**Proposal**

------------------------------------------------------

We are interested in exploring the Boston Crime Dataset which actually depicts the set of crime incidents happened in Boston city over different area. The Dataset is provided by the Boston city police department where all the records were specified from June 14, 2015 to September 3, 2018. The dataset gives us inspiration about predicting future crimes and it will be beneficial in all aspects. The different types of crimes [‘Offense Description’] belongs to the different crime groups [‘Offense\_code\_group’] along with the occurrence timing and with the location area. Our main goal will be to make a predictor model which will be able to predict the future crime incident mapped with its respective area. Now, referencing from the previous work we already have some result of showing the impedance of crime zones in the Boston city which is demonstrated as a Heat-Map. The next Goal will be to show the area’s crime effectiveness more accurately by aligning the Heat-map with data of each area. For a reference we have recently looked a census data contains every zip codes and we have street names in our Boston Data. So which zip code contains which streets we have to find that and after that we need to visualize the crime effectiveness of each area. When we will be able to achieve that then we will able to show future Heat-Map predictions of different crimes on each area of Boston.

If we able to achieve the first two milestones in a good manner then we keep our final milestone which is mapping the demographic distribution with our findings. The census data we have currently tells us about the economic and racial overlays, and the higher-crime areas were typically lower-income, and were predominantly non-white. This part could be challenging for us because there are demographic differences for high and low crime areas. We were thinking we could merge the census data used in those maps if we're feeling ambitious and demonstrate the covariance between crime, neighborhood, race, and income. The inspiration question could be phrased like “what race of people possibly could a particular group of crime in which particular area or neighborhood?”

Our Goal will be a chain of exploratory Data-Analysis and Prediction. We will update accordingly our work progress.

Below are the references for reviewing the dataset and the Project.

[1] Dataset: <https://www.kaggle.com/AnalyzeBoston/crimes-in-boston#crime.csv>

[2] Census Data:  ([https://statisticalatlas.com/place/Massachusetts/Boston/](https://statisticalatlas.com/place/Massachusetts/Boston/Industries))

[3] Reference Findings: <https://www.kaggle.com/heesoo37/boston-crimes-starter-eda>

**Tools we will use**: R studio, Python

disclaimer : We will declare the set the milestones along with the timeline as we are dedicated to this project.

[Team Members: Alex Amoore, Reetayan Das] Dated: 09/24/2019