



Selecting suitable area in Bangkok Thailand

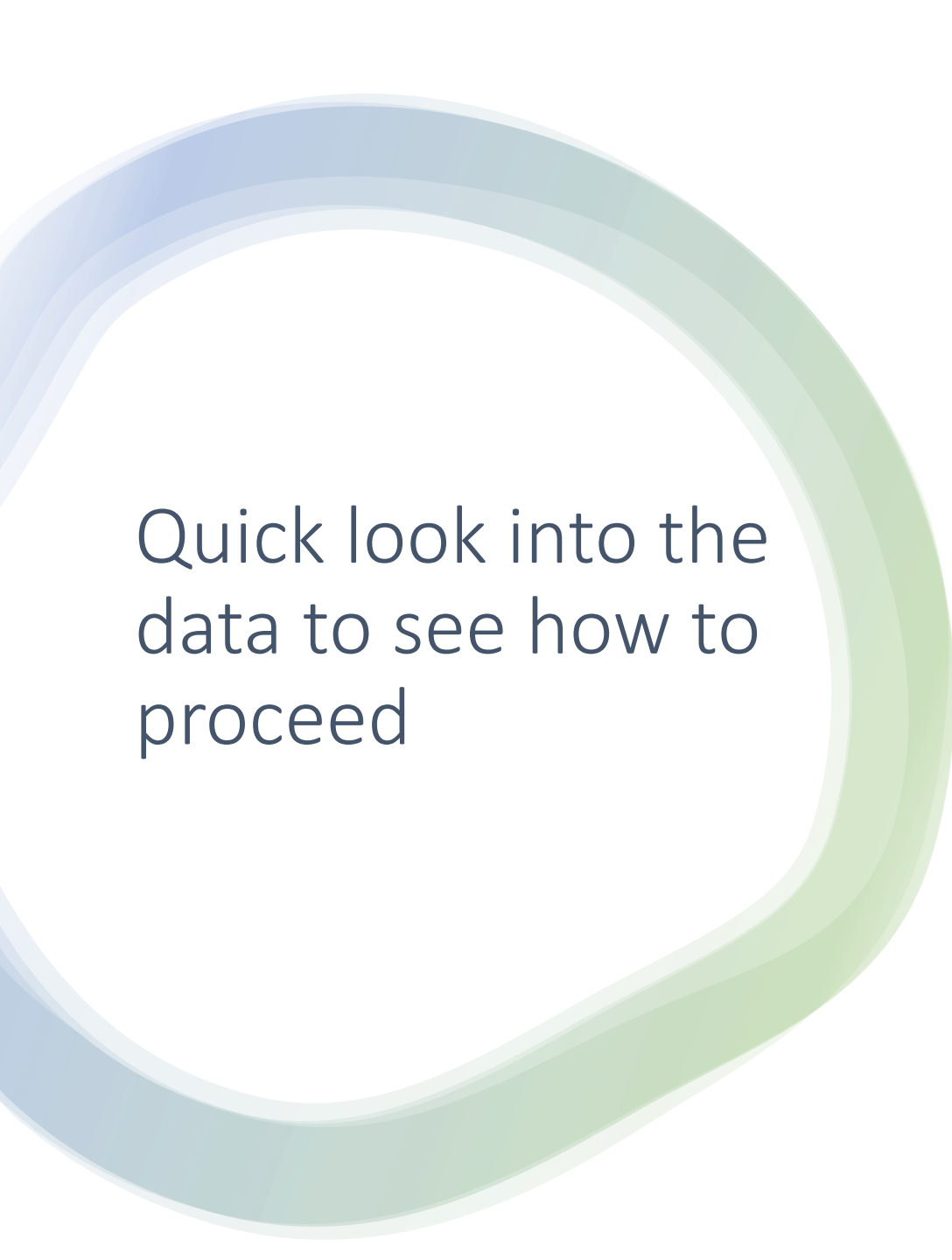
May 2021, 20

Selecting suitable area in Bangkok Thailand

- Generally, moving to a new place is tiresome from a lot of factors. Not only you are not familiar with the culture, but you also don't have any idea about the environment around there.
- In this project, I will help people who are looking for a rental properties in Bangkok for the best suitable location.
- They will be able to see the price per square meter so they can decide according to their budgets.
- They can decide whether to live in residential or commercial area.
- They will be able to see what is the average rental price in Bangkok, Thailand.

Data acquisition and cleaning

- The data was got from <https://www.thailand-property.com/properties-for-rent/bangkok?page=1>
- Using beautifulsoup library to scrap the data from the website.
- Coordinate location of each location in Bangkok from csv file which I got from some online website.
- foursquare data for the near by interesting places near by in each location
- I checked the data set for each location about the bed_rooms, bath_rooms, size, and price.
- After the removing the outlier from price and plot the scatter plot chart between rooms and price.

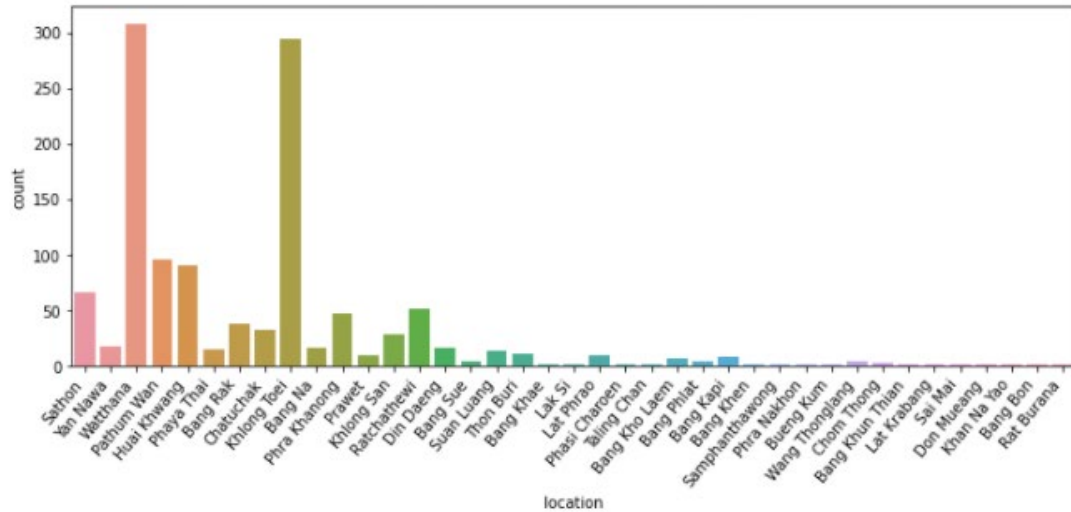
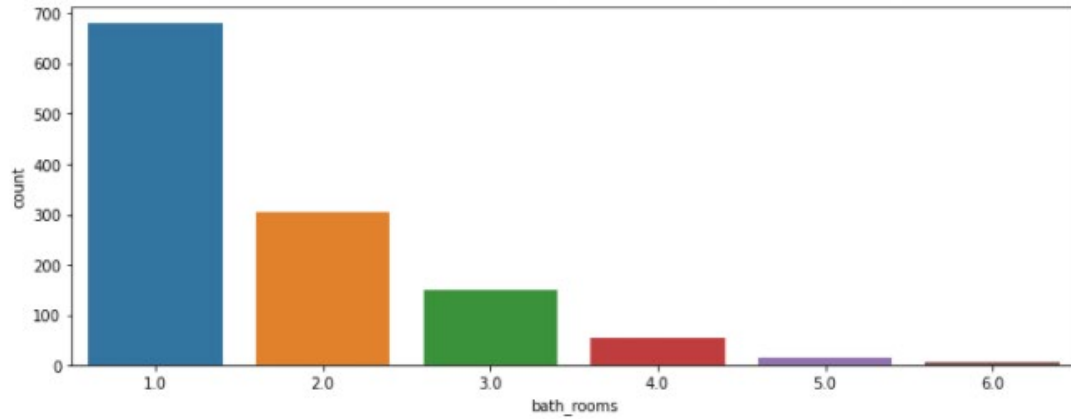
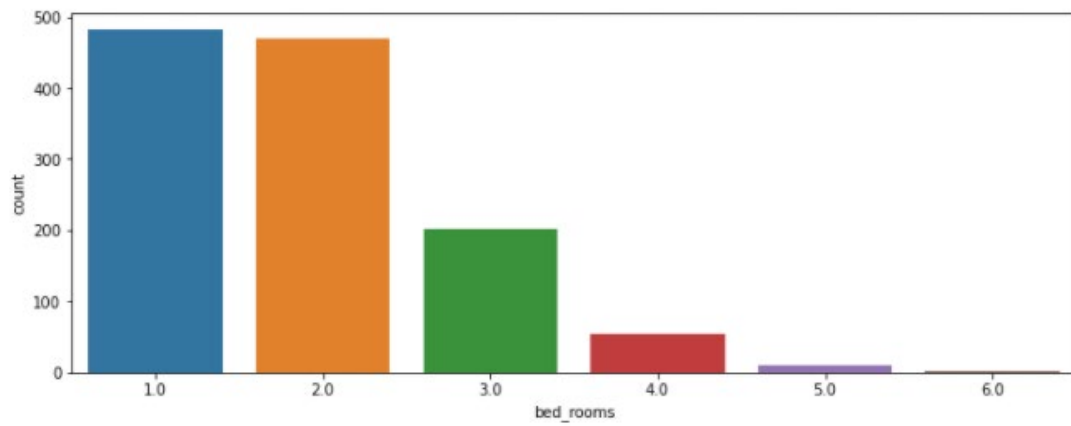


Quick look into the
data to see how to
proceed

```
In [10]: rental_properties.describe(include='all')
```

Out[10]:

	price	location	bed_rooms	bath_rooms	size
count	1226.000000	1226	1226.000000	1226.000000	1226.000000
unique	NaN	40	NaN	NaN	NaN
top	NaN	Watthana	NaN	NaN	NaN
freq	NaN	308	NaN	NaN	NaN
mean	51318.329527	NaN	1.909462	1.757749	104.020285
std	46417.247883	NaN	1.006489	1.104001	97.685080
min	2000.000000	NaN	1.000000	1.000000	1.000000
25%	22000.000000	NaN	1.000000	1.000000	43.000000
50%	36000.000000	NaN	2.000000	1.000000	70.000000
75%	65000.000000	NaN	2.000000	2.000000	120.000000
max	450000.000000	NaN	16.000000	11.000000	800.000000

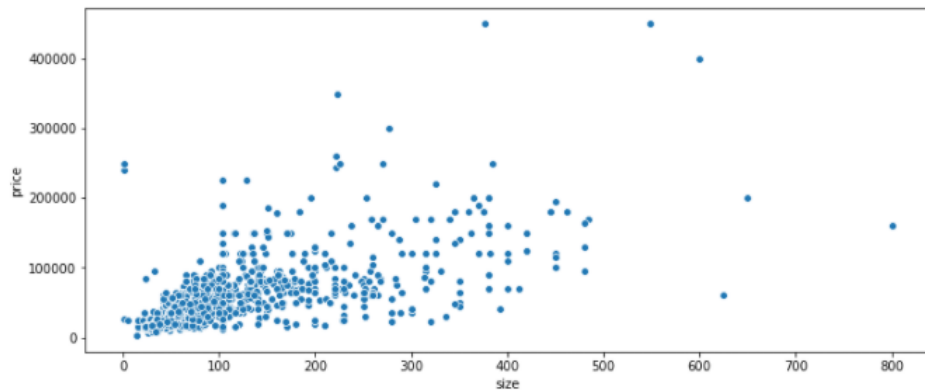


I checked the data set for each location about the bedrooms, bathrooms, size, and price.

Plot scatter plot between price and size of the room then remove outliers

```
In [17]: plt.figure(figsize=(12, 5))  
sns.scatterplot(x='size', y='price', data=rental_properties)
```

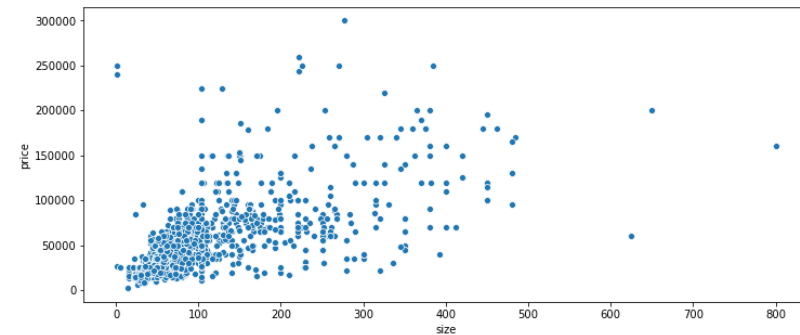
Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x7f72cf629e90>



```
In [19]: # remove outliers  
rental_properties = rental_properties[rental_properties['price'] <= 300000]
```

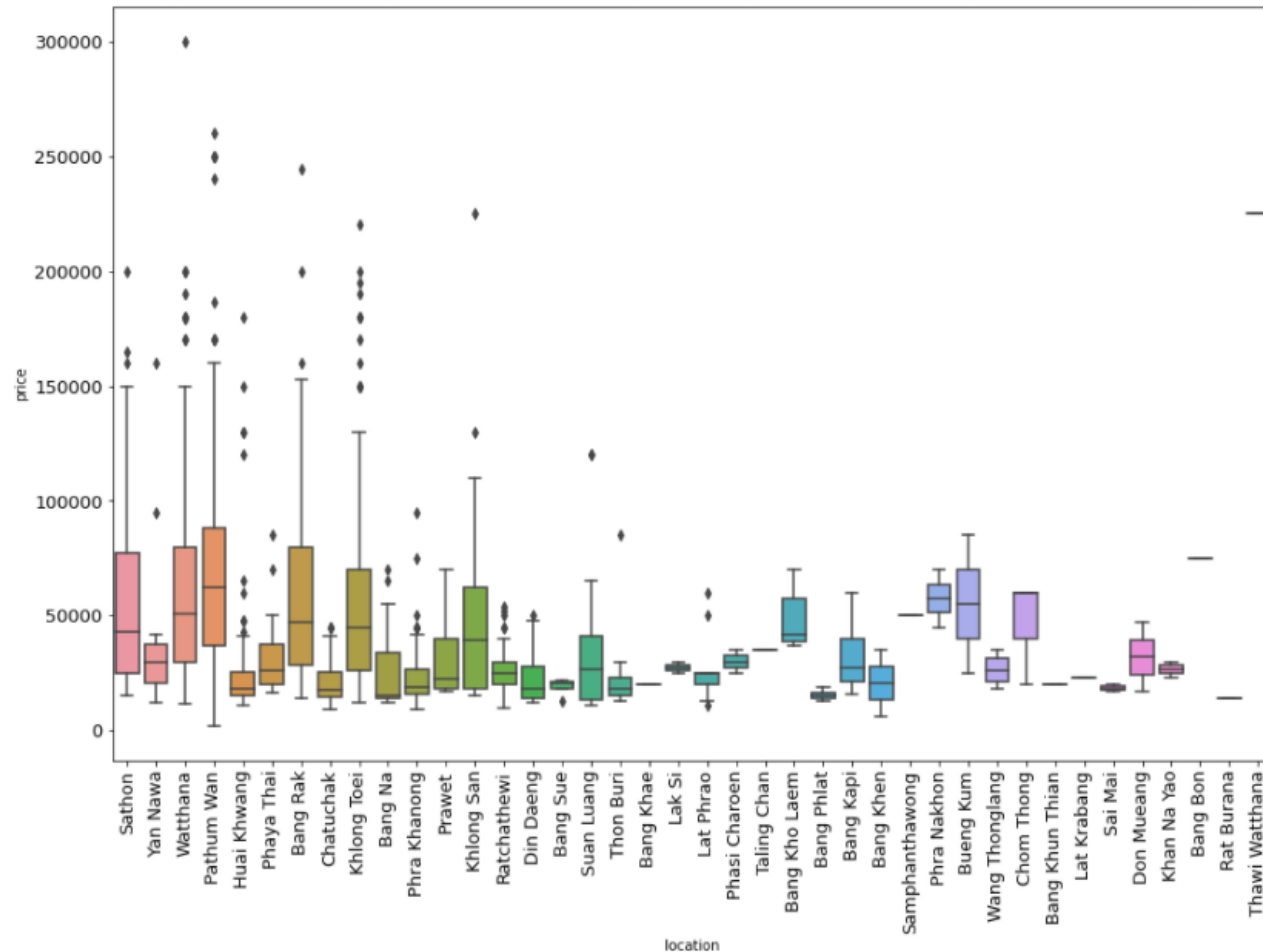
```
In [20]: plt.figure(figsize=(12, 5))  
sns.scatterplot(x='size', y='price', data=rental_properties)
```

Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x7f72cf051b90>

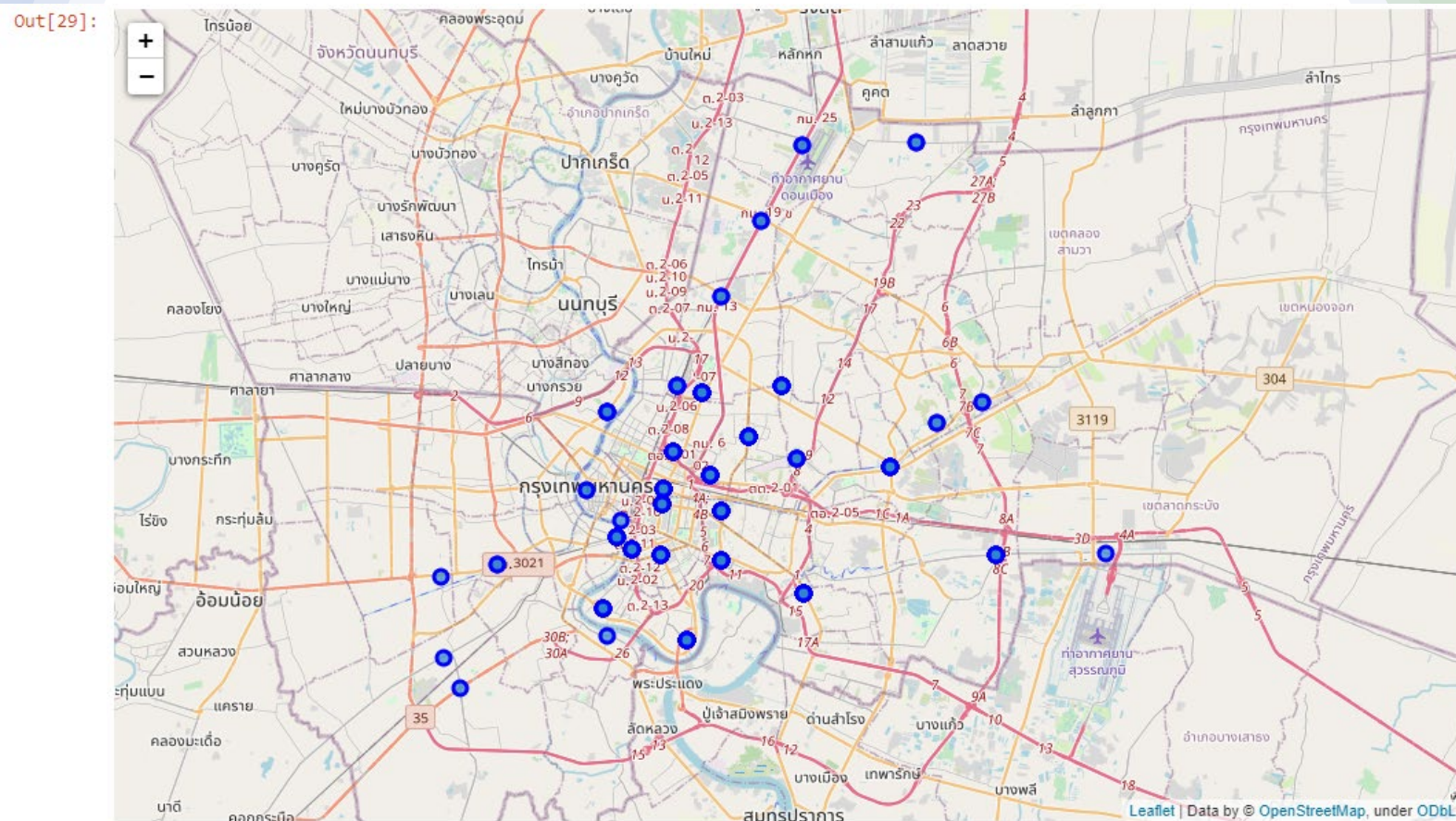


The areas which have the highest price are Patum Wan, Watthana, Bang Rak, and Sathorn

```
In [22]: plt.figure(figsize=(15,10))
ax = sns.boxplot(x='location', y='price', data=rental_properties)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90)
ax.tick_params(labels=13)
```



Then I combine coordinate location that I got from internet to the dataframe and I ran foursqaure to get the interesting place near by all the location



Next I ran the k-mean clustering model to separate the area into 2 areas - commercial and residential area

```
In [41]: # set number of clusters
kclusters = 2

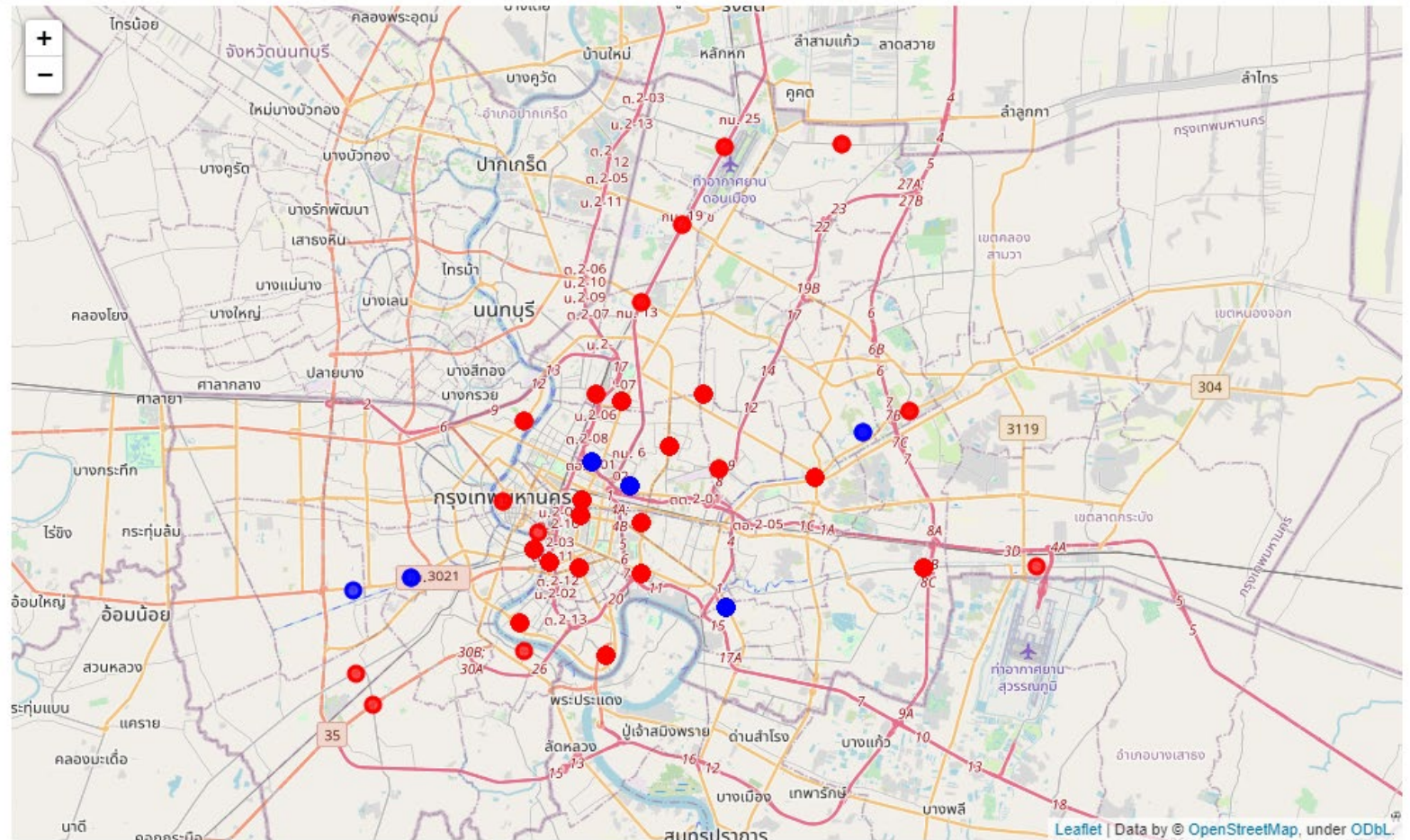
bkk_grouped_clustering = bkk_grouped.drop('location', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(bkk_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[41]: array([1, 1, 0, 1, 1, 1, 1, 1, 1, 1], dtype=int32)

Out[63]:



Conclusion

- Finally, based on this information the stakeholder will be able to see which area is most suitable for them in terms of price, interesting places, whether they will live in industrial or commercial areas.
- Based on the last picture, we can see that the red points are the commercial areas which have higher rental prices while the blue ones are residential areas which are less expensive.