Tata Steel Project

4th June '19 -2nd July '19
By:- Anubhav Kumar Singh

ACKNOWLEDGEMENT

I am using this opportunity to express my deepest gratitude and special thanks to my project guide and mentor, Mr. Deepak Sharma who took time out to hear, guide and keep me on the correct path and allowing me to carry out my project at their esteemed organization and extending during the training.

GOALS

To study the dataset and visualize various patterns.

INTRODUCTION

The dataset provided to me was from the UCI Machine Learning repository named 'Absenteeism at work Data set'. The given dataset had various columns linking to time of absence such as 'ID', 'Reason for absence', 'Month of absence', 'Day of the week', 'Seasons', 'Transportation expense', 'Distance from Residence to Work','Service time', 'Age', 'Workload Average/day ', 'Hit target', 'Disciplinary failure', 'Education', 'Son', 'Social drinker', 'Social smoker', 'Pet', 'Weight', 'Height', 'Body mass index'.

MILESTONES

On studying the dataset it was found that most people gave the reason for absence as 'Medical Consultation'. The mean time for absence was 6.92 hours and only 0.05% of the total of 740 people were never absent. People living in the range of 12 to 18km were absent for the most time. The mean age of people working there was 36.5 years and people above 53 yrs of age were absent for the most time.

Expense/Distance ratio — Son Variation(Mean)
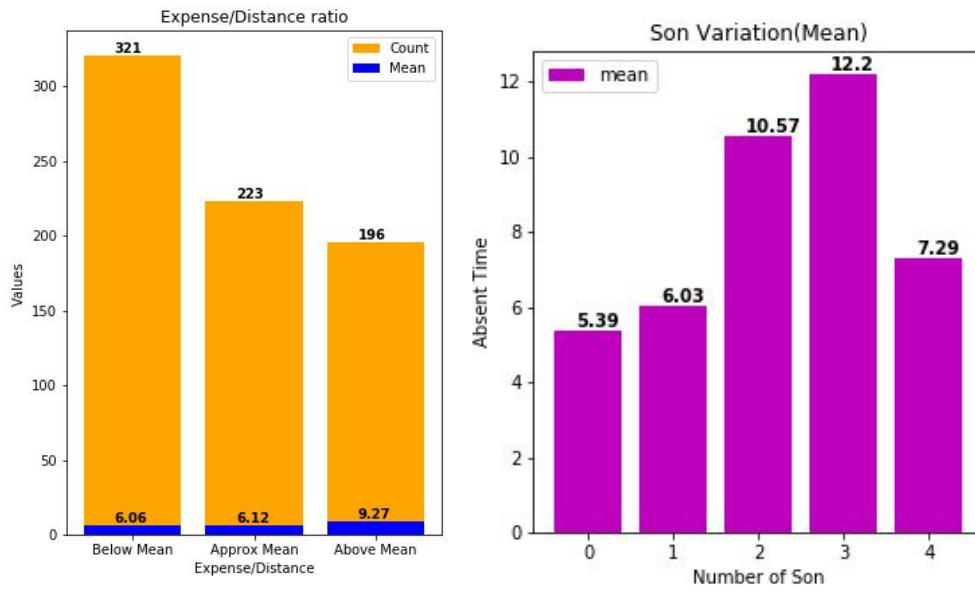
- From the graph, it is visible that people with Sons more than 2 were absent for most amount of time.
- The people who stated the Expense/Distance ratio more were the people who were absent for more amount of time.


Variation due to month
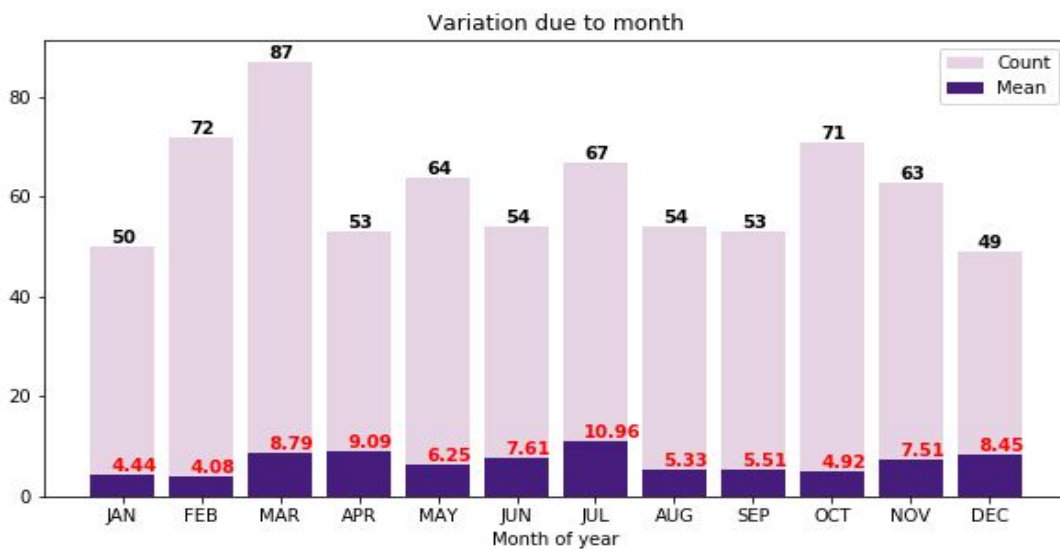
- It is visible that most people were absent in the month of March but the mean time of absence is greater in the month of July.
- Monday saw the most number of absentees with Winter being the laziest month for the people with the most time of absence.

Days of week variation
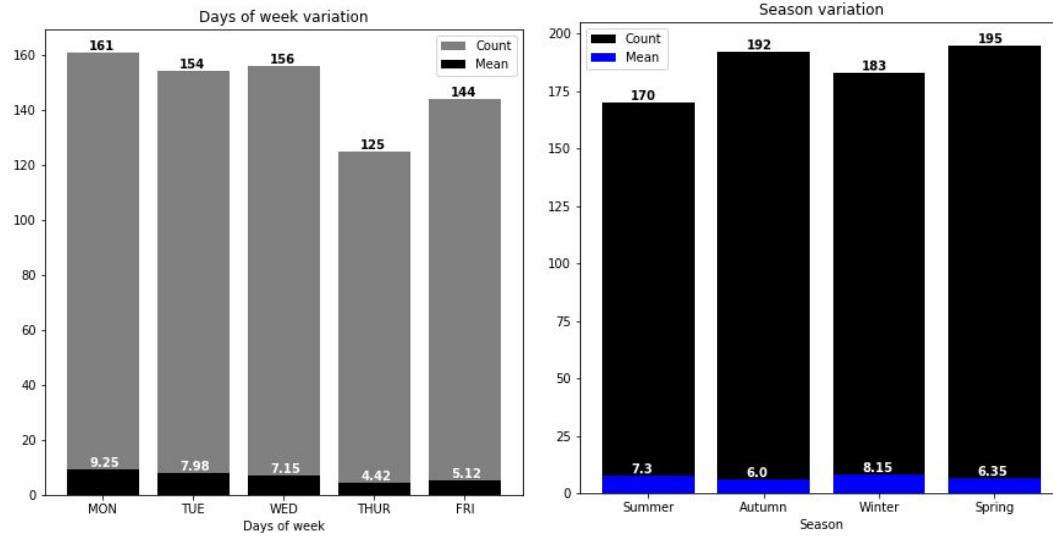
Season variation

-
- BMI variation shows that people who are overweight are more absent at work.
- We can see that transportation expense is bigger in winter months than in other seasons.

BMI variation
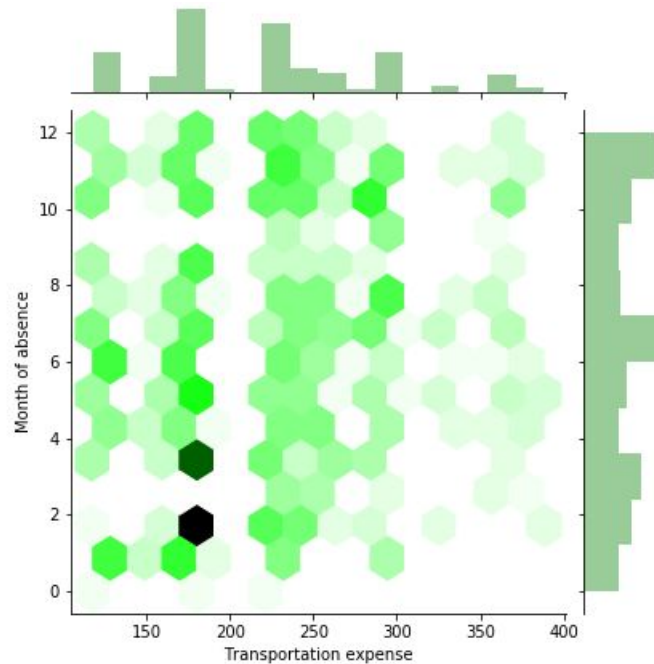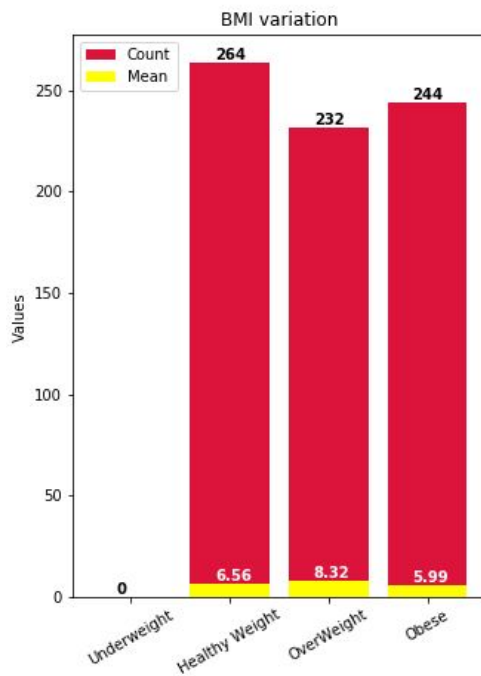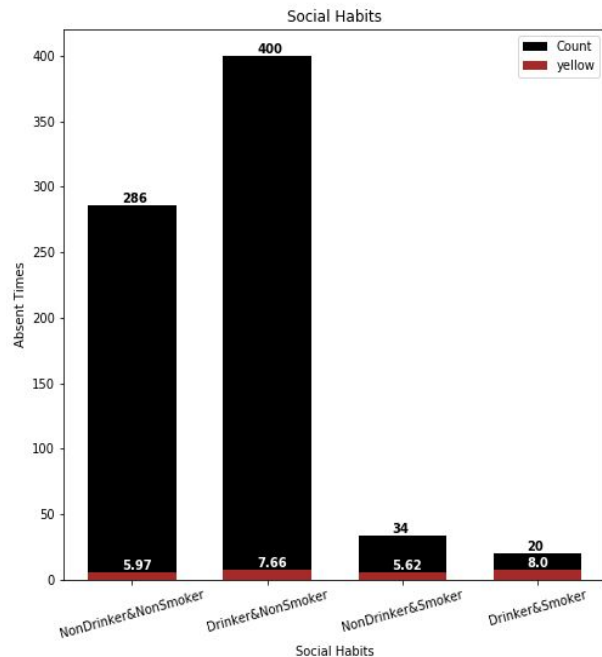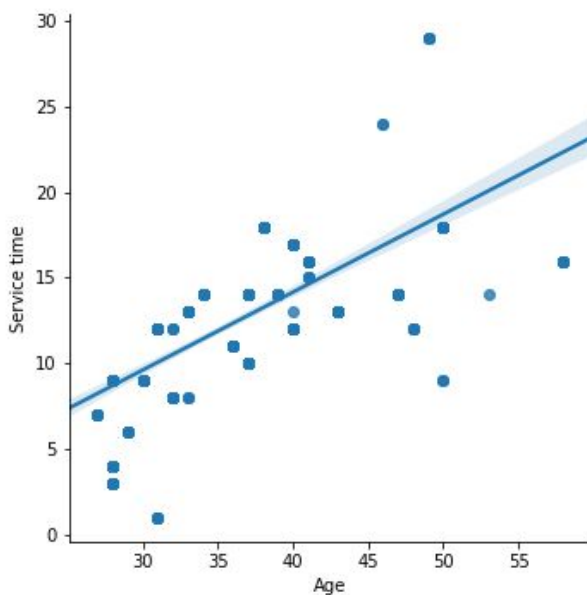
How does this help TataSteel?

This would help TataSteel in maintaining a good employer-employee relation. The company would get to know about the various reasons why the employees are absent. For example, the social habits of the employees will help the company create a better work environment for all the other employees. The various factors that define the social status such as number of pet, number of sons, travel expense will help the company know more about the reason for the absence of its employees. Distance from home to work accounts as a major reason for the increased absence time of the employees. A correlation between Service time and age as seen in the dataset and various other attributes of the dataset accounts for a correlation that can be used to create and predict the model.

References

https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work

Project Link

https://github.com/Anubhavkumar45/Absenteeism-at-work