Update One, The Final Project (Draft)

By: Daniel Lewis

For: IST 687 – Introduction to Data Science

Draft Table of Contents

[1 Introduction 3](#_Toc1649584)

[2 Business Questions 3](#_Toc1649585)

[3 Data Acquisition, Cleansing, Transformation, Munging 3](#_Toc1649586)

[3.1 Describe your data acquisition process 3](#_Toc1649587)

[3.2 What data did you select, all, subset, why 3](#_Toc1649588)

[3.3 What was your initial quality assessment 3](#_Toc1649589)

[3.4 What fields/variables did you finally decide on, why 3](#_Toc1649590)

[3.5 Provide a data dictionary 3](#_Toc1649591)

[3.6 Provide data descriptive statistics, rows, str, 4](#_Toc1649592)

[3.7 Did you have to do any cleansing, describe 4](#_Toc1649593)

[3.8 Graphs, charts, tables, text 4](#_Toc1649594)

[3.9 Interesting findings ? 4](#_Toc1649595)

[4 Descriptive statistics 4](#_Toc1649596)

[4.1 Provide demographic statistics ie. 4](#_Toc1649597)

[4.2 Any early observations, nuggets of interest, interpretation, interesting findings 4](#_Toc1649598)

[4.3 Graphs, charts, tables, visuals, text 4](#_Toc1649599)

[5 Use of modeling techniques 4](#_Toc1649600)

[5.1 Linear modeling 4](#_Toc1649601)

[5.2 Association Rules 4](#_Toc1649602)

[5.3 Support vector 4](#_Toc1649603)

[5.4 Provide key statistics of interest and interpretation for each model 4](#_Toc1649604)

[6 Overall interpretation of results/Actionable Insights 4](#_Toc1649605)

[6.1 Are the results actionable (as compared to just interesting). 4](#_Toc1649606)

[6.2 Are your recommendations/actionable insights supported by your models, if yes, indicate which model supports which recommendation. 5](#_Toc1649607)

[6.3 Actionable recommendations to management 5](#_Toc1649608)

[7 Conclusion 5](#_Toc1649609)

[8 Appendix 5](#_Toc1649610)

[9 References 5](#_Toc1649611)

# 1 Introduction

In hopes to provide safety data to those moving to the City of New York in the year 2019, this report will evaluate the shooting incidents within the city. In order to do this the student has referenced the New York Police Department (NYPD) public facing records for shooting crime violations during the year 2018. This sourced file was provided free to all on the internet and has used the open source program R to analyze this data and present it in a way that anyone can see and understand. This file contains 115,326 data points that will analyzed.

# 2 Business Questions

This project will attempt to answer the following two questions:

* What time of day on average do most of the shootings take place?
* Which borough and precinct has reported the most shootings during this period?
* How outside the median of this data set is this precinct?

# 3 Data Acquisition, Cleansing, Transformation, Munging

## 3.1 Describe your data acquisition process

The data for this analysis was acquired from the State of New York’s public dataset repository. Specifically, the subsection for the [NYPD’s section](https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Year-To-Date-/5ucz-vwe8) of the website. This data can be found on the database hyperlinked to this text.

## 3.2 What data did you select, all, subset, why

For this report I have selected a subset of this data. Specifically, the time, borough, and precinct.

## 3.3 What was your initial quality assessment

The initial quality assessment of this data was that it needed to be cleaned up a bit to be used by R due to extra data being present as well as data being listed as a number but showing that it has commas within.

## 3.4 What fields/variables did you finally decide on, why

The following columns will be valid in order to process this information.

* Incident\_KEY – this a Primary Key that could be used to align data
* Occur Date and Occur Time – these two columns are key to plotting out a chart showing the crimes on a xy line graph.
* Boro – This will be key to group by the data into the boros the are in
* Precinct – To further identify which departments have the worst crime rate
* The qualities columns of each can be used to bring out more data based on race gender etc.
* Finally, it could be possible to take the lat long data and generate a heatmap of sorts over the city.

## 3.5 Provide a data dictionary

TBD

## 3.6 Provide data descriptive statistics, rows, str,

TBD

## 3.7 Did you have to do any cleansing, describe

TBD

## 3.8 Graphs, charts, tables, text

TBD

## 3.9 Interesting findings ?

TBD

# 4 Descriptive statistics

## 4.1 Provide demographic statistics ie.

**Can you explain more about this in class?**

## 4.2 Any early observations, nuggets of interest, interpretation, interesting findings

TBD

## 4.3 Graphs, charts, tables, visuals, text

Line Graph Showing Each Burroughs and amount of shootings reported throughout the day

Second Line Graph Showing In only the shootings that resulted in injuries reported

Charts generated for each borough specifically

# 5 Use of modeling techniques

## 5.1 Linear modeling

TBD

## 5.2 Association Rules

TBD

## 5.3 Support vector

TBD

## 5.4 Provide key statistics of interest and interpretation for each model

TBD

# 6 Overall interpretation of results/Actionable Insights

## 6.1 Are the results actionable (as compared to just interesting).

TBD

## 6.2 Are your recommendations/actionable insights supported by your models, if yes, indicate which model supports which recommendation.

TBD

## 6.3 Actionable recommendations to management

TBD

# 7 Conclusion

Summarize here

# 8 Appendix

Code will be entered here.

# 9 References

Name etc