MALAYSIAN UNIVERSITIES COMPARING

BY THEIR NEARBY SERVICES AND VENUES

INTRODUCTION/BUSINESS PROBLEM:

- By 2019 there were 130,110 international students in Malaysia from 136 countries and Gross Enrolment Ratio in 2016 of 44% is higher than most of the Asian countries, and higher than the world average of 37%.

- In the university campus you can find restaurants but they serve only Malay food although some of the university has international food but it is not that quality, so most of students would go outside for eating or order food from outside using grab food app. Same goes for other facilities like barbershops, universities don't have these ones inside the campus.

THE SOLUTION

- Give the students a clue about the area around the universities so that they can choose the one with more facilities around and easy to be reached or order services from

- Give people who are interested in business a good idea about the area around the university, so that they can choose the best business to invest their money in, since there will be so many students coming to these places because it is near to the campus

DATA ACQUISITION

- We have a Wikipedia page that has 2 tables of public and private universities in Malaysia
- https://en.wikipedia.org/wiki/List_of_universities_in_Malaysia

Universities [edit]

Name in English	Name in Malay	Acronym	Foundation	Location	Link
University of Malaya	Universiti Malaya[8]	UM	1905	Kuala Lumpur Nilam Puri, Kelantan	[1]
International Islamic University Malaysia	Universiti Islam Antarabangsa Malaysia[9]	IIUM	1983	Gombak, Selangor	[2]
National University of Malaysia	Universiti Kebangsaan Malaysia[10]	UKM	1970	Bangi, Selangor	[3]
University of Malaysia Kelantan	Universiti Malaysia Kelantan[11]	UMK	2007	Pengkalan Chepa, Kelantan Jeli, Kelantan Bachok, Kelantan	[4]
University of Malaysia Pahang	Universiti Malaysia Pahang[12]	UMP	2002 (as KUKTEM)	Pekan, Pahang	[5]
University of Malaysia Perlis	Universiti Malaysia Perlis[13]	UniMAP	2001 (as KUKUM)	Arau, Perlis	[6]
University of Malaysia Sabah	Universiti Malaysia Sabah[14]	UMS	1994	Kota Kinabalu, Sabah	[7]
University of Malaysia Sarawak	Universiti Malaysia Sarawak[15]	UNIMAS	1992	Kota Samarahan, Sarawak	[8]

DATA WRANGLING

- website gives the data in 2 tables: Public universities and private universities, and each one of these tables has two sub-tables in it
- Data cleaning:
  - Remove the unwanted rows, and these are the rows that have values
  - The next step was to give the columns meaningful names because they were just numbers
  - Then we edit the name in each cell of name so that we remove the brackets and numbers

FEATURE SELECTION

- Next I remove the unwanted columns, and I chose the features to be: MalayName , Acronym , Location. I chose the Malay name for the sake of consistency with other tables , because other ones have only Malay name without English name, the acronym was selected also because some universities can be found by its acronym

RESULTED DATASET

	MalayName	Acr	Loc
1	Universiti Malaya	UM	Kuala Lumpur Nilam Puri, Kelantan
2	Universiti Islam Antarabangsa Malaysia	IIUM	Gombak, Selangor
3	Universiti Kebangsaan Malaysia	UKM	Bangi, Selangor
4	Universiti Malaysia Kelantan	UMK	Pengkalan Chepa, Kelantan Jeli, Kelantan Bacho...
5	Universiti Malaysia Pahang	UMP	Pekan, Pahang
6	Universiti Malaysia Perlis	UniMAP	Arau, Perlis
7	Universiti Malaysia Sabah	UMS	Kota Kinabalu, Sabah
8	Universiti Malaysia Sarawak	UNIMAS	Kota Samarahan, Sarawak
9	Universiti Malaysia Terengganu	UMT	Kuala Terengganu, Terengganu
10	Universiti Pendidikan Sultan Idris	UPSI	Tanjung Malim, Perak
11	Universiti Pertahanan Nasional Malaysia	UPNM	Kuala Lumpur
12	Universiti Putra Malaysia	UPM	Serdang, Selangor Bintulu, Sarawak
13	Universiti Sains Islam Malaysia	USIM	Nilai, Negeri Sembilan
14	Universiti Sains Malaysia	USM	George Town, Penang
15	Universiti Sultan Zainal Abidin	UniSZA	Kuala Terengganu, Terengganu
16	Universiti Teknikal Malaysia Melaka	UTeM	Durian Tunggal, Malacca
17	Universiti Teknologi Malaysia	UTM	Skudai, Johor Jalan Semarak, Kuala Lumpur
18	Universiti Teknologi MARA	UiTM	Shah Alam, Selangor
19	Universiti Tun Hussein Onn Malaysia	UTHM	Batu Pahat, Johor

DATA EXPANDING:

- First of all, we add new 2 columns to the dataset and these are: latitude and longitude information for each row or university:

```
unis["lat"] = float("NaN")
unis["long"]= float("NaN")
unis.astype({'lat': 'float64'} , inplace=True).dtypes
unis.astype({'long': 'float64'},inplace = True).dtypes
```

- We used "geolocator " package to retrieve the coordinates of each university and we tried it first on one of the universities to check the results:

```
#testign the function  on one uni
geolocator = Nominatim(user_agent="kk")
location = geolocator.geocode("Universiti Malaysia Sarawak")
print(location.address)
print(location.latitude, location.longitude)
```

```
Universiti Malaysia Sarawak, Jalan Datuk Mohamad Musa, Taman Samarindah, Kampung Mangka, Sarawak, 94300, Malaysia
1.4648777 110.4270042
```

GETTING COORDINATES

- Using loops to get coordinates of each university:

```
i = 0
for index , data in unis.iterrows():
    loc = data["Loc"].replace(',', '')
    add = data['MalayName']+ " "+ loc
    location = geolocator.geocode(data['MalayName'])
    if (location == None):
        location = geolocator.geocode(loc)
        if (location == None):
            unis.drop(index , axis=0 , inplace=True )
            unis.reset_index()
            continue

    print(location.latitude , location.longitude )
    unis.iloc[index , unis.columns.get_loc('lat')] = location.latitude
    unis.iloc[index , unis.columns.get_loc('long')] = location.longitude

    i += 1
print(i)
```

- using two type of combination only and in case there was a Value of None returned back by the locater then we will drop the row of that specific university. Following this methodology gave us very good results, it returned back 103 coordinates out of 121

FOURSQUARE JOBS:

- First, I set up my API call credentials
- Then I defined one function that goes into the dataset and take each university and fetch the venues around it, limited to 100 venues, as it is shown below:

```
def getNearbyVenues(names, latitudes, longitudes, radius=5000):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

EXPLORING NEW DATASET

- After doing group by function on the dataset to see a summary of this new dataset:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Kolej Universiti Bandar Utama	29	29	29	29	29	29
Kolej Universiti Islam Antarabangsa Selangor	100	100	100	100	100	100
Kolej Universiti Islam Melaka	100	100	100	100	100	100
Kolej Universiti Komunikasi Han Chiang	100	100	100	100	100	100
Kolej Universiti Linton	5	5	5	5	5	5
Kolej Universiti New Era	20	20	20	20	20	20
Kolej Universiti Saito	100	100	100	100	100	100
Kolej Universiti TATI	90	90	90	90	90	90
Kolej Universiti Teknologi Sarawak	39	39	39	39	39	39
Kolej Universiti Widad	100	100	100	100	100	100
Kolej Universiti Yayasan Sabah	100	100	100	100	100	100

CLUSTERS AND MACHINE LEARNING

- We are using K-means algorithm with 5 clusters then are fitting the model with dataset and getting the cluster for each row, then appending these cluster labels to each row in the dataset:

	MalayName	Acr	Loc	lat	long	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	Universiti Malaya	UM	Kuala Lumpur Nilam Puri, Kelantan	3.122674	101.653561	2	Café	Indian Restaurant	Malay Restaurant	Shopping Mall	Ice Cream Shop	Convenience Store	Coffee Shop	Hotel	Spa	Steakhouse
1	Universiti Islam Antarabangsa Malaysia	IIUM	Gombak, Selangor	3.253190	101.735714	2	Malay Restaurant	Indonesian Restaurant	Café	Sandwich Place	Asian Restaurant	Burger Joint	Thai Restaurant	Gym	Spa	Middle Eastern Restaurant
2	Universiti Kebangsaan Malaysia	UKM	Bangi, Selangor	2.924087	101.781385	2	Malay Restaurant	Coffee Shop	Restaurant	Asian Restaurant	Café	Burger Joint	Hotel	Soccer Field	Indonesian Restaurant	Japanese Restaurant
3	Universiti Malaysia Kelantan	UMK	Pengkalan Chepa, Kelantan Jeli, Kelantan Bacho...	5.995766	102.402764	2	Beach	Restaurant	Grocery Store	Resort	Breakfast Spot	Soup Place	Soccer Field	Other Great Outdoors	Caribbean Restaurant	Fast Food Restaurant
4	Universiti Malaysia Pahang	UMP	Pekan, Pahang	3.722724	103.122982	1	Malay Restaurant	Chinese Restaurant	Light Rail Station	Thai Restaurant	Asian Restaurant	Breakfast Spot	Bus Station	Cafeteria	Metro Station	Baseball Field

BUILDING THE MAP

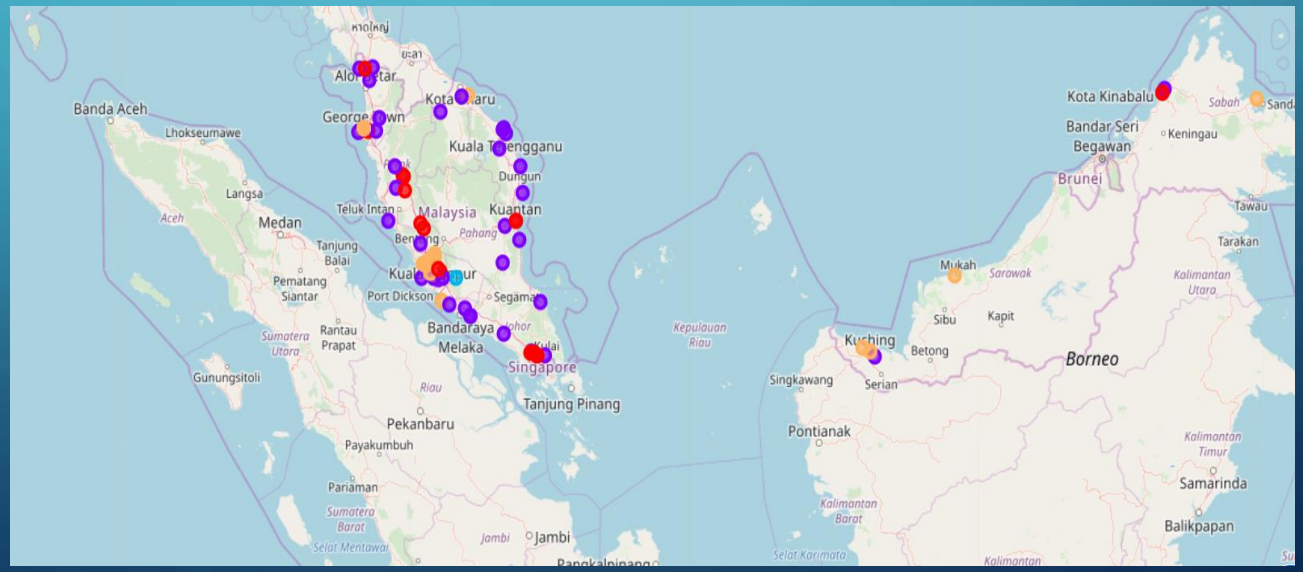
- Since we are looking for the universities in all Malaysia, we are going to build full Malaysia map, so we need to get the coordinates of Malaysia:

```
address = 'Malaysia Kuala lumpur'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
```

BUILDING THE MAP

- Then we start to build or Folium map using looping through the universities and their clusters, then shows group these universities by colouring them depending on their clusters, and this is the result we get on the map:



CONCLUSION

- from the final results we can see that students who prefer international dishes to choose cluster 2. However, stakeholders who want to start international restaurants bushiness, they would have high chance of success in cluster 3 and 4