

Ishika Prasad

[ip1262@rit.edu](mailto:ip1262@rit.edu) (<mailto:ip1262@rit.edu>)

## Import Libraries ¶

```
In [7]: ➤ import pandas as pd
from pandas import ExcelWriter
from pandas import ExcelFile
from matplotlib import pyplot as plt
import seaborn as sns
```

## Read Excel file

```
In [8]: ➤ df = pd.read_excel('places.xlsx', sheet_name='Sheet1')
```

## Print the data read from excel file

```
In [9]: ➤ print(df)
```

	Climate	HousingCost	HlthCare	Crime	Transp	Educ	Arts	Recreat
0	521	6200	237	923	4031	2757	996	1405
1	575	8138	1656	886	4883	2438	5564	2632
2	468	7339	618	970	2531	2560	237	859
3	476	7908	1431	610	6883	3399	4655	1617
4	659	8393	1853	1483	6558	3026	4496	2612
5	520	5819	640	727	2444	2972	334	1018
6	559	8288	621	514	2881	3144	2333	1117
7	537	6487	965	706	4975	2945	1487	1280
8	561	6191	432	399	4246	2778	256	1210
9	609	6546	669	1073	4902	2852	1235	1109
10	885	16047	2025	983	3954	2843	5632	3156
11	195	12175	601	1223	5091	2414	2346	3000
12	530	5704	580	878	2865	2469	430	838
13	591	5725	820	975	2707	2772	169	613
14	546	11014	2508	1067	3433	3346	7559	2288
15	560	5530	598	1125	3051	2189	268	1165
16	396	7877	833	525	3298	2844	1166	2315
17	604	6733	1304	566	5086	3000	1301	1543

## Descriptive statistics for climate

```
In [10]: ▶ sum_climate = df['Climate'].sum()
print('Sum of climate =', sum_climate)
print('Mean of climate =', df['Climate'].mean())
print('Median of climate =', df['Climate'].median())
print('Mode of climate =', df['Climate'].mode())
print('Standard deviation of climate =', df['Climate'].std())
```

```
Sum of climate = 177243
Mean of climate = 538.7325227963526
Median of climate = 542.0
Mode of climate = 0      536
dtype: int64
Standard deviation of climate = 120.80825946472612
```

## Descriptive statistics for Housing

```
In [11]: ▶ sum_housing_cost = df['HousingCost'].sum()
print('Sum of housing cost =', sum_housing_cost)
print('Mean of housing cost =', df['HousingCost'].mean())
print('Medidan of housing cost =', df['HousingCost'].median())
print('Mode of housing cost =', df['HousingCost'].mode())
print('Standard deviation of housing cost =', df['HousingCost'].std())
```

```
Sum of housing cost = 2746018
Mean of housing cost = 8346.559270516716
Medidan of housing cost = 7877.0
Mode of housing cost = 0      6697
1      6760
2      7078
3      7143
4      7767
5      7778
6      8083
7      8263
8      8310
9     11652
dtype: int64
Standard deviation of housing cost = 2385.2626223806665
```

## Descriptive statistics for Health Care

```
In [12]: ▶ sum_health_care = df['HlthCare'].sum()
print('Sum of health care =', sum_health_care)
print('Mean of health care =', df['HlthCare'].mean())
print('Median of health care =', df['HlthCare'].median())
print('Mode of health care =', df['HlthCare'].mode())
print('Standard deviation of health care =', df['HlthCare'].std())
```

```
Sum of health care = 390108
Mean of health care = 1185.7386018237082
Median of health care = 833.0
Mode of health care = 0      593
dtype: int64
Standard deviation of health care = 1003.0020358412526
```

## Descriptive statistics for Crime

```
In [13]: ▶ sum_crime = df['Crime'].sum()
print('Sum of crime =', sum_crime)
print('Mean of crime =', df['Crime'].mean())
print('Median of crime =', df['Crime'].median())
print('Mode of crime =', df['Crime'].mode())
print('Standard deviation of crime =', df['Crime'].std())
```

```
Sum of crime = 316187
Mean of crime = 961.0547112462006
Median of crime = 947.0
Mode of crime = 0      413
1      633
2      892
dtype: int64
Standard deviation of crime = 357.15418638343317
```

## Descriptive statistics for Transport

```
In [14]: ▶ sum_transport = df['Transp'].sum()
print('Sum of transport =', sum_transport)
print('Mean of transport =', df['Transp'].mean())
print('Median of transport =', df['Transp'].median())
print('Mode of transport =', df['Transp'].mode())
print('Standard deviation of transport =', df['Transp'].std())
```

```
Sum of transport = 1385117
Mean of transport = 4210.082066869301
Median of transport = 4080.0
Mode of transport = 0      2989
1      3092
2      3197
3      3496
4      4186
5      4399
dtype: int64
Standard deviation of transport = 1451.1792395557711
```

## Descriptive statistics for Education

```
In [15]: ▶ sum_education = df['Educ'].sum()
print('Sum of education =', sum_education)
print('Mean of education =', df['Educ'].mean())
print('Median of education =', df['Educ'].median())
print('Mode of education =', df['Educ'].mode())
print('Standard deviation of education =', df['Educ'].std())
```

```
Sum of education = 926098
Mean of education = 2814.887537993921
Median of education = 2794.0
Mode of education = 0      2772
1      3029
dtype: int64
Standard deviation of education = 320.79295255865276
```

## Descriptive statistics for Arts

```
In [16]: ▶ sum_arts = df['Arts'].sum()
print('Sum of arts =', sum_arts)
print('Mean of arts =', df['Arts'].mean())
print('Median of arts =', df['Arts'].median())
print('Mode of arts =', df['Arts'].mode())
print('Standard deviation of arts =', df['Arts'].std())
```

```
Sum of arts = 1036641
Mean of arts = 3150.884498480243
Median of arts = 1871.0
Mode of arts = 0      87
1      268
2      334
3      373
4      529
5      904
6      996
7     1921
8     2111
9     3596
10    7559
dtype: int64
Standard deviation of arts = 4642.283737956128
```

## Descriptive statistics for Recreation

```
In [17]: ▶ sum_recreation = df['Recreat'].sum()
print('Sum of recreation =', sum_recreation)
print('Mean of recreation =', df['Recreat'].mean())
print('Median of recreation =', df['Recreat'].median())
print('Mode of recreation =', df['Recreat'].mode())
print('Standard deviation of recreation =', df['Recreat'].std())
```

```
Sum of recreation = 607320
Mean of recreation = 1845.9574468085107
Median of recreation = 1670.0
Mode of recreation = 0      3000
1      3300
dtype: int64
Standard deviation of recreation = 807.888171075795
```

## Descriptive statistics for Economics

```
In [18]: ▶ sum_economics = df['Econ'].sum()
print('Sum of economics =', sum_economics)
print('Mean of economics =', df['Econ'].mean())
print('Median of economics =', df['Econ'].median())
print('Mode of economics =', df['Econ'].mode())
print('Standard deviation of economics =', df['Econ'].std())
```

Sum of economics = 1817845

Mean of economics = 5525.364741641338

Median of economics = 5384.0

Mode of economics = 0      4247

1      4491

2      4631

3      5097

4      5154

5      5165

6      5187

7      5204

8      5537

9      5594

10     5656

11     5671

12     5739

13     6019

14     6307

15     6527

16     7060

dtype: int64

Standard deviation of economics = 1084.4685225095445

## Descriptive statistics for Case Number

```
In [19]: ▶ sum_case_number = df['CaseNum'].sum()
print('Sum of case number =', sum_case_number)
print('Mean of case number =', df['CaseNum'].mean())
print('Median of case number =', df['CaseNum'].median())
print('Mode of case number =', df['CaseNum'].mode())
print('Standard deviation of case number =', df['CaseNum'].std())
```

Sum of case number = 54285

Mean of case number = 165.0

Median of case number = 165.0

Mode of case number = 0                      1

1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10
10	11
11	12
12	13
13	14
14	15
15	16
16	17
17	18
18	19
19	20
20	21
21	22
22	23
23	24
24	25
25	26
26	27
27	28
28	29
29	30

...

299	300
300	301
301	302
302	303
303	304
304	305
305	306
306	307
307	308
308	309
309	310
310	311
311	312
312	313
313	314

```
314    315
315    316
316    317
317    318
318    319
319    320
320    321
321    322
322    323
323    324
324    325
325    326
326    327
327    328
328    329
Length: 329, dtype: int64
Standard deviation of case number = 95.11834733635777
```

## Descriptive statistics for Longitude

```
In [20]: ▶ sum_long = df['Long'].sum()
print('Sum of long =', sum_long)
print('Mean of long =', df['Long'].mean())
print('Median of long =', df['Long'].median())
print('Mode of long =', df['Long'].mode())
print('Standard deviation of long =', df['Long'].std())
```

```
Sum of long = -29668.74697875977
Mean of long = -90.17856224547042
Median of long = -86.806999206543
Mode of long = 0    -80.218002
1    -71.461998
dtype: float64
Standard deviation of long = 14.625079713369603
```

## Descriptive statistics for Latitude

```
In [21]: ▶ sum_lat = df['Lat'].sum()
print('Sum of lat =', sum_lat)
print('Mean of lat =', df['Lat'].mean())
print('Median of lat =', df['Lat'].median())
print('Mode of lat =', df['Lat'].mode())
print('Standard deviation of lat =', df['Lat'].std())
```

```
Sum of lat = 12561.673992156982
Mean of lat = 38.18137991537077
Median of lat = 39.6450004577637
Mode of lat = 0    34.237000
1    40.693001
dtype: float64
Standard deviation of lat = 5.123344468052448
```



## Descriptive statistics for Population

```
In [22]: ▶ sum_population = df['Pop'].sum()
print('Sum of population =', sum_population)
print('Mean of population =', df['Pop'].mean())
print('Median of population =', df['Pop'].median())
print('Mode of population =', df['Pop'].mode())
print('Standard deviation of population =', df['Pop'].std())
```

```
Sum of population = 171776970
Mean of population = 522118.4498480243
Median of population = 241617.0
Mode of population = 0      62820
1      66100
2      67640
3      68807
4      71856
5      73762
6      79988
7      80696
8      81582
9      81717
10     82636
11     83490
12     83919
13     84690
14     84784
15     85949
16     87000
```

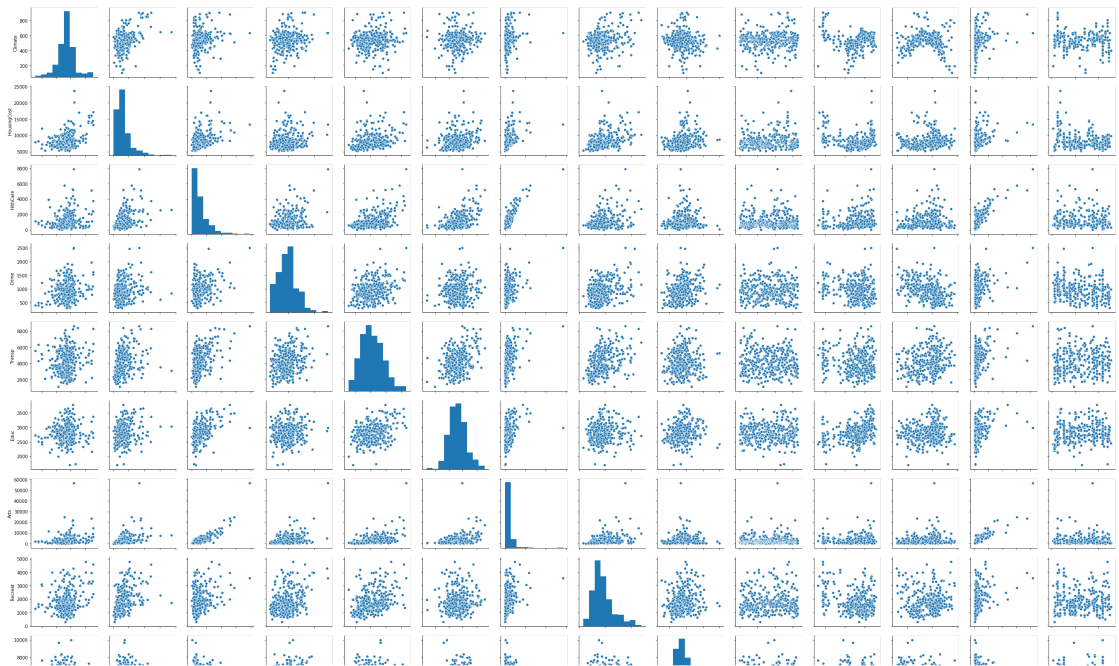
## Descriptive statistics for Street Number

```
In [23]: ▶ sum_stnum = df['StNum'].sum()
print('Sum of stnum =', sum_stnum)
print('Mean of stnum =', df['StNum'].mean())
print('Median of stnum =', df['StNum'].median())
print('Mode of stnum =', df['StNum'].mode())
print('Standard deviation of stnum =', df['StNum'].std())
```

```
Sum of stnum = 8434
Mean of stnum = 25.635258358662615
Median of stnum = 25.0
Mode of stnum = 0      44
dtype: int64
Standard deviation of stnum = 14.93627020685531
```

## Pairwise comparison plots

```
In [24]: sns.pairplot(df)  
plt.show()
```



## Conclusion

Comparison between attributes:

- 1) Street Number vs Housing cost: Based on the street Number, how the housing cost varies. If the street number is between 0-10 and then housing cost goes up to 25000.
- 2) Population vs Crime: Based on the population, how the crime rate changes. If the population is 0-2000000, then crime goes till 2000.
- 3) Lat vs Climate: Based on latitude, how the climate changes.
- 4) Long vs Climate: Based on longitude, how the climate changes.

```
In [ ]:
```