

WHAT IF WE ADD PREPROCESSING?

```
1 lm_spec = LinearRegression()  
2 log_tf = FunctionTransformer(log)  
3  
4 pipeline = make_pipeline(  
5     log_tf,  
6     lm_spec  
7 )  
8  
9 pipeline.fit(X = penguins[["bill_length_mm", "bill_depth_mm"]],  
10             y = penguins["body_mass_g"])
```


WHAT IF WE ADD PREPROCESSING?

```
1 lm_spec = LinearRegression()  
2 log_tf = FunctionTransformer(log) ← Preprocessing decision made here  
3  
4 pipeline = make_pipeline(  
5     log_tf,  
6     lm_spec  
7 )  
8  
9 pipeline.fit(X = penguins[["bill_length_mm", "bill_depth_mm"]],  
10             y = penguins["body_mass_g"])
```

Predictor inclusion decision made here

The diagram illustrates decision points in the pipeline. A blue arrow points to the `FunctionTransformer(log)` step, labeled "Preprocessing decision made here". Two red arrows point to the feature names `"bill_length_mm"` and `"bill_depth_mm"` in the `fit` method, labeled "Predictor inclusion decision made here".