DS

LE CONG BINH

BANKING TRANSACTION TEMINAL NETWORK OPTIMIZE

OBJECTIVE

Apply data science technical to support manage all transaction point: Branches, ATM and POS Merchant

DATA

Foursquare API
Transaction Network of bank
Folium data to visualization

INTRODUCTION

Ha Noi, the capital of Vietnam, where over 8 million people live and it has a population density of 2.398 people per square kilometer. The city have 30 district but all transaction only do at five centers district. So, each of bank need control transaction network. That include, branch, ATM and POS Merchant. There are more 40 bank in Ha Noi; customer of any bank can do transaction in every bank. Therefore, We need a system to easy control transaction network.

I am working in a Bank. We have data about all transaction point of my Bank. However, I do not have information of their bank. Other side, I need more information about each of place to decide which place should be setup new agent, which place need to remap for control network. I chose Foursquare data and unsupervised cluster to solve that problem.

METHODOLOGY

BUSINESS UNDERSTAND.

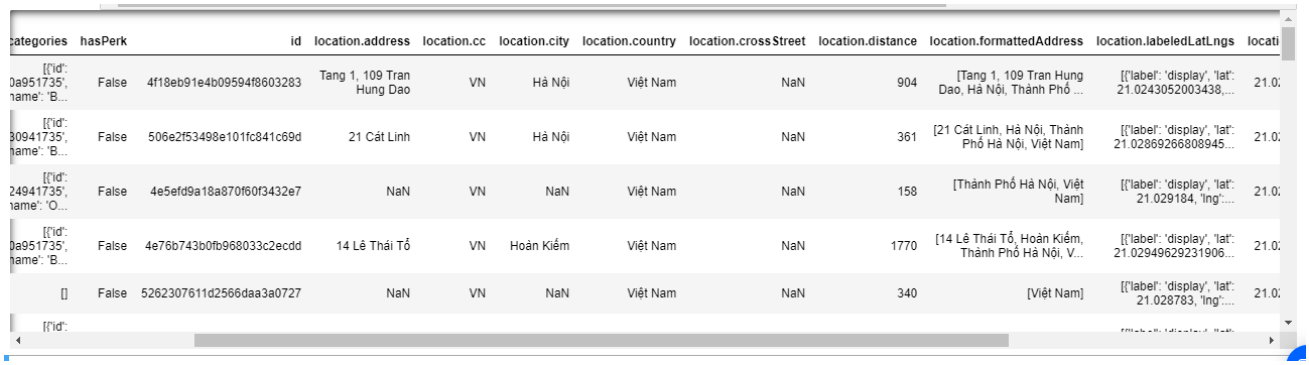
We focus about collect all information about banking network to support deciding rebuilt transaction network

DATA PROCECING

As I show in notebook, There are many technical to preprocessing data. This is step, which need more time to complete. After processing data I use K-Mean to cluster and use that information merge with my branch network to decide next action.

EMAIL | TWITTER HANDLE | TELEPHONE | LINKEDIN URL

I use google service to define current coordinates. (Latitude = 21.027763 Longitude = 105.834160)
After that used Regular Call with search query to get json content.



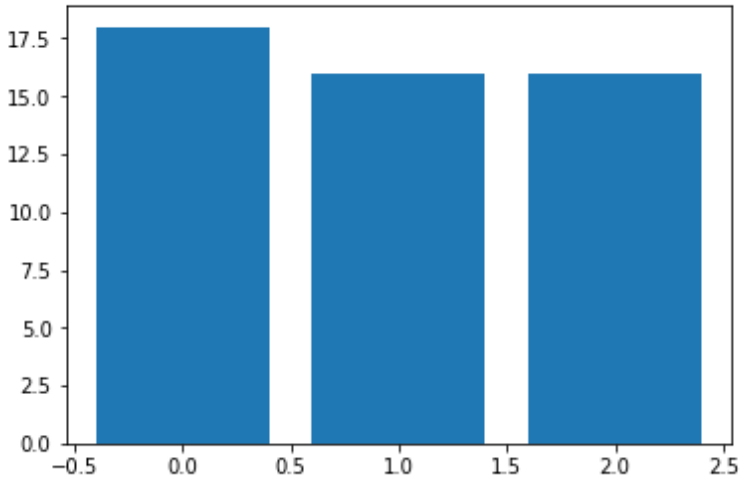
categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.distance	location.formattedAddress	location.labeledLatLngs	locati
[{'id': '0a951735', 'name': 'B...	False	4f18eb91e4b09594f8603283	Tang 1, 109 Tran Hung Dao	VN	Hà Nội	Việt Nam	NaN	904	[Tang 1, 109 Tran Hung Dao, Hà Nội, Thành Phố ...	[{'label': 'display', 'lat': 21.0243052003438,...	21.0
[{'id': '30941735', 'name': 'B...	False	506e2f53498e101fc841c69d	21 Cát Linh	VN	Hà Nội	Việt Nam	NaN	361	[21 Cát Linh, Hà Nội, Thành Phố Hà Nội, Việt Nam]	[{'label': 'display', 'lat': 21.02869266808945...	21.0
[{'id': '24941735', 'name': 'O...	False	4e5efd9a18a870f60f3432e7	NaN	VN	NaN	Việt Nam	NaN	158	[Thành Phố Hà Nội, Việt Nam]	[{'label': 'display', 'lat': 21.029184, 'lng':...	21.0
[{'id': '0a951735', 'name': 'B...	False	4e76b743b0fb968033c2ecdd	14 Lê Thái Tổ	VN	Hoàn Kiếm	Việt Nam	NaN	1770	[14 Lê Thái Tổ, Hoàn Kiếm, Thành Phố Hà Nội, V...	[{'label': 'display', 'lat': 21.02949629231906...	21.0
[]	False	5262307611d2566daa3a0727	NaN	VN	NaN	Việt Nam	NaN	340	[Việt Nam]	[{'label': 'display', 'lat': 21.028783, 'lng':...	21.0
[{'id':											

Although Foursquare return JSON format, but it only raw data sow I need use python script to process data to my work. I am attack full source code in Github to verify.

Each of Bank endpoint I use loop process to get all venues nearby bank location. Foursquare returns many categories. However, it has not much venues in Ha Noi. Therefore, I need sum technical to reduce missing feature and use total venues by category to cluster.

	Asian Restaurant	BBQ Joint	Bakery	Bank	Boutique	Building	Café	Clothing Store	Coffee Shop	Convenience Store	...	Shoe Store	Snack Place	Spa	Tea Room	VenuesID	VenuesName	Vietnamese Restaurant
0	7.0	0.0	1.0	8.0	1.0	6.0	14	2.0	3.0	2.0	...	2.0	1.0	1.0	2.0	4f18eb91e4b09594f8603283	GP Bank	14
1	4.0	1.0	1.0	3.0	1.0	5.0	11	2.0	2.0	0.0	...	1.0	1.0	1.0	3.0	506e2f53498e101fc841c69d	MB Bank - Head Office	4
2	4.0	2.0	1.0	2.0	0.0	1.0	11	1.0	3.0	1.0	...	2.0	1.0	4.0	4.0	4e5efd9a18a870f60f3432e7	Việt Tín Bank	14
3	2.0	0.0	0.0	8.0	0.0	2.0	17	1.0	7.0	1.0	...	0.0	0.0	2.0	3.0	4e76b743b0fb968033c2ecdd	ANZ Bank	10
4	4.0	2.0	2.0	3.0	1.0	5.0	11	2.0	3.0	0.0	...	1.0	1.0	2.0	3.0	5262307611d2566daa3a0727	MB Bank	6
5	3.0	2.0	5.0	5.0	0.0	8.0	15	2.0	3.0	0.0	...	1.0	1.0	0.0	2.0	5122f262e4b04244022e651a	Shinhan Bank Lotte Centre	9
6	3.0	2.0	2.0	0.0	1.0	4.0	11	3.0	5.0	1.0	...	1.0	1.0	4.0	0.0	4e6ae657b993d5439f84bba6	ACB Bank - Giang Vo	14
7	3.0	2.0	1.0	3.0	5.0	6.0	8	1.0	3.0	0.0	...	0.0	0.0	4.0	2.0	4e97ac6577c8d1cccb6cb711	TechcomBank	12
8	1.0	1.0	0.0	1.0	0.0	3.0	7	1.0	2.0	0.0	...	0.0	1.0	4.0	0.0	52a5701d498edfa4a88ee6ba	TP Bank	13

I will use this information to cluster each of place by K-Mean algorithm. Because of missing data in Vietnam from Foursquare, I only have 3-cluster segment. If I have enough data I can elbow method it ensured me the degree for optimum k of the K-Means. Before run K-Mean cluster, I need clean all input by transform, mapping and standard by Sklearn StandardScaler method

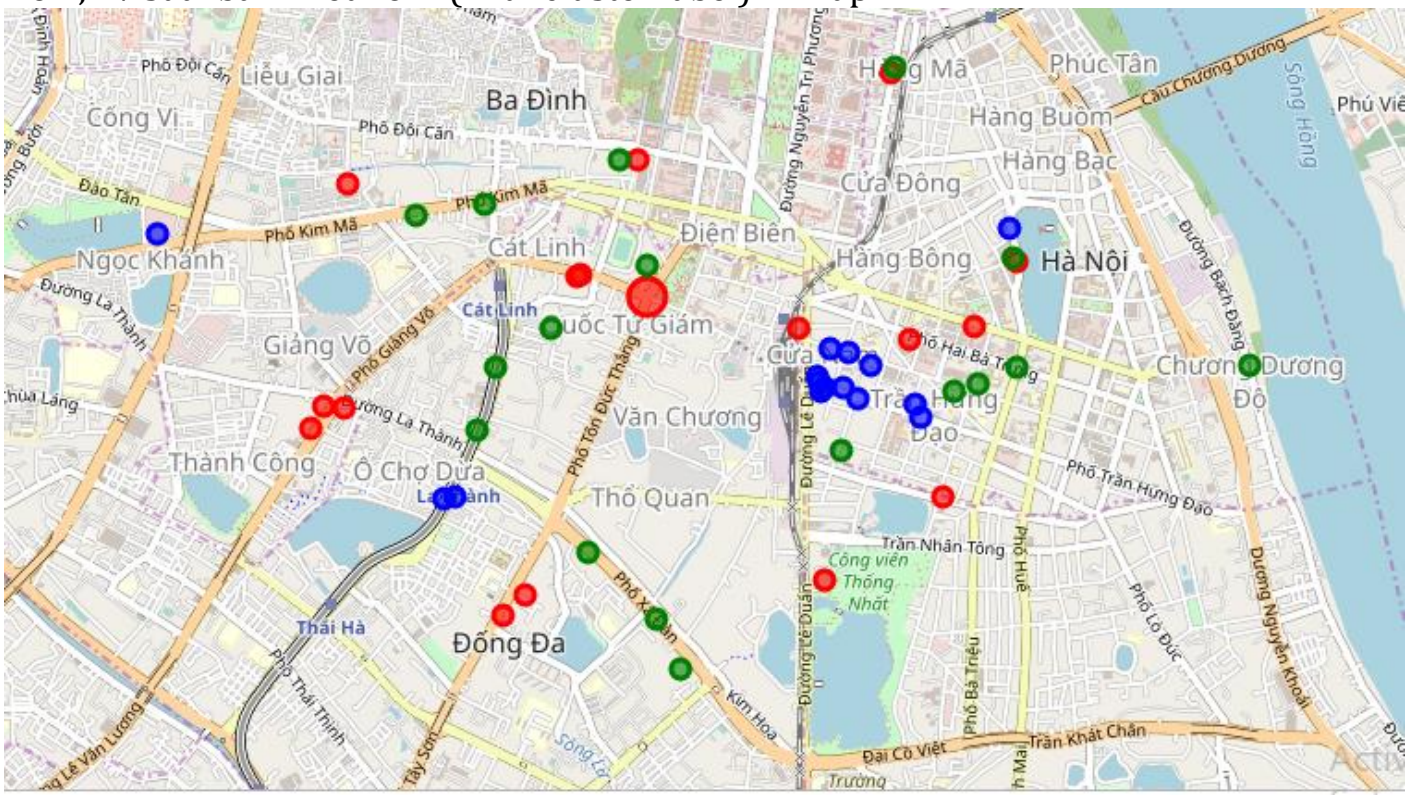


This is chart number of bank venue by cluster auto by K-Mean.

DS

PROJECT DETAIL

Now, I visual bank network (with cluster label) in map



From map, I can see all transaction place of bank locate at center. Blue terminal is place with many office and travel place. K-Mean algorithm some place it far away center. That is a signal to review network transaction at that place.

Base on list of transaction place auto cluster by K-Mean, I need merge with my data in the bank for adjust. Example: position of transaction place, marketing by place customer segment, characteristics of each cluster to support deciding transaction number, transaction position of bank in Ha Noi

CONCLUSIONS

In this project, I only demonstration my knowledge, that I take after nine courses of IBM Data Science Professional Certificate. It include methodology of data science, python technical to process data, visualization data, using model to support business, the machine learning technical and many tool support for data science: IBM Cloud, Foursquare, many library. It may be do not have enough to apply because missing of data, missing of business requirement, and not enough to research.

Thank Coursera and Instructor to help me complete this course. That is giving me very interesting time and much love about data science.

The best interesting thing of me in this course: Data Science is the sexiest job in 21st century.