O'REILLY®

GenAI Powered Data Analysis with Python Bootcamp

Dr. Chester Ismay

Welcome and Introduction

Dr. Chester Ismay

- PhD in Statistics
- Worked in academia, online education, corporate training, tech bootcamps, and independent consulting
- Currently, freelance data scientist and educator
- Fun Fact: Slept a night or eaten a meal in all 50 US states

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Learning Objectives

By the end of this course, you will be able to:

- Build proficiency with Python libraries (Pandas, Matplotlib, Seaborn, Plotly) to wrangle, analyze, and visualize data.
- Apply exploratory data analysis to clean, summarize, and transform datasets, revealing key insights.
- Leverage Generative AI to speed up learning, debug code, and refine workflows through prompt engineering.
- Create interactive visualizations that clearly communicate complex information.
- Use AI-assisted summarization and storytelling to craft compelling data narratives for stakeholders.

Day 1 (Prompt to Wrangle) Agenda

- Intro: Data Analytics Kickoff + Course Goals
- Module 1: Pandas for Data Wrangling
- Module 2: Transforming and Aggregating Data with Pandas
- Module 3: Exploring and Learning from Mistakes
- Wrap-Up & Reflection

Day 2 (Visualize to Tell) Agenda

- Module 4: Visualizing with Matplotlib & Seaborn
- Module 5: Interactive Plotly Visuals
- Module 6: Storytelling with Data with Support from GenAI
- How GenAI Solves Problems
- Course close

Intro

Data Analytics Kickoff + Course Goals

What is Data Analysis?

- Examining, cleaning, transforming, and interpreting data
- Deriving insights
- Improving decision-making

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

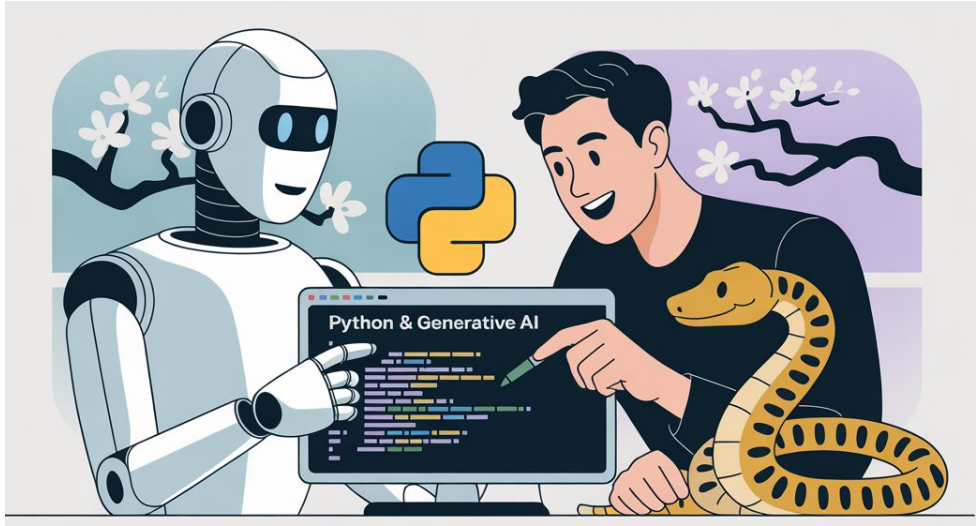
Python for Data Analysis

- Powerful programming language
- Popular in data fields
- Rich ecosystem of libraries
  - Pandas
  - Matplotlib
  - Seaborn
  - Plotly

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Generative AI for Assistance

- 24/7 learning companion
- Code debugging assistant
- Can explain complex concepts in simpler terms



Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Framing GenAI: Your Thinking Partner

Not Magic

It's a tool that complements your skills, not replaces them

Ask Clearly

Specific prompts yield better results than vague requests

Versatile Helper

Explains concepts, simplifies code, guides exploration

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Setting Up Python with GenAI

"Help me install pandas and matplotlib on my system. I'm using [Windows/Mac/Linux] and getting this error: [paste error message]"

"I'm getting an import error with seaborn. Can you walk me through troubleshooting this step by step?"

"What are the essential packages for data analysis in Python? Show me how to check if they're installed correctly."

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Walkthrough and Exercise #1
Setting Up the Python Environment

By completing this exercise, you will be able to

1. Import necessary Python packages
2. Check for successful package loading

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Day 1: Prompt to Wrangle

Data Wrangling

with Pandas

Pandas for Handling Data

- `pandas`: flexible and efficient data wrangling tool
- Key features: Series and DataFrame
- Works with data from various sources

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Pandas for Cleaning and Preparing Data

- Handling missing data
- Converting data types
- Renaming columns
- Changing index
- Filtering

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Using LLMs for Code Help

Explain & Elaborate

"Can you walk me through this `df.groupby(...).agg(...)` line step-by-step? What's happening under the hood here?"

Refactor & Improve

"I have this nested loop for data processing. Is there a more Pythonic or efficient way to achieve the same result, perhaps using pandas vectorization?"

Debug & Resolve

"I'm encountering a `KeyError` when trying to access a column in my DataFrame. I've checked the column name, and it seems correct. What might be going on with my data loading?"

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Walkthrough and Exercise #2
Cleaning and Preparing Data with Pandas

By completing this exercise, you will be able to use `pandas` to

1. Import data from a CSV or from an Excel file
2. Perform an initial exploration of the data
3. Handle missing data
4. Convert a column to a different data type
5. Rename a column
6. Filter a DataFrame

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Transforming and Aggregating Data

with Pandas

Data Transformation and Aggregation with Pandas

- Applying functions
- Grouping data
- Creating pivot tables
- Analyzing categorical data

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

LLM Conversation Strategies

Explain to a Novice

Ask ChatGPT to simplify complex data analysis concepts or explain them as if you're entirely new to the topic. For example: "Explain the `groupby()` function in Pandas as if I've never used it before, with a simple example."

Multiple Solutions

Request various ways to solve a data analysis problem or different code approaches, along with their pros and cons. For example: "Show me 3 different ways to perform data aggregation in Python and explain the tradeoffs of each."

Iterative Refinement

Engage in a dialogue by asking clarifying or deeper questions based on ChatGPT's previous responses to build understanding. "Can you explain why [specific part of the code/explanation] works that way, or provide an alternative approach?"

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Walkthrough and Exercise #3
Summarizing Data with Pandas

By completing this exercise, you will be able to use `pandas` to

1. Aggregate data effectively by grouping it
2. Transform data by applying functions element-wise or to groups
3. Create summary tables
4. Analyze categorical data using cross-tabulation

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Exploring and Learning from Mistakes

Some Common pandas Pitfalls

How LLMs can support

- Index issues explained
- Missing value debugging
- Type error clarification

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Debugging Tips

- Read from bottom to top
- Confirm column names
- Check variable types/spelling

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Walkthrough and Exercise #4
Debug with GenAI

By completing this exercise, you will be able to fix broken code with AI assistance by

1. Pasting the buggy code into the chatbot
2. Prompting the LLM to fix the code
3. Reviewing the completion provided by the LLM
4. Iterating until explanations from the AI model are clear

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Reflection: Learning from Experience

Share your insights in the chat!

1. What specific prompts or prompting techniques yielded the most valuable insights for you?

2. What made some explanations clearer than others?

3. How will you refine your questions next time?

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Day 1 Recap

pandas basics

Loading, inspecting, and transforming data

Data summarization

groupby(), agg(), and custom calculations

GenAI assistance

Explaining concepts, debugging code, suggesting improvements

You now have the foundation for efficient data wrangling!

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Take Home Exercise A (Optional)
Prompt Practice

By completing this exercise, you will be able to ask AI for guidance on any of the following questions:

1. "Suggest 3 different ways to summarize this dataset"
2. "What 3 interesting questions could we investigate with this data?"
3. "What visualization would best show the relationship between **X** and **Y**?"

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Looking Ahead to Day 2

Visualization

Turn numbers into compelling stories

Interactivity

Build exploratory dashboards

Narrative

Craft insights that drive decisions

Challenge: Use GenAI to generate sample plots in matplotlib, seaborn, or plotly before tomorrow!

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Day 2: Visualize to Tell

Data Visualization Basics

with Matplotlib and Seaborn

Fundamentals of Data Visualization with Matplotlib

- Importance of data visualization
- `matplotlib`: versatile library for static plots
- Common chart types
- Customization options

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Enhancing Visualizations with Seaborn

- `seaborn` vs. `matplotlib`
- Advanced visualizations (heatmaps, pair plots, violin plots)
- Color palettes and themes

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Walkthrough and Exercise #5
Data Visualization Techniques

By completing this exercise, you will be able to use `matplotlib` and `seaborn` to

1. Create line plots and bar charts
2. Add labels and titles
3. Adjust axes and tick marks
4. Create histograms
5. Design boxplots and violin plots

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Choosing Plots with GenAI

Ask for guidance

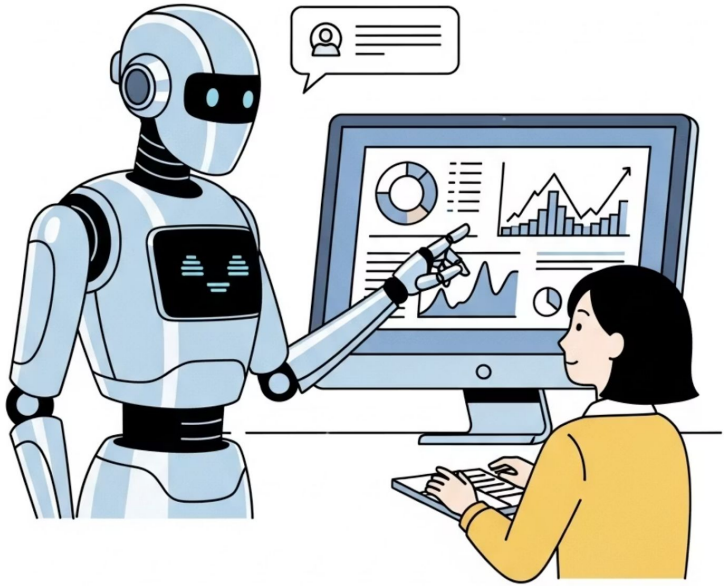
"What type of plot best shows the relationship between sales and customer age?"

Get explanations

"Explain what this boxplot is showing about our data distribution"

Create headlines

"Suggest a headline that captures the main insight from this chart"



Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Interactive Data Visualization

with Plotly

Building Data Graphics with Plotly

- `plotly`: interactive visualization library
- Benefits of interactive visualizations
- Overview of Plotly's features and capabilities

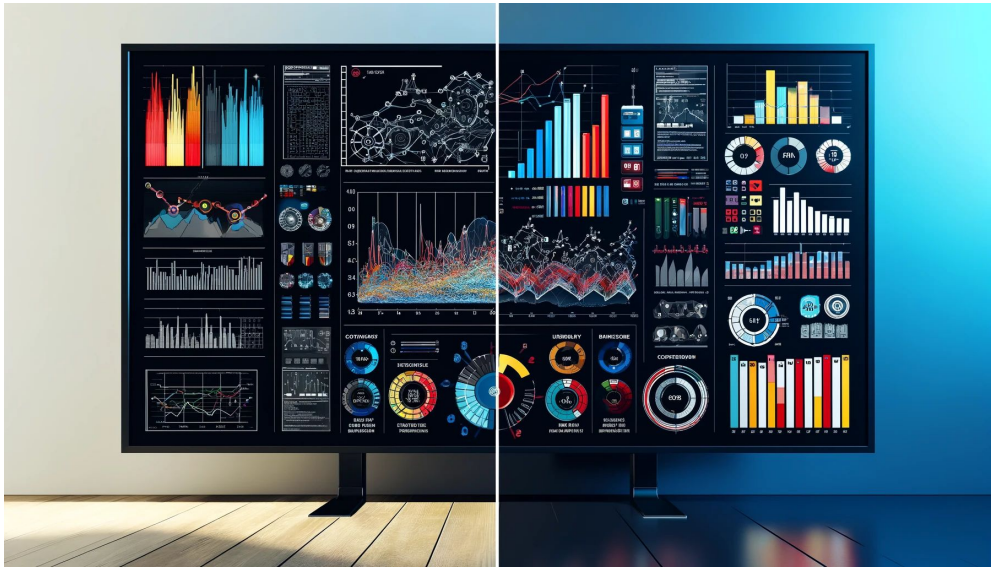
Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Audience-specific visualizations

- Match visuals to audience expertise
- Focus on metrics that matter to them
- Adjust detail and style for context

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Customizing Plotly Visuals for Different Audiences



- Customizing colors, layouts, and annotations
- Using templates and themes for consistency

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Walkthrough and Exercise #6
Interactive Charts and Dashboards with Plotly

By completing this exercise, you will be able to use `plotly` to

1. Create a basic interactive chart
2. Add interactive elements: hover, zoom, and selection tools
3. Design a simple dashboard with multiple charts

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Improving Visuals with GenAI

How can I make this chart more executive-friendly? Focus on the key message and remove unnecessary details.

Suggest 3 design improvements for clarity. What colors, fonts, or layout changes would help?

Write a compelling headline and caption that explains the key insight from this visualization.

GenAI can be your design consultant and clarity coach

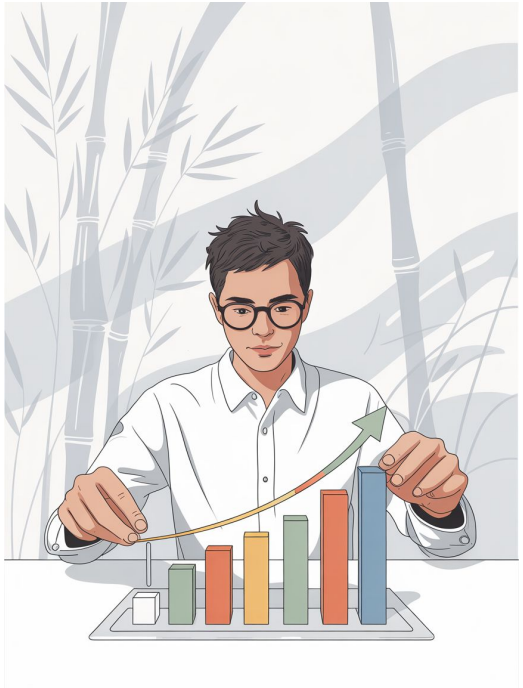
Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Storytelling with Data with Support from GenAI

What Makes Insights Compelling

- Show comparisons
- Add context
- Align with story

Numbers alone don't persuade—stories do



Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

From Insight to Story

- Translate findings into clear, audience-focused messages
- Use visuals and context to highlight the “so what”
- Frame insights as part of a larger narrative leading to decisions

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Storytelling with Data - How to Present Findings

- Importance of storytelling in data presentation
- Techniques for effective data storytelling
- Tailoring the story to your audience

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Alternate Narratives

For Executives

"Sales increased 24% in Q3, driven primarily by the new product line, suggesting our innovation strategy is working."

For Data Team

"The sales data shows statistically significant growth ($p<0.01$) across all segments, with strongest correlation to our marketing spend."

For Customers

"You're part of something big! Our community grew by 10,000 members this quarter, all enjoying our new features."

Same data, different emphasis!

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Walkthrough and Exercise #7
Craft multiple stories

By completing this exercise, you will be able to use GenAI to craft multiple story angles from your data. You'll use an analysis from earlier exercises as the foundation. Choose each of these three to prompt and then review the differences.

1. Executive Summary (Brief, action-oriented, focuses on business impact)
2. Technical Deep-Dive (Methodology, statistical significance, limitations)
3. Marketing Copy (Engaging, benefit-focused, customer-centric)

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

How GenAI Solves Problems

How GenAI Works (High-Level)

The Basics

- Trained on vast text data to predict what comes next
- Recognizes patterns, not facts
- No actual understanding or reasoning

Key Capabilities

- Speed (instant responses)
- Creativity (novel combinations)
- Exploration (tries many paths)

Limitations

- Accuracy
- Bias

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Benefits and Boundaries

Coding Partner

Suggests approaches, explains errors, simplifies complex code

Storytelling Assistant

Helps articulate insights, reframes for different audiences

Not the Source of Truth

Verify factual claims, especially about your data

Critical Thinking Required

You're still the expert on your analysis context

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Take Home Exercise B
Real-World Prompt

By completing this exercise, you will be able to

1. Create a prompt template for your actual work
2. Have a conversation with GenAI to solve an actual problem
3. Feel confident that you can reliably test the output/completion for reliability and correctness

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Reflection Time

1. Which part of your workflow benefits most from GenAI assistance?
2. What prompt template will you use regularly?
3. How will you balance AI assistance with your own expertise?

Share your insights with the class in the chat!

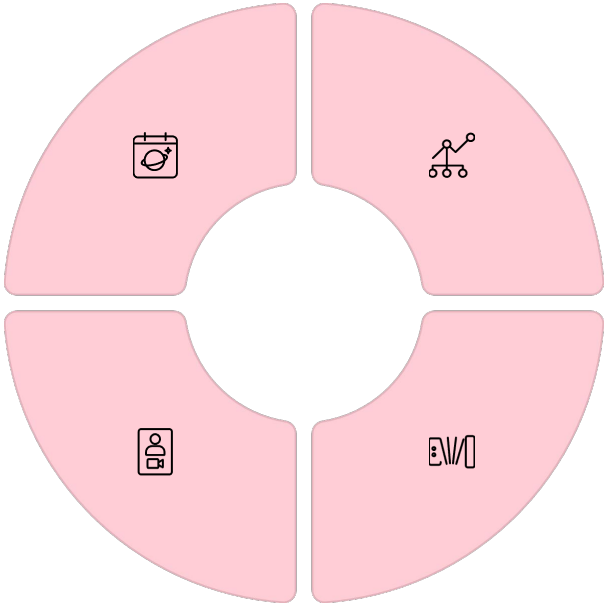
Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Course Close

Technical Recap

pandas

Data wrangling and transformation



matplotlib & seaborn

Statistical visualization fundamentals

Storytelling

Turning insights into action

plotly

Interactive dashboards and exploration

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

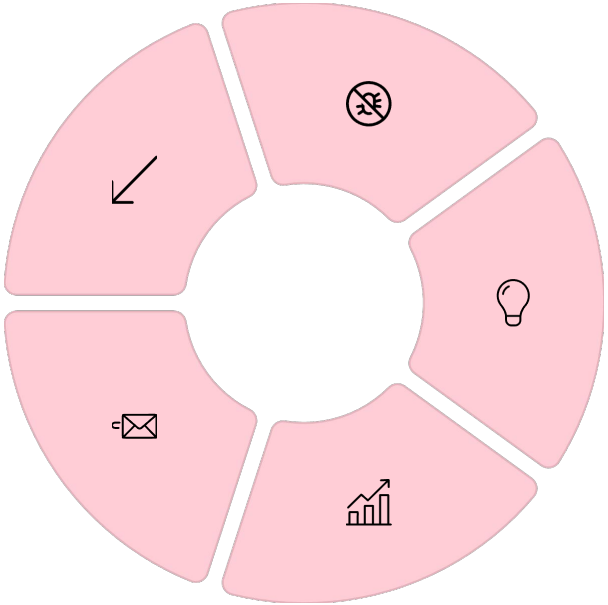
GenAI Recap

Effective Prompting

Crafting clear, specific questions for better AI responses

Clear Communication

Articulating insights for different audiences



AI-Assisted Debugging

Using GenAI to identify and fix code errors

Alternative Approaches

Exploring multiple solutions and methodologies

Visual Enhancement

Refining charts and visualizations for clarity

You now have a powerful partner in your data workflow

Course repo: https://github.com/ismayc/oreilly-genai-powered-data-analysis-with-python

Learning Objectives

By the end of this course, you will be able to:

- Build proficiency with Python libraries (Pandas, Matplotlib, Seaborn, Plotly) to wrangle, analyze, and visualize data.
- Apply exploratory data analysis to clean, summarize, and transform datasets, revealing key insights.
- Leverage Generative AI to speed up learning, debug code, and refine workflows through prompt engineering.
- Create interactive visualizations that clearly communicate complex information.
- Use AI-assisted summarization and storytelling to craft compelling data narratives for stakeholders.

Resources & References

Prompt Templates

- Code explanation: "Explain this pandas code line by line..."
- Debugging: "This code gives error X. What's wrong?"
- Visualization: "What's the best plot type for showing..."

Further Learning

"Storytelling with Data" by Cole Knaflic

"Python for Data Analysis" by Wes McKinney

"Prompt Engineering Guide" by dair.ai

Conclusion

Additional resources:

- pandas
- matplotlib
- seaborn
- plotly

LinkedIn: https://www.linkedin.com/in/chesterismay/

Personal website: https://chester.rbind.io/

Images generated with ChatGPT/Imagen or available via Creative Commons Google Image search

Final Reflection & Wrap-Up

1. What will you try first with Python + GenAI?

2. What was your biggest "aha" moment?

3. How will this change your workflow?

Thank you for joining us!


O'REILLY®