# Battle of the Neighborhoods

## Introduction:

#### Life School Phuket

I own a small alternative international school in Phuket, Thailand. I am fitness and nature junky and an obsessed surfer. The school has just finished its 5th full year in operation. I started the school to provide an education to my own two boys first and foremost, and they are nearly to the point where they will finish the school and begin higher education online. The school was started with very little capital and provides a community environment for children who might otherwise be home-schooled or a cheaper alternative to very expensive international schools.

The long term plan is to take my children on the road with me and start 4-5 new small schools in strategic locations that are similar to Phuket and have similar challenges for expat families. Rather than compare two cities or neighborhoods within a city, I compared the neighborhoods of the top surfing spots in the world to attempt to identify the top choices to expand our school brand

## Data:

The main data I used for this analysis was a list of the top 50 surf spots in the world that was published in a CNN report in 2013. The link to the article is as follows: <https://edition.cnn.com/travel/article/50-surf-spots/index.html>

I also used venue data from Foursquare accessed via the GPS coordinates of the which I needed to manually scrape to a CSV file.

## Methodology:

The first step of the analysis was to web-scrape the CNN website with the article in a Jupyter notebook on IBM Watson and use Beautiful Soup to clean the data. It was then be transformed into a Pandas data frame . Once the data set of the top 50 surf spots was in a data frame format I needed to get the GPS coordinates of each and add them to the data frame.

Once I had uploaded my self generated CSV file to the IBM Watson Studio project I merged the two into a single data frame. This was used to generate a list of the top ten venue types for each location.

Finally using the Foursquare API obtained venue location I performed a k-means clustering exercise to divide the locations into 5 clusters. This narrowed down my selection to the cluster which includes Phuket and the potential next location will be chosen based on its closeness to Phuket.

## Results:

The clustering process resulted in Phuket being among a cluster with a total of 11 spots, thus 10 which would be “most like Phuket” according to the venue data from Foursquare. Of these I can easily identify which surf spots were in highly developed economies with good English language based schools already, and those in developing economies similar to Phuket.

My short list of places to consider based upon this exercise has been narrowed down to Puerto Escondido, Montanita, Hainan Island, Arugam Bay. Interestingly Bali did not make the same cluster and my belief is that it is because I had to choose between Kuta and Uluwatu as the article listed both spots which are an hour away from each other. I chose Uluwatu for the wave but Kuta is probably a better location for a school as it is nearest to the main expat community.

## Discussion:

There were some interesting findings brought about by this in that 3 out of 4 of the sites were already in consideration in the back of my mind for various reasons. A friend of mine owns land in Arugam Bay so these results are encouraging to explore that possibility more. I had previously almost co-bought some land near Montanita as well so it seems my instincts have not been too bad.

A surprise in this project was that Bali did not make the list since having spent time in both places I expected Bali to be the closest. I also believe I made an error by choosing Uluwatu over Kuta which was a choice forced upon me by the article. I could have added both to make a total of 51 and an area for further research may be to add Kuta.

## Conclusion:

In conclusion, of the 50 best surf spots in the world there are 4 spots I need to do some research on to consider them as the next locations to open schools. Arugam Bay being geographically close is a great candidate whereas Hainan may be difficult due to increasing Chinese regulation in the education industry.

For further research I intend to run a similar clustering exercise without using Foursquare but instead creating a list of other factors I feel are important for my business model such as cost to start a business, amount of regulation, etc.