



STUDYDADDY

**Get Homework Help
From Expert Tutor**

Get Help

A.B. Clark
ALY 6100, Data-Driven Decision Making
Week 2 Assignment: Data Collection

From: abc@huskey.neu.edu
To: c.moore-kochlacs@northeastern.edu
Subject: Project Oxygen: Data Analysis

Caroline,

Well, getting the data for our first phase of analysis was a bit more complicated than we hoped, but we will have enough to start analyses by the beginning of next week. The focus in for the next two weeks will be getting the data organized, calculating descriptive distributions, and beginning to build a model of employee turnover.

First though, I wanted to confirm what we've decided re employee privacy. Given the detailed nature of the employee information we have, we proactively reached out to HR for training and establish confidentiality expectations. We agreed on the following:

- The project team is limited to 4 analysts. Each has signed non-disclosure paper about the data in this project.
- The data collected is being stored on an HR server and will only be accessible via a set of computers in a locked room.
- Managers are identified in the data by their IDs. While there is data by which managers may be directly identified, e.g. department and number of direct reports, we wanted to minimize our knowledge of the names behind the data we're working with. This will both reduce the likelihood of violations of employee privacy as well as minimize the potential for personal biases in our analyses.

In later work, deanonymization will be necessary, and we will continue to work with HR to be respectful of employee privacy and limit the number of people who see private information.

To start off the analysis, here is what I propose:

(1) Data organization. We need to join together all the data we have about individual managers.

For each manager ID, we will combine datasets so that we have:

1. Position and department history
2. Performance review history
 - Performance and satisfaction ratings from direct reports
 - Performance and satisfaction ratings from their managers
 - Other ratings, including self-review
3. Team history (for the teams they've managed)
 - Team performance history

- Team turnover history
 - Team member promotions
 - Team member lateral internal moves
 - Team members who left
 - Exit interview ratings of manager
 - Exit interview comments re manager (see below)
 - Exit interview transcript comments re manager

(1b) Exit interview comments re manager. For the moment, we'd like to take the text from the exit interview transcripts and turn it into a quantitative measure. Specifically:

- -1 = No mention of manager
 - -1 will be assigned to all exit interviews in which keyword search did not turn up OR upon inspection of interview transcript, keyword search turned up a false positive
- 0 = Comments about the manager are positive or neutral
- 1 = Comments about manager indicate manager had some role in employee leaving
- 2 = Comments about manager indicate manager had primary role in employee leaving

To keep exit interview information private, HR will have their employee retention specialist Marie do this inspection.

(2) Descriptive statistics and distributions

To start, we want to gather some descriptive statistics and distributions from the data set in **(1)**. Specifically, for each department we want to measure:

1. Exit interviews. Are exiting employees citing managers as why they are leaving?
 - a. What is the distribution of exiting employees in the (1b) categories?
2. Performance reviews. How well/not well are managers currently being rated?
 - a. What are the distributions of manager's performance and satisfaction ratings from direct reports?
 - b. What are the distributions of manager's performance and satisfaction ratings from their managers?

These data will give us a good sense of current employee perception of managers.

(3) Do low performing managers have higher team turnover? (Gavin, 2013)

The first question we will ask with a statistical model is whether low performing managers have a significantly higher team turnover rate than high performing managers. We will use the following definitions:

- Low performing manager: top quartile of managers, based historical average of performance + satisfaction rankings
- High performing manager: top quartile of managers, based historical average of performance + satisfaction rankings

- Turnover: Average % employees on a manager's team that leave Google or move laterally internally

Our covariates will be:

- Manager performance (high/low)
- Department
- Employee level

We will look at the significance and explained variance of manager performance, to answer the question. (We expect to build more complicated models in the longer term, with this a starting point.)

Let me know what you think!

Best,

A.B.



STUDYDADDY

**Get Homework Help
From Expert Tutor**

Get Help