

Data Science Final Project

CIT/CSC 1105
Computer Science Colloquium
Data Science Infusion
Dr. Kanampiu

E-Mail: @gmail.com
Spring 2022

Hello, my name is _____, I am a senior information technology major that attends WSSU. Today I will be talking about the step-by-step process of the data science infusion labs we have completed during Computer Science Colloquium and the possibilities on how it could help others in the future such as Data Scientist in describing data correlations. Our lab 1 kicked off the start of the data science labs, by getting us familiar with Google Colab. In this lab it was system print line where all we had to do was to enter our names as the declared name variable and to execute the following code that was given. The output would be “Hello, World! 5, John, my name is Heshimu”. Continuing over to lab 2, this program simply would add up two numbers to execute the sum. In lab 3 we started to go through the different import modules that was available in python. In this program we imported the “random” module and printed out a random number in between 0 and 100. Next would be lab 4, this would be the start of where we would be importing data and printing out plots. We would first import our selected modules, next would be to generate the data into the declared “x” variable, and then to plot the histogram from our data we used the “plt” module that was imported in the beginning. To generate our data for “x” and “y” we listed an array of numbers that represented each axis, and lastly plotted the scatter plot with the generated data inside. Our next lab would be lab 5, our outcome would be to generate and plot a scatter for x, and y values using the numpy and plt modules to plot the scatter plot. We added a linear regression that assumes a linear relationship between the input variables (x) and the single output variable (y). Below the plot we printed out r value that indicates strength of the correlation. Lab 6 we learned how to import/access an existing data set and imported a downloaded file into google colab to print the size and rows of the data set. In lab 7 we would continue with the same code from last lab but added a drop columns method that would drop the specific columns that were listed. Next would be to display the shape of the new data set with the dropped columns gone. In lab 8 we continue with the same code from lab 7 but now we are adding a visualization of the data set by making our own plot. We named the x and y axis with selected column needed and corrected the size of the plot if needed and printed the output of the plot. In lab 9 we took what we did in lab 5 and combined it with lab 8 to plot a regression graph and computed the correlation strength. Lab 10 we used a new dataset to complete this lab. The plot would visualize the correlation between Covid-19 daily deaths (number of deaths vs day). Lastly for lab 11 We computed the linear regression on the selected data categories using the python's stats library, dropped multiple data set columns that were not needed, and plotted a scatter plot with “new-deaths vs people_fully_vaccinated” in the data set and contained a linear regression line and an actual values line to finish off the lab.