

Bangkok, Thailand



1. Introducton

Bangkok, the capital city of Thailand, have various places to travel and also have very low cost of living. So, this city becomes popular among tourists. Then, this lures contractors who are trying to start their own business. When people want to start their own business, they need to explore the places and try to fetch as much information as possible regarding the city. Thus, it would be great for them if there is information which tell them about venues of each neighborhoods in this city.

This project will help find venues and classify them for in each district by using k-means clustering to cluster the venues based on the place category. So, business constructors can gain more insights about types porportion and number of businesses in Bangkok which can help them to determind which types of businesses they should do. Also, tourists can determind which district they should travel to for the place they prefer.

2.Data sets and APIs

2.1)Data

Beautiful soup web scraping or read html from pandas can be used to scrap the list of 50 districts of Bangkok from the following url:https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok
Note: I'll try to fill the row with NA in Latitude and Longitude by using Geocoder. if I cannot get lititude and longitude for that row, I will drop that row.

2.2)Foursquare API

This project will use Foursquare API as its main data collecting source. This API gives the ability to perform location search, location sharing and details about a business and then will be used to search the nearby venues. After that, vanues will be used for k-means clustering analysis to know each types of businesses

2.3)Python packages

- Pandas:Library used to work with data frame
- NumPy:Library used to handle data in a vector
- JSON: Library used to handle with JSON files
- Geopy: To retrieve Location Data
- Requests: Library to used handle http requests
- Matplotlib: used for python plotting
- Sklearn: Library for machine learning
- Folium: Map rendering Library which will be used to visualizing the result from clustering analysis

3.Methodology

The work flow of this project begins with web scrapping by using one of the functions from pandas from link above. Note: this project will consider 'District' column as nightnborhoods

	District(Khet)	Code	Thai	Population	No. ofSubdistricts(Khwaeng)	Latitude	Longitude
0	Bang Bon	50	บางบอน	105161	4	NaN	NaN
1	Bang Kapi	06	บางกะปิ	148465	2	13.765833	100.647778
2	Bang Khae	40	บางเขน	191781	4	13.696111	100.409444
3	Bang Khen	05	บางเขน	189539	2	13.873889	100.596389
4	Bang Kho Laem	31	บางคอแหลม	94956	3	13.693333	100.5025

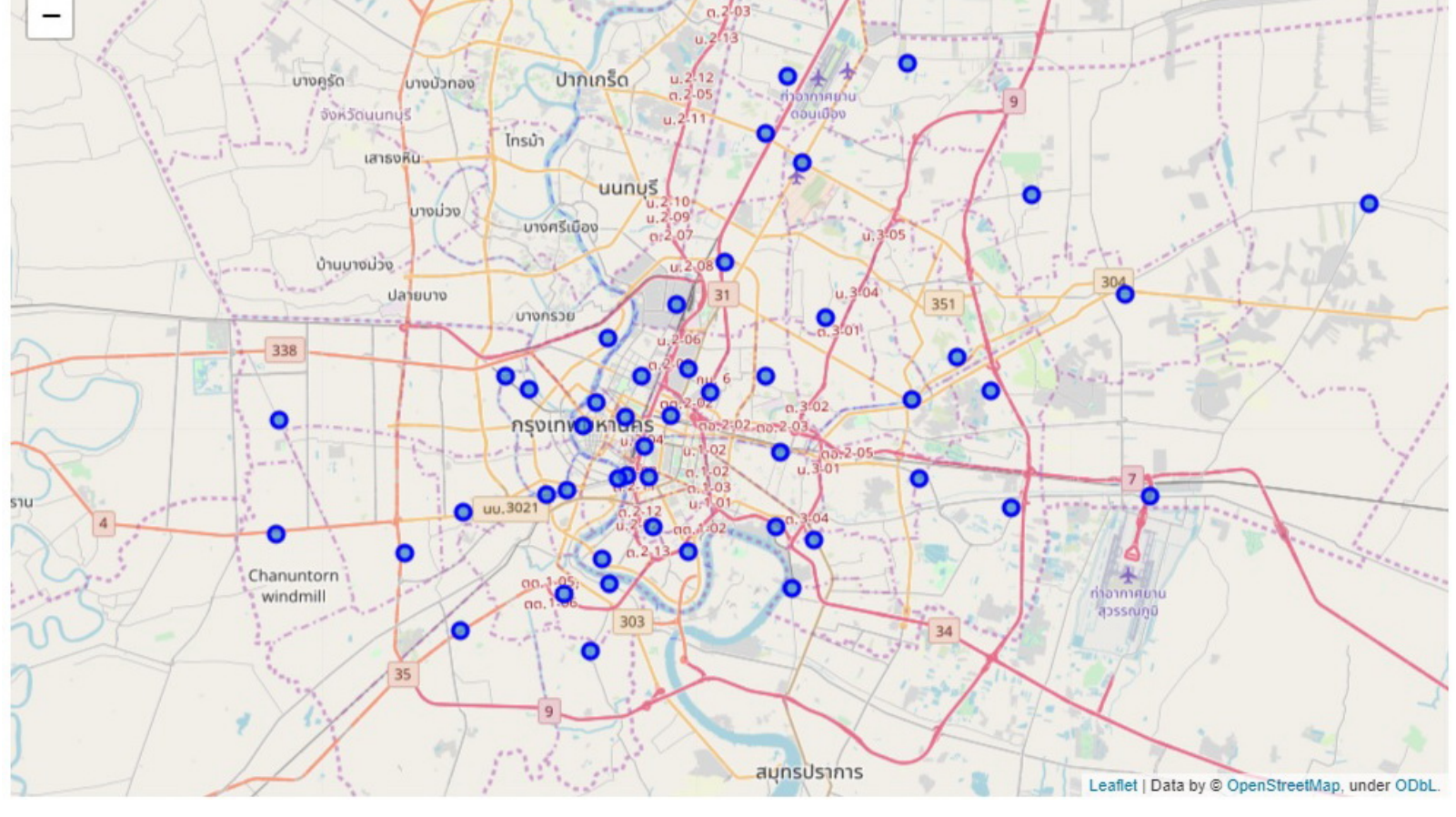
(The data set have no NaN value in 'District' and 'Code' column)
After that, rows which have 'Latitude' or 'Longitude' columns with NaN value will be filled by using geocoder. Then, the NaN value after this process will be dropped.

The next process is retriving the column used to analyze which are 'District','Code','Latitude','Longitude'. However, when plotting these location to the map, the row named Bang Bon in district column has wrong location. So, this row will be dropped.

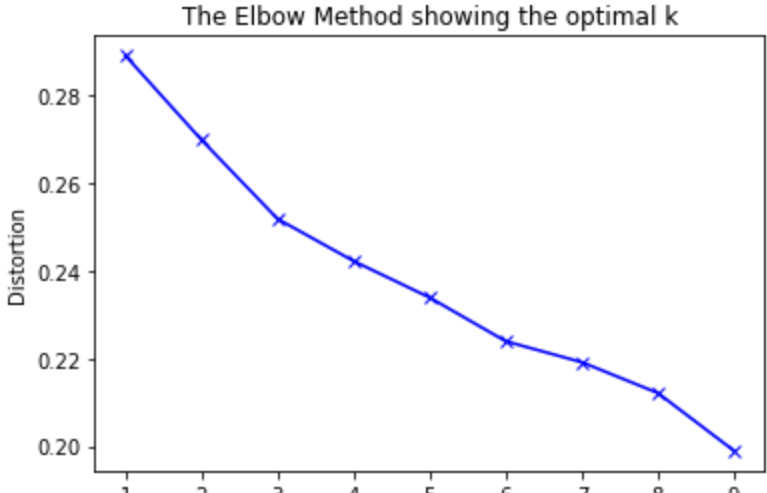
	District	Code	Latitude	Longitude
0	Phra Nakhon	01	13.764444	100.499167
1	Dusit	02	13.776944	100.520556
2	Nong Chok	03	13.855556	100.8625
3	Bang Rak	04	13.730833	100.524167
4	Bang Khen	05	13.873889	100.596389
5	Bang Kapi	06	13.765833	100.647778
6	Pathum Wan	07	13.744942	100.5222
7	Pom Prap Sattru Phai	08	13.758056	100.513056
8	Phra Khanong	09	13.702222	100.601667
9	Min Buri	10	13.813889	100.748056
10	Lat Krabang	11	13.722317	100.759669
11	Yan Nawa	12	13.696944	100.543056
12	Samphanthawong	13	13.731389	100.514167
13	Phaya Thai	14	13.78	100.542778
14	Thon Buri	15	13.725	100.485833
15	Bangkok Yai	16	13.722778	100.476389
16	Huai Khwang	17	13.776667	100.579444
17	Khlong San	18	13.730278	100.509722
18	Taling Noi	19	13.776944	100.456667
19	Bangkok Noi	20	13.770867	100.467933
20	Bang Khun Thian	21	13.660833	100.435833
21	Phasi Charoen	22	13.714722	100.437222
22	Nong Khaem	23	13.704722	100.348889

This project will analyze by using data above.

Before analyzing, the data will be plot in the map.

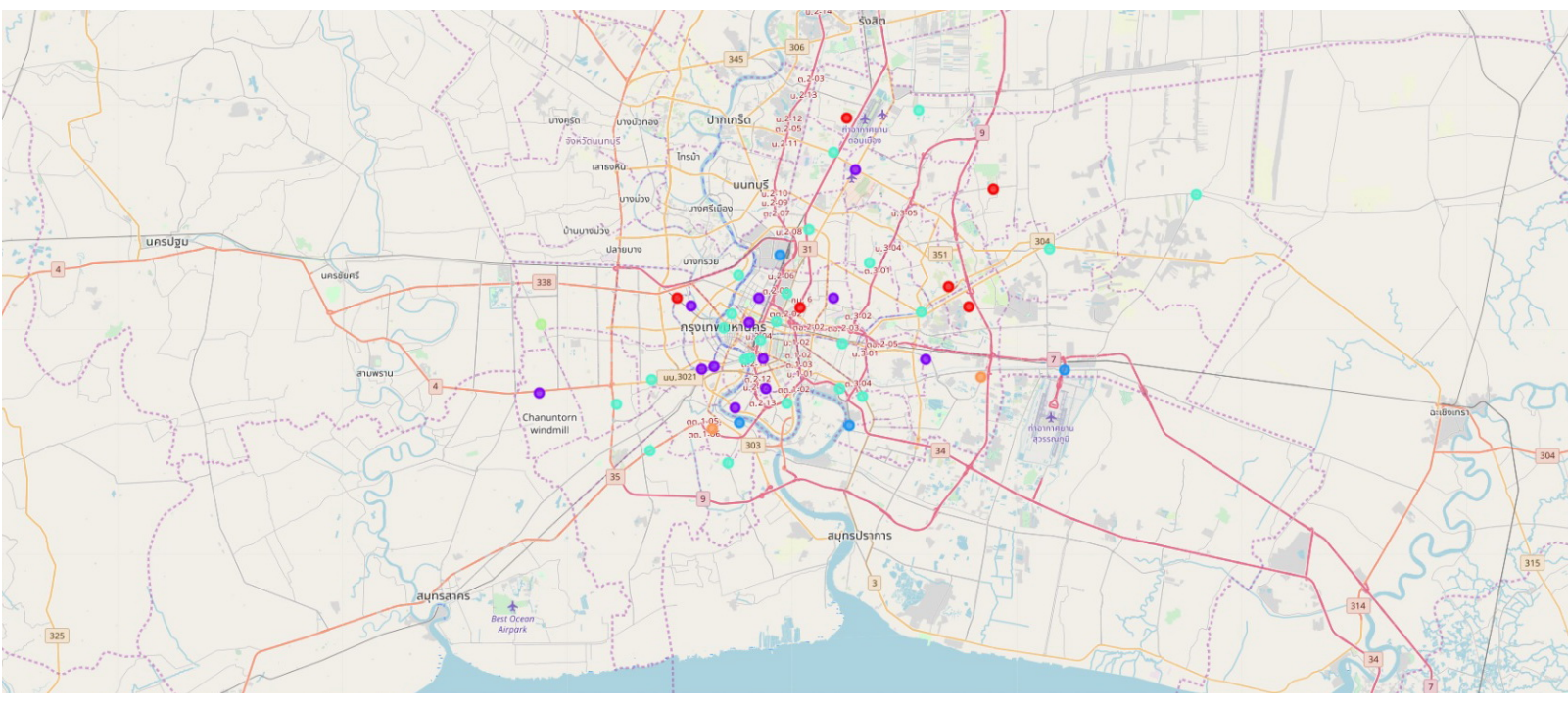


Four-square api is used to retrieve top vanues for each neighborhood.
K-means clustering is used to cluster each district by their top venues.
Elbow analysis is used to find an optimum number of K for K-means clustering (the number is not obvious in this case)



Note: this project use K = 6

4. Result



Explore each cluster

Cluster 0

Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
18	18	18	18	18	18	18	18	18	18	18
19	19	19	19	19	19	19	19	19	19	19
20	20	20	20	20	20	20	20	20	20	20
21	21	21	21	21	21	21	21	21	21	21
22	22	22	22	22	22	22	22	22	22	22
23	23	23	23	23	23	23	23	23	23	23
24	24	24	24	24	24	24	24	24	24	24
25	25	25	25	25	25	25	25	25	25	25
26	26	26	26	26	26	26	26	26	26	26
27	27	27	27	27	27	27	27	27	27	27
28	28	28	28	28	28	28	28	28	28	28
29	29	29	29	29	29	29	29	29	29	29
30	30	30	30	30	30	30	30	30	30	30
31	31	31	31	31	31	31	31	31	31	31
32	32	32	32	32	32	32	32	32	32	32
33	33	33	33	33	33	33	33	33	33	33
34	34	34	34	34	34	34	34	34	34	34
35	35	35	35	35	35	35	35	35	35	35
36	36	36	36	36	36	36	36	36	36	36
37	37	37	37	37	37	37	37	37	37	37
38	38	38	38	38	38	38	38	38	38	38
39	39	39	39	39	39	39	39	39	39	39
40	40	40	40	40	40	40	40	40	40	40
41	41	41	41	41	41	41	41	41	41	41
42	42	42	42	42	42	42	42	42	42	42
43	43	43	43	43	43	43	43	43	43	43
44	44	44	44	44	44	44	44	44	44	44
45	45	45	45	45	45	45	45	45	45	45
46	46	46	46	46	46	46	46	46	46	46
47	47	47	47	47	47	47	47	47	47	47
48	48	48	48	48	48	48	48	48	48	48
49	49	49	49	49	49	49	49	49	49	49
50	50	50	50	50	50	50	50	50	50	50
51	51	51	51	51	51	51	51	51	51	51
52	52	52	52	52	52	52	52	52	52	52
53	53	53	53	53	53	53	53	53	53	53
54	54	54	54	54	54	54	54	54	54	54
55	55	55	55	55	55	55	55	55	55	55
56	56	56	56	56	56	56	56	56	56	56
57	57	57	57	57	57	57	57	57	57	57
58	58	58	58	58	58	58	58	58	58	58
59	59	59	59	59	59	59	59	59	59	59
60	60	60	60	60	60	60	60	60	60	60
61	61	61	61	61	61	61	61	61	61	61
62	62	62	62	62	62	62	62	62	62	62
63	63	63	63	63	63	63	63	63	63	63
64	64	64	64	64	64	64	64	64	64	64
65	65	65	65	65	65	65	65	65	65	65
66	66	66	66	66	66	66	66	66	66	66
67	67	67	67	67	67	67	67	67	67	67
68	68	68	68	68	68	68	68	68	68	68
69	69	69	69	69	69	69	69	69	69	69
70	70	70	70	70	70	70	70	70	70	70
71	71	71	71	71	71	71	71	71	71	71
72	72	72	72	72	72	72	72	72	72	72
73	73	73	73	73	73	73	73	73	73	73
74	74	74	74	74	74	74	74	74	74	74
75	75	75	75	75	75	75	75	75	75	75
76	76	76	76	76	76	76	76	76	76	76
77	77	77	77	77	77	77	77	77	77	77
78	78	78	78	78	78	78	78	78	78	78
79	79	79	79	79	79	79	79	79	79	79
80	80	80	80	80	80	80	80	80	80	80
81	81	81	81	81	81	81	81	81	81	81
82	82	82	82	82	82	82	82	82	82	82
83	83	83	83	83	83	83	83	83	83	83
84	84	84	84	84	84	84	84	84	84	84
85	85	85	85	85	85	85	85	85	85	85
86	86	86	86	86	86	86	86	86	86	86
87	87	87	87	87	87	87	87	87	87	87
88	88	88	88	88	88	88	88	88	88	88
89	89	89	89	89	89	89	89	89	89	89
90	90	90	90	90	90	90	90	90	90	90
91	91	91	91	91	91	91	91	91	91	91
92	92	92	92	92	92	92	92	92	92	92
93	93	93	93	93	93	93	93	93	93	93
94	94	94	94	94	94	94	94	94	94	94
95	95	95	95	95	95	95	95	95	95	95
96	96	96	96	96	96	96	96	96	96	96
97	97	97	97	97	97	97	97	97	97	97
98	98	98	98	98	98	98	98	98	98	98
99	99	99	99	99	99	99	99	99	99	99
100	100	100	100	100	100	100	100	100	100	100

In cluster 0 , there are 6 neighborhoods and most of them have Convenience Store at its first most common venue. So, it can be implied that cluster 0 is the cluster of Convenience Store.

Cluster 1

3	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----