Project Proposal
Identifying Key Variables to Explain Employee Performance

MGT 6203

Fall 2024

10/3/2024

Ben Mascott (Analysis Co-lead), Erick Ordonez (Project Manager),
Jerry Fang (Data Visualization Lead), Isabella Reyna (Analysis Co-lead)

Georgia Tech

Problem Statement

Business Problem:
We've been asked to use analytics to help the HR (Human Resources) department strategically allocate resources to support existing lower-performing employees. We are aiming to determine which factors, such as years of experience, job satisfaction, education level, and commute time, significantly impact employee performance. By controlling for variables such as education level and department, we can determine predictors that best explain the variability in employee performance. This analysis will help HR improve performance among existing employees who are currently under-performing.

Context and Importance:
This problem is crucial as it directly impacts the existing employee performance. Understanding the variables that are most correlated with performance across different groups in the data will help the business save time and money.

Industry Impact:
Identifying key predictors of performance, such as years of experience, job satisfaction, and education levels, helps HR to grow existing talent within the company. We're specifically helping the HR industry to efficiently utilize resources in a quantitative manner that bolsters decision-making with analytics.

Georgia Tech

Objectives and Goals

Project Objectives:
- Determine which factors, such as years of experience, job satisfaction, education level, etc., significantly impact employee performance to help HR tailor support and intervention efforts.
- Develop strategies for HR to allocate their limited time and budget effectively to support and improve the performance of lower-performing employees.

Expected Outcomes:
- A clear understanding of the key factors that explain employee performance will help HR strategically allocate resources to target support at underperforming employees, which ultimately *save time and money*.
- Obtain insights on how various factors affect employee performance, which will help HR to implement *tailored interventions* based on employee groups or job roles within the company.
- We expect those to lead to *improved employee performance* and satisfaction, which will contribute to *business growth* and *operational efficiency*.

Georgia Tech

Data Sources

Data Overview

- The Employee Performance Dataset, a pre-approved dataset, will be used for this project.
- Dataset Characteristics:
  - **10,000** total observations
  - **13** predictor variables, **1** response variable
  - **Qualitative Predictors:** Education_Level, Department, Job_Satistfaction
  - **Quantitative Predictors:** Age, Salary, Experience, Work_Life_Balance, Annual_Bonus, Commute_Time, Work_Hours_Per_Week, Promotions_Last_5_Years, Training_Hours_Last_Year, Overtime_Hours_Per_Week
  - **Response Variable:** Performance_Score (quantitative)

Data Relevance

- For of this project, each observation in the dataset will represent a data point collected from a current employee at our company.
- To assist our company with supporting employees that are underperforming, we will identify which factors best predict low employee performance.

Georgia Tech

Methodology

1: Data Exploration

- Visualize data, identify categorical and quantitative predictors
- Preliminarily assess **multicollinearity** using a **correlation matrix** to identify if any predictors are correlated with each other

2: Data Preparation

- Identify and **factorize** categorical variables
- Compute the **correlation coefficient and R-squared** between the response and each predictor
- **Transform** predictors that violate the linearity assumption using the boxcox method and trial and error

3: Simple Linear Regression and ANOVA

- Create a **simple linear regression model for each quantitative predictor** and use a **t-test** to determine the statistical significance of each predictor when it _alone_ is used to predict Employee Performance Score.
- Create an **ANOVA model for each qualitative predictor**. Then for each ANOVA, use an **F-test** to identify if there are unequal group means for any of the qualitative variables.

4: Full Regression Model (uses all Predictors)

- Use an **F-test** to assess overall significance of the model
- **Residual Analysis** to assess the model assumptions (linearity, constant variance, normality, independence of errors)
- Calculate **Variance Inflation Factors** for each predictor in the presence of all other predictors to assess multicollinearity
- Identify outliers using **Cook's Distance** test

5: Multiple Linear Regression Models with Subsets of Predictors
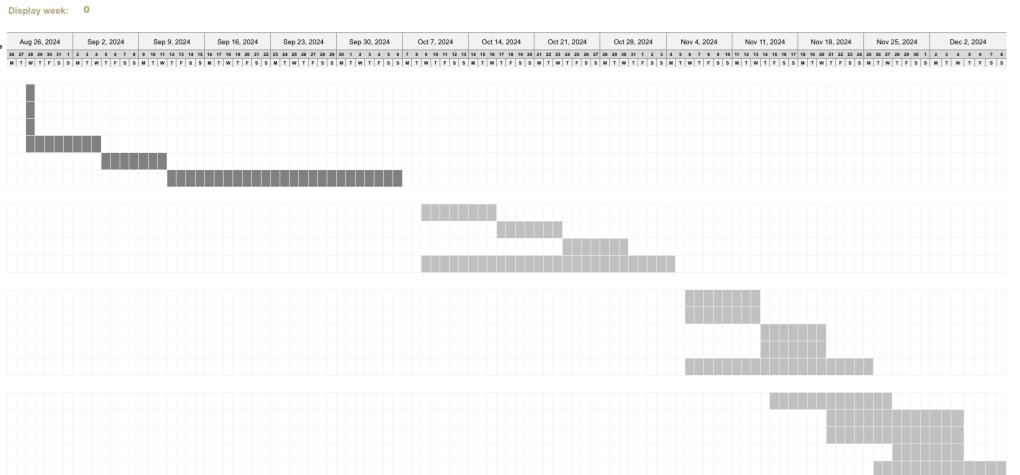
- Partial F-tests will be used to compare regression models with different subsets of predictors
- Create Subset models for different combinations of controlling variables (such as department or education level)

Expected Challenges

Potential Obstacles	Potential Solutions
Multicollinearity. Correlation between independent variables could impact the statistical significance of some predictors. For example, there may be a negative linear trend between Work_Life_Balance and Overtime_Hours_Per_Week.	By utilizing VIF to detect multicollinearity, we can identify if multicollinearity is present and use a variable selection method to remove highly correlated predictor variables.
Bias selection. For example, a more ambitious employee may sign up for more overtime, but if HR incentivizes low performing employees to work more, we may not observe improved performance.	By developing multiple regression models, we can control for variables like overtime to minimize bias.
Group differences. For example, different factors may explain performance in the Engineering department compared to employees in the Finance department.	By performing ANOVA F-tests with qualitative predictors, we can determine if there are differences in group means that will need further analysis

Georgia Tech

Value Proposition

• Potential Business Value

- **Improve Employee Retention** – By figuring out what goes into having a higher satisfaction and performance , human resources can address key issues that'll improve employee retention rates at the company.

- **Increased Employee Productivity** – A more productive and satisfied company can increase overall profitability and long-term growth.

- **Reduced Hiring Cost** – If human resources can predict what goes into a high employee performance and increase retention rates, they'll find that they'll spend less in recruitment in training costs.

• Stakeholder Benefit

- **Company Management** – With a more data-backed hiring process, management can benefit from a higher employee retention, satisfaction, and productivity, leading to a higher-performing organization as a whole.

- **Employees** – Addressing the factors of commute time and satisfaction will lead to a more engaged and satisfied company, contributing to people staying longer and improving overall productivity.

- **Customers** – A more productive company will ensure higher-quality products or services, strengthening the customer experience and loyalty.

Team Roles and Responsibilities

Name	Role	Responsibilities	Skills / Expertise
Erick Ordonez	Project Manager	- Create Project plan and timeline - Ensure team is on schedule and stays on track throughout the semester	- Planning - SQL - Python
Benjamin Mascott	Analysis Co-lead, Videographer	- Direct and Post Final Presentation YouTube Video - Lead Data-Preprocessing and post-regression V&V efforts	- Videographer, 1 year of SQL, data analysis and visualization
Isabella Reyna	Analysis Co-lead, Deliverable Submitter	- Submit deliverables - Support github account - Lead Regression Analysis	- 1 year of R programming experience, 7 years experience programming for data science applications
Jerry Fang	Data Visualization Lead	- Create plots and graphs for data exploration and visualization of results	-3+ years of R and Python (some SQL) programming experience, data analysis and visualization

All team members will support report writing and analysis tasks

Georgia Tech.