

ZOMATO'S GLOBAL RECIPE

A Data Understanding Journey



THE JOURNEY BEGINS

- In this project, I embarked on a data exploration journey to lay the groundwork for understanding Zomato's global presence.
- Unlike typical data analysis projects, I prioritized building a strong foundation by focusing on the essentials: data cleaning and understanding.
- Through meticulous coding and attention to detail, I refined Zomato's dataset to unlock its secrets. I ensured data consistency by standardizing column names and transforming them into a readable format.
- This project showcases the importance of data preparation and understanding in uncovering meaningful insights. By mastering the basics, we can unlock a deeper understanding of Zomato's customer preferences, restaurant trends, and market dynamics.

THE RECIPE FOR INSIGHTS: DATA EXPLORATION STEPS

Data Ingestion

Importing the Zomato dataset, unlocking its potential

Data Cleaning

Standardizing column names, replacing spaces with underscores for consistency

Data Understanding

Exploring data structure, content, and quality to unlock insights

IMPORT NECESSARY LIBRARIES FOR DATA ANALYSIS AND VISUALIZATION

```
# pandas for data manipulation and analysis
```

```
import pandas as pd
```

```
# numpy for numerical computations and data manipulation
```

```
import numpy as np
```

```
# matplotlib for creating static, animated, and interactive visualizations
```

```
import matplotlib.pyplot as plt
```

```
# seaborn for creating informative and attractive statistical graphics
```

```
import seaborn as sns
```

LOAD THE ZOMATO DATASET FROM A CSV FILE

```
# File path: C:\Users\HP\Desktop\zomato.csv  
# Encoding: latin-1 (to handle special characters and foreign languages)
```

```
df = pd.read_csv(r"C:\Users\HP\Desktop\zomato.csv", encoding="latin-1")
```

CREATING A DATA BACKUP: ENSURING SAFETY

```
df_info = df.copy()
```


DISPLAY THE TOP 5 ROWS TO UNDERSTAND DATA STRUCTURE AND CONTENT

```
df_info.head()
```

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	Average Cost for two	Currency	Has Table booking	Has Online delivery
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443	French, Japanese, Desserts	1100	Botswana Pula(P)	Yes	No
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708	Japanese	1200	Botswana Pula(P)	Yes	No
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	121.056831	14.581404	Seafood, Asian, Filipino, Indian	4000	Botswana Pula(P)	Yes	No
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.056475	14.585318	Japanese, Sushi	1500	Botswana Pula(P)	No	No
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.057508	14.584450	Japanese, Korean	1500	Botswana Pula(P)	Yes	No

CONFIGURE PANDAS TO DISPLAY ALL COLUMNS WHEN VIEWING DATAFRAMES

```
pd.set_option("display.max_columns", 999)
```

EXPLORE THE ZOMATO DATASET: COLUMN NAMES

```
df_info.columns
```

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address',  
      'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',  
      'Average Cost for two', 'Currency', 'Has Table booking',  
      'Has Online delivery', 'Is delivering now', 'Switch to order menu',  
      'Price range', 'Aggregate rating', 'Rating color', 'Rating text',  
      'Votes'],  
      dtype='object')
```

STANDARDIZE COLUMN NAMES: CONVERT TO LOWERCASE AND REPLACE SPACES WITH UNDERSCORES

```
df_info.columns = df_info.columns.str.lower()
```

```
df_info.columns = df_info.columns.str.replace(" ", "_")
```

```
df_info.columns
```

```
Index(['restaurant_id', 'restaurant_name', 'country_code', 'city', 'address',  
      'locality', 'locality_verbose', 'longitude', 'latitude', 'cuisines',  
      'average_cost_for_two', 'currency', 'has_table_booking',  
      'has_online_delivery', 'is_delivering_now', 'switch_to_order_menu',  
      'price_range', 'aggregate_rating', 'rating_color', 'rating_text',  
      'votes'],  
      dtype='object')
```


DISPLAY INFORMATION ABOUT DATA TYPES, MISSING VALUES, AND MEMORY USAGE

```
df.info.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   restaurant_id         9551 non-null   int64
1   restaurant_name       9551 non-null   object
2   country_code          9551 non-null   int64
3   city                  9551 non-null   object
4   address               9551 non-null   object
5   locality              9551 non-null   object
6   locality_verbose      9551 non-null   object
7   longitude             9551 non-null   float64
8   latitude              9551 non-null   float64
9   cuisines              9542 non-null   object
10  average_cost_for_two  9551 non-null   int64
11  currency              9551 non-null   object
12  has_table_booking     9551 non-null   object
13  has_online_delivery   9551 non-null   object
14  is_delivering_now     9551 non-null   object
15  switch_to_order_menu  9551 non-null   object
16  price_range           9551 non-null   int64
17  aggregate_rating      9551 non-null   float64
18  rating_color          9551 non-null   object
19  rating_text           9551 non-null   object
20  votes                 9551 non-null   int64
dtypes: float64(3), int64(5), object(13)
memory usage: 1.5+ MB
```



SUMMARY STATISTICS OF NUMERICAL COLUMNS (INT AND FLOAT)

df.info.describe()

	restaurant_id	country_code	longitude	latitude	average_cost_for_two	price_range	aggregate_rating	votes
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763	1.804837	2.666370	156.909748
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073	0.905609	1.516378	430.169145
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000	1.000000	0.000000	0.000000
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000	1.000000	2.500000	5.000000
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000	2.000000	3.200000	31.000000
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000	2.000000	3.700000	131.000000
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000	4.000000	4.900000	10934.000000

COUNT THE NUMBER OF NULL VALUES IN EACH COLUMN

```
df_info.isnull().sum()
```

```
restaurant_id      0
restaurant_name    0
country_code       0
city               0
address            0
locality           0
locality_verbose   0
longitude          0
latitude           0
cuisines            9
average_cost_for_two 0
currency           0
has_table_booking  0
has_online_delivery 0
is_delivering_now  0
switch_to_order_menu 0
price_range        0
aggregate_rating    0
rating_color        0
rating_text         0
votes              0
dtype: int64
```


COUNT THE NUMBER OF DUPLICATE ROWS

```
df_info.duplicated().sum()
```

0

COMBINE THE TWO TABLES BASED ON "COUNTRY_CODE" COLUMN USING AN "INNER JOIN"

```
df_final = pd.merge(df_info, df_country, on = "country_code", how = "inner")
```

```
df_final.head()
```

currency	has_table_booking	has_online_delivery	is_delivering_now	switch_to_order_menu	price_range	aggregate_rating	rating_color	rating_text	votes	country
Botswana Pula(P)	Yes	No	No	No	3	4.8	Dark Green	Excellent	314	Phillipines
Botswana Pula(P)	Yes	No	No	No	3	4.5	Dark Green	Excellent	591	Phillipines
Botswana Pula(P)	Yes	No	No	No	4	4.4	Green	Very Good	270	Phillipines

CONCLUSION: A SOLID FOUNDATION FOR INSIGHTS

KEY TAKEAWAYS:

- **Standardized Data:** We refined the dataset by transforming column names to lowercase and replacing spaces with underscores, ensuring consistency and readability.
- **Data Clarity:** Through meticulous data cleaning, we gained a deeper understanding of the dataset's structure and content, paving the way for future insights.

IMPACT:

- By prioritizing data cleaning and understanding, we:
- Ensured data quality and consistency
- Unlocked the potential for meaningful insights
- Set the stage for further analysis and discovery

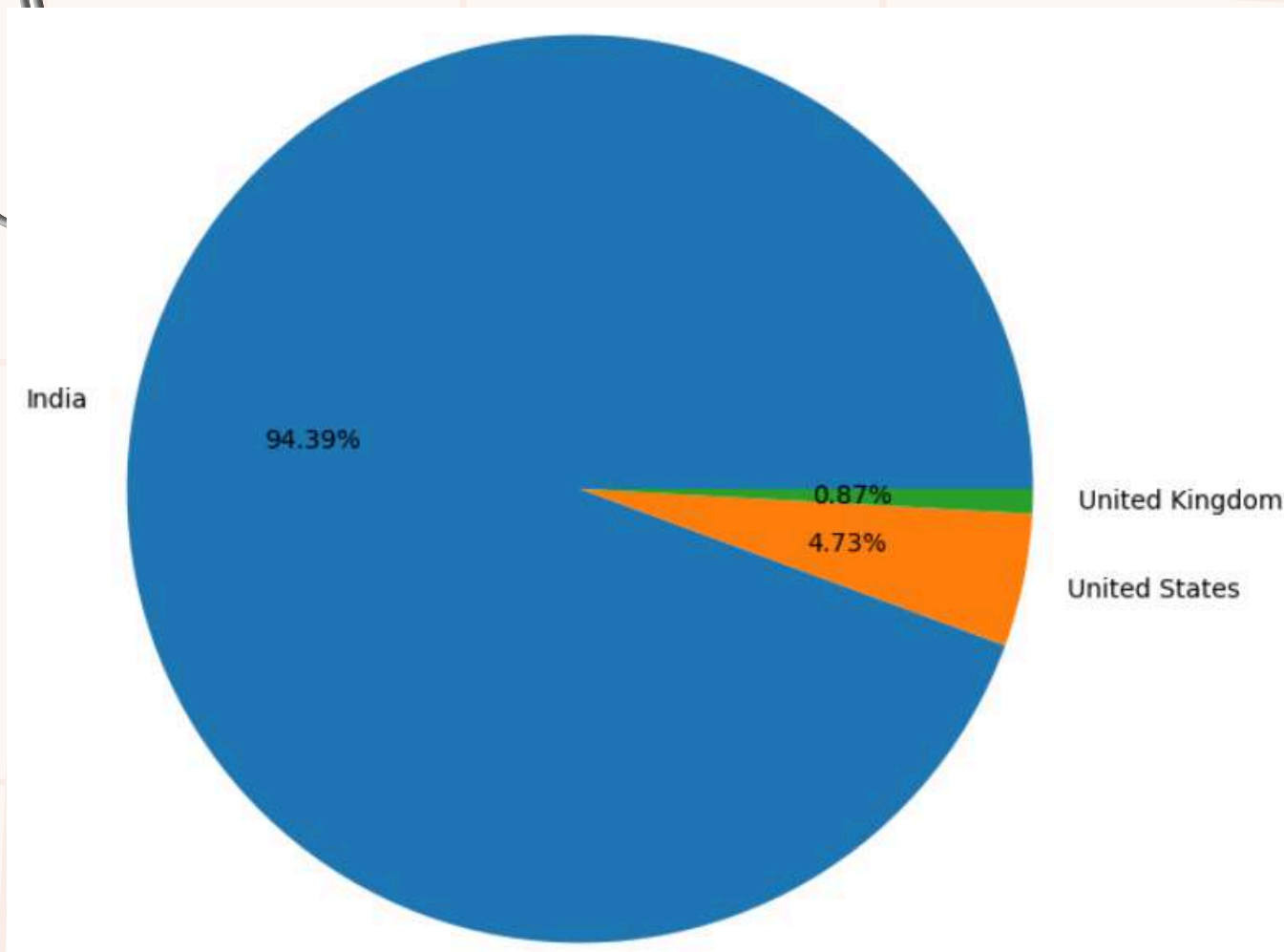
NEXT STEPS:

- With a solid foundation in place, we can now:
- Explore customer preferences and restaurant trends
- Analyze market dynamics and uncover hidden patterns
- Drive business decisions with data-driven insights

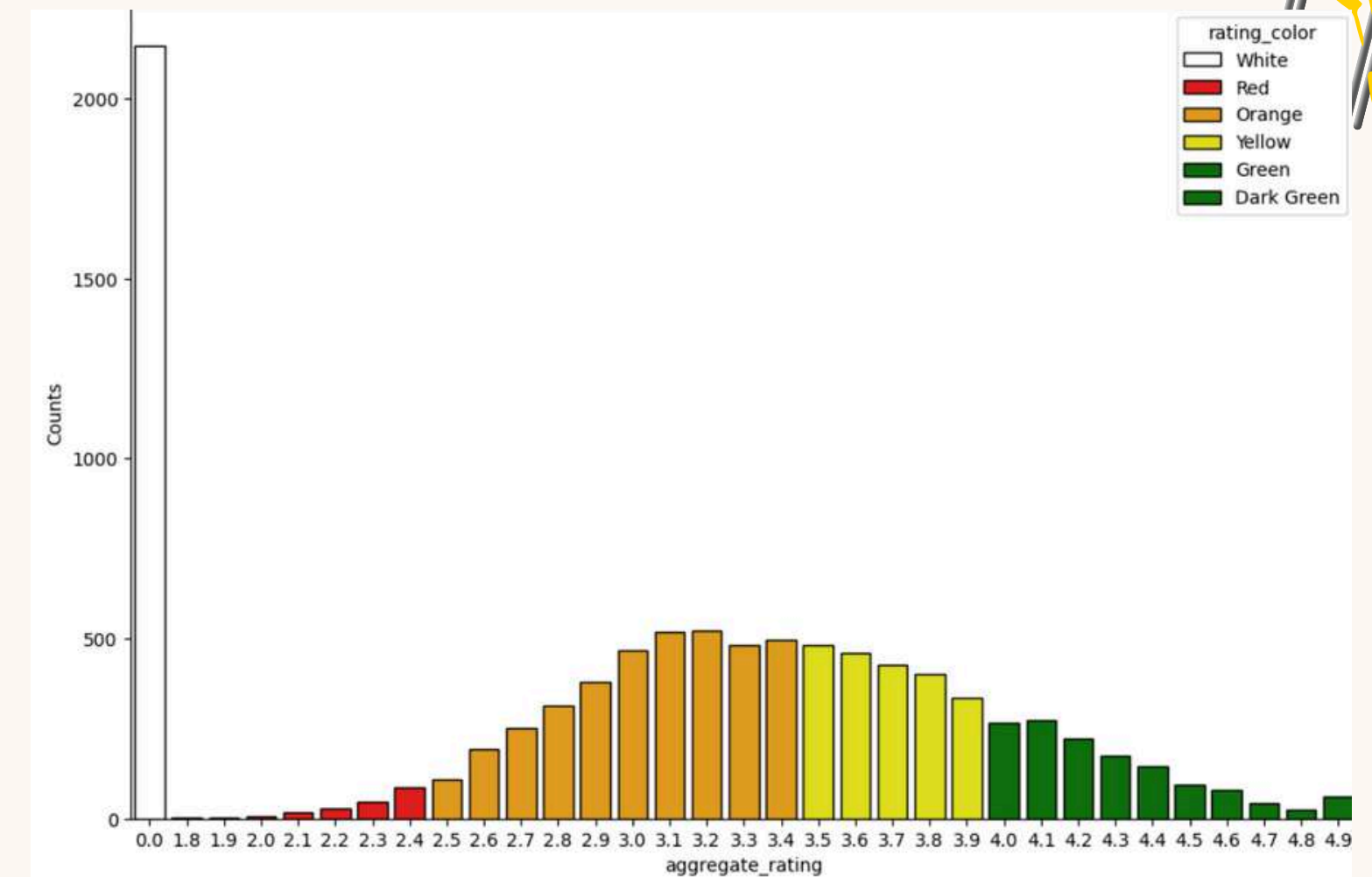
FOOD FOR THOUGHT

A TASTE OF WHAT'S TO COME...

**GLOBAL FLAVORS:
COUNTRY-WISE ORDER DISTRIBUTION**



**RATING RAINBOW:
AGGREGATE RATINGS & COLORS**



BON APPÉTIT!

Thank You for Savoring the Flavor of Data!

We hope you enjoyed this culinary journey through Zomato's data as much as we did!

The feast doesn't end here! Stay tuned for our next project, where we'll:

Dive Deeper into the world of Zomato's data, uncovering hidden gems and surprising insights!
Ignite Innovation, pushing the boundaries of what's possible with data-driven discovery!
Serve Up the Future of food exploration, where technology meets taste buds!

