**Applied Data Science Capstone Project Report**

1. **Introduction/Business Problem**

**London is a diverse place. It is both a famous tourism destination and a financial and political capital. There are many opportunities to invest in catering in London. From stakeholders to self-business starters, many people would be interested in this. An investment needs to take into account a lot of aspects – location, scalability, target people, concepts, etc.**

**And I am here to help.**

1. **Data**

**The data used here are an excel file of London Borough information and a csv file of all London postcodes and geographical information. Both data are downloaded directly from open sources such as UK government website.**

1. **Methodology**

**I am going to combine these 2 data files and filter down to important boroughs. Then with use of Foursquare API and machine learning techniques, in particular clustering, I can make an appropriate suggestion for my client ABC.**

**The reason for using k-means clustering is that firstly, it is good in solving our problem of finding a best location of investment. And secondly, it is efficient especially for large data sets, which is suitable for this problem. Although it potentially returns deferent clusters each time it is run due to random initialisation of centroids, this can be compensated by running this algorithm a few times to reduce randomness and improve accuracy.**

1. **Results**

**The detailed map is shown in notebook.**

1. **Discussion**

**As we can see from the map, London Boroughs are clustered into 5 areas, one of which is centred in Kensington, surrounded by South Kensington, Earls Court, etc.**

**As I have also computed a data frame consisting of top 5 venues of each borough. By search this, I get to know that café is the top first most common venue in Kensington, with parks, Italian restaurants and exhibits being also popular.**

**Based on these, I would suggest investing in a small Italian styled café near park or museums.**

1. **Conclusion**

**The recommendation of a small Italian styled café made upon data resources gathered and clustering mechanisms.**

**If more data of people group (age, career, etc) and purpose of visit are collected, the recommendation can be more precise.**