Devyang Sikotariya

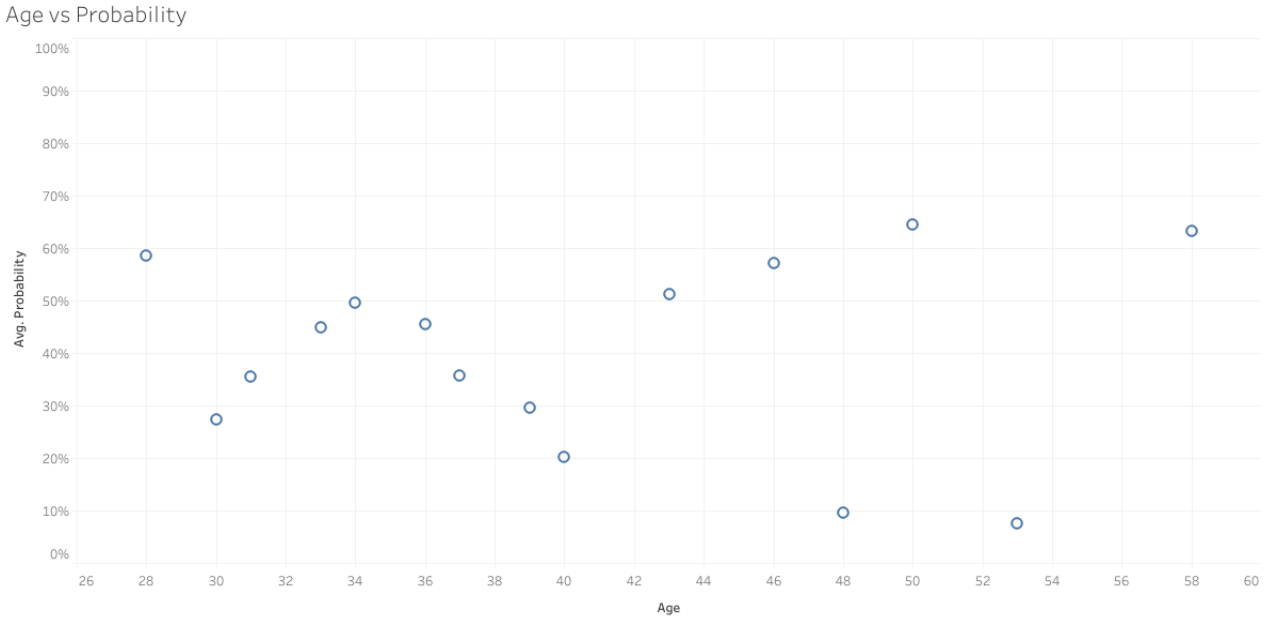
Business Intelligence Report

24 March 2021

Report on Absenteeism of Employees

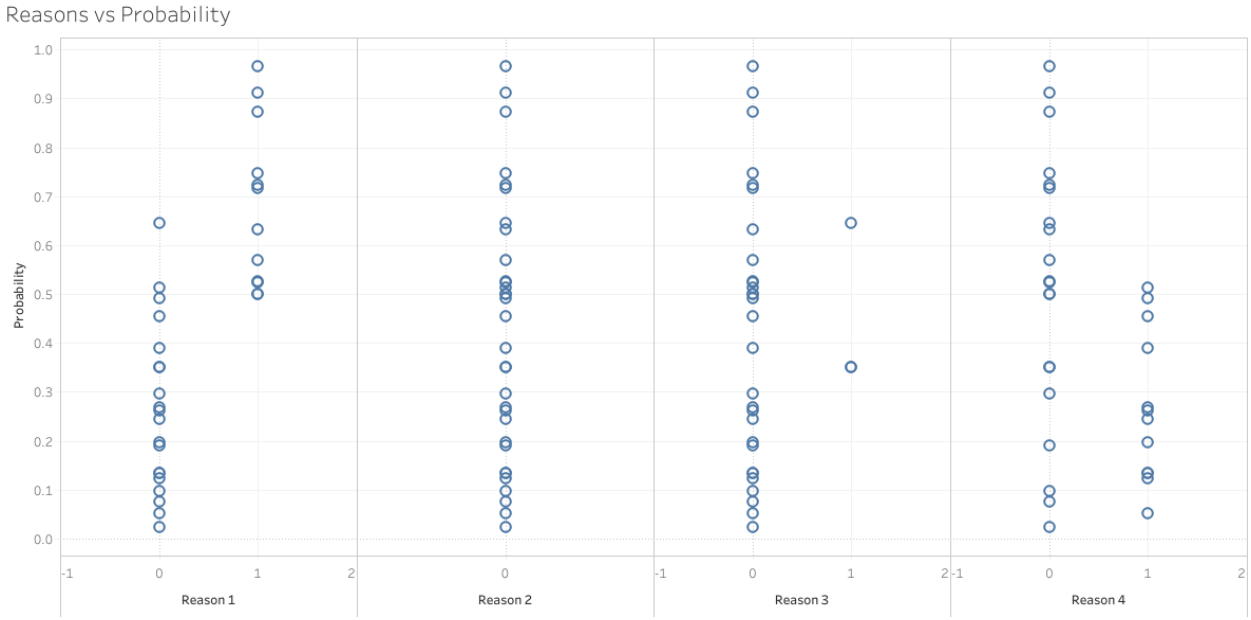
During the modelling(Logistic Regression), the employees who are absent for more than 3 hours have been categorized as 'Excessively Absent'. Data of the past two years was available (secondary data), and accordingly the data was divided into train (0.8) and test (0.2) datasets. The Model gave accuracy around 75% to 80%. The following graphs have been developed based on entirely new data which the model had never either been trained or tested on.

Age vs. Probability



There is only a 29% chance that an individual who is 31 will be absent from work for more than 3 hours. While the chances of people who are in their 40's have more than 50% to be excessively absent.

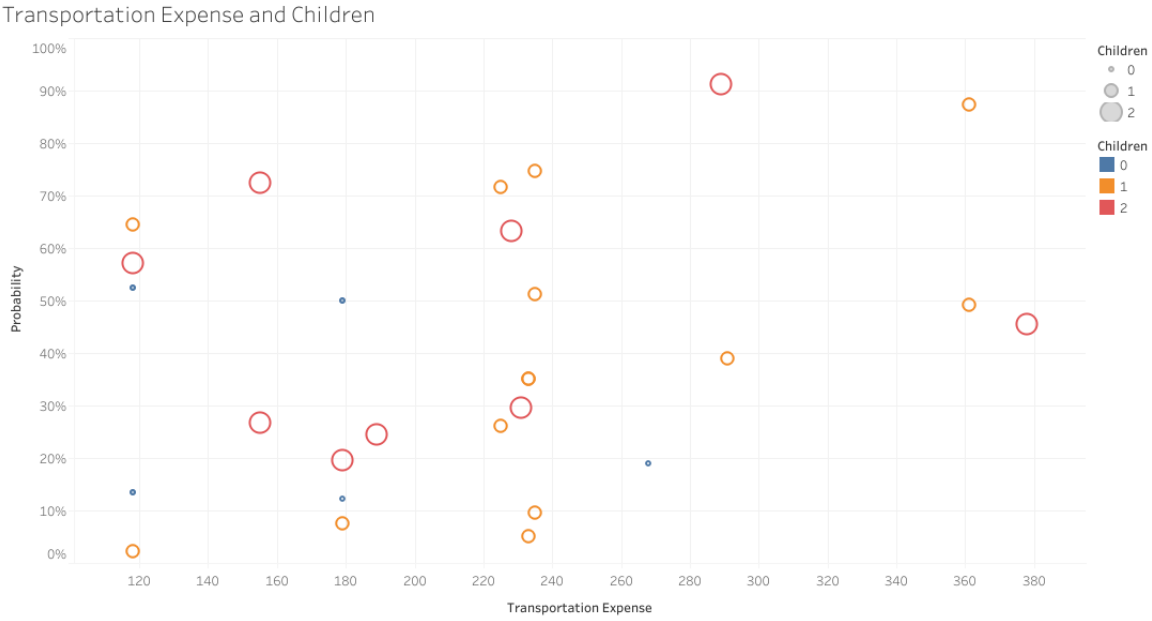
Reasons vs. Probability



Observing one or zero along the horizontal axis indicates whether an individual has been or has not been absent for each particular reason. From the graph its clear that all the reasons contain the values 0 and 1, except for 'Reason 2'. This means that none of the employees from the new data (40 observations) has been away from work because of 'Reason 2'. Consequently, from the graph, it can be deduced that, if a person is supposed to be absent due to a 'Reason 1' from group 1 as marked with dots associated with value of 1 he probability that he/she will be excessively absent, is above 50%. Hence such an individual is expected to be away from work for more than three hours. Whereas, looking at the 'Reason 3', very few people who had specified this reason for excessive absence, as well as the fact that the probabilities have been distributed both in the lower and upper parts of the vertical line (50%). Similarly to 'Reason 2' this class does not tell much about what to expect from the individuals being absent because of the 'Reason 3'. Finally, 'Reason 4' leads to an opposite conclusion compared to 'Reason 1'. The people who have been selected a reason for absence from this class exhibit a probability below 50% to be excessively absent. 'Reason 1' represents very serious diseases. That's why the numbers tell that the expected probability of an individual to be excessively absent because of the reason from this class is

higher than 50%. ‘Reason 4’ represents light reasons for absence like dental appointment, physiotherapy, a medical consultation and others, which concludes that this is not serious and the employee does not need to be absent for the entire day, and as it can be seen from the graph that their probabilities are below 50%.

Transportation Expense and Children



- Loosely, there is a positive correlation between transportation expense and absence. Perhaps if there were more observations, could have solidified this observation by saying, higher the transportation costs higher the chances they will be away from work.
- People with no children dont exhibit a high probability for excessive absence. Also, apart from one small exception none of them have high transportation costs.
- The probability of people having one child is expected to be excessively absent varies across observations. However, most of the data is clustered around the average transportation expenses spending between 220$ and 240$ per month.
- People with 2 children shows the same similarity with the people bearing one child.

- Thus the data is about people who dont generally spend more than $240 on transportation. Only 5 observations lay beyond the average spending line.