

Data Collection and Labeling

CS 203: Software Tools and Techniques for AI

Module Overview

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Why Does Data Matter?

Part 1: Data Collection

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart Labeling	Tabular Data:
Conversion	59	SMOTE
What is Data Labeling?	61	Exercise: Create Your Augmentation Pipeline
Image Classification: Labeling Cats vs Dogs	65	77
Text Classification Example	67	Key Takeaways
Label Studio: Professional Tool	69	81
Inter-Annotator Agreement	71	Tools Summary
	73	83
		Common Pitfalls to Avoid
		85
		Your Assignment
		87
		Resources for Further Learning
		89
		Thank You!
		91

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Sampling Strategies

Part 2: Data Validation

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart Labeling	Tabular Data:
Conversion	59	SMOTE
What is Data Labeling?	61	Exercise: Create Your Augmentation Pipeline
Image Classification: Labeling Cats vs Dogs	65	77
Text Classification Example	67	Key Takeaways
Label Studio: Professional Tool	69	81
Inter-Annotator Agreement	71	Tools Summary
	73	83
		Common Pitfalls to Avoid
		85
		Your Assignment
		87
		Resources for Further Learning
		89
		Thank You!
		91

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart Labeling	Tabular Data:
Conversion	59	SMOTE
What is Data Labeling?	61	Exercise: Create Your Augmentation Pipeline
Image Classification: Labeling Cats vs Dogs	65	77
Text Classification Example	67	Key Takeaways
Label Studio: Professional Tool	69	81
Inter-Annotator Agreement	71	Tools Summary
	73	83
		Common Pitfalls to Avoid
		85
		Your Assignment
		87
		Resources for Further Learning
		89
		Thank You!
		91

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart Labeling	Tabular Data:
Conversion	59	SMOTE
What is Data Labeling?	61	Exercise: Create Your Augmentation Pipeline
Image Classification: Labeling Cats vs Dogs	65	77
Text Classification Example	67	Key Takeaways
Label Studio: Professional Tool	69	81
Inter-Annotator Agreement	71	Tools Summary
	73	83
		Common Pitfalls to Avoid
		85
		Your Assignment
		87
		Resources for Further Learning
		89
		Thank You!
		91

Data Type Conversion

Part 3: Data Labeling

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Weak Supervision Example

Part 4: Data Augmentation

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart Labeling	Tabular Data:
Conversion	59	SMOTE
What is Data Labeling?	61	Exercise: Create Your Augmentation Pipeline
Image Classification: Labeling Cats vs Dogs	65	77
Text Classification Example	67	Key Takeaways
Label Studio: Professional Tool	69	81
Inter-Annotator Agreement	71	Tools Summary
	73	83
		Common Pitfalls to Avoid
		85
		Your Assignment
		87
		Resources for Further Learning
		89
		Thank You!
		91

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Exercise: Create Your Augmentation Pipeline

Summary & Best Practices

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	

Outline

What We'll Learn Today	3	Web Analytics: Google Analytics Example . . .	17	Common Data Problems	33
Why Does Data Matter?	5	Web Scraping: Collecting Public Data	19	Data Validation with Python	35
Real-World Example: Building a Chatbot	9	Exercise: Scrape a Simple Website	21	Pydantic: Automatic Validation	37
Python Logging: Your First Tool	11	API Data Collection ..	23	Exercise: Create Your Own Validator	39
Exercise: Add Logging to Your Code	13	Database Logging Example	25	Pandas Data Validation	41
Structured Logging with JSON	15	Sampling Strategies .	27	Handling Missing Data	43
		Why Validate Data? . . .	31		

Data Type	Active Learning: Smart	Tabular Data:
Conversion	Labeling 59	SMOTE 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling?	Example 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline 77
Labeling Cats vs	Image Augmentation	Key Takeaways 81
Dogs	with PIL 67	Tools Summary 83
Text Classification	Image Augmentation	Common Pitfalls to
Example	with OpenCV 69	Avoid 85
Label Studio:	Text Augmentation:	Your Assignment 87
Professional Tool	Synonym	Resources for Further
Inter-Annotator	Replacement 71	Learning 89
Agreement	Text Augmentation:	Thank You! 91
	Back Translation 73	