Jonathan Mulleda

March 9, 2023

**PROJECT PROPOSAL: Data Analysis and Information Visualization**

CPSC 101 Introduction to Programming II

**Background**

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends, and outliers in large data sets.  
  
**Objectives**

* To effectively communicate information in a universal manner using visual information
* To pinpoint areas that need to be improved or need more attention
* To help organizations in increased understanding of the next steps that must be taken
* To help audience absorb information quickly, improve insights and make faster decisions

**Scope and Limitations**

The project will focus on the data set of distribution of students (domestic and international) in Selkirk College, Castlegar Campus. Data set from other campuses may not be available thus may not be included in the report.

**Timeline**

|  |  |  |
| --- | --- | --- |
|  | **Description of Work** | **Start and End Dates** |
| Phase One | Collection of data set | 1st week of March |
| Phase Two | Data validation and data clean up | 2nd week of March |
| Phase Three | Coding and creation of Jupyter notebook | 3rd week of March to 1st week of April |

**Approach and Possible Issues**

|  |  |  |
| --- | --- | --- |
|  | **Approach** | **Possible Issues** |
| Phase One | Consult with school admin to get the data set | Data may not be available or restricted |
| Phase Two | Arrange data in useful format | Data may be corrupted or in a non-useful format |
| Phase Three | Coding and debugging | Coding and visualization errors |

**Possible Libraries**

|  |  |
| --- | --- |
| Pandas and Numpy | Adds support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions |
| Seaborn and BeautifulSoup | Pulls data out of HTML and XML files. Provides a high-level interface for drawing attractive and informative statistical graphics. |
| Matplotlib | Comprehensive library for creating static, animated, and interactive visualizations in Python |

**Stretch goals and future lines of work**

Data analysis and visualization have grown to become an integral part of the business realm that governs many aspects of our lives in this modern-day and age. Emerging technologies like machine learning, augmented reality, virtual reality, and AI play a more significant role in effective data visualization.