# [Data Wrangling (Module 2)](https://www.coursera.org/learn/data-analysis-with-python/lecture/1wWCA/pre-processing-data-in-python)

Table of Contents

[Data Wrangling (Module 2) 1](#_Toc187337790)

[Lesson Summary 1](#_Toc187337791)

### Lesson Summary

Congratulations! You have completed this lesson. At this point in the course, you know:

* Data formatting is critical for making data from various sources consistent and comparable.
* Master the techniques in Python to convert units of measurement, like transforming "city miles per gallon" to "city-liters per 100 kilometers" for ease of comparison and analysis.
* Acquire skills to identify and correct data types in Python, ensuring the data is accurately represented for subsequent statistical analyses.
* Data normalization helps make variables comparable and helps eliminate inherent biases in statistical models.
* You can apply Feature Scaling, Min-Max, and Z-Score to normalize data and apply each technique in Python using pandas’ methods.
* Binning is a method of data pre-processing to improve model accuracy and data visualization.
* Run binning techniques in Python using numpy's "linspace" and pandas' "cut" methods, particularly for numerical variables like "price."
* Utilize histograms to visualize the distribution of binned data and gain insights into feature distributions.
* Statistical models generally require numerical inputs, making it necessary to convert categorical variables like "fuel type" into numerical formats.
* You can implement the one-hot encoding technique in Python using pandas’ **get\_dummies** method to transform categorical variables into a format suitable for machine learning models.