# Turnover Analysis

TLDR (Brief Summary):

1. Strategy and Operations has the shortest tenures – dominated by hourly paid employees at L2 level
2. Pay type change from hourly to salary improved tenure, those who switched to salary stay longer
3. Some locations (due to departments they harbor) have higher turnovers (Such as Moon Mart 1 and Moon Mart 2 which has S&OP), whereas some like New York, Seattle (mixed departments) have better tenures
4. Levels L2 and L3 are more likely to turnover (again same department S&OP), senior levels have longer tenure, except for L10 which is somewhere in the middle
5. 2025 forecast shows that the turnover will stay high and so looking at subgroups (like S&P, L10 employees, entry-level hourly employees) that have higher turnover risk can lower cost for hiring

Problem Statement and Hypothesis Testing: XYZ wants to estimate 2025 strategic headcount budget and to support it is using the turnover analysis. Our Hypothesis is that some job functions, pay types, locations, promotions might be driving most of the attrition and can help reduce some additional headcount needed for next year.

Data Description: XYZ data is panel dataset with employees tracked from 05-2022 to 01-2025 for employees hired from 2018 to 2025. The key columns in the dataset are job function, location, job level, role type, pay rate type and employee status. The data is right censored as active employees' status at effective date is taken.

Methods: For analysis, some transformations I used were panel to cross-sectional and time-series data. Performed descriptive and statistical analysis to see patterns in data and difference in tenure or turnover rate for different department, levels, locations etc using Chi-Square and ANOVA. Implemented Survival and Cox Analysis for different in tenure for various subgroups and also isolating the payrate transition vs no transition. Time-Series Forecasting (using Prophet) for projecting turnovers in 2025 with scenario line for 10% decrease in turnover.

Results: Some sub-groups identified that show higher turnover, such as S&OP + entry level, L10 engineers, Engineers in cities like Seattle and New York, more in Appendix. Predicted department wise forecast for better prediction with scenario testing (what-if) of 10% decrease in turnover.
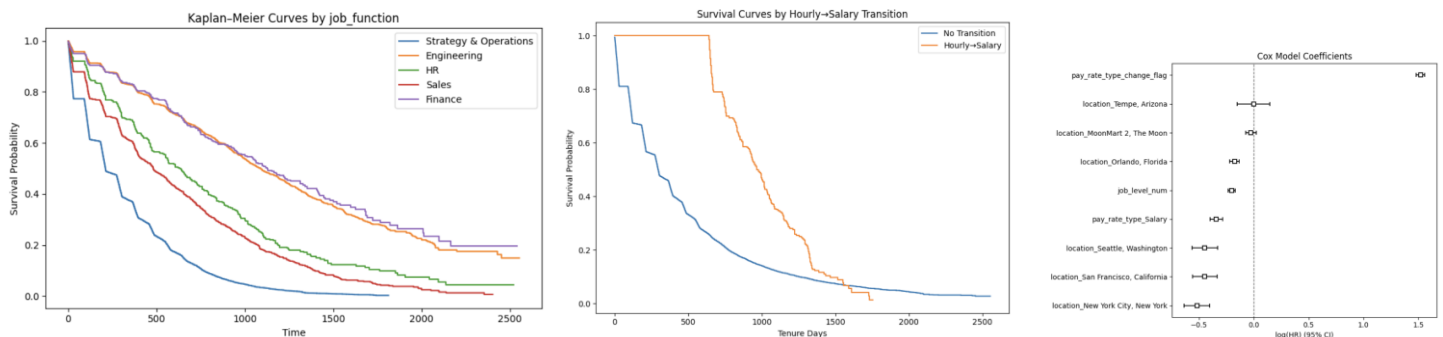
Recommendations: Differentiating the good vs bad attrition, prioritize L10 Engineers turnover, finding the optimal solution for S&OP Hourly L2 employees, turnover vs hourly to salary cost, using the forecast to focus on sub-groups where retention activities are necessary for lower future staffing cost.

Appendix:

Link to the code

Data Preparation & Assumptions

Date Columns:
- Converted effective_date and hire_date to datetime format.

Panel to Cross-Section:
- Since each employee appears multiple times in the dataset (one row per quarter). To fix this duplication, performed sorting by employee_id and effective_date, and then took the last row per employee to obtain a final snapshot (one row/employee).
- This snapshot was used for ANOVA, Tukey HSD, and turnover rate calculations.

Promotion & Pay Rate Changes:
- To check for a promotion, I flagged a promotion when a job level for an employee (job_level_num) in the current row exceeded the previous row's "job_level_num" for the same employee. This method helped to understand if an employee was promoted.
- Similarly, for pay rate change, I flagged a pay rate change (Hourly ↔ Salary) using a lag operation.

Censoring:
- In the survival analysis, the employees which were still active at their last observation date are censored. If the turnover_flag is 1, it means they eventually terminated, else 0.

Analytical Methods

Descriptive Statistics:
- Cross-tabulations - Turnover by pay_rate_type, role_volume_type, location, job_function, and job_level.
  Output – Lower tenure for S&OP in Job function, also for high role_volume_type, lower tenure for hourly pay rate and entry job level like L2 and L3.

- Correlation heatmaps- For numeric features (like tenure_days, job_function_encoded, location_encoded, etc.) No multicollinearity or strong associations.
  Output - No pairwise correlations are extreme, suggesting minimal multicollinearity issues.

Statistical Tests:
- Chi-Square Test - Used it for comparing turnover rates across categorical groups such as location, job level etc.
  Output – Mostly differences are significant but that is also due to large sample size so I did post hoc analysis for the ones where I could see the difference as well.

ANOVA & Tukey HSD

- Used these for comparing the tenure_days across job functions, locations, and job levels.

Output – 1. **Job Function -** *Salaried, Strategy & Operations have a lower mean tenure compared to Engineering or Finance. Sales and HR also show significantly lower tenure compared to Engineering or Finance.*

Output – 2. **Locations -** *Moon Mart 1 and Moon Mart 2 have a lower mean tenure as they also contain employees at entry level position in S&OP department*

Output – 3. **Job Levels -** *Entry level there is lower tenure level than for senior level except for L10 which also has a lower tenure level when compared to other senior and mid level employees*

Survival Analysis
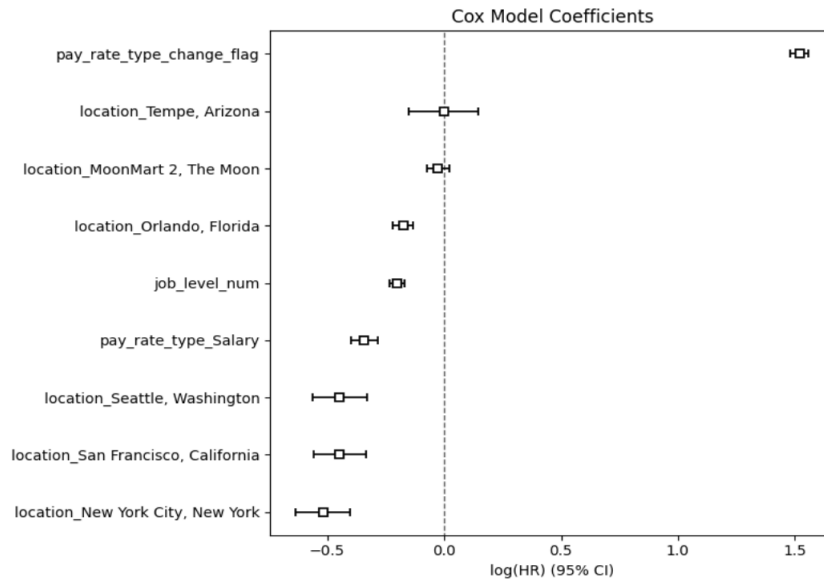
- **Kaplan–Meier curves:**
  Used Kaplan–Meier curves to subgroup comparisons such as job function and pay_rate_type.
  [Output:]
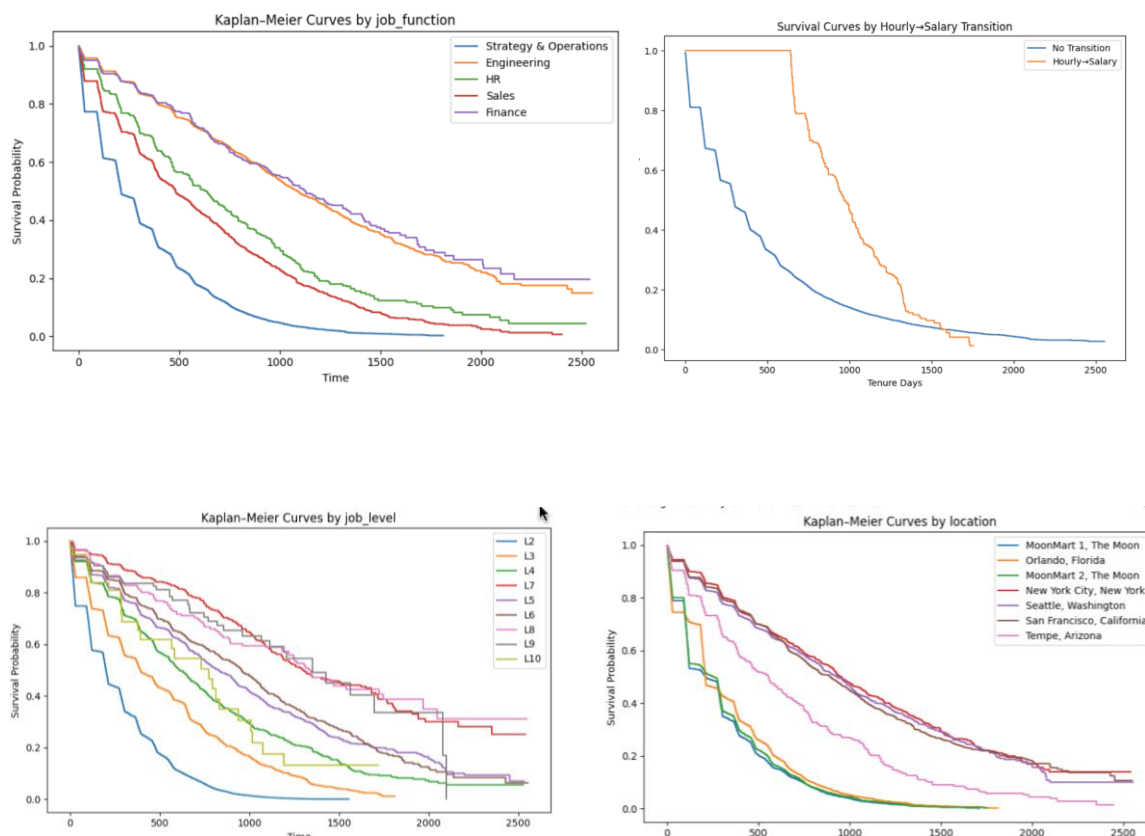
- **Cox Proportional Hazards modeling:**
  Used Cox modeling to quantify each predictor's effect on time-to-exit such as location, job level, pay rate transitions.
  [Output:]

Cox Model Coefficients

- **Time-Series Forecast (Prophet or ARIMA):**
  For Time-Series Forecast, aggregated monthly and quarterly terminations. Constructed Scenario lines (example: 10% reduction in turnover) to check how it lowers next year's hiring needs.
 *Output:*

Strategy & Operations: Quarterly Turnover Forecast (Q4 '24 to Q4 '25)

Terminations in Strategy & Operations are projected to continue increasing through Q4 2025

Sales: Quarterly Turnover Forecast (Q4 '24 to Q4 '25)

Turnover in Sales shows historical volatility but is forecast to slightly decrease and stabilize by Q4 2025.

HR: Quarterly Turnover Forecast (Q4 '24 to Q4 '25)
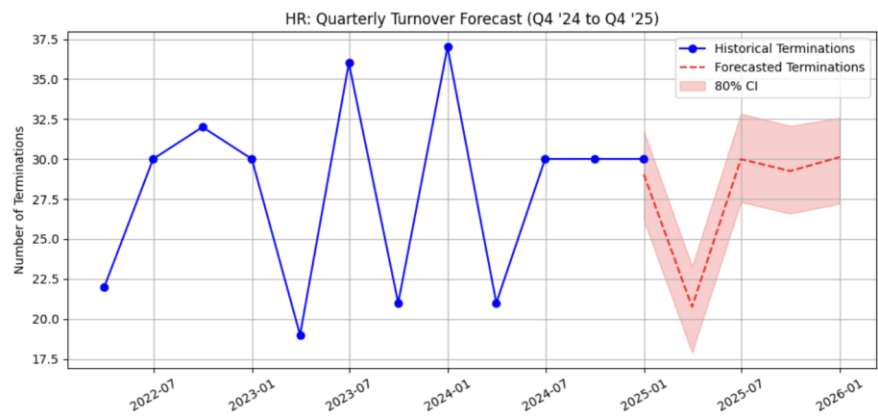
HR turnover historically fluctuates more yet is projected to decline steadily through Q4 2025.

**Engineering: Quarterly Turnover Forecast (Q4 '24 to Q4 '25)**
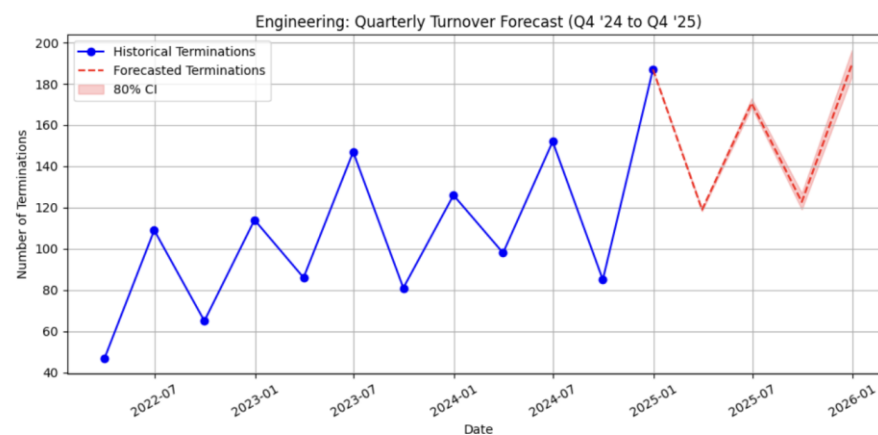
Engineering turnover is projected to spike by Q4 2025, reflecting a higher potential for volatility in the near term.
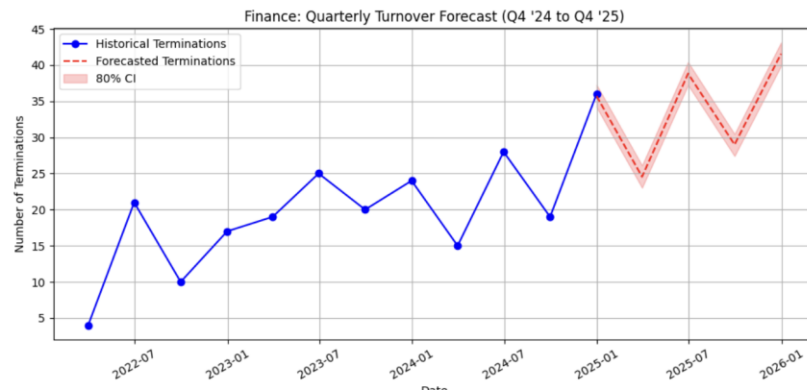
Finance turnover is projected to trend upward through Q4 2025, showing moderate volatility
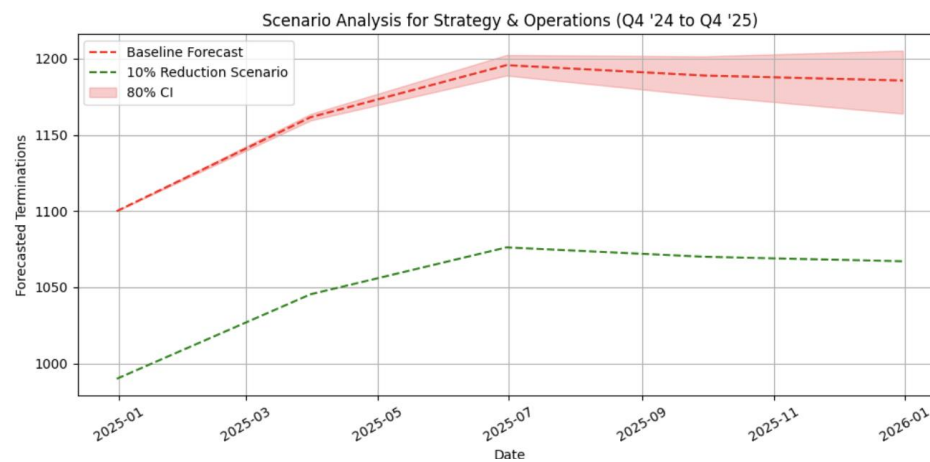

A scenario-based analysis shows that 10% turnover reduction in Strategy & Operations leads to fewer terminations than the baseline by Q4 2025, with moderate uncertainty around the baseline projection.
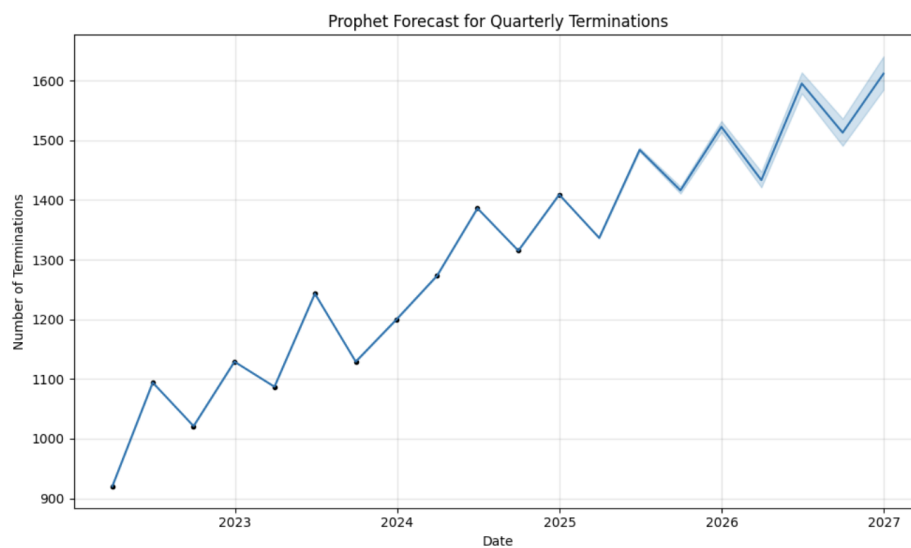
- **Fixed Effect and Random effect modeling:** Hausman Test suggests that Fixed effect model will come into effect, maybe due to the omitted variable bias but current model suggests that promotion might even lead to employee turnover maybe due to competitive market. Further analysis required.

Key Observations

1. **Strategy & Ops:** Strategy and Ops shows a large gap in average tenure vs. other functions (can be seen by the graph in results)
2. **Hourly to Salary Transitions:** About 400 employees changed pay rate, showing improved retention (Cox model hazard ratio < 1). Graph in Results section.
3. **Location Differences:** Large cities fared better than other small cities like Orlando, Moon Mart 1 and Moon Mart 2 (MM1 and MM2 has S&OP employees working at hourly or salaried pays at entry level)

Distribution of Tenure Days by Location
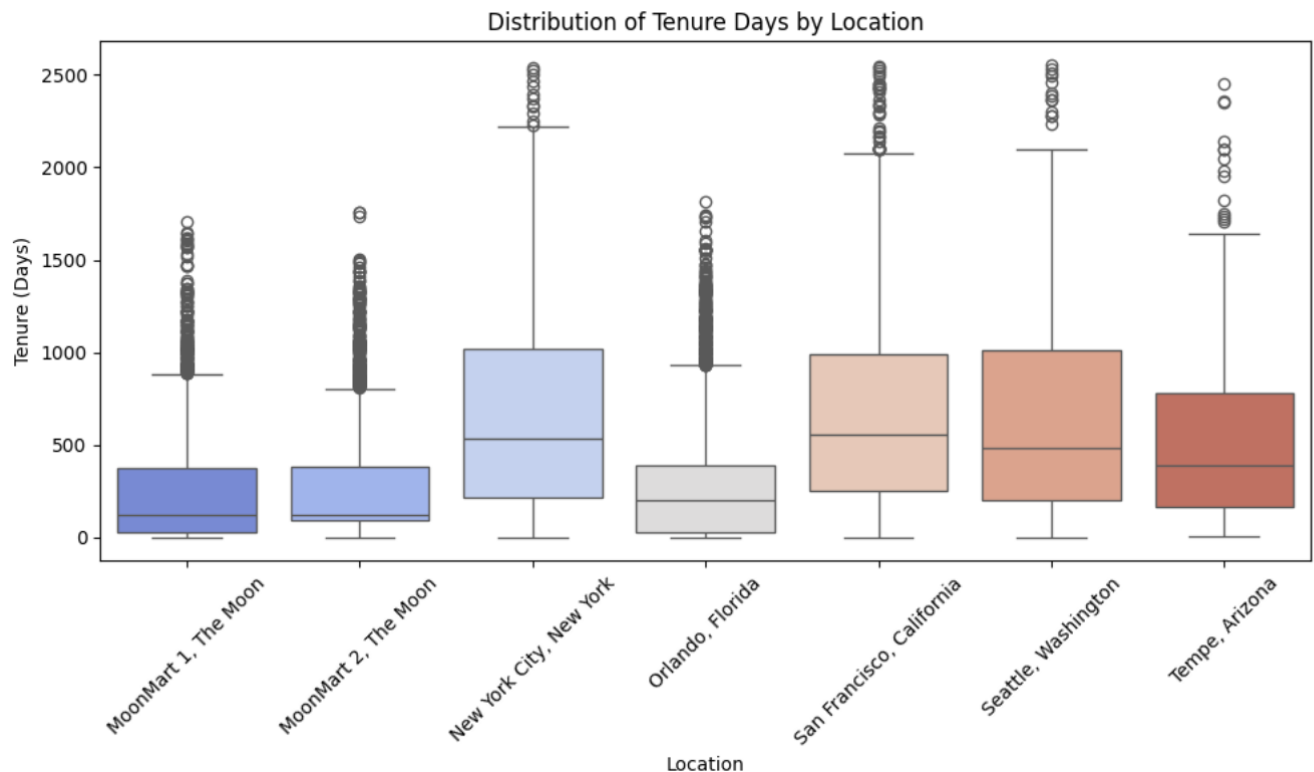
Since MM2 and MM2 have mostly S&OP entry, I did a post hoc analysis for the other major cities like New York and Seattle and found that Engineering tenure is relatively smaller here when compared to HR and Sales by around 140+ days.

4. **Promotion Effects:** Employees with zero promotions over the timeframe often displayed significantly higher turnover rates or shorter tenure.

Turnover Proportions by Promotion Status