1. <u>Introduction</u>

<u>1.1 Background</u>

Opening a new business is hard for anyone. It is a complex task that requires a lot of research. Deciding on a location is one of the hardest tasks to accomplish when it comes to opening a new small business or relocating. London is one of the most popular cities in the world with a large number of restaurants and small businesses. Any person looking to open up a business here will be concerned about the safety in the city. The safety in that area should be one of the prime concerns when starting a new venture. Businesses need to make proper accommodations regarding the safety of their business as well as the people who visit. Different types of crimes are reported in different parts of the city so any new business owner must be aware of these and must make the necessary safety arrangements in case of a mishap.

To lower the risk we must always conduct proper research on choosing a location that is best suited for our needs. This report focuses on the safety aspect of either opening a new business or restaurant.

<u>1.2 Problem</u>

The crime statistics dataset that is found on Kaggle has crimes that have been recorded for each Borough from 2008 to 2016. We will focus on the year 2016 since that will be the most up to date data that is available.

This project aims to find the most crime ridden areas of the city of London based on the total crimes, explore those neighborhoods and look at what businesses already exist in those neighborhoods and find the top 10 most common venues in each neighborhood and then finally cluster them using k-means clustering

<u>Data acquisition</u>

The data acquired for this project is a combination of two sources. The first source uses a data sheet containing the London crime data from Kaggle: https://www.kaggle.com/jboysen/london-crime

That shows the crime per borough in London. The dataset has the following columns:

- Lsoa_code
- Borough: Common name for London borough.
- Major_category: category for a major crime

- Minor_category: category for a minor crime
- Value: Reported count of crimes for each borough in a month.
- Year: Year of reported counts(2008-2016)
- Month: Month of reported counts(1-12)

The second source of data is scraped from the Wikipedia page that contains the list of London Boroughs which we will need to explore the neighborhoods in the boroughs.

The dataset has the following columns:

- Borough : Common name for a London borough
- Inner
- Status
- Local authority
- Political Control
- Headquarters
- Area in sq miles
- Population
- Coordinate data

The Foursquare API will then be used to explore the venues around these neighborhoods with an 800m radius to have a look at what venues exist in and around these neighborhoods.