# Applying Logistic Regression Models: Takeaways



## Syntax

- Using the split-apply-combine workflow to quickly summarize many variables in a dataset:

```
auto.groupby("high_price").agg(
    {
        "horsepower": "mean",
        "width": "mean"
    }
)
```

- Creating a confusion matrix to see how a model misclassifies cases and non-cases:

```
from sklearn.metrics import confusion_matrix
predictions = model.predict(X)
confusion_matrix(y, predictions)
```

## Concepts

- Overfitting is a phenomena that occurs as we use increasingly more complex models. Instead of learning the underlying laws that "govern" the relationship between the outcome and predictor, the model is just learning the structure of the data itself. Evidence of overfitting can be seen in large discrepancies between optimistic training errors and much-worse test errors.
- Multi-class classification can be done in `LogisticRegression` models the same way as binary classification. Just be aware that the model metrics will change, as will the overall structure of the model itself.