Data Science TA Hiring Assignment – Cuvette

Objective

We're looking for Teaching Assistants who not only understand Data Science concepts but can **clearly explain** them to students. This assignment evaluates your technical knowledge, problem-solving skills, and communication ability.

Duration: ~3–4 hours

Submit:

1. Google Drive folder or GitHub repo with:
    - Jupyter notebook (`.ipynb`)
    - SQL query file
    - Tableau Public link
    - Excel file
    - `README.md` summarizing your steps
2. **Screen recording (10–15 mins)** where you:
    - Share your screen and explain your solution step-by-step
    - Walk through your code, visualizations, logic, and thought process
    - Upload it to Google Drive or YouTube (unlisted) and share the link

---

✅ Section 1: Python + Machine Learning (1.5 hours)

Dataset: Student Performance Dataset (Math & Portuguese)

Tasks:

1. Load the dataset and perform **data cleaning**.
2. Perform **EDA** using visualizations (use `seaborn`, `matplotlib`).

3. Predict whether a student will **pass or fail** based on scores and other features.
4. Use **Logistic Regression** or **Random Forest** for classification.
5. Evaluate the model using **accuracy, confusion matrix, F1-score**.
6. Add comments/markdown to explain the steps.

✅ Section 2: SQL (30–45 mins)

Dataset: Chinook Database (Music Store)

You can run it on SQLite Online or import into any SQL IDE.

Tasks:

1. List the top 5 customers by total purchase amount.
2. Find the most popular genre in terms of total tracks sold.
3. Retrieve all employees who are managers along with their subordinates.
4. For each artist, find their most sold album.
5. Write a query to get monthly sales trends in the year 2013.

Submit a `.sql` file with queries and output screenshots.

✅ Section 3: Tableau (30 mins)

Dataset: Airbnb Listings in NYC

Tasks:

1. Create a dashboard showing:
    - Listings count by neighborhood
    - Price distribution per room type
    - Availability trends

2. Include filters for room type and neighborhood.
3. Publish your dashboard to Tableau Public and share the link.

✅ Section 4: Excel (30 mins)

Dataset: Online Retail Dataset (UCI)

Tasks:

1. Clean the data (remove nulls, handle duplicates).
2. Create a pivot table to show **total sales by country and month**.
3. Add formulas to calculate:
    - Average order value
    - % contribution of each country to total sales
4. Highlight top 5 countries by revenue using conditional formatting.
5. Create a chart to visualize monthly revenue trend.

✅ Section 5: Bonus (Optional, +15 mins)

Write ~150 words on:

- How you'd support students struggling with concepts or deadlines.
- How you would break down a complex topic like "Gradient Descent" for beginners.

✅ Section 6: AI Tools & LLMs (30 mins)

Objective:

We want to see how comfortable you are with the rapidly evolving AI landscape, especially tools powered by LLMs like ChatGPT or Claude.

Tasks (Pick any 1 of the 2 options below):

🧠 **Option A: Prompt Engineering**
Use **ChatGPT or any LLM** to assist with one of your earlier tasks (e.g., EDA, SQL query building, Excel formula generation).

1. Share the **exact prompt(s)** you used.
2. Copy the **response** from the AI.
3. Reflect briefly (3–4 lines):
    - What did the AI do well?
    - What did you have to modify or improve?

🤖 **Option B: Build with an AI Tool**
Use any AI-based tool (e.g., Notion AI, Excel Copilot, Tableau GPT, or ChatGPT Code Interpreter) to solve a small part of your assignment.

1. Mention the tool and what you used it for.
2. Upload a screenshot or output file.
3. Briefly explain how this tool could help students in the course.

Would you like me to update the Google Doc and PDF with this section included and re-share them?

📹 Video Recording Instructions (Mandatory)

Record a 10–15 minute video where you:

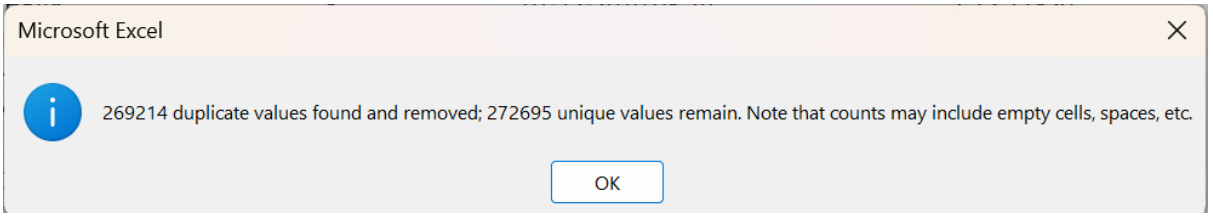
- Share your screen and explain each part of your solution
- Talk through your approach, challenges, and learnings
- Ensure your audio is clear and your screen is visible
- **Keep your camera ON** so we can see you while you explain

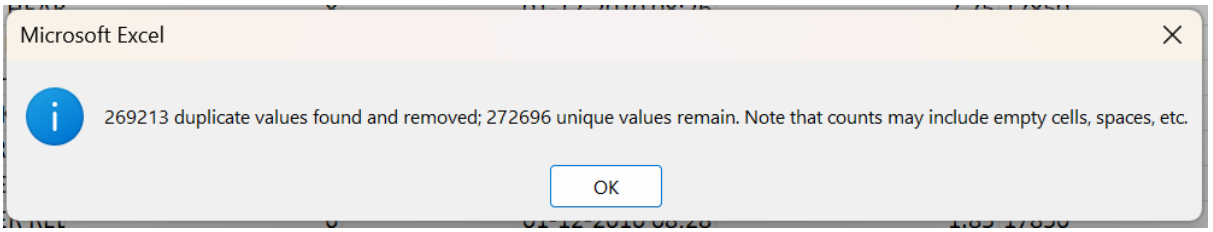
Upload the video to:

- Google Drive (with view access) or
- YouTube (set to **unlisted**)
  Share the link along with your submission.

Best of Luck!

Section 3
Tableau Link -
https://public.tableau.com/app/profile/anil.kumbhar/viz/Cuvette_Tech_TAInternship_Tablue_Anil/Dashboard1





Section 5: Bonus

Supporting Students Struggling with Concepts or Deadlines:

To support students struggling with concepts or deadlines, I would first ensure they feel comfortable asking for help without fear of judgment. I'd offer clear explanations, break down complex ideas into simpler steps, and use real-life examples to make the concepts relatable. I'd also recommend additional resources like tutorials or practice exercises. If deadlines are a concern, I'd help them create a manageable schedule, allowing extra time for tough tasks, and encourage them to tackle smaller portions of their work regularly instead of leaving it all until the last minute.

Breaking Down "Gradient Descent" for Beginners:

Imagine you are standing at the top of a hill, and your goal is to find the lowest point in the valley. The best way to reach the bottom is by taking small steps downhill. In the same way, **Gradient Descent** helps find the lowest point of a function (minimum) by adjusting the model's parameters step by step. You calculate the slope (gradient) at your current position, then move a little downhill (adjust the parameters) until you reach the bottom. By repeating this process, you get closer and closer to the best solution.

Section 6: AI Tools & LLMs

Option A: Prompt Engineering

Task: I used ChatGPT to help me write an SQL query.

Prompt used:
"Write an SQL query to find the average reading score and math score for each gender from the 'StudentsPerformance' dataset. Group the data by gender and order it by the average math score in descending order."

Response from AI:

sql
CopyEdit

```sql
SELECT gender,
       AVG(math_score) AS avg_math_score,
       AVG(reading_score) AS avg_reading_score
FROM StudentsPerformance
GROUP BY gender
ORDER BY avg_math_score DESC;
```

Reflection:

- **What did the AI do well?**
    - The AI correctly wrote the SQL query, using the appropriate `AVG` function, `GROUP BY` to group by gender, and `ORDER BY` to sort the results. The syntax and logic were all correct.
- **What did you have to modify or improve?**
    - The response was accurate, and no modifications were needed. If the query had been more complex, I might have asked for optimizations or improvements on performance.

How this helps students:
This tool can significantly assist students by quickly generating the right SQL syntax,

reducing the time spent on writing queries. It’s particularly useful for beginners, as it helps them understand the structure of SQL commands and provides immediate feedback.