

Pandas II

Learning Objectives

After this lesson, you will be able to:

- Identify and handle missing values with Pandas.
 - Implement groupby statements for specific segmented analysis.
 - Use apply functions to clean data with Pandas.
-

To the notebook!

We actually will commence this lesson directly in the Jupyter Notebook, `pandas-ii.ipynb`, to walkthrough the what, why, and how all at once.

Nonetheless, below, we have included slides reviewing the key concepts.

Teaching Tips:

- Go through the lesson in the notebook first; come back to this presentation for a recap of concepts afterwards.
-

How do we handle missing data?

To handle missing data, we must:

- Identify we have missing data from our DataFrame
- Determine, to the best of our ability, the cause of this missingness
- Justify how we will handle the missing data (drop or fill in with a specific value?)

Pro tip: The faster you understand *why* some observations are missing, the faster and more accurately you can handle them.

Teaching Tips:

- Do a quick recap and check for understanding.
-

A Note on Delivery

- This unit’s lessons will occur in [jupyter notebooks](#)
 - Slides will be an introduction to the lesson (no code, just overview)
 - Then, we’ll open a notebook and start coding!

Teaching Tips: - We could have made this into a speaker note, but it’s helpful to get it out there so everybody’s on the same page - No repl.it for this unit as we’ll be in notebooks

Key Pandas functions for missing data: