Sai Vinay Pabbisetty, YM53858, DATA 606 Capstone Project

**Second Deliverable Summary**

**Literature Review:**

Crime is one of the prevailing and concerning facet in any society. The increase in crime rates is one of the causes for the alarm. Law enforcement agencies, intelligence agencies and police maintain crime databases. The crime data can be analyzed to gain insights and to extract knowledge from it. Several studies have discovered number of techniques to analyze the crime data. The crime data analysis can provide the crime statistics of a region, country or world. The law enforcement agencies can take better decisions for prosperity of the citizens by understanding the various parameters that influence the crimes. There are some existing tools and softwares that are being used by police like Hunchlab, COPLINK by IBM and Accurint Crime Analysis which are based on risk terrain modeling by using several data pointers like location, description of crime etc. In my Project, I will try applying new algorithms and techniques for the crime analysis and predictive policing & patrolling and see how the results differ when compared to the existing tools and softwares.

**Exploratory Data Analysis:**

Before starting the Exploratory data analysis, I cleaned my data in three steps. Firstly, I took care of Null values without losing a large part of the data. And then I dropped some columns which are not useful for my analysis as of now, But I will drop some more columns in future if required by my analysis. Finally, I looked if the data set has any missing values and prepared the data for analysis.

Firstly, to know how each column is affecting other column I have plotted a correlation heat map. After looking at the correlation heat map I understood how each column is affecting other columns or the desired target column. And then I gained some insights by plotting graphs on frequently committed crimes and frequently committed crime locations. These two graphs will give us a good idea about the frequently committed crimes and frequently committed crime locations. After plotting these graphs, I focused on how the number of arrests has changed over the years by plotting weekly arrests, monthly arrests and yearly arrests over different years till now. In future I am planning to explore the data further and gain more useful insights like predicting the type of the crime by analyzing crime in Chicago through machine learning using different clustering and classification algorithms and other things that I have discussed in my project proposal.

**References:**

* **Varan, Shyam. (2007). Crime Pattern Detection Using Data Mining. 41 - 44. 10.1109/WI-IATW.2006.55.**
* **X. Zhang, Z. Hu, R. Li and Z. Zheng, "Detecting and mapping crime hot spots based on improved attribute oriented induce clustering," *2010 18th International Conference on Geoinformatics*, Beijing, 2010**
* **Nathan Holt, Analyzing Crime in Chicago Through Machine Learning, ,Unpublished**