## Model Prediciton

## Step 01: Loading Dataset



This step imports necessary libraries: pandas for data manipulation, train\_test\_split for splitting data, LinearRegression for building the regression model, and SimpleImputer for handling missing values

## Step 02: Feature and Target Variable Separation



In this step, the features (X) and the target variable (y) are separated from the Data Frame. The features are assigned to **X**, while the target variable 'Turnout' is assigned to **y**. Additionally, the 'Year' and 'NA' columns are dropped from the features since they are not used as predictors.

## Step 03: Data Splitting for Training and Testing



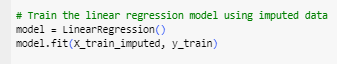
The dataset is split into training and testing sets using the **train\_test\_split()** function. 70% of the data is used for training (**X\_train**, **y\_train**), and 30% is used for testing (**X\_test**, **y\_test**). The **test\_size** parameter specifies the proportion of the dataset to include in the testing split, and **random\_state** is set to ensure reproducibility.

## Step 04: Handling Missing Values in Training Set



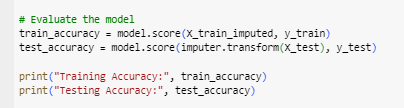
This step handles missing values in the training set (**X\_train**) by imputing them with the mean strategy using the **SimpleImputer** object. The **fit\_transform()** method is used to learn the mean value for each feature and then impute missing values.

## Step 05: Linear Regression Model Training



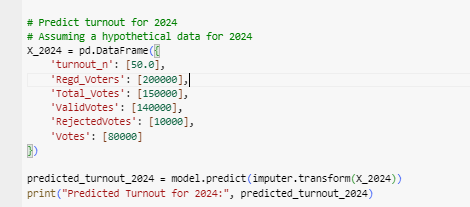
Here, a Linear Regression model is instantiated and trained using the imputed training data (**X\_train\_imputed**) and corresponding target variable (**y\_train**) using the **fit()** method.

## Step 06: Model Performance Evaluation



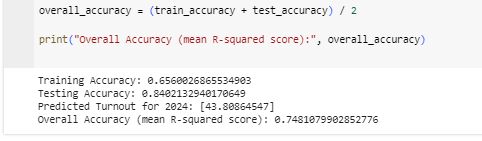
The model's performance is evaluated using the R-squared score. R-squared is a measure of how well the model explains the variability in the target variable. The scores for both the training and testing datasets are calculated using the **score()** method.

## Step 07: Prediction for 2024 Voter Turnout



A hypothetical dataset for the year 2024 is created, and the voter turnout is predicted using the trained model. The features for 2024 are passed into the model's **predict()** method after being imputed.

## Step 08: Overall Model Accuracy



Finally, the overall accuracy of the model is calculated as the mean of the training and testing accuracies obtained earlier, and it is printed out.