

Ethics Guide

Querying Inequality?



MaryAnn Baker works as a data analyst in human relations at a large, multinational corporation. As part of its compensation program, her company defines job categories and assigns salary ranges to each category. For example, the category M1 is used for first-line managers and is assigned the salary range of \$75,000 to \$95,000. Every job description is assigned to one of these categories, depending on the knowledge and skills required to do that job. Thus, the job titles Manager of Customer Support, Manager of Technical Writing, and Manager of Product Quality Assurance are all judged to involve about the same level of expertise and are all assigned to category M1.

One of MaryAnn's tasks is to analyze company salary data and determine how well actual salaries conform to established ranges. When discrepancies are noted, human relations managers meet to determine whether the discrepancy indicates a need to:

- Adjust the category's salary range;
- Move the job title to a different category;
- Define a new category; or
- Train the manager of the employee with the discrepancy on the use of salary ranges in setting employee compensation.

MaryAnn is an expert in creating database queries. Initially she used Microsoft Access to produce reports, but much of the salary data she needs resides in the organization's Oracle database. At first she would ask the IS Department to extract certain data and move it into Access, but over time she learned that it was faster to ask IS to move all employee data from the operational Oracle database into another Oracle database created just for HR data analysis. Although Oracle provides a graphical query interface like that in Access, she found it easier to compose complex queries directly in SQL, so she learned it and, within a few months, was a SQL expert.

"I never thought I'd be doing this," she said. "But it turns out to be quite fun, like solving a puzzle, and apparently I'm good at it."

One day, after a break, MaryAnn signed into her computer and happened to glance at the results of a query that she'd left running while she was gone. "That's odd," she thought. "All the people with Hispanic surnames have lower salaries than the others." She wasn't looking for that pattern; it just happened to jump out at her as she glanced at the screen.

As she examined the data, she began to wonder if she was seeing a coincidence or if there was a discriminatory pattern within the organization. Unfortunately for MaryAnn's purposes, the organization did not track employee race in its database, so she had no easy way of identifying employees of Hispanic heritage other than reading through the list of surnames. But, as a skilled problem solver, that didn't stop MaryAnn. She realized that many employees having Hispanic origins were born in certain cities in Texas, New Mexico, Arizona, and California. Of course, this wasn't true for all employees; many non-Hispanic employees were born in those cities, too, and many Hispanic employees were born in other cities. This data was still useful, however, because MaryAnn's sample queries revealed that the proportion of employees with Hispanic surnames who were also born in those cities was very high. "OK," she thought, "I'll use those cities as a rough surrogate."