**Taste of Chicago:**

**A Research into The Taste of Different Neighborhoods**

* **Introduction**
  + Background

The City of Chicago is the 3rd largest city in the United States, it is home to about 2.7 million people. I lived there for several years in two neighborhoods – from Hyde Park in the southern suburb to West Loop in downtown. I found the types of restaurants so different – in Hyde Park there were mostly fast food restaurants while West Loop might be the most condensed neighborhood of fine dining restaurants. Therefore, I would like to understand the different taste between neighborhoods using data science techniques.

* + Business Problem

The problem I’m trying to analyse is: if someone is looking to open a restaurant in Chicago, where would I recommend that they open it?

The idea is to categorically segment the neighborhoods of Chicago into major clusters based on their food taste. I would then compare groups of neighborhoods by demographic statistics such as population density, per capita income and so on. From these two angles, I will get a better understanding of the taste and potential consumers’ profile. After that, I can give specific recommendations to potential stakeholder based on the type and pricings of his restaurant.

How would we define an area’s food taste? For this problem, we would utilize FourSquare API to find top 100 restaurants within each one neighborhood. We would group them by food types and aggregate weights of numbers of each food type.

* + Stakeholders

The result of this analysis can be utilized by a potential food vendor hoping to open a new restaurant. Also, it can be used to understand the distribution of different cultures and cuisines over Chicago.

* **Data**

Following sources are used:

* + Stanford Digital Repository (https://purl.stanford.edu/xq082nw3443)

The geographic and demographic data source of Chicago downloaded from 'Stanford Digital Repository'. In the file 'hoods3155lite.dbf', it contains major US cities neighborhoods, latitudes, longitudes, average and medium household income and so on. For the purpose of this project, we will focus on the city of Chicago. We will keep and rename relevant features and discard the others. The features to keep are:

* 'NAME': name of neighborhood
* 'X': longitude
* 'Y': latitude
* 'POPDENSITY': population density
* 'DIVERSITY': diversity index
* 'PC\_INCOME': per capita income
* 'MEDAGE\_CY': median age
* 'UNEMPRT\_CY': unemployment pct 2010
* 'MEDVAL\_CY': median home value
  + Foursquare API (https://developer.foursquare.com/docs)

Foursquare API, a location data provider, will be used to make API calls to retrieve data about venues in different neighborhoods. Venues retrieved from all the neighborhoods are categorized broadly into ‘Arts & Entertainment’, ‘College & University’, ‘Event’, ‘Food’, ‘Nightlife Spot’, ‘Outdoors & Recreation’, etc. Under each category, there are detailed subcategories. For example, ‘Food’ category contains ‘Fast Food’, ‘American Restaurant’, ‘Deli/Bodega’, ‘Pizza Place’ and so on, there are more than 200 food subcategories.