|  |  |
| --- | --- |
| Student Name: Gavin O’NeillStudent Number: D15127205 **Phone Number:** 0834587520 | **Stream:** Data Analytics |
| Project Title: An Analysis of Cybercrime in the United States | |
| Introduction This project is being undertaken with the intention of showing how cybercrime has developed over the past 10 – 15 years in the United States and will give a view of which direction trends are moving towards in the near future. It should give the reader a breakdown of the various types of cybercrime that occur with detailed analysis on areas such as:   * Who is most susceptible to becoming a victim (e.g. gender, age, etc.) * Which geographical locations are targeted by criminals * The financial cost to victims * Its effect on organisations and individuals regarding security policies   This subject area is very broad so I hope that my analysis will allow me to narrow down on the main aspects and give the reader a better understanding of online crime. | |
| **Background**  I have chosen to focus my analysis on cybercrime as I currently work as a fraud analyst for a FinTech company and I’m very interested in how both the attacks and security measures are evolving continuously. This evolving nature of Internet crime presents a unique set of challenges, as crimes can overlap jurisdictional boundaries and perpetrators can attack from anywhere on the globe.  The scope of this area is huge and can include a variety of crimes from hacking, copyright infringement, espionage, financial theft and many more. Attacks can take place at various levels, from large corporate or government organisations all the way down to a casual user through common methods such as phishing and malware. The cost of cybercrime worldwide is reported to be in the billions of dollars worldwide and it is expected to grow indefinitely.  Cybercrime data for the United States is the most comprehensive I could find at this time, hence why I have chosen this region specifically but with some more time to search for data I may broaden the scope to include analysis on a more international level. | |
| **Aims and objectives**  The aim of the project is to present an analysis of how Internet crime has evolved over the past 10 – 15 years, breaking the data down into the types of crimes committed, how much damage is caused financially, geographical hotspots for attacks, who are targeted by these attacks and why were they chosen. I’m also hoping I can bring in other relative data such as increases in spending and jobs in the Internet security sector over this time period. With the use of the correct models I want to finally give some predictions or indications of what cybercrime will entail in the coming years.  My learning objectives are to become proficient in the use of R and SAS, learn how to incorporate probability and statistical models into the analysis and gain an understanding in the various methods of data visualisation available. I would like to finish at a level where I would be comfortable carrying out analysis in a work environment. | |
| Methodology **1. Literature Review:** a review of the available literature on cybercrime which will give me the knowledge required for gathering the correct data  **2. Data Acquisition:** at the moment all the data I need is contained in tables within pdf reports and I will need to extract this data so that it will be useable for my analysis. I expect that with more time to search I will find some other data sources and everything will need to be brought together for processing  **3. Data Processing:** the data will be cleaned up and stored in an appropriate  format for analysis  **4. Preliminary Analysis** of the dataset  **5. Choice of Appropriate Model:** this will involve the choice between classification or regression models, with the aim of showing relationships between various data variables and my choices will depend on learning more in this area over the coming weeks.  **6. Modelling:** training and testing the model and evaluating its performance  **7. Recursion:** refining the model  **8. Reporting:** presenting the findings and writing the final report  Appropriate visualisations will be generated at each step. | |
| Deliverables A final report will be submitted to DIT, including:   * A description of the dataset, how it was acquired and what techniques were used to prepare it for analysis * A detailed discussion of the model used, why it was chosen, and the metrics for evaluating it * An evaluation of the model * Data visualisations   This project will finish with a detailed presentation of the findings. | |
| Project from Previous Streams that has inspired me (approx. 400 words): Title: An analysis of breastfeeding in Ireland  Student’s Approach: This project didn’t inspire me regarding the topic but rather how the project was carried out. The main focus of the project was to perform an analysis with an aim of discovering patterns and predictors which is very similar to what I hope to achieve. There are a lot of visualisations used in this project similar to what I think I will need to use, such as histograms for demographics, plotted charts for timelines, maps for geographic data and others. The methods used for predictions are something that I can look back upon when at the same stage of my own project. The conclusion section of this project also gives very clear results and I would like my project to be able to give accurate information to the reader which would help them make decisions regarding cybercrime and security. | |