Executive Summary

Brief

 # Analysis: to determine if performance ratings and promotions are equitably distributed by gender, analysis was conducted using multiple cross-tabulations, tables, charts, machine learning models for anomaly detection and prediction.

 # Results: some gaps were discovered at the highest level of performance. Differences by group sizes and gender distribution within different groups can be compared.

 # Recommendations: A disparity in promotion rates has been observed, with men being promoted less frequently than women at the highest performance levels, and client's business impact is required.

Hypothesis / problem statement

There is evidence of groups of employees where performance ratings and promotions are not equitably distributed by gender.

Data description

- Employee Roster: employee details, hire date, job level, job function, region, gender, and age.
- Mid-Year Outcomes: Includes performance ratings and promotion decisions.
- Employees (10k) and Performance (8k) were merged, resulting in 8k records master dataset.
- Mostly categorical data, which is aggregated between groups.

Assumptions about the data

- It is assumed that the provided values are correct, and that data source validation is not required.
- It is assumed that the current sample represents the entire population.
- It is assumed that the existence of any inequalities by gender within data is unknown.

Methods

Manual investigation of data has been done first to understand data, followed by automated flow in Python for future analysis. The following methods have been used to conduct the analysis:

- Transformations of data for analysis, such as recode variables.
- Calculation of cross-tabulations, and averages.
- Visualization of the patterns using charts and dashboard.
- Machine learning model to measure predictability.
- Anomaly detection.

Analysis

- The performance of 8000 employees has been analyzed focusing on possible inequalities by promotion rate and gender.
- The dataset contains multiple variables and categories, resulting in multiple statistical outputs to review.

Results

- Analysis identified the differences by gender by job function. For example, in Engineers, there are around 90% of men.
- There are fewer promoted men at highest performance rank (5-Redefines Expectations).
- Aggregated data helps compare groups using group averages

Recommendations

- Review and standardize performance evaluation criteria.
- Establish a baseline for promotion rate.
- Instead of forcing the outcome by gender, use a randomized approach.
- Smaller groups tend to have larger differences against the overall group statistics and therefore must be reviewed with more careful consideration.
- Include additional data upon review with the client to strengthen the analysis.
- Move analysis to SQL.

Appendix: Source Code

Please attach the Python/R code, as well as any ancillary files or tools you used to process the data and conduct your analysis. Please add comments to your code so that it is easy to follow along.

Fig.1. Comparison of US1 and US4 distributions of (%employees, %promoted, and %masculine), Job Function, and Type (I, T)

region	job_function	% employees I	% employees T	Promoted, % I	Promoted, % T	Masculine, % I	Masculine, % T
US1	CXI	8.9%	3.6% ▲	4.8% ◆	3.5% ▼	19.3% ▼	21.8%
	Engineering	14.1%	8.6% ▲	4.8% ▲	4.4% ▲	79.5% ▲	88.2%
	G&A	6.5%	3.1% ▲	5.4% ◆	3.3% ▼	18.6% ▼	22.8%
	Product	6.5%	3.5% ◆	3.1% ▲	4.3% ▬	65.3% ▲	76.1%
	S&O	9.3%	4.9% ▲	4.1% ▲	4.1% ▬	43.6% ▬	58.0%
	Sales	7.4%	4.1% ▲	4.7% ▲	4.9% ▬	51.9% ▬	61.1%
US1 Total		52.8%	27.7% ▲	4.5% ▲	4.2% ▬	49.9% ▬	61.4%
US4	CXI	2.0%	1.0% ◆	2.6% ◆	2.5% ▼	17.9% ▼	22.5%
	Engineering	3.3%	2.2% ◆	1.5% ◆	2.3% ▲	80.9% ▲	88.6%
	G&A	1.5%	0.8% ●	6.9% ▲	6.1% ▼	31.0% ▼	27.3%
	Product	1.2%	0.8% ◆	2.0% ◆	3.2% ▬	61.2% ▲	71.0%
	S&O	2.3%	1.1% ▲	4.4% ●	8.9% ▬	45.6% ▬	55.6%
	Sales	2.2%	1.1% ◆	3.4% ◆	2.2% ▬	54.0% ▼	40.0%
US4 Total		12.4%	7.1% ◆	3.2% ◆	3.9% ▬	51.9% ▬	57.1%
Grand Total		65.2%	34.8% ▲	4.3% ▲	4.1% ▬	50.3% ▬	60.6%