

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

Thừa Thiên-Huế is a province in the North Central Coast region of Vietnam, approximately in the centre of the country. 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 Huế.



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.

Conclusion 1st Week:

Description of **Data** Preparation

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

```

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 # transform 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 l y data t  wiki - https://utinokati.com/en/details/land-market-value/area/Tokyo/
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/Ira-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%A3%20%E1%BA%A1%1%XC%3%A3%20ph%C3%A0nh%20th%E1%BB%9c%20th%C3%A0nh%20ph%E1%BB%91%20h%E1%BA%8F').text

```

```
[ ] 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
```

	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

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.