Python Course Exercises: Accidents & Bikers Dataset

1) Load + Basic Inspection

- Load both datasets into pandas DataFrames.
- Print head(), info(), and describe() for relevant columns.
- Count missing values per column.
- Identify categorical vs numeric vs datetime columns.

2) Clean Types: Date + Time → Datetime Features

- Convert Date to datetime.
- Combine Date and Time into a DateTime column.
- Create Hour and Month columns.
- Print earliest and latest DateTime.

3) Groupby Statistics

- Compute average Number_of_Casualties and Number_of_Vehicles by Road_type.
- Repeat grouped by Light_conditions.
- Which Road_type has highest average casualties?

4) Frequency Tables

- Compute counts and percentages for Road_conditions.
- Repeat for Weather_conditions and Day.
- Create a summary table with count and percent.

5) Merge the Datasets

- Merge accidents and bikers on Accident_Index (inner join).
- Compare row counts before and after merge.
- Create Is_Serious column (Severity != 'Slight').

6) Plot: Severity Distribution

- Create a bar chart of Severity counts.
- Add title and axis labels.
- Optional: display count labels on bars.

7) Plot: Severity by Gender

- Create a crosstab of Gender × Severity.
- Plot as grouped or stacked bar chart.
- Interpret differences between genders.

8) Speed Limit Analysis

- Compare Speed_limit across Severity levels.
- Create a boxplot grouped by Severity.
- Compute median Speed_limit per severity level.
- Interpret findings.

9) Time-of-Day Pattern

- Use Hour feature to count accidents per hour.
- Create line plot or histogram (0–23).
- Optional: Plot Slight vs Serious as two lines.

10) Risk Scoring (Advanced)

- Create RiskScore based on casualties and risk factors (darkness, wet/snow, weather, speed >= 40).
- Compute average RiskScore by Road_type or Day.
- Plot top 5 categories by mean RiskScore.
- Write a short interpretation of results.