

## **Project Proposal:**

- - **Project Topic:**  
Absenteeism at Work:
    - <https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work>
  - **Business Value and Questions:**
    - Every workplace has issues with absenteeism. While employers expect employees to miss a certain amount of workdays - excessive absenteeism can lead to a decrease in productivity, finances, and morale.
    - We plan on finding out what causes absenteeism, the cost of absenteeism, and how might employers reduce absenteeism.
  - **Target Variable:**
    - Our target variable is Absenteeism in Hours - We will be able to calculate the number of hours each employee has missed by summing their ID #s.
  - **Dataset-Rows and Columns:**
    - Rows 740
    - Columns 21

## **Data Dictionary**

Serial No#	Attribute	Possible values	Data Type	Attribute Information
1	ID		Nominal	ID corresponds to the Identification Codes for individual employees at the Courier Company
2	Reason for absence	28 unique values	Categorical / Nominal	This attribute represents the medical reasons given by employees while taking leaves. The concerned courier company recognizes 28 reasons for absence, out of which 21 are attested by the International Code of Diseases (ICD) and 7 are categories without ICD attestation.
3	Month of absence	1 to 12	Categorical / Nominal	Corresponds to the months of a year, with 1 representing January; 2 representing February and so on. Each row relative to this variable represents the month when an employee was absent from work.
4	Day of the week	(Monday (2), Tuesday (3), Wednesday (4), Thursday (5), Friday (6))	Categorical / Nominal	This column describes days when individual employees were absent, with each instance (corresponding row) representing a distinct day.
5	Seasons	(summer (1), autumn (2), winter (3), spring (4))	Categorical / Nominal	This column represents seasons.
6	Transportation expense		Continuous	Continuous variable describing the transportation expenses.

7	Distance from Residence to Work		Continuous	Continuous variable describing the distance from residence to workplace of individual employees in Kilometers.
8	Service time		Continuous	The documented duration of service in the week when the concerned employee took leave from work.
9	Age		Continuous	Age of the employee.
10	Workload Average/day		Continuous	The average workload per day.