**Timeline:**

April 4: Go through some papers and see what kind of clustering algorithms to use and find the code, at least use 3 algorithms for comparison (done)

April 5-10: Code, run algorithms using multiple clustering algorithms

* RQ 1: April 10: What are the most accident-prone areas of the city?
* RQ 2: April 6: When do the crashes happen more often?
* RQ3: April 7: What is the most common factor for road accidents?
* RQ4: April 8: Is there any correlation between different contributing factors?
* RQ5: April 9: Is there any pattern in traffic accidents considering different contributing factors?

April: 11: Evaluate the clusters and RQ6: Is there any outlier in the dataset?

April 12: Prepare the demo and presentation

**Update:**

**April 4:**

1. Python cluster evaluation method that I am going to use: Rand Index

<https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation>

1. Python clustering methods in scikit learn: [https://scikit-learn.org/stable/modules/clustering.html#](https://scikit-learn.org/stable/modules/clustering.html)
   1. I am going to use: K-means, Hierarchical Clustering, DBSCAN, Optics, BIRCH, Agglomerative (if not covered under Hierarchical clustering)
2. Useful resources:
   1. <https://builtin.com/data-science/data-clustering-python> Implements an example using k-means and finds suitable k using the elbow method
   2. Kmeans and agglomerative clustering medium article: <https://towardsdatascience.com/an-introduction-to-clustering-algorithms-in-python-123438574097> Also saved as a pdf in resources folder named K means medium
   3. Example of 10 clustering algorithms in python <https://machinelearningmastery.com/clustering-algorithms-with-python/>
   4. Another downloaded pdf inside resources on spatial analysis
   5. <https://dges.carleton.ca/CUOSGwiki/index.php/Analysing_Traffic_Accidents_Using_QGIS_-_Heatmaps,_Hotspot_Analysis,_and_the_Time_Controller_Panel>
   6. Label encoding: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>

**Report guidelines:**

The project report should be in the style of a research paper, and must be submitted as a pdf file. The report should be between 10 and 20 pages (maximum!) in a 12 pt font (not smaller), and with a 1.5 line spacing. It must include at least the following thematic sections:

An Introduction that includes the problem definition and how it relates to previous work.

A Methods section that describes your proposed approach or solution, challenges faced, etc.

An Experimental Evaluation section that describes the experimental setup, data sets, data processing, etc, and discusses the obtained results.

A Conclusions section that summarizes and discusses what you did and its results.