Project-1:

Insurance Cost Prediction

The dataset contains individual medical costs incurred by a health insurance provider based on a handful of categorical and numeric attributes.

- **age**: age of the primary beneficiary
- **sex**: insurance contractor gender, female, male
- **BMI**: Body mass index, providing an understanding of the body, weights that are relatively high or low relative to height, objective index of body weight (kg/m^2) using the ratio of height to weight, ideally 18.5 to 24.9
- **children**: Number of children covered by health insurance/Number of dependents
- **smoker**: Smoking? (yes/no)
- **region**: the beneficiary's residential area in the US (northeast, southeast, southwest, northwest)
- **charges**: Individual medical costs billed by health insurance

Download Link: click here.
(https://raw.githubusercontent.com/lnbspace/datasets/csv/Insurance%20data.csv)

Tasks

- ➢ *Analyse the dataset and do EDA(Exploratory Data Analysis).*
- ➢ *Plotting of various graphs & correlations.*
- ➢ *Model Building using Linear Regression.*
- ➢ *Calculating the R-Squared, RMSE, and MSE for the model.*

Note: Please note that this is a real world dataset. The code should be well commented and submitted with outputs in an Ipython notebook (Jupyter Notebook).

Hints: EDA refers to exploring the dataset from various facets such as Null values, duplicates, wrong data types, types of features (categorical/numerical), distribution, skewness, measures of central tendency, quartiles, outliers, correlations, categorical feature to numeric feature conversion, univariate and bivariate analysis, etc.

Feel free to use any kind of exploratory data analysis on the dataset and present your insights about the data.