**Introduction/Business Problem**

To reduce the frequency of bicycle-related collisions in a community, an algorithim will be developed to identify collision hotspots and the type of collisions so that appropriate cycling infrastructures (i.e. bike lanes, traffic lights, signage) can be implemented to improve cyclist safety at those hotspots.

Cluster analysis will be used to identify hotspots with a high concentration of collisions so that resources can be efficiently allocated to these hotspots. The government can use the density of clusters to prioritize which hotpsots to work on first.

**Data**

To perform the analysis, the following data will be needed:

1. A subset of the data with cycling-related collisions only. This will be achieved by applying a filter to the attribute “COLLISIONTYPE” where the value is “Cycles”.
2. Geographical coordinates of the collision, by using attributes “X” and “Y”.
3. The attributes “JUNCTIONTYPE” will provide the type of location where the collision occurred. This can provide insight to the type of infrastrucure to adopt to improve safety.