**Chapel Hill Crime Data Analysis Report**

Team member: Chenlu Meng, Cai Chen, Yimei Fan

10/6/2019

**Problem Statement**

The main three problems answered by the data analysis are:

1. How’s the satisfaction level of public safety in Chapel Hill from 2011 to 2015? How are the incidents and police arrests situation in these years?
2. For each safety-related issue, could we have the in-depth analysis to help with public safety?

**Data Resources**

The data resources is Chapel Hill open data. The dataset includes [Police Incidents Reports Written](https://www.chapelhillopendata.org/explore/dataset/police-incident-reports-written/information/?disjunctive.offense&disjunctive.reported_as&sort=date_of_occurrence), [Police Arrests Made](https://www.chapelhillopendata.org/explore/dataset/police-arrests-made/information/?sort=date_of_arrest) and [Q5 Public Safety Police Services](https://www.chapelhillopendata.org/explore/dataset/q5-public-safety-police-services/information/).

**Analysis Process**

Firstly, after reviewing the public safety community survey data, we found out that the survey includes many dimensions of evaluation. In order to figure out which dimension is the most representative one among other dimensions, we conducted the correlation matrix among all different assessment variables in public safety data and figure out that Q5g, the overall performance of the police department, is the one which has the highest correlation with other assessment features. We decide to use Q5g as the evaluation variable of public safety.

Secondly, based on taking Q5g as the evaluation variable, we calculated the percentages of satisfaction and dissatisfaction on Chapel Hill’s public safety in 2011, 2013 and 2015 and plot the trend. We find that satisfaction increased by year and dissatisfaction decreased by year. Then we conducted a distribution analysis of both incidents happening based on their location and public safety evaluations in tableau. We found that locations that had the most likelihood of incidents were regions that have zip code 27517, and 27514 and most people are satisfied with the safety environment. We could also observe a stable trend of incidents happening by year and a slightly increasing trend of arrests, which means police were actively engaging in law enforcement at that time. Thus, increasing public safety satisfaction is probably due to the police’s improved visibility and attitude.

The third part is an in-depth analysis of arrest data. We trained a K-nearest-neighbor model to predict potential crime based on existing crime records. The results show that for some Primary Charge category like Impaired driving DWI, the model could predict the results in relatively good accuracy. Thus we could further prevent the town from potential crime categories and contribute to a better public safety environment.

**Conclusion**

Based on the above analysis, we reached two main conclusions:

1. The satisfaction level of public safety was improving from 2011 to 2015, given constant number of crimes.
2. The best primary charge category to be predicted are crimes are failing to appear in court and driving while intoxicated in Chapel Hill given the existing database.