**Assignment 3**

**1**. According to the video "Pandas for Data Science in 20 Minutes," CRUD stands for "Create; Read; Update; Delete."

**2**. The datasets shown in the video "Pandas for Data Science in 20 Minutes" are as follows.

State object

Account length int64

Area code int64

International plan object

voice mail plan object

Number mail messages int64

Total day minutes float64

Total day calls float64

Total day charge float64

Total eve minutes float64

Total eve calls float64

Total eve charge float64

Total night minutes float64

Total night calls float64

Total night charge float64

Total intl minutes float64

Total intl calls int64

Total intl charge float64

Customer service calls float64

Churn object

D type object

**3**. Users will struggle to find the relevant information, resulting in inefficient processes.

**4**.Restricting user access to data promotes data privacy and security, fosters collaboration and communication, and improves data governance and documentation. Working with smaller quantities of data, applying past knowledge, and ensuring compliance, albeit doing so does not always make developing models more complex.

**5**.These are the four Python library functions not covered in the class: Requests, SQLite 3, Kera’s, and TensorFlow.

• TensorFlow: TensorFlow, an open-source library for deep learning applications, was designed by the Google Brain Team. It was originally intended for mathematical computations, but it now provides a wide range of tools, libraries, and community resources for developers to use when developing and deploying machine learning-based applications. It has now been updated.

• Kera’s: The open-source Kera’s TensorFlow library interface enables rapid testing of deep neural networks. Francois Chollet invented it, and it was initially introduced in 2015. Kera’s provides tools for constructing models, showing graphs, and analyzing datasets. There are also prelabeled datasets that may be imported and loaded right away. It's simple to use, adaptable, and great for exploratory inquiry.

Among the four library functions not discussed in class, I considered KERAS to be one of the most important, and I loved it because it offers a wide range of options such as building models, displaying graphs, and analyzing datasets, and datasets can be immediately imported and loaded.

**6**. SciKit-Learn was built using NumPy, SciPy, and Matplotlib.

Scikit-learn was built using NumPy, SciPy, and Matplotlib for several reasons:

1. NumPy handles and computes multidimensional arrays efficiently, which is critical for describing and manipulating data in machine learning.

2. SciPy extends NumPy's usefulness by providing new scientific and technical computing tools, such as optimization techniques and statistical functions, hence increasing scikit-learns capabilities.

3. Matplotlib is used to create visualizations that help users understand data, model performance, and results, which is critical in machine learning.

Overall, these libraries offer optimized implementations, comprehensive functionality, and compatibility with the larger Python scientific computing ecosystem, resulting in scikit-learn being a powerful and adaptable machine learning toolkit.

**7**. To effectively analyze missing data and uncover potential biases, the author devised the category "Not Answered" to distinguish unintentionally missing values from purposely unanswered questions. Keep track of missing data, maintain analytical precision, and describe data constraints in a clear way. Each missing value is assigned to a specific category, emphasizing the significance of dealing with missing data while reducing prejudice and wrong interpretation.

**8**. Correlations are statistical techniques for determining the relationship or connection between two variables. They assess how closely changes in one variable correlate to changes in another. The correlation coefficients vary from -1 to 1, with -1 indicating no link.

There is a negative correlation; there is a perfect positive correlation; both variables increase or fall together; and there is no correlation at all. By identifying patterns, dependencies, and probable relationships between variables, correlations shed light on interactions and affecting variables.