Objective: to build a machine learning solution to predict the Estimated\_Annual\_Salary of a customer.

Steps to follow:

1. Study the data dictionary (available in root directory. Name synthetic\_regression\_data\_dictionary.docx) in detail so that you have sound idea of the dataset.
2. Import the dataset (available in root directory. Name synthetic\_regression\_dataset.csv) and perform primary inspection of data like number of rows, columns, field types etc (add as per your wish)
3. Perform a detailed exploratory data analysis (descriptive and visual, both univariate and bivariate) – save all the results to EDA folder in word document and pdf
4. Perform a detailed feature engineering for categorical and continuous data – save all the results to feature-engineering folder in word document and pdf
5. Split the data into train and test data. Keep proportion of 80:20. Save the datasets into model-data folder
6. Build a linear regression model on the top of training data and validate it. Save all results into model-performance folder in word document and pdf
7. Build a xgboost model on the top of training data and validate it. Save all results into model-performance folder in word document and pdf
8. Compare both the models based on performance, fairness and explainability and document your observations to into model-performance folder in word document and pdf
9. Save the best model to model folder and write a scoring script