

IBM DATA SCIENCE
APPLIED DATA SCIENCE CAPSTONE
APPLIED DATA SCIENCE CAPSTONE
REPORT
THE BATTLE OF NEIGHBORHOODS



EXPLORING VENUES IN LAHORE (PAKISTAN)



**EXPLORING VENUES IN LAHORE
(PAKISTAN)**

AMIN NADIM

OCOTBER, 2019

**IBM Data Science
Professional Certificate**



EXPLORING VENUES IN LAHORE (PAKISTAN)

INTRODUCTION OF LAHORE CITY



Lahore (/ləˈhɔːr/; Punjabi: لاہور Urdu: لاہور, pronounced [ləˈɦɔːr]) is the capital of the Pakistani province of Punjab. Lahore is the country's second-most populous city and is one of Pakistan's wealthiest cities, with an estimated GDP of \$58.14 billion (PPP) as of 2015. Lahore is the largest city, and historic cultural centre of the Punjab region, and one of Pakistan's most socially liberal, progressive and cosmopolitan cities.

Following independence in 1947, Lahore was declared capital of Pakistan's Punjab province.

Lahore exerts a strong cultural influence over Pakistan. Lahore is a major center for Pakistan's publishing industry, and remains the foremost center of Pakistan's literary scene. The city is also a major centre of education in Pakistan, with some of Pakistan's leading universities based in the city. Lahore is also home to Pakistan's film industry, Lollywood, and is a major centre of *Qawwali* music. The city also hosts much of Pakistan's tourist industry with major attractions including the Walled City, the famous Badshahi and Wazir Khan mosques and Sikh shrines. Lahore is also home to the Lahore Fort and Shalimar Gardens, both of which are UNESCO World Heritage Sites.

LHR_LATITUDE = '31.520370' LHR_LONGITUDE = '74.358749'

(Source: <https://en.wikipedia.org/wiki/Lahore>)

AIM

- To identify venues in Lahore, Pakistan based on their rating and average prices.
- We identify various venues in the city of Lahore, Pakistan using Foursquare API and Zomato API to help visitors select the restaurants that suit them the best.
- Helpful for User, whenever a user is visiting a city they start looking for places to visit during their stay.
- Primarily look for places based on the venue ratings across all venues and the average prices such that the locations fits in their budget.
- We will identify places that are fit for various individuals based on the information collected from the two APIs and Data Science.
- Once we have the plot with the venues, any company can launch an application using the same data and suggest users such information.

IMPLEMENTATION

EXPLORING VENUES in LAHORE ,PAKISTAN

LAHORE

Lahore is composed of a number of sectors spread across a total area of 404 sq Km. There are many venues (especially restaurants, hotels and cafes) which can be explored. 📍

We can use the geopy library to extract the latitude and longitude values of LAHORE but it seems off and thus, we Will directly supply the values in this case beacuse number of status not shown.

```
In [145]: LHR_LATITUDE = '31.528378'
LHR_LONGITUDE = '74.358749'
print('The geographical coordinates of Lahore are {}, {}'.format(LHR_LATITUDE, LHR_LONGITUDE))

The geographical coordinates of Lahore are 31.528378, 74.358749.

In [146]: !pip install folium==0.5.0

Requirement already satisfied: folium==0.5.0 in /opt/conda/envs/Python36/lib/python3.6/site-packages (0.5.0)
Requirement already satisfied: six in /opt/conda/envs/Python36/lib/python3.6/site-packages (from folium==0.5.0) (1.12.0)
Requirement already satisfied: branca in /opt/conda/envs/Python36/lib/python3.6/site-packages (from folium==0.5.0) (0.5.1)
Requirement already satisfied: Jinja2 in /opt/conda/envs/Python36/lib/python3.6/site-packages (from folium==0.5.0) (2.10)
Requirement already satisfied: requests in /opt/conda/envs/Python36/lib/python3.6/site-packages (from folium==0.5.0) (2.21.0)
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/envs/Python36/lib/python3.6/site-packages (from Jinja2->folium==0.5.0) (1.1.0)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python36/lib/python3.6/site-packages (from requests->folium==0.5.0) (2019.9.11)
Requirement already satisfied: chardet<3.1.0, >=3.0.2 in /opt/conda/envs/Python36/lib/python3.6/site-packages (from requests->folium==0.5.0) (3.0.4)
Requirement already satisfied: urllib3<1.25, >=1.21.1 in /opt/conda/envs/Python36/lib/python3.6/site-packages (from requests->folium==0.5.0) (1.24.1)
Requirement already satisfied: idna<2.9, >=2.5 in /opt/conda/envs/Python36/lib/python3.6/site-packages (from requests->folium==0.5.0) (2.8)

In [147]: import folium
lahore_map = folium.Map(location=[31.528378, 74.358749], zoom_start=13)
folium.Marker([31.528378, 74.358749]).add_to(lahore_map)
```

DATA COLLECTION FROM APIS

- Take a look at Lahore, Pakistan on the Map using the folium library.
- Fetch the data from two different APIs.

1. Foursquare API: We will use the Foursquare API to fetch venues in Chandigarh starting from the middle upto 44 Kilometers in each direction.

2. 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.

FOURSQUARE API IMPLEMENTATION

Foursquare API

By fetching a total of all venues in Chandigarh upto a range of 4 Kilometers using the Foursquare API.

```
In [148]: FOURSQUARE_CLIENT_ID = '9HHZ8WPF82F4LD3CL3SQMEFSKPCV083DS08OKP4HNUC70'
FOURSQUARE_CLIENT_SECRET = 'DS4XPFK1RAFCW8SUS3VCU0IA1IA2XVC0W1P1A3K3D1LB0QS'
RADIUS = 4000 # 4 km
NO_OF_VENUES = 100
VERSION = '20191012' # Current date
```

```
In [149]: def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']
```

```
In [150]: import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors

from pandas.io.json import json_normalize
import requests

pd.set_option('display.max_rows', None)

offset = 0
total_venues = 0
```

ZOMATO API IMPLEMENTATION

Zomato API

The Zomato API allows using its search API to search for any given venue based on certain search filters such as query, latitude, longitude and more

```
In [151]: headers = {'user-key': 'b3a1c26ed20422bfaae7ada559744e1f'}
venues_information = []

for index, row in foursquare_venues.iterrows():
    print("Fetching data for venue: {}".format(index + 1))
    venue = {}
    url = 'https://developers.zomato.com/api/v2.1/search?q=' + \
        '&start=0&count=1&lat={}&lon={}&sort=real_distance'.format(row['name'], row['lat'], row['lng'])
    result = requests.get(url, headers=headers).json()
    if (len(result['restaurants']) > 0):
        venue.append(result['restaurants'][0]['restaurant']['name'])
        venue.append(result['restaurants'][0]['restaurant']['location']['latitude'])
        venue.append(result['restaurants'][0]['restaurant']['location']['longitude'])
        venue.append(result['restaurants'][0]['restaurant']['average_cost_for_two'])
        venue.append(result['restaurants'][0]['restaurant']['price_range'])
        venue.append(result['restaurants'][0]['restaurant']['user_rating']['aggregate_rating'])
        venue.append(result['restaurants'][0]['restaurant']['location']['address'])
        venues_information.append(venue)
    else:
        venues_information.append(np.zeros(6))

zomato_venues = pd.DataFrame(venues_information,
                             columns=['venue', 'latitude',
                                       'longitude', 'price_for_two',
                                       'price_range', 'rating', 'address'])
```

```
Fetching data for venue: 1
Fetching data for venue: 2
Fetching data for venue: 3
Fetching data for venue: 4
Fetching data for venue: 5
Fetching data for venue: 6
```

DATA CLEANING

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

DATA CLEANING

Data Cleaning

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

```
In [152]: lahore_map = folium.Map(location = [31.528378, 74.358749], zoom_start = 13)

for name, latitude, longitude in zip(foursquare_venues['name'], foursquare_venues['lat'], foursquare_venues['lng']):
    label = '{}'.format(name)
    label = folium.Popup(label, parse_html = True)
    folium.CircleMarker(
        [latitude, longitude],
        radius = 5,
        popup = label,
        color = 'green',
        fill = True,
        fill_color = '#3186cc',
        fill_opacity = 0.7,
        parse_html = False).add_to(lahore_map)

lahore_map
```



METHODOLOGY

This project aims at identifying the venues in Lahore, Pakistan based on their rating and average costs.

- This would enable any visitor to identify the venues he/she wants to visit based on their rating and cost preference.
- As a first step, we retrieved the **data from two APIs (Foursquare and Zomato)**.
- We extract venue information from the center of Lahore, Pakistan up-to a distance of 4 Km. The latitude and longitude values are then used to fetch venue rating and price from Zomato.

METHODOLOGY

Methodology

This project aims at identifying the venues in *IAHORE* based on their rating and average costs. This would enable any visitor to identify the venues he/she wants to visit based on their rating and cost preference.

Analysis

The complete dataset is now in its final form.

We will inspect these venues based on their rating.

Categories

We have various types of venues in the final dataset.

```
In [173]: venue_distribution = selected_venues["categories"].value_counts()
          colors = cm.rainbow(np.linspace(0, 1, len(venue_distribution.index)))
          plt.figure(figsize = (20, 12))
          plt.xticks(rotation = 90)
          plt.xlabel("Venue category", fontsize = 16)
          plt.ylabel("Count", fontsize = 16)
          plt.title("Count of venues of each category", fontsize = 16)
          plt.bar(venue_distribution.index, venue_distribution.values, color = colors)
```

```
Out[173]: <BarContainer object of 0 artists>
```



ANALYSIS

- We will inspect these venues based on their rating.
 - The rating of a venue are based on user reviews and belongs to a range from 1 to 5.
 - We will also analyze the venues based on their price per person as well as the price range.
1. Categories
 2. Rating
 3. Price
 4. Clustering

ANALYSIS

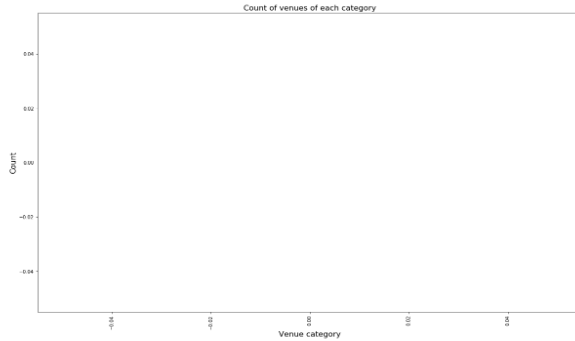
Categories

Categories

We have various types of venues in the final dataset.

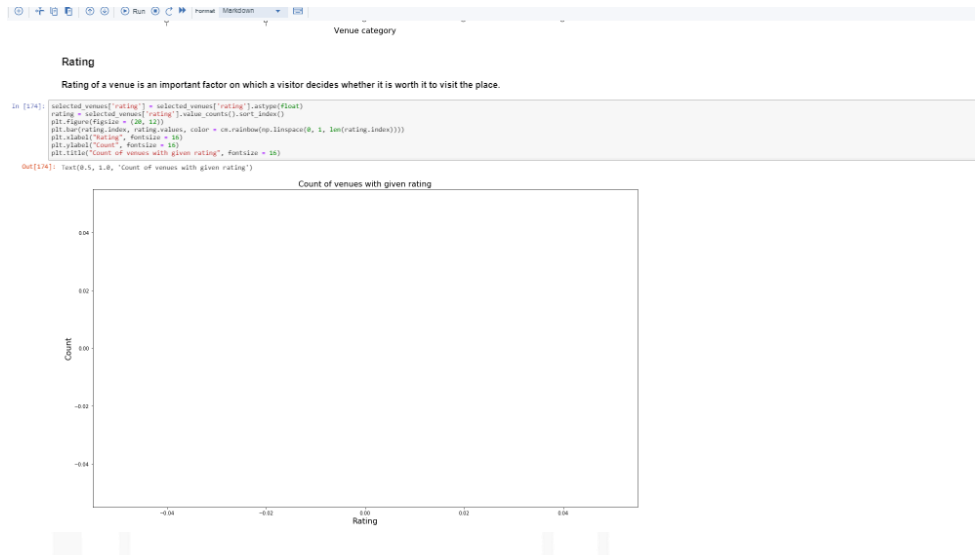
```
In [173]: vnames_distribution = selected_volumes['categories'].value_counts()
           colors = cm_rainbow(np.linspace(0, 1, len(vnames_distribution.index)))
           plt.figure(figsize=(20, 12))
           plt.xticks(rotation = 90)
           plt.xlabel('Volume category', fontsize = 16)
           plt.ylabel('Count', fontsize = 16)
           plt.title('Count of volumes of each category', fontsize = 16)
           plt.bar(vnames_distribution.index, vnames_distribution.values, color = colors)

Out[173]: <matplotlib.figure.Figure at 0xab710000>
```



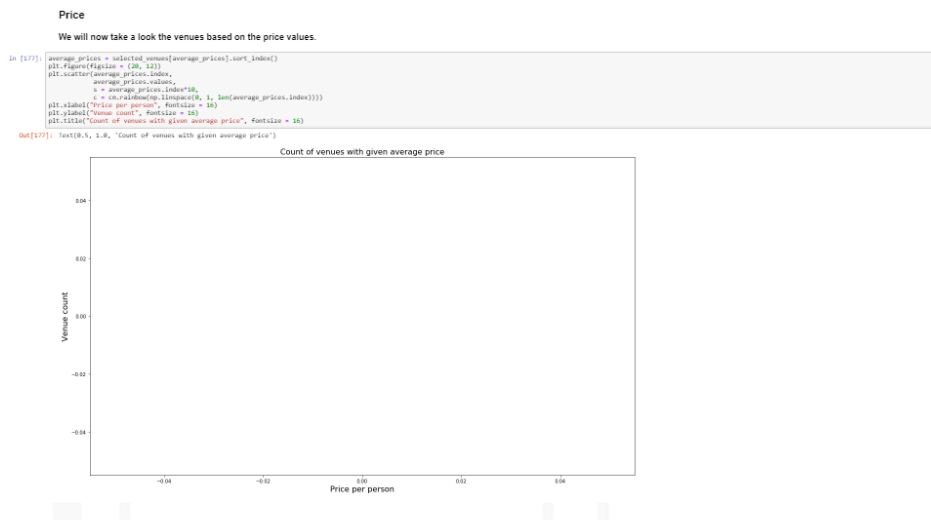
ANALYSIS

Rating



ANALYSIS

Price



CONCLUSION

- Purpose of this project was to explore the places that a person visiting Lahore, Pakistan could visit.
- The venues have been identified using Foursquare and Zomato API and have been plotted on the map.
- The map reveals that there are major areas a person can visit: Old Lahore Sector , New Lahore Sector & Industrial area.
- Based on the visitor's venue rating and price requirements, he/she can choose amongst the three places.

CONCLUSION

- Purpose of this project was to explore the places that a person visiting Lahore, Pakistan could visit.
- The venues have been identified using Foursquare and Zomato API and have been plotted on the map.
- The map reveals that there are major areas a person can visit: Old Lahore Sector , New Lahore Sector & Industrial area.
- Based on the visitor's venue rating and price requirements, he/she can choose amongst the three places.

RESULT

- Based on our analysis above, we can draw a number of conclusions that will be useful to aid any visitor visiting the city of Lahore, Pakistan.
- After collecting data from the Foursquare and Zomato APIs, we got a list of 120 different venues.
- However, not all venues from the two APIs were identical. Hence, we had to inspect their latitude and longitude values as well as names to combine them and remove all the outliers. This resulted in a total venue count of 167.

RECOMENDATAIONS

The following analysis can be improved with following extensions:

- In the Locality itself, it can also be computed the distance between all the venues in order to find a place with the most number of potential customers.
- Using smaller geographical areas like Neighborhoods could improve the accuracy for the scores.

FUTURE WORK

A company can use this information to build up an online website/mobile application, to provide users with up to date information about various venues in the city based on the search criteria (name, rating and price).



Thank you!

**EXPLORING VENUES IN LAHORE
(PAKISTAN)
AMIN NADIM**

OCTOBER 2019