Diversity in the Workplace

(Referenced from http://datascientistjobinterview.com/)

Goal

Diversity, unconscious bias in the workplace, and the way companies treat their employees are a very important topic. Data science can help discover potential discrimination by looking at the data and seeing what is says.

Challenge Description

There has been lots of talking about diversity in the workplace, especially in technology. The Head of HR at your company is very concerned about that and has asked you to analyze internal data about employees and see whether results suggest that the company is treating its employees fairly.

Specifically, she gave you the following tasks:

1. In the company there are 6 levels (described below). Identify, for each employee, their corresponding level:
   - Individual Contributors (IC) - they don't manage anyone
   - Middle Managers (MM) - they are the direct bosses of IC
   - Directors (D) - they are the direct bosses of MM
   - Vice Presidents (VP) - D direct bosses
   - Executives (E) - VP direct bosses
   - Chief Executive Officer (CEO) - The direct boss of E

2. How many people does each employee manage? You should have a table with employee ids and the number of people managed. Consider that if John directly manages 2 people and these two people manage 5 people each, then we conclude that John manages 12 people.

3. Build a model to predict the salary of each employee.

4. Describe the main factors impacting employee salaries. Do you think the company has been treating its employees fairly? Do you see any bias? What are the next steps you would suggest to the Head of HR?

Data

We have 2 tables that will be provided.

The first table is:
"company_hierarchy_mod" - info about each employee, their direct boss, and their department

Columns:
- employee_id : the id of the employee. It is unique by employee and can be joined to employee id in the other table
- boss id : the id of the boss of employee id. It is unique by employee and can also be joined to employee id in the other table
- dept : employee department:
    - Engineering (data science is under engineering)
    - Marketing
    - Sales
    - HR
    - CEO

The second table is:
"employee_mod" - info about each employee

Columns:
- employee_id : the id of the employee. It is unique by employee and can be joined to employee id and/or boss id in the other table
- signing_bonus : whether the employee got a signing bonus when she joined the company (1 -> yes, 0 -> no)
- salary : the current salary of that employee in USD
- degree level: the highest degree received by the employee.
- sex: Male/Female
- yrs_experience: employee years of work experience

Example

Let's check a random employee, say employee 51535:

Company_hierarchy_mod, employee_id == 51535

Column Name	Value	Description
employee_id	51535	Id of the employee
boss_id	61554	Id of the employee's boss
dept	engineering	Employee 51535 works in engineering

Employee_mod, employee_id == 51535

Column Name	Value	Description
employee_id	51535	Id of the employee
signing_bonus	1	He got a signing bonus when he started
salary	650000	Yearly salary
degree_level	PhD	Education level
sex	M	male
yrs_experience	33	Has worked for 33 years