**COMP 4106Project Proposal**

Tyler McBride

*100888344*

**K-Means Clustering of Latitude & Longitude Coordinates**

**Problem Domain**

Given the set of latitude and longitudes coordinates,

The problem is to split into clusters of roughly size , where ,

Each cluster has a respective center coordinate, and the problem is to minimize the sum of the squared Euclidean distances to these respective cluster centers. In other words, achieve the minimal distance between each cluster center and that of each element within the respective cluster. In mathematical terms, the problem is to find the minimum for:

Where is the respective cluster center for .

**Motivation**

I have volunteered with the Peterborough Youth Soccer Club for many years, and I am currently developing an application for their convenors to create and manage teams. Growing up, I too played in the Peterborough Youth Soccer Club league and found that often times my family had to drive across town to my soccer game.

With the solution to this problem, I am trying to both reduce gas consumption and save time by having members of each team live in proximity to one another; therefore, teams can be scheduled to play games at fields within close proximity and even potentially carpool to said games.

**AI Techniques**

Most likely, a variant of the k-means clustering algorithm will be used; however, since this algorithm is NP-hard, a heuristic algorithm will have to be used to provide an optimal solution, such as Lloyd's algorithm. The only issue is that the k-means clustering algorithm does not enforce equal sized clusters; therefore, in the worst case possible, some post-processing logic may be required. More in depth research will be required to find a variant that works for the proposed problem.