Course Description

A vital component of data science involves acquiring raw data and getting it into a form ready for analysis. In fact, it is commonly said that data scientists spend 80% of their time cleaning and manipulating data, and only 20% of their time actually analyzing it. This course will equip you with all the skills you need to clean your data in Python, from learning how to diagnose your data for problems to dealing with missing values and outliers. At the end of the course, you'll apply all of the techniques you've learned to a case study in which you'll clean a real-world Gapminder dataset!

1 Exploring your data FREE

0%

So you've just got a brand new dataset and are itching to start exploring it. But where do you begin, and how can you be sure your dataset is clean? This chapter will introduce you to the world of data cleaning in Python! You'll learn how to explore your data with an eye for diagnosing issues such as outliers, missing values, and duplicate rows.

VIEW CHAPTER DETAILS

Continue Chapter

2 Tidying data for analysis

0%

Here, you'll learn about the principles of tidy data and more importantly, why you should care about them and how they make subsequent data analysis more efficient. You'll gain first hand experience with reshaping and tidying your data using techniques such as pivoting and melting.

VIEW CHAPTER DETAILS

Continue Chapter

3 Combining data for analysis

0%

The ability to transform and combine your data is a crucial skill in data science, because your data may not always come in one monolithic file or table for you to load. A large dataset may

be broken into separate datasets to facilitate easier storage and sharing. Or if you are dealing with time series data, for example, you may have a new dataset for each day. No matter the reason, it is important to be able to combine datasets so you can either clean a single dataset, or clean each dataset separately and then combine them later so you can run your analysis on a single dataset. In this chapter, you'll learn all about combining data.

VIEW CHAPTER DETAILS

Continue Chapter

4 Cleaning data for analysis

0%

Here, you'll dive into some of the grittier aspects of data cleaning. You'll learn about string manipulation and pattern matching to deal with unstructured data, and then explore techniques to deal with missing or duplicate data. You'll also learn the valuable skill of programmatically checking your data for consistency, which will give you confidence that your code is running correctly and that the results of your analysis are reliable!

VIEW CHAPTER DETAILS

Continue Chapter

5 Case study

0%

In this final chapter, you'll apply all of the data cleaning techniques you've learned in this course towards tidying a real-world, messy dataset obtained from the Gapminder Foundation. Once you're done, not only will you have a clean and tidy dataset, you'll also be ready to start working on your own data science projects using the power of Python!

VIEW CHAPTER DETAILS