

Zomato Pune Restaurant Analysis

Project by: Divyansh Deora

Overview :

This project analyzes restaurant trends in Pune using Zomato data. Using Python and Power BI, I processed over 12,000+ entries to uncover cuisine preferences, top localities, and spending patterns. This end-to-end data analytics pipeline transforms messy data into valuable business insights.

 **Dataset:** Zomato Pune (12,189 records)

 **Features:** Ratings, Votes, Charges, Cuisine, Locality

Objectives

1. Identify top-rated cuisines and popular localities
2. Understand cost vs rating patterns (affordability analysis)
3. Analyze cuisine popularity and restaurant density
4. Deliver visual insights for customer and business decisions

Table of Contents

1. Project Overview
2. Import Libraries
3. Load Dataset
4. Data Cleaning & Transformation
5. Feature Engineering
6. Top Localities
7. Data Visualization
8. Conclusion

What I Learned

Handling real-world messy data

Cleaning multi-value columns (like cuisines)

Visualizing actionable trends

End-to-end storytelling through data

1. Import Libraries

```
#importing libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from collections import Counter
```

2. Load Dataset

```
df = pd.read_csv('/content/zomato_pune_V002.csv')
df.head()
```



	Restaurant_Name	Web_Link	Locality	Sponsored	Ratings_out
0	AB's - Absolute Barbecues	https://www.zomato.com/pune/abs-absolute-barbe...	Hinjawadi	Casual Dining	
1	Cafe Co2 Resto Lounge	https://www.zomato.com/pune/cafe-co2-resto-lou...	Bhugaon	Lounge, Casual Dining	
2	Paasha - JW Marriott Pune	https://www.zomato.com/pune/paasha-jw-marriott...	Senapati Bapat Road	Fine Dining	
3	I Amsterdam	https://www.zomato.com/pune/i-amsterdam-hinjawadi	Hinjawadi	Casual Dining, Bar	
4	FC Road Social	https://www.zomato.com/pune/fc-road-social-shi...	Shivaji Nagar	Bar, Casual Dining	

5 rows × 104 columns

✓ 3. Data Cleaning

```
# Clean Charges_for_two
df['Charges_for_two'] = (
    df['Charges_for_two']
    .replace('Not Present', np.nan)
    .str.replace('₹', '', regex=False)
    .str.replace(',', '', regex=False)
)
df['Charges_for_two'] = pd.to_numeric(df['Charges_for_two'], errors='coerce')
```

```
# Clean "Number of votes"
df['Number of votes'] = (
    df['Number of votes']
    .astype(str)
    .str.replace('votes', '', regex=False)
    .str.replace(',', '', regex=False)
    .str.strip()
)
df['Number of votes'] = pd.to_numeric(df['Number of votes'], errors='coerce')
```

```
# Clean Ratings column
df['Ratings_out_of_5'] = pd.to_numeric(df['Ratings_out_of_5'], errors='coerce')
```

✓ 4. Feature Engineering

```
# Clean Cuisine list
df['Cuisine_List'] = df['Cuisines'].astype(str).str.split(',')
df['Cuisine_List'] = df['Cuisine_List'].apply(lambda x: [c.strip() for c in x])
```

```
# Clean Locality
df['Locality'] = df['Locality'].astype(str).str.strip()
```

```
# Final selected data
df_cleaned = df[['Restaurant_Name', 'Locality', 'Ratings_out_of_5', 'Number of votes', 'C
df_cleaned.head()
```



	Restaurant_Name	Locality	Ratings_out_of_5	Number of votes	Charges_for_two	Cuisine_List
0	AB's - Absolute Barbecues	Hinjawadi	4.9	7029.0	1400.0	[Continental, North Indian, Chinese]
1	Cafe Co2 Resto	Rhugdaon	4.6	2578.0	1500.0	[North Indian, Chinese,

Next
steps:

[Generate code with df_cleaned](#)[View recommended plots](#)[New interactive sheet](#)

5. Top Localities

```
df['Locality'] = df['Locality'].astype(str).str.strip()

# View top 10 localities
df['Locality'].value_counts().head(10)
```



	count
Locality	
Hadapsar	575
Kothrud	560
Wakad	523
Hinjawadi	481
Sinhgad Road	458
Viman Nagar	457
Kharadi	454
Baner	384
Chinchwad	314
Wagholi	305

dtype: int64



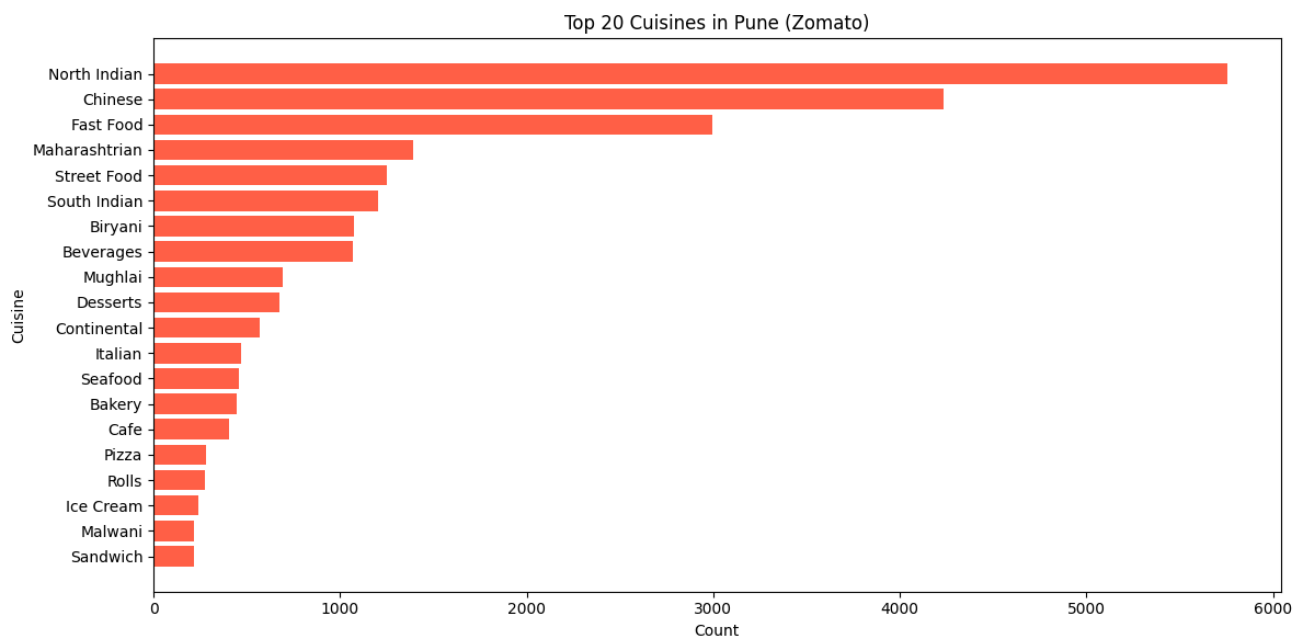
6. Data Visualizations



Top 20 Cuisines in Pune

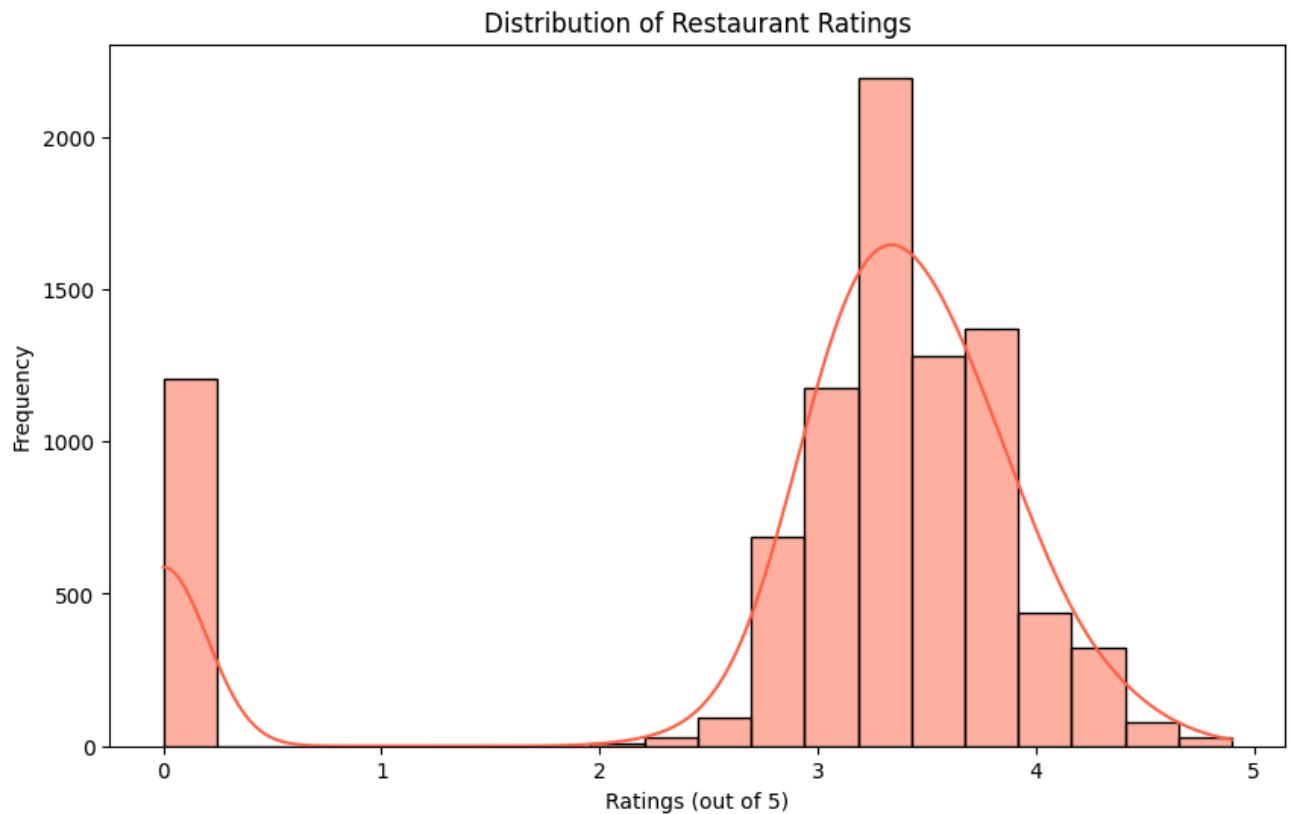
```
# Top 20 Cuisines in Pune (Zomato)
cuisine_counter = Counter([cuisine for sublist in df['Cuisine_List'] for cuisine in sublist])
common_cuisines = pd.DataFrame(cuisine_counter.most_common(20), columns=['Cuisine', 'Count'])

# Plot
plt.figure(figsize=(12, 6))
plt.barh(common_cuisines['Cuisine'], common_cuisines['Count'], color='tomato')
plt.gca().invert_yaxis() # Highest on top
plt.title('Top 20 Cuisines in Pune (Zomato)')
plt.xlabel('Count')
plt.ylabel('Cuisine')
plt.tight_layout()
plt.show()
```



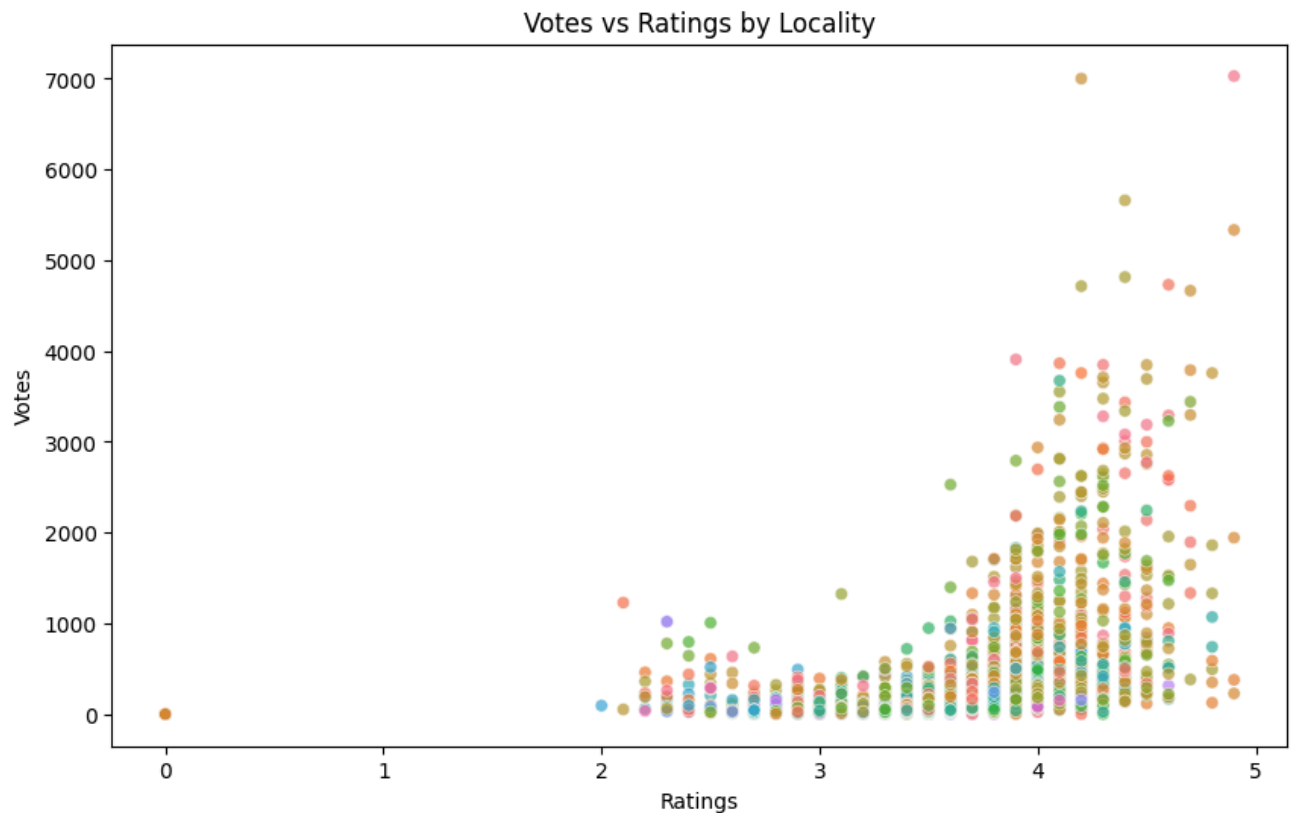
✓ ★ Distribution of Ratings

```
plt.figure(figsize=(10,6))
sns.histplot(df["Ratings_out_of_5"], bins=20, kde=True, color="tomato")
plt.title("Distribution of Restaurant Ratings")
plt.xlabel("Ratings (out of 5)")
plt.ylabel("Frequency")
plt.show()
```



✓ Votes vs Ratings by Locality

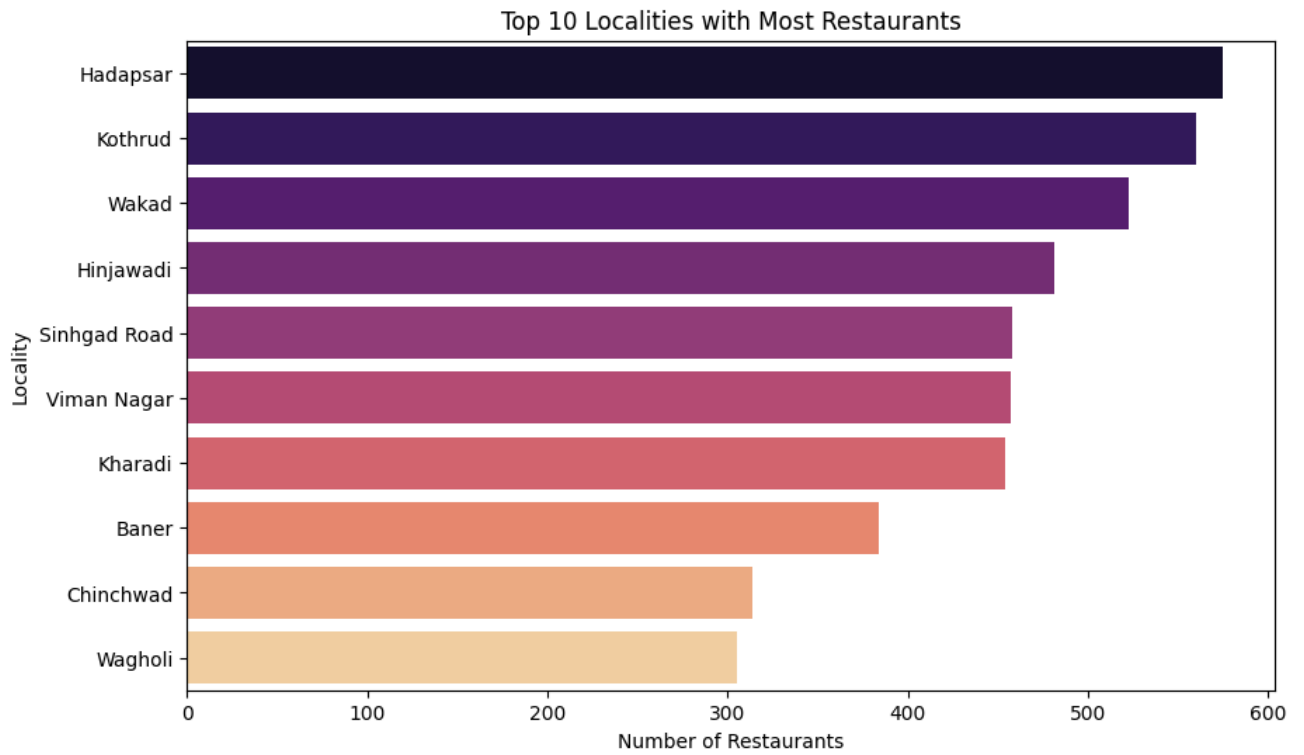
```
# Votes vs Ratings
plt.figure(figsize=(10,6))
sns.scatterplot(x="Ratings_out_of_5", y="Number of votes", data=df, hue="Locality", alpha=0.5)
plt.title("Votes vs Ratings by Locality")
plt.xlabel("Ratings")
plt.ylabel("Votes")
plt.show()
```



✓ 🏠 Top 10 Localities by Restaurant Count

```
# Prepare data for plot
top_localities_df = df['Locality'].value_counts().nlargest(10).reset_index()
top_localities_df.columns = ['Locality', 'Count']

# Plot
plt.figure(figsize=(10, 6))
sns.barplot(data=top_localities_df, x='Count', y='Locality', hue='Locality', palette='magma')
plt.title("Top 10 Localities with Most Restaurants")
plt.xlabel("Number of Restaurants")
plt.ylabel("Locality")
plt.show()
```



```
df_cleaned.to_csv("zomato_pune_cleaned.csv", index=False)
```



7. Conclusion

- Hinjawadi, Kothrud, and Hadapsar have the highest number of restaurants
- North Indian, Chinese, and Continental are the most common cuisines
- Higher cost does not always guarantee better ratings
- Some cuisines like "Biryani" and "Kebab" are top-rated in premium zones