**Coursera Capstone Project**

**The Battle of Neighborhoods (Week 1)**

**Coursera Capstone - REPORT CONTENT**

1. Introduction Section : ⁃ Discussion of the business problem and the interested audience in this project.
2. Data Section:⁃ Description of the data that will be used to solve the problem and the sources.
3. Methodology section ⁃ Discussion and description of exploratory data analysis carried out, any inferential statistical testing performed, and if any machine learnings were used establishing the strategy and purposes.
4. Results section ⁃ Discussion of the results.
5. Discussion section ⁃ Elaboration and discussion on any observations noted and any recommendations suggested based on the results.
6. Conclusion section ⁃ Report Conclusion.

**1. Introduction Section :**

**Discussion of the business problem and the audience who would be interested in this project.**

**Description of the Problem and Background**

**Scenario:**

My name Bharat wanwari, currently live in Delhi, India. My family runs a cake bakery shop in delhi.

The cakes are of different shapes, taste, customised and prices. Also the material used in preparation is obtained from various part of india and are mostly in bulk.

**Business Problem:**

My Family wants to grow the business in india and then in other countries. The challenge is to find the starting point for the growth. where the population in other states is more as compared with Delhi is high and preferences of people are technically suitable to sweet/bread/cream products.

**Interested Audience**

I believe the audience will all those person those who wants to grow business in other regions and that Data analysis.

**2. Data Section:**

**Description of the data and its sources that will be used to solve the problem**

**Description of the Data:**

The following data is required to answer the issues of the problem:

1. List of states of India with census.

2. List of states of India with temperature.

3. List of preferences of people of India state wise.

**How the data will be used to solve the problem**

Use of Pandas to form dataframes

Use Foursquare and geopy data to map top 10 venues

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**from** **geopy.geocoders** **import** Nominatim

**import** **requests**

address\_Delhi = 'Delhi, India'

geolocator = Nominatim()

location = geolocator.geocode(address\_Delhi)

latitude = location.latitude

longitude = location.longitude

print('The geograpical coordinate of India home are **{}**, **{}**.'.format(latitude, longitude))

The geograpical coordinate of India home are 28.6517178, 77.2219388.

**Methodology**

**import** **pandas** **as** **pd**

url1 = 'https://raw.githubusercontent.com/BHARATW1993/mygit/master/data**%20c**apstone%20project1.csv'

data1 = pd.read\_csv(url1)

data1.head(10)

|  | **State** | **Population 2018** |
| --- | --- | --- |
| **0** | Uttar Pradesh | 228959599 |
| **1** | Maharashtra | 120837347 |
| **2** | Bihar | 119461013 |
| **3** | West Bengal | 97694960 |
| **4** | Madhya Pradesh | 82342793 |
| **5** | Rajasthan | 78230816 |
| **6** | Tamil Nadu | 76481545 |
| **7** | Karnataka | 66165886 |
| **8** | Gujarat | 63907200 |
| **9** | Andhra Pradesh | 52883163 |

url2 = 'https://raw.githubusercontent.com/BHARATW1993/mygit/master/data**%20c**apstone%20project2.csv'

data2 = pd.read\_csv(url2)

data2.head(10)

|  | **State** | **Place** | **High °C** | **Low °C** |
| --- | --- | --- | --- | --- |
| **0** | Uttar Pradesh | Agra | 33 | 19 |
| **1** | Uttar Pradesh | Allahabad | 32 | 19 |
| **2** | Punjab | Amritsar | 30 | 15 |
| **3** | Madhya Pradesh | Bhopal | 32 | 19 |
| **4** | Chandigarh | Chandigarh | 30 | 17 |
| **5** | Uttarakhand | Dehradun | 28 | 15 |
| **6** | Madhya Pradesh | Indore | 32 | 18 |
| **7** | Uttar Pradesh | Lucknow | 32 | 18 |
| **8** | Punjab | Ludhiana | 30 | 17 |
| **9** | Uttarakhand | Mukteswar | 19 | 8 |

url3 = 'https://raw.githubusercontent.com/BHARATW1993/mygit/master/data**%20c**apstone%20project3.csv'

data3 = pd.read\_csv(url3)

data3.head(10)

|  | **State** | **Preference** | **Bread Type** |
| --- | --- | --- | --- |
| **0** | Uttar Pradesh | Yes | Simple bread |
| **1** | Maharashtra | Yes | Bun bread |
| **2** | Bihar | Yes | Simple bread |
| **3** | West Bengal | Yes | Simple bread |
| **4** | Madhya Pradesh | Yes | Simple bread |
| **5** | Rajasthan | Yes | Simple bread |
| **6** | Tamil Nadu | Yes | Simple bread |
| **7** | Karnataka | Yes | Simple bread |
| **8** | Gujarat | Yes | Simple bread |
| **9** | Andhra Pradesh | Yes | Simple bread |

**we have to find the highest population with temperature between 20-30 having bread preferences.**

data1.nlargest(10, 'Population 2018')

|  | **State** | **Population 2018** |
| --- | --- | --- |
| **0** | Uttar Pradesh | 228959599 |
| **1** | Maharashtra | 120837347 |
| **2** | Bihar | 119461013 |
| **3** | West Bengal | 97694960 |
| **4** | Madhya Pradesh | 82342793 |
| **5** | Rajasthan | 78230816 |
| **6** | Tamil Nadu | 76481545 |
| **7** | Karnataka | 66165886 |
| **8** | Gujarat | 63907200 |
| **9** | Andhra Pradesh | 52883163 |

data2.nlargest(10, 'High °C')

|  | **State** | **Place** | **High °C** | **Low °C** |
| --- | --- | --- | --- | --- |
| **43** | Tamil Nadu | Madurai | 34 | 24 |
| **0** | Uttar Pradesh | Agra | 33 | 19 |
| **17** | Orissa (Odisha) | Bhuba­neshwar | 33 | 22 |
| **24** | Chhattisgarh | Raipur | 33 | 21 |
| **28** | Tamil Nadu | Chennai | 33 | 24 |
| **35** | Pondicherry | Puducherry | 33 | 24 |
| **38** | Andhra Pradesh | Vishak­hapatnam | 33 | 24 |
| **40** | Tamil Nadu | Chennai | 33 | 24 |
| **48** | Pondicherry | Puducherry | 33 | 24 |
| **51** | Andhra Pradesh | Vishak­hapatnam | 33 | 24 |

data2.nsmallest(10, 'High °C')

|  | **State** | **Place** | **High °C** | **Low °C** |
| --- | --- | --- | --- | --- |
| **12** | Himachal Pradesh | Shimla | 17 | 10 |
| **9** | Uttarakhand | Mukteswar | 19 | 8 |
| **13** | Jammu and Kashmir | Srinagar | 20 | 7 |
| **18** | Meghalaya | Cherrapunji | 21 | 14 |
| **26** | Meghalaya | Shillong | 21 | 12 |
| **22** | Nagaland | Kohima | 22 | 15 |
| **38** | Andhra Pradesh | Vijayawada | 23 | 23 |
| **51** | Andhra Pradesh | Vijayawada | 23 | 23 |
| **5** | Uttarakhand | Dehradun | 28 | 15 |
| **19** | Assam | Dibrugarh | 28 | 18 |

address\_Delhi = 'Uttar Pradesh, India'

geolocator = Nominatim()

location = geolocator.geocode(address\_Delhi)

latitude = location.latitude

longitude = location.longitude

print('The geograpical coordinate of India Uttar Pradesh are **{}**, **{}**.'.format(latitude, longitude))

The geograpical coordinate of India home are 27.1303344, 80.859666.

In [40]:

address\_Delhi = 'Tamilnadu, India'

geolocator = Nominatim()

location = geolocator.geocode(address\_Delhi)

latitude = location.latitude

longitude = location.longitude

print('The geograpical coordinate of India Tamil Nadu are **{}**, **{}**.'.format(latitude, longitude))

The geograpical coordinate of India Tamil Nadu are 10.9094334, 78.3665347.

**Results**

**Found that Tamilnadu and Uttarpradesh have identical environment have highest in ten population size, with bread preference 'yes'.**

**Discussion**

**The data is taken from Online sources just to represent the data science skills and cannot be used for any other useful purpose.**