**Linear Regression Concepts**

* **Correlation vs Simple Linear Regression**
  + Correlation analysis shows the strength and direction of the linear relationship between two variables. Simple linear regression analysis estimates parameters in a linear equation to predict values of one variable from another.
* **The Dependent Variable**
  + In linear regression, the dependent variable (y) is continuous and numerical. It's the variable being predicted.
* **Linear Regression**
  + A linear approach to modeling the relationship between a scalar response (dependent variable) and one or more explanatory variables (independent variables). It's used to predict a continuous outcome based on one or more predictor variables.
* **Scalar**
  + A physical quantity that is completely described by its magnitude (e.g., volume, density, speed, energy, mass, and time). It's a single number, as opposed to a vector, which has both magnitude and direction.
* **Simple Linear Regression**
  + A type of linear regression where there's only one explanatory (independent) variable used to predict a single scalar response (dependent variable).
* **Line of Best Fit**
  + A straight line drawn through a scatter plot of data points that best represents the relationship between those points. It's often determined using the least squares method.
* **Residuals**
  + The vertical distance between a data point and the regression line (line of best fit). It represents the error in the prediction made by the regression model for that specific data point.
* **Mean Squared Error (MSE)**
  + A measure of the average squared difference between the predicted values and the actual values in a regression model. It's commonly used to evaluate the performance of a regression model.
* **Linear Regression with Scikit-learn**
  + Scikit-learn is a Python library that provides tools for machine learning, including linear regression. It can be used to build and train linear regression models.
* **Assumptions of Linear Regression**
  + Linear regression models have several assumptions that should be met for the model to be reliable, including linearity, no multicollinearity, normality of residuals, homoscedasticity, and independence of residuals.
* **Multiple Linear Regression**
  + A type of linear regression where there are two or more explanatory (independent) variables used to predict a single scalar response (dependent variable).
* **Confidence Intervals**
  + A range of values within which it is estimated that a population parameter (like a regression coefficient) lies, with a certain level of confidence (e.g., 95%). It's used to express the uncertainty associated with an estimate.
* **Types of Regression**
  + There are various types of regression analysis beyond simple and multiple linear regression, including polynomial regression, support vector regression, decision tree regression, and random forest regression, each with its own approach to modeling relationships between variables.