Final Project

Final Project

At-risk behaviors

While performing EDA upon my At-risk behaviors data set, my statistical question was:

* What shift do most at-risk behaviors occur and are employees with less than a year employment the most at-risk?

My hypothesis that I am trying to prove my statistical question as true is:

* Out of the shifts (Morning, Sunrise, Day, Twilight, & Night), Sunrise will have the most at-risk behaviors (Push/pull, Overreach, Powerzone, Struck by/against, Load stand, Lifting/lowering, Hand-to-surface, Yard control, Egress, Slips/trips/falls, Cornerstone, Firm footing, 3 p.o.c., F.O.D., Hydration) and employees with less than a year of employment will be less at-risk than employees with more than a year of employment.

My conclusion of the EDA I performed upon my At-risk behavior data set was:

* Even though my hypothesis was correct that most at-risk behaviors occur on the sunrise shift and employees under a year of employment as less likely to be more at-risk compared to employees over a year of employment, there is still bias and indeterminants. Basically, the reason why the sunrise shift had more at-risk behaviors compared to others is because that is the shift that has the Safety Committee in it. Also, there is more employees observed who are over a year of employment compared to employees less than a year employment, thus, they are considered complacent towards their job function and therefore, tend to perform more at-risk behaviors.

Since all of my variables in my At-risk behavior data set was categorical, I missed out in operating Python to handle descriptive characteristics, pmf, cdf, covariance, Pearson’s correlation, hypothesis testing, and regression without having to do it manually or draw conclusions beforehand.

I am not sure what other variables could have helped with my analysis, but possibly the ‘Week’ variable could offer some more insight in regard to at-risk behavior within a work-week.

My main assumption I probably got incorrect is my regression. Due to time constraints and conclusions drawn from my hypothesis testing, I believe a regression analysis was unable to be performed upon my At-risk behavior data set. However, given more time, maybe a regression analysis could be exploited.

My main challenges was getting the plots to show up after executing my code. After that though, my EDA was difficult to complete, but I believe I presented the best outcome of my At-risk behavior data set.