

# CS 203: Software Tools and Techniques for AI



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Why Does Data Matter?

# Part 1: Data Collection



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . 87
Professional Tool . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Sampling Strategies

# Part 2: Data Validation

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Data Type Conversion

# Part 3: Data Labeling



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . 87
Professional Tool . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Weak Supervision Example

# Part 4: Data Augmentation



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	<b>Image Augmentation</b>	Key Takeaways ..... 81
Dogs ..... 51	<b>with PIL ..... 67</b>	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

## Exercise: Create Your Augmentation Pipeline

# Summary & Best Practices



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	<b>Key Takeaways ..... 81</b>
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . 87
Professional Tool . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	<b>Tools Summary ..... 83</b>
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . 87
Professional Tool . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	<b>Common Pitfalls to</b>
Example ..... 53	with OpenCV ..... 69	<b>Avoid ..... 85</b>
Label Studio:	Text Augmentation:	Your Assignment . . . 87
Professional Tool . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	<b>Your Assignment . . . . 87</b>
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	



# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	<b>Resources for Further</b>
Inter-Annotator	Replacement ..... 71	<b>Learning ..... 89</b>
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	

# Outline

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

Data Type	Active Learning: Smart	Tabular Data:
Conversion ..... 45	Labeling ..... 59	SMOTE ..... 75
What is Data	Weak Supervision	Exercise: Create Your
Labeling? ..... 49	Example ..... 61	Augmentation
Image Classification:	Why Augment Data? . 65	Pipeline ..... 77
Labeling Cats vs	Image Augmentation	Key Takeaways ..... 81
Dogs ..... 51	with PIL ..... 67	Tools Summary ..... 83
Text Classification	Image Augmentation	Common Pitfalls to
Example ..... 53	with OpenCV ..... 69	Avoid ..... 85
Label Studio:	Text Augmentation:	Your Assignment . . . . 87
Professional Tool . . . . 55	Synonym	Resources for Further
Inter-Annotator	Replacement ..... 71	Learning ..... 89
Agreement ..... 57	Text Augmentation:	Thank You! ..... 91
	Back Translation ..... 73	