# Best competitive location for a bubble tea store in Kuala Lumpur

Week 5 Capstone Project

# Introduction

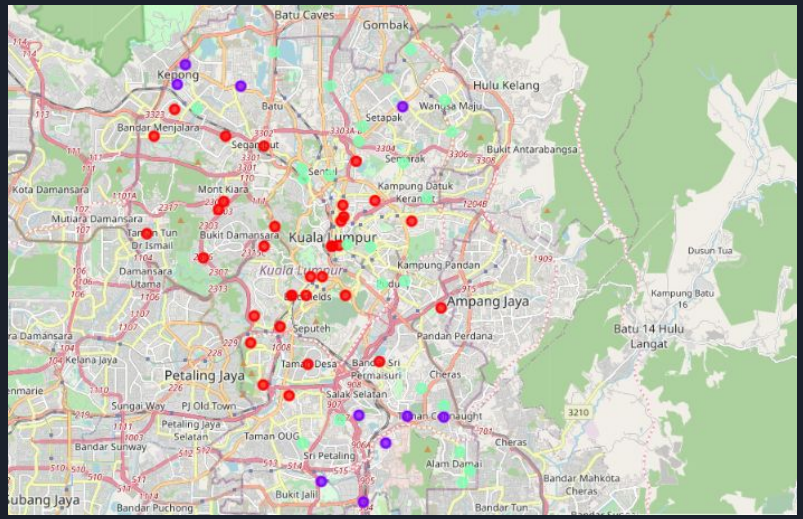
According to Straits Research, the Malaysian bubble tea market was valued at 49.8$ million in 2018 and is expected to grow 6.9 percent from 2018 to 2026. This proves to be both a promising future and an interesting dilemma for bubble tea enthusiasts looking to start their own bubble tea business. As with any business bubble tea can become a very competitive field and especially in a place like Malaysia where proximity is very important due to the nature of their traffic system. Due to the combination of the increase in demand for bubble tea leading to more business owners looking for opportunities to open bubble tea stores, the competitive nature of the bubble tea system and the importance of location to the average bubbler tea consumer in Malaysia, I believe that using data science to determine the best position for a bubble tea shop in the capital city of Malaysia Kuala Lumpur will bring in many different entrepreneurs looking for the prime locations that i can provide using the data.

# Data

I will only be needing 2 seperate data sources for this project. Firstly we will need a method of separating different areas of the cities. At first my initial consideration was to separate the city into a grid and give an analysis based off proximity to downtown. However upon further consideration i've decided since im selling the data to local businessmen and businesswomen who grew up in the area they will have a better understanding of the data if we were to split it up by regions of the city. To find the various region names in Kuala Lumpur we can go to Wikipedia and head over to https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur and used beutiful soup to scrape the list of names into python. After that for the condensation of other businesses we used FourSquare data to identify other Bubble tea places in that neighbourhood

# Methodology

To begin beautiful soup was used to web scrape a list of neighbourhoods from a city in Kuala Lumpur. Combining that with Geopy Coordinate data and foursquare business data, a data frame was created for the amount of certain types of businesses within each neighbourhood. Converting that into the mean amount of items we can sort the data frame to see where the most condensed areas of bubble tea shops are. Using dummies and one hot encoding to convert that to 0's and 1's we are able to run it through. Running the data through a k-cluster algorithm with different amount of pre-set clusters it was found that three is the perfect amount due to the simplicity of the final product. Finally all that was needed to do was make a map with the three clusters colour coded

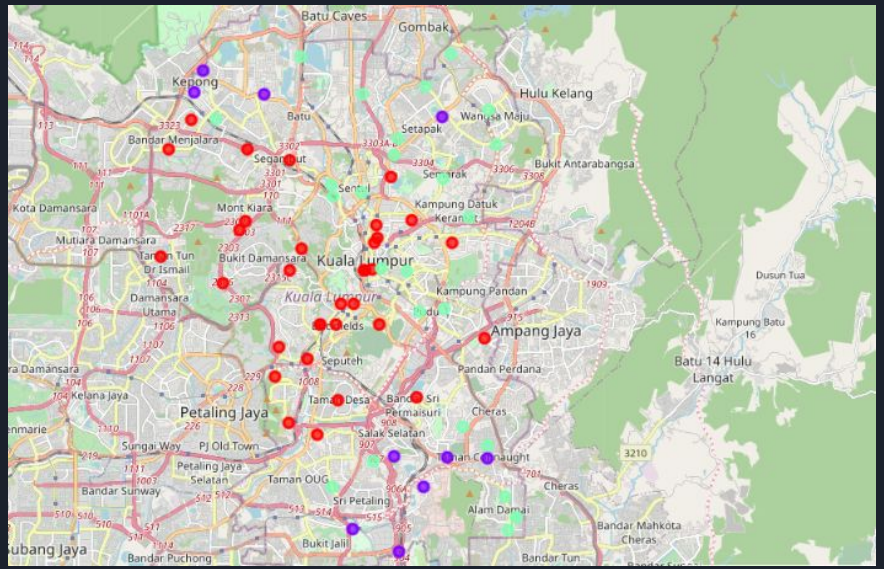



Best Places to start bubble tea business (Least competition)
Average places to start bubble tea business (decent competition)
Worst place to start bubble tea business (High competition)

# Results

As discussed previously the decision to make the final result in a form of a map was made for the simplicity of the business owner. Attached is also part of a graph showing the names of the specific neighbourhoods with their label shows as a zero, one or two. This is also a valid way of showing the result to the owner as they are probably familiar with the areas of the city. This information can be very useful to owners who are looking for the best possible start to their bubble tea business.

| | | | | | |
|---|---|---|---|---|---|
| 16 | Bukit Petaling | 0.000000 | 0 | 3.129290 | 101.698960 |
| 14 | Bukit Kiara | 0.000000 | 0 | 3.143480 | 101.644330 |
| 53 | Sungai Besi | 0.020000 | 1 | 3.049970 | 101.706030 |
| 55 | Taman Connaught | 0.030000 | 1 | 3.082690 | 101.736890 |
| 23 | Desa Petaling | 0.030000 | 1 | 3.083300 | 101.704380 |
| 4 | Bandar Tasik Selatan | 0.020619 | 1 | 3.072750 | 101.714610 |
| 5 | Bandar Tun Razak | 0.020000 | 1 | 3.082760 | 101.722810 |
| 25 | Happy Garden | 0.020000 | 1 | 3.201630 | 101.721070 |
| 27 | Jinjang | 0.020000 | 1 | 3.209500 | 101.658740 |
| 13 | Bukit Jalil | 0.020000 | 1 | 3.057810 | 101.689650 |
| 62 | Taman OUG | 0.020000 | 1 | 3.210051 | 101.634508 |
| 33 | Kepong | 0.020000 | 1 | 3.217500 | 101.637630 |
| 61 | Taman Midah | 0.010000 | 2 | 3.093590 | 101.728370 |
| 58 | Taman Ibukota | 0.010000 | 2 | 3.212310 | 101.715250 |



# Discussion

Several interesting observations were made during the process of this project. For starters as K clusters were added although the map showed a better representation of where other bubble tea locations were located it did not give an easy explanation to business owners of where to put there new business would be. Turning the K clusters down to three showed more obvious representations of what areas of the city were lacking the supply needed to meet the growing demand in that city. It was also found that a Multiple Linear Regression had the same effect on the data and it would have been much harder for business owners to understand and make decisions based off. Finally we found that the north and south poles of the city were similar in the sense they have relatively close number of stores while the east and south contrastingly the west and east  poles of the city are polar opposites and have a high and low populated area accordingly.

# Conclusion

To finish off this report it is crucial to remember that the demand in the city of Kuala Lumpur is showing huge increase in demand for bubble tea and the industry is set only to grow more until 2026. That information combined with the transit system of Malaysia shows how important the information gathered on this report is. On top of that making the information as clean and understandable as possible while maintaining the accuracy and preciseness of the data is crucial to find business owners to be satisfied with the information given to them.