When to Use Mean, Median, and Mode in Data Analysis

---

1. Mean (Average)

Use when:

- Data is normally distributed (symmetric, not skewed).

- No or few outliers.

Example:

```
df["Calories"].mean()
```

Best for: Continuous numerical data with a normal distribution (e.g., height, weight).

Avoid if: Data has extreme outliers.

2. Median (Middle value)

Use when:

- Data is skewed (e.g., income, house prices).

- You want a robust central tendency that isnt affected by outliers.

Example:

```
df["Calories"].median()
```

Best for: Skewed distributions or data with outliers.

3. Mode (Most frequent value)

Use when:

- You have categorical data (e.g., most common category).

- Or when a certain value repeats often in numerical data.

Example:

```
df["Calories"].mode()[0]
```

This returns the most frequent calorie value.

The [0] is used because `.mode()` returns a Series (there can be multiple modes).

Best for: Categorical data or discrete values where frequency matters.

Which one should I use to fill missing values?

- Use .mean() for numeric data with a normal distribution.

- Use .median() for numeric data with skewed distributions or outliers.

- Use .mode()[0] for categorical or repetitive values.

Example: Fill missing values in the "Calories" column

```
df["Calories"].fillna(df["Calories"].median(), inplace=True)
```