# IBM Data Science Professional Certificate-Capstone Project

## 1-Problem Statement

In the country of United Kingdom, London is the one of the hotspots for IT related jobs. In this project, one of the main problems that the people who started their career in IT can face will be targeted. From the beginning of the process, there are various problems. Even though there is a plenty of demand for skilled people out there, landing a good job has been always hard. However, a problem that can be faced after landing a good job will be treated in this project. Supposedly the person has already landed their job, and now it is time to find a place to live. Finding a good neighbourhood is always a big concern to deal with. Especially in big cities like London, the dilemma of price and quality can be challengingly sharp. While the price range can easily widen, the person may also end up living an undesirable neighbourhood. High average of rental prices may be a sign for more enjoyable or safer neighbourhood.

## 2-An Approach to Problem

A reasonable approach to this problem is comparing crime numbers and rental prices of places for whole city of London and then analysing neighbourhood according to their amenities for our IT person. To start with, one assumption is made. Wage of a person who makes their living as an IT specialist will probably be more than average wage and this person will probably want to live somewhere close to central London, to benefit London's beauties. For this reason, the first step is filtering London boroughs by their inner/outer designation. Inner London boroughs will be the matter of this project. Then London boroughs will be compared by using the rental price and crime number aspects. London boroughs will be ordered by the total number of crimes and most dangerous boroughs will be removed from candidate list. Then the remaining boroughs will be ordered by rental price. There is a tricky point which should be taken into consideration here. By knowing the supply and demand relationship from Economy, low price can be an indicator of low demand which can be caused by insufficient social amenities of neighbourhood. Therefore, a medium price band will be focussed, instead of focusing on the lowest price band. After finding our target borough, neighbourhoods of that borough will be analysed by their venues. As a young, relatively comfortable in financial person may have different expectations from where they live. Those neighbourhoods will be clustered by their venues to come up with potential best places to live for our IT person.

## 3-Description of Data

For this project, the following data will be used:

- List of London Boroughs with their designations: https://en.wikipedia.org/wiki/London_boroughs
- Number of Crimes by London Boroughs: https://drive.google.com/file/d/1TTOhuQit4gkenrUizsUn9wloHalRffhE/view

- London Rental Prices by Boroughs: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/adhocs/11100privaterentalmarketinlondonjanuary2019todecember2019

- Foursquare API

However, these sources give the data for multiple purposes. This is why they are needed to be cleaned and wrangled accordingly. Some of the columns and row may be needed to be removed, some others may be combined, some nonsense entity may be handled according to need.
First source is a Wikipedia page. Therefore, it is needed to be scrapped by using python library. The second is a csv file which can easily be turned to Pandas dataframe, however it does not have total number of crimes. It has data for different kind of crimes separately from March 2018 to February 2020. So there will be some calculations to generate useable data for our purpose. The third source is an xls file with some of text before the needed table. Also the table is formed in a different structure. It has data for different types of properties separately, instead of an average rental price for each borough. This source has also rental price data recorded between January 2019 to December 2019. So there will be calculation to find average rental prices for each boroughs as well. Final source is Foursquare's API that we used before, during the course. Course's instructions will be followed to obtain venue data for neighbourhoods of selected borough.