

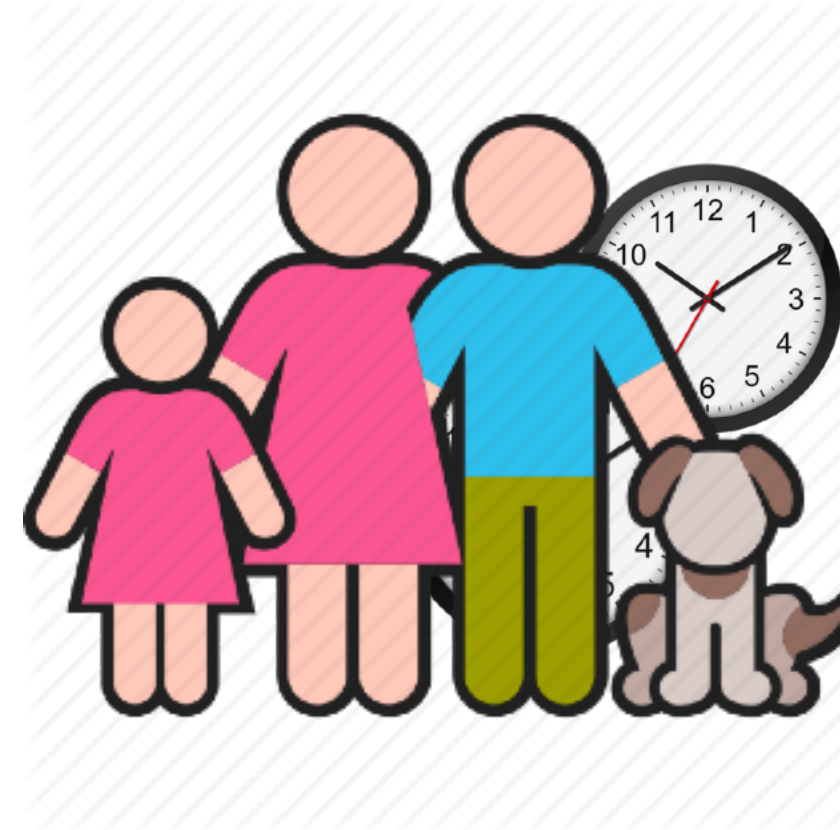
# ABSENTEEISM

Based on what information should we predict whether an employee is expected to be absent or not?

How far they live from the workplace



How many children and pets they have



Do they have higher education?



And so on!



# THE DATASET

```
In [1]: import pandas as pd

In [2]: raw_csv_data = pd.read_csv("Absenteeism-data.csv")

In [3]: type(raw_csv_data)

Out[3]: pandas.core.frame.DataFrame

In [4]: raw_csv_data
```

Out[4]:

	ID	Reason for Absence	Date	Transportation Expense	Distance to Work	Age	Daily Work Load Average	Body Mass Index	Education	Children	Pets	Absenteeism Time in Hours
0	11	26	07/07/2015	289	36	33	239.554	30	1	2	1	4
1	36	0	14/07/2015	118	13	50	239.554	31	1	1	0	0
2	3	23	15/07/2015	179	51	38	239.554	31	1	0	0	2
3	7	7	16/07/2015	279	5	39	239.554	24	1	2	0	4
4	11	23	23/07/2015	289	36	33	239.554	30	1	2	1	2
...	...	...	...	...	...	...	...	...	...	...	...	...
695	17	10	23/05/2018	179	22	40	237.656	22	2	2	0	8
696	28	6	23/05/2018	225	26	28	237.656	24	1	1	2	3
697	18	10	24/05/2018	330	16	28	237.656	25	2	0	0	8
698	25	23	24/05/2018	235	16	32	237.656	25	3	0	0	2
699	15	28	31/05/2018	291	31	40	237.656	25	1	1	1	2

700 rows x 12 columns

## Dataset:

- Absenteeism\_data.csv
- Already existing study about the prediction of absenteeism at work
- 100% application in the business world