Understanding Imputation:   
A CHEAT SHEET

# What is Imputation?

Imputation is the process of filling in missing data in a dataset using reasonable or estimated values. Instead of deleting rows with missing values, imputation helps retain valuable information by replacing the missing entries with typical or representative data.

# What are Mean, Median, and Mode?

## Mean (Average)

The mean is calculated by adding all the values in a dataset and dividing by the number of values.  
Example: For values $2,000, $3,000, $4,000, $5,000, $6,000:  
Mean = (2000 + 3000 + 4000 + 5000 + 6000) ÷ 5 = $4,000

## Median (Middle Value)

The median is the middle number when the data is ordered from smallest to largest.  
Example: For values $2,000, $3,000, $4,000, $5,000, $6,000:  
Median = $4,000 (the third number in the sorted list)

## Mode (Most Frequent)

The mode is the value that appears most frequently in a dataset.  
Example: For values $2,000, $2,000, $3,000, $4,000, $5,000:  
Mode = $2,000 (it appears twice)

# How is Imputation Used?

When data is missing, analysts use the mean, median, or mode to fill in those blanks depending on the nature of the data:  
- Use the MEAN if the data is well-balanced without extreme values.  
- Use the MEDIAN if the data contains outliers that could skew the average.  
- Use the MODE if a single value is very common and likely represents others.