# The Battle of the Neighborhoods - Week 1

## DATA DESCRIPTION

The two suburbs under consideration-New Albany & Dublin, Ohio will be analyzed in this project

## DATASETS

We will be using the below mentioned datasets for analyzing New Albany & Dublin, Ohio

**Data 1**: In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the neighborhoods that exist in each suburb as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. Links to the dataset are :  
<http://worldpopulationreview.com/us-cities/new-albany-oh-population/>  
<http://worldpopulationreview.com/us-cities/dublin-oh-population/>

The above links gives us the demographics of the two suburbs under consideration as well as education level and income level. These will be used to find spending capacity to compare the two locations.

**Data 2** : Second data which will be used is the US Government Census data along with the prediction for future growth to find which market will have potential for growth in the future and to validate the data received from our primary data source.  
<https://www.census.gov/quickfacts/fact/csv/newalbanycityohio,dublincityohio/PST045218>

**Data 3 :**For more analysis, we will get data Indian population data for over 350 zip code from <http://zipatlas.com/us/zip-code-comparison/percentage-indian-population.htm> which are comparable to the two neighborhoods we are analyzing. We will fit various machine learning regressions algorithms to this data and then predict which areas population is underserved.

Ohio has a huge fan base for Indian food. There are about 20 Indian restaurants already in Dublin. Whereas New Albany and its surrounding areas have around 5, only 2 of them in New Albany & Gahanna put together. This makes New Albany an easy market for our entrepreneur to open an Indian restaurant.

**Data 4:** Dublin & New Albany geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in these two suburbs. Here we are querying the Foursquare API to get a count of number of Indian restaurants that already exists in different Zip code to then build a model for how many restaurants serve the population (training data). We then use the model to predict how many restaurants should exist in the two-neighborhood selected to then find the ideal location for opening a new restaurant.