HR Salary Analysis

Business Briefing

From role + education + experience → transparent salary expectations

Compensation planning

Pay equity review

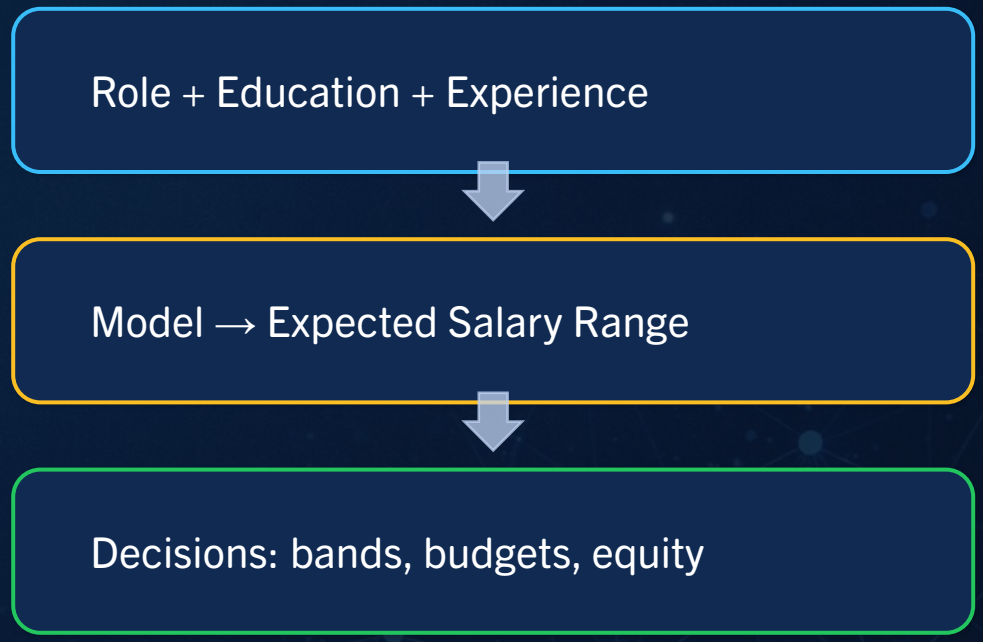
Workforce strategy

Purpose & context

What leadership gets

- A concise view of the salary prediction initiative
- How the analysis supports compensation planning
- Signals to inform workforce strategy (hiring, progression, retention)

High-level flow



Objectives

Transparent expectations

- Clear, data-backed salary expectations
- Consistent logic across roles & levels
- Explainable inputs: role, education, experience

Equity + retention leverage

- Spot gaps between predicted vs actual pay
- Identify high-variance roles for review
- Support fair offers & career progression

Actionable guidance

- Deliver business-ready takeaways
- Complement the technical notebook
- Enable piloting → scale-out with governance

Data & approach (what the notebook operationalizes)

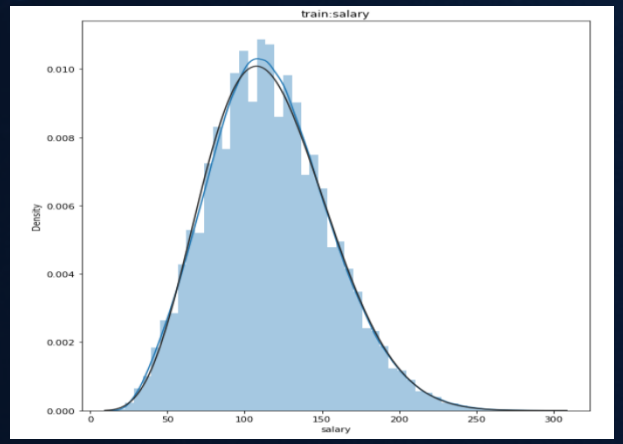
Labeled rows (salary known)

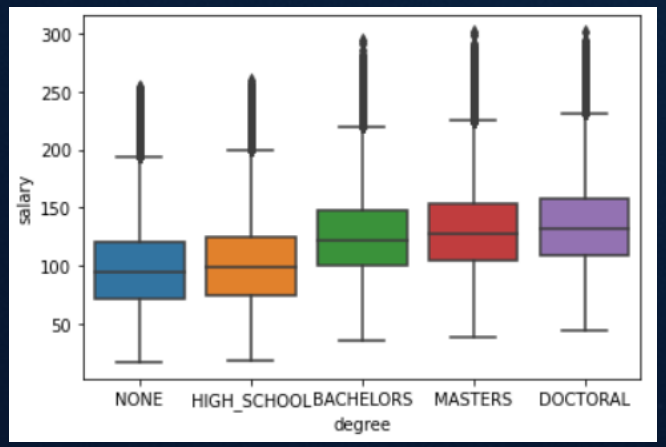
1.0M

Median Salary

120k $

Dataset at a glance





How the notebook works

- Split labeled data into training + validation
- QA checks: nulls, outliers, sparse categories
- Feature engineering via a utility class to keep train/validation aligned
- Grouped statistics (mean/median/variance) computed on training only
- One-hot encoding + scaling with safe reindexing for new categories

Caveat

The repository notes the data appears machine-generated and noisy — so we optimize for stability and guardrails over "perfect" accuracy.

Benefits for HR & leadership

Pay equity reviews, faster

- Compare predicted vs actual pay
- Flag potential inequities early
- Prioritize deeper review where gaps cluster

Career ladders with evidence

- Surface role/education/experience patterns
- Support hiring & promotion ladders
- Make progression logic auditable

Benchmark smarter

- Identify roles with high variance
- Target market pricing where it matters
- Reduce blanket benchmarking spend

Budgeting before requisitions open

- Quantify expected salary ranges
- Improve workforce cost forecasting
- Reduce surprise offers & renegotiations

Robust methods & safeguards

Modeling discipline (to avoid false confidence)

1. Train/validation split to prevent leakage
2. Same feature engineering on both splits
3. Null/outlier/sparsity checks before fitting
4. Grouped stats computed on training only
5. Validation performance tracked to detect overfit

Safeguards we enforce

- No train/test column drift
- Train-only grouped statistics
- Category sparsity checks
- Repeatable preprocessing pipeline
- Holdout validation monitoring

Why leadership should care

Guardrails reduce the risk of “model theater” — pretty charts that fail the moment inputs shift or categories evolve.

Generalization of results

Designed to tolerate new or rare categories

- Encodings are fit on training categories and then applied to new data with a fixed column set.
- Unseen categories are safely ignored (no column shifting).
- Validation on holdout data helps estimate stability before deployment.
- Best generalization happens when incoming data follows similar collection standards.

Operational guidance

- Set a refresh cadence (e.g., quarterly) as roles evolve
- Monitor drift: new titles, new education programs, new markets
- Review fairness metrics by subgroup before widening scope
- Keep a human-in-the-loop policy for exceptions

Recommended next steps

A pragmatic rollout: pilot → scale, with governance baked in

Phase 1
Define tolerances

- Agree acceptable error for bands
- Define salary band rules
- Select fairness checks & owners

Phase 2
Pilot

- Pick a subset of roles
- Compare predictions vs current comp
- Capture exception reasons

Phase 3
Scale + govern

- Expand coverage gradually
- Pair with market intelligence
- Schedule recurring recalibration

Suggested KPIs to track

Median absolute error • % offers outside band • equity gap trend • drift (new titles/levels)