AI Boot Camp

Visualizing DataFrames Using Pandas

Module 7 Day 2

Class Objectives

By the end of class, you will be able to:

1. Create plots by using the `DataFrame.plot()` method.
2. Explain the advantages and disadvantages of creating charts using the `DataFrame.plot()` method.
3. Analyze a complex dataset using Pandas and data visualizations.

Activity:

Pyplot Warmup

Use Pyplot to create a bar chart that shows the average rainfall in different US states.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Questions?

Instructor **Demonstration**

Plotting Pandas Data

Plotting Pandas Data

No matter your data source,

you will almost always need to clean and process your data before analysis, typically by using Pandas. Therefore, we can expect that most real-world data that we analyze from now on will come from within a Pandas DataFrame.

The creators of Pandas

realized that most people using Pandas would then visualize their plots using Matplotlib. In a moment of pure genius, they built Matplotlib methods into their library to allow data analysts to easily generate complex charts in very little time.

Activity:

Union Settlements

Use a dataset about major collective bargaining settlements to create a bar chart.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Instructor **Demonstration**

Plotting Groups

Do you remember how to
group data using Pandas?

We can group and summarize data by using the Pandas `groupby()` function. The output of this is a GroupBy object.

Activity:

Library Usage: Grouped Charts

Create a pair of charts based on library usage collected from San Francisco.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Break

15 mins

Activity:

Miles per Gallon: Scatter Plot

Create a scatter plot by using vehicle data, Pandas, and Matplotlib.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Instructor **Demonstration**

Plotting Multiple Lines

Activity:

Traveling Companions, Part 1

In this activity, you will take three separate CSVs that were gathered from Tourism Malaysia, merge them together, and then create charts to visualize a comparison of travelers to Malaysia from different countries of origin over three years. This is Part 1 of a three-part mini project.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Activity:

Traveling Companions, Part 2

In this activity, you will examine the averages of each column, decide which columns they want to keep in their DataFrame for analysis, and set the index.This is Part 2 of a three-part mini project.

Suggested Time:
5 Minutes

Time’s up!

Let’s review

Activity:

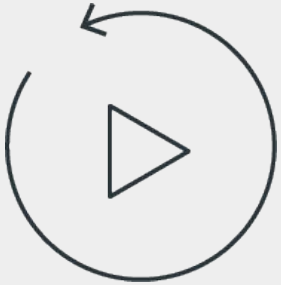
Traveling Companions, Part 3

In this activity, you will take the DataFrame they created and, using Matplotlib, chart a comparison of three different countries for one type of traveling companion between 2016 and 2018. This is the third and final part of a three-part mini project.

Suggested Time:
15 Minutes

Time’s up!

Let’s review



Let’s **recap**

Recap

In this lesson, you learned how to:

1. Create plots by using the DataFrame.plot() method.
2. Explain the advantages and disadvantages of creating charts by using the DataFrame.plot() method.
3. Analyze a complex dataset using Pandas and data visualizations.

Next

In the next class, you will continue exploring data by using visualizations. You will also revisit important statistical concepts used to characterize data.

Questions?

The End