Project Dictionary: Pay Equity Analyzer

Scope: 120-500 employees

Architecture: Hierarchical (N1-N8)

Methodology: Multiple Linear Regression (Log-linear)

1. Job Leveling Architecture (N1 - N8)

The **Market Midpoint** represents the standard competitive salary for a competent performer. **Job Level** is the primary proxy for the "value of work" in the regression model.

Level	Executive Category	Market Midpoint (Annual USD)	Salary Range (80% - 120%)
N1	Entry / Support	$20,000	$16,000 - $24,000
N2	Junior Analyst	$32,000	$25,600 - $38,400
N3	Senior Analyst	$50,000	$40,000 - $60,000
N4	Supervisor / Lead	$75,000	$60,000 - $90,000
N5	Manager	$110,000	$88,000 - $132,000
N6	Director	$165,000	$132,000 - $198,000
N7	VP / Sr. Director	$235,000	$188,000 - $282,000
N8	C-Suite (CEO/CFO)	$350,000	$280,000 - $420,000

2. Job Family Matrix (Career Paths)

Roles are segmented by **Job Family** to account for market premiums specific to each function (e.g., IT vs. HR).

Family	Entry Roles (N1-N2)	Professional Roles (N3-N5)	Leadership Roles (N6-N8)
IT	Help Desk (N1), Jr Dev (N2)	Sr Dev (N3), IT Mgr (N5)	**CTO (N8)**

Commercial	Lead Gen (N1), Sales Rep (N2)	Account Mgr (N3), Sales Mgr (N5)	**VP Sales (N7), CRO (N8)**
Operations	Warehouse (N1), Logistics (N2)	Plant Sup (N4), Ops Mgr (N5)	**COO (N8)**
Finance	Clerk (N1), Accountant (N2)	Controller (N4), Finance Mgr (N5)	**CFO (N8)**
HR	Assistant (N1), Recruiter (N2)	HRBP (N4), HR Manager (N5)	**CHRO (N8)**
Legal	Legal Asst (N1), Paralegal (N2)	Sr Counsel (N3), Legal Mgr (N5)	**General Counsel (N7)**

3. Dataset Specification (CSV Schema)

The file `data.csv` must follow this schema to ensure compatibility with the `project.py` analysis engine.

- `id` : Unique employee identifier (e.g., "EMP001").

- `name` : Full name of the employee.

- `gender` : Protected variable for gap analysis (Male/Female).

- `job_family` : Functional area to control for market differences.

- `job_title` : Standardized taxonomy title.

- `job_level` : Hierarchical grade (N1-N8). Primary predictor.

- `education` : Legitimate factor for pay differentiation (High School, Bachelor, Master, PhD).

- `tenure_months` : Total months in the company/role.

- `annual_salary` : Dependent variable (y) for regression.

- `performance_rating` : Annual score (1.0 - 5.0) to justify merit-based differences.

4. Governance & Metrics

Compa-Ratio Formula

$$\text{Compa-Ratio} = \left(\frac{\text{Annual Salary}}{\text{Market Midpoint}} \right) \times 100$$

Risk Zones (Flags)

- **Underpaid (< 80%)**: Critical flight risk / Potential inequity.

- **Equitable (80% - 120%)**: Green zone.

- **Overpaid (> 120%)**: Budget anomaly / "Red circle" rate.

Outlier Detection Strategy

The program uses the **Interquartile Range (IQR)** method ($Q3 + 1.5 \times IQR$) to identify anomalies statistically, avoiding skew from high C-Suite salaries.

5. Analytics Methodology

The core analysis utilizes **Multiple Linear Regression (MLR)** with a log-transformation of the salary ($ln(\text{Salary})$).

This approach isolates the impact of the `gender` variable while holding "legitimate factors" constant (job level, job family, tenure, and performance) to calculate the **Adjusted Pay Gap** rather than just the raw average difference.