

Zomato

Zomato Data Analysis Project

Analyzing Restaurant Data Using Python

Introduction

Objective:

To analyze Zomato restaurant data to derive meaningful insights regarding restaurant types, ratings, customer preferences, and order methods.

Tools Used:

- Jupyter Notebook
- Python Libraries: Pandas, NumPy, Matplotlib, Seaborn

Libraries Installation

Installing Libraries:

1. Seaborn Installation:

```
```pip install seaborn```
```

2. Importing Libraries:

```
```python
```

```
import pandas as pd
```

```
import numpy as np
```

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

```
import seaborn as sns
```

```
```
```

# Loading the Data

Creating the DataFrame:

```
```python  
dataframe = pd.read_csv('Zomato data .csv')  
print(dataframe.head())  
```
```

Conclusion:

- Data successfully loaded and displayed.

# Data Cleaning

Handling the 'Rate' Column:

```
```python
def handleRate(value):
    value=str(value).split('/')
    value=value[0]
    return float(value)

dataframe['rate'] = dataframe['rate'].apply(handleRate)

print(dataframe.head())
```
```

Summary of the DataFrame:

```
```python
dataframe.info()
```
```

Conclusion:

- No NULL values present.

# Visualization - Restaurant Types

Count of Restaurant Types:

```
```python
sns.countplot(x=dataframe['listed_in(type)'])
plt.xlabel('Type of restaurant')
plt.show()
```
```

Conclusion:

- Majority of restaurants fall into the dining category.

# Visualization - Votes by Restaurant Type

Total Votes by Restaurant Type:

```
```python
grouped_data =
dataframe.groupby('listed_in(type)')['votes'].sum()
result = pd.DataFrame({'votes': grouped_data})
plt.plot(result, c='green', marker='o')
plt.xlabel('Type of restaurant', c='red', size=20)
plt.ylabel('votes', c='red', size=20)
plt.show()
```
```

Conclusion:

- Dining restaurants receive a larger number of votes.

# Visualization - Ratings Distribution

Histogram of Ratings:

```
```python  
plt.hist(dataframe['rate'], bins=5)  
plt.title('Ratings Distribution')  
plt.show()  
```
```

Conclusion:

- Majority of restaurants have ratings between 3.5 and 4.



# Visualization - Approximate Cost

Distribution of Approximate Cost for Two People:

```
```python
couple_data = dataframe['approx_cost(for two
people)']
sns.countplot(x=couple_data)
plt.show()
```
```

Conclusion:

- Most couples prefer restaurants with an approximate cost of 300 rupees.

# Visualization - Online vs Offline Orders

Ratings by Order Type:

```
```python
plt.figure(figsize = (6,6))

sns.boxplot(x = 'online_order', y = 'rate', data = dataframe)

plt.show()
```
```

Conclusion:

- Online orders generally receive higher ratings compared to offline orders.

Heatmap of Order Types:

```
```python

pivot_table = dataframe.pivot_table(index='listed_in(type)', columns='online_order', aggfunc='size', fill_value=0)

sns.heatmap(pivot_table, annot=True, cmap='YlGnBu', fmt='d')

plt.title('Heatmap')

plt.xlabel('Online Order')

plt.ylabel('Listed In (Type)')

plt.show()
```
```

Conclusion:

- Dining restaurants accept more offline orders, while cafes receive more online orders.

# Conclusion

## Key Insights:

- Dining restaurants are preferred by a larger number of individuals.
- Majority of restaurants have ratings between 3.5 and 4.
- Couples prefer restaurants with an approximate cost of 300 rupees.
- Online orders receive higher ratings compared to offline orders.
- Dining restaurants accept more offline orders, while cafes receive more online orders.

# Questions

Q&A Session:

Feel free to ask any questions related to the Zomato Data Analysis Project.

# References

Resources:

- Pandas Documentation
- Seaborn Documentation
- Jupyter Notebook

Thank You