Regression Analysis of Factors Affecting Company Average Package in India

1. Project Objective

This project aims to build a multiple regression model that identifies key factors influencing the average salary package offered by Indian companies. Using a dataset of over 10,000 companies, we will quantify how variables such as industry sector, company size, employee count, and company ratings explain variation in compensation across firms. The goal is to produce a well specified, interpretable model supported by diagnostics and validation.

2. Why This Topic is Interesting

Compensation patterns vary widely across industries, geographies, and organizational structures. Understanding the structural determinants of salary helps job seekers, HR teams, and policymakers identify key market drivers. Regression analysis provides a statistical framework to measure and compare these effects.

3. Data and Variables

We plan to use the Indian Companies Complete Data 2025 (10,000+ companies) dataset. This dataset includes detailed company attributes that can be used to model average salary.

Response Variable: Average Salary
Potential Predictors:

- Locations
- Total Employee Count
- Industry / Sector (categorical)
- Company Type
- Founding Year
- Rating

These variables allow a combination of numeric and categorical predictors, ideal for multiple linear regression.

4. Planned Methodology

We plan to use the following statistical techniques:

1. Exploratory Data Analysis (distribution checks, outlier detection, correlation analysis)
2. Model Building using Multiple Linear Regression
3. Diagnostic Tests (residual analysis, heteroscedasticity, influential points)
4. Model Validation (train–test split or cross-validation, performance metrics)

Group Members:

- Anmol Sai Ramchandran Ramanan
- Devang Chaturvedi
- Kanishk Deshwal
- Shreyas Pantangi