**Lab 6: Data Distribution and Normality Testing**

**Prelab Questions**

1. What is a data distribution, and why is it important in data analysis?
2. Define normal distribution and list its key properties.
3. What is the purpose of normality testing in statistical analysis?
4. Name two statistical tests used to check for normality.
5. Why is normality important in hypothesis testing and machine learning models?

**In-Lab Details**

**Objective**:  
Analyze the distribution of a dataset and test for normality using visualization and statistical tests.

**Resources**:

* Python (Jupyter Notebook).
* Libraries: Pandas, Seaborn, Matplotlib, SciPy.
* Dataset: income\_data.csv with columns for age and annual income.

**Expected Output**:

1. **Histogram with KDE**: A plot showing whether the income data follows a normal distribution.
2. **Shapiro-Wilk and D’Agostino’s K² Tests**:
   * **p-value > 0.05**: Data is normally distributed.
   * **p-value ≤ 0.05**: Data is not normally distributed.

**Postlab Questions**

1. How do you interpret the shape of a histogram to determine if data is normally distributed?
2. What does a p-value > 0.05 in the Shapiro-Wilk test indicate?
3. What are some real-world scenarios where normality testing is essential?
4. How can you transform a non-normal dataset to approximate a normal distribution?
5. Why is a normal distribution preferred in many machine learning algorithms?