Bonertz, Brian

TM Attendance Analysis

Project title – Human Resource Dataset Review

Statistical/Hypothetical Question

The dataset I selected was Human Resource data. I chose this dataset because HR data doesn’t seem to be examined closely or often and I think there is a lot of information a company can use to improve work force and output. When we do see analysis of this type of data, we often look at comparison in salary between different groups. My initial analysis was headed down the same path, however, the more I dug into the dataset, I decided to try to use the data to determine if there were certain variables or factors that may contribute or predict team members that have higher absence or poor attendance.

Therefore, my statistical / hypothetical question is: ‘Are there variables or attributes in employee data that influence team member attendance either positively or negatively?’

Outcome of your EDA

* The outcome of my EDA was inconclusive. With the fictional HR dataset I explored, I did not find any attributes that would affect a team members attendance.

What do you feel was missed during the analysis?

* There were several variables that have still yet to be explored. I think a deeper dive into different departments, or type of work may have an effect. Perhaps certain positions take a different toll on a team members health or stressors on the body that can lead to absence. Also, a closer look at supervisors or managers of these departments can contribute. Are managers taking the right approach to correcting absence or are team members taking advantage?

Were there any variables you felt could have helped in the analysis?

* I am not sure how additional variables would help the dataset or analysis. Maybe a better idea or description of the type of work the team members were doing or perhaps the employees work history would help?

Were there any assumptions made you felt were incorrect?

* I was a bit surprised that the employee satisfaction survey didn’t have a better correlation with absences. It seemed that team members that were completely satisfied with their job were absent just as much as team members who were not. Also, salary didn’t seem to have any affect on absence, which sounds a bit odd to me. I would like to see how these two variables compare in a real life dataset.

What challenges did you face, what did you not fully understand?

* I don’t know if I had too much challenge with the project overall, however, my biggest challenge in the class was trying to understand the flow of these functions. It took me a while to catch on to the operation of these functions, but the book explained a lot and other resources on the internet helped.