In this assignment, you will work with two real-world inspired case studies. Each case study provides a dataset and a set of tasks requiring statistical visualization, interpretation, and critical thinking. Your goal is to use programming (Python) to produce meaningful plots and summaries.

Case Study 1: Tech Industry Salaries & Skills

Background

The technology sector is rapidly growing, but salaries vary by job title, programming skills, and experience. Understanding these patterns helps both companies (for recruitment) and students (for career planning).

A survey of 500 tech professionals across the U.S. was conducted, collecting data on:

* Job Title
* Years of Experience
* Primary Programming Language
* Annual Salary

Tasks

1. Show the distribution of programming languages. Which language is most common?
2. Display the distribution of salaries. Is it symmetric, skewed, or multimodal?
3. Plot salary vs. years of experience. Describe the relationship (linear? plateauing?).
4. Compare salaries across programming languages for each experience group. Which language offers the highest returns at higher experience levels?
5. Summary – Write 2–3 paragraphs summarizing which factors (job title, experience, or language) seem most important in predicting salaries.

Case Study 2: Wearable Sensor Data & Athlete Performance

Background

Wearable devices (e.g., smartwatches) generate data on training, sleep, and health. Sports scientists analyze this data to improve performance and prevent injuries.

You are provided with a dataset of 100 athletes, with variables including:

* Sport
* Training Load (minutes/day)
* Resting Heart Rate (bpm)
* VO₂ Max (ml/kg/min)
* Sleep Hours
* Performance Score (0–100 scale)
* Injury Status (Yes/No)

Tasks

1. Show the distribution of athletes by sport. Which sport dominates the sample?
2. Plot the distribution of VO₂ Max. Does it look normal or skewed? What does it suggest about athlete fitness?
3. Plot training load vs. performance score. Do you see a point of diminishing returns?
4. Compare performance scores across sports. Which sport shows the widest variation?
5. Plot sleep hours vs. average performance score. Is performance strongly linked to sleep?
6. Show injury occurrence across different training load categories. What patterns do you notice?
7. Summary – Write 2–3 paragraphs explaining how training, sleep, and workload affect both performance and injuries.

Submission Instructions

* Submit your Python (.py) code along with generated plots.
* Answer all written interpretation questions in a short report (PDF or Word).
* Use clear titles, axis labels, and legends in all plots.