Analyzing Warehouse Productivity

Ankita Devasia, Lydia Yu, Deekshita Kacham

Background and Significance

Supplement work done by the Warehouse Work, Health, and Well-Being study conducted under Professor Erin Kelly at Sloan

Purposes of study:

- Understand what it's like to work at a fulfillment center for an online retailer
- Identify how changes in the workplace affect workers' experiences
- Investigate how work conditions may affect workers' decisions to stay or leave job

Team aimed to build on study by finding relationships between turnover and warehouse productivity

Objective

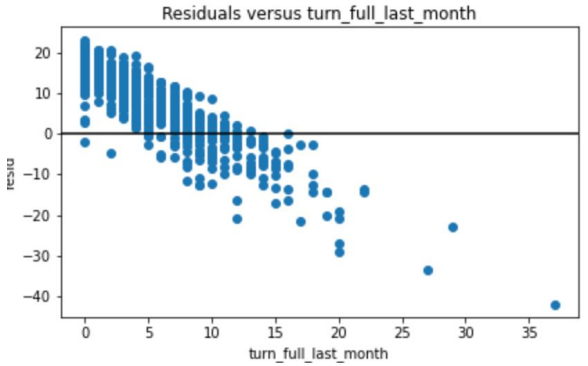
- Support the Warehouse Work, Health, and Well-Being study by studying the effect of turnover on productivity within warehouses of a national online retailer
- Finding an optimal level of turnover that maximizes productivity
    - Have data from an E-commerce retailer as well as monthly turnover data for the retailer's warehouses, which corresponds to how many workers leave each fulfillment center

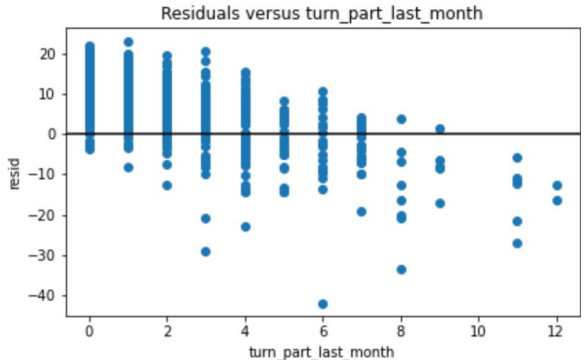
Data

- Data on labor productivity, taken weekly, for each fulfillment center (FC) between 2011 and 2020
- Monthly data on turnover for each FC between 2017 and 2020
  - Aggregated two datasets
- Final dataset has 36 features
  - **Turn_full_last_month:** turnover of full-time employees from previous month
  - **Boxes_per_hr:** boxes moved per hour, primary measure of productivity
  - **Ot_hrs:** total overtime hours worked
  - **Tot_sales:** total sales

Exploratory Data Analysis

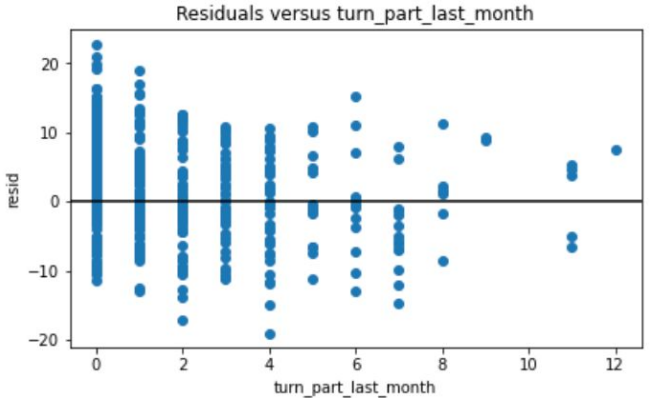
- Plotted turnover (FT (full time) and PT (part time)) versus productivity (boxes_per_hr).
  - No clear correlation
- Simple regression of productivity on FT and PT turnover.
  - R^2 value of 0.394, probably due to outliers
- Removed outliers and ran regression again.
  - The R^2 was still quite low.
  - Noted non-linear relationship with FT/PT turnover since the residuals not randomly distributed
- Introduced features $turnoverFT^2$ and $turnoverPT^2$
  - Residual plots looked better, decided to introduce more features

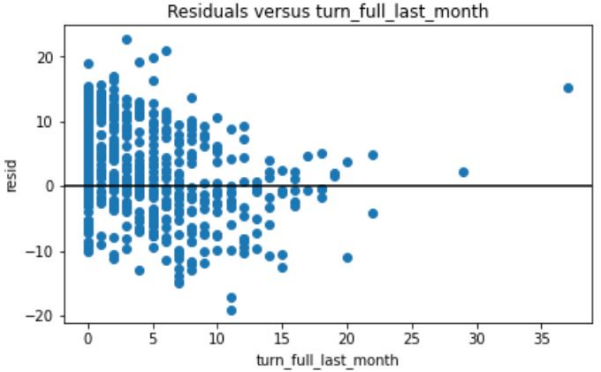




Model 1: Linear Regression

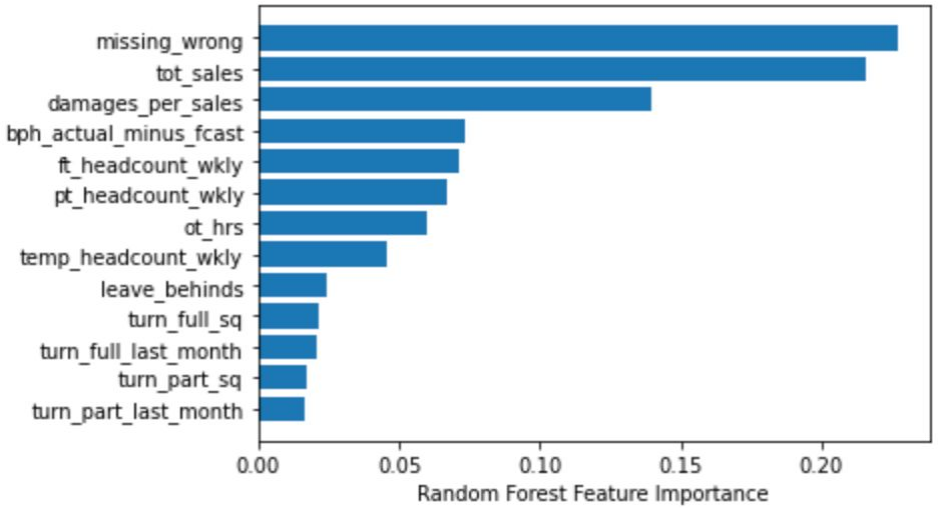
- **Response variable:** productivity (boxes_per_hr)
- **Explanatory variables:** full time turnover, part time turnover, turnover_FT^2 and turnover_PT^2.
- **70%** of the data in training and **30%** of the data in testing and build a model on **training**
- Features with the highest significance are **turnover variables, bph_actual_minus_fcast, ft_headcount_wkly, ot_hrs**
- Residuals are much more randomly distributed in this model





Model 2: Random Forest

- Random forest regression using 30% train/test split
  - R^2 of 0.953
  - MSE of 1.935
- Feature importances rank turnover variables as least important



OLS vs. RFR

- OLS has more interpretable results, but it is not as accurate
- RFR doesn't know how to extrapolate outside of the data set, which could be problematic
- RFR probably performs better since there isn't a linear relationship between turnover and productivity and it is able to identify nonlinear relationships
    - Risk that it could be overfitting
- If we wanted a more accurate model that is able to extrapolate, we could train a neural network
    - Ideally done with more data

Model 3: Optimization

- Aimed to maximize the productivity for a given FC for a given month. We used coefficients from the train-test-split model

  Objective Function:

  maximize productivity for a given FC for a given month
  max 1.163F + 1.570P - 0.037F2 - 0.145P2
  F and *P* are the levels of full time and part time turnover per month

  Constraints:

  Turnover (positive or negative integer) for next month cannot exceed the headcount that the FC starts out with (equal to headcount for the last week of the previous month).

  Results for FC 6710 2020-06: a turnover of 16 FT and 5 PT maximizes productivity

Results and Insight

- Having some amount of turnover, usually associated with **negative worker well-being,** actually **improves productivity**
- The social science aspect of this study would look at **“which workers are the ones that should go?” “Is it ethical to lay off workers to increase productivity?”**
- It would be great to have **more turnover data for a longer time period**.
  - Focus on other features affecting productivity
  - Seasonality component
  - Separate data on those who left due to work conditions and those who left for other reasons
- Ultimately, the goal of this project is to **improve worker well-being**, not necessarily productivity
  - Turnover is usually associated with negative well-being, but it improves productivity, which is usually associated with positive well-being

Thank you!

Questions?