1. A description of the data and how it will be used to solve the problem. (**15 marks)**

We need data from reliable sources for further analysis. Focusing on this project, we will use following data:

1) Pulling the data of neighborhood of the city of San Diego out of HTML <https://web.archive.org/web/20080509171047/http://www.sandiego.gov/neighborhoodmaps/> . The neighborhoods of the City of San Diego are divided into regions. These regions are based on the neighborhood policing and community services areas.

2) Foursquare Developers Access to venue data. Foursquare is a US tech company from New York focusing on location data. I will use foursquare data such as the venue name, location and category.

3) The house price comes from Zillow. We use Zillow home value index, which means the typical value for a geographic area, to represent house price in that neighborhood.

4) The latitude and longitude data of each neighborhood comes from Google search, due to recently Google started charging for their API.