O'REILLY®

Data Analysis with Python

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,
  - Vice President of Data and Automation, MATE Seminars
  - Freelance data scientist
- Fun Fact: Slept a night or eaten a meal in all 50 US states

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

Learning Objectives

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

- Utilize Python's data wrangling tools to clean and prepare data, making it ready for analysis and visualization.
- Design and implement a variety of charts and plots that effectively communicate your data's story.
- Conduct explorations of your datasets, identifying key trends, patterns, and outliers.
- Leverage your analytical and visualization skills to support informed decision-making related to organizational/business goals

Agenda

- Intro: Foundations of Data Analysis with Python
- Module 1: Data Wrangling with Pandas
- Module 2: Data Visualization Basics with Matplotlib and Seaborn
- Module 3: Advanced Data Visualization with Plotly
- Module 4: Review and Effective Data Storytelling

Intro

Foundations of Data Analysis with Python

What is Data Analysis?

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

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

Practical Application

- Data analysis approach
- Insights derived
- Informed decision-making
- Impact

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

Discussion/Poll Question #1 (For On24)
What are your most looking forward to in the course?

1. **Fundamental Understanding**: Gain a basic understanding of data wrangling and visualization with Python libraries.
2. **Hands-on Practice**: Apply theoretical knowledge through hands-on exercises and case studies.
3. **Tool Proficiency**: Become proficient in using Pandas, Matplotlib, Seaborn, and Plotly for different stages of data analysis.
4. **Effective Storytelling**: Develop skills to create visual data narratives that clearly communicate insights to stakeholders.
5. **Other**

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

Questions and Answers

Anything I can clear up regarding the *Intro* content?

Data Wrangling

with Pandas

Discussion/Poll Question #2 (For On24)
Which of the following do you think are key characteristics of data wrangling with pandas? (Select all that apply)

1. Cleaning data by handling missing values and errors
2. Performing complex mathematical operations on datasets
3. Merging and concatenating multiple datasets
4. Enhancing the graphical interface of Python scripts
5. Using pandas to directly manipulate HTML content

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

Pandas for Handling Data

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

Walkthrough and Exercise #2
Loading and Inspecting 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.

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

Pandas for Cleaning and Preparing Data

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

Walkthrough and Exercise #3
Cleaning and Preparing Data with Pandas

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

1. Handle missing data
2. Convert a column to a different data type
3. Rename a column
4. Change a DataFrame's index
5. Filter a DataFrame

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

Data Transformation and Aggregation with Pandas

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

Walkthrough and Exercise #4
Transforming and Aggregating 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-data-analysis-with-python

Questions and Answers

Anything I can clear up regarding the
Module 1: Data Wrangling with Pandas content?

Data Visualization Basics

with Matplotlib and Seaborn

Discussion/Poll Question #3 (For On24)
What do you think is the primary purpose of data visualization?

1. To make complex data look simple and remove all complicated details
2. To create visually appealing reports that prioritize aesthetics
3. To facilitate the understanding of complex data by presenting it in a graphical format
4. To completely replace traditional methods of data analysis

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

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-data-analysis-with-python

Walkthrough and Exercise #5
Creating Basic Plots with Matplotlib

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

1. Create line plots and bar charts
2. Add labels and titles
3. Adjust axes and tick marks

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

Walkthrough and Exercise #6
Data Visualization Techniques with Seaborn

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

1. Create heatmaps
2. Design pair plots and violin plots
3. Customize Seaborn plots

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

Questions and Answers

Anything I can clear up regarding the
Module 2: Data Visualization Basics with Matplotlib and Seaborn
content?

Review of Modules 1 & 2

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-data-analysis-with-python

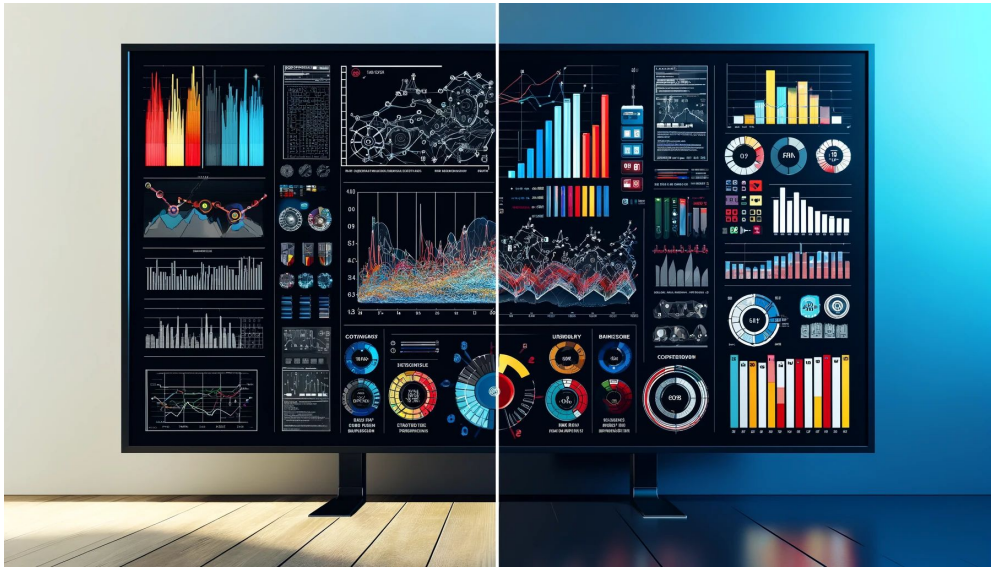
Walkthrough and Exercise #7
Interactive Charts and Dashboards with Plotly

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-data-analysis-with-python

Customizing Plotly Visuals for Different Audiences



- Importance of audience-specific visualizations
- Customizing colors, layouts, and annotations
- Using templates and themes for consistency

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

Walkthrough and Exercise #8
Creating a Dynamic Data Report

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

1. Select relevant data
2. Build a dynamic report
3. Add contextual text and summaries

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

Questions and Answers

Anything I can clear up regarding the *Module 3: Interactive Data Visualization with Plotly* content?

Review and Effective Data Storytelling

Discussion/Poll Question #4 (For On24)
Which of the following steps do you think comes first in the data analysis workflow?

1. Selecting the right tools and techniques for data analysis.
2. Defining clear, actionable questions that the analysis aims to answer.
3. Gathering all data available.
4. Extracting actionable insights from analyzed data.

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

Applying Your Skills – From Data to Insights

- Overview of the data analysis workflow
- Identifying and defining the analysis problem
- Selecting the right tools and techniques for data analysis

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

Walkthrough and Exercise #9
Working on a Real-World Data Analysis Project

By completing this exercise, you will be able to identify ways to best

1. Select a dataset
2. Apply cleaning, transforming, and analysis techniques
3. Review initial findings and interpretation

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

Walkthrough and Exercise #10
Finalizing and Presenting Your Data Analysis Project

By completing this exercise, you will be able to identify the best strategies to

1. Integrate feedback to refine the analysis
2. Finalize the presentation with impactful visuals and narrative
3. Rehearse the presentation

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

Review of Modules 3 & 4

Questions and Answers

Anything I can clear up regarding any of our content today?

Learning Objectives

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

- Utilize Python's data wrangling tools to clean and prepare data, making it ready for analysis and visualization.
- Design and implement a variety of charts and plots that effectively communicate your data's story.
- Conduct explorations of your datasets, identifying key trends, patterns, and outliers.
- Leverage your analytical and visualization skills to support informed decision-making related to organizational/business goals

Conclusion

Additional resources:

- pandas
- matplotlib
- seaborn
- plotly
- Fun game that uses some of the data we used in this course: https://www.geogridgame.com/

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

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

Images generated with DALL-E or available via Creative Commons Google Image search


O'REILLY®