



Data Zomato

DI KOTA-KOTA METROPOLITAN INDIA

Data Science
Final Project

Adhi Sukmawan
23B

Business Understanding

Zomato adalah perusahaan multinasional dari India yang bergerak di bidang restaurant aggregator dan food delivery

Semakin berkembangnya teknologi telekomunikasi membuat industri food and beverage juga berkembang pesat karena adanya dukungan layanan delivery dan kemudahan untuk melakukan pencarian online

Konteks

Disini saya akan memproses data, mengerjakan Exploratory Data Analysis dan Deep Dive Question

Targetnya adalah memberikan insight kepada:

1. Pemilik bisnis, yang ingin membuka bisnis restaurant atau mengembangkan bisnis restaurantnya untuk mengetahui area yang ramai disetiap kota pada dataset ini dan jenis makanan apa yang populer
2. Partner delivery service, untuk menentukan mana area paling ramai
3. Pendetang atau turis, yang ingin mencari tempat kuliner yang tepat untuk mereka

Data Understanding dan Preprocessing

```
zomato.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123657 entries, 0 to 123656
Data columns (total 12 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Restaurant Name      123657 non-null object
1   Dining Rating        91421 non-null  float64
2   Delivery Rating      122377 non-null float64
3   Dining Votes         123657 non-null int64
4   Delivery Votes       123657 non-null int64
5   Cuisine              123657 non-null object
6   Place Name           123657 non-null object
7   City                 123657 non-null object
8   Item Name            123657 non-null object
9   Best Seller          27942 non-null  object
10  Votes                123657 non-null int64
11  Prices               123657 non-null float64
dtypes: float64(3), int64(3), object(6)
memory usage: 11.3+ MB
```

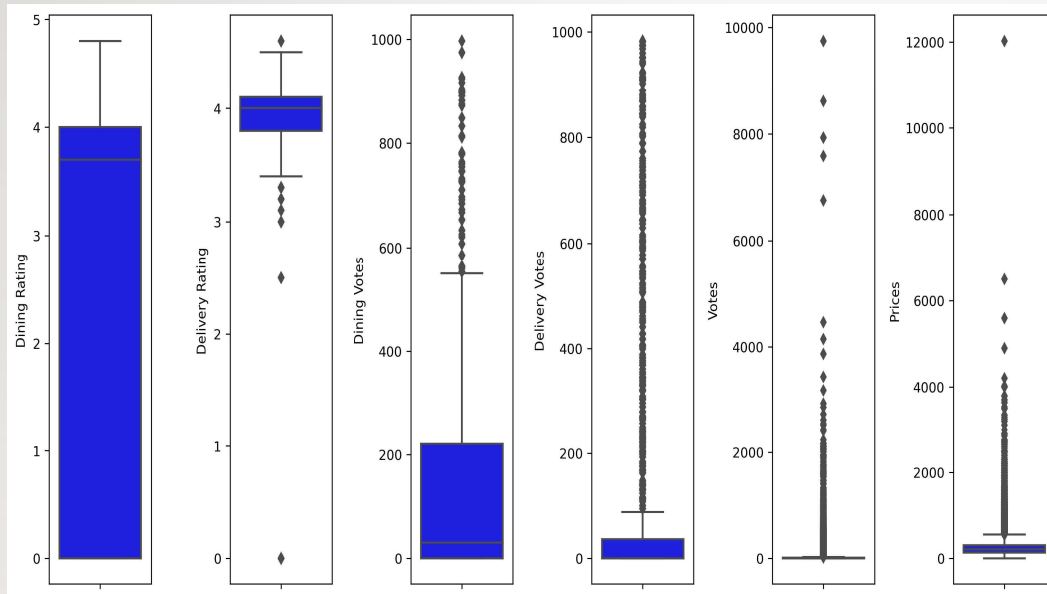
- Dataframe terdiri dari 12 kolom dengan 123657 baris yang terdiri dari tipe object, int64 dan float64
- Jumlah data yang terduplikasi ada 22127 baris dan redundant datanya dihilangkan
- Nilai NaN (null) ada 110598 entry

- Null pada data numerical diubah menjadi 0 sedangkan pada data categorical diubah menjadi 'No'
- Nilai pada kolom 'Best Seller' diubah menjadi 'Yes' dan 'No' kemudian untuk proses selanjutnya diubah menjadi '1' dan '0'
- Setelah melakukan data cleaning, dataframe menjadi 12 kolom dengan 99307 baris

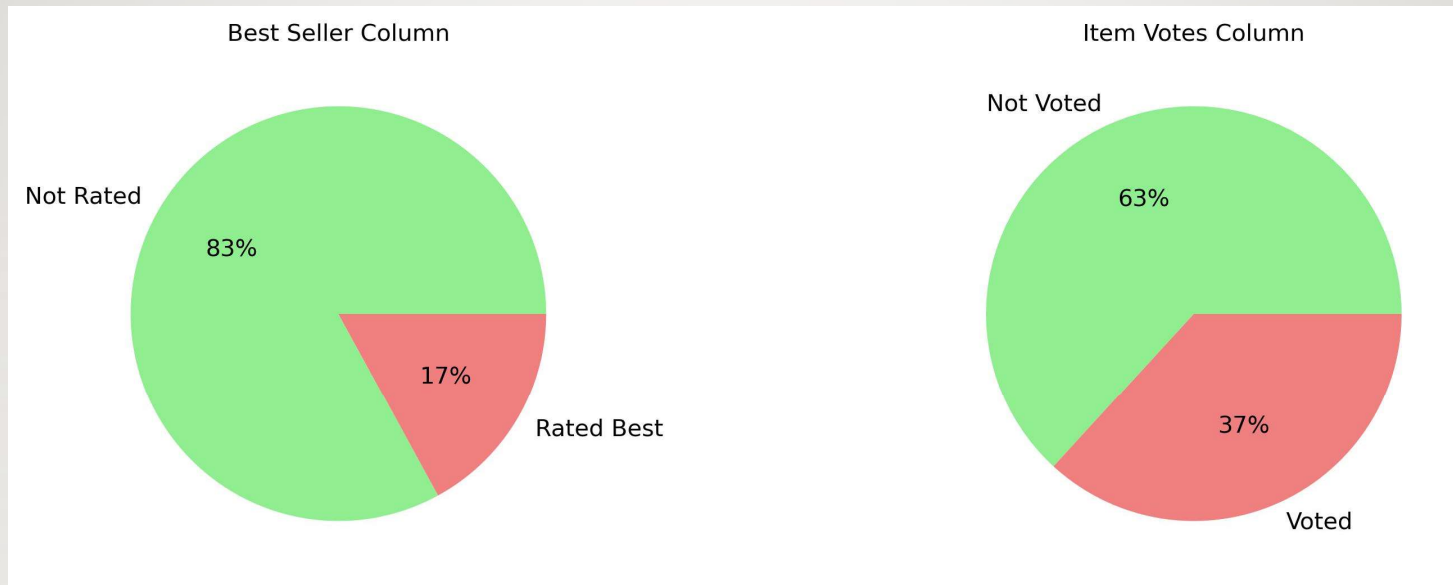
```
zomato.info()

<class 'pandas.core.frame.DataFrame'>
Index: 99307 entries, 0 to 123634
Data columns (total 12 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Restaurant Name      99307 non-null  object  
1   Dining Rating        99307 non-null  float64  
2   Delivery Rating      99307 non-null  float64  
3   Dining Votes         99307 non-null  int64  
4   Delivery Votes       99307 non-null  int64  
5   Cuisine              99307 non-null  object  
6   Place Name           99307 non-null  object  
7   City                 99307 non-null  object  
8   Item Name            99307 non-null  object  
9   Best Seller          99307 non-null  int64  
10  Votes                99307 non-null  int64  
11  Prices               99307 non-null  float64  
dtypes: float64(3), int64(4), object(5)
memory usage: 9.8+ MB
```

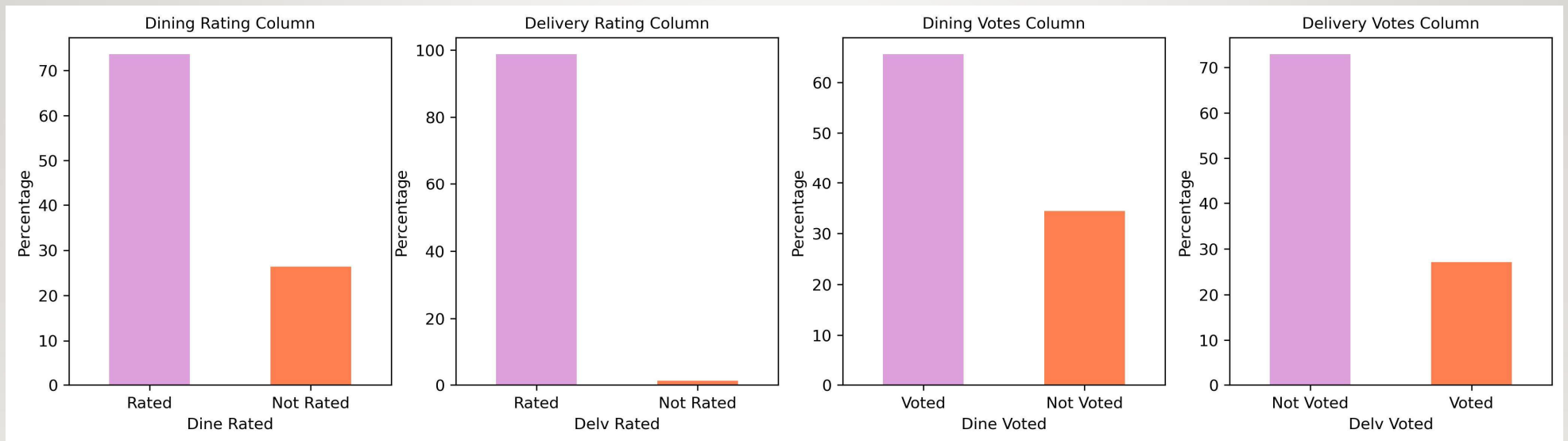
Exploratory Data Analysis



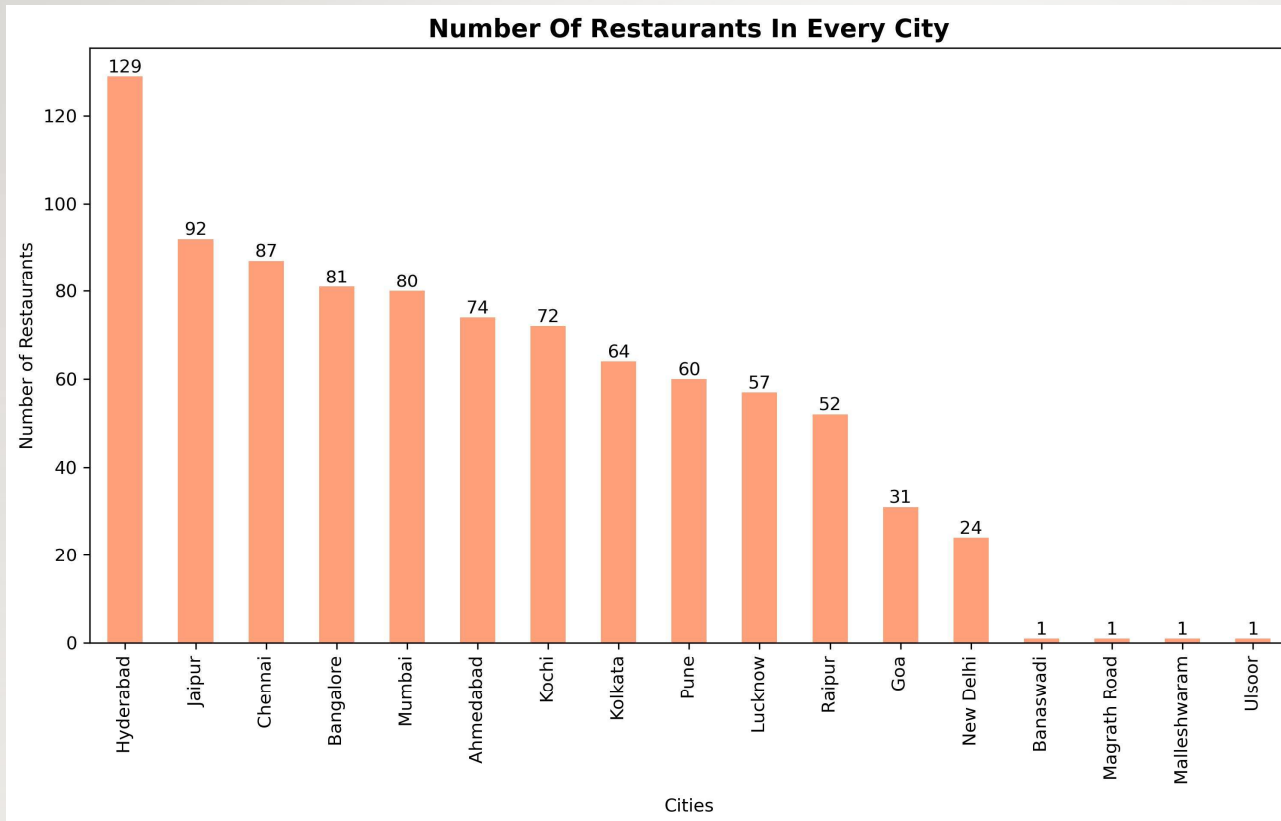
- Penyebaran data rating dan votes sangat tidak merata sehingga banyak entry yang dianggap sebagai outlier, tetapi dibiarkan karena memang nilai2 tersebut masuk akal



- Makanan yang di vote sebagai Best Seller adalah 17% dari total makanan yang terdaftar
- Makanan yang di vote untuk direkomendasikan dibeli adalah 37% dari total makanan yang terdaftar

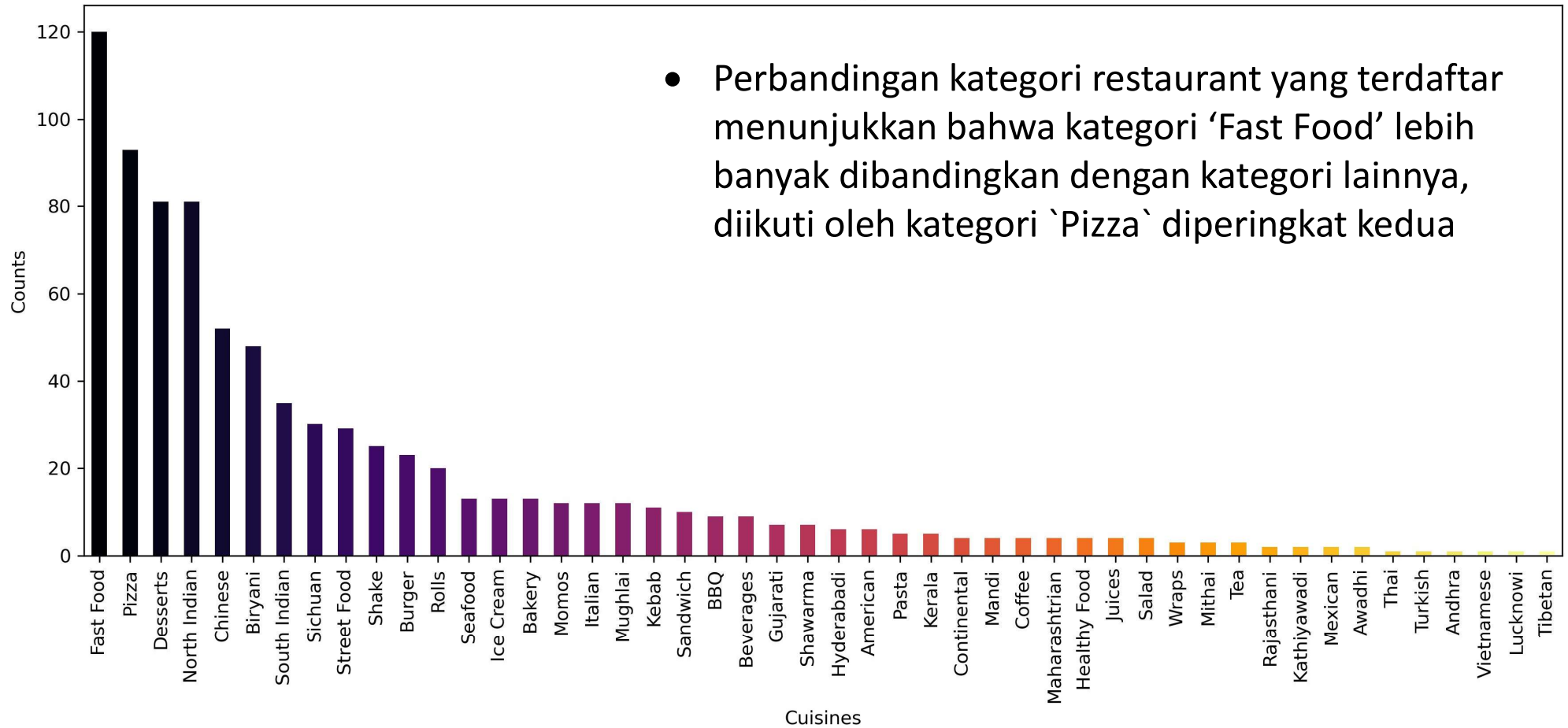


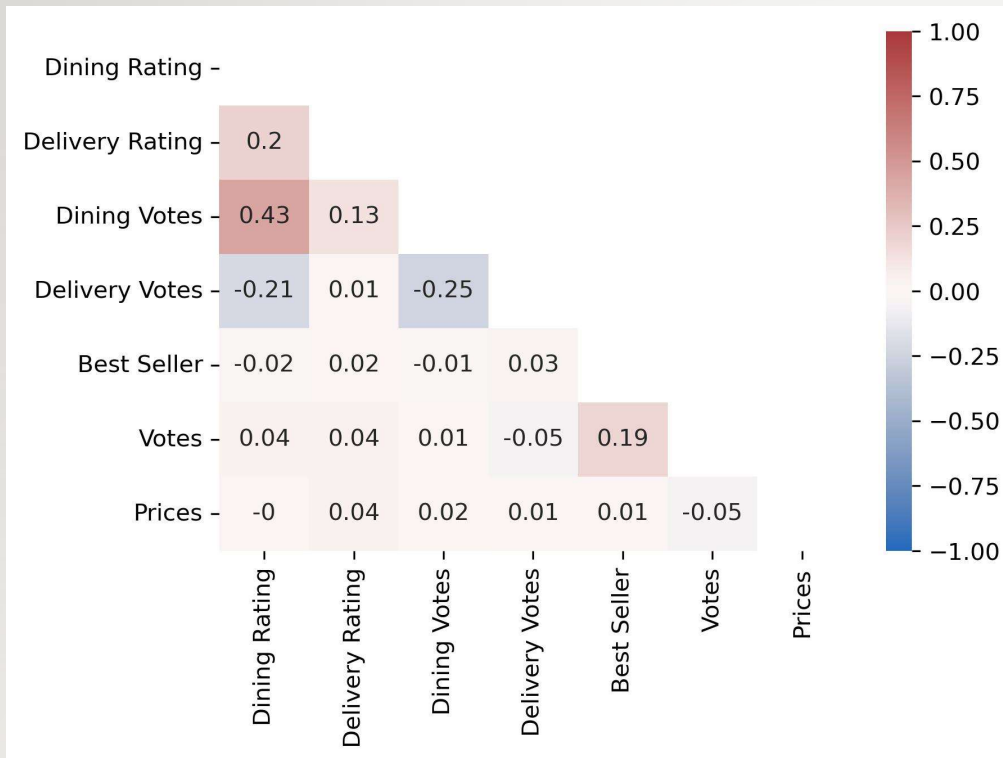
- Data untuk rating dan votes juga sangat imbalance



- Distribusi jumlah restaurant di setiap kota menunjukkan kota Hyderabad sebagai kota dengan jumlah restaurant yang terbanyak dari seluruh 17 kota metropolitan yang terdaftar dengan 129 restaurant dari total keseluruhan 826 restaurant

Cuisine Comparison

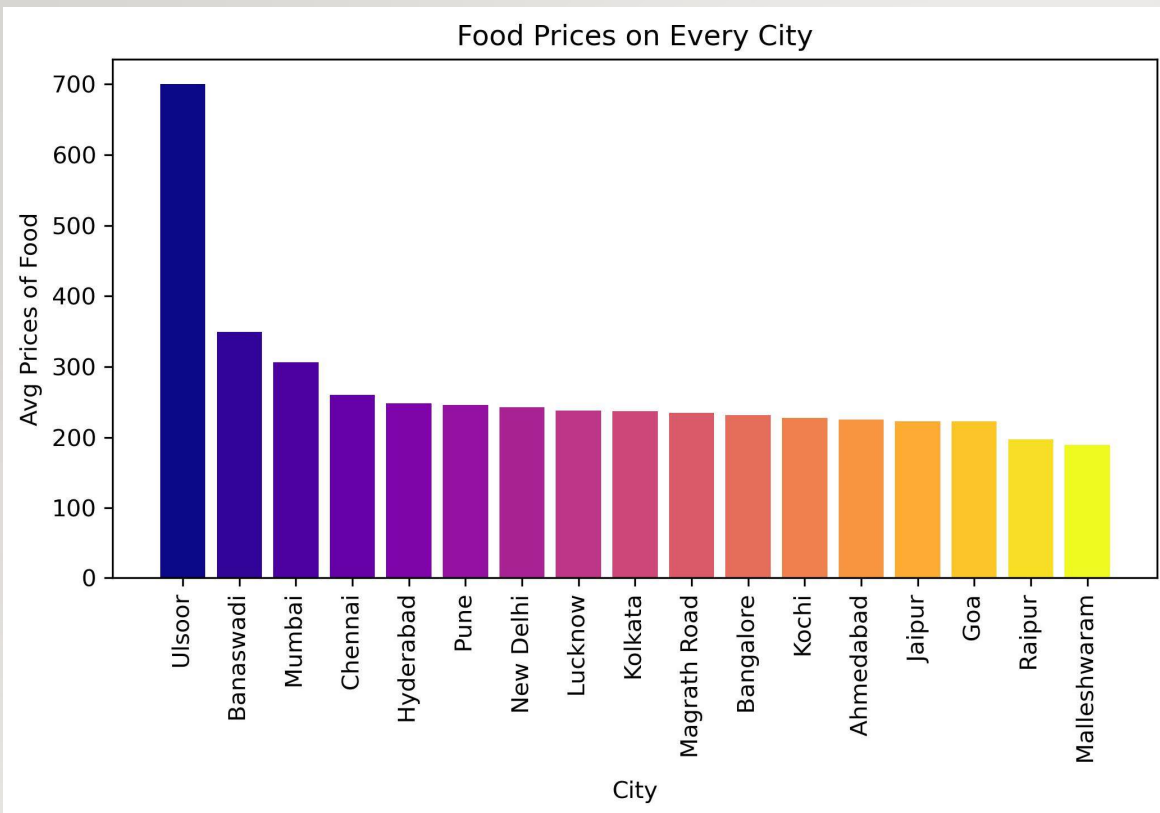




- Multicollinearity check menunjukkan tidak ada kolom yang memiliki korelasi yang kuat dengan kolom lainnya

Deep Dive Question

- Saya akan melanjutkan dengan pemrosesan data yang diharapkan dapat memberikan insight yang diperlukan dengan deep dive question
- Disini saya akan memberikan contoh dengan kota Hyderabad yang memiliki jumlah restaurant terbanyak yakni 129 restaurant dan mencakup 37 area
- Insight untuk kota lainnya dapat dilihat dalam script code saya



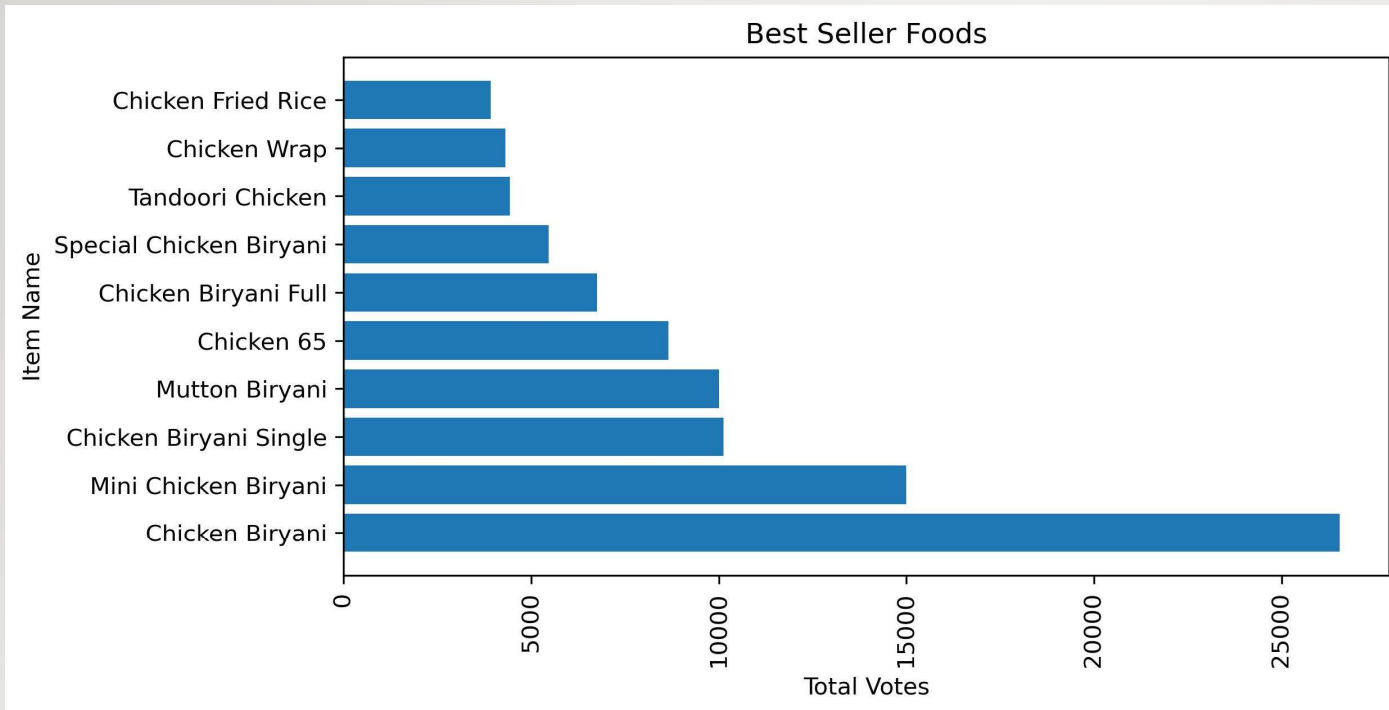
Highlight:

Kota Ulsoor adalah kota dengan harga makanan rata-rata termahal

Insight:

Harga rata-rata makanan di kota-kota metropolitan di India adalah sekitar 200 – 300 Rupee

	City	max	min	median	mean
0	Ulsoor	1999.00	139.00	529.0	700.389831
1	Banaswadi	1250.00	47.65	330.0	349.466471
2	Mumbai	3800.00	0.95	280.0	306.135515
3	Chennai	6500.00	0.95	225.0	259.927833
4	Hyderabad	3250.00	0.95	210.0	248.443447
5	Pune	2560.00	0.95	219.0	245.684117
6	New Delhi	2069.00	2.00	200.0	242.382990
7	Lucknow	5593.22	1.00	199.0	237.712971
8	Kolkata	12024.00	0.95	200.0	237.164973
9	Magrath Road	864.79	75.42	259.0	234.553250
10	Bangalore	3999.00	0.95	200.0	231.262688
11	Kochi	2749.00	0.95	190.0	227.801003
12	Ahmedabad	2599.00	2.00	195.0	225.199184
13	Jaipur	3000.00	0.95	195.0	222.676428
14	Goa	2300.00	8.00	200.0	222.521372
15	Raipur	3999.00	5.00	175.0	197.490858
16	Malleshwaram	325.00	70.00	210.0	189.354839

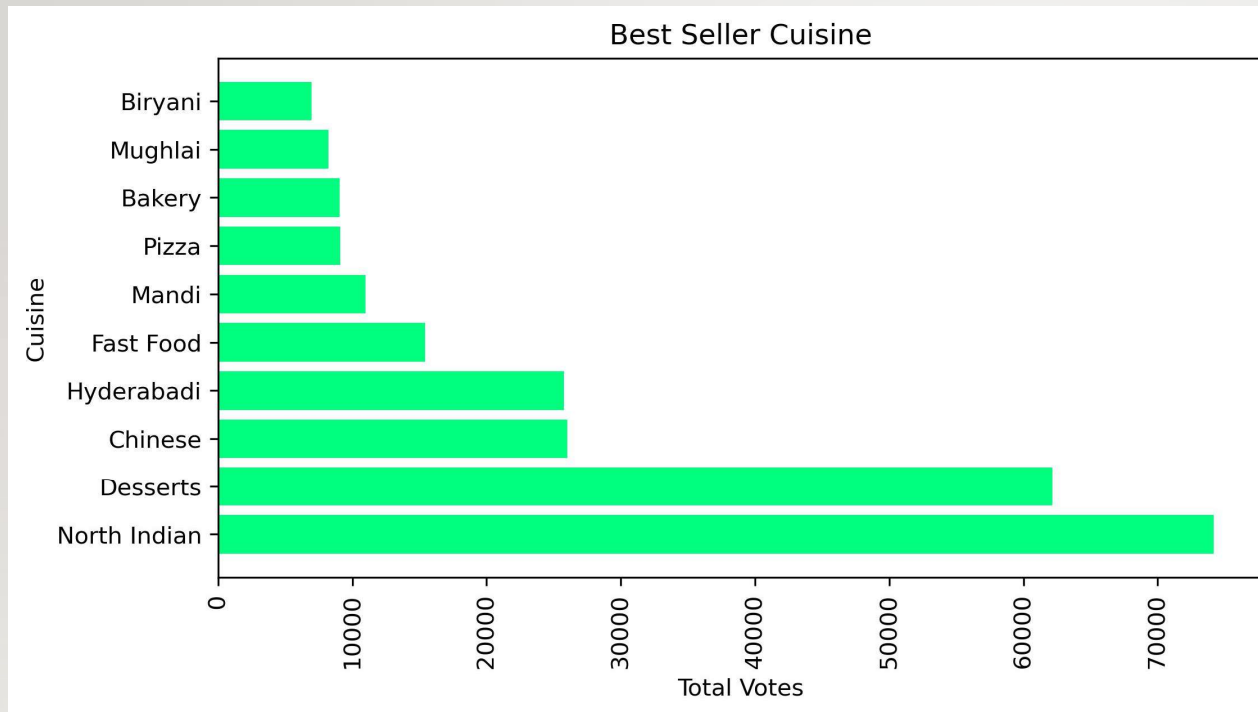


Highlight:

Di kota Hyderabad, Chicken Biryani adalah menu best seller

Insight:

Chicken Biryani mendominasi menu best seller di kota Hyderabad dengan total 88 menu Chicken Biryani dan turunannya dengan total 81344 votes yang merekomendasikan menu-menu tersebut

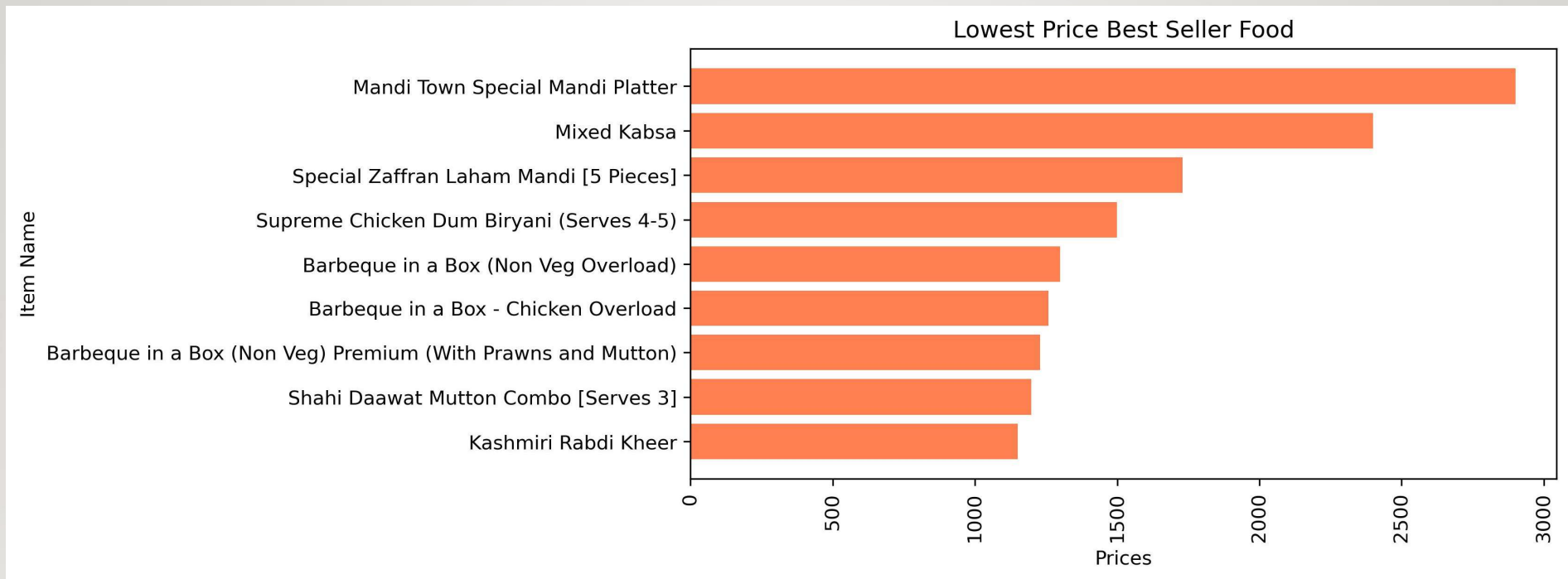


Highlight:

Kategori makanan 'North Indian' adalah yang menjadi best seller di kota Hyderabad, diikuti oleh kategori 'Desserts'

Insight:

Makanan lokal mendominasi kategori best seller di kota Hyderabad

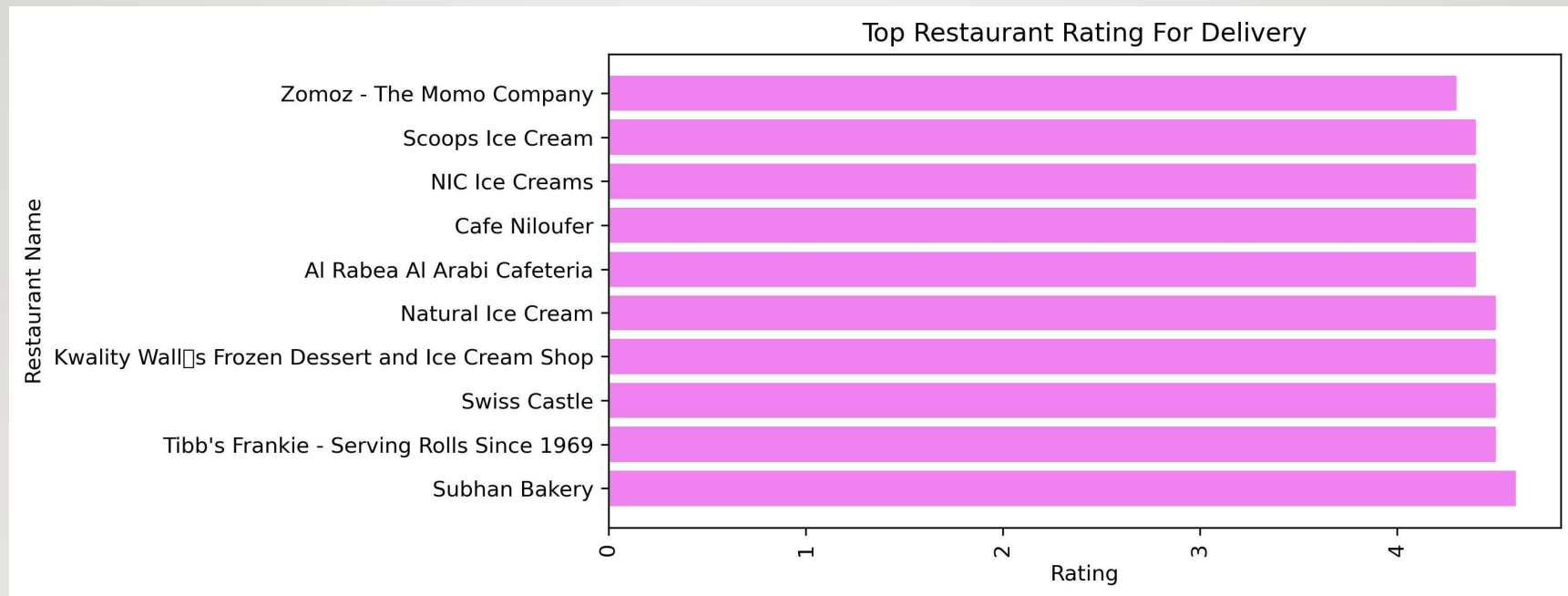


Highlight:

Kashmiri Rabdi Kheer dari restaurant 'Spice 9' merupakan menu best seller termurah diikuti oleh Shahi Daawat Mutton Combo [Serves 3] dari restaurant 'Art Of Biryani'

Insight:

1. Harga menu 10 besar best seller termurah di kota Hyderabad adalah sekitar 1100 – 3000 Rupee
2. Beberapa menu barbeque dari restaurant 'Barbeque Nation' masuk dalam kategori 10 besar ini



Highlight:

‘Subhan Bakery’ merupakan restaurant dengan rating delivery tertinggi

Insight:

Kategori fast food dan dessert mendominasi restaurant dengan delivery rating terbaik



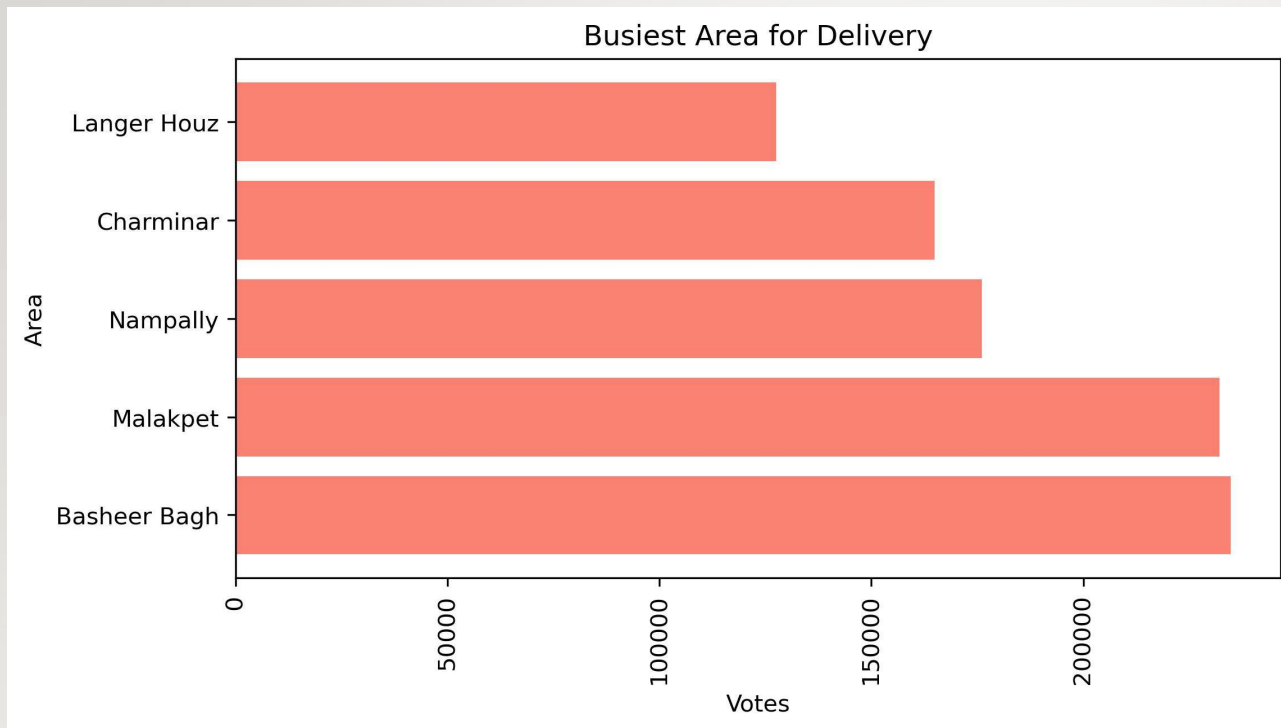
Highlight:

Restaurant seafood 'Exotica' di area 12th Square Building merupakan restaurant yang memiliki rating tertinggi untuk kategori dine-in di kota Hyderabad diikuti oleh restaurant 'Govind Dosa' dan 'Shiddique Kabab Centre'



Highlight:

Restaurant dessert 'Zaffran Mataam Alarabi' di area Falaknuma merupakan yang paling sibuk di kota Hyderabad dilihat dari jumlah rekomendasi dine-in dan delivery diikuti oleh restaurant dengan menu south Indian 'Govind Dosa' di area Charminar

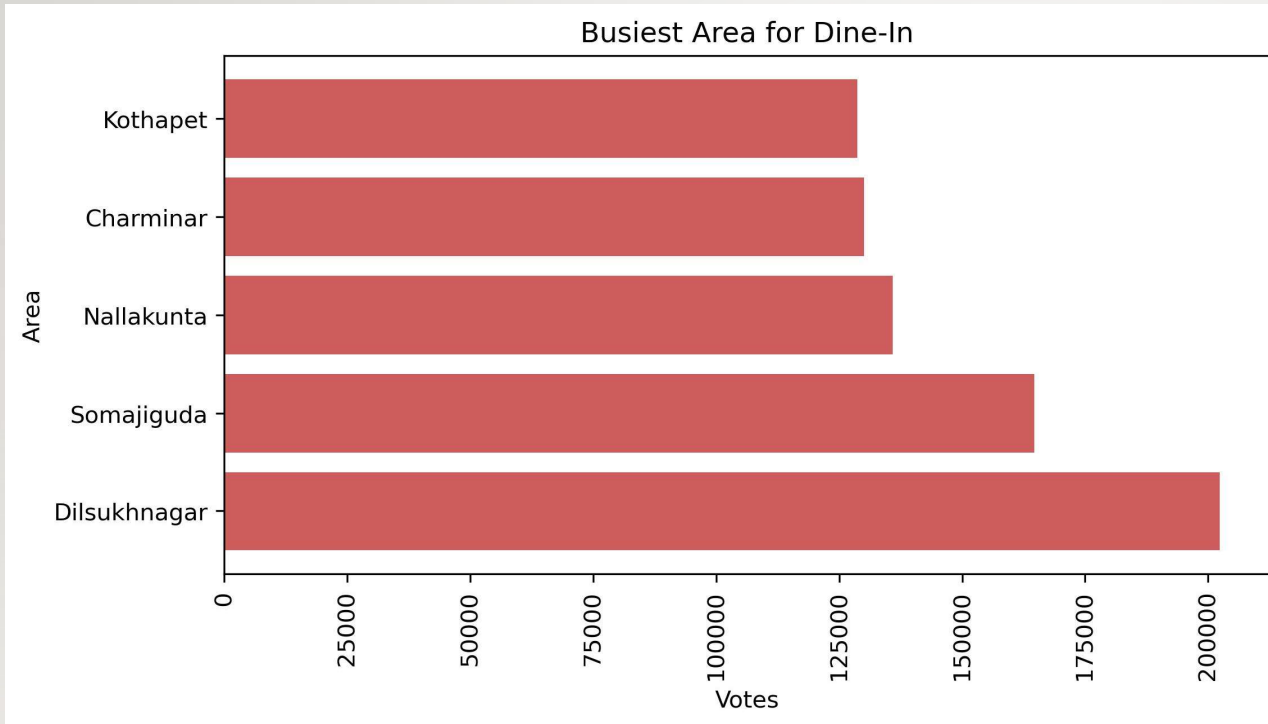


Highlight:

Area Basheer Bagh dan kemudian Malakpet merupakan area yang paling sibuk untuk delivery di kota Hyderabad

Insight:

Diperlukan banyak armada delivery untuk standby di area-area tersebut



Highlight:

Area Dilsukhnagar dan kemudian Somajiguda merupakan area-area yang paling ramai untuk dine-in

Insight:

Pendatang atau turis dapat mendatangi area-area tersebut untuk kuliner



Highlight:

Area di kota Hyderabad dengan harga makanan rata-rata paling murah adalah area Taj Mahal Hotel diikuti oleh area Koti dan area Attapur dengan harga rata-rata dibawah 150 Rupee

Insight:

Pendatang atau turis yang ingin menikmati kuliner dengan harga yang murah dapat pergi ke area-area tersebut

Kesimpulan

- Insight yang dapat diberikan kepada pemilik bisnis untuk menentukan area mana dari kota tersebut yang paling ramai dan apa jenis makanan populer juga harga rata-rata makanan di area tersebut
- Insight untuk partner delivery service adalah menentukan area mana yang ramai untuk delivery sehingga dapat mengatur penempatan kendaraan untuk mencegah keterlambatan pengiriman
- Insight untuk pendatang atau turis adalah informasi harga rata-rata makanan berdasarkan kota dan area juga informasi area yang ramai untuk dine-in

Terima kasih



Sumber data:

<https://www.kaggle.com/datasets/narsingraogoud/zomato-restaurants-dataset-for-metropolitan-areas/data>