2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data from three sources. The first data

source of the project uses a London crime data that shows the crime per borough in London.

The dataset contains the following columns:

● LSOA\_CODE: CODE FOR LOWER SUPER OUTPUT AREA IN GREATER LONDON.

● BOROUGH: COMMON NAME FOR LONDON BOROUGH.

● MAJOR\_CATEGORY: HIGH LEVEL CATEGORIZATION OF CRIME

● MINOR\_CATEGORY: LOW LEVEL CATEGORIZATION OF CRIME WITHIN MAJOR CATEGORY.

● VALUE: MONTHLY REPORTED COUNT OF CATEGORICAL CRIME IN GIVEN BOROUGH

● YEAR: YEAR OF REPORTED COUNTS, 2008-2016

● MONTH: MONTH OF REPORTED COUNTS, 1-12

The second source of data is scraped from a wikipedia page that contains the list of London

boroughs. This page contains additional information about the boroughs, the following are

the columns:

● BOROUGH: THE NAMES OF THE 33 LONDON BOROUGHS.

● INNER: CATEGORIZING THE BOROUGH AS AN INNER LONDON BOROUGH OR AN OUTER LONDON

BOROUGH.

● STATUS: CATEGORIZING THE BOROUGH AS ROYAL, CITY OR OTHER BOROUGH.

● 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 (2013 EST): 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.

The 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.