Groups members:

Shabab Karim 1629484

Eduardo Sanchez 1528670

Mubashir Khan 1521657

MATH 4323, Project: Data Questions & Models.

**1. Main objectives/goals for our data selection & their importance**

The goal for our data set is to be able to classify streets into a binary classification (dangerous/not dangerous). We plan to use all the predictors in the dataset in order to classify the streets. Some of the predictors are straight forward and can tell us the severity of the crimes, but more importantly, the natural predictors such as *occurrence date* and *occurrence hour* will help separate streets apart. Some streets might have more high-risk at certain times of the day vs others.

**2. Models and methods selection in order to analyze and answer our questions**

Supervised, Classification

Models: KNN and SVM variations for non-linear model

Kernels reduce computational expenses over enlarging feature spaces in SMVs

1. Support Vector Classifier allows some violations
2. Polynomial Kernel
3. Radial Kernel

Model Validation: Validation Set, LOOCV, K-fold cross-validation,

**3. Workload distribution**

Eduardo Sanchez – holder

Shabab Karim – holder

Mubashir Khan – holder