Data Analysis with Python

# Importing Datasets: Dataframe, Feature, Target

## Python packages

* Scientific Computing Libraries

1. Pandas: Data Structures
2. Numpy: Array & Matrices
3. SciPy: Integrals, solving differential equations, optimization

* Visualization Libraries

1. Matplotlib
2. Seaborn

* Algorithmic Libraries

1. Scikit-learn: Machine Learning (Regression, classification, etc.)
2. Statsmodels: Explore data, estimate statistical models and perform statistical tests

## Getting Started with Python Pandas

* df.dtypes()
* df.describe()
* df.info()

## Accessing databases with Python

### SQL API

Graphical user interface

Description automatically generated

### DB-API

Diagram

Description automatically generated

### Connection Methods (using DB-API)

Graphical user interface, text, application

Description automatically generated

## Summary

Graphical user interface, application

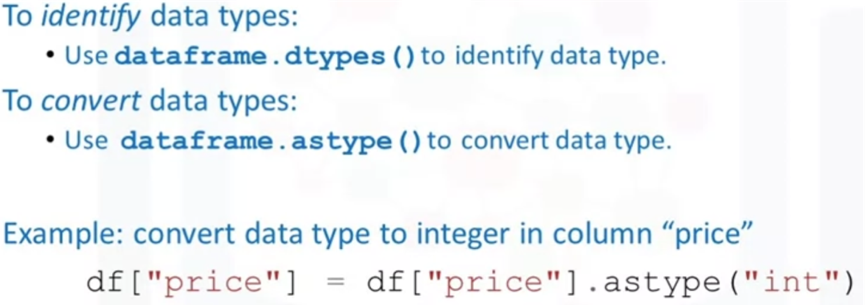
Description automatically generated

# Data Wrangling

## Dealing with Missing Values

* Drop the missing Values (Variable or Data Entry): df.dropna()
* Replace the Missing Values: df.replace()

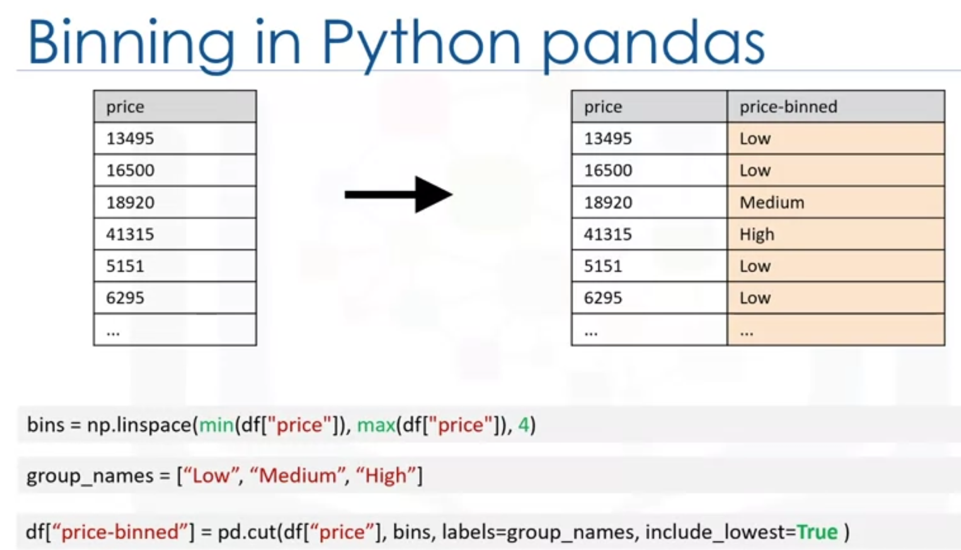
## Data Formatting: Correcting Data types



## Data Normalization

* Simple Feature Scaling
* Min-Max
* Z-Score

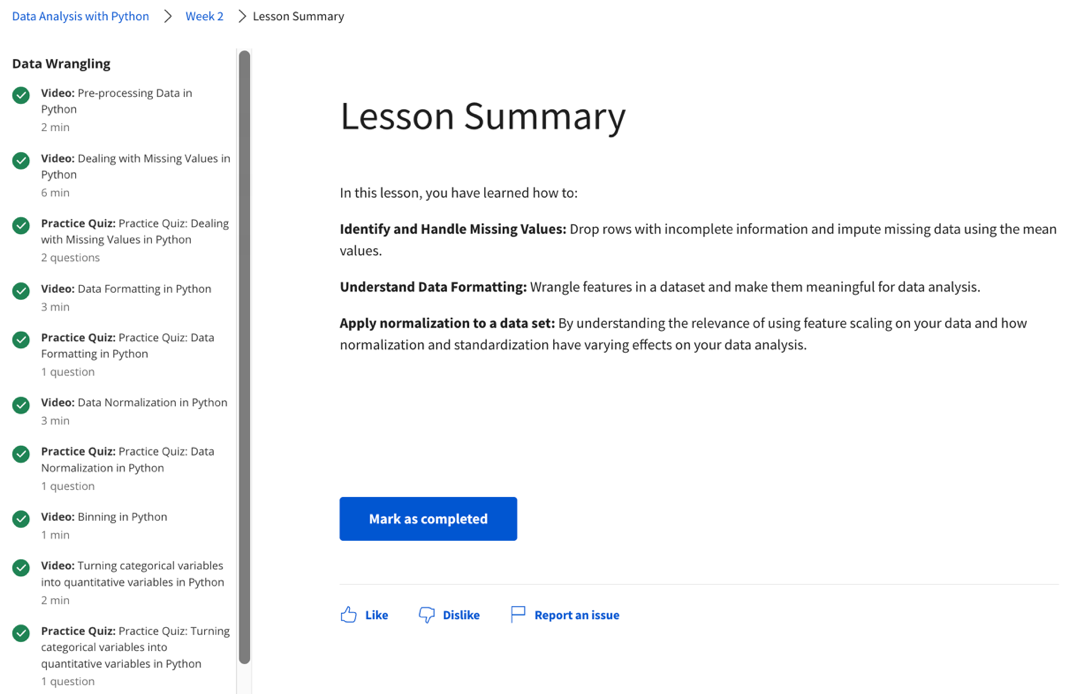
## Binning



## Turning Categorical Variables into Quantitative Variables

* pd.get\_dummies(df[“column”])

## Summary



# Exploratory Data Analysis

## Descriptive Statistics

* Short Summaries: df.describe(), Box Plot, Scatter Plot

## Grouping Data

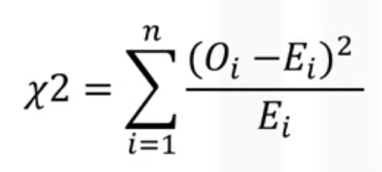
* groupby(), pivot(), heatmaps

## Correlation

* How are the variables dependent to each other?
* Pearson correlation (using SciPy package)

1. Correlation coefficient (0 to 1)
2. P-value (0.01, 0.05 etc)

## Chi-square: Association between 2 categorical variables



## Summary

Graphical user interface, text, application

Description automatically generated

# Model Development

## Simple and Multiple Linear Regression’

Text

Description automatically generated with medium confidence

### Fitting a Linear Model

Graphical user interface, text

Description automatically generated

Table

Description automatically generated with low confidence

### Fitting a Multiple Linear Model

Table

Description automatically generated

### Estimated Linear Model

Graphical user interface, text, application

Description automatically generated

## Model Evaluation using Visualization

* Regression Plot
* Residual Plot

Diagram

Description automatically generated

* Distribution Plots

Chart

Description automatically generated

## Polynomial Regression and Pipelines

Diagram, text

Description automatically generated with medium confidence

### Pipelines

Graphical user interface, diagram, text, chat or text message

Description automatically generated

## R-squared and MSE for In-Sample Evaluation

* Mean Squared Error (MSE)
* R-Squared

## Predictions and Decision Making

Timeline

Description automatically generated

## Summary

Graphical user interface, text, application

Description automatically generated

# Model Evaluation

## Model Evaluation and Refinement

### Splitting Data into train and test

Text

Description automatically generated

### Evaluation Metrics

* K-cross validation
* Cross val score

## Overfitting, Underfitting and Model Selection

Underfitting (left) and Overfitting (right)

A picture containing diagram

Description automatically generated

### Checking for the most optimal n value

Graphical user interface, text, application

Description automatically generated

## Ridge Regression

Ridge regression is a regression that is employed in a Multiple regression model when Multicollinearity occurs. Multicollinearity is when there is a strong relationship among the independent variables. Ridge regression is very common with polynomial regression. The next video shows how Ridge regression is used to regularize and reduce the standard errors to avoid over-fitting a regression model

A screenshot of a computer

Description automatically generated with low confidence

## Grid Search

Graphical user interface, text, chat or text message

Description automatically generated

### Code

Graphical user interface, text, application, email

Description automatically generated

## Summary

Graphical user interface, text, application

Description automatically generated

Check ipynb notebook for 5\_Model Evaluation