Report

November 6, 2019

1 Introduction

1.1 Business Problem

Determining the most popular places in *Karachi, Pakistan* for incoming tourists and immigrants.

1.2 Stakeholders

- The State Government : for potential foreign currency inflow
- Tourism firms : to run their business and take the opportunity to build client relationships

2 Data

- **Area names** and their respective **postal codes** from *Karachi Metropolitan Corporation's* website:

```
[1]: import pandas as pd
     import numpy as np
     #!pip install lxml
     url = 'http://www.kmc.gos.pk/Contents.aspx?id=13'
     df = pd.read_html(url)
```

```
[2]: #By trial, error and inspection the most relevant table is retained:
     df = df[13]
     df.head()
```

```
[2]:                                             0            1
     0                                  POSTAL CODE          NaN
     1  AREA  POSTAL CODE  AIRPORT  72500  BALDIA TOWN…         NaN
     2                                         AREA  POSTAL CODE
     3                                      AIRPORT        72500
     4                                  BALDIA TOWN        75760
```

Note: This data frame will need to be cleaned and properly formatted.

- Manually form a data frame for geo-coordinates, by manually inputting the names on Google Maps and copying the coordinate pairs:
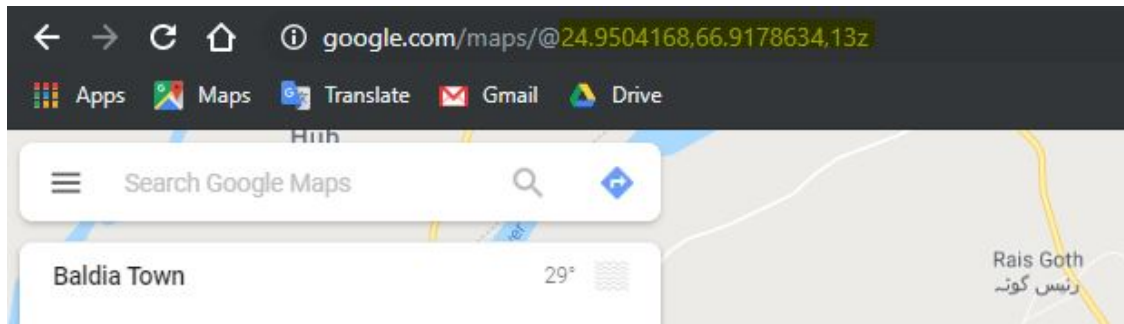
This *geo-data* frame should then be merged with `df`.

- **Foursquare data** to obtain popular places; according to user's feedback and corresponding to the *geo-data* obtained previously.