Shane Matsushima

CS352

Project Proposal

The project I will be doing is showcasing the Python DSL’s of Pandas and Streamlit. Pandas is a database library that helps with organizing and collecting specific data from files such as csv. Streamlit is another library that is utilized to visualize and create web apps for displaying data information from databases in graph or other media forms. The goal of my project is to utilize and showcase both Pandas and Streamlit in Python on a database to see how visualizing data helps with better understanding of different information. The database I will be using will be pulled from Kaggle, a website used for data scientists to practice and get new databases to work on. For pandas, I will be going over the basics and how pandas are used to parse and grab data from databases. I may explain some parts of Numpy as well as Numpy and Pandas go hand in hand with one another. For Streamlit, I will be going over how to display and run streamlit with the database parsed from pandas and how there are different widgets that can be used to display that information.

More specifically, I will be explaining the reading and format of csv files used in pandas, such as delimiters and how rows / headers work. I will also be explaining how to grab specific information and how to manipulate rows or columns of data grabbed, to get a better meaning out of it (i.e average of data, etc). This is where numpy maybe used and explain as numpy arrays and pandas database object go hand in hand with manipulating data.

From there, the database or information being utilized will be showcased on a webapp created with streamlit. I will be showcasing different graphs and widgets that are able to be produced suing streamlit while displaying the data found / grabbed from the database utilizing pandas. With streamlit, I will also be showcasing findings of what the database had as well as explaining the importance of visualizing data.