

INVESTING A REAL ESTATE FOR THE STARTUP "BLACK GARLIC CAFE CHAIN" IN HUE CITY, VIETNAM

1. Introduction

Thừa Thiên-Hue is a province in the North Central Coast region of Vietnam, approximately in the centre of the country. It borders Quảng Trị Province to the north and Đà Nẵng to the south, Laos to the west and the East Sea to the east. The province has 128 km of coastline, 22,000 ha of lagoons and over 200,000 ha of forest. There is an extensive complex of imperial tombs and temples in Hue.



Hue is intimately connected to the imperial Nguyễn Dynasty, based in Hue, who ruled from 1802 to 1945, when the Emperor Bao Dai abdicated in favor of Ho Chi Minh's revolutionary government.



Hue is also famous for its delicious sweet desserts such as Lotus seeds sweet soups, Lotus seed wrapped in logan sweet soup, Areca flower sweet soup, Grilled pork wrapped in cassava flour sweet soup, and Green sticky rice sweet soup.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Hue, Vietnam to open a new Cafe chain. However, the choices are severely affected by the restrictions placed by misunderstanding and speculative bubbles.



Source: Internet

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Hue, Vietnam, if a property developer is looking to open a new coffee chain, where would you recommend that they open it?

Target Audience

This project is particularly useful to property developers and investors looking to open or invest in new coffee chain in the capital city of Hue. This project is timely as the city is currently suffering from oversupply of coffee shop.

- Business personnel who wants to invest or open a cafe chain. This analysis will be a comprehensive guide to start or expand cafe business targeting the large pool of office workers in Hue during working hours.
- Freelancer who loves to have their own cafe shop as a side business. This analysis will give an idea, how beneficial it is to open a cafe store and what are the pros and cons of this business.
- Budding Data Scientists, who wants to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it and, finally be able to tell a story out of it.

Conclusion 1st Week:

Description of Problem and Data Preparation

```

Data Science Capstone Course

[ ] 1 import numpy as np # library to handle data in a vectorized manner
    2
    3 import pandas as pd # library for data analysis
    4 pd.set_option("display.max_columns", None)
    5 pd.set_option("display.max_rows", None)
    6
    7 import json # library to handle JSON files
    8
    9 from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
   10
   11 import requests # library to handle requests
   12 from bs4 import BeautifulSoup # library to parse HTML and XML documents
   13
   14 from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe
   15
   16 # Matplotlib and associated plotting modules
   17 import matplotlib.cm as cm
   18 import matplotlib.colors as colors
   19 import matplotlib.pyplot as plt
   20
   21 # import k-means from clustering stage
   22 from sklearn.cluster import KMeans
   23
   24 import folium # map rendering library

[ ] 1 #lây data từ wiki - https://utinokati.com/en/details/land-market-value/area/Tokyb/
    2 #data = requests.get('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M').text
    3 #data = requests.get('https://thuathienhue.gov.vn/vi-vn/Tra-cuu-gia-dat').text
    4 #data = requests.get('https://www.geopostcodes.com/Hue').text
    5 data = requests.get('https://vi.wikipedia.org/wiki/Th%E1%BB%83_lo%E1%BA%A1i:X%C3%A3,ph%C6%B0%E1%BB%9Dng_thu%E1%BB%99c_th%C3%A0nh_ph%E1%BB%91_Hu%E1%BA%8F').text
  
```

We get the Initial Data-Frame with Names of Major Wards, and corresponding wards in those Major Wards

```
[ ] 1 #chuyển data thành đối tượng object của gói beautifulsoup
2 soup = BeautifulSoup(data,'html.parser')

[ ] 1 #Tìm tất cả các [hàng (dòng) thẻ 'tr':table row] - thẻ 'td':table data: là các cell
2 #table = soup.find('tbody')
3 #table
4 #new_feed = soup.find('table', class_='list2')
5 #new_feed.find('thead').find_all('tr')[0].find_all('th')[0].text.rstrip('\r\n')
6 new_feed = soup.find_all('li')

[ ] 1 phungs = []
2 for p in range(0,28):
3     text=new_feed[p].text.split()
4     if(len(text)>1):
5         text = text[0]+" "+text[1]
6         phungs.append(text.rstrip(","))
7         #text1 = text.rstrip("(phường)")
8         #text2 = text.rstrip("Huế")
9     #phungs.append('Thành Phố Huế')
10 phungs[24]='Vỹ Dạ'

[ ] 1 df=pd.DataFrame(phungs,columns=['Ward'])
2 df.index = np.arange(1, len(df) + 1) # reset the index so that it starts from 1.
3 df.head()
```

	Ward
1	An Cựu
2	An Đông
3	An Hòa
4	An Tây
5	Hương Long

```
+ Code + Text

[ ] 1 wards = []
2 latitudes = []
3 longitudes = []
4 geolocator = Nominatim(user_agent="Thua Thien Hue, Vietnam")
5 for p in phungs:
6     if(p not in ('An Tây','Phú Cát','Phú Hậu','Phú Hiệp','Phú Hội','Phú Nhuận','Phước Vĩnh','Phường Đức','Trường An','Xuân Phú')):
7         address = p+',Thừa Thiên Huế, Việt Nam'#'thanh pho Hue, Thua Thien Hue, Vietnam' - 'Phường Kim Long, Thành phố Huế, Thừa Thiên Huế, Việt Nam'
8         #print(address)
9         location = geolocator.geocode(address)
10        latitude = str(location.latitude)
11        longitude = str(location.longitude)
12        wards.append(p)
13        latitudes.append(latitude)
14        longitudes.append(longitude)
15        #print('The geographical coordinates of {} are {}, {}'.format(address,latitude, longitude))
16
17 df=pd.DataFrame(wards,columns=['Ward'])
18 df['Latitude']=latitudes#str(latitudes)#map(str,latitudes)
19 df['Longitude']=longitudes#str(longitudes)#map(str,longitudes)
20 #df.index = np.arange(1, len(df) + 1) # reset the index so that it starts from 1.
21 #df.dtypes

[ ] 1 df

[ ] 1 df
2 df
3 df
4 df
5 df
6 df
7 df
8 df
9 df
10 df
11 df
12 df
13 df
14 df
15 df
16 df
17 df
18 df
19 df
20 df
21 df
22 df
23 df
24 df
25 df
26 df
27 df
28 df
29 df
30 df
31 df
32 df
33 df
34 df
35 df
36 df
37 df
38 df
39 df
40 df
41 df
42 df
43 df
44 df
45 df
46 df
47 df
48 df
49 df
50 df
51 df
52 df
53 df
54 df
55 df
56 df
57 df
58 df
59 df
60 df
61 df
62 df
63 df
64 df
65 df
66 df
67 df
68 df
69 df
70 df
71 df
72 df
73 df
74 df
75 df
76 df
77 df
78 df
79 df
80 df
81 df
82 df
83 df
84 df
85 df
86 df
87 df
88 df
89 df
90 df
91 df
92 df
93 df
94 df
95 df
96 df
97 df
98 df
99 df
100 df
101 df
102 df
103 df
104 df
105 df
106 df
107 df
108 df
109 df
110 df
111 df
112 df
113 df
114 df
115 df
116 df
117 df
118 df
119 df
120 df
121 df
122 df
123 df
124 df
125 df
126 df
127 df
128 df
129 df
130 df
131 df
132 df
133 df
134 df
135 df
136 df
137 df
138 df
139 df
140 df
141 df
142 df
143 df
144 df
145 df
146 df
147 df
148 df
149 df
150 df
151 df
152 df
153 df
154 df
155 df
156 df
157 df
158 df
159 df
160 df
161 df
162 df
163 df
164 df
165 df
166 df
167 df
168 df
169 df
170 df
171 df
172 df
173 df
174 df
175 df
176 df
177 df
178 df
179 df
180 df
181 df
182 df
183 df
184 df
185 df
186 df
187 df
188 df
189 df
190 df
191 df
192 df
193 df
194 df
195 df
196 df
197 df
198 df
199 df
200 df
201 df
202 df
203 df
204 df
205 df
206 df
207 df
208 df
209 df
210 df
211 df
212 df
213 df
214 df
215 df
216 df
217 df
218 df
219 df
220 df
221 df
222 df
223 df
224 df
225 df
226 df
227 df
228 df
229 df
230 df
231 df
232 df
233 df
234 df
235 df
236 df
237 df
238 df
239 df
240 df
241 df
242 df
243 df
244 df
245 df
246 df
247 df
248 df
249 df
250 df
251 df
252 df
253 df
254 df
255 df
256 df
257 df
258 df
259 df
260 df
261 df
262 df
263 df
264 df
265 df
266 df
267 df
268 df
269 df
270 df
271 df
272 df
273 df
274 df
275 df
276 df
277 df
278 df
279 df
280 df
281 df
282 df
283 df
284 df
285 df
286 df
287 df
288 df
289 df
290 df
291 df
292 df
293 df
294 df
295 df
296 df
297 df
298 df
299 df
300 df
301 df
302 df
303 df
304 df
305 df
306 df
307 df
308 df
309 df
310 df
311 df
312 df
313 df
314 df
315 df
316 df
317 df
318 df
319 df
320 df
321 df
322 df
323 df
324 df
325 df
326 df
327 df
328 df
329 df
330 df
331 df
332 df
333 df
334 df
335 df
336 df
337 df
338 df
339 df
340 df
341 df
342 df
343 df
344 df
345 df
346 df
347 df
348 df
349 df
350 df
351 df
352 df
353 df
354 df
355 df
356 df
357 df
358 df
359 df
360 df
361 df
362 df
363 df
364 df
365 df
366 df
367 df
368 df
369 df
370 df
371 df
372 df
373 df
374 df
375 df
376 df
377 df
378 df
379 df
380 df
381 df
382 df
383 df
384 df
385 df
386 df
387 df
388 df
389 df
390 df
391 df
392 df
393 df
394 df
395 df
396 df
397 df
398 df
399 df
400 df
401 df
402 df
403 df
404 df
405 df
406 df
407 df
408 df
409 df
410 df
411 df
412 df
413 df
414 df
415 df
416 df
417 df
418 df
419 df
420 df
421 df
422 df
423 df
424 df
425 df
426 df
427 df
428 df
429 df
430 df
431 df
432 df
433 df
434 df
435 df
436 df
437 df
438 df
439 df
440 df
441 df
442 df
443 df
444 df
445 df
446 df
447 df
448 df
449 df
450 df
451 df
452 df
453 df
454 df
455 df
456 df
457 df
458 df
459 df
460 df
461 df
462 df
463 df
464 df
465 df
466 df
467 df
468 df
469 df
470 df
471 df
472 df
473 df
474 df
475 df
476 df
477 df
478 df
479 df
480 df
481 df
482 df
483 df
484 df
485 df
486 df
487 df
488 df
489 df
490 df
491 df
492 df
493 df
494 df
495 df
496 df
497 df
498 df
499 df
500 df
501 df
502 df
503 df
504 df
505 df
506 df
507 df
508 df
509 df
510 df
511 df
512 df
513 df
514 df
515 df
516 df
517 df
518 df
519 df
520 df
521 df
522 df
523 df
524 df
525 df
526 df
527 df
528 df
529 df
530 df
531 df
532 df
533 df
534 df
535 df
536 df
537 df
538 df
539 df
540 df
541 df
542 df
543 df
544 df
545 df
546 df
547 df
548 df
549 df
550 df
551 df
552 df
553 df
554 df
555 df
556 df
557 df
558 df
559 df
560 df
561 df
562 df
563 df
564 df
565 df
566 df
567 df
568 df
569 df
570 df
571 df
572 df
573 df
574 df
575 df
576 df
577 df
578 df
579 df
580 df
581 df
582 df
583 df
584 df
585 df
586 df
587 df
588 df
589 df
590 df
591 df
592 df
593 df
594 df
595 df
596 df
597 df
598 df
599 df
600 df
601 df
602 df
603 df
604 df
605 df
606 df
607 df
608 df
609 df
610 df
611 df
612 df
613 df
614 df
615 df
616 df
617 df
618 df
619 df
620 df
621 df
622 df
623 df
624 df
625 df
626 df
627 df
628 df
629 df
630 df
631 df
632 df
633 df
634 df
635 df
636 df
637 df
638 df
639 df
640 df
641 df
642 df
643 df
644 df
645 df
646 df
647 df
648 df
649 df
650 df
651 df
652 df
653 df
654 df
655 df
656 df
657 df
658 df
659 df
660 df
661 df
662 df
663 df
664 df
665 df
666 df
667 df
668 df
669 df
670 df
671 df
672 df
673 df
674 df
675 df
676 df
677 df
678 df
679 df
680 df
681 df
682 df
683 df
684 df
685 df
686 df
687 df
688 df
689 df
690 df
691 df
692 df
693 df
694 df
695 df
696 df
697 df
698 df
699 df
700 df
701 df
702 df
703 df
704 df
705 df
706 df
707 df
708 df
709 df
710 df
711 df
712 df
713 df
714 df
715 df
716 df
717 df
718 df
719 df
720 df
721 df
722 df
723 df
724 df
725 df
726 df
727 df
728 df
729 df
730 df
731 df
732 df
733 df
734 df
735 df
736 df
737 df
738 df
739 df
740 df
741 df
742 df
743 df
744 df
745 df
746 df
747 df
748 df
749 df
750 df
751 df
752 df
753 df
754 df
755 df
756 df
757 df
758 df
759 df
760 df
761 df
762 df
763 df
764 df
765 df
766 df
767 df
768 df
769 df
770 df
771 df
772 df
773 df
774 df
775 df
776 df
777 df
778 df
779 df
780 df
781 df
782 df
783 df
784 df
785 df
786 df
787 df
788 df
789 df
790 df
791 df
792 df
793 df
794 df
795 df
796 df
797 df
798 df
799 df
800 df
801 df
802 df
803 df
804 df
805 df
806 df
807 df
808 df
809 df
810 df
811 df
812 df
813 df
814 df
815 df
816 df
817 df
818 df
819 df
820 df
821 df
822 df
823 df
824 df
825 df
826 df
827 df
828 df
829 df
830 df
831 df
832 df
833 df
834 df
835 df
836 df
837 df
838 df
839 df
840 df
841 df
842 df
843 df
844 df
845 df
846 df
847 df
848 df
849 df
850 df
851 df
852 df
853 df
854 df
855 df
856 df
857 df
858 df
859 df
860 df
861 df
862 df
863 df
864 df
865 df
866 df
867 df
868 df
869 df
870 df
871 df
872 df
873 df
874 df
875 df
876 df
877 df
878 df
879 df
880 df
881 df
882 df
883 df
884 df
885 df
886 df
887 df
888 df
889 df
890 df
891 df
892 df
893 df
894 df
895 df
896 df
897 df
898 df
899 df
900 df
901 df
902 df
903 df
904 df
905 df
906 df
907 df
908 df
909 df
910 df
911 df
912 df
913 df
914 df
915 df
916 df
917 df
918 df
919 df
920 df
921 df
922 df
923 df
924 df
925 df
926 df
927 df
928 df
929 df
930 df
931 df
932 df
933 df
934 df
935 df
936 df
937 df
938 df
939 df
940 df
941 df
942 df
943 df
944 df
945 df
946 df
947 df
948 df
949 df
950 df
951 df
952 df
953 df
954 df
955 df
956 df
957 df
958 df
959 df
960 df
961 df
962 df
963 df
964 df
965 df
966 df
967 df
968 df
969 df
970 df
971 df
972 df
973 df
974 df
975 df
976 df
977 df
978 df
979 df
980 df
981 df
982 df
983 df
984 df
985 df
986 df
987 df
988 df
989 df
990 df
991 df
992 df
993 df
994 df
995 df
996 df
997 df
998 df
999 df
1000 df
```

So as the next step we will use Foursquare data and obtain information on coffee chains. With these, we can start with our battle of neighborhoods for opening a coffee shop in Hue city.

2. Initial Data Preparation

2.1. Web-Scrapping and Cleaning

2.1.1. Get The Names of Wards from Wikipedia

The Wikipedia page of Hue Wards contains the table of 27 wards of Hue. I have used BeautifulSoup4 and pandas library to create the initial data-frame. For a clean and understandable data-frame some of the wards are renamed for example ‘Vĩ Dạ’ to Vỹ Dạ. Here I have taken the first entry of the major ward column in the Wikipedia table to concentrate on. Even though not complete but it gives us quite a detailed picture of the corresponding ward, as later on I have considered top most venues within 500 meter radius of the major ward. After this initial preparation, I moved on to the next step to obtain coordinates using Geopy library.

2.1.2. Get the Coordinates of the Major Wards

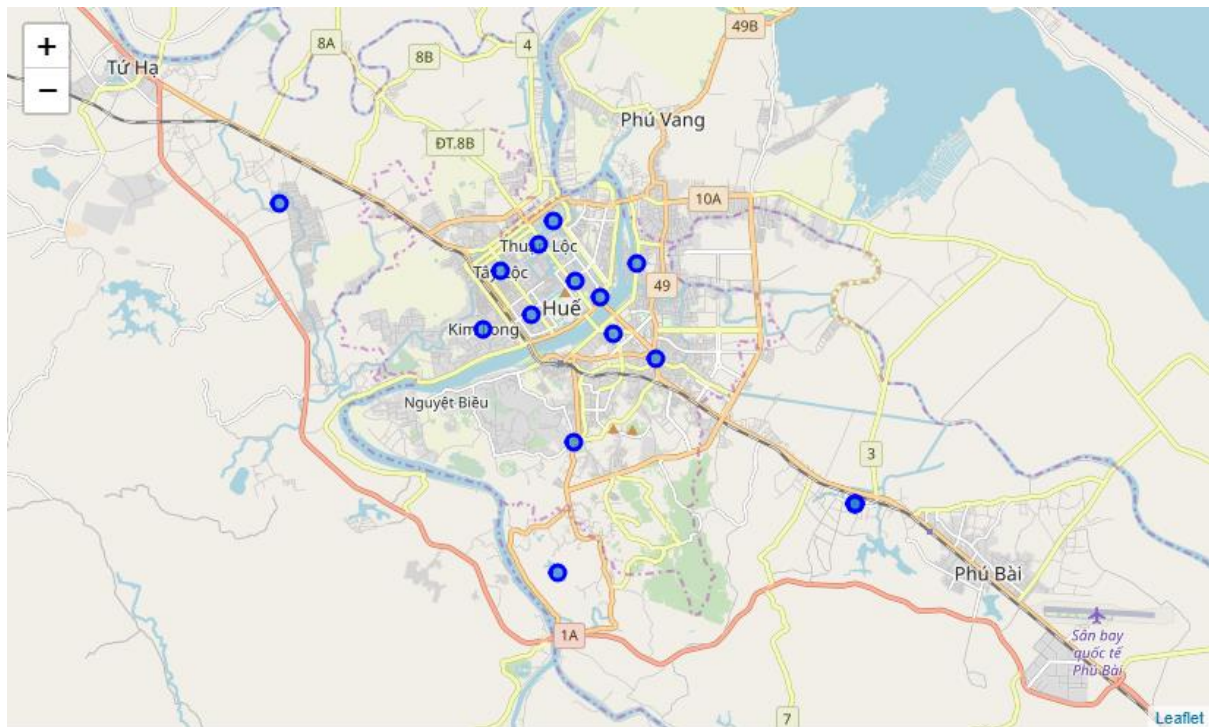
Some of the coo-ordinates of the major wards returned by Geopy are wrong and I have figured out the reason for this is the name of the major wards in the data-frame are different from their actual names. In these cases (4 of them), I had to google search and replace using pandas library. After little manipulation the obtained data-frame looks as below after remove several wards which could not receive locations from code “geolocator.geocode(address)”, I guess that geocoders package did update it’s data.

[] 1 df

	Ward	Latitude	Longitude
0	An Cựu	16.4574892	107.6002023
1	An Đông	16.4248801	107.6469586
2	An Hòa	16.4248801	107.6469586
3	Hương Long	16.4248801	107.6469586
4	Hương Sơ	16.4924722	107.5119899
5	Kim Long	16.4642271	107.559751
6	Phú Bình	16.4885976	107.5762144
7	Phú Hòa	16.471216	107.5868958
8	Phú Thuận	16.4248801	107.6469586
9	Tây Lộc	16.4772416	107.5637146
10	Thuận Hòa	16.4674072	107.5710217
11	Thuận Lộc	16.4831435	107.572616
12	Thuận Thành	16.4748538	107.5812484
13	Thủy Biều	16.4094987	107.5770314
14	Thủy Xuân	16.4388	107.5809431
15	Vỹ Dạ	16.4790295	107.5958008
16	Vĩnh Ninh	16.4631912	107.5901125

2.2. Foursquare Data

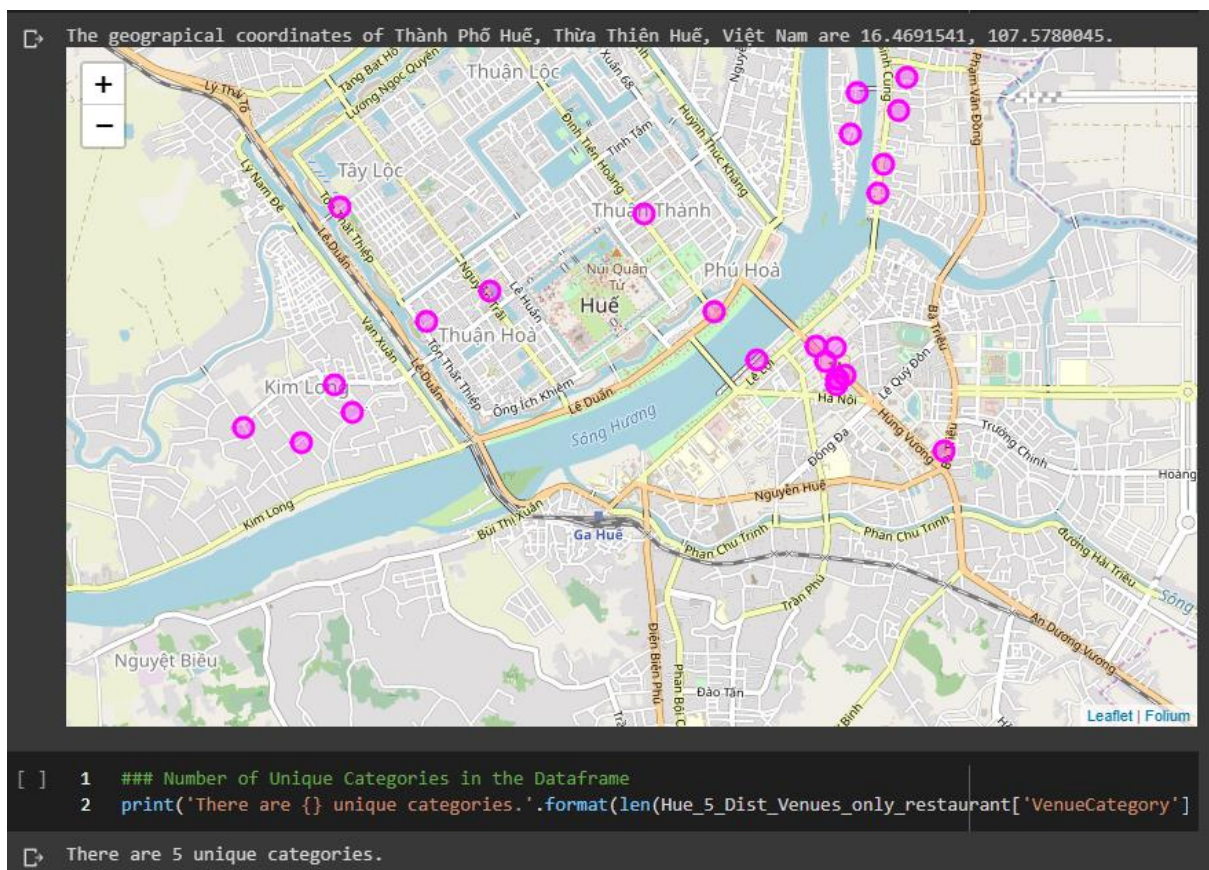
Finally, I make use of Foursquare API to obtain the 100 most common venues within 1000 meter of each wards.



3. Exploring the Data

3.1. Exploring the Wards of Hue

From the Foursquare data, we could see that there are 5 unique categories of Restaurant and 2 unique categories of coffee

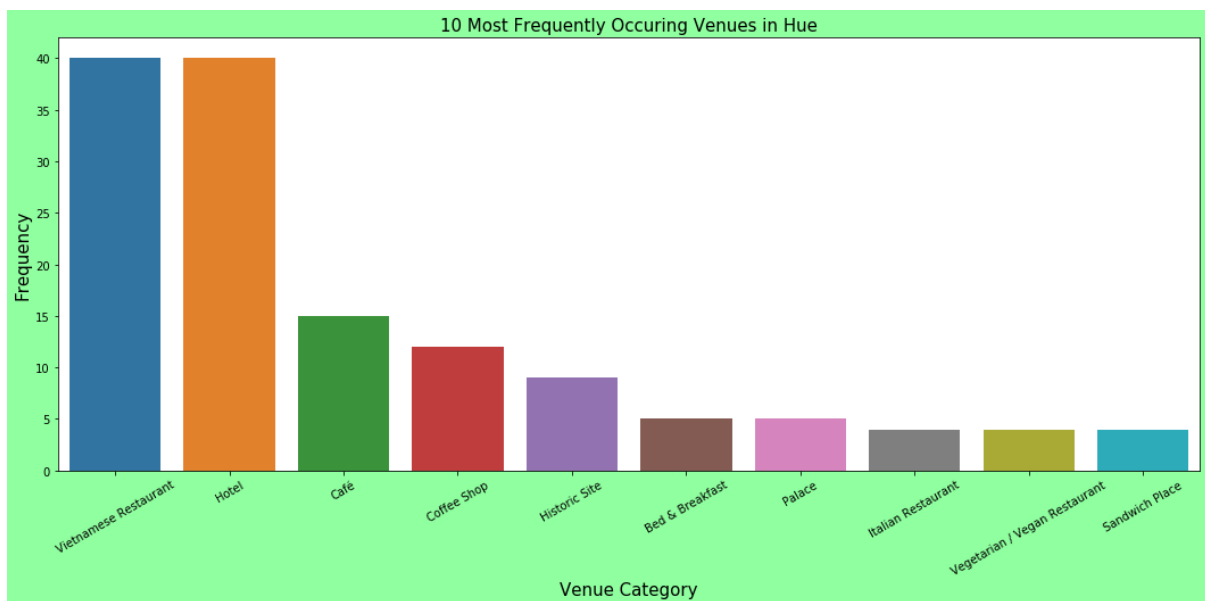


```
[40] 1 print (Hue_Venues_Coffee_Cafe ['VenueCategory'].value_counts())
```

Café	15
Coffee Shop	12

Here we have found out that

- Vietnamese restaurants are top the charts of most common venues in top 10 categories, followed by hotels, café and Coffee Shop
- A plot of the ten most frequent venues in Hue are as below



Since this reminded of Hue restaurant and coffee so you need to take a break







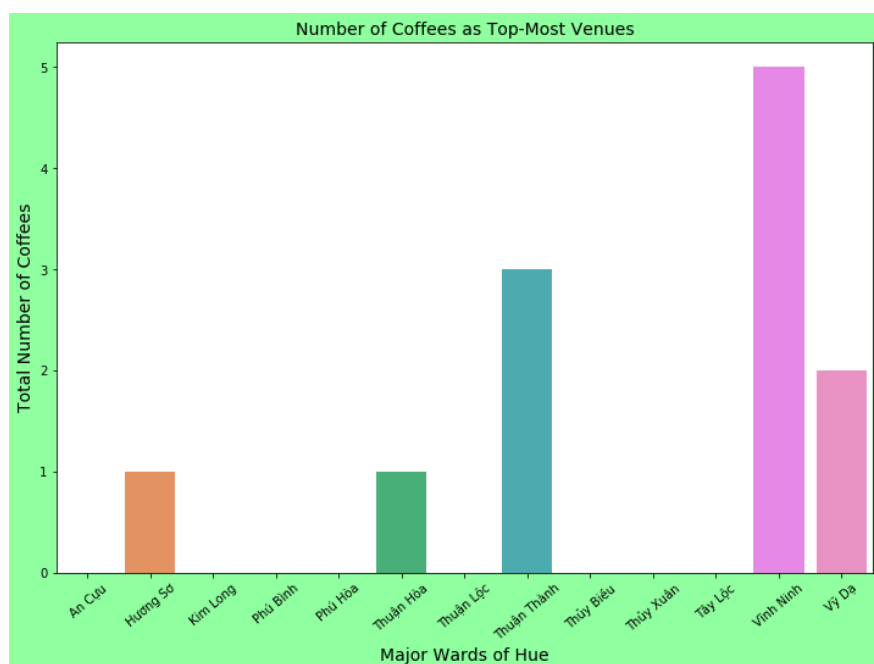
Next step was to obtain information about the top 5 venues of each district. And to do that, I proceed as follows

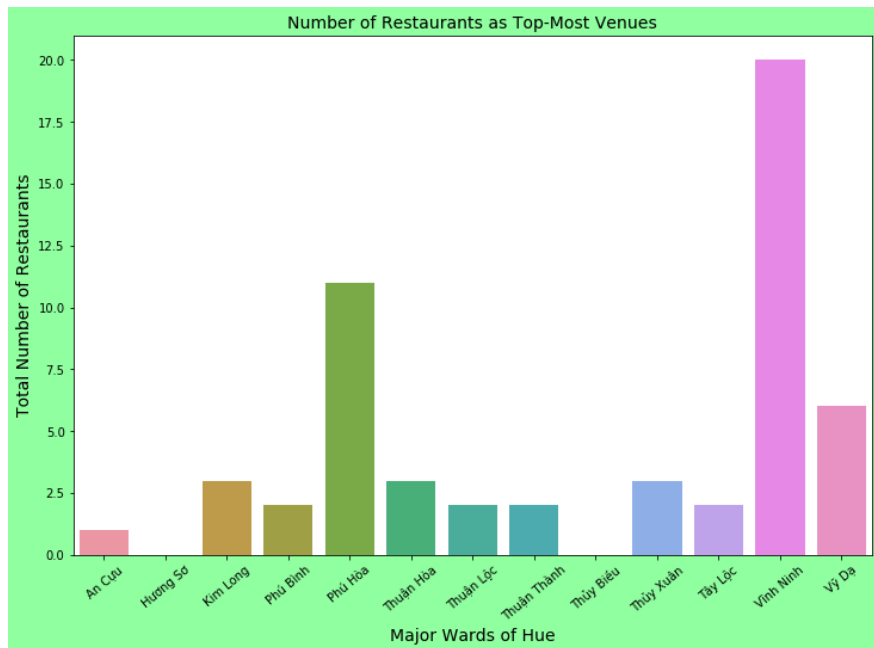
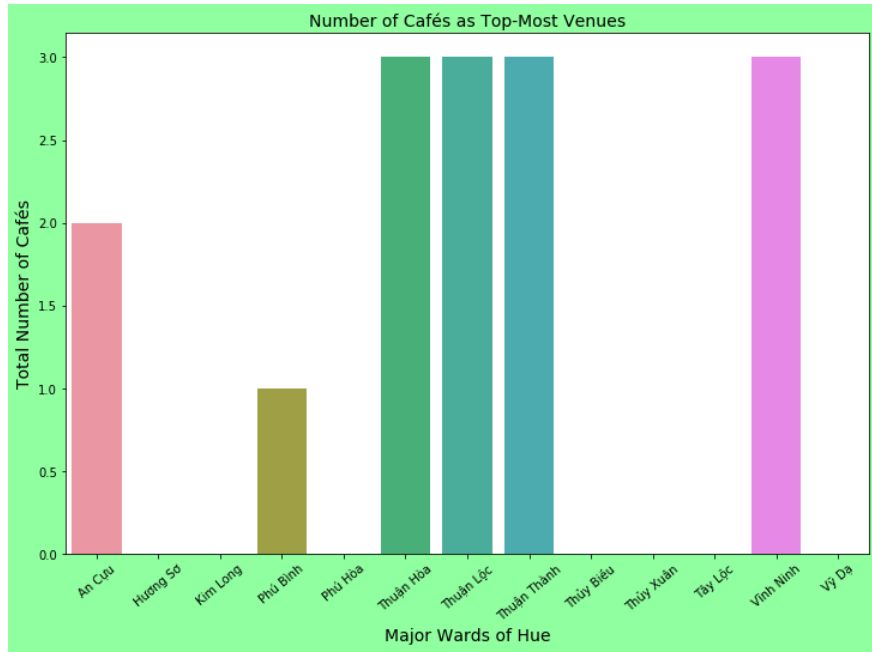
- Create a data-frame with pandas one hot encoding for the venue categories.
- Use pandas groupby on District column and obtain the mean.
- Transpose the data-frame at step 2 and arrange in descending order.

Implementing them in Pandas outputs the following:

<p>-----An Cựu-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Café</td><td>0.25</td></tr> <tr><td>1</td><td>Movie Theater</td><td>0.25</td></tr> <tr><td>2</td><td>Hotel</td><td>0.12</td></tr> <tr><td>3</td><td>Vegetarian / Vegan Restaurant</td><td>0.12</td></tr> <tr><td>4</td><td>Bed & Breakfast</td><td>0.12</td></tr> </tbody> </table>		Venue	Freq	0	Café	0.25	1	Movie Theater	0.25	2	Hotel	0.12	3	Vegetarian / Vegan Restaurant	0.12	4	Bed & Breakfast	0.12	<p>-----Hương Sơ-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Coffee Shop</td><td>1.0</td></tr> <tr><td>1</td><td>Art Gallery</td><td>0.0</td></tr> <tr><td>2</td><td>Hotel</td><td>0.0</td></tr> <tr><td>3</td><td>Indian Restaurant</td><td>0.0</td></tr> <tr><td>4</td><td>Italian Restaurant</td><td>0.0</td></tr> </tbody> </table>		Venue	Freq	0	Coffee Shop	1.0	1	Art Gallery	0.0	2	Hotel	0.0	3	Indian Restaurant	0.0	4	Italian Restaurant	0.0	<p>-----Thuận Hòa-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Café</td><td>0.20</td></tr> <tr><td>1</td><td>Vietnamese Restaurant</td><td>0.13</td></tr> <tr><td>2</td><td>Palace</td><td>0.13</td></tr> <tr><td>3</td><td>Juice Bar</td><td>0.07</td></tr> <tr><td>4</td><td>Historic Site</td><td>0.07</td></tr> </tbody> </table>		Venue	Freq	0	Café	0.20	1	Vietnamese Restaurant	0.13	2	Palace	0.13	3	Juice Bar	0.07	4	Historic Site	0.07
	Venue	Freq																																																						
0	Café	0.25																																																						
1	Movie Theater	0.25																																																						
2	Hotel	0.12																																																						
3	Vegetarian / Vegan Restaurant	0.12																																																						
4	Bed & Breakfast	0.12																																																						
	Venue	Freq																																																						
0	Coffee Shop	1.0																																																						
1	Art Gallery	0.0																																																						
2	Hotel	0.0																																																						
3	Indian Restaurant	0.0																																																						
4	Italian Restaurant	0.0																																																						
	Venue	Freq																																																						
0	Café	0.20																																																						
1	Vietnamese Restaurant	0.13																																																						
2	Palace	0.13																																																						
3	Juice Bar	0.07																																																						
4	Historic Site	0.07																																																						
<p>-----Thuận Lộc-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Café</td><td>0.33</td></tr> <tr><td>1</td><td>Vietnamese Restaurant</td><td>0.11</td></tr> <tr><td>2</td><td>Vegetarian / Vegan Restaurant</td><td>0.11</td></tr> <tr><td>3</td><td>Souvenir Shop</td><td>0.11</td></tr> <tr><td>4</td><td>Historic Site</td><td>0.11</td></tr> </tbody> </table>		Venue	Freq	0	Café	0.33	1	Vietnamese Restaurant	0.11	2	Vegetarian / Vegan Restaurant	0.11	3	Souvenir Shop	0.11	4	Historic Site	0.11	<p>-----Thuận Thành-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Historic Site</td><td>0.17</td></tr> <tr><td>1</td><td>Café</td><td>0.12</td></tr> <tr><td>2</td><td>Coffee Shop</td><td>0.12</td></tr> <tr><td>3</td><td>Vietnamese Restaurant</td><td>0.08</td></tr> <tr><td>4</td><td>Palace</td><td>0.08</td></tr> </tbody> </table>		Venue	Freq	0	Historic Site	0.17	1	Café	0.12	2	Coffee Shop	0.12	3	Vietnamese Restaurant	0.08	4	Palace	0.08	<p>-----Vỹ Dạ-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Hotel</td><td>0.37</td></tr> <tr><td>1</td><td>Vietnamese Restaurant</td><td>0.21</td></tr> <tr><td>2</td><td>Coffee Shop</td><td>0.11</td></tr> <tr><td>3</td><td>Resort</td><td>0.05</td></tr> <tr><td>4</td><td>BBQ Joint</td><td>0.05</td></tr> </tbody> </table>		Venue	Freq	0	Hotel	0.37	1	Vietnamese Restaurant	0.21	2	Coffee Shop	0.11	3	Resort	0.05	4	BBQ Joint	0.05
	Venue	Freq																																																						
0	Café	0.33																																																						
1	Vietnamese Restaurant	0.11																																																						
2	Vegetarian / Vegan Restaurant	0.11																																																						
3	Souvenir Shop	0.11																																																						
4	Historic Site	0.11																																																						
	Venue	Freq																																																						
0	Historic Site	0.17																																																						
1	Café	0.12																																																						
2	Coffee Shop	0.12																																																						
3	Vietnamese Restaurant	0.08																																																						
4	Palace	0.08																																																						
	Venue	Freq																																																						
0	Hotel	0.37																																																						
1	Vietnamese Restaurant	0.21																																																						
2	Coffee Shop	0.11																																																						
3	Resort	0.05																																																						
4	BBQ Joint	0.05																																																						
<p>-----Vĩnh Ninh-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Hotel</td><td>0.27</td></tr> <tr><td>1</td><td>Vietnamese Restaurant</td><td>0.23</td></tr> <tr><td>2</td><td>Coffee Shop</td><td>0.07</td></tr> <tr><td>3</td><td>Café</td><td>0.04</td></tr> <tr><td>4</td><td>Noodle House</td><td>0.03</td></tr> </tbody> </table>		Venue	Freq	0	Hotel	0.27	1	Vietnamese Restaurant	0.23	2	Coffee Shop	0.07	3	Café	0.04	4	Noodle House	0.03	<p>-----Kim Long-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Vietnamese Restaurant</td><td>0.75</td></tr> <tr><td>1</td><td>Noodle House</td><td>0.25</td></tr> <tr><td>2</td><td>Vegetarian / Vegan Restaurant</td><td>0.00</td></tr> <tr><td>3</td><td>Movie Theater</td><td>0.00</td></tr> <tr><td>4</td><td>Indian Restaurant</td><td>0.00</td></tr> </tbody> </table>		Venue	Freq	0	Vietnamese Restaurant	0.75	1	Noodle House	0.25	2	Vegetarian / Vegan Restaurant	0.00	3	Movie Theater	0.00	4	Indian Restaurant	0.00	<p>-----Phú Bình-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Vietnamese Restaurant</td><td>0.25</td></tr> <tr><td>1</td><td>Historic Site</td><td>0.12</td></tr> <tr><td>2</td><td>Lottery Retailer</td><td>0.12</td></tr> <tr><td>3</td><td>Boat or Ferry</td><td>0.12</td></tr> <tr><td>4</td><td>Sandwich Place</td><td>0.12</td></tr> </tbody> </table>		Venue	Freq	0	Vietnamese Restaurant	0.25	1	Historic Site	0.12	2	Lottery Retailer	0.12	3	Boat or Ferry	0.12	4	Sandwich Place	0.12
	Venue	Freq																																																						
0	Hotel	0.27																																																						
1	Vietnamese Restaurant	0.23																																																						
2	Coffee Shop	0.07																																																						
3	Café	0.04																																																						
4	Noodle House	0.03																																																						
	Venue	Freq																																																						
0	Vietnamese Restaurant	0.75																																																						
1	Noodle House	0.25																																																						
2	Vegetarian / Vegan Restaurant	0.00																																																						
3	Movie Theater	0.00																																																						
4	Indian Restaurant	0.00																																																						
	Venue	Freq																																																						
0	Vietnamese Restaurant	0.25																																																						
1	Historic Site	0.12																																																						
2	Lottery Retailer	0.12																																																						
3	Boat or Ferry	0.12																																																						
4	Sandwich Place	0.12																																																						
<p>-----Phú Hòa-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Hotel</td><td>0.34</td></tr> <tr><td>1</td><td>Vietnamese Restaurant</td><td>0.23</td></tr> <tr><td>2</td><td>Italian Restaurant</td><td>0.06</td></tr> <tr><td>3</td><td>Historic Site</td><td>0.03</td></tr> <tr><td>4</td><td>Sports Bar</td><td>0.03</td></tr> </tbody> </table>		Venue	Freq	0	Hotel	0.34	1	Vietnamese Restaurant	0.23	2	Italian Restaurant	0.06	3	Historic Site	0.03	4	Sports Bar	0.03	<p>-----Thủy Biều-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Art Gallery</td><td>0.33</td></tr> <tr><td>1</td><td>Lake</td><td>0.33</td></tr> <tr><td>2</td><td>Campground</td><td>0.33</td></tr> <tr><td>3</td><td>Night Market</td><td>0.00</td></tr> <tr><td>4</td><td>Indian Restaurant</td><td>0.00</td></tr> </tbody> </table>		Venue	Freq	0	Art Gallery	0.33	1	Lake	0.33	2	Campground	0.33	3	Night Market	0.00	4	Indian Restaurant	0.00	<p>-----Thủy Xuân-----</p> <table> <thead> <tr> <th></th><th>Venue</th><th>Freq</th></tr> </thead> <tbody> <tr><td>0</td><td>Vietnamese Restaurant</td><td>0.50</td></tr> <tr><td>1</td><td>Historic Site</td><td>0.25</td></tr> <tr><td>2</td><td>Seafood Restaurant</td><td>0.25</td></tr> <tr><td>3</td><td>Movie Theater</td><td>0.00</td></tr> <tr><td>4</td><td>Indian Restaurant</td><td>0.00</td></tr> </tbody> </table>		Venue	Freq	0	Vietnamese Restaurant	0.50	1	Historic Site	0.25	2	Seafood Restaurant	0.25	3	Movie Theater	0.00	4	Indian Restaurant	0.00
	Venue	Freq																																																						
0	Hotel	0.34																																																						
1	Vietnamese Restaurant	0.23																																																						
2	Italian Restaurant	0.06																																																						
3	Historic Site	0.03																																																						
4	Sports Bar	0.03																																																						
	Venue	Freq																																																						
0	Art Gallery	0.33																																																						
1	Lake	0.33																																																						
2	Campground	0.33																																																						
3	Night Market	0.00																																																						
4	Indian Restaurant	0.00																																																						
	Venue	Freq																																																						
0	Vietnamese Restaurant	0.50																																																						
1	Historic Site	0.25																																																						
2	Seafood Restaurant	0.25																																																						
3	Movie Theater	0.00																																																						
4	Indian Restaurant	0.00																																																						

To explore which wards has the highest number of Cafés, Coffee shops and restaurants as the most common venue and the plot below is the answer

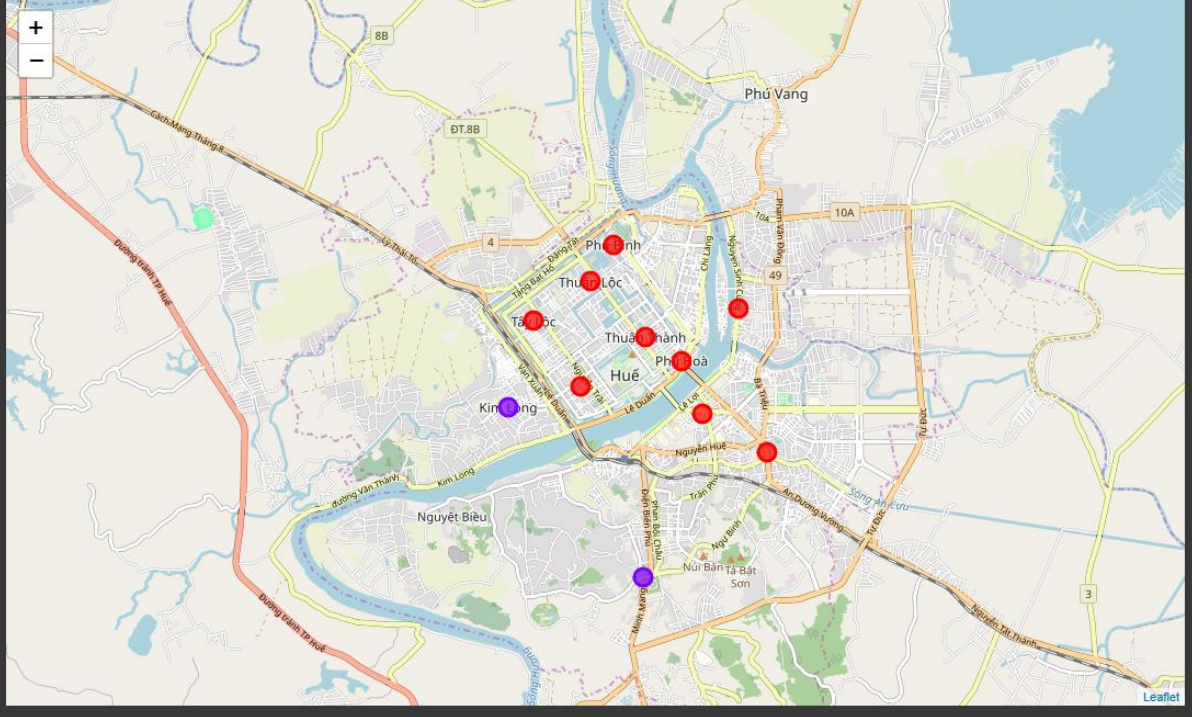




3.2. Clustering the Major Wards of Hue

Finally, we try to cluster these wards based on the frequency of venue categories and, use K-Means clustering. So our expectation would be based on the similarities of venue categories, these wards will be clustered. Using K-Means algorithm from Scikit-learn library we obtain 3 clusters as shown below.

The geographical coordinates of Thành Phố Huế, Thừa Thiên Huế, Việt Nam are 16.4691541, 107.5780045.



Cluster 1

```
[ ] 1 Hue_merged_not_NAN.loc[Hue_merged_not_NAN['Cluster Label'] == 0, Hue_merged_not_NAN.columns[[0]+list(range(4, Hue_merged_not_NAN.shape[1]))]]
```

	Ward	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	An Cựu	Movie Theater	Café	Vegetarian / Vegan Restaurant	Hotel	Bed & Breakfast	Bus Station	Castle	Food Truck	Food Service	Food Court
6	Phú Bình	Vietnamese Restaurant	Sandwich Place	Food Service	Lottery Retailer	Boat or Ferry	Café	Historic Site	Bakery	French Restaurant	Food Truck
7	Phú Hòa	Hotel	Vietnamese Restaurant	Italian Restaurant	BBQ Joint	Bakery	Bar	Bed & Breakfast	Beer Garden	Breakfast Spot	Historic Site
9	Tây Lộc	Food Truck	Shopping Mall	Hotpot Restaurant	Food Court	Fast Food Restaurant	Noodle House	Breakfast Spot	Vietnamese Restaurant	Campground	Food Service
10	Thuận Hòa	Café	Vietnamese Restaurant	Palace	Historic Site	Juice Bar	Coffee Shop	Castle	French Restaurant	Monument / Landmark	Shopping Mall
11	Thuận Lộc	Café	Vietnamese Restaurant	Sandwich Place	Vegetarian / Vegan Restaurant	Food Service	Historic Site	Souvenir Shop	Bar	Coffee Shop	BBQ Joint
12	Thuận Thành	Historic Site	Café	Coffee Shop	Vietnamese Restaurant	Palace	Monument / Landmark	Bookstore	Castle	Creperie	Dessert Shop
13	Thủy Biều	Art Gallery	Lake	Campground	Bakery	Coffee Shop	French Restaurant	Food Truck	Food Service	Food Court	Fast Food Restaurant
15	Vỹ Dạ	Hotel	Vietnamese Restaurant	Coffee Shop	Lounge	BBQ Joint	Vegetarian / Vegan Restaurant	Bed & Breakfast	Resort	Italian Restaurant	Food Service
16	Vinh Ninh	Hotel	Vietnamese Restaurant	Coffee Shop	Café	Bakery	Bed & Breakfast	Noodle House	Breakfast Spot	Dive Bar	Vegetarian / Vegan Restaurant

Cluster 2

```
[ ] 1 Hue_merged_not_NAN.loc[Hue_merged_not_NAN['Cluster Label'] == 1, Hue_merged_not_NAN.columns[[0]+list(range(4, Hue_merged_not_NAN.shape[1]))]]
```

	Ward	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
8	Kim Long	Vietnamese Restaurant	Noodle House	Campground	French Restaurant	Food Truck	Food Service	Food Court	Fast Food Restaurant	Dive Bar	Dessert Shop
14	Thủy Xuân	Vietnamese Restaurant	Seafood Restaurant	Historic Site	Campground	French Restaurant	Food Truck	Food Service	Food Court	Fast Food Restaurant	Dive Bar

Cluster 3

```
[ ] 1 Hue_merged_not_NAN.loc[Hue_merged_not_NAN['Cluster Label'] == 2, Hue_merged_not_NAN.columns[[0]+list(range(4, Hue_merged_not_NAN.shape[1]))]]
```

	Ward	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
4	Hương Sơ	Coffee Shop	Vietnamese Restaurant	Campground	French Restaurant	Food Truck	Food Service	Food Court	Fast Food Restaurant	Dive Bar	Dessert Shop

4. Results

The results of the exploratory data analysis and clustering are summarized below:

- Most of the neighborhoods fall into Cluster 2 which are mostly business areas with Vietnamese Restaurant. Cluster 1 is only 2 1st Most Common Venue with Cafe. In cluster 3, Coffee Shop is the number one of most common Venue
- Restaurants top the charts of most common venues in the wards.
- Phú Bình, Kim Long and Thủy Xuân wards are dominated by Vietnamese restaurants as the the most common venues.
- Võ Dạ, Vĩnh Ninh and Phú Hòa areas are dominated by hotel as most common venues.

5. Discussion

According to this analysis, Vinh Ninh area will provide most competition for an upcoming café chain as hotel is the most common venue in this area and the frequency of café as common venue are very low compared to the remaining wards.

Some drawbacks of this analysis are the clustering is completely based on the most common venues obtained from Foursquare data. Since land price, distance of the venues from closest stations, number of potential customers, benefits and drawbacks of Shinagawa being a port region, could all play a major role and thus, this analysis is definitely far from being conclusory. However, it definitely gives us some very important preliminary information on possibilities of opening restaurants around the major wards of Hue.

Also, another pitfall of this analysis could be consideration of only one major district of each ward of Hue, taking into account of all the areas, wards would give us an even more realistic picture.

Furthermore, this results also could potentially vary if we use some other clustering techniques like DBSCAN.

6. Conclusion

Finally to conclude this project, we have got a small glimpse of how real life data-science projects look like. I have made use of some frequently used python

libraries to scrap web-data, use Foursquare API to explore the major wards of Hue and saw the results of segmentation of wards using Folium leaflet map. Potential for this kind of analysis in a real life business problem is discussed in great detail. Also, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned. Finally, since my analysis were mostly concentrated on the possibilities of opening a café chain targeting the huge pool of office workers, some of the results obtained are surprisingly what I have expected after staying in Hue.