What is the best neighbourhood to live in as a student at Imperial College London?

Abstract

1. Introduction

1.1. Background

Imperial College London is a STEM university situated in South Kensington: a cultural area full of museums, parks, restaurants, shops and hotels. As such, the university's campus is fairly small, and has minimal accommodation to offer students after their first year at university. This results in the vast majority finding private accommodation, typically a couple of miles away from South Kensington, due to the high flat prices (per meter squared) in the area.

With this, commute to university becomes much longer, with many opting to cycle, use the bus or the tube (underground train) instead, while student safety decreases due to increased crime rates further away from university. Of course, other areas have their own positives, as they typically have younger demographics, are more diverse, can be more 'trendy' and offer access to other amenities, such as nightclubs.

At the moment, students are able to find correct properties (in terms of price and location) through trial and error (on website such as rightmove.co.uk and zoopla.co.uk) or by consulting with real-estate agencies, such as Foxtons or Dexters. However, there exists no mechanism by which students can *understand* the type of neighbourhood or area wherein they look for housing. Ideally, students would be able to answer a couple of questions regarding what they consider to be important features in a neighbourhood, and get answers in the form of 'suitable' neighbourhoods for them.

This report aims to explore this problem in more detail and finding a solution.

1.2. Literature Review

Definitions:

a neighbourhood is referred to as a London postcode district.

Parameters to consider:

- Average house price per unit area in the area
- Trendiness in the areas
- Availability of restaurants / nightlife
- Distance from uni (accounting for walk vs. cycle vs. bus vs. train)
- Average age in the area
- Crime rate
- Closeness of supermarkets

1.3. Scope

This project has two objectives:

1) Cluster the neighbourhoods to find any that are similar, and thus 'appropriate' for students
2) Create a model that quantifies neighbourhoods, giving them a score for how suitable they are for students. This model should be deployable such that a student could enter an address that they are considering, and it would give them a score, as well as highlight the key parameters affecting that score (for example, high crime rate).

Note that this project is partly motivated by the final assignment (Capstone Project) on the IBM Professional Data Science Certificate. **For peer assessors, please only consider Objective 1 (Clustering) as a submission for the Capstone.**

2. Data

This project requires more data than that available from Foursquare.

A description of the data required for this project will be given here, as well as links. Note that the methodology for how the data will be used is given under **Methodology**.

- **Foursquare:** will be used for exploring venues in postal districts, as well as trendiness
- **Rent Barometer:** includes data on average rent prices (£) in London by postcode districts for different property types (Studios, 1—5 beds)
- **Office of National Statistics (Population):** includes age data (male and female) by districts for different years
- **UK Crime Stats:** includes crimes (of different types) per district, as well as information on districts (such as population and land area)
- **Google Maps:** will be used to explore distances and durations from locations to Imperial College

Note that links tot he datasets of Rent Barometer, Office of National Statistics and UK Crime Stats are given under **Data Sources**. Both Google Maps and Foursquare data are accessed through API's, and as such have no 'datasets'.

The next subsections will discuss the datasets in more detail in terms of the parameters that they contain and how they will be used to achieve the objectives of the project.

2.1. Foursquare

Foursquare gives developers a certain number of calls (cite Alex Akklson from IBM) to retrieve information regarding venues.

This project will use the following approaches:

- Postcode districts will be explored for specific venues relevant to students, these include: nightclubs, bars, pubs, restaurants, coffee shops and grocery stores.
- The 'trending' feature in Foursquare will be used to gather information about the trendiness of a postcode district at different times of the day.

The data described above will be used to create 'metrics', for example how 'trendy' the area is, or what the 'nightlife' score is. This is described in more detail under the methodology.

2.2. Rent Barometer

The rent barometer dataset is a webpage containing tables which present information on weekly rent for different property types in each postcode district.

Webscraping will be used to extract this information. There are some missing values in the datasets, and these will be approximated by suitable models.

The data collected here would also be converted to a metric relevant to students.

2.3. Office of National Statistics (Population)

This dataset is a .CSV file, and thus easily accessible.

It's important to note that the data is outdated, containing data from 2008.

This data will be used to create a metric for the average age in each postcode district.

2.4. UK Crime Stats

The UK crime statistics can be exported as a .XLS file. It will be used to determine a ‘safety score’ for each postcode district.

2.5. Google Maps

Google gives developers access to Map data.

Google maps will be used to determine average durations from postcode districts to Imperial College London (bus, tube, walk and cycle).

This data will be used to create a ‘closeness score’ for each postcode district.

3. Methodology

4. Results

5. Discussion

6. Conclusion

7. Further Work

8. Resources

8.1. Data Sources

https://www.rentbarometer.com/london/all-prices/by-postcode.html#BR

https://data.london.gov.uk/dataset/office-national-statistics-ons-population-estimates-borough

https://www.ukcrimestats.com/Postcode_Districts/

https://developers.google.com/maps/documentation/distance-matrix/overview

8.2. Other