### Part 2: Data description & how it will be used to solve the problem

There are two sources identified--

1. It is logical to assume that neighborhoods/towns/cities are not homogeneous. They

will have different demographics profile, different distribution of wealth,

infrastructure, etc … Hence, a logical method would be to analyse by postal codes or

zip codes as they are called in the U.S. Hence, the first data set will be Dallas Zip

codes, along with their population. The name Dallas can refer to a county, in which there are cities and towns as listed below:

1. Addison

2. Balch Springs

3. Carrollton

4. Cedar Hill

5. Coppell

6. Dallas

7. Desoto

8. Duncanville

9. Garland

10. Grand Prairie

11. Hutchins

12. Irving

13. Lancaster

14. Mesquite

15. Richardson

16. Rowlett

17. Sachse

18. Seagoville

19. Sunnyvale

20. Wilmer

Each zip code will have a corresponding latitude and longitude value assigned.

Unfortunately, a city, by virtue of its size, can have multiple zip codes, and the results

have to be grouped by city/town. For the purpose of aligning the nomenclature,

cities and towns will be called ‘neighborhood’, and Dallas county will be referred to

‘city’ per the the data from Foursquare.

2. The second data set will be the venue data. Details on the venues will be derived

from Foursquare.com website via an API to the application. Foursquares provides a

rough guide on the types of cuisine according to a predefined set of categories as

documented on its website https://developer.foursquare.com/docs/resources.

While it also returns the venues’ frequency by neighborhoods which is defined by

their zip codes and their respective latitude and longitude. This information can only

be used as a rough guide as Foursquare returns the findings based on a specified

radius from a given latitude and longitude. This already assumes that all

neighborhoods are circular and of a fixed size with its latitude and longitude in the

centre of the circle and it is not capable of limiting its search within the boundaries

of a given city or town or neighbourhood. In this exercise, we will not attempt to.

‘scrub’ the information coming back from Foursquare for duplicates, or venues returned that don’t match the search criteria.

#### Analytical Methods::

The basic “Demand and Supply” approached will be used. To achieve this:

1. Statistical analysis has to be performed on the types of venues and its frequency by

neighborhoods.

2. Understanding of how the Asian cuisine market segments itself such as a generic

Chinese restaurant vs a Sushi or Peking Duck restaurant. As Chinese forms the

greatest land mass in Asia, and the highest population count, understanding will

have to be derived from how Foursquare defines this.

3. Any correlation between Asian and non-Asian cuisine to derive acceptance rates and

opportunities for growth.

### Interpretation of results::

1. If an area has a high volume of Asian cuisine with respect to the overall venues

count, it could mean that the market is moving towards saturation

2. If an area has a very low count of Asian cuisine, with respect to the overall venues

count, it could mean that the penetration rate is poor.

3. There could be potential correlation between population density, the various type of cuisine and their frequencies.

4. Again, the data cannot be interpreted strictly and it should only be used as guide