15.780 Final Project Proposal
Ankita Devasia, Deekshita Kacham, Lydia Yu

Optimizing Productivity for Warehouse Worker Wellbeing

Project Topic

Our project will supplement work done by the Warehouse Work, Health, and Well-Being study conducted by the Sloan School of Management under Professor Erin Kelly. The purpose of this study is to to understand what it's like to work at a fulfillment center for an online retailer and to understand how changes in the workplace affect workers' experiences. It also hopes to investigate how work conditions may affect health and well-being, personal and family life, and decisions about whether to stay in the job or leave.

Honing in on one subset of the problem, our team hopes to supplement the work already done by the lab and find relationships between particular variables and the productivity of the warehouse. Upon doing some initial research, we have decided on the problem statement discussed in the next section. More information on the study itself can be found at https://mitsloan.mit.edu/programs/phd/mit-warehouse-work-study.

Problem Statement

Our goal is to support the Warehouse Work, Health, and Well-Being study by studying the effect of turnover on productivity within warehouses of a national online retailer and finding an optimal level of turnover that maximizes productivity. We have labor productivity 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 (either voluntarily or through lay-offs) each month.

Significance

The Warehouse Worker Well-being study seeks to document ways to improve workers' experience in warehouses, a setting rapidly being affected by e-commerce and more recently by the pandemic. This project aims to investigate how work redesign – deliberate changes in work conditions – may improve the health and well-being of workers.

The pandemic has significantly increased consumers' reliance on online shopping, putting strain on many warehouse workers. In light of this, Professor Kelly's lab partnered with the e-commerce division of a national retail firm to introduce a participatory workplace intervention and its impact on employees' health and wellbeing and on key organizational outcomes. The study has just finished initial rounds of survey data collection in which workers of the various warehouses of this retailer were surveyed on topics related to their wellbeing and work conditions. The results of these surveys are currently being communicated to the managers of each of the fulfillment centers at which the surveys were sent out.

A significant measure of worker wellbeing is how likely they are to leave their jobs or be laid off. If turnover has a critical impact on the productivity of a warehouse, then this relationship could be an important target for our study because productivity is directly tied with profits and other measures of business outcomes.

Data

Through this study, we have acquired data on labor productivity, taken weekly, for each fulfillment center (FC) between 2011 and 2020. We also have monthly data on turnover for each FC between 2017 and 2020, which we will lag by 1 month because we want to observe how turnover from the previous month affects current productivity. We can aggregate the weekly labor productivity data into monthly data by grouping by FC and month and taking the average values per FC per month, which can then be merged with the monthly turnover data to get one dataset with the following features:

- **fc_id:** unique identifier for each fulfillment center
- **year_month:** month of current observation in YYYY-MM format
- **turn_full_last_month:** turnover of full-time employees from previous month
- **turn_part_last_month:** turnover of part-time employees from previous month

- **turn_sum_last_month:** turnover of full and part-time employees from previous month
- **boxes_per_hr:** boxes moved per hour, the primary measure of productivity
- **bph_actual_minus_fcast:** difference between actual and forecasted boxes per hour
- **cargo_loss:** boxes lost
- **damages_per_sales:** $ value of damages incurred as a percentage of total $ value of sales
- **fcast_cargo_loss:** forecasted cargo loss
- **fcast_damages_per_sales:** forecasted damages value as a percentage of sales
- **fcast_labor_per_sales:** forecasted labor value as a percentage of sales
- **fcast_ot_hrs_pct_tot_prod_hrs:** forecasted overtime hours as a percentage of total hours
- **fcast_prod_avg_wage:** forecasted average wage
- **fcast_receiving_cartons:** forecasted cartons received
- **fcast_vol_boxes_wkly:** forecasted boxes processed per week
- **ft_headcount_wkly:** forecasted headcount per week
- **labor_per_box:** $ value of labor per box handled (total wage bill / total boxes)
- **labor_per_sales:** $ value of labor as a percentage of sales (total wage bill / total sales)
- **leave_behinds:** boxes that were left behind and didn't make it onto a truck
- **missing_wrong:** boxes that were missing or wrong
- **ot_hrs:** total overtime hours worked
- **ot_hrs_pct_tot_prod_hrs:** overtime hours as a percentage of total hours worked
- **prod_avg_wage:** average wage of workers
- **pt_headcount_wkly:** headcount of part-time workers per week
- **pt_ot_hrs:** overtime hours worked by part-time workers
- **pt_tot_hrs:** total hours worked by part-time workers
- **receiving_cartons:** cartons received per week
- **temp_headcount_wkly:** headcount of temp workers per week
- **temp_hrs:** total hours worked by temp workers
- **temp_ot_hrs:** overtime hours worked by temp (3rd party contracted) workers
- **tot_damages:** $ value of total damages incurred
- **tot_headcount_wkly:** total headcount of all workers
- **tot_prod_hrs:** total hours worked by all workers
- **tot_sales:** total sales
- **vol_boxes_wkly:** boxes processed per week

In the labor productivity data, there exists a measure of forecasted productivity (this is the target level that the FCs aim to reach), which can potentially be used as a measure of how close the FCs are to their optimal productivity.

Modeling & Decision Approaches

Broadly, we plan to predict the effect of turnover of different types of workers, full-time (FT) and part-time (PT), on productivity for the retailer. This can be done by running a simple regression of productivity on turnover of FT and PT workers. We will likely also include other covariates such as number of workers, sales, and damages as features in this regression. We will then use the results of this predictive problem within an optimization problem to maximize productivity, where decision variables are the number of FT/PT workers “turning over” and their coefficients come from the predictive step. This step will suggest an optimal level of turnover per month, which may be applied to the business as perhaps a suggestion on how many people an FC should be laying off (positive turnover) or hiring (negative turnover) on a monthly basis. We can also extend this analysis to observe how productivity, and therefore turnover, affects profits.

In terms of application, we have two options: we can analyze how turnover affects productivity in general for all FCs and come up with an average level of optimal FT/PT turnover across all FCs that yields maximum productivity. If we include features like FC size into the regression, this could potentially take into account the differences between FCs. Our second option would be to filter the data by FC and analyze each FC on its own.

Challenges

In the scope of this project, there are a few challenges. First, relative to labor productivity, there isn’t as much data on turnover; it is monthly and we only have data from 2017-2020. In addition, if we were to analyze each FC individually to find the optimal turnover level for each one, we would need to further filter by FC and possibly create FC-specific models that would have limited amounts of data. However, if we were to analyze the productivity of all FCs together, we would miss nuances and trends that may differ across the FCs.

Potential Approaches

1. Data Preprocessing and Initial Analysis

In our warehouse data set, some of our variables have missing values. An important initial step we need to take is to impute missing values that we may need for our final analysis. The exact approach we take to do this needs to be well-informed by correlations with other variables. After we do this, we will look at the data and see if there are any interesting relationships between variables to be mindful of (examine correlations).

2. Forecasting

We will primarily consider using a linear regression to make a prediction. In this case, we anticipate that the decision variable will be productivity and that the independent variable is mainly turnover--we will account for residuals by fitting other potential variables that affect productivity. If the results from our regression model do not work well, we may consider using a random forest technique to generate a prediction.

3. Evaluation

Once we get results from our forecasting, we will use a Root Mean Squared Logarithmic Error (RMSLE) to analyze our results.

4. Optimization

Based on our results from the forecasting step, we will use an integer program with the objective of maximizing productivity where the decision variables are the levels of turnover for full-time and part-time workers. Constraints for this problem may include a minimum/maximum number of employees in a warehouse or a minimum level of productivity that must be reached.