

```
In [1]: # In this lab we will find the regression model for the
# given dataset to estimate the arriving time.
# the Goal is to predict the travel time (the last column) using 8 variables
# 'avg. Speed', 'Morning', 'Afternoon', 'Evening', 'weekend', 'rain', 'fog', 'distance to travel',
# requirements:
#   pandas library
#   LinearRegression from sklearn
#   dataset file "DatasetLab1.csv"

# Task-1: import the libraries

import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
```

```
In [2]: # Task-2: Read the file DatasetLab1.csv, and show the first five rows
```

```
In [3]: # Task-3: save the values of the first 8 variables in x1 and the output variable in y1
```

```
In [5]: # Task-4: standardize the data (both x1 and y1) using the StandardScaler().fit_transform function
from sklearn.preprocessing import StandardScaler

x_Stand = StandardScaler().fit_transform(x1)
y_Stand = StandardScaler().fit_transform(y1.reshape(-1, 1))
```

```
In [4]: # Task-5: split the data into train and test datasets with test size = 20%
# for this use the train_test_split () function, this function will return two datasets for each x_Stand
# and y_Stand  xTrain, xTest, yTrain, yTest
```

```
In [5]: # Task_5 : apply LR model on xTrain, yTrain. then, print the r_squared, the intercept, and the slopes
# for this you need first to create the LR model then use the fit() function, visit the following link for more info
# https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

# 5.1- print r_squared (b0)

# 5.2- print y-intercept (b0)

# 5.3- print the slopes (coef for each variable)
```

```
In [6]: # Task-6: use the model to predict the output for the test data set (x_test) , then find the r squared.
```