**LAB EXERCISE 1**

Data Acquisition: pd.read\_csv ()

Data Exploration: df.head(), df.describe(), df.info(), df.vaule\_counts()

Data Wrangling: df.isna(), df.dropna, df.replace()

* + handling missing values
  + data standardization
  + data normalization

EDA (Exploratory Data Analysis)

Feature pattern

* + - * + Continuous variables (positive/weak linear relationship): sns.regplot()
        + Categorical variables: sns.boxplot
        + Grouping, Pivot & Correlation: df.groupby(), df.corr()

Descriptive Analysis: df.unique()

* Create bins for the values in the column: plt.pylot.hist(), plt.pyplot.xlabel(), plt.pyplot.ylabel(), plt.pyplot.title()
  + - Replacing column values using replace function: df.replace()

**LAB EXERCISE 2**

Prepare Data

df.drop(), df.rename(), df.sum(),

df.set\_index(), df.index.name,

df.loc[], df.iloc[], df.columns

Visualization

Get a dataset: df.loc[]

Plot Chart: df.transpose()

df.plot(kind=’**line’**), df.plot(kind=’**box**’)

plt.title()

plt.ylabel()

plt.xlabel()

plt.show()

**LAB EXERCISE 3**

from sklearn import linear\_model

regr = linear\_model.LinearRegression()

train\_x = np.asanyarray(train[['ENGINESIZE']])

train\_y = np.asanyarray(train[['CO2EMISSIONS']])

regr.fit (train\_x, train\_y)