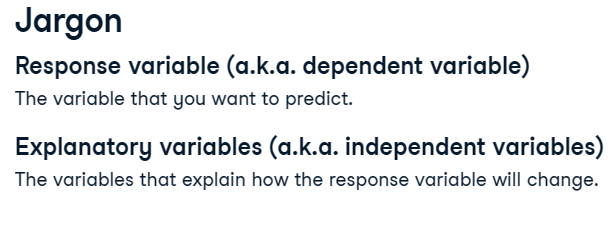
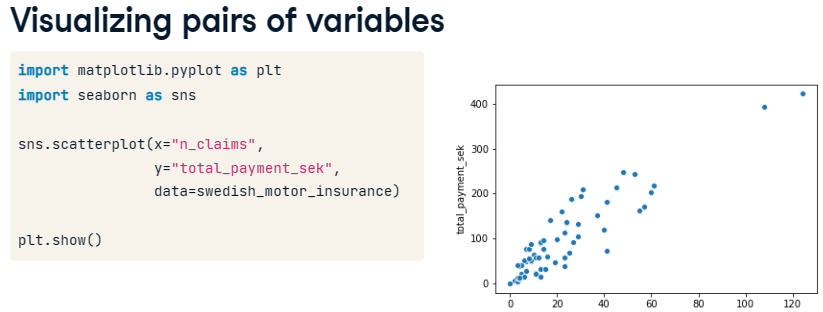
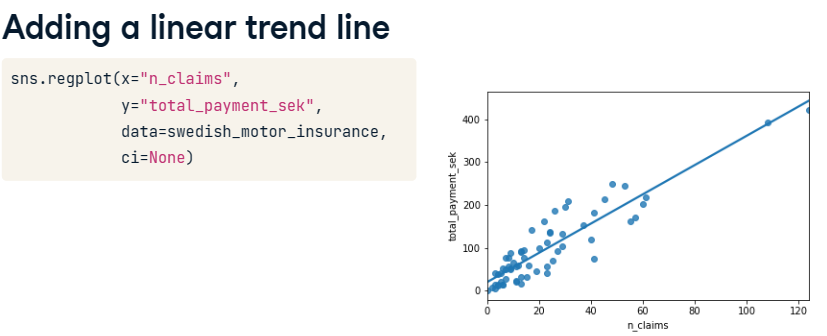
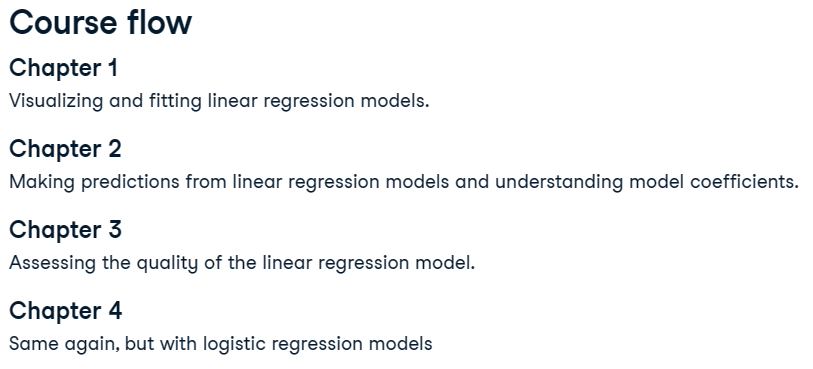
**Statmodels**

****

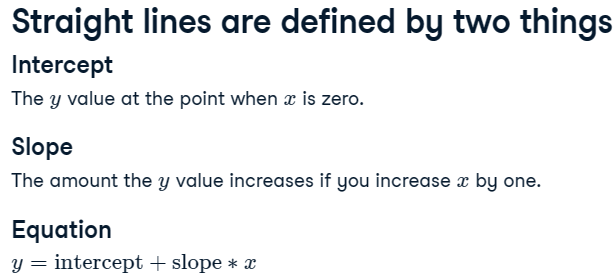
Before running regression model, it is good behavior to visualize your data

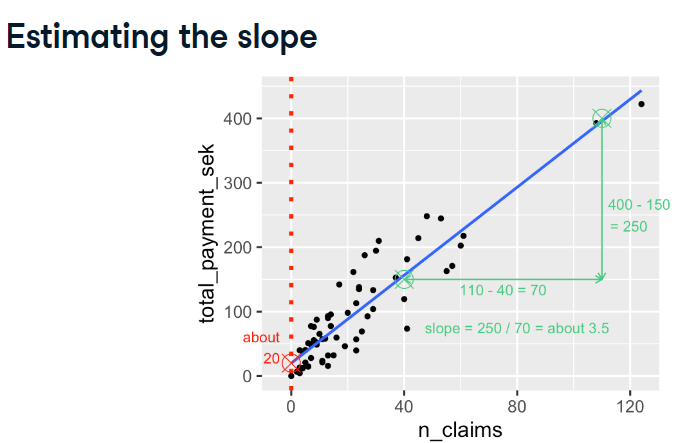






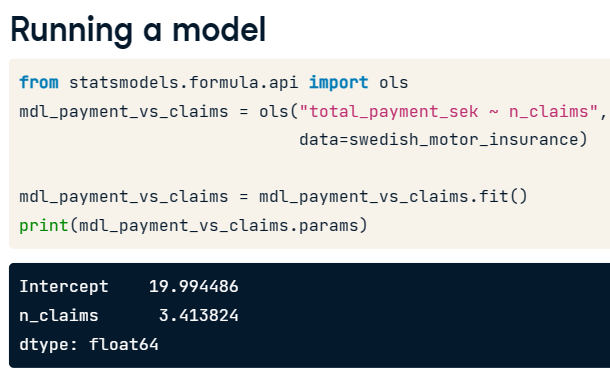
**Fitting a linear regression**

****

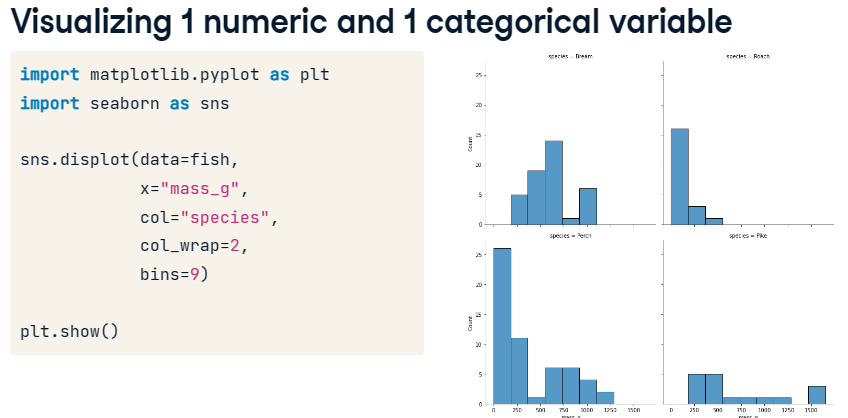


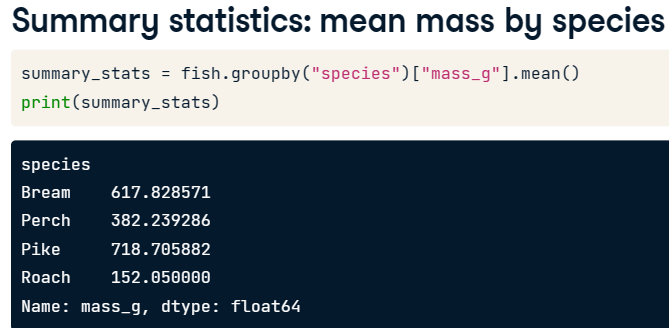
**ols** – ordinary released squares

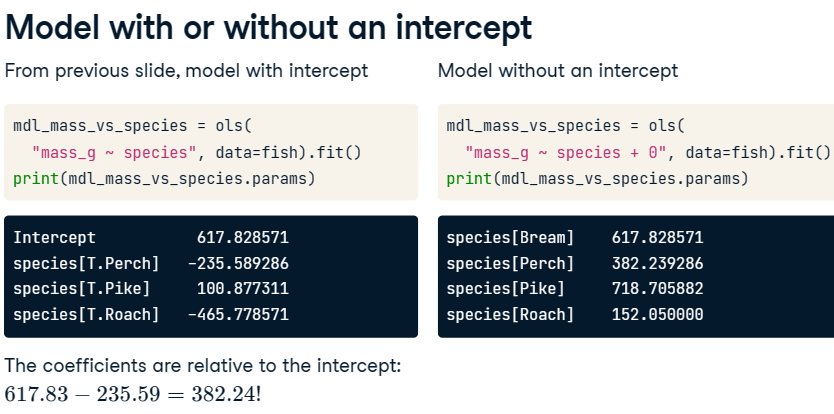
**.params** – parameters



**Categorical Explanatory variables**

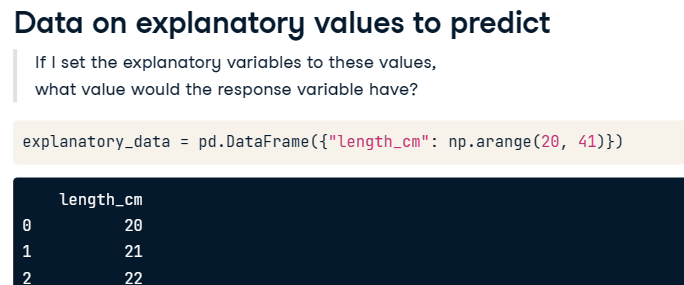
****

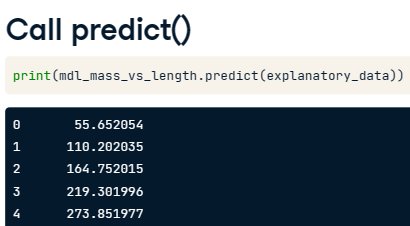
****

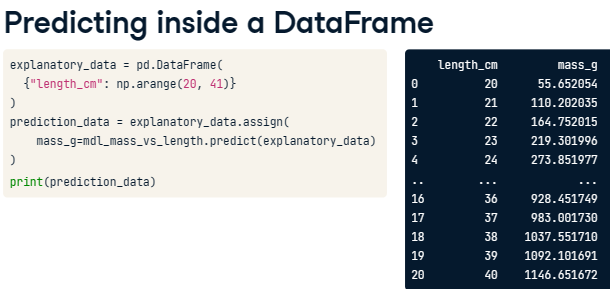


As we see on the left the slope is negative, so we add ‘0’ on the right. By doing this we ensure that our data gets numbers above ‘0’

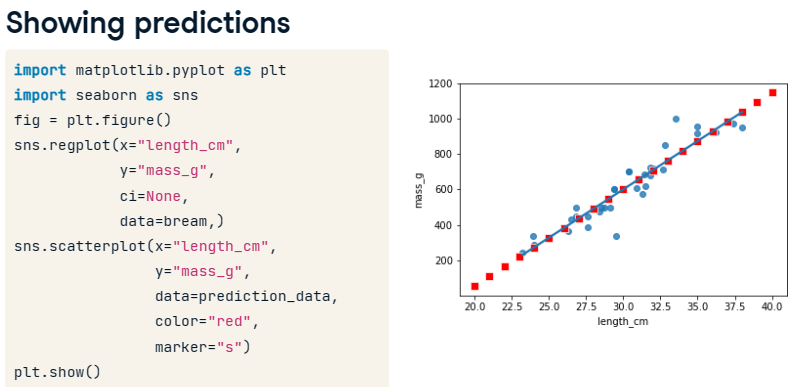
**MAKING PREDICTIONS**

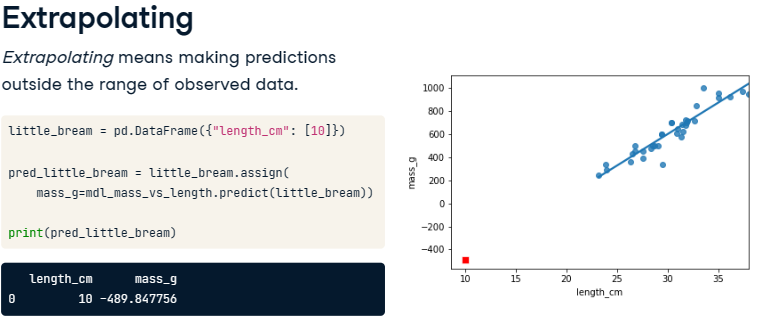
****

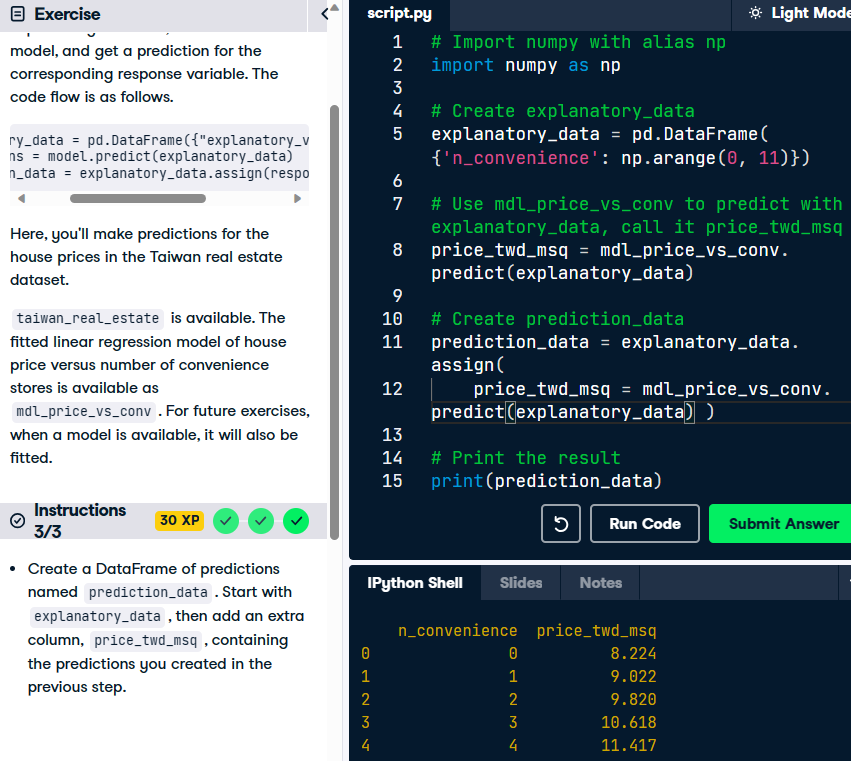
**predict()**

****

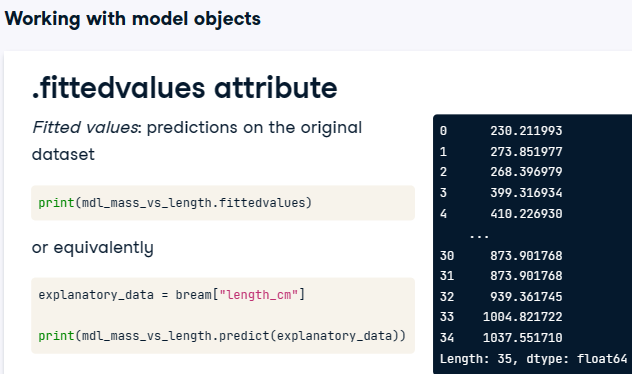
**assign()**

****

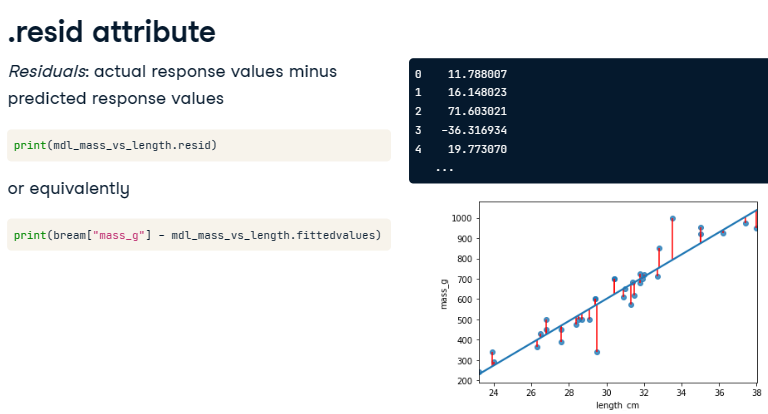




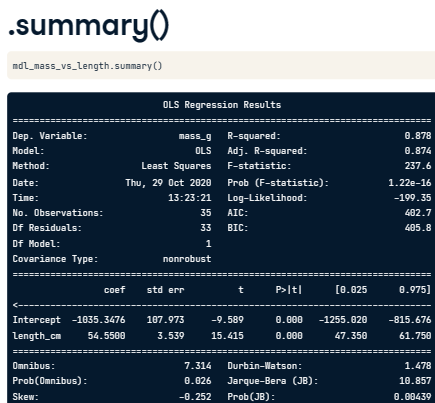
**MODEL OBJECTS**

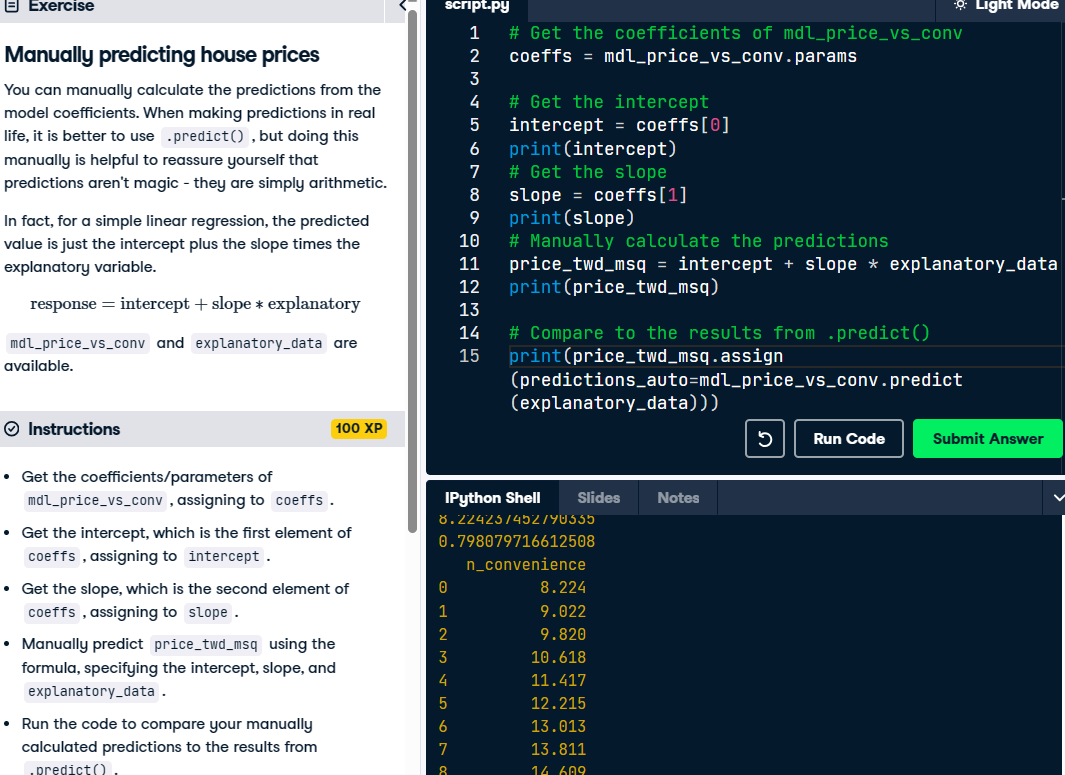
****

**Fittedvalues()**

****

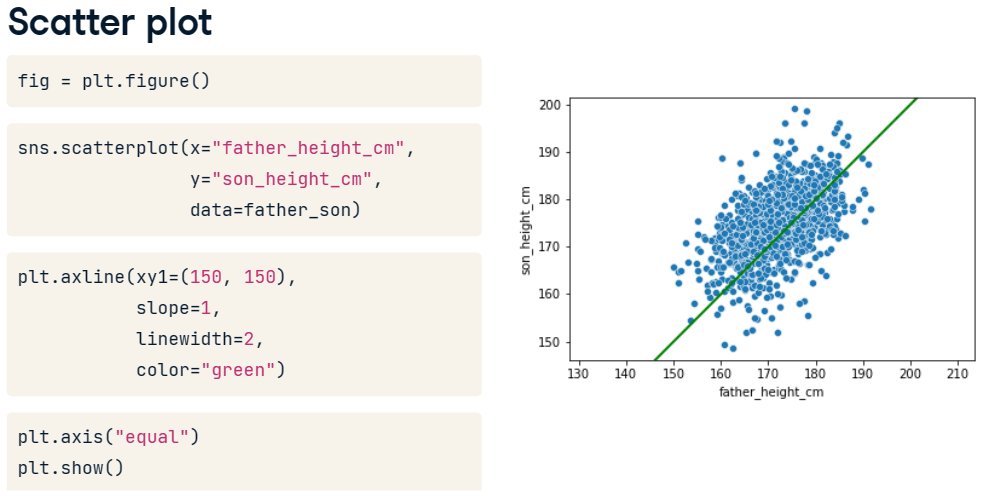
**resid()**

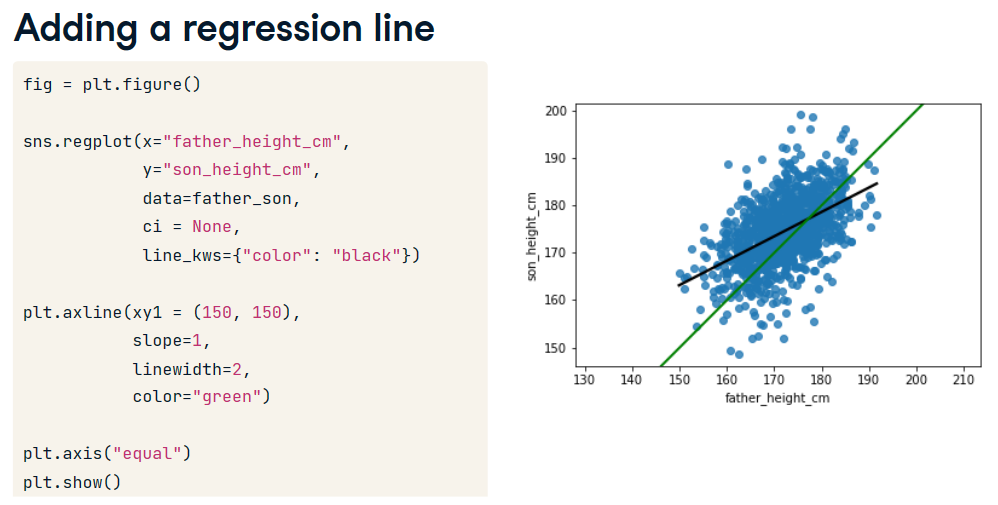
**summary()**

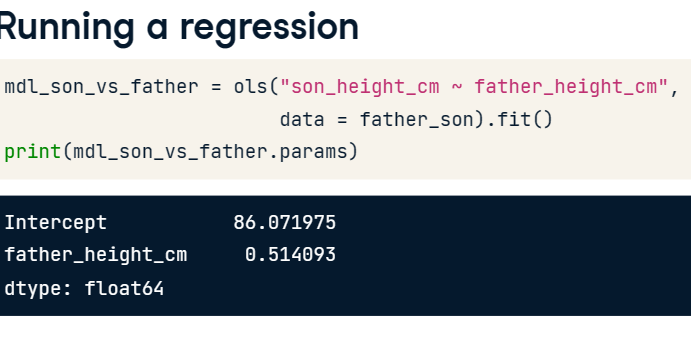


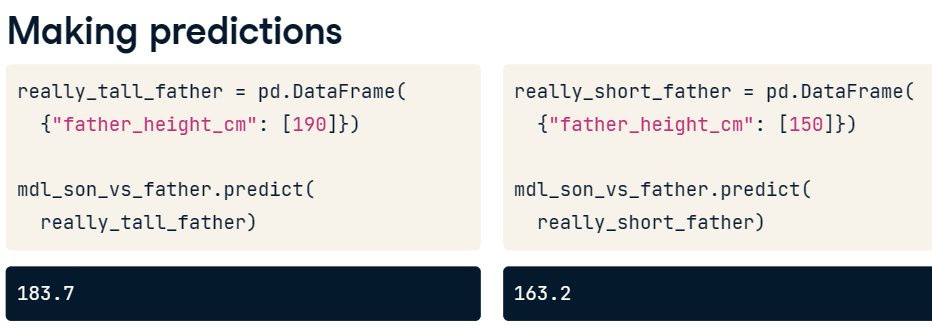
Formula of calculating prediction model of linear regression

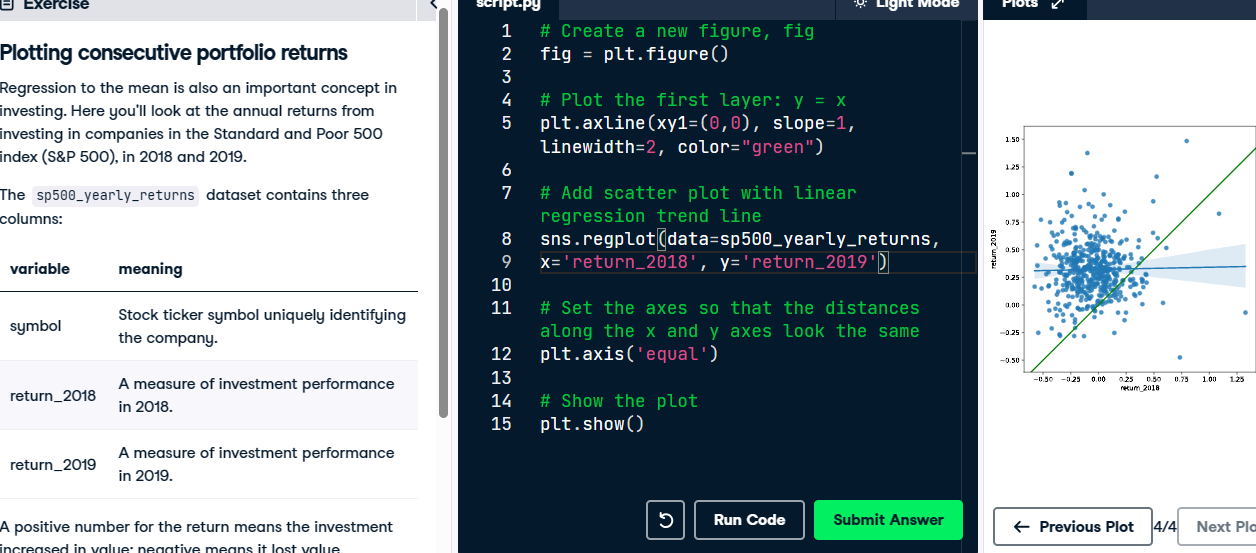
**REGRESSION TO THE MEAN**

****

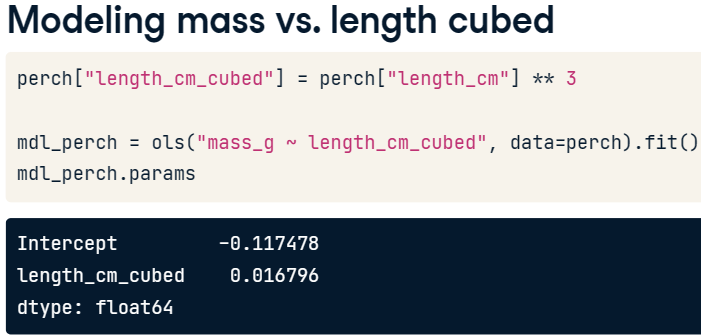
****

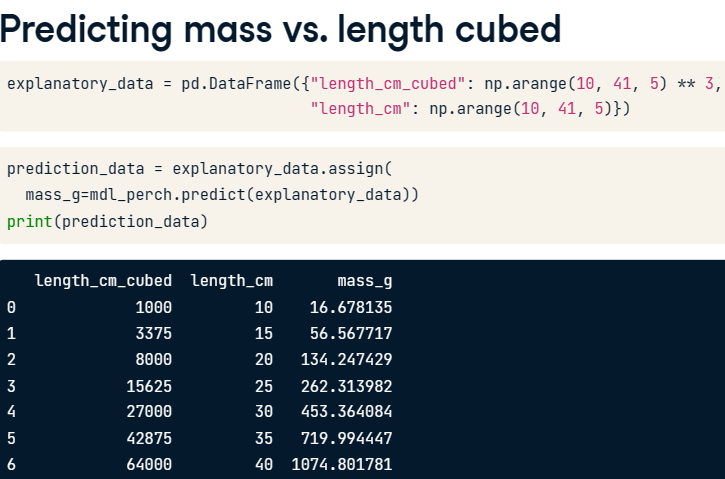
****

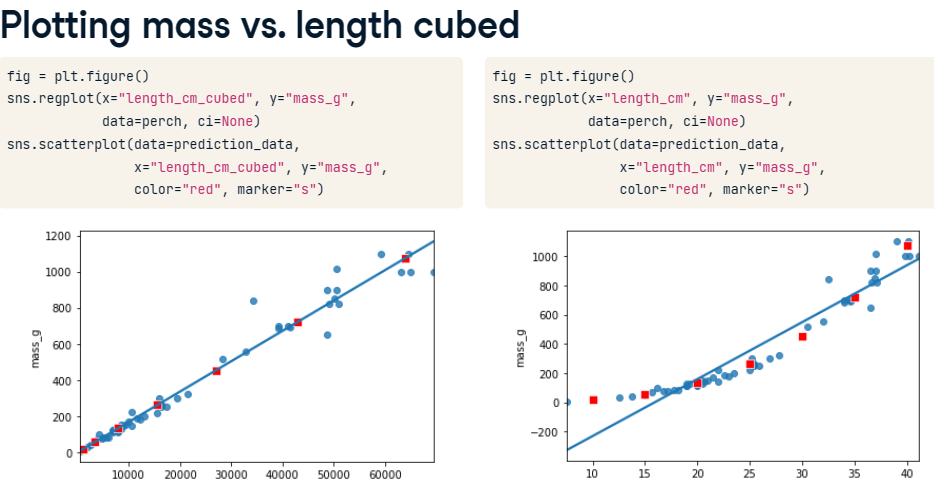
****



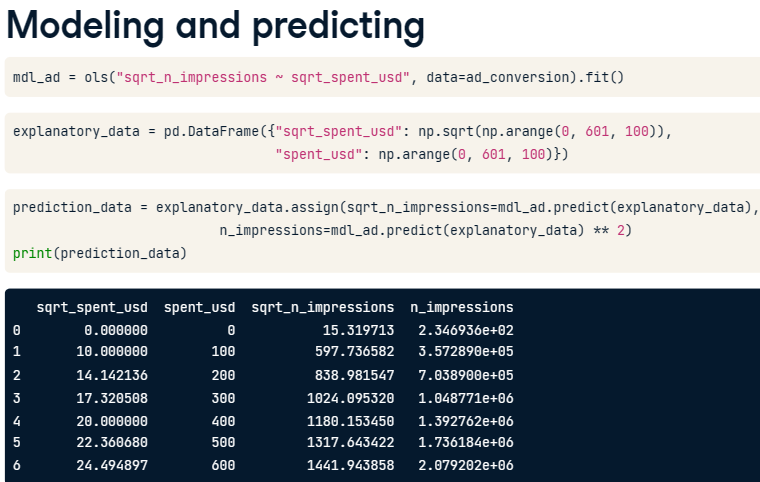
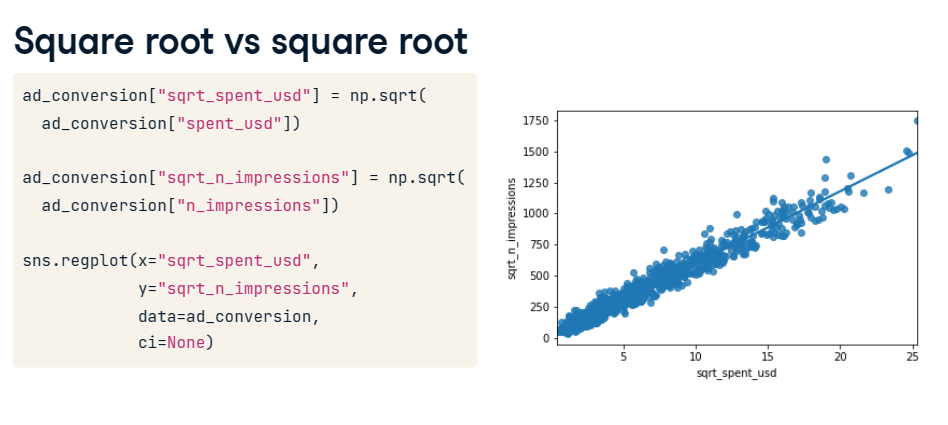
**TRANSFORMING THE MODELS TO BETTER FIT THE MODEL**

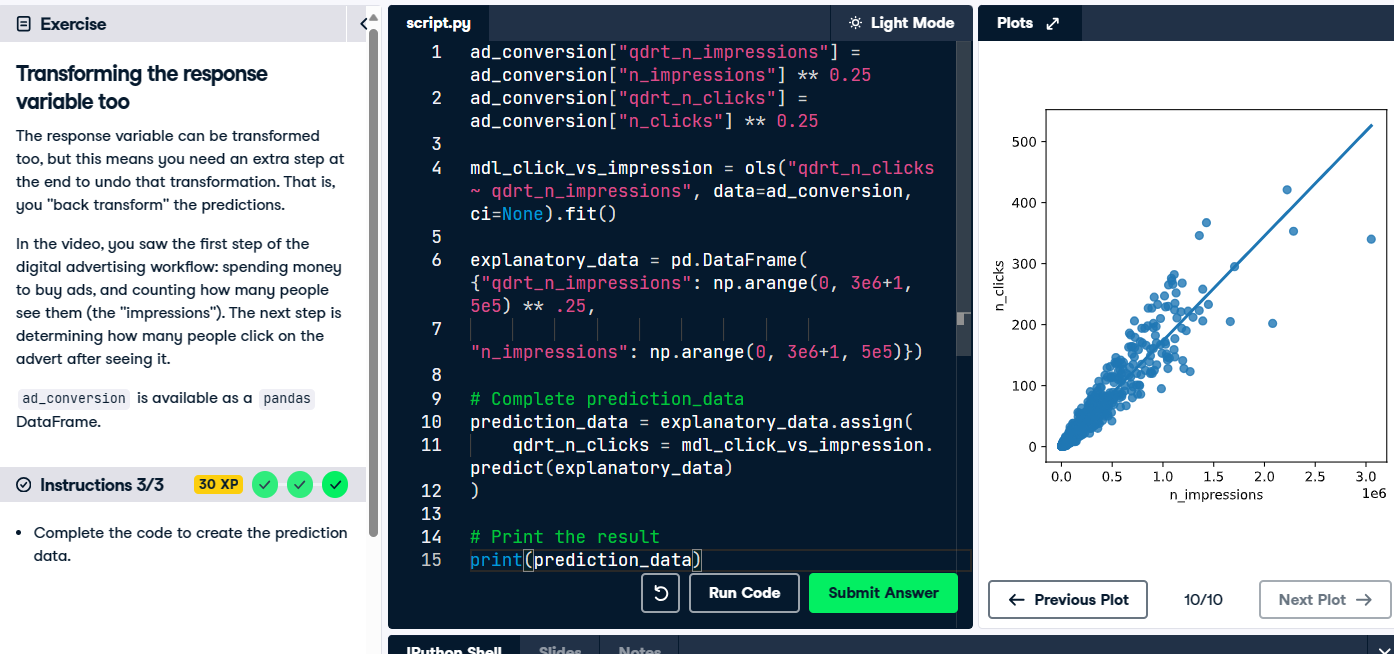
****

****

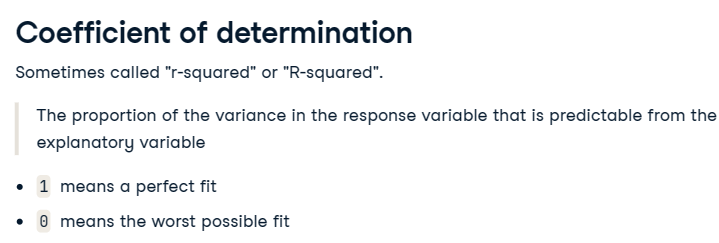
****

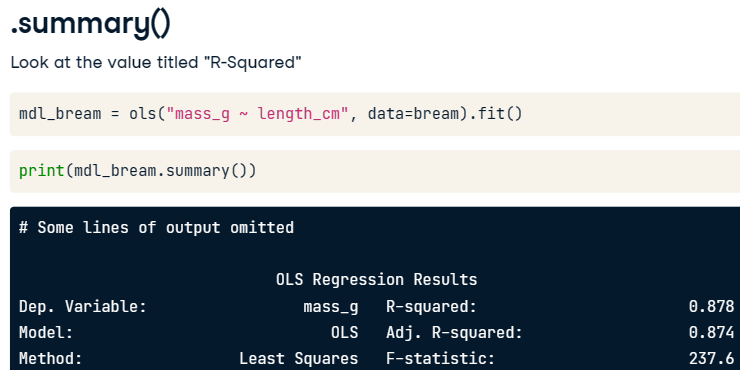
Facebook data

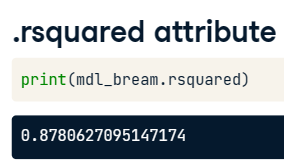


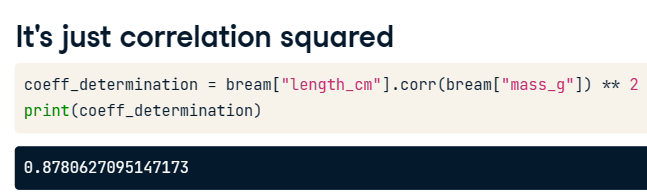


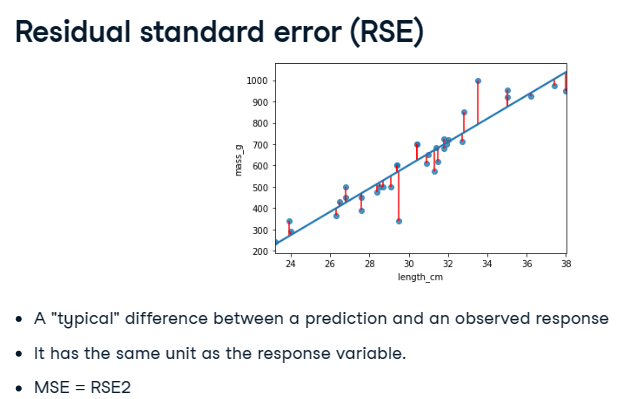
**QUANTIFYING MODEL FIT(HOW GOOD IT IS)**



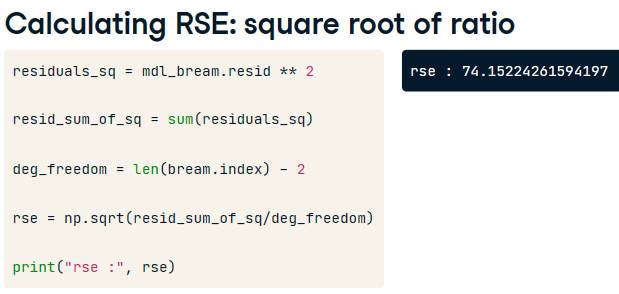


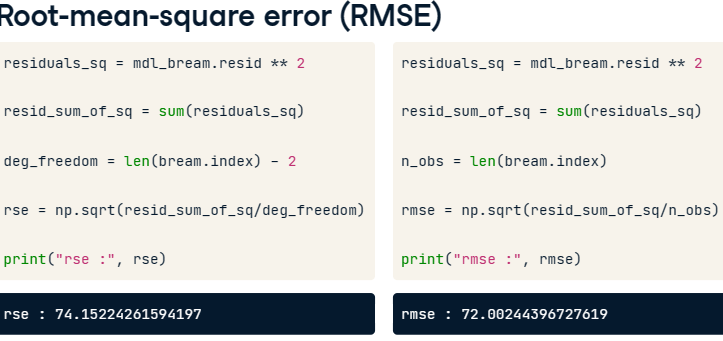
**rsquared()**



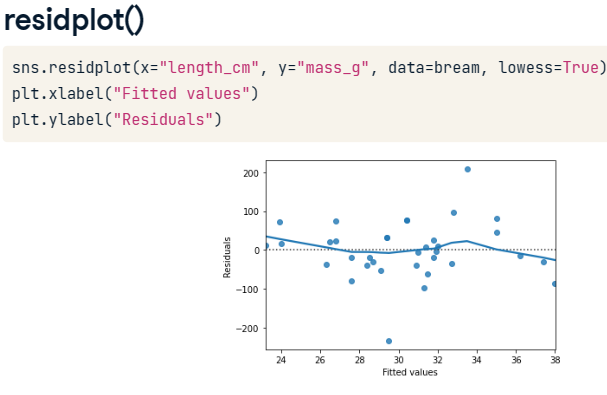


mse\_resid()

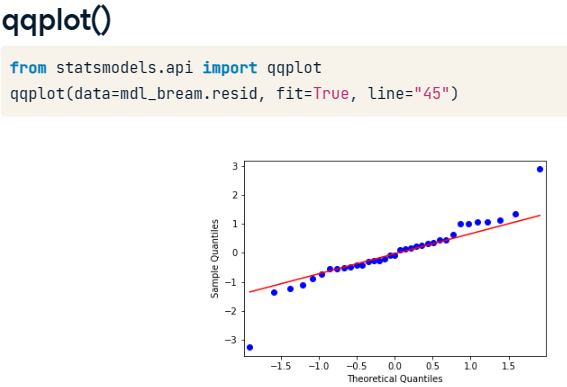
rse

rmse

**VISUALIZING MODEL FIT**

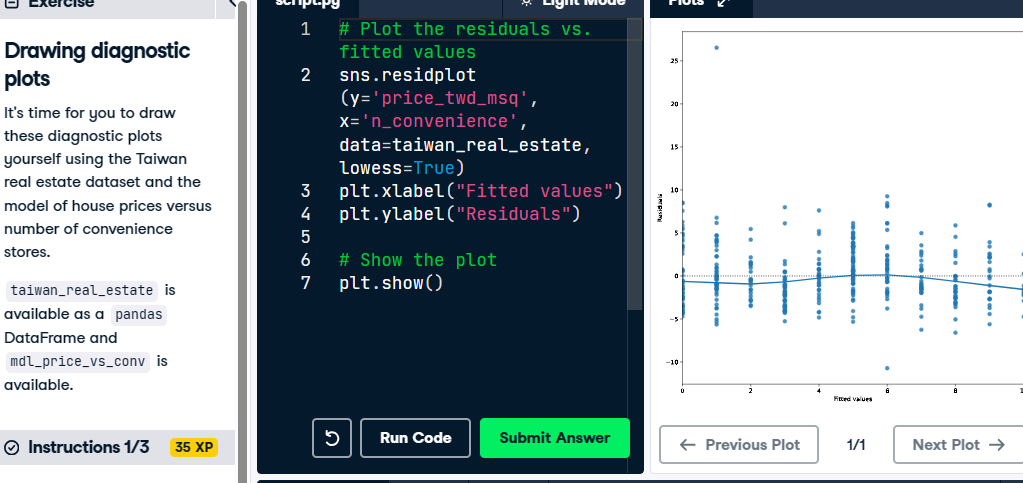
****

**residplot()**

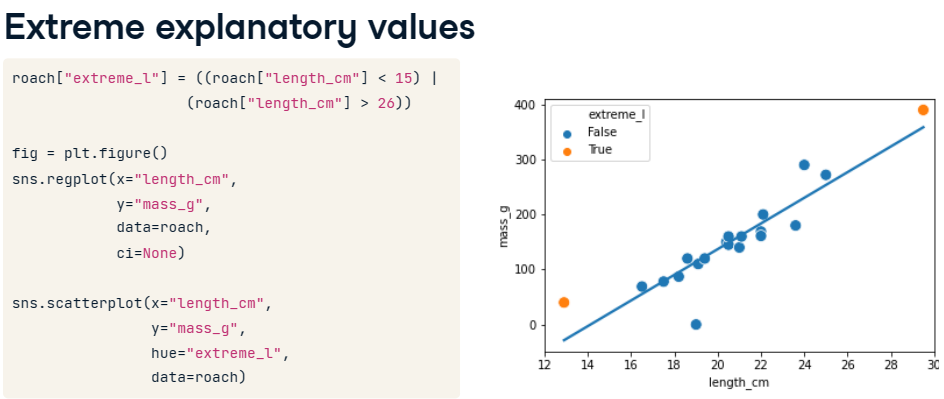
****

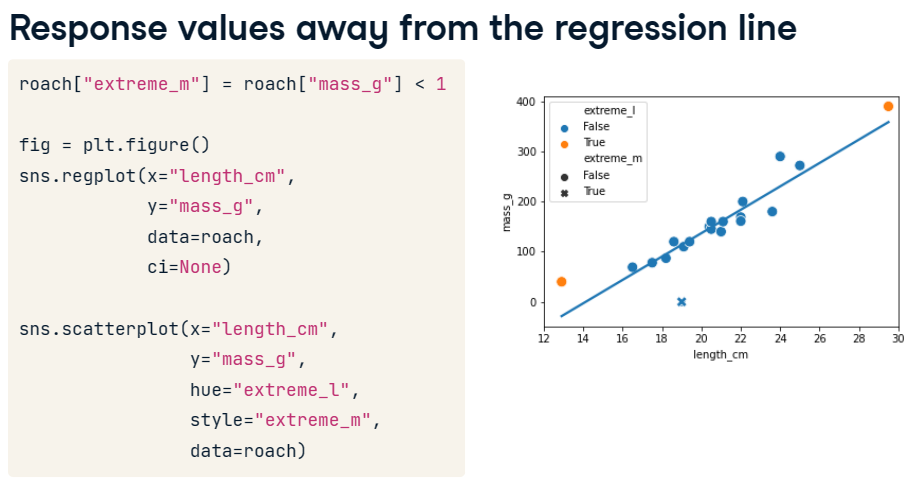
**qqplot()**

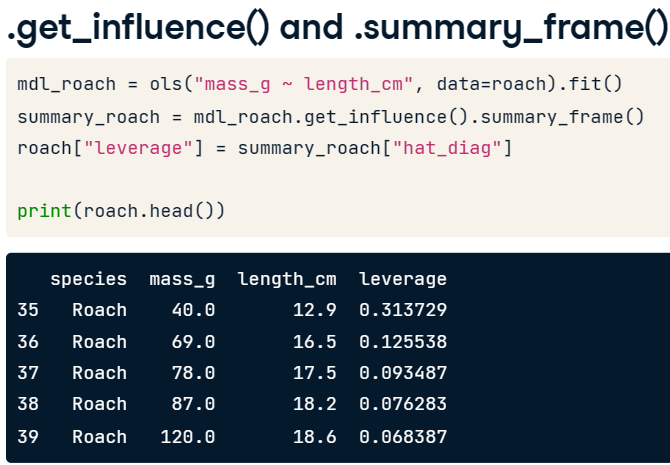




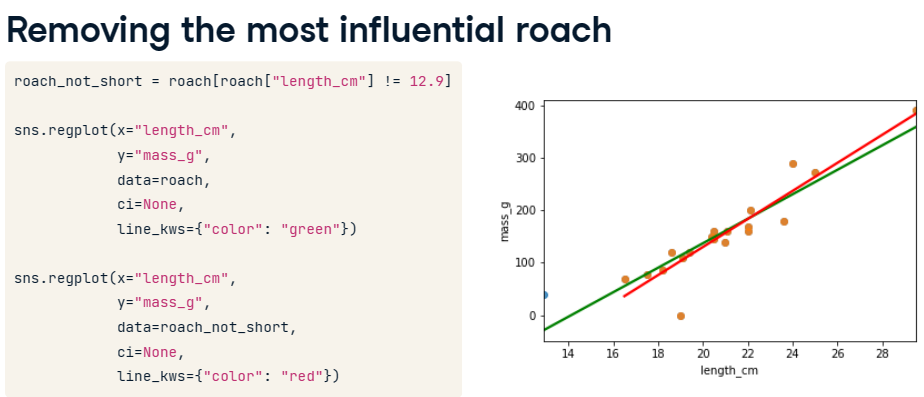
**OUTLIERS IN DATA**

****

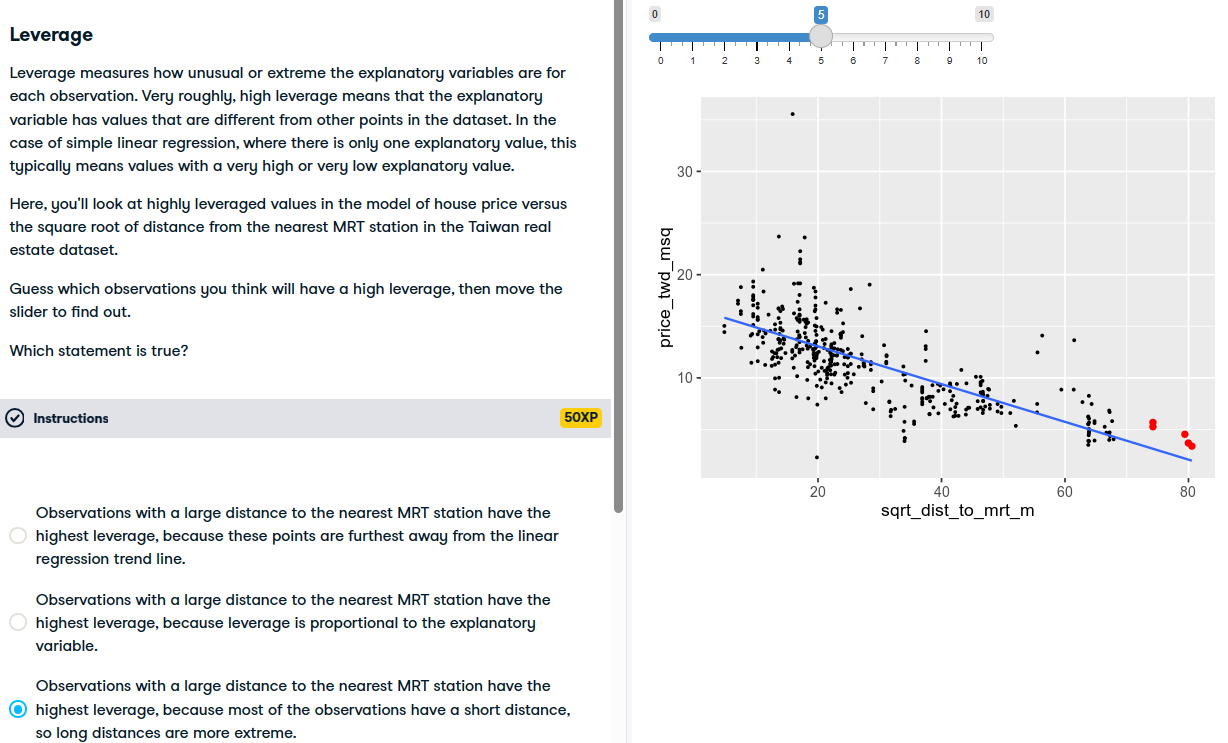


**get\_influence()**

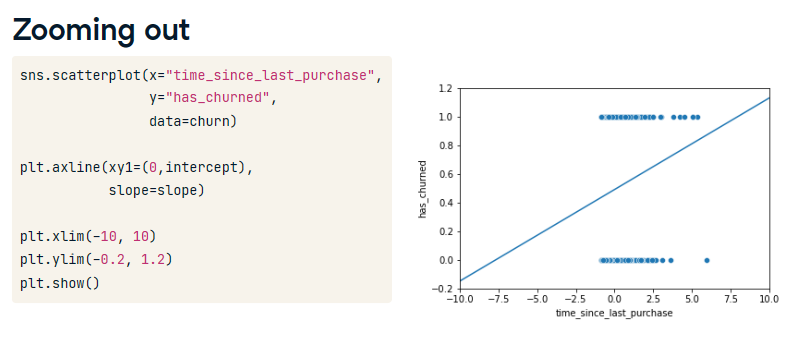
**summary\_frame()**



By removing the most influential data(shortest fish) in our column we can see that the slope has changed

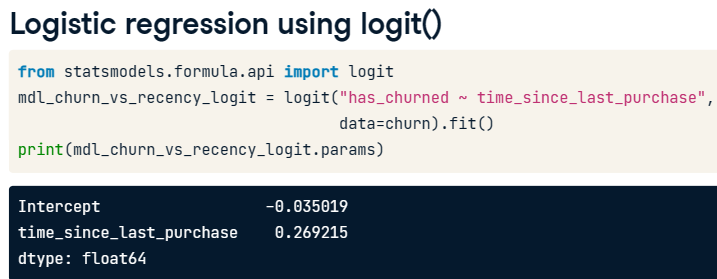


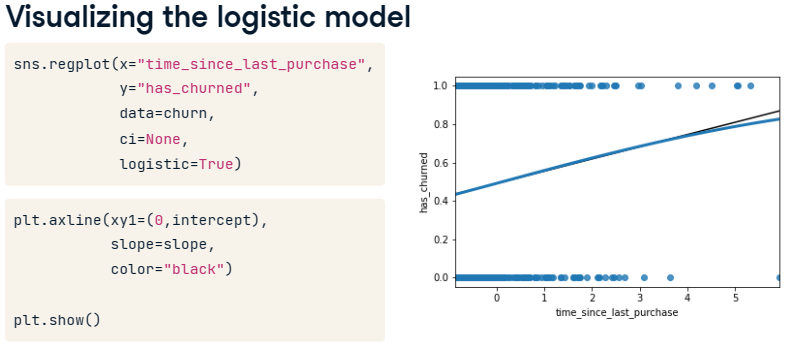
**LOGISTIC REGRESSION**

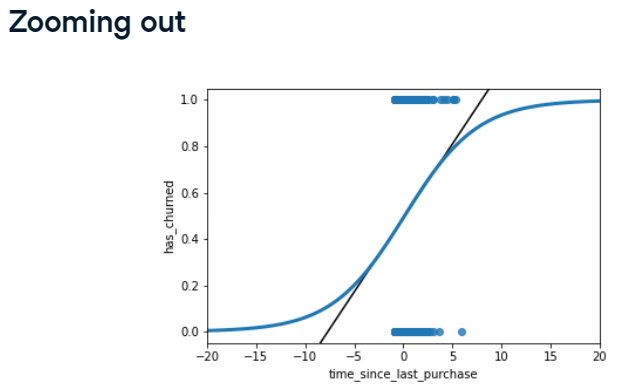
****

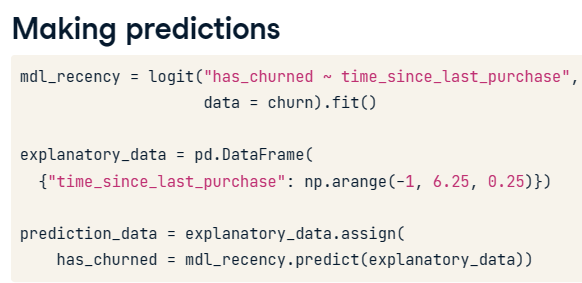
**Axline() –** for the line not to be edged withinh the dots on the plot

**Xlim() and ylim()** – for zooming the plot

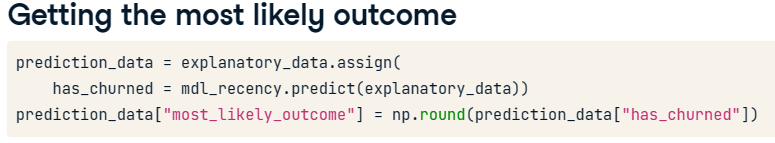


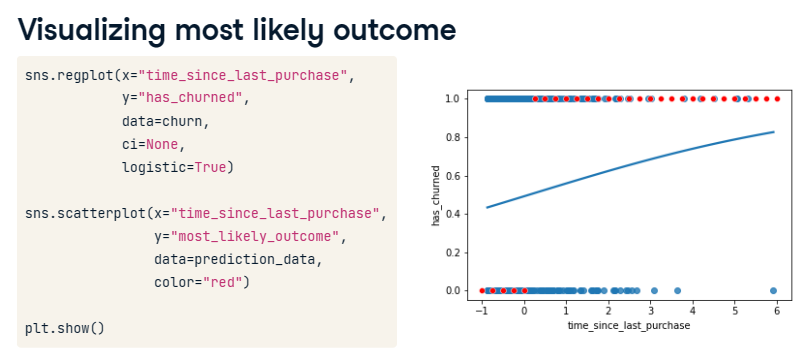


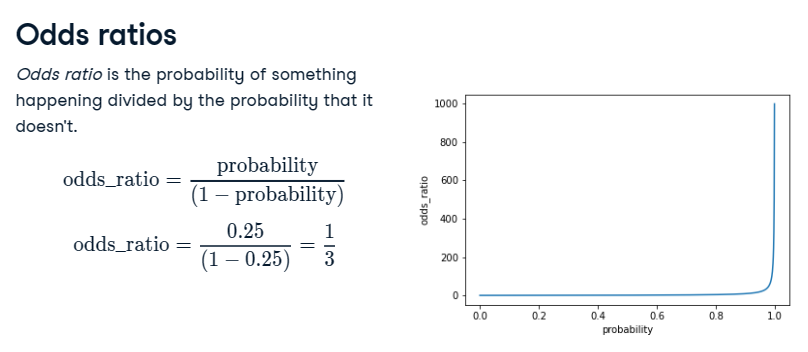
the customers who bought recently are less likely to churn, however customers who buy less frequently are more likely to churn



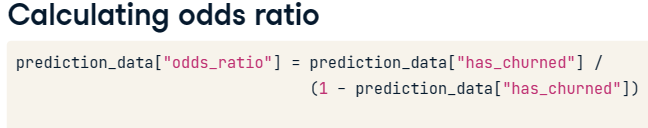


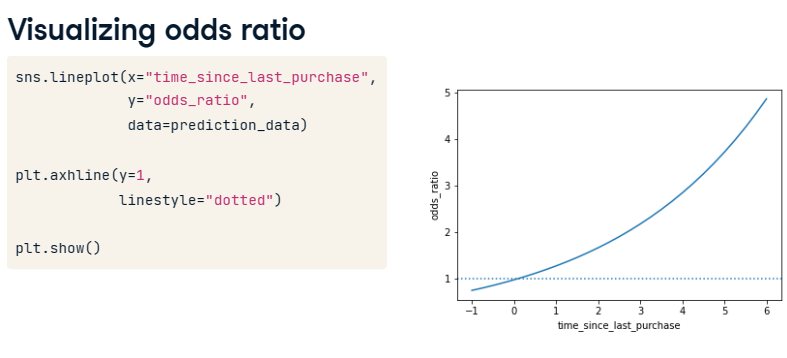


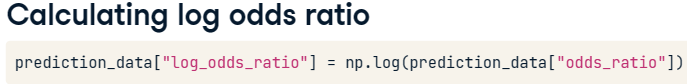


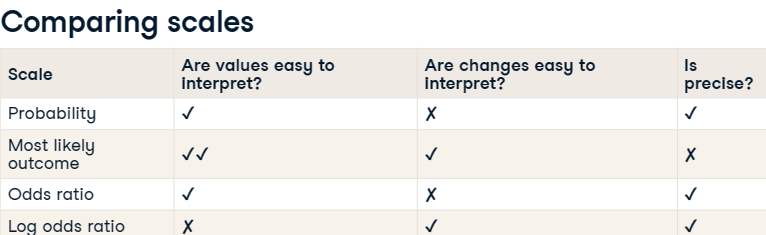
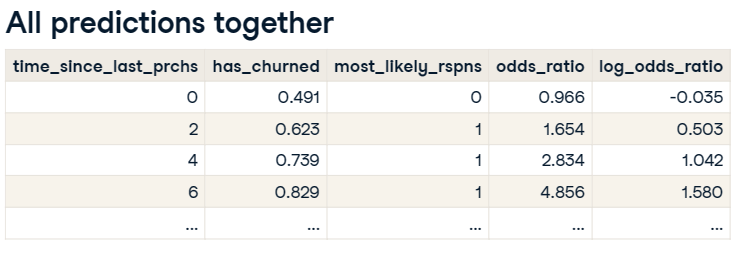


Odds ratio

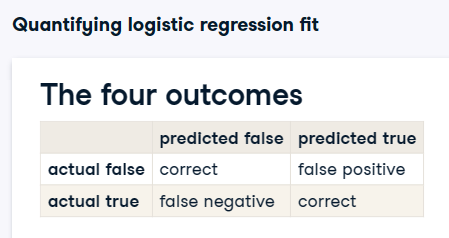


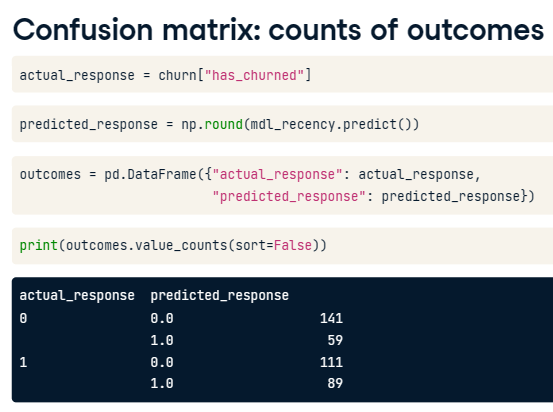


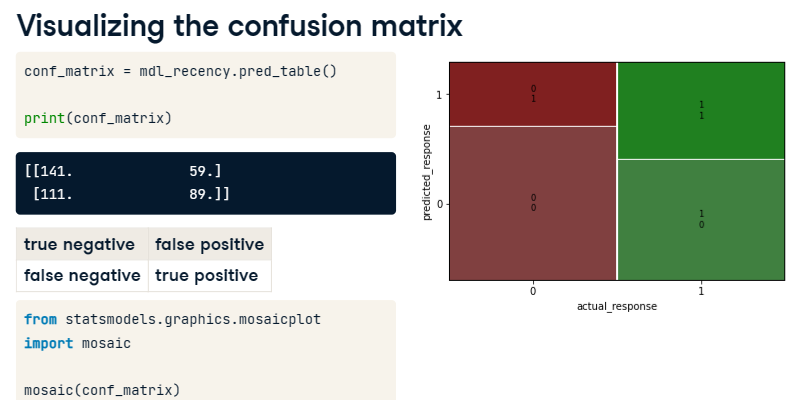




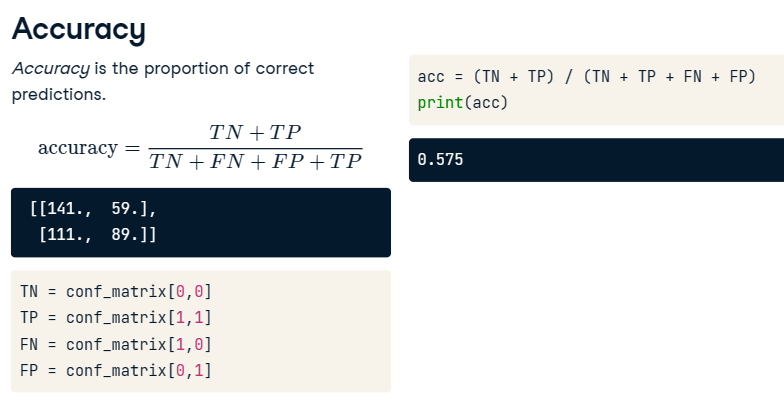
**QUANTIFYING LOGISTIC REGRESSION FIT**

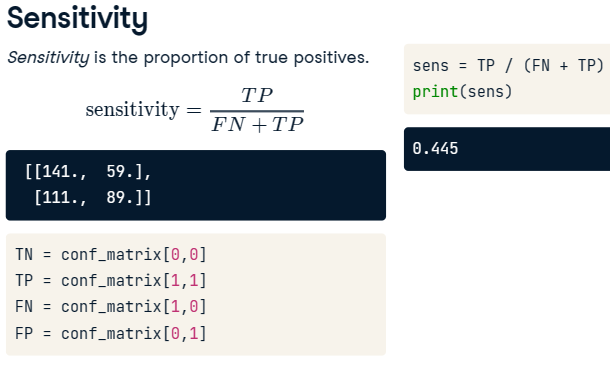
****

****



ACCURACY



SENSITIVITY

