Executive Summary

Brief

 # Analysis: Initial analysis performed in Excel: manual review, data transformations, derived variables, cross-tabulations, tables, charts. Analysis is automated in Python, adding machine learning models for anomaly detection and prediction.

 # Results: A disparity in promotion rates has been observed for specific categories, for example, men are promoted less frequently than women at the highest performance levels.

 # Recommendations: Review and standardize performance evaluation criteria. Establish a baseline for promotion rate and deviations between specific category and group categories.

Hypothesis / problem statement

The main challenge is identifying if and where gender disparities exist.

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.

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 in advance.

Methods

The following methods have been used to conduct the analysis:

- Transformations: recode variables.
- Calculations: counts, averages, ratios, and cross-tabulations.
- Visualization: a dashboard in Excel and static charts in Python.
- Advanced: anomaly detection.
- Advanced: predictive model.
- Advanced: clustering.

Analysis

- Prepared and screened data contains around 3800 records.
- Dataset has been analyzed focusing on possible inequalities by promotion rate and gender.
- Dataset contains multiple variables and categories, resulting in multiple statistical outputs to review.

Results

- Analysis identified the differences in gender distribution by job function. For example, there are ~90% of men in Engineers.
- The differences in the promotions and gender are not significant overall, however some differences appear when looking deeper.
- Group size impacts the probability: smaller group, less promotions.
- There are fewer promoted men at top performance rank (5).
- The developed analytical frame is suitable for a subsequent in-depth analysis using Excel dashboard to review smart charts, and Python for statistical outputs and advanced analysis..

Recommendations

- Review and standardize performance evaluation criteria.
- Establish a baseline for promotion rate and allowed deviations.
- Smaller groups tend to have larger differences from 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 data collection, merging, and processing, to SQL.
- Use advanced analytics to validate findings.

Appendix: Source Code

Data analysis assets are available in repository https://github.com/aleurb/empl_perf/.

Appendix: Dashboard

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

		% employees		Promoted, %		Masculine, %		
region	job_function	I	T	I	T	I	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%**	