IBM Capstone Project - Allegheny County / City of Pittsburgh 2017 Crash Data

## Introduction / Business Problem

Alleghney County in the City of Pittsburgh government official wants to analyze the data collected for car accidents. They want to know by using data science methodologies and algorithms of the circumstances recorded, whether there is a way to decrease the frequency of accidents. The decrease of accidents can help city officials better plan their personnel hours and resources where needed most. The goal is to plan resources more efficiently and effectively based on the predicted severity type of an accident. If an accident severity is level 1 (killed) or 2 (major injury), then more resources should be directed to help accident assistance or to educate the drivers. If an accident severity is level 3 (moderate injury) or 4 (minor injury), then less resources can be planned.

## Data

The historical data of the crash data in 2017 contains locations and information about every crash incident reported to the police in Allegheny Country in 2017. This will be used to train and test the developed model.

Data will be used as an input to build a machine learning mechanism to predict the severity of the types of the accident based on key factors that might caused the accident or contribute to the accidents. The key factors can include “alcohol usage,” “cell phone usage,” “day of the week,” “driver age,”

“icy road,” or even “deer indicator”, etc. The prediction is the “Max\_Severity\_Level”.

The results will be evaluated using the logistic regression methodology if possible.