# ASSIGNMENT 2 : LINEAR REGRESSION

Link for dataset: Assignment 2 dataset
Name:_____

## Part 1: The Datasets

You will be provided with two synthetic datasets:

1. Plot the two datasets A and B separately and note down their shapes.

---

## Part 2: Tasks

### 1. Linear Modeling

- Using scipy.optimize.minimize, define a linear objective function (y = mx + c) and fit it to both the datasets.
- **Challenge:** Even though the data isn't a line, the optimizer will still find the "best" straight line. Your task is to find those optimal parameters (m and c).

### 2. Visualization

- For each dataset, plot the original data points as a scatter plot.
- Overlay the "Best Fit Line" found by your linear regression.

### 3. Statistical Analysis & Residuals

For both datasets, calculate and report:

- **The Point Balance:** How many points are above the line vs. below the line?
- **Mean Distance:** The average vertical distance of points on either side of your line.

---

## Part 3: Discussion Questions

1. **Underfitting:** Do you think the fitting is good enough, what do you think is a better way to fit things for these?