



Project Brief **Clustering - Employee Segmentation on Absenteeism**

Employee Absenteeism Clustering

Employees across the world are entitled for holidays / vacations. It has also been noticed that some employees are absent during work hours due to various reasons. This may have adverse effects to the team / department in its daily progress.

The clustering technique will help an organization to understand some reasons why employees tend to be away from work. By creating these clusters, a lot of attention can be given to address issues, take corrective actions, or implementing better working conditions, whatever may be the reason.

Business Objective

- To create an employee cluster to show absenteeism at work and try to identify the top contributors for it



Data Dictionary: AbsenteeismAtWork Data

The Employee dataset has about 700 records in a CSV format. Click [here](#) to download the datafile for this project.

S. No.	Feature	Data Type	Description
1	ID	Numeric	Unique record identification number
2	Reason for absence	Numeric	Reason for employee's absenteeism
3	Month of absence	Numeric	Month of employee's absenteeism
4	Day of the week	Numeric	Day of the week of employee's absenteeism
5	Seasons	Numeric	Season of employee's absenteeism
6	Transportation expense	Numeric	Employee's transportation expense
7	Distance from residence to work	Numeric	Employee's distance from residence to work

Data Dictionary: AbsenteeismAtWork Data (contd.)

S. No.	Feature	Data Type	Description
8	Service time	Numeric	Employee's service time
9	Age	Numeric	Employee's age
10	Workload average/day	Numeric	Employee's workload average/day
11	Hit target	Numeric	Employee's hit target in current role
12	Disciplinary failure	Numeric	Employee's disciplinary failure
13	Education	Numeric	Employee's education
14	Son	Numeric	Employee's number of sons
15	Social drinker	Numeric	Whether employee is drinker
16	Social Smoker	Numeric	Whether employee is smoker

Data Dictionary: AbsenteeismAtWork Data (contd.)

S. No.	Feature	Data Type	Description
17	Pet	Numeric	Number of pets that employee has
18	Weight	Numeric	Employee's weight
19	Height	Numeric	Employee's height
20	Body mass index	Numeric	Employee's bmi
21	Absenteeism time in hours	Numeric	Employee's absenteeism time in hours

Technical Goals

- Understand the data very well. Do all transformations / data engineering / etc. wherever applicable
- Perform Exploratory Data Analysis (EDA)
- Carry out all the Data mining tasks
- Identify the salient features that will determine the best results
- Perform the model evaluation to select the appropriate algorithms





Thank you!