User Requirements Doc:

Absenteeism at Work Dashboard

# Objective

To analyze the data to uncover trends in absenteeism by day of the week, month, or season and reward a healthy employee lifestyle.

# Problems identified

* Provide a list of Healthy individuals & low absenteeism for our healthy bonus program – Total budget $1000 USD.
* Calculate a wage increase or annual compensation for non-smokers for
* Insurance budget of $983,221 for all non-smokers.
* Create a dashboard for HR to understand absenteeism at work based on approved wireframe.

# Target audience

* Primary – Head of HR
* Secondary - HR team members department

# Use cases

## 1. Identify the healthy individuals & low absenteeism employees with

### User story

As the Head of HR, I want to identify the healthy individuals & low absenteeism employees for our healthy bonus program that been budgeted around $1000 USD. I also want to identify wage increase for non-smokers for insurance budget of $983,221 fro all non-smokers.

### Acceptance criteria

The dashboard should

* List the employee and categories, time and trends, reason and comparison.
* Display key metrics (absenteeism time, reason for absence, day of the week, month, and season should be included and cleaned.)
* Clear insights into how absenteeism affects **workload** and **transportation expenses**, and correlate these to absenteeism trends.
* Be user-friendly and easy to filter/sort
* Use the most recent data possible

# Information needed

Head of HR needs the healthy individuals & low absenteeism employee in the company, and the key metrics needed include:

* Average of absenteeism time in hours
* Sum of absenteeism time in hours
* Count of education
* Count of pet
* Count of social smoker
* Count of social drinker
* Average of transportation expenses
* Work load average per day

# Data needed

The dataset to produce the information we need should include the following fields

* ID (string)
* Reason of absence (tinyint)
* Month of absence (tinyint)
* Day of the week (tinyint)
* Seasons (tinyint)
* Transportation expense (tinyint)
* Work load average per day (tinyint)
* Disciplinary failure (bit)
* Education (tinyint)
* Social drinker (bit)
* Social smoker (bit)
* Pet (tinyint)
* Weight (tinyint)
* Body mass index (tinyint)
* Absenteeism time in hours (tinyint)
* Compansation (tinyint)
* Number (tinyint)
* Reason (varchar)

# Data quality checks

We need to add measures in place to confirm the dataset contains the data required without any issues – here are some of the data quality checks we need to conduct:

* Row count check
* Column count check
* Data type check
* Duplicate check
* Join table

# Additional requirements

* Document the solution and include the data sources, transformation processes and walk through on analysis conclusions
* Make source code and docs available on GitHub
* Ensure the solution is reproducible and maintainable so that it can support future updates