Capstone Project Report

Part 1: Description of the problem and background discussion

Introduction

When the Asians migrated from distant Eastern countries, i.e China Hong Kong, Japan, etc. throughout the world, it opened an altogether a new skill of food industry and cuisines for the food hospitality industry. It is safe to say that Asian cuisine is well known and enjoyed everywhere in the U.S.A and a method to help provide some statistical data can prove useful to understand the existing market in a particular location.

Problem Description

For this project, we will study in depth the greater Dallas area (or Dallas county) to derive the statistical results for various cuisine/foods for the purpose of getting to know of various opportunities of establishing a new Asian cuisine restaurant. The term Asian cuisine is used loosely here to cover the wide range of cuisines found in the greater Asia and can include fusion for variety and to increase the offering.

The definition of Asian cuisine incorporates a extensive range of cooking practises, techniques and traditions followed throughout the mainland South Asian and Eastern Asian countries and there is no enforcement on how the term is being used. Just in the Asia region alone, there is Chinese cuisine which varies greatly in taste and flavour at different locations, to Japan, countries in South Asia, and to western parts including India. Rest, it will all depend on how the data used is initially prepared for the analysis.

For this exercise, Dallas was pick for the study for the following reasons:

1. It has a large white racial makeup, about 50% and while Asian constitutes about 6%.
2. It has mild weather with summer in the mid-30s and its dry, while winter is in the teens with at worse 1 inch of snow or so.
3. While it has issues with tornado, the inland is less vulnerable to
4. The current unemployment rate is about 10% below the nationwide numbers and it is at its lowest since 2002.

In this exercise, we will attempt to understand the numbers around the frequency of various categories of venues and the ratio to the Asian venues for the cities that falls within the Dallas greater area.

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

Data Analytics

Data sources and description of the data

There are two sources identified

1. It is logically safe to assume that neighbourhoods/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:

    a. Addison
    b. Balch Springs
    c. Carrollton
    d. Cedar Hill
    e. Coppell
    f. Dallas
    g. Desoto
    h. Duncanville
    i. Garland
    j. Grand Prairie
    k. Hutchins
    l. Irving
    m. Lancaster
    n. Mesquite
    o. Richardson
    p. Rowlett
    q. Sachse
    r. Seagoville
    s. Sunnyvale
    t. 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 '**neighbourhood'**, and Dallas county will be referred to '**city**' per 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 neighbourhoods 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 neighbourhoods 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. Statistics based analysis is to be used on the various data like the types of venues and its frequency by neighbourhoods.
2. Understand how the Asian food market has segmented itself into different domains like a generic Chinese restaurant vs a Sushi or Peking Duck restaurant or an Indian Restaurant. As Chinese is of greater land mass in Asia as compared to rest of the countries, and has the highest population count, a analogy has to be derived to understand how third party APIs like Foursquare defines this scenario.
3. Any correlation between Asian and non-Asian cuisine to derive acceptance rates and opportunities for growth.

Interpretation of results

1. If there is a high volume of Asian cuisine/food chains with respect to the overall venues count in a particular area, it could mean that the market in this location is moving towards its saturation
2. If the count of Asian cuisine/food outlet is low, with respect to the overall venues count, it could mean that there is poor penetration rate of this cuisine.
3. There could be potential correlation between population density, the various types of cuisine and their frequencies.
4. Again, the data cannot be interpreted strictly and it should only be used as guide.