UNDERSTANDING TURNOVER

AGENDA

Two distinct approaches

Data & Models Summary

Individual turnover Drivers

Aggregate Trends Drivers

TWO PREDICTIVE MODELING TECHNIQUES HELP US APPLY DESCRIPTIVE ANALYTICS TO UNDERSTAND TURNOVER

Individual Behavior Model

By using detailed individual factors to train advanced models we can understand:

- Who is likely to resign?
- Why did [employee] leave?
- What levers do we have to minimize the risk of turnover?

Aggregate Trend Model

By aggregating what we know about our staff at location, store and year levels, we can use advanced modeling techniques to attempt to understand:

- Why did X% of my staff leave last year?
- What is driving my location's turnover?
- How can I begin to minimize my store's turnover risk?

DATA SOURCES

Distinct problems require distinct approaches-

- To accommodate both types of use cases, I've developed two approaches, both using fictional HR data as a base from Kaggle:
  - Sample Individual HR data from here
    - Demographic information (i.e age, gender, education, distance from work)
    - Role requirements (i.e department, education field
  - Sample aggregate HR data from here
    - Required manual aggregation into city, store number, and year
    - Joined with some publicly available from the US Bureau of Labor Statistics, and St. Louis Federal Reserve to attempt to model overall macroeconomic trends (i.e real median wages, job openings, avg hire rates)

*See fictional dataset used

MODEL OVERVIEW

Both models are unique, but leverage boosting techniques

- Individual model: CatBoost classifier predicting if employee leaves or stays
- Aggregate model: XGBoost Regressor predicting store's annual turnover rate

Boosting model rough overview-

- Imagine I give a room of 100 people some data to sequentially predict if it will rain in the next hour 1000 times. Each person can observe the other's predictions and attempt to account for others' mistakes.
- At first, I average predictions together to aggerate them. Over the course of the 1000 times, I can observe who appears to perform worse, and who appears to perform better and create a weighted sum of everyone's predictions based on their performance to create as strong a model as possible.
- This is how Boosting models work, only its 100 models instead of people, and it goes way faster, with less arguing!

By summarizing how each sub model's predictions changed based on the given data (i.e. marginal impact of adding a given feature), we can use these advanced modeling techniques to provide descriptive analytics on why the model thinks turnover will be 3% or that employee id 24 will resign.

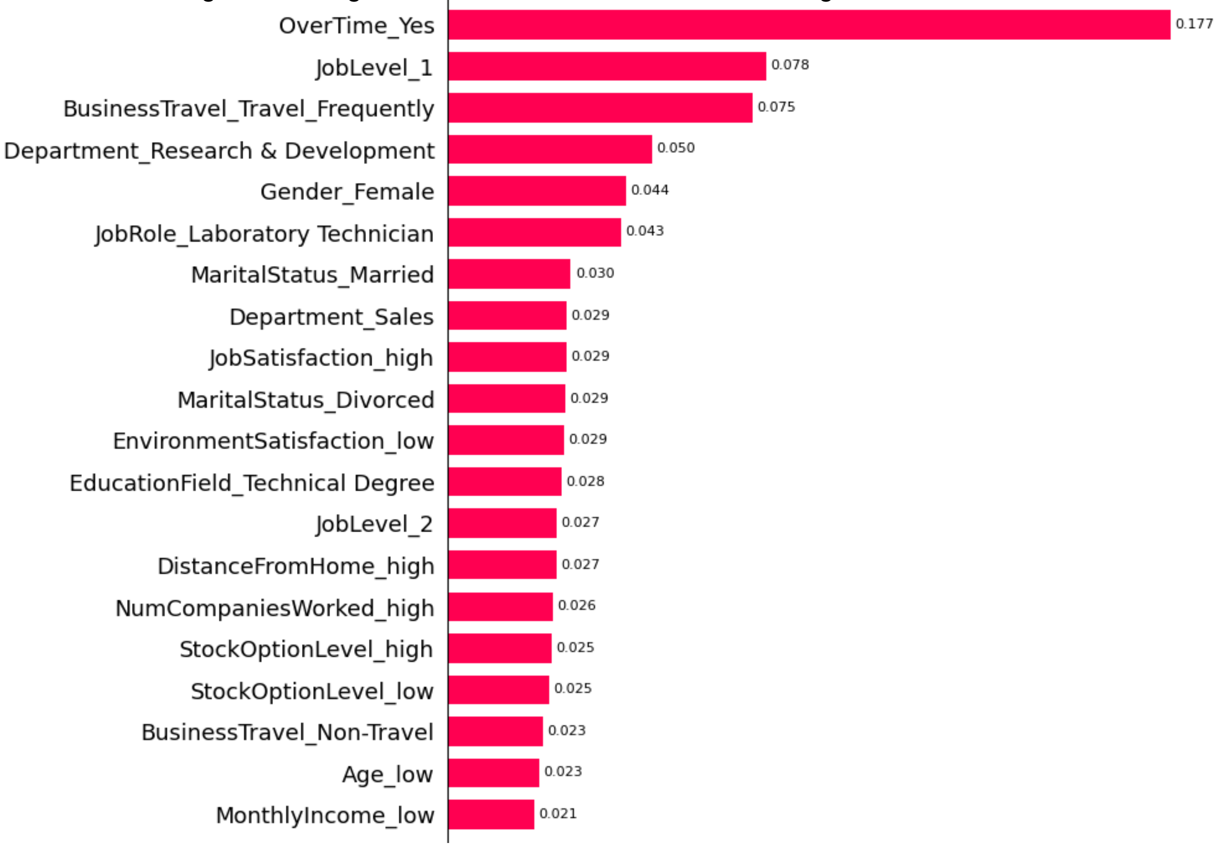
INDIVIDUAL MODEL RESULTS

FEATURE IMPORTANCE

- Top drivers of employee turnover appear to be overtime, entry level positions and travel
  - Overtime appears to on average change the likelihood of employees resigning by ~18%
  - Entry level appears to impact
  - Frequent business Travel on average impacts 7.5% of resignations
- *This measures magnitude, rather than direction (i.e this overtime is most impactful, but this does not say overtime increases resignations)*

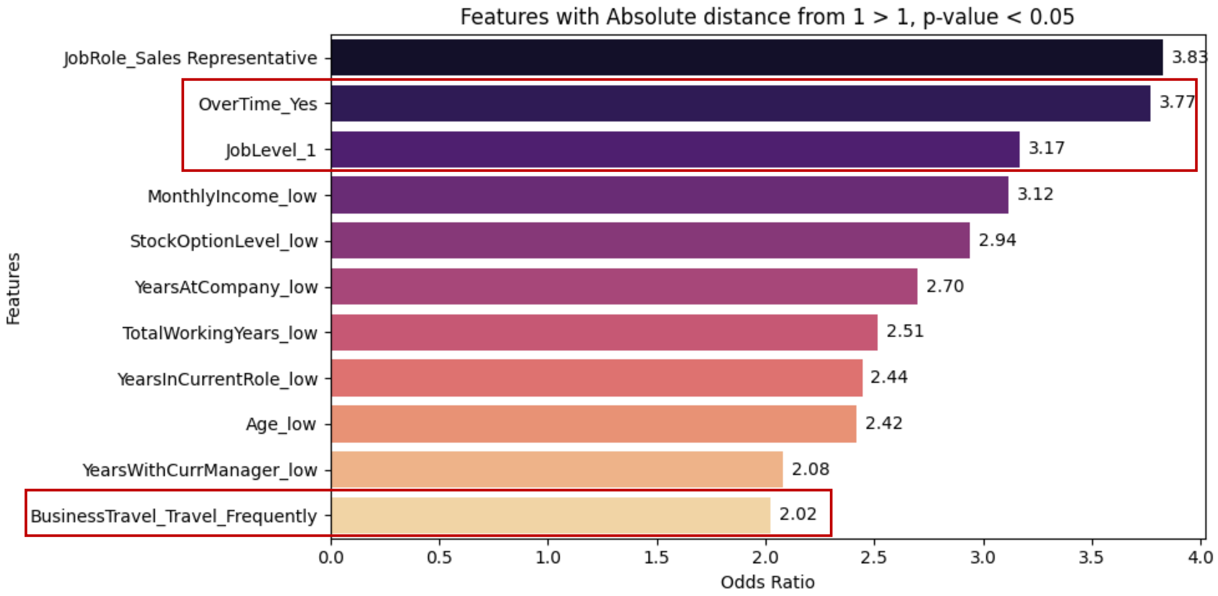
What factors are most impactful to an individual's turnover?

Based on the weighted average absolute value of each features marginal contribution



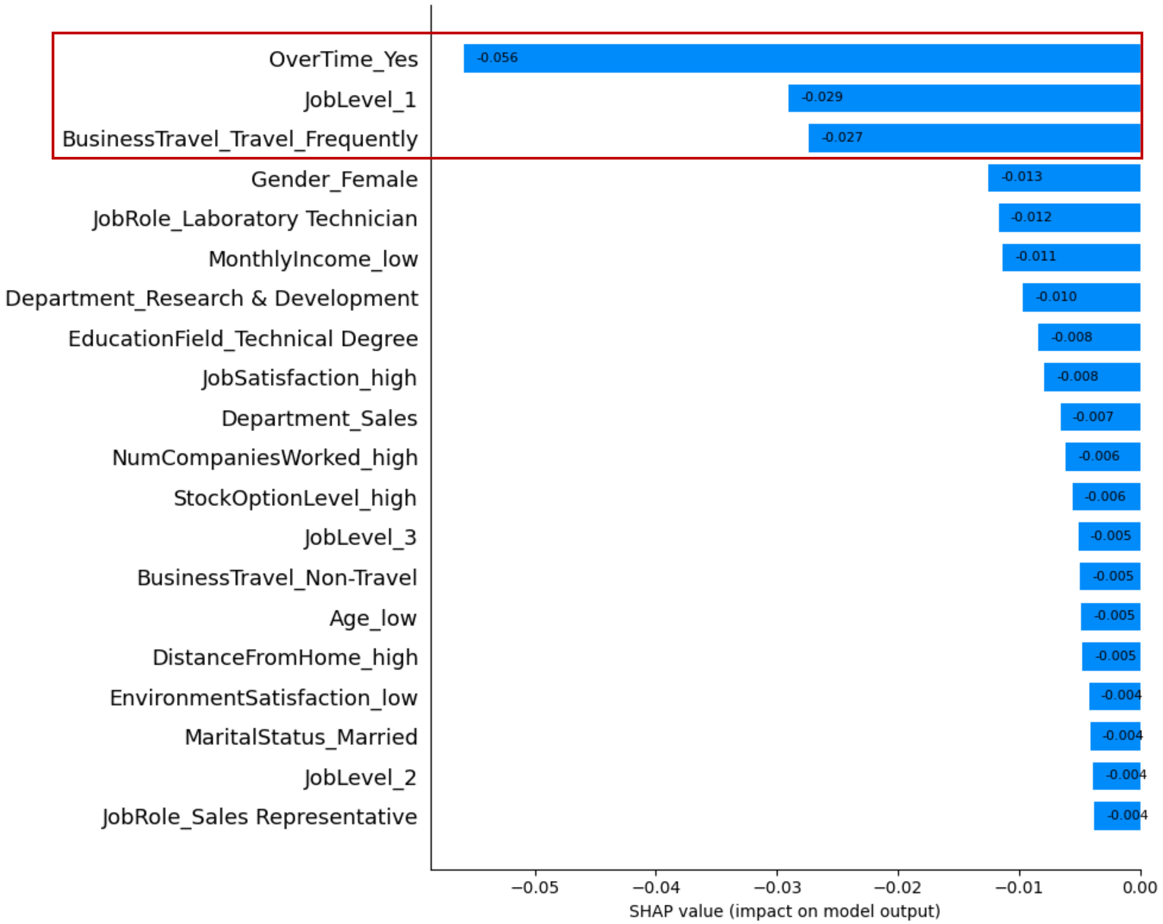
FEATURE RELATIONSHIPS

- Our top drivers (overtime, entry level positions and travel) of employee turnover appear at the top again, and all appear to increase the likelihood of resignations
- We can also use odds ratios from a chi squared contingency test to depict this relationship:
  - From this, we can see that over time makes employees 3.7 times as likely to resign!
  - Entry level employees are 3.2 times as likely
  - Frequent business travelers are 2 times as likely



How do these features impact an individual's turnover?

Based on the weighted average value of each features marginal contribution



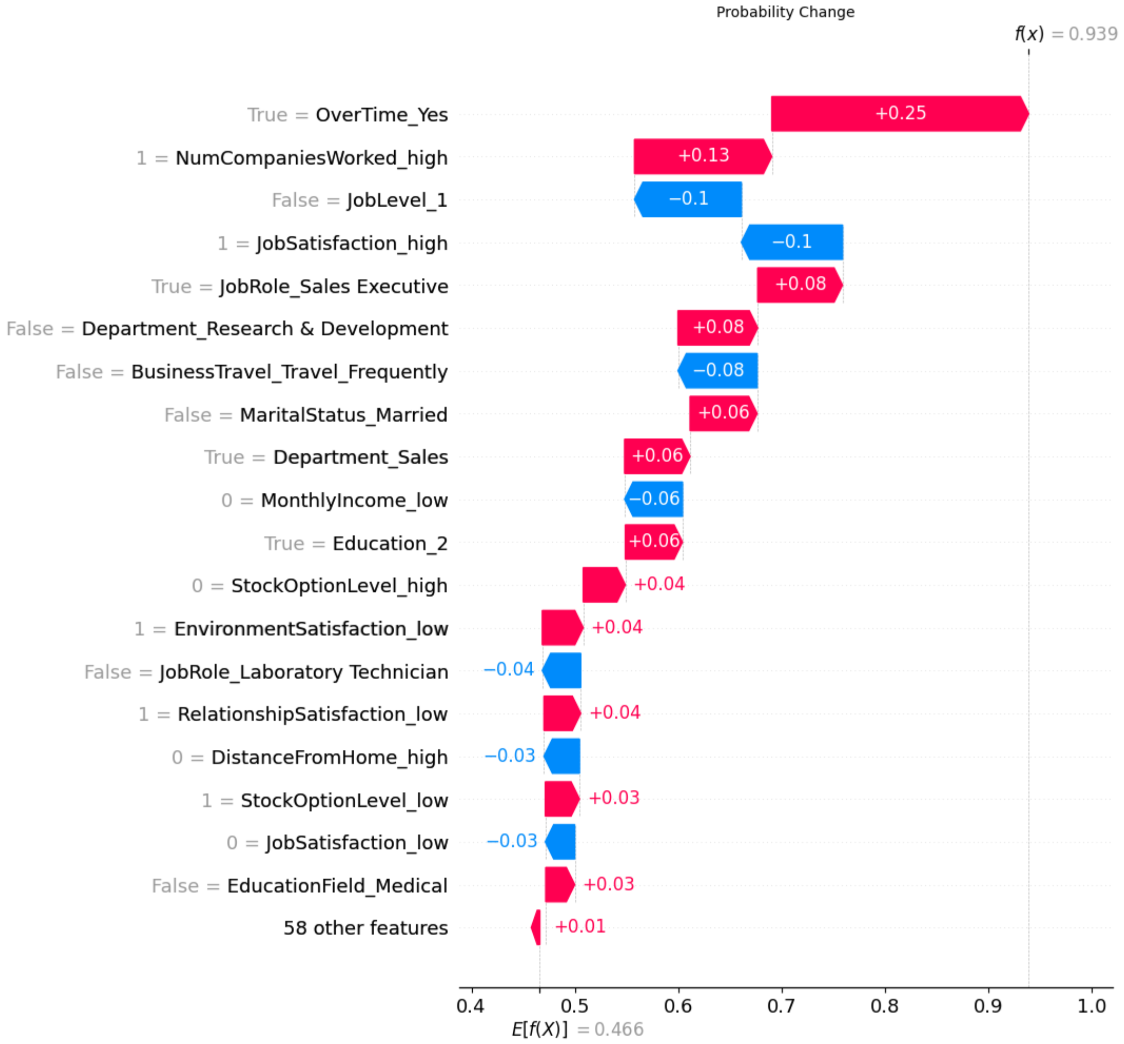
INDIVIDUAL INSIGHTS

Another benefit of this model, is we can examine how each feature impacted a given final prediction

- In this example, we can see again how large the impact of overtime is to the probability of resignation for a specific person
  - Overtime increased the probability that this employee would resign by 25 percentage points!
  - These insights can inform our decision making- *is the value from this employee's overtime work worth the added risk of their resignation?*

Why did [person] resign?

Based on the features marginal contribution to the overall prediction



POTENTIAL MODEL IMPROVEMENTS

- All number features were converted into categorical variables for easy interpretation- wages were translated from actual wages to `high` wages or `low` wages
  - This was not sophisticated in its approach
    - Wages are high/low with respect to the job market and specific requirements of the role
    - Understanding how an additional marginal dollar or percentage increase in wage might impact the likelihood of resignation is useful, but this model cannot provide that with this technique

- Additional factors might drive turnover exogenous to the model
  - Potential features might include
    - Sentiment analysis on any available natural language from employees
    - Recent increases in sick/PTO/vacation days might provide and early indication of potential turnover
    - Social network graph analysis might provide some useful insight on how specific managers or teams effect the likelihood of turnover
    - Some feature depicting resignation momentum may improve the model – recent increased resignations might increase the likelihood of following
    - Lack of seasonal awareness or trend in the model – staffing planning might benefit from some understanding of how seasons impact resignations

AGGREGATE MODEL RESULTS

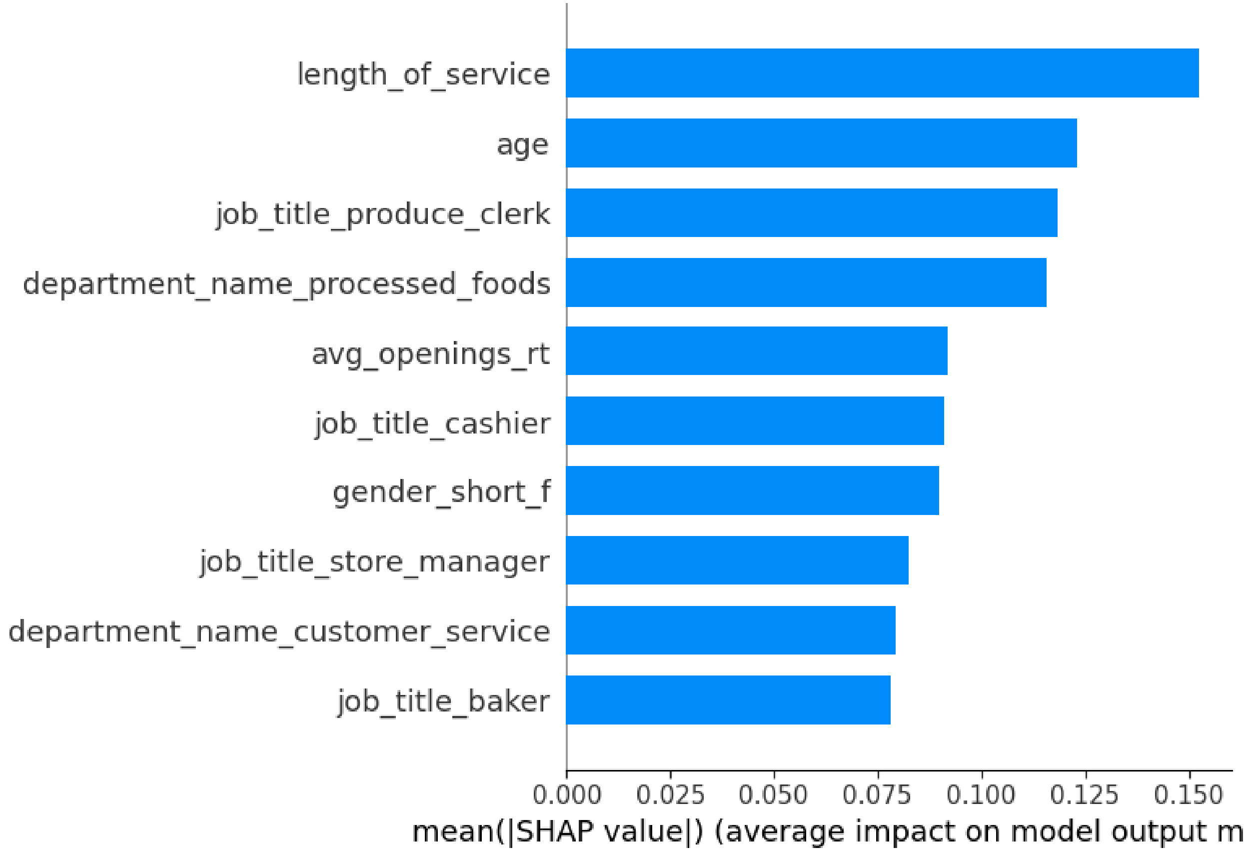
FEATURE IMPORTANCE

The trend has less detailed data than the individual model's sampled data, however there is more sample

To get the data in terms of city / store / year

- The trend model has different inputs- it seems like the top 4 are the store's average length of service, age, percentage of workforce that are produce clerks or work in processed foods

What factors are most impactful a stores annual turnover?



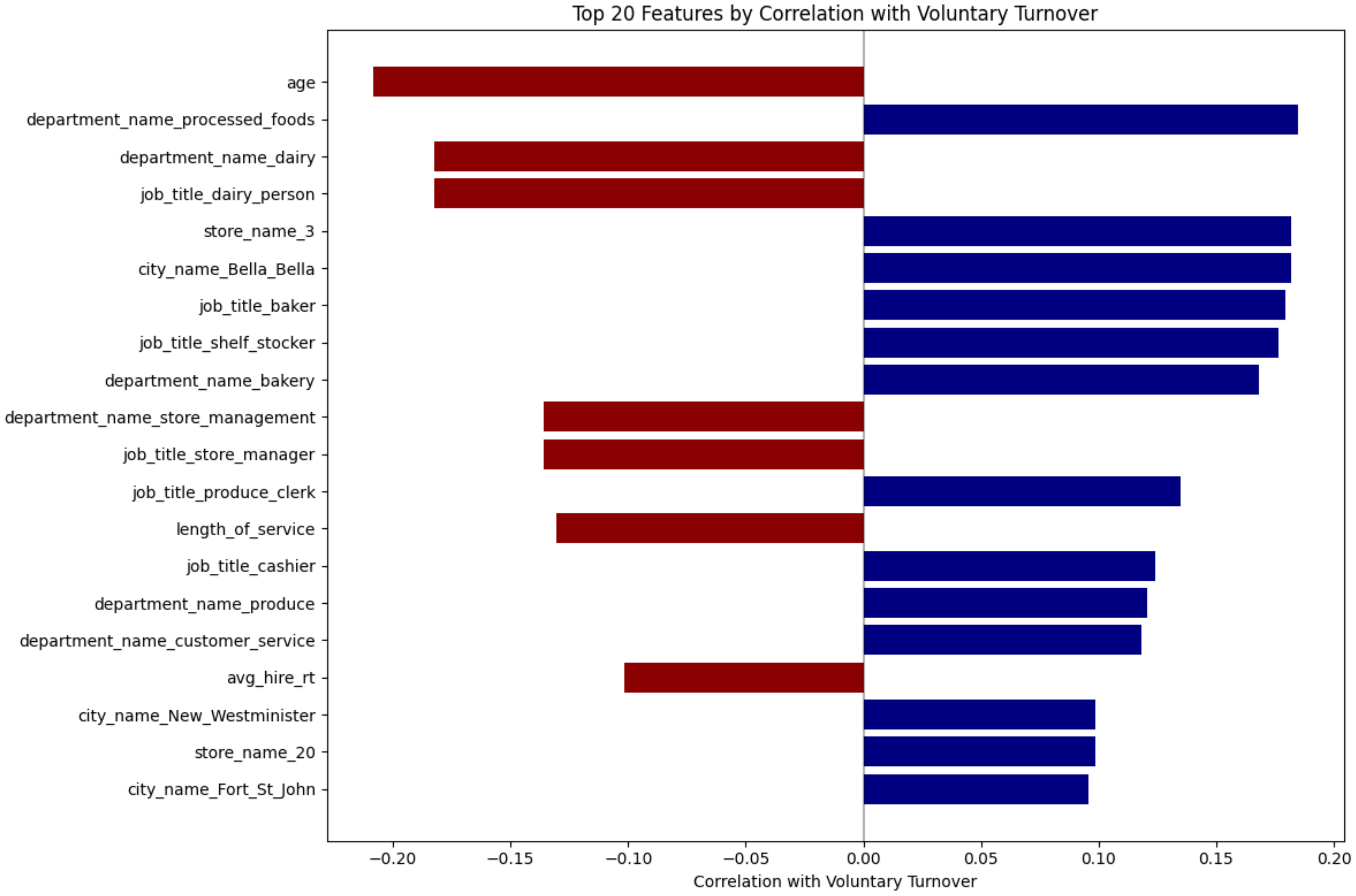
FEATURE IMPORTANCE

Next, we can view correlations to understand how the model might apply these features

- It seems the higher the stores average age, the less expected turnover
- The percentage of processed food employees increases the estimated turnover
- As percentage of produce clerks increase, the turnover rate increases

How do these features impact a stores annual turnover?

Based on the Pearsons correlations between features and annual turnover



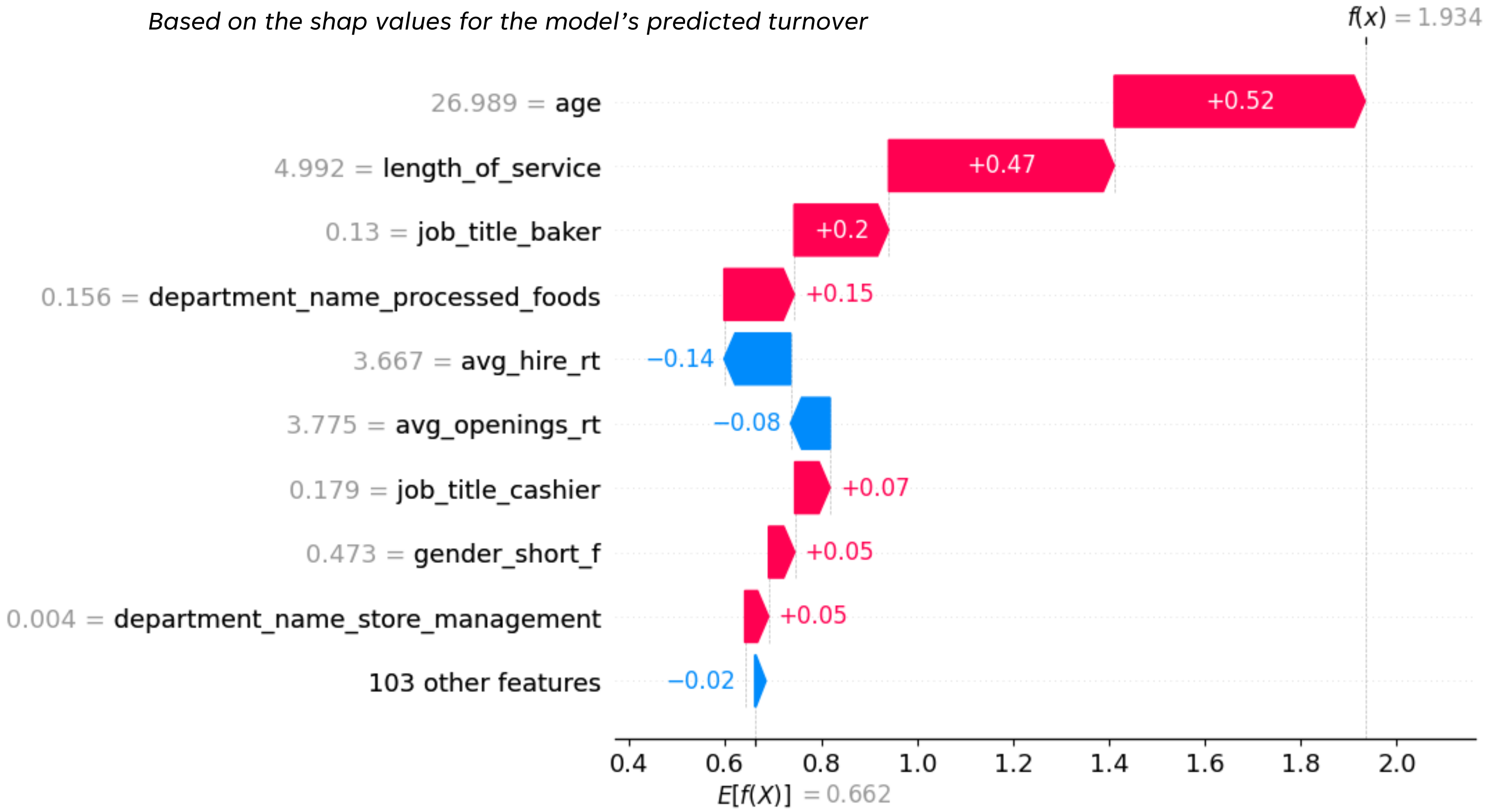
WHY DID MY STORE HAVE THIS TURNOVER?

Like the individual model, we can use this technique to highlight why the model expects a given turnover for a store/year/location

- In the example of store 44 in Vancouver in 2015, we can see that the average age of the workforce drives the expected turnover much higher, as well as the length of service.

Why did store 44 in Vancouver have this turnover in 2015?

Based on the shap values for the model's predicted turnover

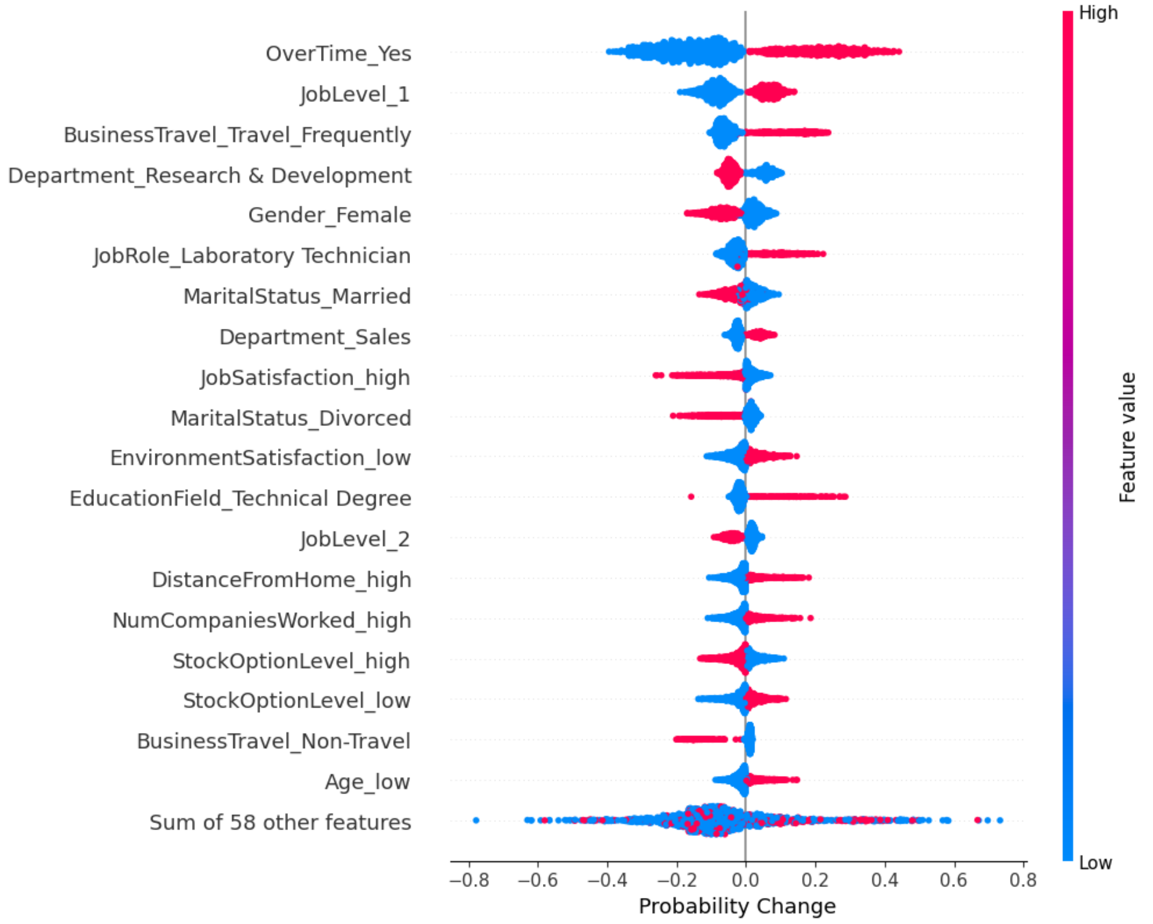


APPENDIX

Individual model details

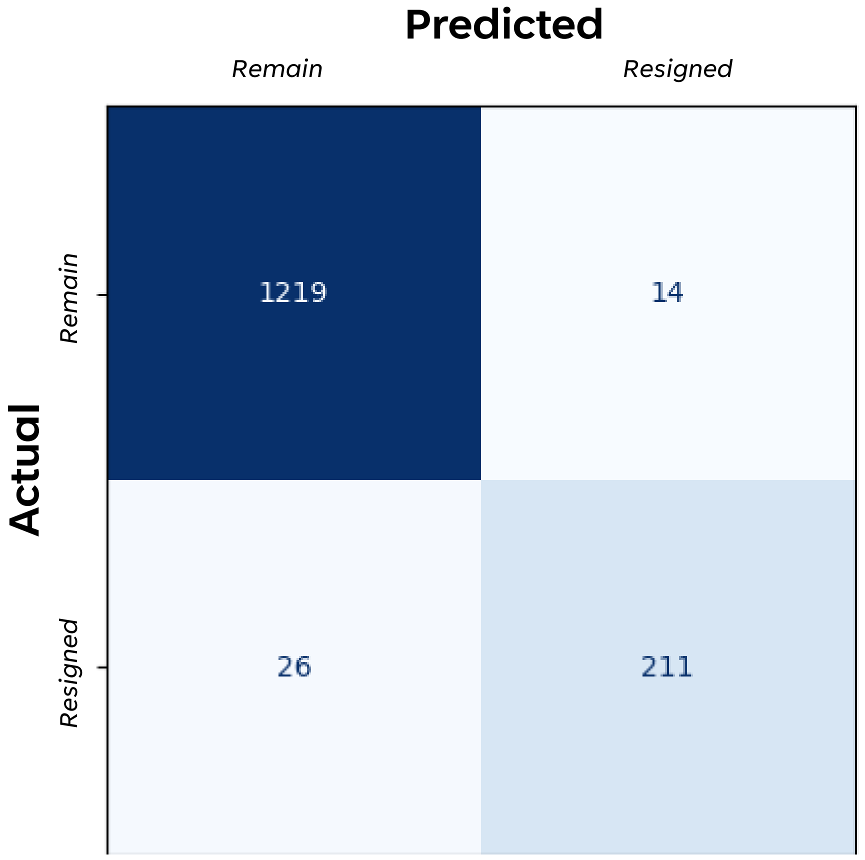
Individual model beeswarm

Here each sample (employee) is plotted in a distribution where the color is the value of the feature (high = red) and the x-axis is the features marginal change to the prediction



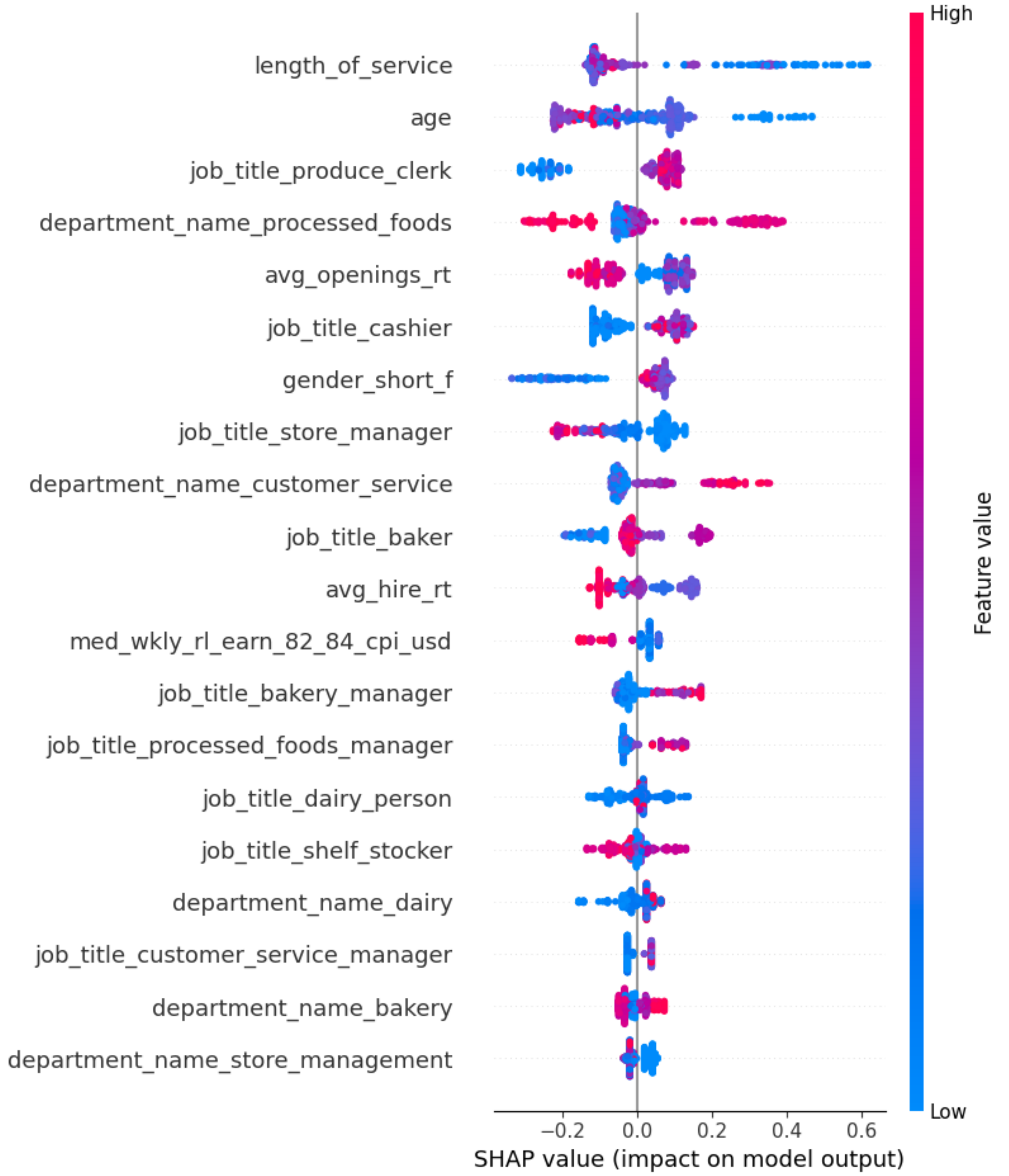
Individual model confusion matrix

Model had 86% accuracy in predicting employee turnover, capturing 89% of overall actual turnover

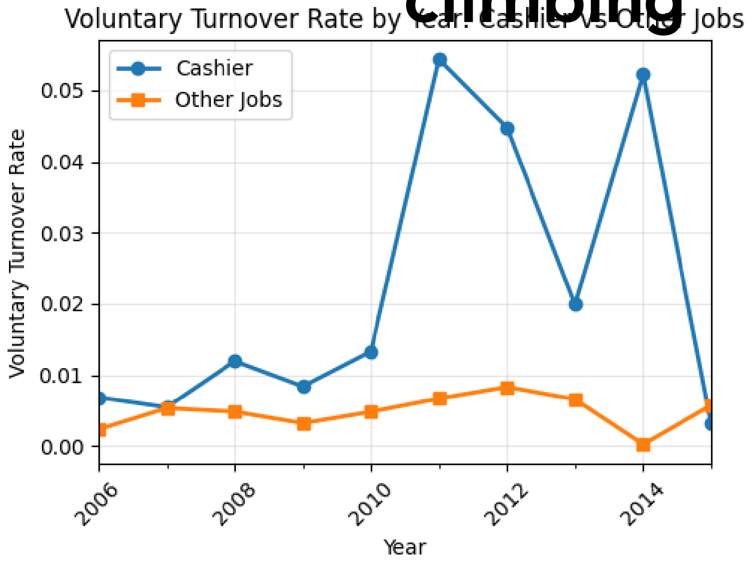


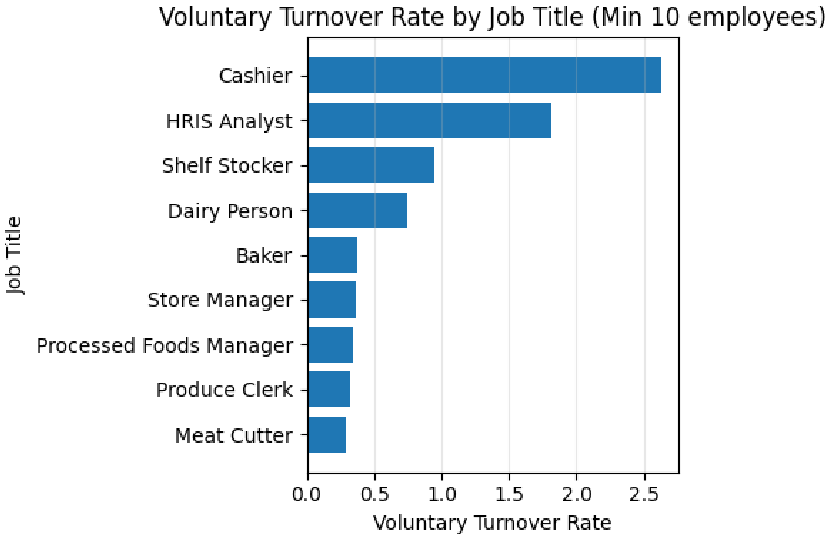
Trend model beeswarm

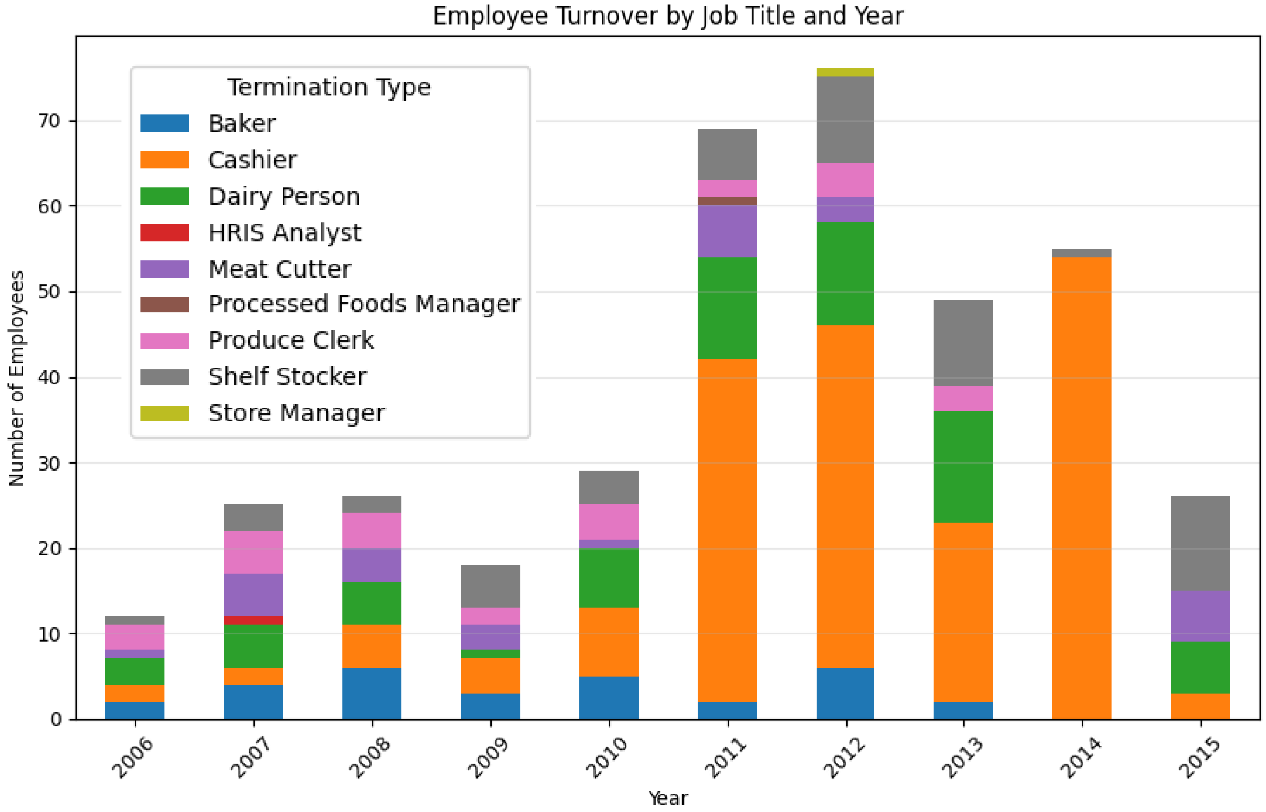
Here each sample (city, store, year) is plotted in a distribution where the color is the value of the feature (high = red) and the x-axis is the features marginal change to the prediction



Trend Dataset: Cashier resignations climbing

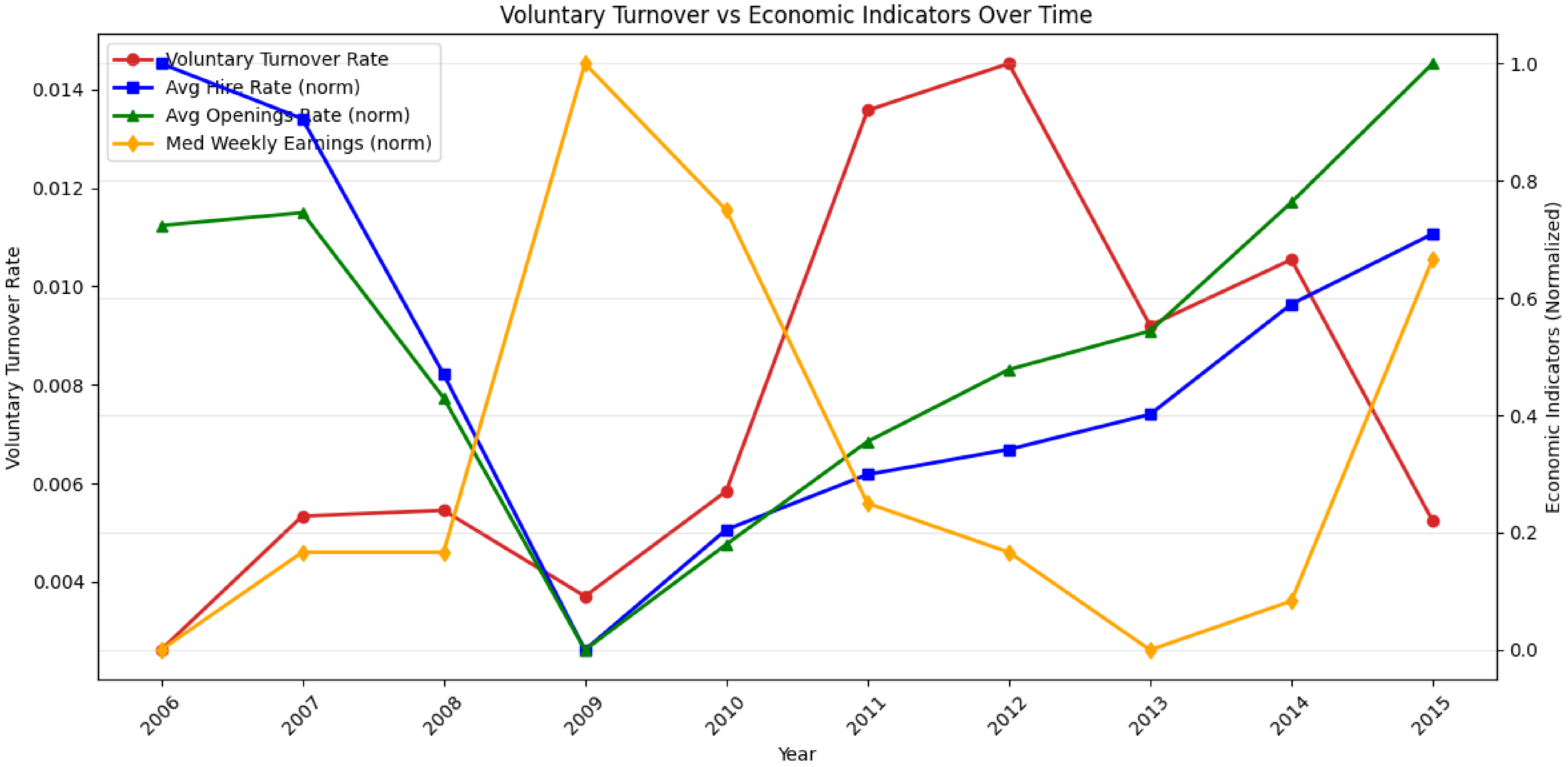






Economic Indicators by Resignations

At this aggregated level, the relationships between these features are unintuitive, requiring more analysis at a lower level of detail for valuable insights



DYNAMIC DELIVERY

Learn to infuse energy into your delivery to leave a lasting impression

One of the goals of effective communication is to motivate your audience

METRIC	MEASUREMENT	TARGET	ACTUAL
Audience attendance	# of attendees	150	120
Engagement duration	Minutes	60	75
Q&A interaction	# of questions	10	15
Positive feedback	Percentage (%)	90	95
Rate of information retention	Percentage (%)	80	85

FINAL TIPS & TAKEAWAYS

Practice makes perfect

- Consistent rehearsal
  - Strengthen your familiarity
- Refine delivery style
  - Pacing, tone, and emphasis
- Timing and transitions
  - Aim for seamless, professional delivery
- Practice audience
  - Enlist colleagues to listen & provide feedback

Continue improving

Seek feedback

Reflect on performance

Explore new techniques

Set personal goals

Iterate and adapt

SPEAKING ENGAGEMENT METRICS

IMPACT FACTOR	MEASUREMENT	TARGET	ACHIEVED
Audience interaction	Percentage (%)	85	88
Knowledge retention	Percentage (%)	75	80
Post-presentation surveys	Average rating	4.2	4.5
Referral rate	Percentage (%)	10	12
Collaboration opportunities	# of opportunities	8	10