The Battle of Neighborhoods

MADE BY -
ADNAN BEG

Introduction

- **Background:** in order To minimize the possibility of wrong searching a new London county, it is convenient to carry out a proper investigation when planning our next move.
- **The Problem:** This project aims to select the safest district London based on total crime, explore neighborhoods in that district to find the 10 most common places in each neighborhood, and finally group neighborhoods using k-mean grouping.
- **Interest:** People who are considering moving to London will be interested in getting to know the safest districts of London.

Data Acquisition

- The data acquired for this project is a combination of data from three sources.
  - 1. London crime data (https://www.kaggle.com/jboysen/london-crime) that shows the crime per borough in London. The dataset contains the following columns:
  - 2. The list of London boroughs (https://en.wikipedia.org/wiki/List_of_London_boroughs). This page contains additional information about the boroughs, the following are the columns:
  - 3. The list of Neighborhoods in the Royal Borough of Kingston upon Thames (https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames). This dataset is created from scratch using the list of neighborhood available on the site, the following are columns:

Data Cleaning

- From the London crime data we use only the crimes during the most recent year (2016). The major categories of crime are pivoted to get the total crimes per borough as per the category. The second data is scraped from a wikipedia page using the Beautiful Soup library in python. Using this library we can extract the data in the tabular format as shown in the website.
  After the web scraping, string manipulation is required to get the names of the boroughs in the correct form (we will be merging the two datasets together using the Borough names).
- Once crime is obtained in each district, we can determine the district with the lowest crime rate, and therefore label that district as the safest district. The third source of data comes from Wikipedia's list of neighborhoods in the safest district. This data set is created from scratch, the pandas data frame is created with the names of the neighborhoods and the name of the municipality with the latitude and longitude are left blank.
- The neighborhood coordinates have been obtained using the Google Maps API geocoding to obtain the final data set. The new dataset is used to generate the locations for each neighborhood using the Foursquare API.

Methodology

- Exploratory Data Analysis
    1. Statistical summary of crimes
    2. Boroughs with the highest crime rates
    3. Boroughs with the lowest crime rates
    4. Neighborhoods in Kingston upon Thames

- **Modelling:** Using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.

Results - 1

- After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306282	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Record Shop	Cosmetics Shop	Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant	Indian Restaurant	Fish & Chips Shop	Dry Cleaner
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform	Grocery Store	Farmers Market	Dry Cleaner
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café	Wine Shop	Fast Food Restaurant	Chinese Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store	Gastropub	French Restaurant	Train Station
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

- The cluster one is the biggest cluster with 9 of the 15 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores.

Results - 2

- Looking into the neighborhoods in the second, third and fifth dusters, we can see these clusters have only one neighborhood in each. Thisis because of the unique venues in each of the neighborhoods, hence they couldn't be clustered into similar neighborhoods

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Fast Food Restaurant	Wine Shop	Golf Course	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant

- The second cluster has one neighborhood which consists of Venues such as Restaurants, Golf courses, and wineshops.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Train Station	Pub	Food	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Deli / Bodega

Results - 3

- The third cluster has one neighborhood which consists of Venues such as Train stations, Restaurants, and Furnitureshops.

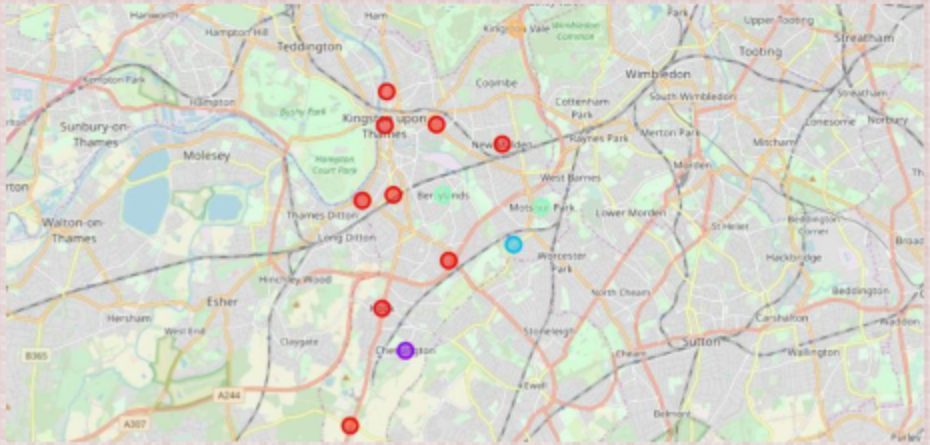
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Italian Restaurant	Soccer Field	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Department Store

- The fifth cluster has one neighborhood which consists of Venues such as Grocery shops, Bars, Restaurants, Furniture shops, and Department stores. We will look into the neighbourhoods in the fourth cluster

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Café	Bus Stop	Wine Shop	Fish & Chips Shop	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Park	Gym	Restaurant	Soccer Field	Bus Stop	Wine Shop	Fast Food Restaurant	Dry Cleaner	Electronics Store

Results - 4

- The fourth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.



- Each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood falls in the red cluster which is the first cluster. Three neighborhoods have their own cluster (Blue, Purple and Yellow), these are clusters two three and five. The green cluster consists of two neighborhoods which is the 4th cluster.

Discussion

- The aim of this project is to help people who want to relocate to the safest borough in London
- The choices of neighborhoods may vary from person to person.

Conclusion

- This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood.
- We have just taken safety as a primary concern to shortlist the safest borough of London.
- The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.