LINEAR REGRESSION
CASE

ACTIVE LEARNING

ANALYSING LABOUR PROBLEM

BACKGROUND

Business Problem

"The labour problem" is the economics term widely used toward the turn of the twentieth century with various applications. It has been defined in many ways, such as "the problem of **improving the conditions of employment** of the wage-earning classes." It encompasses the difficulties faced by wage-earners and employers who began to cut wages for various reasons including increased technology, desire for lower costs or to stay in business.

Now the labour organization, **https://labour.gov.in/lcandilasdivision/india-ilo** the Indian version of this published the dataset used in this case.

Dataset contains data for the years 1974 and 1975 and wanted to demo data science models such that the data of 1978 labour can be predicted or not?

Data Contains Age, Race, Educational detail and Labour earning for 1974, 1975. The problem we are solving is the prediction of the

future labours earning. The earning can be dependent on many of the variables.

We have data for following

Age of the person.

Race: Is he/she is black or not black.

Education Details: How qualified the person is?

Hispanic: Is that person is Hispanic or not?

Married: Does marriage affect the earnings. And other information.

Solution Approach:

As we need to predict Labour earning for 1978 which is continuous in nature, Linear Regression can be used for prediction

PART 1: EDA (Exploratory data analysis)

Use all the functions learned like head, tail, describe, shape, size, info, correlation, pair plot, scatter plot, box plot, violin plot,

PART 2: Convert categorical into numerical and merge

This is most important step you will be required to do in most of your projects. Creation of dummies using get_dummies, then merge using concat function.

PART 3: Divide the data into 2 parts

Split the data into 2 parts using train_test_split, run linear regression model, find coefficients, interpret the coefficients

PART 4: Understand performance metrics

Determine the credibility of your model using r square, RMSE