**IBM Data Science Professional Certificate Capstone**

Ha Nguyen

**I – Business problem**

As a lazy person, I would always love to check out as many signature sights in the city as possible when I go out. Yet, due to time constraints, I need to find the most effective routes to take to maximize the venues visited and minimize the time and money spent on transportation.

The first part of the assignment explores the clusters of places of interest in San Francisco. The second part attempts to find the shortest path to traverse through each cluster.

**II – Data**

This assignment uses zip code and neighborhood information from healthysf.org. Specifically, I use the BeautifulSoup4 library to scrape the zip code – neighborhood table from the webpage. To retrieve the coordinates of each zip code, I use Google geocoder. To retrieve popular venues in each zip code, I use the Foursquare API.

**Bibliography**

Katcher, B., Reiter, R. (2004). San Francisco Neighborhoods as ZIP Codes. Retrieved from <http://www.healthysf.org/bdi/outcomes/zipmap.htm>