



Extract, Transform, Load

Data Boot Camp

Lesson 13.1



The background is a dark charcoal gray with a series of parallel diagonal lines running from the top-left to the bottom-right. Overlaid on this are several teal-colored geometric shapes: a large central triangle pointing right, a smaller triangle to its left, and a square to its right. Scattered around these shapes are various white line-art symbols, including a plus sign, a minus sign, a circle with a dot, a circle with a horizontal line, a circle with a vertical line, a circle with a diagonal line, a circle with a cross, a circle with a dot, a circle with a horizontal line, a circle with a vertical line, a circle with a diagonal line, a circle with a cross, a circle with a dot, a circle with a horizontal line, a circle with a vertical line, a circle with a diagonal line, and a circle with a cross.

WELCOME

Class Objectives

By the end of today's lesson, you'll be able to:



Extract data by using Python and Pandas.



Transform and clean data by using Python and Pandas.



Parse string data into a Python dictionary.



Use list comprehensions to make code more readable.



Use regular expressions to manipulate string data.



Instructor Demonstration

Introduction to ETL

Introduction to ETL

ETL: Extract, Transform, and Load



Introduction to ETL: Extract

Data may come from disparate sources, such as:



CSV files



JSON files



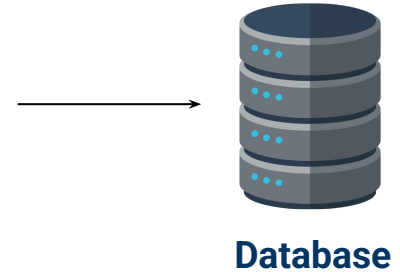
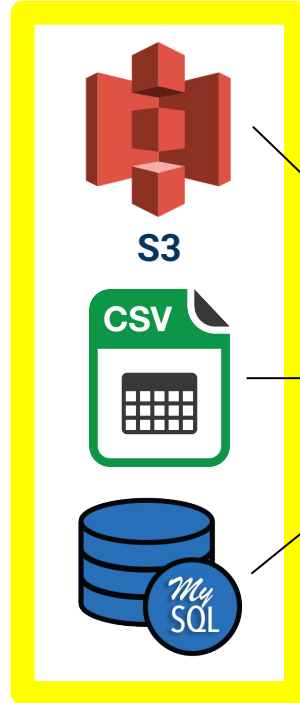
HTML tables



SQL databases



Spreadsheets



Extract

Introduction to ETL: Transform

Transform the data to suit business needs, including:



Data cleaning



Summarization



Selection



Joining



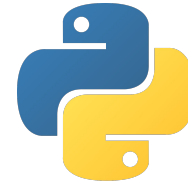
Filtering



Aggregating



S3



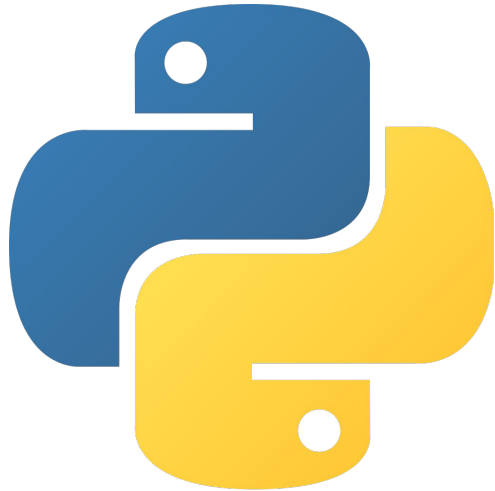
Python



Database

Transform

Note: We will use Python and Pandas for transformation, which can also be done with SQL or a specialized ETL tool.



Introduction to ETL: Load

Load the data into a final database that can be used for future analysis or business applications:



Can be a relational or non-relational database



Can be local or in the cloud



Can be a data lake or data warehouse



S3



Python



Database

Load

Questions?



Data Transformation and Cleaning



Instructor Demonstration

Data Transformation and Cleaning

Questions?





Activity: Transform and Clean Online Orders

In this activity, you'll transform and clean a dataset that consists of online orders.

Suggested Time:

15 minutes



Time's Up! **Let's Review.**

Questions?



List Comprehensions



Instructor Demonstration

List Comprehensions Review

Questions?





Activity: List Comprehensions Review

In this activity, you'll practice using list comprehensions.

Suggested Time:

15 minutes



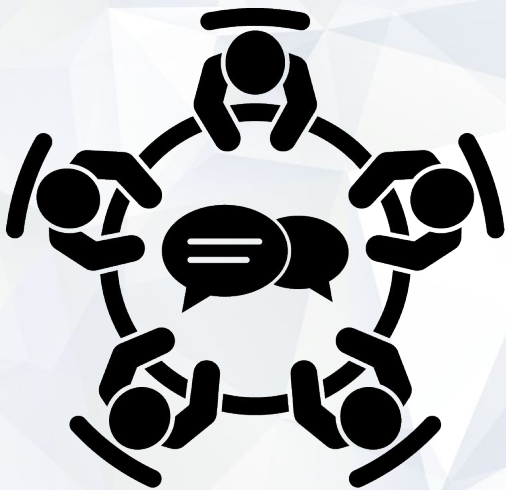
Time's Up! **Let's Review.**

Questions?





Break



Group Programming Activity:

Transform and Clean Grocery Orders

In this activity, you'll transform and clean grocery order data and then merge the data with another dataset.

Suggested Time:

15 Minutes

Questions?



Basic Regex Pattern Matching



Instructor Demonstration

Basic Regex Pattern Matching

Questions?





Group Programming Activity:

Regex Matching with Pandas

In this activity, you'll load a text dataset from *The Adventures of Sherlock Holmes* and then use regular expressions to find matching text.

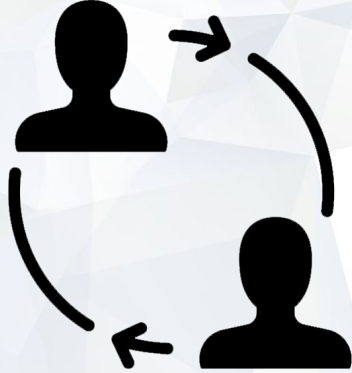
Suggested Time:

15 Minutes

Questions?



ETL Mini Project



Partner Activity: ETL Mini Project

For this mini project, you'll practice building an ETL pipeline by extracting and transforming a crowdfunding dataset and then loading the data into a PostgreSQL database.

Suggested Time:

20 minutes

Questions?



*The
End*