**INTRODUCTION**

We live in a global world where opportunities are not limited to the country we reside in. We apply for jobs in new cities, start business endeavors across of the globe, and have a seemingly endless list of places we can travel to.

New York City and London are no exception. Both financial centers are filled with booming economies, lively cultural experiences, and a wealth of resources. As such, it is not uncommon for people to consider visiting or even moving there. Say an employee must move from New York City to London to take on a new role at their company. She wants to live in a neighborhood that offers many of the same amenities as her hometown. While the decision to move may be a difficult one, it may be harder to know what neighborhood to settle down in. This comes as no surprise as these metropolises are home to a population of more than 8 million and span roughly 300 square miles and 600 square miles, respectively.

With so many neighborhoods to choose from, it’s tough for an individual to weigh all the options. That’s where the role of data science comes in: we will use the computing power of machines and the insights from statistical models to help the user make a more informed decision. Through this project, a user will be able to see how the various neighborhoods in New York City and London are similar based on the venues they offer.

**DATA SOURCE**

* For the New York City portion of this project, I will be using the data set provided in the IBM lab that is sourced from NYU. The [NYC data set](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json) is a json file that contains the names of the neighborhoods as well their boroughs and geographic coordinates.
* For the London portion of this project, I will be using this [Wikipedia](https://en.wikipedia.org/wiki/List_of_areas_of_London) page on areas in London. The data set will be web scraped from the table under “Other use of place names” and for the sake of comparison I will assume the district location is synonymous to the neighborhood value in the NYC data set. This table also contains the names of the boroughs and grid ref geographic coordinates.
* To identify venues within the neighborhoods, I will be using the Foursquare API. I will then conduct an analysis using K-Means clustering across both cities to find similar neighborhoods on the basis of mean venue frequency.