**CIS 3200 Term Project Tutorial**

**Authors:** Dylan Mora; Heang Chua

**Instructor:** Jongwook Woo

**Date:** 12/12/2020

**Lab Tutorial**

**Chicago Crime Data (2015 - 2017)**

**Objectives:**

The main purpose of this project tutorial is to find a dataset which exceed 150MB requirements and provided a detailed step-by-step instructions for others to follow our example.

**Step 1: (How we located the data and how to download it and add into the programs used)**

We started by first researching for a large dataset that would reach the 150MB requirements needed for the term project. We first downloaded the dataset onto our machine which took about 4 minutes. The file downloaded was a large file as it consisted of having the data from 2001 up to Decemeber 3rd,2020. The data ends on December 3rd because it is always being updated with recent crimes.

After having difficulty opening the file on Excel, we decided to download three separate data sets: 2015, 2016 and 2017. By doing so, we were able to view our data in a .csv file which was then uploaded to Kibana and Azure Machine Learning.

**Adding Data Set to Kibana:**

1. After we had the three individual .csv files, we logged into Kibana where we selected the “Upload Data From Log File” link.

Graphical user interface, application

Description automatically generated

1. We drag our 2015.csv file onto the interface and then it will present a preview of our dataset. At the bottom of the preview, we can see the file stats presented which will have the attributes and also give the top values for each. After confirming the data, we select the “Import” button.

Graphical user interface, text, application

Description automatically generated

1. Here we would give an index name, we kept it simple and named it after the year of the dataset and then press “Import”. Then Kibana will upload the data which will take a few seconds.

Graphical user interface, text, application, email

Description automatically generated

1. Once the upload is complete, repeat the first three steps to add the remaining two files, 2016 and 2017.

**Adding Data Set to Azure ML:**

1. After logging into Azure ML, we want to add our dataset by selecting the “New”.

Graphical user interface, application, table

Description automatically generated

1. We will then select the “2015.csv” file from our local documents and leave all the settings default followed by selecting the checkmark to continue. After about 1 minute, the upload will be complete.

Graphical user interface, application

Description automatically generated

1. Once the upload is complete, repeat the first three steps to add the remaining two files, 2016 and 2017.

**Step 2: (Visualizations with code)**

**Reference:**

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/data