This exercise is part of the course The Data Analyst Course: Complete Data Analyst Bootcamp 2023 from 365Careers. The goal is to clean and preprocess the dataset so later some statistical models can be applied in order to predict “Absenteeism Time in Hours”. While the course provides solutions to the exercise, I tried to solve it on my own.

The AbsenteeismEDAandPreprocessing.ipynb contains the code I wrote to solve the exercise.

Before opening the ipynb file, please read the data-homework.pdf to see the tasks of the exercise.

The Absenteeism-data.csv is the original csv file.

The Absenteeism-data-refined.csv is the cleaned and preprocessed csv file.