CTEC 298-101 Fall 2024 2nd 7w December 5, 2024

Name:\_\_\_\_Genesis Grant\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Define the terms below using the definitions found in the tutorials that were provided to the class. Each answer will consist of the definition and the tutorial where it was found.**

Jupyter Notebook

-Jupyter notebook is a compiler that allows for R, Python and other language computations to be done within the software. This software allows for code to be executed within the notebook and see in real time how the code is working in relation to the environment. Within class we used Jupyter Notebook through Anaconda software.

-This can be found within the Anaconda/Jupyter Notebook tutorial where we also submitted our Anaconda/Jupyter notebook proof of completions.

Anaconda

-This is the software that holds tools like Jupyter Notebook and other code compilation software.

--This can be found within the Anaconda/Jupyter Notebook tutorial where we also submitted our Anaconda/Jupyter notebook proof of completions.

Kernel

-This is the powerhouse of an operating system. Within Jupyter Notebook it is what powers the notebook. It will run the code and remember any stored computations and variables.

-This can also be found with Anaconda/Jupyter Notebook Tutorials.

Ipynb file

-This is the file extension for jupyter notebooks. It will note that the file is in a jupyter notebook file.

-This can also be found with Anaconda/Jupyter Notebook Tutorials.

JSON

-A JSON file is a format used for holding data and specifically for allowing computers to read this data.

Metadata

-Metadata is data about other data. It can describe the contents, file type, create date, author, etc. This will give you descriptive information about the contents of the

Code cell

-Within Jupyter notebook a code cell is a cell that holds code and can be executed within the cell. This is beneficial as users are able to see in real time how the code interacts with the environment.

Markdown cell

-A markdown cell is a code cell that can be used for specifically writing comments or notes about the code.

Matplotlib Plots

-Matplotlib is a python library that allows coders to use the classes within it to create a variety of different plots.

-This can be found in the Matplotlib tutorial where we chose a plot to create a presentation on and develop ourselves.

Visualization

-Data visualization is the process of showcasing large amounts of data.

-This can be found in the tutorial videos, where they define what visualizations are and their significance.

1. Describe each of the 6 plots that we have used in Matplotlib

-Bar Plot: bar chart that can be used to compare different values

-Line Plot: line plot that can be used to show a trend over time

-Pie Plot: pie plot that can be used to show how different factors relate to each other as a whole.

-Scatter Plot: scatter plot can be used to show how pairs and data relate to each other

-Multiplots: use multiple plots to show different types of data relating to each other.

1. What is a Panda in this class?

-Panda is the python library that is used to data wrangle and finetune data. It can be used to hold data as well as manipulate the data like choosing specific rows/columns to showcase.

-This can be found in the Pandas Tutorials where we went through 4 videos and implemented the code into our Jupyter Notebooks.

1. How are Pandas related to Excel?

-Pandas is able to upload data from Excel into the software. This can be beneficial as the large data can be stored within the excel file, but pandas is able to finetune the data and return it with more clean results.

1. Pandas are used for what major data science concept.?

-Pandas is used for data wrangling.

1. Compare Matplotlib and Tableau for creating plots.

-Matplotlib and Tableau both create data visualizations like bar charts, scatter plots, etc. But, Matplotlib is a Python Library that can be used in code, while Tableau is a software that allows users to drag and drop data to create visualizations without having to use code.

1. Describe the numpy function of Matplotlib.

-The numpy function of matplotlib is similar to pandas, except that it can use arrays to hold data. This function allows for users to finetune data as well as compute and/or merge data. It has many uses and can be used with Matplotlib to create visualizations.

1. What is a dataframe? Give an example.

-A dataframe is a way to store data in a variable within code. Data can be uploaded different ways to the dataframe. For example, if you had an excel file and you wanted to use a dataframe to store it, you could use pandas to upload the excel file into a dataframe (df) and manipulate the data within the code.

1. Name and give an example of 3 statistics concepts that can be used with dataframes?

-Mean: If you want to find the average of a certain column of data. Ex. Average rainfall within a year

-Median: If you want to find the middle most data within a dataset. Ex.

Median income in a town

-Mode: If you want to find the most displayed value within a dataset. Ex.

Most sold product in a store

1. Describe 3 techniques for creating a dataframe using Pandas.

-You can create a dataframe using Pandas by uploading a csv file

-You can create a dataframe using Pandas by uploading an excel file

-You can create a dataframe using Pandas by defining it in a dictionary within the code.

1. 1 page summary of class(single spaced).

As I matriculate throughout Bowie State and earn my Computer Technology Degree, I will take classes that will focus on my concentration. My concentration being Data Science and Database Administration, learning about the concepts within it will allow me to explore a more subject specific area. In this paper I will summarize one of my focus specific classes, COMPUTER TECHNOLOGY 298.

In CTEC 298, Symbolic Computation using Big Data, we met virtually on Tuesdays and in Person on Thursdays @4p, to discuss data science concepts. Both online and in person, as a class I would interact with my peers to understand class concepts as well as learn from each other’s mistakes. Within the first few weeks we began with understanding data science concepts like data visualizations and data wrangling. To apply these concepts, we first had to revisit python skills. We began with visiting LearnPython.org and going through tutorials to ensure we understood basic syntax of python. They we continued with using new tools like Jupyter Notebook and Anaconda, to display this work in more organized software and become familiar with the tools. We continued with our refined python skills to data wrangle and create data visualizations using Matplotlib, Pandas and Numpy. Every week we would go through different tutorials to learn about how to apply these concepts. We put this work into two folders within Google Drive and Github to organize the work. To showcase our understanding and work throughout the class, we are assigned a project. Our project consists of using data from our prerequisite class (CTEC 128), where we data wrangled using Excel, and using new concepts (Pandas, Numpy, Matplotlib, Tableau) to showcase this same data but using the new skills we have learned throughout the class. As well as present our findings. This is beneficial as the tools we have learned in class, can be applied to real world applications and projects and with these tools we are able to enhance our skillset to employers.

In conclusion, within CTEC 298 we revisited prior subjects and topics from CTEC 128, to apply the new concepts and tools we learned to old projects and understand how using these new concepts can be more efficient and effective when creating data visualizations.