Musa Practicum Scope of Work, Spring, 2022
A human resources turnover intelligence system (Guilford County, NC)

Instructors:
Michael Fichman & Matthew Harris
Team:
Jiali Yao, Jonathan Clementi, Olivia Scalora
Client Contacts:
Sherri Bigelow, Enterprise Application Manager (sbigelow@guilfordcountync.gov)
June Harley, Human Resources Manager, (jharley@guilfordcountync.gov)
Graham Rothrock, Talent Acquisition Manager (grothro@guilfordcountync.gov)
Jason Jones, Assistant County Manager (jjones6@guilfordcountync.gov)
David Albright, IT - ERP Specialist (dalbrig@guilfordcountync.gov)

Governments or firms invest precious time and money into training their workforce. Each employee also holds institutional knowledge that is crucial to the organization functioning effectively. When an employee leaves (when there is “turnover”), they take with them that knowledge, and the impact of their departure is felt not just during their absence, but during the training of their successor. Guilford County, NC is seeking to lessen this issue by creating a more pro-active approach to avoiding employee turnover.

The goal of this project is to predict the likelihood that an employee of Guilford County will leave their job in the near future. Assuming such a model is robust, it can be used to prioritize “stay interviews” and other interventions Guilford County’s Human Resources professionals use to improve employee retention. This project will rely on administrative data including employee pay, work history, history and other info. In order to use these data, you will have to complete a data-use agreement form, and take an online CITI certification in human subject research from the University.

You will explore these data for insightful patterns; learn the various paths of employees through the County workforce; understand the factors associated with turnover. You will need to formulate some insights into how the client operates - this is, in some ways, a management consulting style project. What is the current business process - the current turnover-avoidance detection scheme, and interventions? How are these interventions deployed? How would the client like to deploy them? What are the costs associated with turnover?

Your job is to create a model that can predict the probability of turnover in the employee pool. With a robust model in hand, you will then develop a web-based data visualization tool to enable HR to make decisions about resource allocation and interventions to prevent turnover.

Project Management - Designate one person to take the lead on the dashboard, one person to lead on the markdown, one person to do field research on organizational and management approaches to modeling and intervening on employee turnover (you can start with this research review).

Data wrangling - The extracts you will be using in this project come from longitudinal data systems used by the client to log events in employee work history. You will also have data with attributes about those employees and data dictionaries. The first order of business is to characterize turnover - how do you detect it in your data? Is it just an employee leaving or being terminated? You will also need to figure out how to construct and parameterize work histories, events, and employee characteristics in order to inform a useful, accurate, generalizable model of turnover.

Data sets: There are several data sets which will comprise the basis for your project. (These are sensitive data subject to the highest human subject research protocols and restrictions on storage and privacy in representation. Misuse of these data will result in severe consequences - read the Data Security section of this scope for more.)

Personnel actions: Data from Lawson (HR systems vendor) which contains records of various "actions" affecting tens of thousands of current and historical employees at the County. These actions include hiring, firing, changes in supervisor, pay status, and much, much more. This is the key data set from which you will be able to create your dependent variable, and paramaterize your employee history. (Not in hand as of 1/10/2021)

Employee characteristics: There are also data about the characteristics of employees (who have unique ID numbers to link data sets together) which may serve as useful parameters. These data sets come with detailed data dictionaries that explain the numerous variables. EEO information (Equal Employment Opportunity) has information about employee demographics, veteran status, addresses, and other characteristics that might be of use in a model. Spatially - you might consider a hypothesis associated with the relationship between turnover and the distances between homes and employment locations, or employee moves – but that would require geocoding.

Pay history: A record of pay rates for County employees, which a row associated with each rate of pay and its effective date. A single employee will have multiple rows if they have been given multiple raises.

Job code history: A history of job codes (e.g. titles, ranks) associated with a particular employee. Pay rates reflect the last pay rate known for each employee - you will need to refer to Pay history for more accurate timeseries info.

Halogen performance reviews: Information on employee performance which can be text mined for feature engineering. (Not in hand as of 1/10/2021)

Exploratory Analysis - One of the central challenges to this project will be using data to tell the story of employment at Guilford County. Explore and visualize the process of employees into, through, and out of the HR system. Understand the nature and volume of "turnover" and explore the factors which correlate with turnover. How has turnover changed through time? Does the story vary amongst groups? Are there indicators of a likely candidate for turnover?

Learn about the “business process” of intervention in the turnover issue. You should be prepared to ask some of these questions in your kickoff interview, but afterwards, out of respect for the client’s time, your communication with the client should be either a) mediated through your professors or b) in a manner proposed by the client.

Modeling: Your model should be able to predict the likelihood that somebody in Guilford County employment is going to turnover. The nuances of this model will be a product of your ability to understand the use case - what is a time frame that is relevant to the client’s business process? How far in advance do they want to intervene with “stay interviews” or incentives? What is the business-as-usual approach that your model has to outperform? Does the model generalize to different employee types or demographic profiles?

Your app - You will build an app that has a dashboard of estimated turnover risk. Beyond that, the nature of your app will reflect your understanding of the way in which the client’s informational needs are best served pursuant to your deeper analysis of their desired process. Keep in mind that there is a likely cost/benefit case to be made for interventions in this space.

Data security

The data for this project are highly sensitive. We will be following medical-level data protocols. Your participation in the project will require the following data security measures:

1. Sign a data use agreement with the School and the Client.
2. Take the CITI biomedical human subject research certification course and exam (takes 2-3 hours)
3. Access the data only using PennBox API and the boxr R package and refrain from saving raw data on your personal computer
4. Consult your instructors about the level of aggregation appropriate for presenting raw or point-level data in public deliverables
5. Do not use any real names, addresses, employee records, or personal characteristics in published materials - instead use demonstration examples with simulated data.