**The Battle of Neighborhoods –**

**A Comparative Analysis of Neighborhoods in San Francisco, CA & Chicago Illinois.**

1. **Introduction/Business Problem.**

United States of America is a land of opportunities where many from all over the world want to either visit temporarily or immigrate to permanently and call a new home. As a result of this, many prospective immigrants or newcomers attempt to compare various cities to determine where best suits their needs which could range from job seeking opportunity, business opportunity or a peaceful area to live in etc. We do know that starting life afresh in a new city can be a problem.

This work is a capstone project for the completion of the IBM Data Science Certification, and it will involve a comparative analysis of different neighbourhoods in two different cities. It includes a segmentation of downtown areas, identify various venues and cluster them together. San Francisco, California and Chicago, Illinois are two US cities which has been selected for this project. The analysis will also explore their similarities and differences and the results can be a useful piece of information for new settlers and or prospective immigrants.

1. **Data Description.**

Various sets of data will be required to solve the problem highlighted above therefore, the following data sets will be used to complete this project.

Data sets containing;

* United States cities zip codes together with their corresponding geographical coordinates. This will be used to generate data frames for San Francisco and Chicago. (<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude>)
* San Francisco zip codes and the area name.

(<http://www.healthysf.org/bdi/outcomes/zipmap.htm>)

* Chicago zip codes and the area names.

(<https://www.chicagotribune.com/chi-community-areas-htmlstory.html>)

* These data containing the zip codes and area names will be merged to create a new data frame for identifying the neighborhoods in both cities which will then be used for segmentation and clustering analysis.
* Foursquare API will be required to generate venues in the selected areas.