A glossary of terms and concepts from the document you provided is presented below.

* **Mean Squared Error (MSE)**
  + A measure of the average squared difference between the predicted values and the actual values in a regression model.
* **R-squared**
  + The proportion of the variation in the dependent variable predictable from the independent variable(s). It's a statistical measure that indicates how well a regression model fits the data.
* **Adjusted R-squared**
  + A modified version of R-squared that adjusts for the number of predictors in a regression model. It's used to compare models with different numbers of predictors.
* **Least Squares**
  + A standard approach in regression analysis for approximating the solution of overdetermined systems by minimizing the sum of the squares of the residuals. It's used to find the line of best fit in linear regression.
* **Statsmodels**
  + A Python module that provides tools for estimating different statistical models, conducting statistical tests, and statistical data exploration. It's used for in-depth statistical analysis.
* **Ordinary Least Squares (OLS)**
  + A type of linear least squares method for estimating the unknown parameters in a linear regression model. It's a common method for fitting linear regression models.
* **Sklearn Linear Regression Model**
  + A machine learning library that provides tools for building and training linear regression models. It's used for creating linear regression models in Python.
* **Standard Error of the Estimate**
  + The average distance that the observed values fall from the regression line. It's a measure of the accuracy of the regression model.

OLS Regression Results Explanation

Endog(enous): Similar to the dependent variable

Exog(enous): Similar to the independent variable

https://www.statisticshowto.com/endogenous-variable/

https://medium.com/swlh/interpreting-linear-regression-through-statsmodels-summary-4796d359035a

Model Info

* Dep. Varialble: the response variable, dependent, outcome, etc.
* Model: what model are we using (ordinary least squares) for the training
* Method: how the parameters (coefficients) were calculated
* No. Observations: the number of observations, rows... (n)
* DF Residuals: degrees of freedom of the residuals
* DF Model: number of parameters in the model excluding the constant if present
* Covariance Type: deals with violations of assumptions

Goodness of Fit

* R-Squared: coefficient of determination, how well the regression fits the data
* Adj R-Squared: R-squared adjustment based on number of parameters and df residuals
* F statistic: a measure of how significant the fit is
* Prop F statistic: the probability that you would get the F stat given the null hypothesis
* Log-Liklihood: can be used to compare the fit of different coefficients, the higher value is better
* AIC: Akaike Information Criterion is used to compare models, a lower score is better (doesn't address features, just the overall model)
* BIC: Bayesian Information Criterion is similar to AIC but uses a higher penalty

Coefficients

* coef: the estimated value of the coefficient
* std error: the basic standard error of the estimate of the coefficient
* t: the t-statistic value, how significant the coefficient is
* P>|t|: the p-value, indicates a statistically significant relationship to the dependent variable if less than the confidence level, usually 0.05
* 95% confidence interval: the lower and upper values

Statistical Tests

* Skewness: A measure of the symmetry of the data about the mean
* Kurtosis: A measure of the shape of the data
* Omnibus: D'Angostino's test provides a combined test for the presence of skewness and kurtosis
* Prob(Omnibus): probability of Omnibus
* Jarque-Bera: Another test for skewness and kurtosis
* Prob(Jarque-Bera): probability of Jarque-Bera
* Durbin-Watson: A test for the presence of autocorrelation, if the errors aren't independent
* Cond No: A test for multicollinearity

**Understanding the Basics**

* **Coefficients:** These are the 'weights' assigned to each feature variable (x1, x2, etc.) in your model. They determine the strength and direction of the relationship between a feature and the outcome.
* **Standard Error:** This measures the accuracy of your coefficient estimates. A smaller standard error means your coefficient estimate is likely more precise.
* **t-statistic:** This value helps determine the significance of your coefficients. It's calculated by dividing the coefficient by its standard error. A larger absolute t-value suggests the feature is a more significant predictor.
* **p-value:** This indicates the probability of observing the given t-value if there were no real relationship between the feature and the outcome. A small p-value (typically < 0.05) suggests that the feature is statistically significant in predicting the outcome.
* **Confidence Intervals:** These provide a range within which the true population coefficient is likely to fall (with a certain level of confidence, e.g., 95%).

**How They Contribute to Predicting 'y'**

In a linear regression model, the outcome 'y' is predicted using an equation like this:

y = β0 + β1\*x1 + β2\*x2 + ... + βn\*xn

* β0 is the intercept (the value of y when all features are 0)
* β1, β2, ... βn are the coefficients for each feature
* x1, x2, ... xn are the values of your features

The coefficients, guided by their significance (t-value, p-value) and confidence intervals, determine how much each feature contributes to the final prediction of 'y'.

**Example**

Let's imagine a model predicting house prices ('y') based on features like size (x1) and location (x2):

* Coefficient for size (β1) = 1000
* Coefficient for location (β2) = 50000

This means that:

* For every 1 unit increase in size, the house price is predicted to increase by 1000 units (holding location constant).
* For a 1 unit increase in the location's desirability score, the house price is predicted to increase by 50000 units (holding size constant).

The standard error, t-value, p-value, and confidence intervals would further refine these estimates, indicating the precision and significance of these relationships.