

## **Course Overview**

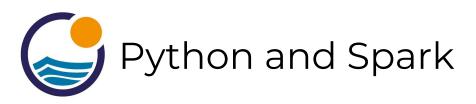
Let's learn something!





- Welcome to the course!
- In this lecture we will quickly cover some FAQs and what to expect during the course!
- Please don't skip this lecture! It will help you understand the course!





- Let's discuss a few things that will help give you the best course experience possible!
- Lectures can be speed up to your liking up to 2x and they can also be viewed at higher quality, click on the settings icon.



- Who is this course for and what does it cover?
  - Some experience with programming.
  - Interested in using Spark and Python for analyzing big data.





- If you are interested in just general Data Science and Machine Learning with Python, check out my other course:
  - "Python for Data Science and Machine Learning Bootcamp"



- This course is ideal for students also enrolled in that course, or with some general knowledge of basic data analysis with Python.
- Designed to add Spark and PySpark to your professional resume.





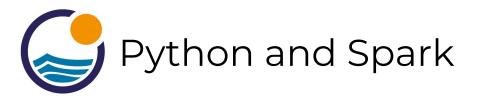
- What this course is not:
  - We don't cover typical Python libraries:
    - NumPy, Pandas, Visualization,
      Scikit-Learn, etc.
  - This course is for Big Data that those technologies can't handle.





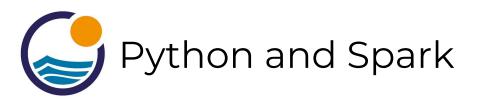
- This course is for people looking to build skills in:
  - PySpark DataFrames
  - Spark 2.0
  - Working with very large datasets
  - Spark specifics, like Spark
    Streaming





- Hopefully this gives you an idea of the target audience!
- A few more things before we get started!





 We have various installation lectures for different ways of setting up PySpark, don't feel obligated to use any of them if you already have your own setup.

- We use the Jupyter Notebook for the video lectures, but if you prefer .py files with an IDE or Text Editor that is no problem!
- I've also provided the .py files for you to follow along with, it is totally up to you!



- You can grab the notes as a zip file as a resource in this lecture (or in the FAQ lecture or the Installation overview)
- They are all the same .zip files, just put in several lectures for convenience, any of them will work.

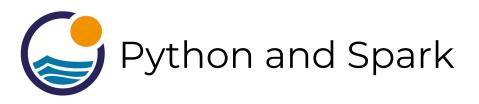
- The slides shown in the course are also linked to as a resource in their respective lecture.
- They are hosted as Google Slides you can just visit online to see any updates.



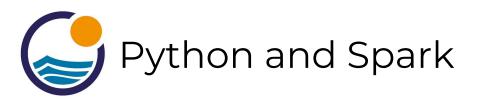
- How to get help during the course:
  - Double check that your code matches the notes exactly!
  - Do a quick check on StackOverflow for a fast answer!



- How to get help during the course:
  - Check the course FAQ lecture.
  - Search the QA forums of the course to see if someone else has already asked your question!



- How to get help during the course:
  - Post a new question to the QA forums if you still can't resolve the issue given the previous methods, we're happy to help out!



 Alright let's quickly cover the course curriculum!





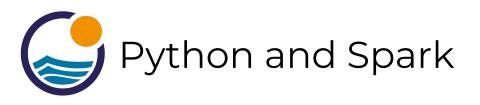
- We will cover the following topics:
  - Spark and Big Data basics.
  - Setting up Spark in various ways.
  - Python Crash Course
  - Python and Spark 2.0 DataFrames
  - PySpark Project Exercise





- Then the second half of the course:
  - Introduction to Machine Learning
  - Linear Regression
  - Logistic Regression
  - Decision Trees and Random Forests
  - Gradient Boosted Trees





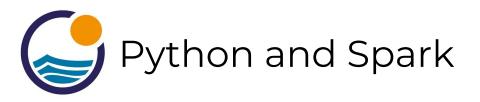
- Then the second half of the course:
  - K-means Clustering
  - Recommender Systems
  - Natural Language Processing
  - Spark Streaming (Local and Twitter)



- Within these topics we introduce other things, like Pipelines or Cross-Validation.
- Thanks for listening to all of that! It should make your course experience much smoother!



- One last thing A HUGE THANK YOU!
  - I truly appreciate you enrolling in this course!



- Let's get started learning one of the latest technologies:
  - Spark with Python

