|  |  |
| --- | --- |
| requirement specification  Gun Crimes in America | Analysis of the American Gun Crimes. American gun laws appear to be out of control. Is this true, or are their gun laws the easy blame for high crime rates? Are states with stricter approaches to gun laws achieving lower crime rates?  Charlene Moore  Data Analytics x15412048@student.ncirl.ie BSc (Hons) in Technology Management |

Table of Contents

[1 INTRODUCTION 2](#_Toc529901688)

[2 METHODOLOGY 2](#_Toc529901689)

[3 USER REQUIREMENTS DEFINITION 2](#_Toc529901690)

[4 REQUIREMENTS SPECIFICATIONS 2](#_Toc529901691)

[5 INTERFACE REQUIREMENTS 3](#_Toc529901692)

[6 SYSTEM ARCHITECTURE 4](#_Toc529901693)

[7 SYSTEM EVOLUTION 4](#_Toc529901694)

[BIBLIOGRAPHY 4](#_Toc529901695)

Document Control

Revision History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Version** | **Scope of Activity** | **Prepared** | **Reviewed** | **Approved** |
| 13/11/18 | 1 | Create | Charlene |  |  |
| 23/11/18 | 2 | Update | Charlene |  |  |

Distribution List

|  |  |  |
| --- | --- | --- |
| **Name** | **Title** | **Version** |
| Eugene O’Loughlin | Supervisor |  |
| Eamon Nolan | Lecturer |  |
| Unknown | Second Marker |  |

Related Documents

|  |  |
| --- | --- |
| **Title** | **Comments** |
| Balsamic used for Mock ups for Tableau Dashboard |  |
| Use Case for Tableau Dashboard created using Draw.io @ @ <https://www.draw.io/> |  |
| GitHub Account @ https://github.com/charlene2225/Final-Year-Project |  |

## 

## **INTRODUCTION**

## **Purpose**

The purpose of this document is to set out the requirements for the Statistical analyse I am required to conduct on a large amount of data all based on a topic of my choice. The topic I chose to analyse is Gun Crimes in America. I hope to gain insights into the data and create visualisations displaying my data for users to view.

## **Project Scope**

Anticipated project start: 18/09/2018

Project Pitch Video: 07/10/2018

Project Proposal: 21/10/2018

Ethics Document: 18/11/2018

Requirements Document: 18/11/2018

Mid-point and prototype presentation: 17/12/2018

Estimated completion date: 06/05/2019 (approx.)

Final Presentation date: 13/05/2019 (approx.)

Project Objectives: To analyse a large amount of data obtained from more than one dataset. The data must be cleansed. Analysis must then be conducted to determine trends, links, patterns and prediction between the datasets to gain a depth into the topic of Gun crimes in America. I will introduce machine learning into my project to predict things like, when crime is most likely to occur and where it may happen also. These predictions may be used by the American Bureau to help lower the crime rates in America in the future.

Goals: My main goal in this project is to incorporate what I have learned over my 4 years of college to produce a structured report on gun grimes in America. Constructed from intensive research on the topic and a thorough analysis on datasets containing real life data on the matter. I hope to achieve an informative report for users to read and gain a clear picture of gun crimes in the areas my analysis is conducted on. My reasoning behind this is due to people not being educated on the subject yet voicing an opinion on the matter. I understand everyone has an opinion and is fully entitled to it, I’m not disagreeing. I just think an opinion on such a delicate topic should come from a bit of research into the topic. I hope this report provides those interested with an insight into the matter to help them form a opinion based on facts rather than hear say on social media or other unreliable sources a majority of the time. I must point out though, that the report I am constructing from my analysis is based on Data I have accessed from Kaggle. (Kaggle, 2018) I will be using RStudio for the majority of my project as it is a new language for me to explore with.

Tasks: There are many tasks I will be undergoing to achieve my goal of a detailed analysis on Gun Crimes in America. These tasks are outlined in more detail in my Project plan further down in this document.

Key tasks are the following:

1. Pick a topic to base my analysis on.
2. Search for a dataset suitable for the topic chose.
3. Download the dataset(s) and start the cleansing process. A combination of Excel, SPSS, MySQL and RStudio will be used for the cleansing and sorting of data.
4. Conduct research on Gun Crimes in America for suitable case studies or articles to help support the analysis being carried out.
5. My descriptive Statistical analysis will be carried out with both Excel and RStudio in the preliminary stages of my analysis.
6. Create Visualisations such as bar charts, pie charts and histograms on the data to visually provide insight into Gun crimes in America. I hope to create a dashboard of interactive visuals in Shiney and Tableau. But a few of my preliminary analysis may be conducted creating visuals using RStudio. (Shiney.RStudio.com)
7. In semester 2 I hope to deepen my search into my datasets with the use of Machine learning algorithms such as k-Nearest Neighbours (kNN), Random Forest, Decision Tree and Regression. I will discuss these in a bit more detail further down in the document as to why I will be using them.

## **Definitions, Acronyms, and Abbreviations**

* (RS) Requirement Specifications
* (kNN) K-Nearest Neighbour
* DB = Database
* Approx. = Approximately
* Etc., = Et cetera
* MySQL: My Structured Query Language
* KDD = Knowledge Discovery and Data Mining

Some of these aren’t in this doc yet, so edit them before final submission.

## **METHODOLOGY**

Using a combination of research methodologies, Quantitative and Qualitative Analysis. I will be doing Quantitative Analysis as this conclusive, for example how many times a certain crime occurs, in this case gun crime. This doesn’t measure why it occurs. This is where the combination of Quantitative and Qualitative comes in. I will then support my analysis with Qualitative analysis on factors to consider about the data seeking a deeper understanding of why it may occur. These factors may be nationality, gender, poverty, education etc., factors that are descriptive of the Quantitative data such as the temperature, Weight, height etc.,

Quantitative Analysis quantifies data to test Hypothesis or predict the future whereas qualitative analysis seeks to get a deeper understanding of why certain things occur.

The methodology approach I have chosen to use for my project is Knowledge discovery and Data Mining Methodology (KDD) My reason for choosing this approach is because it ….

## **Selection**

## **Pre-processing**

## **Transformation**

## **Data Mining**

## **Evaluation**

## **USER REQUIREMENTS DEFINITION**

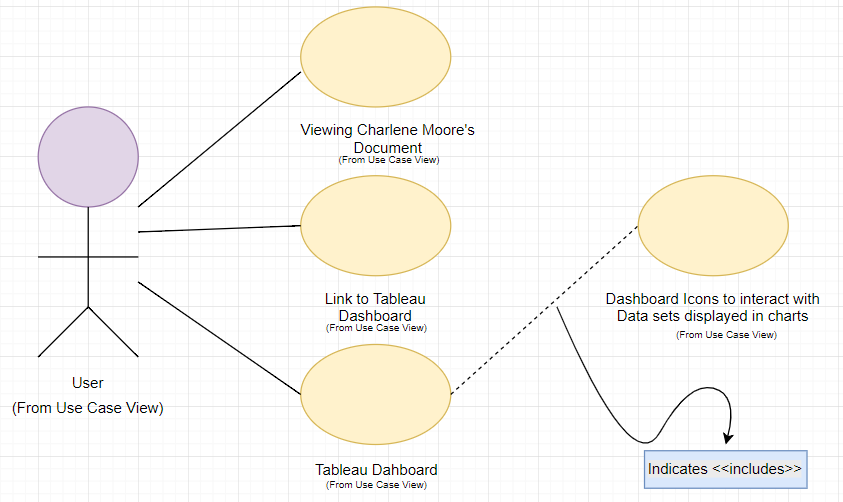
A small bit of research was conducted to see what US Bureaus use to help predict their crimes. And it appears that this seems to happen with a combination of

## **REQUIREMENTS SPECIFICATIONS**

## **Functional Requirements**

The below use case is to display the interaction of the user with the Tableau dashboard that will contain my visuals.

* + 1. Use Case Diagram



* + 1. Requirement 1

Get datasets? Will I need to do use cases for the fact I’m pulling from different datasets? Should I use link to pull data directly from Kaggle for updates? Dashboard on Tableau?

3.1.2.1 Description & Priority

3.1.2.2 Use Case

**Flow Description**

**Precondition**

**Activation**

**Main flow**

**Alternate flow**

**Termination**

**Post condition**

* 1. Non-Functional Requirements

Specifies any other particular non-functional attributes required by the system. Examples are provided below. **Remove the requirement headings that are not appropriate to your project.**

3.2.1 Performance/ Response time requirement

3.2.2 Availability Requirement

3.2.3 Recover Requirement

3.2.4 Robustness Requirement

3.2.5 Security Requirement

3.2.6 Reliability Requirement

3.2.7 Maintainability Requirement

3.2.8 Portability Requirement

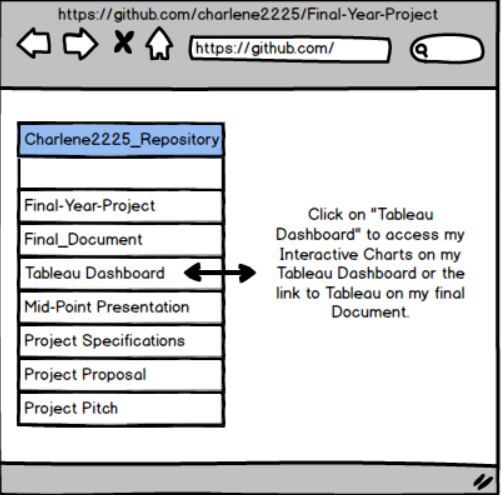
3.2.9 Extendibility Requirement

3.2.10 Reusability Requirement

3.2.11 Resource utilization Requirement

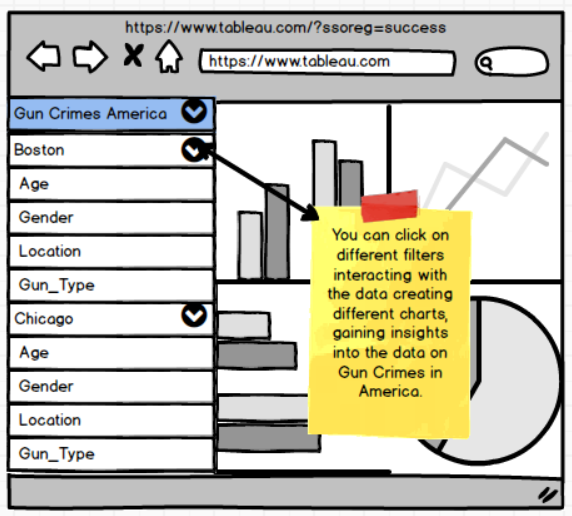
## **INTERFACE REQUIREMENTS**

## **GUI**



Below is a mock-up of how the data will look on a dashboard in Tableau. There is a filter system that users can use to filter the variables of the data they wish to have displayed on the charts.

Include screenshots of Tableau Dashboard when completed



## **SYSTEM ARCHITECTURE**

I will be using R Studio as the programming language for my project and I am considering using MySQL to store the data in a database. Or continue to pull from the CSV files stored in my working directory on RStudio.

**R** is a vectorized programming language that is used mainly for data mining and analytics of statistics. It was first developed in 1993. **R:The R Project for Statistical ComputingR-project.org. (2017)**

**MySQL** is an open-source relational database management system. Its name is a combination of "My", the name of co-founder Michael Widenius's daughter, and "SQL", the abbreviation for Structured Query Language. **Dev.mysql.com. (2017)**

## **SYSTEM EVOLUTION**

The potential evolution of this tool would be to expand the???

## **BIBLIOGRAPHY**

* <https://www.macmillanihe.com/studentstudyskills/page/choosing-appropriate-research-methodologies/> (used this page to help me choose what type of research methodology I am using for my project and to gain an understanding on the different ones I can chose between) [website, online]
* (Kaggle, 2018) `Dataset’s`, Public [Online] Available at: <https://www.kaggle.com/datasets?sortBy=relevance&group=public&search=crime&page=1&pageSize=20&size=sizeAll&filetype=fileTypeAll&license=licenseAll> (Accessed: 29 Sep 2018)
* (Pyle D, no date) [Book online.] `Data Preparation for Data Mining` <https://books.google.ie/books?hl=en&lr=&id=hhdVr9F-JfAC&oi=fnd&pg=PR17&dq=facets+of+data+mining&ots=6gdS7NHtcs&sig=4zW6TgTUtmu1AWB5r_iVhtX4WrI&redir_esc=y#v=onepage&q=facets%20of%20data%20mining&f=false>