Rule to fill in Missing values for your Dataset

- Step 1:
  Look into description of your data eg: df.describe()

- Step 2: Compare mean and median
  - • **mean ≈ median** → roughly symmetric → consider **mean**.
  - • **mean < median** → left-skewed → consider **median**.
  - • **mean > median** → right-skewed → consider **median**.
    - **Note: This is a first hint, not a final decision**.
- Step 3: Visual check
  - df['column'].hist(bins=30) # or
  - sns.boxplot(x=df['column'])
    - Look for:
      - Outliers
      - Multiple peaks (bimodal)
      - Skew direction
- Step 4: Check for subgroups
  - If the column is affected by categories (contract type, region, gender, etc.), compute median/mean **per subgroup**:
    - df.groupby('Contract')['column'].median()
- Step 5: Business context
  - • Ask: “What makes sense for the problem?”
  - • Example:
    - Monthly charges → median gives typical customer value
    - Age → mean might be more representative
- **Step 6: Final Decision**
  - Decide imputation value using combination of:
    - Step 2 (mean vs median hint)
    - Step 3 (visual confirmation)
    - Step 4 (subgroup variation)
    - Step 5 (domain/business sense)