PROBLEM STATEMENTs

Problem Statement 1 (Technical)

The Scale Mirage

A dataset has been recovered from a distorted signal stream.
At first glance, the points appear chaotic, scattered, stretched, and deceptive.
But this disorder is intentional.

Hidden within the data lies a **visual codeword**, masked by a phenomenon known only as the **Scale Mirage**, where a few misleading points dominate perception and drown out the real structure.

Your mission is to see through the mirage.

Objective

Uncover the hidden visual codeword embedded in the dataset by carefully restoring balance to the coordinate space.

Dataset

🔗 https://drive.google.com/file/d/160rlxVw10mhX577Hb0XhfCoVJ-rtlxL7/view?usp=sharing

---

Instructions

1. Begin by visualizing the dataset as it is.
   Plot 1: Raw data (x vs y)
2. Devise a strategy to eliminate deceptive points that distort the true structure of the data.
3. Once the dataset is cleansed, bring it into a comparable coordinate frame using appropriate normalization.
4. Visualize the restored structure.
   Plot 2: Filtered and normalized data (x vs y)
5. Observe carefully.
   When the noise fades, the signal speaks.

Deliverables

- Python script containing:
    - Code for plotting the raw data
    - Code for filtering and normalizing the dataset
    - Final plot revealing the hidden pattern

Final Question

What is the codeword revealed once the illusion collapses?

Problem Statement 2 (Technical)

The Broken Timeline

A system log records events along with their timestamps.
 Due to a synchronisation glitch, the log is **partially corrupted**.

Some events:

- Appear out of chronological order
- Are duplicated
- Contain timestamps that are **technically valid**, but **logically inconsistent** when viewed as a sequence

The data is not completely broken, but it cannot be trusted as-is.

Objective

Restore a logically consistent event timeline and extract basic insights from it.

Dataset

https://drive.google.com/file/d/1uEey4__c4pw24weXuQpNIbF8xHB-Z0OJ/view?usp=sharing

Instructions

1. Load and inspect the dataset.

2. Identify and remove duplicate events.

3. Analyse the timestamp sequence and identify inconsistencies.

4. Apply a reasonable strategy to restore logical order
   (simple sorting may not be sufficient on its own).

5. Create two visualizations:

   - **Plot 1:** Event index vs timestamp (raw data)

   - **Plot 2:** Event index vs timestamp (cleaned timeline)

6. Compute:

   - Average time gap between consecutive events

   - Longest time gap between events

Deliverables

- Python script with clear, well-commented logic

- Two plots (before and after cleaning)

- Printed outputs for the calculated metrics

Final Question

What inconsistencies existed in the original timeline, and how did your cleaning strategy resolve them?

Problem Statement 3 (Non-Technical)

Databiz Society Case: The Meesho Unit Economics Paradox

Background

As part of a Databiz Society business analysis challenge, participants are examining disruptive business models in the Indian e-commerce ecosystem.

In India, the cost of shipping a parcel (last-mile delivery, handling, and returns) typically ranges between **₹60 and ₹90**.
For platforms like Amazon or Flipkart, a **₹1,500+ order value** absorbs this cost easily.

Meesho, however, operates with:

- An **average order value of ~₹270**
- A **0% commission model**, meaning the platform takes no cut from sellers

The Conflict

If Meesho were to rely on traditional logistics costing ~₹70 per delivery:

- Shipping alone would consume **~25% of the order value**
- With zero commission, the platform appears to lose money on every order

Yet, Meesho continues to scale rapidly.

Core Question

How does a platform like Meesho operate sustainably at scale despite these apparent unit-economics constraints?

Instructions

Analyze the situation and explain:

- Where the surface-level math is misleading
- What hidden mechanisms or strategic decisions enable sustainability
- How scale and structure change the economics

Deliverable

- A concise written analysis (maximum 1 page)

Final Question

What fundamentally enables Meesho's business model to work when it appears mathematically impossible?

Problem Statement 4 (Non-Technical)

Databiz Society Case: Growth vs Trust

Background

As part of a Databiz Society strategic thinking challenge, participants are asked to evaluate a fast-growing ed-tech platform offering free digital certificates for online workshops.

To scale quickly, the platform introduces **instant, bulk auto-issued certificates** with minimal verification.

The impact is immediate:

- Certificate issuance triples
- User registrations surge
- Growth metrics look excellent

However, within months, employers and institutions begin questioning the **credibility** of these certificates.

The Dilemma

Tightening verification will:

- Increase operational costs
- Slow down issuance
- Hurt short-term growth

Keeping verification loose will:

- Preserve rapid growth
- Risk long-term trust and brand value

Objective

Evaluate the trade-off between rapid growth and long-term credibility.

Instructions

As a decision-maker:

- Identify risks on both sides
- Propose a strategy to balance growth and trust
- Justify your approach using clear business reasoning

Deliverable

- A structured written response (maximum 1 page)

Final Question

How would you design a system that scales growth without destroying trust?