Applied Data Science Capstone

Capstone Project - The Battle of Neighbourhoods (Week 1)

1. Purpose

   This document provides the details of my final peer reviewed assignment for the IBM Data Science Professional Certificate program – Coursera Capstone.

2. Problem statement

   Singapore is City state located in busiest shipping lanes in the world; it's a small island nation with a thriving tourism industry. The internet provides host of information to visit and stay in and around Singapore. However, most of the recommendations are simply based on usual tourist attractions or residential areas that are mostly expensive or already known for travellers based on certain keywords like "Hotel", or "Backpackers" etc. The intention on this project is to collect and provide a data driven recommendation that can supplement the recommendation with statistical data. This will also be utilizing data retrieved from Singapore open data sources and FourSquare API venue recommendations.

3. Data Sources
   This demonstration will make use of the following data sources:

   - Singapore Towns and median residential rental prices

   Data will retrieved from Singapore open dataset from median rent by town and flattype from https://data.gov.sg website.

   - Singapore Towns location data retrieved using Google maps API
   - Singapore Top Venue Recommendations from FourSquare API

4. Process

   Singapore Towns List with median residential rental prices.

   The idea is to run analysis by using the average rental prices of all available flat type.

   Data Cleanup and re-grouping.

   The retrieved table contains some un-wanted entries and needs some cleanup.The following tasks will be performed:

   - Drop/ignore cells with missing data.
   - Use most current data record.
   - Fix data types.