**Mean, Median, MAD & Standard Deviation – Real World**

**Problem:**  
You are analyzing monthly sales data for two shops.

shop1\_sales = [2200, 2250, 2300, 2350, 2400, 4000]

shop2\_sales = [2000, 2100, 2300, 2500, 2700, 2800]

* a. Calculate the mean, median, MAD, and standard deviation for both shops.
* b. Which shop shows higher **consistency** in sales and why?
* c. Suppose the outlier in shop1 is corrected to 2450. How do the metrics change?

**2. Effect of Data Transformation on Spread**

**Problem:**  
Given the dataset:

data = [25, 30, 35, 40, 45, 50]

* a. Compute the mean and standard deviation.
* b. Now apply:
  + Addition: Add 5 to each value
  + Multiplication: Multiply each value by 2
  + Log transformation: Apply np.log(data)
* c. Discuss the effect of each transformation on center and spread.

**3. Density Curve vs Histogram**

**Problem:**  
Generate 1000 height values assuming a normal distribution (mean=160 cm, std=10 cm).

* a. Plot histogram and KDE using seaborn.
* b. Manually create bins (intervals of 5 cm) and compute relative frequency.
* c. Approximate the area under the density curve between 150–170 cm. What does it represent?

**4. Skewness & Kurtosis Comparison**

**Problem:**  
Generate three datasets:

* A left-skewed dataset
* A right-skewed dataset
* A symmetric dataset (normal)

For each:

* a. Plot histogram and calculate skewness and kurtosis using scipy.stats
* b. Interpret the shape and tail behavior

**5. Chebyshev's Inequality on Unknown Distribution**

**Problem:**  
A dataset has mean income = ₹50,000 and standard deviation = ₹12,000.

* a. Without knowing the distribution, use Chebyshev’s theorem to estimate what percentage of individuals earn between ₹26,000 and ₹74,000.
* b. Compare this with the Empirical Rule (assuming normal distribution).

**6. Real-World Log Transformation**

**Problem:**  
Load a CSV file (or generate synthetic) containing:

* Population sizes of 1000 cities
* Income distribution of households

Tasks:

* a. Plot histograms for both features
* b. Apply log transformation and re-plot
* c. Explain how log helps in compressing skewed data
* d. Comment on interpretability after transformation

**7. SciPy Applications – Linear Algebra and Interpolation**

**Problem:**

* a. Create a 3x3 matrix and compute its determinant using scipy.linalg.det()
* b. Use scipy.interpolate.interp1d() to interpolate the following points and estimate y at x = 3.5:

x = [1, 2, 4, 5]

y = [1, 4, 2, 5]