

Venues in Dehradun, India

Contents

- Introduction
- Data Collection
- Data Cleaning
- Methodology
- Analysis
- Results & Discussion
- Conclusion

Data

Collection: To begin with, let's look at Dehradun on the Map using the folium library. Latitude & Longitudes where the map is centered are supplied *manually*.

The data is fetched from two different APIs.

- **Foursquare API:** The Foursquare API to fetch venues in Dehradun starting from the centre upto 10 Kilometers in each direction.
- **Zomato API:** The Zomato API provides information about various venues including the complete address, user ratings, price for two people, price range and a lot more.

Cleaning: The data from multiple resources might not always align. Thus, it is important to combine the data retrieved from multiple resources properly.

First plot the two data points on the map. Then try to combine data points that have their latitude and longitude values very close to one another. From the remaining selected venues, inspect the venues to ensure that any remaining mismatched venues are also removed from the final dataset of venues before any analysis.

- To combine the two (Foursquare & Zomato) datasets, check that the latitude and longitude values of each corresponding venue match. Thus, round both the latitude and longitude values up to 4 decimal places. Then, calculate the difference between the corresponding latitude and longitude values and check if the difference is less than 0.0004 which should ideally mean that the two locations are same.
- Drop the venues which have 0.0 rating as it's significant or not been rated yet.

```
In [11]: foursquare_venues['lat'] = foursquare_venues['lat'].apply(lambda lat: round(float(lat), 4))
foursquare_venues['lng'] = foursquare_venues['lng'].apply(lambda lng: round(float(lng), 4))
zomato_venues['latitude'] = zomato_venues['latitude'].apply(lambda lat: round(float(lat), 4))
zomato_venues['longitude'] = zomato_venues['longitude'].apply(lambda lng: round(float(lng), 4))
```

```
In [12]: dataset = pd.concat([foursquare_venues, zomato_venues], axis = 1)
#dataset
#dataset['lat_diff'] = dataset['latitude'] - dataset['lat']
#dataset['lng_diff'] = dataset['longitude'] - dataset['lng']
```

```
In [13]: #selected_venues = dataset[(abs(dataset['lat_diff']) <= 0.0004) & (abs(dataset['lng_diff']) <= 0.0004)].reset_index(drop = True)
selected_venues=dataset
selected_venues
```

```
Out[13]:
```

	name	categories	lat	lng	venue	latitude	longitude	price_for_two	price_range	rating	address
0	The Buffet	Fast Food Restaurant	30.3270	78.0452	Ambrosia	30.3279	78.0453	1200.0	4.0	3.6	21, Centre Point, Near Astley Hall, Rajpur Roa...
1	Anandam	Indian Restaurant	30.3353	78.0532	Anandam	30.3351	78.0533	650.0	3.0	4.3	69, Krishna Tower, Rajpur Road, Hathibarkala S...
2	KFC	Fast Food Restaurant	30.3338	78.0515	KFC	30.3336	78.0516	450.0	2.0	4.3	65A, Rajpur Road, Hathibarkala Salwala, Dehradun

Drop the venues which have 0.0 rating as it's significant or not been rated yet.

```
In [15]: selected_venues = selected_venues[selected_venues['rating'] != 0.0]
print("Total venues available: {}".format(selected_venues.shape[0]))
```

Total venues available: 29