Capstone Project - Week 5

Visit notebook on IBM Watson Studio via this link
>> https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/2da2c5b8-ce12-42e9-a1a8-23f4abaf7a15/view?access_token=69e8a02e7955f690439536e4503c6ab8827a389484474ae48403ac3cbb79b295

1. Introduction

The 2019 Thai general election was held in Thailand on 24 March 2019. The elections selected the five hundred members of the new House of Representatives, the previous House having been dissolved by the coup. The 350 constituency seats are won by first-past-the-post voting as in previous elections. However, the 150 party list seats act as levelling seats, and are allocated so as to give each party a total number of seats proportional to the nationwide number of votes they received.

In 2018, the Election Commission (ECT) was tasked with drawing up new district boundaries. The move sparked outrage from the Pheu Thai and Democrat parties and the watchdog organisation Open Forum for Democracy Foundation. They argued the purpose of the delay was to enable the ECT to draw constituency boundaries that favoured PM Prayut's Palang Pracharath Party.

Therefore, the objective of this project is to investigate the boundary issue by studying correlations between neighborhoods in Thailand and voters' behavior. In this project, we may measure voters' behavior using the most popular party for each constituency. To simplify the analysis, we will focus only on the 350 constituency seats.

Target Audience

People who are interested in 2019 Election fever and wondered about the constituency boundary controversy

2. Data

2.1 Constituency Data

It was the first election since the 2014 Thai coup d'état that installed coup leader General Prayut Chan-o-cha as prime minister, and the first held in accordance with the 2017 constitution, which was drafted under the ruling military junta. Therefore, several news agencies had developed real-time election monitoring website. For example:

- https://elect.thematter.co/
- https://vote.workpointnews.com/
- https://web.facebook.com/thestandardth/photos/a.1725541161072102/2074459142846967/?type=3&theater
- https://vote62.com/

In this project, we will use JSON data file from https://github.com/codeforthailand/dataset-election-62-candidates/blob/master/data/election-zones.json which was used to produce constituency query service page (https://elect.in.th/candidates/) Below is the example of JSON file.

Geospatial Data (Latitude/Longitude)

For each constituency (zone), area (AMPHOE) and province (CHANGWAT) parameters will be used as a key for extracting latitude and longitude from Thailand Open Government Data (Link).

	A	B	C	D	E	F	G	H	I	J	K	L
1	AD_LEVEL	TA_ID	TAMBON_T	TAMBON_E	AM_ID	AMPHOE_T	AMPHOE_E	CH_ID	CHANGWAT_T	CHANGWAT_E	LAT	LONG
2	4	910106	ต. เกาะสาหร่าย	Ko Sarai	9101	อ. เมืองสตูล	Mueang Satun	91	จ. สตูล	Satun	6.546	99.706
3	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.646	101.171
4	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.645	101.170
5	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.649	101.174
6	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.644	101.169
7	4	860116	ต. หาดทรายรี	Hat Sai Ri	8601	อ. เมืองชุมพร	Mueang Chumphon	86	จ. ชุมพร	Chumphon	10.459	99.403
8	4	860706	ต. ด่านสวี	Dan Sawi	8607	อ. สวี	Sawi	86	จ. ชุมพร	Chumphon	10.291	99.241
9	4	440103	ต. ท่าตูม	Tha Tum	4401	อ. เมืองมหาสารคาม	Mueang Maha Sarakham	44	จ. มหาสารคาม	Maha Sarakham	16.162	103.484
10	4	840401	ต. อ่างทอง	Ang Thong	8404	อ. เกาะสมุย	Ko Samui	84	จ. สุราษฎร์ธานี	Surat Thani	9.708	99.675
11	4	840401	ต. อ่างทอง	Ang Thong	8404	อ. เกาะสมุย	Ko Samui	84	จ. สุราษฎร์ธานี	Surat Thani	9.796	99.713
12	4	840401	ต. อ่างทอง	Ang Thong	8404	อ. เกาะสมุย	Ko Samui	84	จ. สุราษฎร์ธานี	Surat Thani	9.677	99.659
13	4	200901	ต. สัตหีบ	Sattahip	2009	อ. สัตหีบ	Sattahip	20	จ. ชลบุรี	Chon Buri	12.611	100.889
14	4	921003	ต. ตะเสะ	Tase	9210	อ. หาดสำราญ	Hat Samran	92	จ. ตรัง	Trang	7.033	99.449
15	4	860116	ต. หาดทรายรี	Hat Sai Ri	8601	อ. เมืองชุมพร	Mueang Chumphon	86	จ. ชุมพร	Chumphon	10.477	99.369
16	4	860306	ต. ปากคลอง	Pak Khlong	8603	อ. ปะทิว	Pathio	86	จ. ชุมพร	Chumphon	10.862	99.486
17	4	860801	ต. ปากตะโก	Pak Tako	8608	อ. ทุ่งตะโก	Thung Tako	86	จ. ชุมพร	Chumphon	10.165	99.176
18	4	860306	ต. ปากคลอง	Pak Khlong	8603	อ. ปะทิว	Pathio	86	จ. ชุมพร	Chumphon	10.800	99.538

2.2 Neighborhood Data

We apply Foursquare API to explore the nearby venues according to the latitude/longitude data of each constituency. Then, venue categories are used as a main feature to cluster neighborhoods in Thailand (similar analysis as Manhattan and Toronto examples).

2.3 Election Result

We use the election result from The Election Commission of Thailand (ECT) official website (Link). The election result is sorted by constituency and popularity of candidate as the picture below (translation in red color). However, data preparation is needed since the file is in PDF format.

คะแนนผู้สมัครรายเขต

Province	Constituency	Number	Candidate Name	Party	Score
จังหวัด	เขตเลือกตั้ง	หมายเลข	ชื่อ-สกุลผู้สมัคร	พรรคการเมือง	คะแนน
กระบี่	1	18	นายสาคร เกี่ยวข้อง	ประชาธิปัตย์	44,346
กระบี่	1	2	นายแสงชัย วสุนธรา	พลังประชารัฐ	30,381
กระบี่	1	15	นายยุทธนา อ่าวลึกน้อย	อนาคตใหม่	17,783
กระบี่	1	16	นายทวีเกียรติ ใจดี	ประชาชาติ	8,484
กระบี่	1	10	ว่าที่ พ.ต.กิตติศักดิ์ กิตติสิทโธ	ภูมิใจไทย	4,784
กระบี่	1	5	นายวศิน สิริเกียรติกุล	เสรีรวมไทย	3,474
กระบี่	1	11	นายจักรพรรดิ สุคนธกนิษฐ	รวมพลังประชาชาติไทย	2,032
กระบี่	1	8	นางวิชญาดา บุญฤทธิ์	แทนคุณแผ่นดิน	1,986
กระบี่	1	24	นางสาววรฤทัย ชัยภักดี	เศรษฐกิจใหม่	1,596
กระบี่	1	7	นายไชยา ด่าดี	ชาติไทยพัฒนา	1,128

After converting PDF into XLSX format using Adobe Acrobat Pro with some Microsoft Excel functions, the file will be like this below.

province	zone	party	num_votes	rank
กรุงเทพมหานคร	1	พลังประชารัฐ	23,246	1
กรุงเทพมหานคร	2	พลังประชารัฐ	26,909	1
กรุงเทพมหานคร	3	อนาคตใหม่	28,444	1
กรุงเทพมหานคร	4	พลังประชารัฐ	27,620	1
กรุงเทพมหานคร	5	เพื่อไทย	27,897	1
กรุงเทพมหานคร	6	พลังประชารัฐ	28,690	1
กรุงเทพมหานคร	7	พลังประชารัฐ	25,180	1
กรุงเทพมหานคร	8	พลังประชารัฐ	29,090	1
กรุงเทพมหานคร	9	พลังประชารัฐ	34,907	1
กรุงเทพมหานคร	10	เพื่อไทย	30,800	1
กรุงเทพมหานคร	11	เพื่อไทย	34,679	1
กรุงเทพมหานคร	12	เพื่อไทย	30,254	1
กรุงเทพมหานคร	13	พลังประชารัฐ	27,489	1
กรุงเทพมหานคร	14	เพื่อไทย	31,445	1
กรุงเทพมหานคร	15	พลังประชารัฐ	31,551	1

For further analysis, this result may be modelled using classification algorithm since there are a few parties holding majority from 350 constituencies.

3. Methodologies

3.1 Import Libraries

Import necessary libraries for our projects including:

- BeautifulSoup: scraping data from website
- Numpu: library to handle data in a vectorized manner
- Pandas: library for data analsysis
- Json: library to handle JSON files
- Geopy: convert an address into latitude and longitude values
- Request: library to handle requests
- Matplotlib: plotting modules
- Sklearn.cluster (KMeans): for clustering stage
- Folium: map rendering library
- Statistics: for statistic calculation

3.2 Data Preprocessing

3.2.1 Geospatial Data (Latitude/Longitude)

- Import XLSX data and convert into dataframe format

	AD_LEVEL	TA_ID	TAMBON_T	TAMBON_E	AM_ID	AMPHOE_T	AMPHOE_E	CH_ID	CHANGWAT_T	CHANGWAT_E	LAT	LONG
0	4	910106	ต. เกาะสาหร่าย	Ko Sarai	9101	อ. เมืองสตูล	Mueang Satun	91	จ. สตูล	Satun	6.546	99.706
1	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.646	101.171
2	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.645	101.170
3	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.649	101.174
4	4	210114	ต. มาบตาพุด	Maptaphut	2101	อ. เมืองระยอง	Mueang Rayong	21	จ. ระยอง	Rayong	12.644	101.169

- Select some useful columns

	CHANGWAT_T	CHANGWAT_E	AMPHOE_T	AMPHOE_E	LAT	LONG
0	สตูล	Satun	เมืองสตูล	Mueang Satun	6.546	99.706
1	ระยอง	Rayong	เมืองระยอง	Mueang Rayong	12.646	101.171
2	ระยอง	Rayong	เมืองระยอง	Mueang Rayong	12.645	101.170
3	ระยอง	Rayong	เมืองระยอง	Mueang Rayong	12.649	101.174
4	ระยอง	Rayong	เมืองระยอง	Mueang Rayong	12.644	101.169

- Group dataframe by province (CHANGWAT) and area (AMPHOE)

		LAT	LONG
CHANGWAT_E	AMPHOE_E		
	Chanuman	16.130000	104.938400
	Hua Taphan	15.688750	104.524750
	Lue Amnat	15.714857	104.693286
Amnat Charoen	Mueang Amnat Charoen	15.874000	104.630684
	Pathum Ratchawongsa	15.885429	104.899571
	Phana	15.691250	104.870750
	Senangkhanikhom	16.053000	104.680500
	Chaiyo	14.674667	100.468222
	Mueang Ang Thong	14.583000	100.449000
	Pa Mok	14.490250	100.453250
Ang Thong	Pho Thong	14.670067	100.369000

3.2.2 Constituency Data

- Convert from json file into dataframe

	areas	prefixes.area	prefixes.sub_area	province	zone
0	[{'exterior': [], 'subinterior': [], 'area': '...	อำเภอ	ตำบล	กระบี่	1
1	[{'exterior': [], 'subinterior': [], 'area': '...	อำเภอ	ตำบล	กระบี่	2
2	[{'exterior': ['ถนนนครไชยศรี'], 'subinterior':...	เขต	แขวง	กรุงเทพมหานคร	1
3	[{'exterior': [], 'subinterior': [], 'area': '...	เขต	แขวง	กรุงเทพมหานคร	2
4	[{'exterior': [], 'subinterior': [], 'area': '...	เขต	แขวง	กรุงเทพมหานคร	3

- Drop unnecessary columns and rearrange column order

	province	areas
0	กระบี่	[{'exterior': [], 'subinterior': [], 'area': '...
1	กระบี่	[{'exterior': [], 'subinterior': [], 'area': '...
2	กรุงเทพมหานคร	[{'exterior': ['ถนนนครไชยศรี'], 'subinterior':...
3	กรุงเทพมหานคร	[{'exterior': [], 'subinterior': [], 'area': '...
4	กรุงเทพมหานคร	[{'exterior': [], 'subinterior': [], 'area': '...

- Flatten 'areas' from json format into comma-separated format & Translate province and area names from Thai to English

	province	areas
0	Krabi	[Plai Phraya, Ao Luek, Khao Phanom, Mueang Krabi]
1	Krabi	[Khlong Thom, Lam Thap, Ko Lanta, Khao Phanom,...
2	Bangkok	[Dusit, Pom Prap Sattruphai, Phra Nakhon, Samp...
3	Bangkok	[Bang Rak, Pathum Wan, Sa Thon]
4	Bangkok	[Bang Kho Laem, Yannawa]

- Find average coordinate for each row

	province	areas	lat	long
0	Krabi	[Plai Phraya, Ao Luek, Khao Phanom, Mueang Krabi]	8.29504	98.86
1	Krabi	[Khlong Thom, Lam Thap, Ko Lanta, Khao Phanom,...	7.94758	99.1607
2	Bangkok	[Dusit, Pom Prap Sattruphai, Phra Nakhon, Samp...	13.7546	100.509
3	Bangkok	[Bang Rak, Pathum Wan, Sa Thon]	13.7275	100.528
4	Bangkok	[Bang Kho Laem, Yannawa]	13.6941	100.522
5	Bangkok	[Khlong Toei, Watthana]	13.723	100.578
6	Bangkok	[Din Daeng, Huai Khwang]	13.7752	100.574

- Join back with zone

	province	areas	lat	long	zone
0	Krabi	[Plai Phraya, Ao Luek, Khao Phanom, Mueang Krabi]	8.29504	98.86	1
1	Krabi	[Khlong Thom, Lam Thap, Ko Lanta, Khao Phanom,...	7.94758	99.1607	2
2	Bangkok	[Dusit, Pom Prap Sattruphai, Phra Nakhon, Samp...	13.7546	100.509	1
3	Bangkok	[Bang Rak, Pathum Wan, Sa Thon]	13.7275	100.528	2
4	Bangkok	[Bang Kho Laem, Yannawa]	13.6941	100.522	3

- Finalize neighborhood coordinate dataframe

	province	zone	lat	long
0	Amnat Charoen	1	15.7814	104.578
1	Amnat Charoen	2	15.8949	104.817
2	Ang Thong	1	14.617	100.375
3	Bangkok	1	13.7546	100.509
4	Bangkok	2	13.7275	100.528

3.2.3 Election Result

- Import XLSX data and convert into dataframe format

	province	zone	party	num_votes	rank
0	กรุงเทพมหานคร	1	พลังประชารัฐ	23246	1
1	กรุงเทพมหานคร	2	พลังประชารัฐ	26909	1
2	กรุงเทพมหานคร	3	อนาคตใหม่	28444	1
3	กรุงเทพมหานคร	4	พลังประชารัฐ	27620	1
4	กรุงเทพมหานคร	5	เพื่อไทย	27897	1

- Eliminate 'rank' and 'num_votes' column since we focus only on the winner for each constituency

	province	zone	party
0	กรุงเทพมหานคร	1	พลังประชารัฐ
1	กรุงเทพมหานคร	2	พลังประชารัฐ
2	กรุงเทพมหานคร	3	อนาคตใหม่
3	กรุงเทพมหานคร	4	พลังประชารัฐ
4	กรุงเทพมหานคร	5	เพื่อไทย

- Translate province from Thai to English

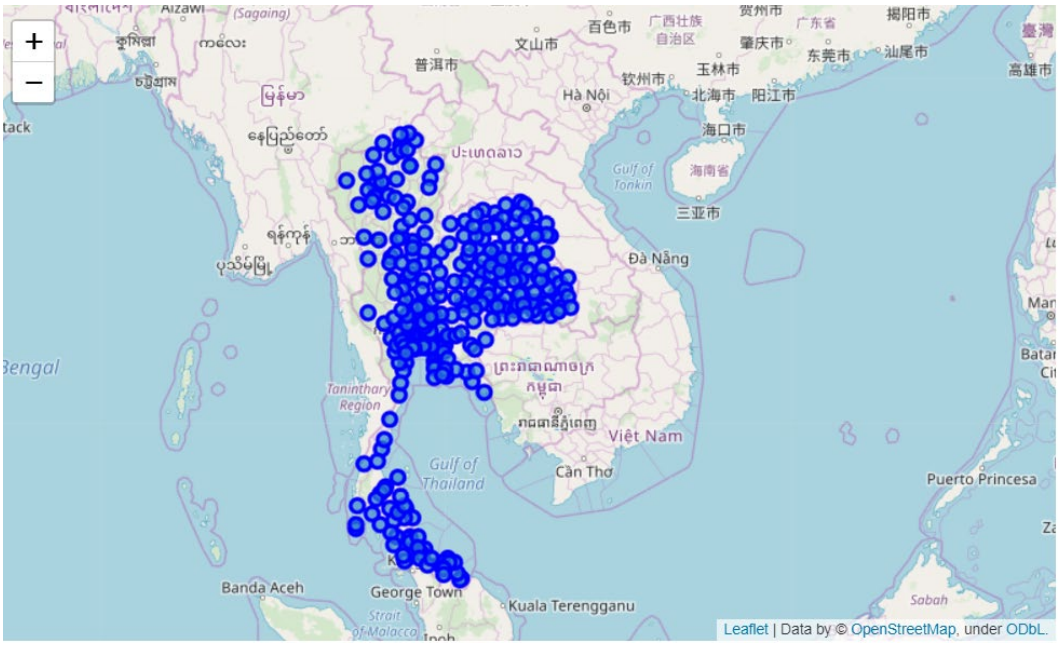
	province	zone	party
0	Bangkok	1	พลังประชารัฐ
1	Bangkok	2	พลังประชารัฐ
2	Bangkok	3	อนาคตใหม่
3	Bangkok	4	พลังประชารัฐ
4	Bangkok	5	เพื่อไทย

- Finalize dataframe & Create new 'key' column

	province	zone	lat	long	party	key
0	Amnat Charoen	1	15.7814	104.578	เพื่อไทย	Amnat Charoen 1
1	Amnat Charoen	2	15.8949	104.817	เพื่อไทย	Amnat Charoen 2
2	Ang Thong	1	14.617	100.375	ภูมิใจไทย	Ang Thong 1
3	Bangkok	1	13.7546	100.509	พลังประชารัฐ	Bangkok 1
4	Bangkok	2	13.7275	100.528	พลังประชารัฐ	Bangkok 2

3.3 Exploring Dataset

- Create map using latitude and longitude values



- Define function to get nearby venues for each constituency
- View example of venues nearby the constituencies

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Amnat Charoen 1	15.781375	104.577717	Fai Kid Hotel (โรงแรมฟ้าย่ิด)	15.860842	104.620121	Hotel
1	Amnat Charoen 1	15.781375	104.577717	แผ่นดินทองฟาร์มฟิชชิ่งปาร์ค	15.834809	104.606749	American Restaurant
2	Amnat Charoen 1	15.781375	104.577717	ก๋วยเตี๋ยวเป็ดซิตี้	15.863962	104.584794	Noodle House
3	Amnat Charoen 1	15.781375	104.577717	Shell (หอมกาแฟ)	15.862201	104.599185	Café
4	Amnat Charoen 1	15.781375	104.577717	Sirin Cafe	15.853887	104.631146	Coffee Shop
5	Amnat Charoen 1	15.781375	104.577717	ส้มตำยายพัน	15.858168	104.625800	Thai Restaurant
6	Amnat Charoen 2	15.894907	104.816501	ผาชิ่นวาริน	15.910299	104.796067	National Park
7	Amnat Charoen 2	15.894907	104.816501	ภูเปาะ จ.อำนาจเจริญ	15.940458	104.850145	Mountain
8	Amnat Charoen 2	15.894907	104.816501	ห้วยสีโท	15.886124	104.740519	Travel & Transport
9	Amnat Charoen 2	15.894907	104.816501	วัดดอยนกยูงทอง	15.954864	104.872673	Forest

- Group venues by 'Neighborhood' and count number of venues

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Amnat Charoen 1	6	6	6	6	6	6
Amnat Charoen 2	4	4	4	4	4	4
Ang Thong 1	30	30	30	30	30	30
Bangkok 1	95	95	95	95	95	95
Bangkok 10	56	56	56	56	56	56

- Find number of unique categories in the data extracted

3.4 Analyze each neighborhood

3.4.1 Use One Hot Encoding to create dummie variables

	Neighborhood	Airport	Airport Lounge	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Arepa Restaurant	Art Gallery	Art Museum	& Crafts Store	Asian Restaurant	Astrologer	Athletics & Sports	Auto Dealership	Auto Garage	Auto Workshop	BBQ Joint
0	Amnat Charoen 1	0.0	0.0	0.0	0.166667	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000
1	Amnat Charoen 2	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000
2	Ang Thong 1	0.0	0.0	0.0	0.033333	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.033333	0.0	0.0	0.0	0.000000	0.0	0.033333
3	Bangkok 1	0.0	0.0	0.0	0.010526	0.0	0.0	0.0	0.0	0.010526	0.010526	0.0	0.042105	0.0	0.0	0.0	0.000000	0.0	0.000000
4	Bangkok 10	0.0	0.0	0.0	0.017857	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.053571	0.0	0.0	0.0	0.000000	0.0	0.017857
5	Bangkok 11	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.044776	0.0	0.0	0.0	0.014925	0.0	0.000000
6	Bangkok 12	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.040816	0.0	0.0	0.0	0.000000	0.0	0.040816
7	Bangkok 13	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.068966	0.0	0.0	0.0	0.000000	0.0	0.000000
8	Bangkok 14	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.025974	0.0	0.0	0.0	0.012987	0.0	0.012987
9	Bangkok 15	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.040000	0.0	0.0	0.0	0.000000	0.0	0.020000

3.4.2 Extract top 10 most frequent venue categories

	Neighborhood	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
0	Amnat Charoen 1	Coffee Shop	Noodle House	American Restaurant	Thai Restaurant	Hotel	Café	Zoo Exhibit	Farm	Farmers Market	Fast Food Restaurant
1	Amnat Charoen 2	Forest	National Park	Travel & Transport	Mountain	Flea Market	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
2	Ang Thong 1	Thai Restaurant	Market	Coffee Shop	Noodle House	Convenience Store	Flea Market	Soccer Stadium	Cafeteria	Asian Restaurant	Restaurant
3	Bangkok 1	Noodle House	Café	Chinese Restaurant	Thai Restaurant	Hotel	Asian Restaurant	Dessert Shop	Convenience Store	Coffee Shop	Hostel
4	Bangkok 10	Noodle House	Thai Restaurant	Convenience Store	Coffee Shop	Restaurant	Flea Market	Asian Restaurant	Steakhouse	Bar	Café

3.5 Clustering Neighborhoods

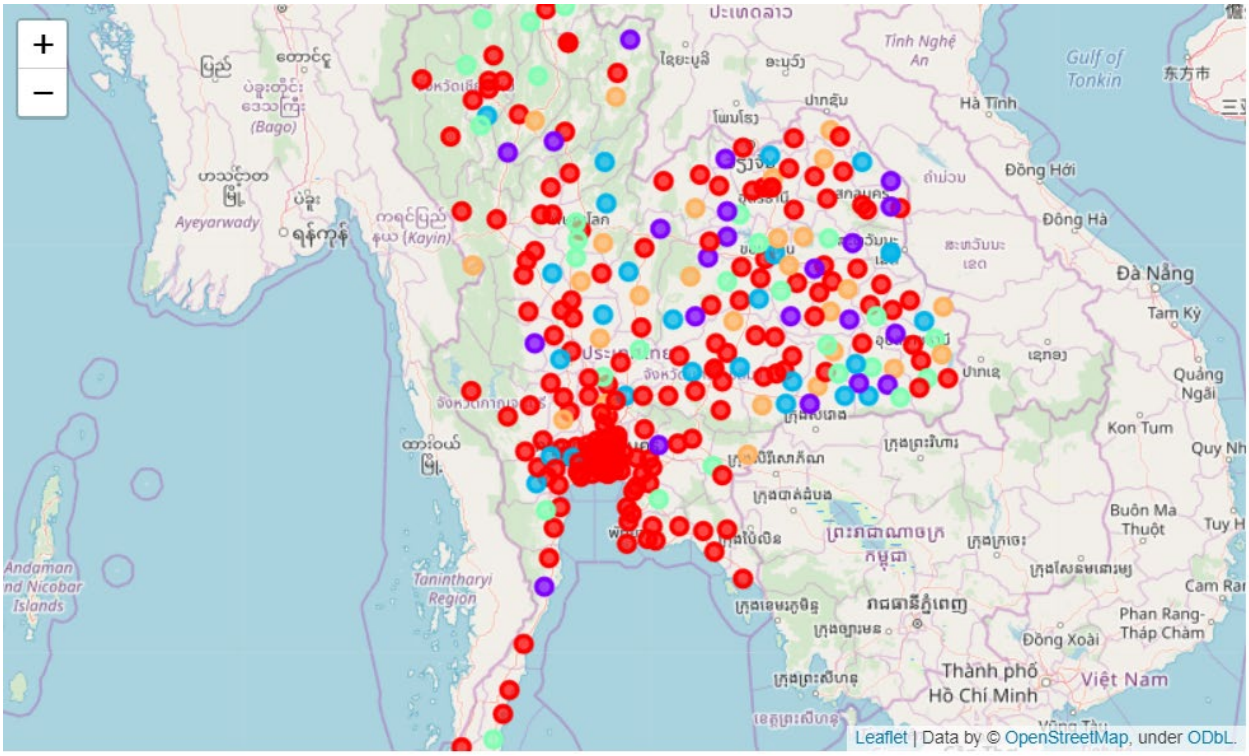
- Run K-Means clustering

	Cluster Labels	Neighborhood	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
0	0	Amnat Charoen 1	Coffee Shop	Noodle House	American Restaurant	Thai Restaurant	Hotel	Café	Zoo Exhibit	Farm	Farmers Market	Fast Food Restaurant
1	0	Amnat Charoen 2	Forest	National Park	Travel & Transport	Mountain	Flea Market	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
2	0	Ang Thong 1	Thai Restaurant	Market	Coffee Shop	Noodle House	Convenience Store	Flea Market	Soccer Stadium	Cafeteria	Asian Restaurant	Restaurant
3	0	Bangkok 1	Noodle House	Café	Chinese Restaurant	Thai Restaurant	Hotel	Asian Restaurant	Dessert Shop	Convenience Store	Coffee Shop	Hostel
4	0	Bangkok 10	Noodle House	Thai Restaurant	Convenience Store	Coffee Shop	Restaurant	Flea Market	Asian Restaurant	Steakhouse	Bar	Café

- Merge dataframes to view add latitude/longitude for each neighborhood

	province	zone	lat	long	party	key	Cluster Labels	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
0	Amnat Charoen	1	15.7814	104.578	เพื่อไทย	Amnat Charoen 1	0.0	Coffee Shop	Noodle House	American Restaurant	Thai Restaurant	Hotel	Café	Zoo Exhibit	Farm	Farmers Market	Fast Food Restaurant
1	Amnat Charoen	2	15.8949	104.817	เพื่อไทย	Amnat Charoen 2	0.0	Forest	National Park	Travel & Transport	Mountain	Flea Market	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
2	Ang Thong	1	14.617	100.375	ภูมิใจไทย	Ang Thong 1	0.0	Thai Restaurant	Market	Coffee Shop	Noodle House	Convenience Store	Flea Market	Soccer Stadium	Cafeteria	Asian Restaurant	Restaurant
3	Bangkok	1	13.7546	100.509	พลังประชารัฐ	Bangkok 1	0.0	Noodle House	Café	Chinese Restaurant	Thai Restaurant	Hotel	Asian Restaurant	Dessert Shop	Convenience Store	Coffee Shop	Hostel
4	Bangkok	2	13.7275	100.528	พลังประชารัฐ	Bangkok 2	0.0	Café	Hotel	Coffee Shop	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Thai Restaurant	Ramen Restaurant	Noodle House	French Restaurant

- View map of clustering results



3.6 Classification Algorithm of Constituency Winner

3.6.1 Test/Train Split

```
from sklearn.model_selection import train_test_split

df_y = df_final
y = th_grouped.join(df_y.set_index('key'), on='Neighborhood')
```

```
X_train, X_test, y_train, y_test = train_test_split(th_grouped.drop('Neighborhood', 1), y['party'], te
st_size=0.2, random_state=4)
print ('Train set:', X_train.shape,  y_train.shape)
print ('Test set:', X_test.shape,  y_test.shape)
```

```
Train set: (278, 335) (278,)
Test set: (70, 335) (70,)
```

3.6.2 Apply Classification Algorithm

```
from sklearn.neighbors import KNeighborsClassifier

k = 5

#Train Model and Predict
neigh = KNeighborsClassifier(n_neighbors = k).fit(X_train,y_train)
```

```
yhat = neigh.predict(X_test)
yhat[0:5]
```

```
array(['พลังประชารัฐ', 'เพื่อไทย', 'เพื่อไทย', 'พลังประชารัฐ',
       'พลังประชารัฐ'], dtype=object)
```

3.6.3 Accuracy Evaluation

```
from sklearn import metrics
print("Train set Accuracy: ", metrics.accuracy_score(y_train, neigh.predict(X_train)))
print("Test set Accuracy: ", metrics.accuracy_score(y_test, yhat))
```

```
Train set Accuracy:  0.503597122302
Test set Accuracy:  0.342857142857
```

4. Result

4.1 Cluster of Each Constituency (Neighborhood)

We can see that each cluster have different most common venue. Let's explore!

4.1.1 Cluster 1

	zone	key	Cluster Labels	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
0	1	Amnat Charoen 1	0.0	Coffee Shop	Noodle House	American Restaurant	Thai Restaurant	Hotel	Café	Zoo Exhibit	Farm	Farmers Market	Fast Food Restaurant
1	2	Amnat Charoen 2	0.0	Forest	National Park	Travel & Transport	Mountain	Flea Market	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
2	1	Ang Thong 1	0.0	Thai Restaurant	Market	Coffee Shop	Noodle House	Convenience Store	Flea Market	Soccer Stadium	Cafeteria	Asian Restaurant	Restaurant
3	1	Bangkok 1	0.0	Noodle House	Café	Chinese Restaurant	Thai Restaurant	Hotel	Asian Restaurant	Dessert Shop	Convenience Store	Coffee Shop	Hostel
4	2	Bangkok 2	0.0	Café	Hotel	Coffee Shop	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Thai Restaurant	Ramen Restaurant	Noodle House	French Restaurant
5	3	Bangkok 3	0.0	Convenience Store	Coffee Shop	Chinese Restaurant	Thai Restaurant	Badminton Court	Spa	Noodle House	Pub	Food Truck	Gym / Fitness Center
6	4	Bangkok 4	0.0	Thai Restaurant	Café	Hotel	Coffee Shop	Noodle House	Japanese Restaurant	Bar	BBQ Joint	French Restaurant	Korean Restaurant
7	5	Bangkok 5	0.0	Hotel	Thai Restaurant	Som Tum Restaurant	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Coffee Shop	Dim Sum Restaurant	Noodle House	Other Nightlife
8	6	Bangkok 6	0.0	Noodle House	Asian Restaurant	Som Tum Restaurant	Restaurant	Thai Restaurant	Bank	Convenience Store	Japanese Restaurant	Coffee Shop	Café
9	7	Bangkok 7	0.0	Thai Restaurant	Convenience Store	Coffee Shop	Noodle House	BBQ Joint	Fast Food Restaurant	Chinese Restaurant	Som Tum Restaurant	Steakhouse	Clothing Store

4.1.2 Cluster 2

	zone	key	Cluster Labels	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
50	2	Chaiyaphum 2	1.0	River	Farm	Flea Market	Food Stand	Food Service	Exhibit	Fair	Food Truck	Farmers Market	Fast Food Restaurant
53	5	Chaiyaphum 5	1.0	Restaurant	Bus Station	Flea Market	Fair	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Fish Market
86	2	Kalasin 2	1.0	Flea Market	Furniture / Home Store	Plaza	Historic Site	Coffee Shop	Fish Market	Fair	Farm	Farmers Market	Fast Food Restaurant
89	5	Kalasin 5	1.0	Restaurant	Asian Restaurant	Market	Deli / Bodega	Fair	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
103	5	Khon Kaen 5	1.0	Scenic Lookout	Science Museum	Restaurant	National Park	Park	Mountain	Zoo Exhibit	Fish Market	Fair	Farm
114	4	Lampang 4	1.0	Restaurant	Market	Thai Restaurant	Juice Bar	Mountain	Zoo Exhibit	Fish Market	Fair	Farm	Farmers Market
127	3	Maha Sarakham 3	1.0	Restaurant	Café	Resort	Convenience Store	Flea Market	Bar	History Museum	Exhibit	Fair	Farm
139	2	Nakhon Phanom 2	1.0	Restaurant	Flea Market	BBQ Joint	Thai Restaurant	Racetrack	Airport Terminal	Food Court	Food & Drink Shop	Fair	Farm
141	4	Nakhon Phanom 4	1.0	Restaurant	Flea Market	Food Truck	Fish Market	Exhibit	Fair	Farm	Farmers Market	Fast Food Restaurant	Field
163	2	Nakhon Si Thammarat 2	1.0	Farm	Resort	Breakfast Spot	Bistro	Flea Market	Night Market	Deli / Bodega	Cruise	Field	Fish & Chips Shop

4.1.3 Cluster 3

	zone	key	Cluster Labels	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
39	5	Buri Ram 5	2.0	Thai Restaurant	Hotel	Historic Site	Som Tum Restaurant	Zoo Exhibit	Fish & Chips Shop	Exhibit	Fair	Farm	Farmers Market
41	7	Buri Ram 7	2.0	Restaurant	Historic Site	Thai Restaurant	Noodle House	Cruise	Fishing Spot	Farm	Farmers Market	Fast Food Restaurant	Creperie
48	2	Chai Nat 2	2.0	Thai Restaurant	Pharmacy	Soccer Field	Asian Restaurant	Event Space	Fair	Farm	Farmers Market	Fast Food Restaurant	Field
51	3	Chaiyaphum 3	2.0	Cave	Department Store	Thai Restaurant	Zoo Exhibit	Fishing Spot	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
93	4	Kamphaeng Phet 4	2.0	Thai Restaurant	Resort	Diner	Comfort Food Restaurant	Fruit & Vegetable Store	Fair	Farm	Farmers Market	Fast Food Restaurant	Field
100	2	Khon Kaen 2	2.0	Thai Restaurant	Convenience Store	Seafood Restaurant	Café	Coffee Shop	Fish & Chips Shop	Fair	Farm	Farmers Market	Fast Food Restaurant
107	9	Khon Kaen 9	2.0	Thai Restaurant	Restaurant	Coffee Shop	Noodle House	Food Truck	Dance Studio	Flea Market	Farmers Market	Fast Food Restaurant	Field
110	2	Krabi 2	2.0	Thai Restaurant	Hot Spring	Dim Sum Restaurant	Gourmet Shop	Waterfall	Coffee Shop	Fish Market	Fair	Farm	Farmers Market
115	1	Lamphun 1	2.0	Thai Restaurant	Zoo Exhibit	Fishing Spot	Fair	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Fish Market
130	1	Mukdahan 1	2.0	Thai Restaurant	Zoo Exhibit	Fishing Spot	Fair	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Fish Market

4.1.4 Cluster 4

	zone	key	Cluster Labels	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
11	9	Bangkok 9	3.0	Coffee Shop	Café	Convenience Store	Art Museum	Soccer Field	Grocery Store	Burger Joint	Badminton Court	Seafood Restaurant	Pool
54	6	Chaiyaphum 6	3.0	Coffee Shop	Convenience Store	Food Court	Café	Fishing Spot	Fair	Farm	Farmers Market	Fast Food Restaurant	Field
62	5	Chiang Mai 5	3.0	Hot Spring	Coffee Shop	Buddhist Temple	Flea Market	Fair	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
63	6	Chiang Mai 6	3.0	Resort	National Park	Historic Site	Coffee Shop	Fishing Spot	Fair	Farm	Farmers Market	Fast Food Restaurant	Field
68	2	Chiang Rai 2	3.0	Coffee Shop	Convenience Store	Diner	Noodle House	Bakery	Café	Lingerie Store	Restaurant	Museum	Dessert Shop
70	4	Chiang Rai 4	3.0	Coffee Shop	Food Truck	Market	Pedestrian Plaza	Thai Restaurant	Warehouse Store	Zoo Exhibit	Fish & Chips Shop	Fair	Farm
77	4	Chon Buri 4	3.0	Coffee Shop	Convenience Store	Zoo Exhibit	Flea Market	Fair	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
84	3	Chumphon 3	3.0	Coffee Shop	Market	Thai Restaurant	Intersection	Train Station	Garden	Flea Market	Paper / Office Supplies Store	Diner	Department Store
88	4	Kalasin 4	3.0	Coffee Shop	Restaurant	Fried Chicken Joint	Buffet	Farmers Market	Food Truck	Convenience Store	Deli / Bodega	Floating Market	Fast Food Restaurant
102	4	Khon Kaen 4	3.0	Coffee Shop	Convenience Store	Campground	Fishing Spot	Fair	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop

4.1.5 Cluster 5

	zone	key	Cluster Labels	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
33	1	Bueng Kan 1	4.0	Resort	Asian Restaurant	Farm	Coffee Shop	Convenience Store	Food Stand	Fishing Spot	Fair	Food Truck	Farmers Market
42	8	Buri Ram 8	4.0	Convenience Store	Farm	Market	Garden Center	Zoo Exhibit	Flea Market	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
52	4	Chaiyaphum 4	4.0	Convenience Store	National Park	Coffee Shop	Mountain	Scenic Lookout	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
71	5	Chiang Rai 5	4.0	Convenience Store	Rental Car Location	Bar	Outlet Store	Zoo Exhibit	Fishing Spot	Fair	Farm	Farmers Market	Fast Food Restaurant
87	3	Kalasin 3	4.0	Convenience Store	Coffee Shop	Night Market	Food Truck	Café	Flea Market	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
101	3	Khon Kaen 3	4.0	Convenience Store	Campground	Zoo	Hotel	Pub	Exhibit	Coffee Shop	Fish Market	Farm	Farmers Market
113	3	Lampang 3	4.0	Convenience Store	Intersection	Asian Restaurant	Gift Shop	Flea Market	Coffee Shop	Soccer Field	Trail	Restaurant	Art Gallery
118	2	Loei 2	4.0	Noodle House	Garden	Flea Market	Sculpture Garden	Convenience Store	Fair	Farm	Farmers Market	Fast Food Restaurant	Field
129	5	Maha Sarakham 5	4.0	Convenience Store	Intersection	Coffee Shop	Jewelry Store	Noodle House	Farm	Deli / Bodega	Restaurant	Arepa Restaurant	Cupcake Shop
146	5	Nakhon Ratchasima 5	4.0	Convenience Store	Coffee Shop	Supermarket	Shopping Mall	Noodle House	Dance Studio	Flea Market	Farm	Farmers Market	Fast Food Restaurant

4.2 Accuracy of Classification Model

As seen below, train and test set accuracy is 0.50 and 0.34 respectively.

```
from sklearn import metrics
print("Train set Accuracy: ", metrics.accuracy_score(y_train, neigh.predict(X_train)))
print("Test set Accuracy: ", metrics.accuracy_score(y_test, yhat))
```

```
Train set Accuracy:  0.503597122302
Test set Accuracy:  0.342857142857
```

5. Discussion

From the result of exploring each cluster, we can notice some observations for each clusters:

- Cluster 1: Famous landmark, Downtown, High venue density
- Cluster 2 and 5: Countryside, Low venue density
- Cluster 3: Lake, River, and restaurant/hotel around them
- Cluster 4: Mountain and restaurant/hotel around them

Each neighborhood environment may affect election result. To investigate this issue, we can measure accuracy of classification model. Train and test set accuracy is only 0.50 and 0.34, respectively. Low accuracy can indicate that neighborhood may not affect which party people choose.

Recommendation

- Vary number of cluster (K) using in the K-Means algorithm
- Plotting number of neighbors (K) using in the KNN algorithm to optimize accuracy
- Select different number of venues' categories using in the K-Means (may be more than 10 most common categories

6. Conclusion

We observe weak correlations between neighborhoods in Thailand and voters' behavior. From classification model, the most popular party for each constituency doesn't depend on neighborhood environment.