# 12\_2: **Workplace Diversity Analysis**

## **Goal**

Diversity, unconscious bias in the workplace, and overall treatment of employees are crucial issues. Data science can play a key role in identifying potential discrimination by analyzing data to determine if certain groups of employees are being treated unfairly.

## **Challenge Description**

There has been considerable discussion about workplace diversity, particularly in the tech industry. The Head of HR at your company is very concerned about this issue and has tasked you with analyzing internal employee data to determine whether the company treats all employees fairly. Specifically, you have been assigned the following tasks:

1. **Identify Employee Levels:** Classify each employee into one of six levels based on their role:
   * Individual Contributors (IC): Employees who do not manage anyone.
   * Middle Managers (MM): Direct supervisors of ICs.
   * Directors (D): Direct supervisors of MMs.
   * Vice Presidents (VP): Direct supervisors of Directors.
   * Executives (E): Direct supervisors of VPs.
   * CEO: Direct supervisor of Executives.
2. **Determine the Number of People Managed by Each Employee:** Calculate the total number of people managed by each employee, including indirect reports. For instance, if John directly manages 2 people, and those 2 people manage 5 people each, John effectively manages 12 people.
3. **Predict Employee Salaries:** Develop a model to predict the salary of each employee.
4. **Analyze Salary Factors and Fairness:** Identify the main factors influencing employee salaries. Assess whether the company is treating all employees fairly based on these factors. Finally, provide recommendations to the Head of HR on the next steps.

**Data**

We have two downloadable tables:

**Table 1: company\_hierarchy**This table provides information about each employee's direct boss and their department. The columns are:

* **employee\_id**: The unique ID of the employee.
* **boss\_id**: The ID of the employee's direct boss.
* **dept**: The department of the employee. The possible departments are:
  + Engineering (data science is under engineering)
  + Marketing
  + Sales
  + HR
  + CEO (for the CEO, who belongs to all departments)

**Table 2: employee**This table provides additional details about each employee. The columns are:

* **employee\_id**: The unique ID of the employee.
* **signing\_bonus**: Indicates if the employee received a signing bonus when they joined the company (1 for yes, 0 for no).
* **salary**: The current salary of the employee in USD.
* **degree\_level**: The highest degree obtained by the employee.
* **sex**: The gender of the employee (Male/Female).
* **yrs\_experience**: The number of years of work experience the employee has.

These tables can be used to perform the following tasks:

1. Classify each employee into one of the six levels: IC, MM, D, VP, E, CEO.
2. Calculate the total number of people managed by each employee.
3. Build a model to predict employee salaries.
4. Analyze the main factors impacting employee salaries and determine if the company treats all employees fairly. Provide recommendations to the Head of HR based on the findings.