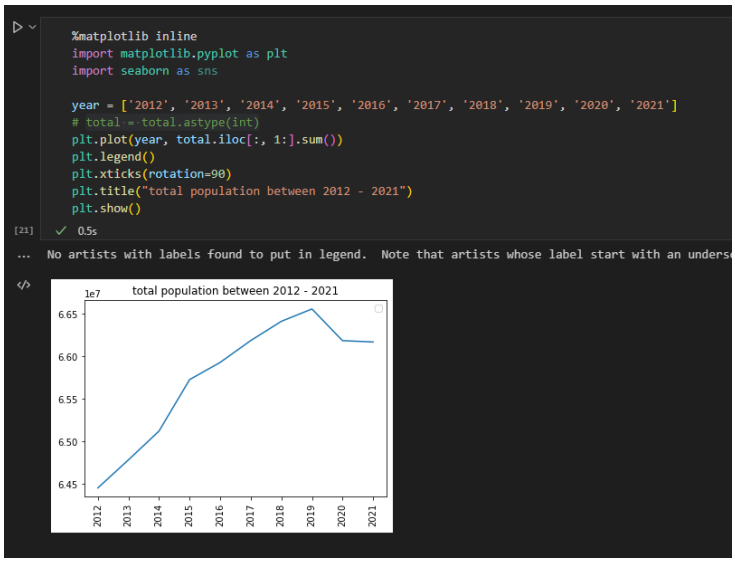
UMAP-MicroCredential: Data Visualization

CapStone Project

At the start of this project, I start by asking questions to myself. And The question is " Why isn't my country is growing too slow " and I search for the information about this and found some growth indicators of each country It's called " GDP " stands for Gross Domestic Product but GDP is used to see that country is growing? But I know my GDP country in 2019 is 544.3 billion USD(Most of the latest data in Thailand is 2019). So what about GPP? GPP is like GDP but calculated by Province instead. Then I try to search about GPP and Find the data and table but the latest data is 2019. Then I use the data between 2012 - 2019 to visualize it. I use a number of population from 2012 to 2019 and this is a graph.

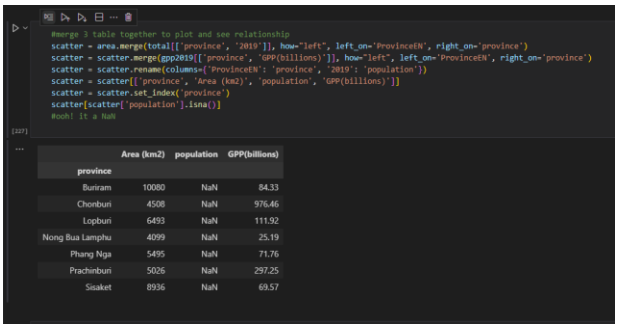


This graph shows after 2019 population In Thailand is decreased.

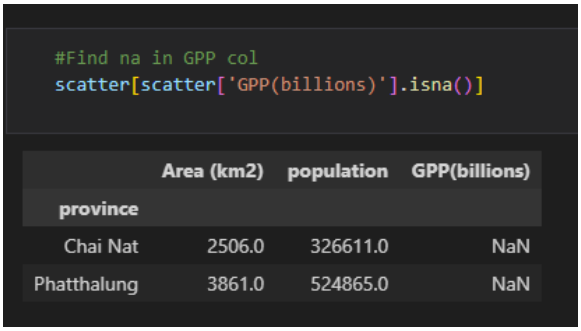
And Why? Maybe from my opinion maybe it because the government management is too bad that make people want to work in other countries like USA or Singapore.

After this we'll focus on the year 2019 I finding each of the areas by province and using population per province from the last table and GPP in 2019 to find the relationship and hope it shows something that I can use to answer my question.

Before I visualize them. I found the NaN in some cells like below picture



province	Area (km2)	population	GPP(billions)
Buriram	10080	NaN	84.33
Chonburi	4508	NaN	976.46
Lopburi	6493	NaN	111.92
Nong Bua Lamphu	4099	NaN	25.19
Phang Nga	5495	NaN	71.76
Prachinburi	5026	NaN	297.25
Sisaket	8936	NaN	69.57



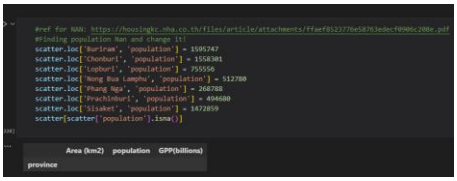
```
#Find na in GPP col
scatter[scatter['GPP(billions)'].isna()]
```

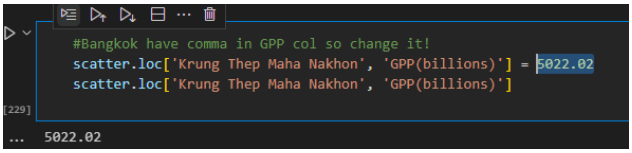
province	Area (km2)	population	GPP(billions)
Chai Nat	2506.0	326611.0	NaN
Phatthalung	3861.0	524865.0	NaN

Oh! not only NaN value but “Krung Thep Maha Nakon” GPP columns is a string so we can’t plot them with string format

```
1   Province,Population  (millions),GPP (billions ?),% of national GDP,GPP (billions US$ nc
2   Krung Thep Maha Nakhon,8.912,"5,022.02",31.88,173.76,434.4,"573,907","19,749","55,297"
3   Samut Prakan,2.171,717.05,4.71,25.69,64.22,"343,215","12,176","34,092"
4   Pathum Thani,1.729,380.69,2.39,13.02,32.55,"254,627","8,039","22,509"
5   Samut Sakhon,1.042,398.10,2.40,13.10,32.75,"411,326","12,914","36,159"
```

Now I solve all NaN, Na, and wrong format data by this one.





```
#Bangkok have comma in GPP col so change it!
scatter.loc['Krung Thep Maha Nakhon', 'GPP(billions)'] = 5022.02
scatter.loc['Krung Thep Maha Nakhon', 'GPP(billions)']
```

```
[229]
...  5022.02
```

```
#Change it!
scatter.loc['Chai Nat', 'GPP(billions)'] = 31.85
scatter.loc['Phatthalung', 'GPP(billions)'] = 36.48
scatter[scatter['GPP(billions)'].isna()]
```

```
[241]
...
```

province	Area (km2)	population	GPP(billions)

Now I got one table that I need to visualize!

```
scatter = scatter.sort_values('GPP(billions)')
topGPP = scatter.nlargest(11, 'GPP(billions)', keep='all')
bottomGPP = scatter.nsmallest(11, 'GPP(billions)', keep='all')
scatt = pd.concat([topGPP, bottomGPP])
scatt
```

province	Area (km2)	population	GPP(billions)
Krung Thep Maha Nakhon	1564.0	5666264.0	5022.02
Rayong	3666.0	734753.0	984.98
Chonburi	4508.0	1558301.0	976.46
Samut Prakan	947.0	1344875.0	717.05
Phra Nakhon Si Ayutthaya	2548.0	820188.0	403.60
Samut Sakhon	866.0	584703.0	398.10
Pathum Thani	1520.0	1163604.0	380.69
Chachoengsao	5169.0	720113.0	341.12
Nakhon Pathom	2142.0	920030.0	332.63
Nonthaburi	637.0	1265387.0	316.63
Prachinburi	5026.0	494680.0	297.25
Mae Hong Son	12765.0	284138.0	13.00
Amnat Charoen	3290.0	378438.0	17.65
Samut Songkhram	414.0	193305.0	21.88
Nong Bua Lamphu	4099.0	512780.0	25.19
Mukdahan	4126.0	353174.0	25.80
Yasothon	4131.0	537299.0	26.04
Sing Buri	817.0	208446.0	26.50
Ranong	3230.0	193370.0	26.77
Nakhon Nayok	2141.0	260751.0	26.84
Bueng Kan	4003.0	424091.0	27.17
Ang Thong	950.0	279654.0	27.79

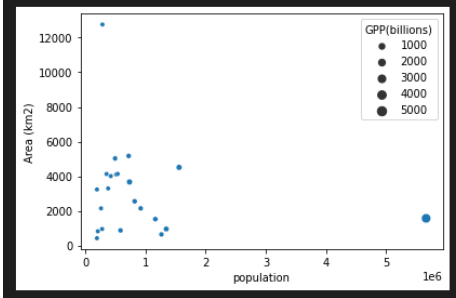
I choose the top 11 and bottom 11 by GPP sorting to visualize and the graph is the picture below.(don’t use Krung Thep Maha Nakon)

```
# plt.scatter(scatt['population'], scatt['Area (km2)'], s=scatt['GPP(billions)'])
# plt.xlabel("population 100k")
# plt.ylabel("Area (km square)")
# plt.title('Relactionship between area an poppulation by province(circle size = GPP)')
# plt.show()
sns.scatterplot(data=scatt, x='population', y='Area (km2)', size='GPP(billions)')

#Mae Hong Son very wide but low pop and low GPP
```

✓ 0.3s

```
<AxesSubplot:xlabel='population', ylabel='Area (km2)'>
```



And Finally, I found some point of the graph that has many areas but low GPP and population what is that province?

I find by sorting the table by area and then get top 11 and get the smallest GDP and that it.

```
popsort = scatter.sort_values('Area (km2)')
popsort.nlargest(11,columns=['Area (km2)'], keep='all').nsmallest(11, columns=['GPP(billions)'], keep='all')

# Why Mae Hong Son that the high populatin and high-mid area is low GPP

#ans: because the terrain of Mae Hong Son Province is mountainous
#top 5 by most area and then the most area have low GPP is North region(many of high motntain and a conservation area)
#but Phetchabun is Central!
```

province	Area (km2)	population	GPP(billions)
Mae Hong Son	12765.0	284138.0	13.00
Nan	12130.0	478227.0	31.31
Tak	17303.0	665620.0	47.80
Chaiyaphum	12698.0	1137357.0	60.09
Lampang	12488.0	738316.0	68.20
Phetchabun	12340.0	992451.0	76.80
Kanchanaburi	19385.0	895525.0	97.29
Ubon Ratchathani	15626.0	1878146.0	120.49
Surat Thani	13079.0	1068010.0	211.05
Chiang Mai	22135.0	1779254.0	231.73
Nakhon Ratchasima	20736.0	2648927.0	274.90

And I visualize data as table the top 10 Area in 77 province

```
scatter.sort_values('Area (km2)', ascending=False).head(10)
```

[63] ✓ 0.4s

province	Area (km2)	population	GPP(billions)
Chiang Mai	22135.0	1779254.0	231.73
Nakhon Ratchasima	20736.0	2648927.0	274.90
Kanchanaburi	19385.0	895525.0	97.29
Tak	17303.0	665620.0	47.80
Ubon Ratchathani	15626.0	1878146.0	120.49
Surat Thani	13079.0	1068010.0	211.05
Mae Hong Son	12765.0	284138.0	13.00
Chaiyaphum	12698.0	1137357.0	60.09
Lampang	12488.0	738316.0	68.20
Phetchabun	12340.0	992451.0	76.80

This table has some province in the previous table

Summary

Thailand in the northern region has many high terrains. We can't build the building like an industry in high terrain because all of the areas are conservation zone. But we can use this area as a tourist Attraction maybe it makes the province in the North region. and another reason still be the government management. And maybe Covid-19 make the GDP and DPP decreased this table has some provinces in the previous table