
DATA SCIENCE
FINAL CAPSTONE PROJECT
BHARAT WANWARI

Introduction

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.

Data Required

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.

Methodology

Using the Below Libraries in python and using the Data required

1. Pandas
2. Geopy

The data required is stores as below in the Github repository.

1. https://raw.githubusercontent.com/BHARATW1993/mygit/master/data%20capstone%20project1.csv
2. https://raw.githubusercontent.com/BHARATW1993/mygit/master/data%20capstone%20project2.csv
3. https://raw.githubusercontent.com/BHARATW1993/mygit/master/data%20capstone%20project3.csv

Procedure and expectation

Importing 'Data1' in Pandas upto 10 data heads

In [37]:
```
import pandas as pd
url1 = 'https://raw.githubusercontent.com/BHARATW1993/mygit/master/data%20capstone%20project1.csv'

data1 = pd.read_csv(url1)

data1.head(10)
```

Out[37]:

	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

Procedure and expectation

Importing 'Data2' in Pandas upto 10 data heads

In [36]:
```
url2 = 'https://raw.githubusercontent.com/BHARATW1993/mygit/master/data%20capstone%20project2.csv'

data2 = pd.read_csv(url2)

data2.head(10)
```

Out[36]:

	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

Procedure and expectation

Importing 'Data1' in Pandas upto 10 data heads.

In [35]:
```
url3 = 'https://raw.githubusercontent.com/BHARATW1993/mygit/master/data%20capstone%20project3.csv'

data3 = pd.read_csv(url3)

data3.head(10)
```

Out[35]:

	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

Procedure and expectation

Finding Largest in Data set 1.

```
In [32]: data1.nlargest(10, 'Population 2018')
```

Out[32]:

	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

Procedure and expectation

Finding Largest and lowest in data 2 in High Degree Celsius.

```
In [38]: data2.nlargest(10, 'High °C')
```

Out[38]:

	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

```
In [33]: data2.nsmallest(10, 'High °C')
```

Out[33]:

	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

Procedure and expectation

Finding Geo Location using the Geopy.geocode

In [39]:
```
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))
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/ipykernel/__main__.py:2: DeprecationWarning: Using No
minatim with the default "geopy/1.18.1" `user_agent` is strongly discouraged, as it violates Nominatim's T
oS https://operations.osmfoundation.org/policies/nominatim/ and may possibly cause 403 and 429 HTTP error
s. Please specify a custom `user_agent` with `Nominatim(user_agent="my-application")` or by overriding the
default `user_agent`: `geopy.geocoders.options.default_user_agent = "my-application"`. In geopy 2.0 this w
ill become an exception.
  from ipykernel import kernelapp as app
```

```
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.
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/ipykernel/__main__.py:2: DeprecationWarning: Using No
minatim with the default "geopy/1.18.1" `user_agent` is strongly discouraged, as it violates Nominatim's T
oS https://operations.osmfoundation.org/policies/nominatim/ and may possibly cause 403 and 429 HTTP error
s. Please specify a custom `user_agent` with `Nominatim(user_agent="my-application")` or by overriding the
default `user_agent`: `geopy.geocoders.options.default_user_agent = "my-application"`. In geopy 2.0 this w
ill become an exception.
  from ipykernel import kernelapp as app
```

Results

- The first data segment display top 10 highest states having largest population.
- The second data displayed the highest and lowest high degree Celsius states of India.
- After comparing the states it is found that there are two states which are common in both the list which are Tamil nadu, Uttarpradesh and Andhrapradesh.

```
In [32]: data1.nlargest(10, 'Population 2018')
```

Out[32]:

	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

```
In [38]: data2.nlargest(10, 'High °C')
```

Out[38]:

	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

```
In [33]: data2.nsmallest(10, 'High °C')
```

Out[33]:

	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

Thank you