What will happen if we do not impute or replace the missing values with the mean median or mode in the dataset?

When you don't impute or replace missing values in a dataset, it can lead to several potential issues in data science using Python:

Biased analysis: Missing data isn't always random. There might be patterns or reasons why data is missing, which can introduce bias into your analysis if you ignore it. For example, if income data is missing for people with lower incomes, your analysis might overlook important trends.

Reduced accuracy: Most machine learning algorithms can't handle missing values directly. Dropping rows with missing data can significantly reduce your dataset size, leading to inaccurate models. Imputing values helps retain information and potentially improve model performance.

Unreliable interpretations: Analyzing incomplete data can lead to misleading conclusions. Without understanding the reason for missing values, you might misinterpret relationships between variables or draw inaccurate inferences.

Limited applicability: Models trained on data with missing values might not generalize well to new data containing missing values. Imputation ensures your model can handle real-world scenarios where data might be incomplete.

Here's a quick breakdown of the impact of ignoring missing values based on the chosen method:

Ignoring completely: Riskiest option, leading to potential bias, reduced accuracy, and unreliable interpretations.

Dropping rows: Reduces data size, potentially affecting accuracy and generalizability.