EMPLOYEE DATA ANALYSIS----DATA VISUALIZATION

Data Visualization using Python libraries such as pandas and matplotlib.

The Dataset

The dataset is an employee dataset. The dataset consists of 2,000 Unique IDs with 10 variables on each IDs. Some of the variables include;

Unique ID: this is the unique identification number that has been assigned to each of the employees which can not be used by anyone else except them.

Position/Title: this is the job position of each employees and the title attached to each position.

Department: this consists of various departments in the organization.

Base Salary: this is the amount each employee is paid at the end of the month.

Hire date: this is the date an employee is being hired.

Job date: this is the date an employee resumes work

These are few of the variables present in the data. I will be performing data analysis on this dataset to check the compositions and disparities between the variables.

Summary of the Findings.

After the analysis, below are the conclusions of my observation;

1. The highest number employees are Black or African American racial group.
2. There are significant disparities in the representation of different racial/ethnic groups across departments.
3. There are more male than female employees in the organization.
4. There are gender imbalances in all departments.
5. There are more full-time employee compare to other employment type such as part time, temporary etc. and across departments.
6. There are significant differences between base salary and race.
7. There are significant differences between base salary and gender.
8. There's evidence of a racial pay gap.

9. The department with the highest salary is the legal department.
10. The position with the highest salary is the city attorney.

Key Insights for Presentation.

The key insights of my presentation after my data analysis are as follows

1. Racial composition.
2. Gender composition.
3. Racial and Gender disparities across departments and positions.
4. Employee types composition.
5. The significant differences between base salary and race and between base salary and gender.

Limitations.

While data analysis is a good way to know the data using interesting and vivid visualizations, in order to be able to make a final statement about the relationship amongst the variables, one would need to conduct statistical tests and build predictive models.

Resources.

> Project from Cortouch media