1. **Explain how you would handle missing data in a given dataset and provide a code snippet demonstrating this.**

Handling missing data in a dataset is a crucial step in data preprocessing for both supervised and unsupervised learning tasks. Missing data can affect the accuracy and reliability of the analysis. Here are some common techniques for handling missing data:

1. Removing rows or columns: In this approach, you can choose to remove the rows or columns containing missing values from the dataset. This technique is suitable when the missing data is limited and doesn't significantly impact the overall dataset.

2. Mean/median imputation: This method involves replacing the missing values with the mean or median value of the corresponding feature. Mean imputation replaces missing values with the average value of the feature, while median imputation replaces them with the middle value. This approach assumes that the missing values are missing at random and the mean or median provides a reasonable estimate.

3. Mode imputation: Mode imputation replaces missing categorical values with the most frequent value (mode) of the corresponding feature. This approach is applicable to categorical variables.

4. Forward or backward filling: This technique is useful when the data has a temporal or sequential structure. It involves replacing missing values with the most recent or subsequent observed value in the dataset. This method assumes that the missing values do not change abruptly over time.

5. Model-based imputation: Model-based imputation involves using machine learning algorithms or statistical models to predict missing values based on the available data. This approach can be more accurate but is often computationally expensive. Techniques such as regression, decision trees, or K-nearest neighbors can be used for imputation.

Remember, the choice of handling missing data depends on the specific context and the nature of the dataset.