Benjamin Dragenice

CTEC 298

3/4/2021

Quiz 2

1. Jupyter Notebook – a powerful tool for interactively developing and presenting data science projects.

Cell – a container for text to be displayed in the notebook.

Kernel – computational engine that executes the code contained in a notebook document.

Ipynb File – text file that describes the contents of your notebook.

JSON – JavaScript Object Notation; open standard file format and data interchange format

Metadata – set of data that describes your notebook.

Code Cell – code to be executed in the kernel.

Markdown Cell – text formatted using Markdown and displays it output in-place.

Matplotlib Plots – lets you plot graphs using the Matplot library.

All definitions were found here: <https://www.dataquest.io/blog/jupyter-notebook-tutorial/>

1. A simple plot is a line graph used to display a series of data points connected by solid line segments. A histogram is a graph used to show frequency distributions with bars. A bar plot is a chart/graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. A pie chart shows the size of items in one data series, proportional to the sum of the items. A scatter plot is used to plot data points on horizontal and vertical axis. A stack plot is a plot that shows the whole data set with easy visualization of how each part makes up the whole.
2. Pandas is an open-source data analysis and manipulation tool used as a library in Python.
3. Pandas is related to Excel because they are both used to analyze and manipulate data. Pandas is used as a library in Python and Excel is used as its own application.
4. Pandas is used for data manipulation and data analysis.
5. Matplotlib is a plotting library for Python that produces 2D graphs/charts. Tableau is an application that allows you to create graphs/charts/maps and other visuals in 2D and 3D. Tableau allows the visuals to become interactive.
6. With Tableau you can input .xl, .xls, .json, .access, .pdf, .csv files. With Matplotlib you can input .csv, .txt files.
7. A data frame is a 2-dimensional labeled data structure with columns of potentially different types.
8. Three statistic concepts that can be used with data frames are data wrangling, data ingestion, and exploratory data analysis.
9. The first way to create a data frame from Pandas is from a list of lists. The second way to create a data frame from Pandas is from a dictionary of an array/list. The third way to create a data frame from Pandas is from a list of dictionaries.