**Predicting Factors That Impact Absenteeism at Work**

Shreya Prabhu(shreyapr) & Audrey Liu(yalinl)

[**https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work#**](https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work#)

Having worked at smaller companies in the past, our experiences have that companies care a lot about the absenteeism rate in the workplace. Some organizations have entire units developed to just keep track of absent employees and the loss undertaken because of this absenteeism. Clearly, this information is valuable to organizations trying to account for their employees’ productivity and for charting out future growth of the company.

**Motivation:**

The motivation for this project comes from our experience working at various start-ups and how absenteeism on the part of our peers would impact our own work and the final output. It is a well-known fact that absenteeism impacts the final solution that any organization is able to develop. We are cognizant of the idea of the requirement of well-intentioned breaks from work and the essence and need for them. However, when absenteeism is unaccounted for, it causes delays in project deliveries and multitudes of unforeseen losses. Most commonly known reasons for being absent at work is sickness. However, we feel there are other underlying features that could be used as predictors of absenteeism, such as age, distance from work, hitting your target etc. If an organization can access these underlying predictors for people being absent at work, they would be empowered to change business processes and company cultures to suit their employee’s needs.

**Related Work:**

The impact of absenteeism is not trivial. According to an article on Spark, co-workers are 30% less productive with the absence of their colleagues and 47% of overtime is used because of decreasing productivity due to absenteeism. When an employee is absent, not only is his/her own work suspended, the team work is usually deterred as well. The overall productivity of an organization can be affected, especially for smaller institutions.

[**https://www.adp.com/spark/articles/2017/01/the-impact-of-absenteeism.aspx**](https://www.adp.com/spark/articles/2017/01/the-impact-of-absenteeism.aspx)

An article published by Forbes in 2013 states that the top causes of absenteeism are bullying and harassment, burnout and stress, family care, depression, disengagement, and illness and injuries. Apparently, many “sick leave” weren’t used for sickness, at least not physical sickness.

<https://www.forbes.com/sites/investopedia/2013/07/10/the-causes-and-costs-of-absenteeism-in-the-workplace/#978303b3eb65>

Thus far, many tips for company to reduce absenteeism suggests offering mental support, reducing workplace stress, providing effective feedback, and rewarding good attendance. We wonder if we can quantify these tips with our dataset. For example, what is the average threshold of workload for an employee to feel burnout, how will hitting a target or positive affirmation of work progress help employee to stay motivated.

<https://www.highspeedtraining.co.uk/hub/reducing-absenteeism-in-the-workplace/>

**Data:**

Our data is pulled from UCI machine learning repository. The data records employees working in a courier company in Brazil from July 2007 to July 2010. There are 740 unique entries of employees who were granted a sick leave. 20 possible predictors were recorded for each leave and some of them are pre-encoded as digits:

1. Reason for absence: most sick leave were granted upon an attested medical proofs or other medical conditions, with one exception of “unjustified absence” (encoded as 25), which we will take a look at in our analysis.
2. Month of absence: 1 to 12 for January to December
3. Day of the week: 2 to 6 for Monday to Friday
4. Season: 1 to 4 for Summer, Autumn, Winter, Spring respectively
5. Transportation expense: in Brazalian dollar
6. Distance from Residence to Work : in km
7. Service time: in years of how long the employee has been with the company
8. Age: in years
9. Work load Average/day:
10. Hit target: in percentage, i.e. 97% hitting the target
11. Disciplinary failure: yes=1; no=0
12. Education: high school (1), graduate (2), postgraduate (3), master and doctor (4)
13. Son: number of children
14. Social drinker: yes=1; no=0
15. Social smoker: yes=1; no=0
16. Pet: number of pets
17. Weight: in kg
18. Height: in cm
19. Body mass index
20. Absenteeism time: in hours

There are plenty of projects that are done based on this set of data. Some focus on whether seasonality played a role in employee’s absenteeism. Some focus on age and service time. We want to focus on the company-controllable factors such as transportation expense, distance from work, work load, hitting target, and disciplinary failure. We believe evaluating other factors such as social drinker or number of sons would not render actionable policies for the company, and may even result in unwarranted discrimiation against certain groups of people.

**Questions:**

We have learned from related work that burnout, stress, low morale and disengagement are top factors that discourage employees from committing to work. Certainly these factors can be addressed with a more relaxed working experience.

The main question we aim to ask in relation to this dataset, using a machine learning methodology, is given certain working environment related information about an employee, how likely are they to be absent at work.

Question 1: if providing accommodations or commute stipend would reduce employee's pressure of going to work

Question 2: how far off from the intended work target would the absenteeism set in and how long does the employee need to revert back to work

Question 3: how much workload on average would burn out an employee and causes absenteeism

Question 4: among all of these working environment factors , which one results in the longest leave and which one results in the shortest leave.

All questions will be studied along with the factor of age. We believe age is an important factor accompanying other working environment related factors. Older people are usually better at dealing with stress and are more experiences with their work. However, older people are also usually less motivated and spare less focus on work. Therefore, it is important to see how age play a role in our analysis.

If an organization has this analysis handy, they would be in a better position to implement business process and cultural changes accordingly.

**Possible Findings and Implications:**

We anticipate ‘Distance from Work’ to be a big factor contributing to absenteeism from work and people staying farther to be more absent than others on average. We also expect a higher age to be an indicator of more absenteeism, since the more settled you would be at your job, the less need you might feel to show and prove yourself every single day. We expect high workload and hitting target to result in a shorter leave because we assume people on average are responsible and are willing to go back to work. We also expect disciplinary failure to result in a longer leave because a disciplinary failure implies potential judgments from colleagues amd biased treatment from supervisors, all of which aggravate an employee’s stress.