ANOVA Problem Statement

1.Salary Variance by Education Level

Problem Statement:
Analyze how salary varies across different education levels (Undergraduate, Graduate, Postgraduate).

Why this works (Variance focus):

- Compare spread of salaries within each education group
- Identify which education level has higher salary variability

Bivariate Variables:

- Education Level (Categorical)
- Salary (Numerical)

Common Visualisation:

- Box plot
- Variance table

2.Exam Score Variance vs Study Hours

Problem Statement:
Study how the variance in exam scores changes with the number of study hours.

Why this works:

- Shows whether consistent studying reduces score variability
- Easy to explain with real-life logic

Bivariate Variables:

- Study Hours (Numerical)
- Exam Score (Numerical)

Common Visualisation:

- Scatter plot
- Grouped variance by hour ranges

3.Product Price Variance Across Brands

Problem Statement:
Compare the price variation of the same product category across different brands.

Why this works:

- Some brands have stable pricing, others vary widely
- Clear business insight

Bivariate Variables:

- Brand Name (Categorical)
- Product Price (Numerical)

Common Visualization:

- Box plot
- Bar chart of variance

4.Variance in Customer Spending by Age Group

Problem Statement:
Analyze how customer spending variability differs across age groups.

Why this works:

- Shows which age group has predictable vs unpredictable spending
- Useful in marketing analytics

Bivariate Variables:

- Age Group (Categorical)
- Monthly Spending (Numerical)

Common Visualization:

- Box plot
- Variance comparison table

5.Attendance Variance vs Academic Performance

Problem Statement:
Examine whether attendance levels affect the variability of academic performance.

Why this works:

- Higher attendance may reduce performance variance
- Strong educational insight

Bivariate Variables:

- Attendance Percentage (Numerical)
- Exam Marks (Numerical)

Common Visualization:

- Scatter plot
- Variance by attendance bands

- **Low variance** → Data is consistent
- **High variance** → Data is spread out
- **Use box plots** for categorical vs numerical
- **Use scatter plots** for numerical vs numerical