# Insights from Data

# Extra Credit Report

# Maxwell Fundi Njiru

1. **Project code**

All project code for this task may be found on my [Github repository](https://github.com/maxwellfundi/insights_from_data_R/blob/main/project_practice/project2_individual.R)

1. **Explain the concept of linear regression and its functionality.**

Linear regression is a type of supervised learning method that uses an algorithm to understand the relationship between dependent and independent variables by fitting a linear equation to the data. It is the most common type of regression approach. It generally uses the mathematical equation of y = ax+b. There are majorly two types of linear regression namely

* Simple linear regression - This is the most basic form of linear regression which uses one independent variable and one dependent variable.
* Multiple linear regression – This is used where there is more than one predictor variables

1. **Utilize linear regression to predict wine quality in the dataset. Include the "type" variable as a predictor alongside other input variables. Binarize the "type" variable for compatibility with linear regression**
2. **Divide the wine quality dataset into distinct training (80%) and test (20%) sets. Construct your linear model using the training data.**
3. **How can you interpret the coefficients in your linear regression model? Determine the main predictors based on the linear model.**
4. **Assess if any variables are irrelevant in predicting wine quality. Explore opportunities to enhance the linear regression model.**
5. **Execute predictions on the test data. Provide a scatter plot comparing actual vs. predicted wine quality.**
6. **Evaluate the performance of your method on test data and discuss the results.**