Identifying the best hotels of Chennai using Zomato API

- Cyril Bastin M

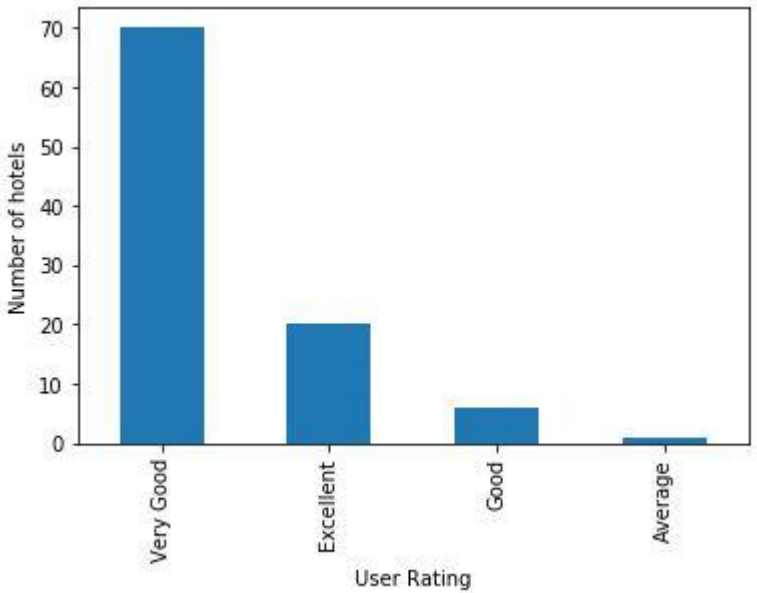
Data Extraction, Transform and Load to DataFrame

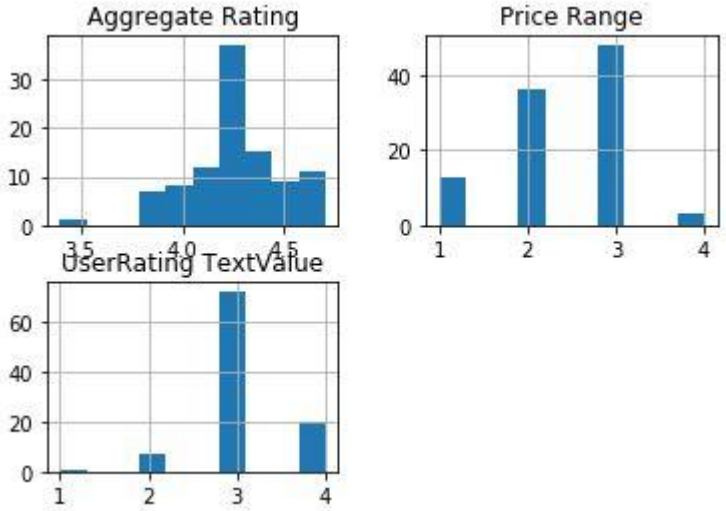

```
{'location': {'entity_type': 'subzone',
  'entity_id': 6008,
  'title': 'Park Town',
  'latitude': '13.0798960000',
  'longitude': '80.2751110000',
  'city_id': 7,
  'city_name': 'Chennai',
  'country_id': 1,
  'country_name': 'India'},
 'popularity': {'popularity': '4.33',
  'nightlife_index': '0.98',
  'nearby_res': ['66330',
   '67408',
   '66334',
   '71679',
   '66342',
   '66312',
   '68329',
   '18428769',
```


```
Index(['restaurant.R.res_id', 'restaurant.all_reviews_count',
       'restaurant.average_cost_for_two', 'restaurant.cuisines',
       'restaurant.establishment', 'restaurant.events_url',
       'restaurant.has_online_delivery', 'restaurant.highlights',
       'restaurant.is_table_reservation_supported',
       'restaurant.location.address', 'restaurant.location.city',
       'restaurant.location.latitude', 'restaurant.location.locality',
       'restaurant.location.longitude', 'restaurant.location.zipcode',
       'restaurant.name', 'restaurant.offers', 'restaurant.opentable_support',
       'restaurant.phone_numbers', 'restaurant.price_range',
       'restaurant.store_type', 'restaurant.timings',
       'restaurant.user_rating.aggregate_rating',
       'restaurant.user_rating.rating_color',
       'restaurant.user_rating.rating_text', 'restaurant.user_rating.votes'],
      dtype='object')
```

City	Latitude	Locality	Longitude	Price Range	Aggregate Rating	UserRating Color	UserRating Text	UserRating Votes
Chennai	13.000380	Alandur	80.200623	3	4.3	5BA829	Very Good	45755
Chennai	12.970815	Pallavaram	80.147182	2	4.5	3F7E00	Excellent	3700
Chennai	13.022305	Kotturpuram	80.242340	3	4.2	5BA829	Very Good	2731

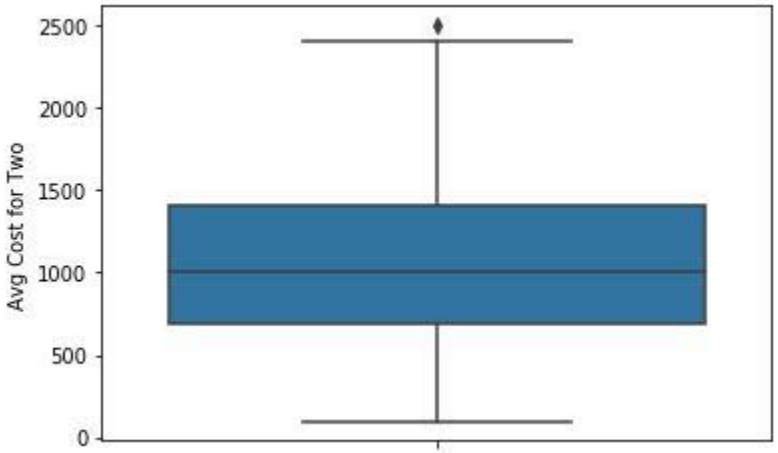
Distribution of data





The data is distributed for user rating and price range of the food that cost for two.

Outliers



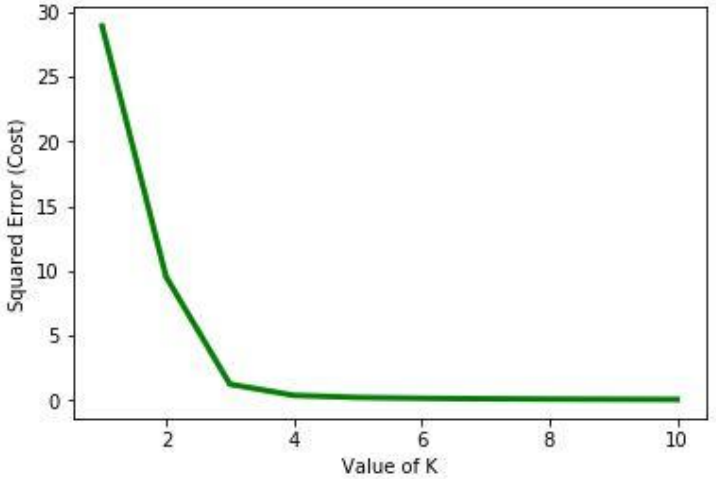
Based on the average cost for two people at a hotel, outliers are removed from the data.

Assumption: Hotels with high price ranges cannot be rated together with that of the lower ones.

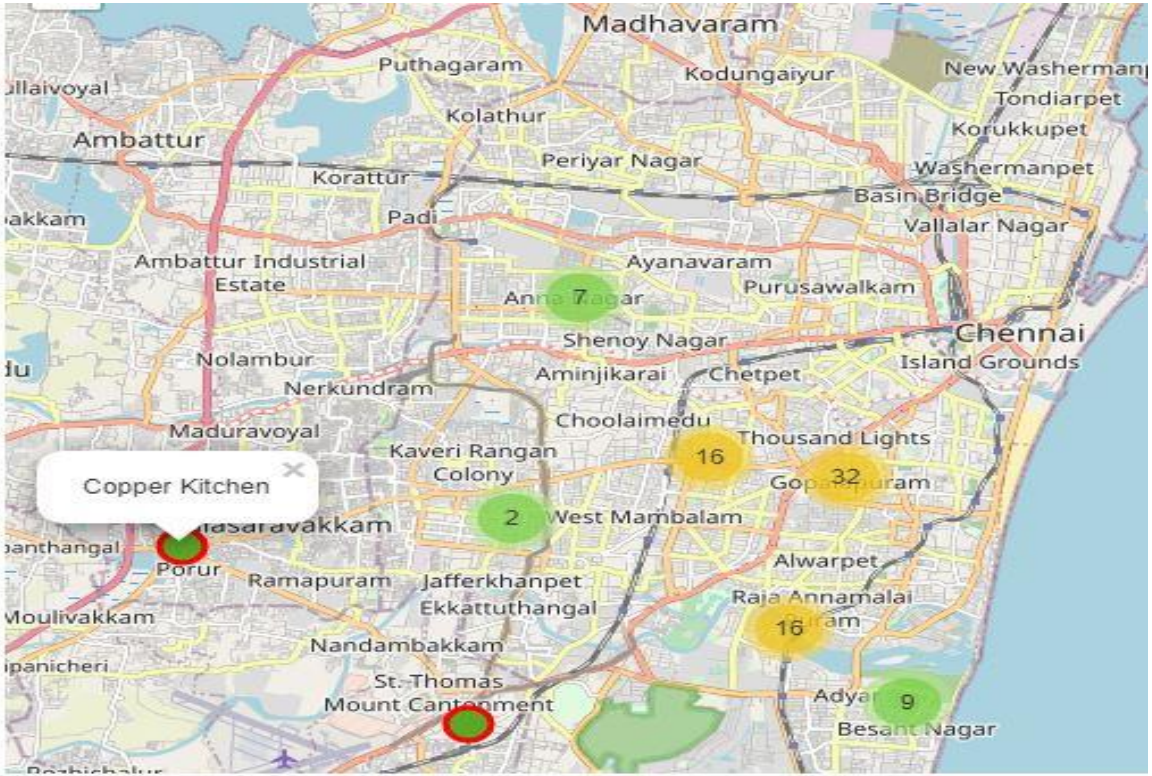
The best value of K and the classified data

```
df_data1['Classifier'].value_counts()
```

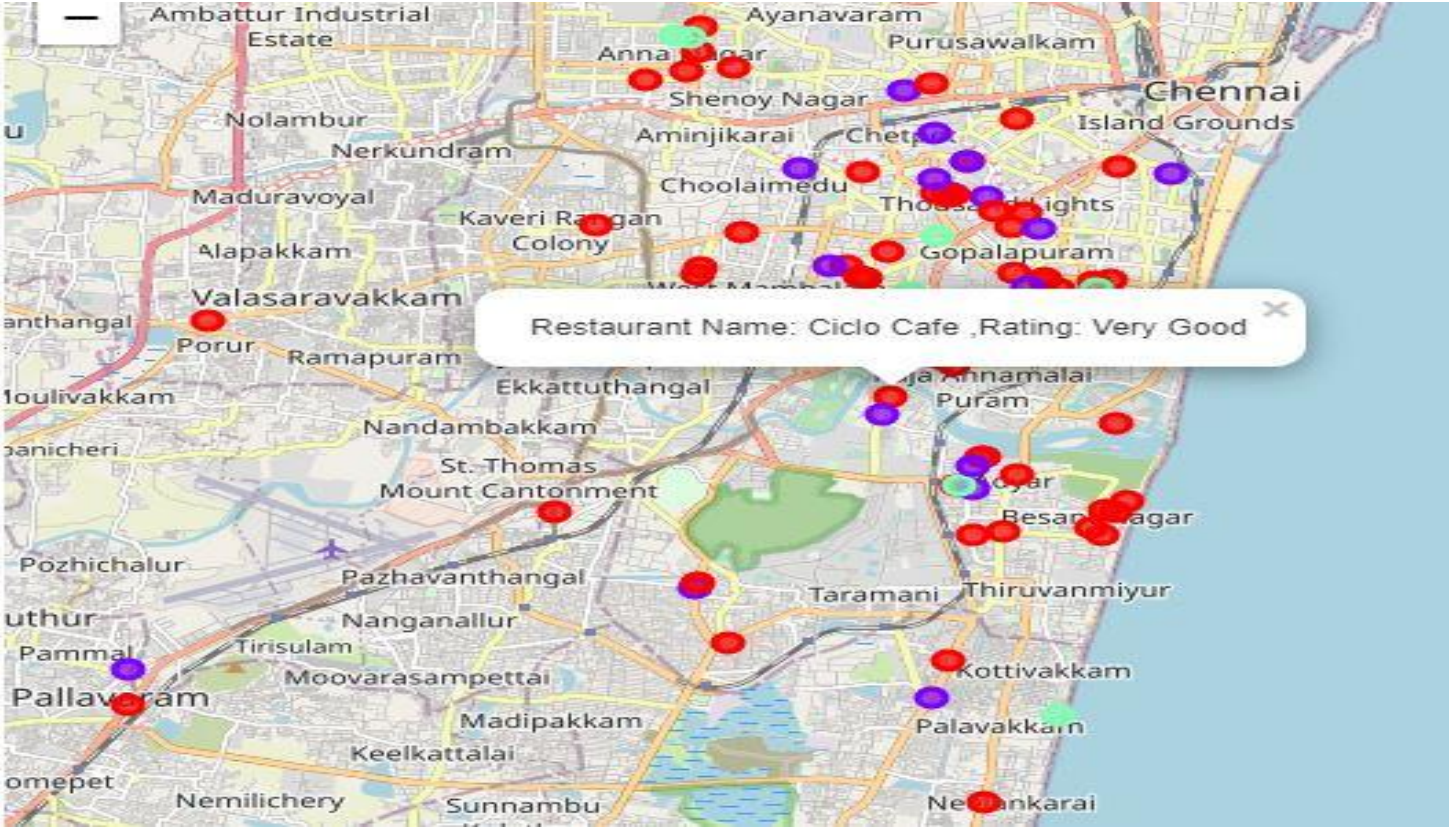
```
0    70
1    20
2     7
Name: Classifier, dtype: int64
```



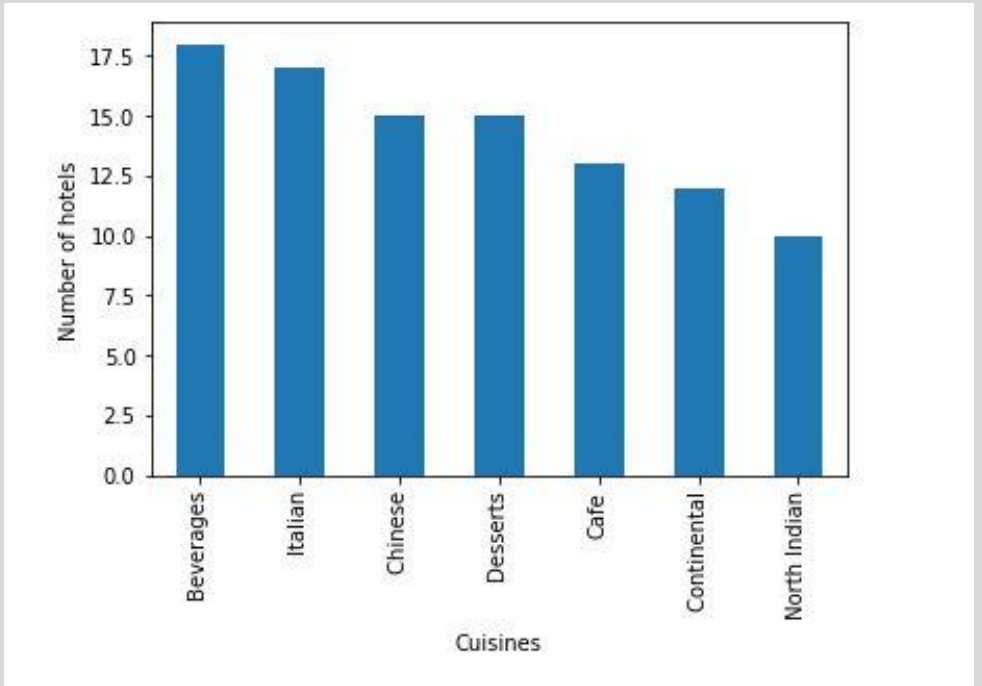
Display of map before clustering



Display of map after clustering



Most served Cuisines in Chennai



Thank you