Capstone Three: Project Proposal
August Perez

- Problem Identification
  - Problem statement formation
    - Produce a model for characterization of road locations combined with other conditions (ex. weather, time, day of week, etc.) where vehicle accidents occur; producing a combination of outputs for locations with highest accident rate (> =60%) and time-series for highest accident rate (>=80%).
  - Context
    - Agencies that respond to vehicle accidents would benefit from pre-planning where to allocate resources and personnel before an accident occurs to decrease response time. Determining general locations to station personnel and also determining peak times for accident risk will be of great use.
  - Criteria for success
    - The model(s) will produce the top 5 location types with >=60% chance of vehicle accident and produce time series predictions for periods where probability of accident is >=80%.
  - Scope of solution space
    - Within the jupyter notebook created to build the model.
  - Constraints
    - Some null values within the dataset may have to be imputed or the entire entry removed if multiple nulls are present.
    - Possible lack of data but preliminary explorations suggest adequate feature count.
  - Stakeholders
    - Agencies that respond to or allocate resources for car accidents
  - Data sources
    - The dataset obtained from kaggle
      - https://www.kaggle.com/datasets/ahmedlahlou/accidents-in-france-from-2005-to-2016

- What is the problem you want to solve?
  - Determine types of locations that have the highest probability of a car accident occurring (ex. T intersection during snow with low visibility) and also performing a time-series analysis for identifying which days of week and also times of day that have the highest accident chance.