaned.csv"
   dataframe.to csv('zomatocleaned.csv')
    return dataframe
#check if mysql table is created using "zomatocleaned.csv"
#Use this final dataset and upload it on the provided database for performing
analysis in MySQL
```

```
#To Run this task first Run the appliation for Terminal to create table named
          'Zomato' and then run test.
         def start():
             remove_the_unknown_character()
         def task_runner():
             start()
In [93]: | check_for_unique_values().columns
Out[93]: Index(['name', 'online_order', 'book_table', 'rating', 'votes', 'location',
                 'rest_type', 'dish_liked', 'cuisines', 'approx_cost', 'type'],
               dtype='object')
In [94]: check_for_unique_values().columns
Out[94]: Index(['name', 'online_order', 'book_table', 'rating', 'votes', 'location',
                 'rest_type', 'dish_liked', 'cuisines', 'approx_cost', 'type'],
               dtype='object')
In [21]: def remove the unknown character():
             #DO NOT REMOVE FOLLOWING LINE
             #call check_for_unique_values() function to get dataframe
             dataframe=check_for_unique_values()
             #remove unknown character from dataset
             dataframe['name']=dataframe['name'].str.replace('[Ãx][^A-Za-z]+','')
               dataframe['approx cost']=dataframe['approx cost'].str.replace(",
               dataframe['approx cost']=dataframe['approx cost'].str.strip('"')
             #export cleaned Dataset to newcsv file named "zomatocleaned.csv"
             dataframe.to csv('zomatocleaned.csv')
             return dataframe
```

In [22]: remove_the_unknown_character()

C:\Users\Ramakrishna\AppData\Local\Temp\ipykernel_15644\929471048.py:6: Futur eWarning: The default value of regex will change from True to False in a futu re version.

dataframe['name']=dataframe['name'].str.replace('[Ãx][^A-Za-z]+','')

Out[22]:

	name	online_order	book_table	rating	votes	location	rest_type	dish_liked
0	Jalsa	Yes	Yes	4.1	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja
1	Spice Elephant	Yes	No	4.1	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G
2	San Churro Cafe	Yes	No	3.8	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc
3	Addhuri Udupi Bhojana	No	No	3.7	88	Banashankari	Quick Bites	Masala Dosa
4	Grand Village	No	No	3.8	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe
56244	The Farm House Bar n Grill	No	No	3.7	34	Whitefield	Casual Dining, Bar	NA
56246	Bhagini	No	No	2.5	81	Whitefield	Casual Dining, Bar	Biryani, Andhra Meal
56247	Best Brews - Four Points by Sheraton Bengaluru	No	No	3.6	27	Whitefield	Bar	NA
56250	Chime - Sheraton Grand Bengaluru Whitefield Ho	No	Yes	4.3	236	ITPL Main Road, Whitefield	Bar	Cocktails, Pizza, Buttermilk
56251	The Nest - The Den Bengaluru	No	No	3.4	13	ITPL Main Road, Whitefield	Bar, Casual Dining	NA

In []: