PART 1 — Visualization Challenge (5 Plots + Insights)

Goal: Students will create **5 different types of plots** using Matplotlib/Seaborn and write insights.

Dataset Options

Students may use:

- Titanic dataset
- Iris dataset
- Students performance dataset
- Sales data
- Mini Netflix dataset created previously

Task 1 — Create 5 unique visualizations

Each visualization **must** have:

- Title
- X/Y labels
- Legends (if needed)
- 1–2 lines of insights (written in markdown or comments)

Students must choose any 5 from below:

Plot 1: Histogram

Example: Distribution of ratings, sales, marks, or ages.

Plot 2: Bar Chart

Example: Total movies per genre, number of students per grade.

Plot 3: Scatter Plot

Example: Age vs Salary, Hours studied vs Marks.

Plot 4: Line Plot

Example: Monthly sales, number of customers per week.

Plot 5: Box Plot

Example: Compare distributions — salaries, ratings, marks.

Plot 6 (Optional Bonus): Heatmap

Example: Correlation matrix of dataset.

Example Insight:

> *"Sci-Fi genre has the highest number of movies, indicating audience preference. Drama is the second most common genre."*

PART 2 — Kaggle Intro + Dataset Exploration

Goal: Students will learn how to find datasets, download them, and explore basic info.

Task 2 — Kaggle Basics

Students must:

1. Create a Kaggle account
2. Search for **3 datasets** they find interesting
3. Download **1 dataset**
4. Write in a text cell:
    - Dataset name
    - Source link
    - Why they chose it
    - Columns available
    - Number of rows

Task 3 — Load & Explore Kaggle Dataset

Write code to:

```
import pandas as pd

df = pd.read_csv("your_dataset.csv")
df.head()
df.info()
df.describe()
df.isnull().sum()
df.nunique()
```

Students must perform:

- Shape (rows, cols)
- Column names
- Missing values check

- Basic statistics
- Unique values

Task 4 — Clean the Dataset

Students must:

- Handle missing values
- Remove duplicates
- Convert datatypes
- Fix string formatting

PART 3 — Mini Task: EDA Storytelling (Write 5 Insights)

Goal: Students build a small storytelling-based analysis report.

Task Instructions

Using the Kaggle dataset they downloaded:

Students must present:

5 meaningful insights, such as:

1. Trend or pattern
2. Comparison between two groups
3. Distribution understanding
4. Correlation between features
5. Any anomaly or outlier found

Examples of insights:

- “Sales were highest during November and December, indicating seasonal demand.”
- “Students who study more than 4 hours tend to score at least 20% higher.”
- “Comedy movies have the highest ratings on average.”
- “Most customers are between ages 20–35, showing a young target audience.”

Output Format

Students must submit a **notebook or PDF** with:

- Dataset description
- Visualizations (optional)
- 5 insights written in bullet format
- A final 4–5 line summary

BONUS CHALLENGES (Optional)

Challenge A: Pairplot

Create a pairplot using Seaborn for multi-variable understanding.

Challenge B: Custom Visualization

Use color themes, annotations, shapes, or style changes.

Challenge C: Compare Two Datasets

Example:

- Compare movie genres between Netflix & Amazon datasets.