APPLIED DATA SCIENCE CAPSTONE PROJECT

REPORT

Introduction

Background

It is always a hassle when someone needs to relocate to a new region. It involves a lot of research and investment. So it is necessary for one to be absolutely content with their selection of residence. This selection involves a lot of research and figuring out what the enmities and facilities are around in a particular neighborhood, does it fulfill and desired satisfaction levels and many more. Safety is one of the factors of foremost concern.

Problem

What is the safest area based on total crimes in a city say London, and what are the most popular venues in the said area?

Data

In this project, the data is from a combination of 3 sources.

1. London Crime Data from Kaggle which contains all the crimes committed per borough from the years ranging from 2008 to 2016. For this case, we will use only the data from 2016 as it is the most current. It consists of the following features:
   - lsoa_code – code for Lower Super Output Area in Greater London.
   - borough – common name for London borough.
   - major_category – High level categorization of the committed crime.
   - minor_category – Low level categorization of the committed crime within major category.
   - value – monthly reported count of categorical crime in the given borough.
   - year – year of reported counts.
   - month – month of reported counts.
2. Second source is scraped from a Wikipedia page containing the list of all the London Boroughs.
   - Borough – Names of all 33 boroughs of London.
   - Inner - categorizing the borough as either Inner London or Outer London borough.
   - Status – categorizing the borough as either Royal, City or other.
   - Local authority: The local authority assigned to the borough.
   - Political control: The political party that control the borough.
   - Headquarters: Headquarters of the Boroughs.
   - Area (sq mi): Area of the borough in square miles.
   - Population: The population in the borough recorded during the year 2013.
   - Co-ordinates: The latitude and longitude of the boroughs.
   - Nr. in map: The number assigned to each borough to represent visually on a map
3. Third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on a Wikipedia page. This dataset is created from scratch using the list of neighborhood available on the site, the following are columns:

- Neighborhood: Name of the neighborhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.