

Zomato Bangalore Restaurants

October 1, 2024

```
[0]: from pyspark.sql import SparkSession
# create session
spark = SparkSession.builder.appName("Zomato analysis").getOrCreate()

[0]: spark

[0]: from pyspark.sql.types import StructField, StructType, IntegerType, StringType, BooleanType, DataType, DecimalType
from pyspark.sql.functions import regexp_replace, col

[0]: # access_key = "use your "
# secret_key = "use your"
spark._jsc.hadoopConfiguration().set("fs.s3a.access.key", access_key)
spark._jsc.hadoopConfiguration().set("fs.s3a.secret.key", secret_key)

spark._jsc.hadoopConfiguration().set("fs.s3a.endpoint", "s3.ap-south-1.
↳amazonaws.com")
zomato_df = spark.read.format("parquet").option("header", "true").load("s3://
↳zomato-csv/zomatofood.parquet")

[0]: df = zomato_df

[0]: df.columns
```

```
Out[6]: ['url',
'address',
'name',
'online_order',
'book_table',
'rate',
'votes',
'phone',
'location',
'rest_type',
'dish_liked',
'cuisines',
'approx_cost(for two people)',
'reviews_list',
```

```
'menu_item',
'listed_in(type)',
'listed_in(city)']
```

```
[0]: df_zomato = spark.read.schema(df_schema).format("parquet").
      ↪option("header","true").load("s3://zomato-csv/zomatofood.parquet")
```

```
[0]: #column is unnecessary for your analysis
columns_to_drop = ["url", "phone", "reviews_list", "menu_item", "dish_liked"]
df_cleaned = df.drop(*columns_to_drop)

# Show cleaned dataframe
# df_cleaned.show(5)
```

```
[0]: df_clean = df_cleaned.withColumn("rate", regexp_replace(col("rate"), "/5", "").
      ↪cast("float"))
```

```
[0]: df_clean.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+
|          address|          name|online_order|book_table|rate|votes|
location|rest_type|cuisines|approx_cost(for two
people)|listed_in(type)|listed_in(city)|
+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+
|942, 21st Main Ro...|Jalsa|Yes|Yes|4.1|
775|Banashankari|Casual Dining|North Indian, Mug...|
800|Buffet|Banashankari|
|2nd Floor, 80 Fee...|Spice Elephant|Yes|No|4.1|
787|Banashankari|Casual Dining|Chinese, North In...|
800|Buffet|Banashankari|
|1112, Next to KIM...|San Churro Cafe|Yes|No|3.8|
918|Banashankari|Cafe, Casual Dining|Cafe, Mexican, It...|
800|Buffet|Banashankari|
|1st Floor, Annaku...|Addhuri Udupi Bho...|No|No|3.7|
88|Banashankari|Quick Bites|South Indian, Nor...|
300|Buffet|Banashankari|
|10, 3rd Floor, La...|Grand Village|No|No|3.8|
166|Basavanagudi|Casual Dining|North Indian, Raj...|
600|Buffet|Banashankari|
|37, 5-1, 4th Floo...|Timepass Dinner|Yes|No|3.8|
286|Basavanagudi|Casual Dining|North Indian|
600|Buffet|Banashankari|
|19/1, New Timbery...|Rosewood Internat...|No|No|3.6|8|
Mysore Road|Casual Dining|North Indian, Sou...|
```

800	Buffet	Banashankari				
2469, 3rd Floor, ...		Onesta	Yes	Yes	4.6	
2556 Banashankari	Casual Dining, Cafe	Pizza, Cafe, Italian				
600	Cafes	Banashankari				
1, 30th Main Road...		Penthouse Cafe	Yes	No	4.0	
324 Banashankari		Cafe Cafe, Italian, Co...				
700	Cafes	Banashankari				
2470, 21 Main Roa...		Smaczego	Yes	No	4.2	
504 Banashankari		Cafe Cafe, Mexican, It...				
550	Cafes	Banashankari				
12,29 Near PES Un...	CafÃ Ã Ã Ã Ã Ã ...		Yes	No	4.1	
402 Banashankari		Cafe		Cafe		
500	Cafes	Banashankari				
941, 3rd FLOOR, 2...		Cafe Shuffle	Yes	Yes	4.2	
150 Banashankari		Cafe Cafe, Italian, Co...				
600	Cafes	Banashankari				
6th Block, 3rd St...		The Coffee Shack	Yes	Yes	4.2	
164 Banashankari		Cafe Cafe, Chinese, Co...				
500	Cafes	Banashankari				
111, Sapphire Toy...		Caf-Eleven	No	No	4.0	
424 Banashankari		Cafe	Cafe, Continental			
450	Cafes	Banashankari				
1112, Next to KIM...		San Churro Cafe	Yes	No	3.8	
918 Banashankari	Cafe, Casual Dining	Cafe, Mexican, It...				
800	Cafes	Banashankari				
2303, 21st Cross,...		Cafe Vivacity	Yes	No	3.8	
90 Banashankari		Cafe		Cafe		
650	Cafes	Banashankari				
241, 4th Floor, 1...		Catch-up-ino	Yes	No	3.9	
133 Banashankari		Cafe Cafe, Fast Food, ...				
800	Cafes	Banashankari				
405, 24th Cross, ...		Kirthi's Biryani	Yes	No	3.8	
144 Banashankari		Cafe Chinese, Cafe, It...				
700	Cafes	Banashankari				
504, CJ Venkata D...		T3H Cafe	No	No	3.9	
93 Banashankari		Cafe Cafe, Italian, Am...				
300	Cafes	Banashankari				
47, 48 &49, 3Rd F...	360 Atoms Restaur...		Yes	No	3.1	
13 Banashankari		Cafe Cafe, Chinese, Co...				
400	Cafes	Banashankari				
+-----+-----+-----+-----+-----+-----+-----						
-----+-----+-----+-----+-----+-----+-----						
+-----+-----+-----+-----+-----+-----+-----						

only showing top 20 rows

```
[0]: df_clean.createOrReplaceTempView("restaurants")

# SQL query to calculate average rating per city
sql_query_ratings = """
SELECT `listed_in(city)`, ROUND(AVG(rate),5) AS avg_rating
FROM restaurants
GROUP BY `listed_in(city)`
ORDER BY avg_rating DESC
"""

# Execute the query and show the results
ratings_by_city = spark.sql(sql_query_ratings)
ratings_by_city.show()

# Visualize (optional)
display(ratings_by_city)
```

```
+-----+-----+
| listed_in(city) | avg_rating |
+-----+-----+
| Church Street | 3.80426 |
| MG Road | 3.80202 |
| Brigade Road | 3.80115 |
| Residency Road | 3.78122 |
| Lavelle Road | 3.77682 |
| Koramangala 6th B... | 3.76315 |
| Koramangala 5th B... | 3.76025 |
| Koramangala 4th B... | 3.75904 |
| Koramangala 7th B... | 3.75841 |
| Indiranagar | 3.74566 |
| Old Airport Road | 3.71904 |
| Malleshwaram | 3.71552 |
| Frazer Town | 3.71484 |
| Basavanagudi | 3.70728 |
| BTM | 3.70724 |
| Jayanagar | 3.69497 |
| Banashankari | 3.66676 |
| HSR | 3.66082 |
| Rajajinagar | 3.65695 |
| Kalyan Nagar | 3.64791 |
+-----+-----+
```

only showing top 20 rows

```
[0]: # SQL query to calculate average cost per city
sql_query_cost = """
SELECT `listed_in(city)`, AVG(`approx_cost(for two people)`) AS avg_cost
FROM restaurants
```

```
GROUP BY `listed_in(city)`
ORDER BY avg_cost DESC
"""

# Execute the query and show the results
cost_by_city = spark.sql(sql_query_cost)
cost_by_city.show()

# Visualize (optional)
display(cost_by_city)
```

```
+-----+-----+
| listed_in(city)| avg_cost|
+-----+-----+
| Brigade Road| 431.0509554140127|
| MG Road|430.82223962411905|
| Brookefield| 428.9642596644785|
| Church Street|428.49378881987576|
| Bellandur|427.37827715355803|
|Koramangala 4th B...| 427.1937730438345|
| Sarjapur Road| 427.1376811594203|
| Kalyan Nagar| 425.9046052631579|
| HSR|425.50995024875624|
| Indiranagar| 424.781121751026|
| Whitefield|423.99421128798843|
| Lavelle Road|423.00158856235106|
| Kammanahalli| 420.6924315619968|
| Residency Road| 420.6663815226689|
|Koramangala 6th B...| 420.2173913043478|
| Marathahalli| 418.2536520584329|
| Old Airport Road| 417.0201577563541|
|Koramangala 5th B...|414.97365220916095|
|Koramangala 7th B...| 414.7888975762314|
| Jayanagar| 412.3905109489051|
+-----+-----+
```

only showing top 20 rows

```
[0]: # SQL query to calculate average cost per city
sql_query_cost = """
SELECT `listed_in(city)`, AVG(`approx_cost(for two people)`) AS avg_cost
FROM restaurants
GROUP BY `listed_in(city)`
ORDER BY avg_cost DESC
"""

# Execute the query and show the results
```

```
cost_by_city = spark.sql(sql_query_cost)
cost_by_city.show()
```

```
# Visualize (optional)
display(cost_by_city)
```

```
+-----+-----+
| listed_in(city)| avg_cost|
+-----+-----+
| Brigade Road| 431.0509554140127|
| MG Road|430.82223962411905|
| Brookefield| 428.9642596644785|
| Church Street|428.49378881987576|
| Bellandur|427.37827715355803|
|Koramangala 4th B...| 427.1937730438345|
| Sarjapur Road| 427.1376811594203|
| Kalyan Nagar| 425.9046052631579|
| HSR|425.50995024875624|
| Indiranagar| 424.781121751026|
| Whitefield|423.99421128798843|
| Lavelle Road|423.00158856235106|
| Kammanahalli| 420.6924315619968|
| Residency Road| 420.6663815226689|
|Koramangala 6th B...| 420.2173913043478|
| Marathahalli| 418.2536520584329|
| Old Airport Road| 417.0201577563541|
|Koramangala 5th B...|414.97365220916095|
|Koramangala 7th B...| 414.7888975762314|
| Jayanagar| 412.3905109489051|
+-----+-----+
only showing top 20 rows
```

```
[0]: # SQL query to count cuisine frequency
sql_query_cuisine = """
SELECT cuisines, COUNT(*) AS count
FROM restaurants
GROUP BY cuisines
ORDER BY count DESC
"""

# Execute the query and show the results
popular_cuisines = spark.sql(sql_query_cuisine)
popular_cuisines.show()

# Visualize (optional)
display(popular_cuisines)
```

```

+-----+-----+
|          cuisines|count|
+-----+-----+
|      North Indian| 2913|
|North Indian, Chi...| 2385|
|      South Indian| 1828|
|      Biryani|    918|
|    Bakery, Desserts|  911|
|      Fast Food|   803|
|      Desserts|   766|
|      Cafe|    756|
|South Indian, Nor...|  726|
|      Bakery|   651|
|      Chinese|   556|
| Ice Cream, Desserts|  417|
|Chinese, North In...|  415|
| Mithai, Street Food|  372|
| Desserts, Ice Cream|  354|
|North Indian, Chi...|  352|
|North Indian, Sou...|  343|
|South Indian, Nor...|  343|
|North Indian, Sou...|  305|
|      Beverages|   301|
+-----+-----+

```

only showing top 20 rows

```

[0]: # SQL query to calculate average rating based on table booking
sql_query_table_booking = """
SELECT book_table, AVG(rate) AS avg_rating
FROM restaurants
GROUP BY book_table
ORDER BY avg_rating DESC
"""

# Execute the query and show the results
table_booking_ratings = spark.sql(sql_query_table_booking)
table_booking_ratings.show()

# Visualize (optional)
display(table_booking_ratings)

```

```

+-----+-----+
|book_table|      avg_rating|
+-----+-----+
|      Yes|4.143464461082432|
|      No|3.621469978125015|
+-----+-----+

```

```
[0]: # SQL query to find top restaurants by votes
sql_query_top_restaurants = """
SELECT name, votes, rate, `listed_in(city)`
FROM restaurants
ORDER BY votes DESC
LIMIT 10
"""

# Execute the query and show the results
top_restaurants = spark.sql(sql_query_top_restaurants)
top_restaurants.show()

# Visualize (optional)
display(top_restaurants)
```

```
+-----+-----+-----+-----+
|          name|votes|rate|          listed_in(city)|
+-----+-----+-----+-----+
|Byg Brewski Brewi...|16832| 4.9|          Sarjapur Road|
|Byg Brewski Brewi...|16832| 4.9|          Sarjapur Road|
|Byg Brewski Brewi...|16832| 4.9|          Sarjapur Road|
|Byg Brewski Brewi...|16345| 4.9|          Bellandur|
|Byg Brewski Brewi...|16345| 4.9|          Bellandur|
|Byg Brewski Brewi...|16345| 4.9|          Bellandur|
|          Toit|14956| 4.7|          Indiranagar|
|          Toit|14956| 4.7|          Indiranagar|
|          Truffles|14726| 4.7|Koramangala 7th B...|
|          Truffles|14723| 4.7|Koramangala 6th B...|
+-----+-----+-----+-----+
```

```
[0]: # SQL query to find cuisine variety by restaurant type
sql_query_cuisine_variety = """
SELECT rest_type, cuisines, COUNT(*) AS count
FROM restaurants
GROUP BY rest_type, cuisines
ORDER BY count DESC
"""

# Execute the query and show the results
cuisine_variety = spark.sql(sql_query_cuisine_variety)
cuisine_variety.show()

# Visualize (optional)
display(cuisine_variety)
```

```
+-----+-----+-----+-----+
```


rest_type	cuisines	count
Quick Bites	South Indian	1532
Quick Bites	North Indian	1480
Quick Bites	North Indian, Chi...	1259
Cafe	Cafe	737
Casual Dining	North Indian, Chi...	712
Casual Dining	North Indian	695
Quick Bites	Fast Food	681
Quick Bites	Biryani	678
Bakery	Bakery, Desserts	561
Dessert Parlor	Desserts	561
Quick Bites	South Indian, Nor...	504
Bakery	Bakery	395
Dessert Parlor	Ice Cream, Desserts	386
Delivery	North Indian	320
Dessert Parlor	Desserts, Ice Cream	310
Quick Bites	Chinese	239
Quick Bites	South Indian, Nor...	238
Quick Bites	Biryani, Kebab	233
Quick Bites	Chinese, North In...	232
Quick Bites	South Indian, Chi...	221

only showing top 20 rows

[0]: