

Python For Data Science Syllabus

Python for Data Science Syllabus

Welcome!

We are delighted to welcome you into the first course of the EdX / UC San Diego MicroMasters in Data Science: Python for Data Science. In this course, you will learn both the basics of conducting data science and how to perform data analysis in python.

Course Staff

Instructors

Ilikay Altintas, Chief Data Science Officer, San Diego Supercomputer Center (SDSC)
Leo Porter, Assistant Teaching Professor, Computer Science and Engineering Department

Teaching Assistants

Alok Singh, Computational Data Science Research Specialist, SDSC
Andrea Zonca, HPC Applications Specialist, SDSC

Prerequisites

This course is intended for learners who have a basic knowledge of programming in any language (Java, C, C++, Pascal, Fortran, Javascript, PHP, python, etc.). You could have learned these basic programming skills on your own or taken a course in programming in high school or college.

Your knowledge need not be extensive, but we'll assume you already know how to:

- Create and assign variables.
- Write programs with loops
- Write programs with conditions
- Author and use functions (methods)

If you are unfamiliar with python, we have an entire week (Week 2) dedicated to getting you up to speed with basic programming in python. If you find that Week 2 progresses too quickly and you need more help with basic programming, you may wish to try an introductory programming course in python before starting this course on Python for Data Science.

Course Overview

This course will introduce you to the field of data science and will prepare you for the next three courses in the MicroMasters: Statistics, Machine Learning, and Spark.

First, and foremost, you'll learn how to conduct data science by learning how to analyze data. That includes knowing how to import data, explore it, analyze it, learn from it, visualize it, and ultimately generate easily shareable reports. We'll also introduce you to two powerful areas of data analysis: machine learning and natural language processing.

To conduct data analysis, you'll learn a collection of powerful, open-source, tools including:

- Python

Python for Data Science syllabus serves as an essential roadmap for anyone looking to navigate the ever-expanding field of data science. With its extensive libraries and frameworks, Python has become the go-to language for data scientists. A well-structured syllabus not only helps learners acquire the necessary skills but also provides a clear pathway to mastering data science concepts. In this article, we will delve into the key components of a typical Python for Data Science syllabus, breaking down essential topics and resources that every aspiring data scientist should be familiar with.

Introduction to Python

Before diving into data science, it is crucial to have a solid foundation in Python

programming. This section typically covers:

- Basic syntax and data types (strings, integers, floats, lists, tuples, dictionaries)
- Control structures (if statements, loops)
- Functions and modules
- Error handling and exceptions

Resources for Learning Python

To get started with Python, you can utilize the following resources:

- Online courses (e.g., Codecademy, Coursera, Udemy)
- Books (e.g., "Automate the Boring Stuff with Python," "Python Crash Course")
- Interactive coding platforms (e.g., LeetCode, HackerRank)

Data Manipulation with Pandas

Pandas is a vital library for data manipulation and analysis in Python. This section will cover:

- Understanding DataFrames and Series
- Reading and writing data files (CSV, Excel, SQL)
- Data cleaning techniques (handling missing values, duplicates)
- Data transformation (filtering, grouping, merging, reshaping)

Hands-On Exercises

Practical exercises can help reinforce learning. Consider the following activities:

1. Loading a dataset and performing initial data exploration
2. Cleaning a messy dataset by handling null values and formatting issues
3. Aggregating data to derive meaningful insights

Data Visualization with Matplotlib and Seaborn

Data visualization is crucial in data science for communicating findings effectively. This segment includes:

- Basic plotting with Matplotlib (line plots, scatter plots, histograms)
- Advanced visualization techniques with Seaborn (heatmaps, box plots, pair plots)
- Customizing plots (titles, labels, legends)
- Saving and exporting visualizations

Project Ideas for Data Visualization

To apply your skills, consider these project ideas:

1. Create a dashboard that visualizes COVID-19 data trends.
2. Analyze a dataset of your choice and present key findings through visualizations.
3. Recreate an existing visualization from a research paper or article.

Statistical Analysis

Understanding statistics is fundamental for data scientists. The syllabus should cover:

- Descriptive statistics (mean, median, mode, standard deviation)
- Inferential statistics (hypothesis testing, confidence intervals)

- Correlation and regression analysis
- Statistical tests (t-tests, chi-square tests)

Resources for Statistical Learning

Consider these resources to bolster your statistical knowledge:

- Online courses (e.g., Khan Academy, edX)
- Books (e.g., "Statistics for Data Science")
- Websites (e.g., StatQuest, Towards Data Science)

Machine Learning Fundamentals

Machine learning is a cornerstone of data science. This section should introduce:

- Supervised learning (classification, regression)
- Unsupervised learning (clustering, dimensionality reduction)
- Key algorithms (linear regression, decision trees, K-means, SVM)
- Model evaluation techniques (cross-validation, confusion matrix)

Machine Learning Libraries

Familiarity with libraries is essential. Focus on:

- Scikit-learn for implementing machine learning algorithms
- TensorFlow or PyTorch for deep learning applications
- Statsmodels for statistical modeling

Data Science Project Workflow

Understanding the data science project lifecycle is crucial for practical application. Key components include:

- Problem definition and project planning
- Data collection and exploration
- Model building and evaluation
- Deployment and monitoring

Capstone Project

To demonstrate your skills, consider undertaking a capstone project that involves:

- Selecting a real-world dataset
- Defining a clear problem statement
- Applying the full data science workflow from data cleaning to model deployment

Ethics in Data Science

As a data scientist, understanding the ethical implications of your work is vital. This section should address:

- Data privacy and security
- Bias in data and algorithms
- Transparency in machine learning models
- Responsible data usage and implications for society

Resources for Ethical Practices

Explore these resources for guidance on ethical considerations:

- Books (e.g., "Weapons of Math Destruction")
- Online courses (e.g., Coursera's "AI for Everyone")
- Research papers on ethics in AI and data science

Conclusion

A comprehensive **Python for Data Science syllabus** encompasses a wide range of topics and skills necessary for success in the field. From mastering Python basics to advanced machine learning techniques, each component is crucial for building a solid foundation. By following a structured syllabus and utilizing available resources, aspiring data scientists can equip themselves with the knowledge and practical experience needed to excel in this rapidly evolving domain. Embrace continuous learning and stay updated with the latest trends to become a proficient data scientist.

Frequently Asked Questions

What are the key topics covered in a Python for Data Science syllabus?

A typical syllabus includes Python basics, data manipulation with Pandas, data visualization using Matplotlib and Seaborn, statistical analysis, machine learning with Scikit-learn, and working with data from APIs or databases.

Is knowledge of statistics required before taking a Python for Data Science course?

While not strictly necessary, a basic understanding of statistics is highly beneficial, as it helps in comprehending data analysis techniques and machine learning concepts taught in the course.

What libraries should I be familiar with for a Python for Data Science course?

Key libraries include NumPy for numerical computations, Pandas for data manipulation, Matplotlib and Seaborn for data visualization, and Scikit-learn for machine learning.

How important is hands-on practice in a Python for Data Science syllabus?

Hands-on practice is crucial as it reinforces learning. Most courses emphasize practical projects and exercises to apply theoretical concepts to real-world data problems.

Are there any prerequisites for enrolling in a Python for Data Science course?

Typically, there are no strict prerequisites, but familiarity with basic programming concepts and an understanding of data analysis will enhance the learning experience.

Find other PDF article:

<https://soc.up.edu.ph/58-view/files?dataid=GFq77-9333&title=the-beginning-of-civilization-answer-key.pdf>

Python For Data Science Syllabus

What does colon equal (:=) in Python mean? - Stack Overflow

Mar 21, 2023 · In Python this is simply =. To translate this pseudocode into Python you would need to know the data ...

*What does asterisk * mean in Python? - Stack Overflow*

What does asterisk * mean in Python? [duplicate] Asked 16 years, 7 months ago Modified 1 year, 6 months ago Viewed ...

What does the "at" (@) symbol do in Python? - Stack Overflow

Jun 17, 2011 · 96 What does the “at” (@) symbol do in Python? @ symbol is a syntactic sugar python provides to ...

Is there a "not equal" operator in Python? - Stack Overflow

Jun 16, 2012 · 1 You can use the != operator to check for inequality. Moreover in Python 2 there was <> ...

Using or in if statement (Python) - Stack Overflow

Using or in if statement (Python) [duplicate] Asked 7 years, 6 months ago Modified 8 months ago Viewed 149k ...

What does colon equal (:=) in Python mean? - Stack Overflow

Mar 21, 2023 · In Python this is simply =. To translate this pseudocode into Python you would need to know the data structures being referenced, and a bit more of the algorithm implementation. ...

What does asterisk * mean in Python? - Stack Overflow

What does asterisk * mean in Python? [duplicate] Asked 16 years, 7 months ago Modified 1 year, 6 months ago Viewed 319k times

What does the "at" (@) symbol do in Python? - Stack Overflow

Jun 17, 2011 · 96 What does the "at" (@) symbol do in Python? @ symbol is a syntactic sugar python provides to utilize decorator, to paraphrase the question, It's exactly about what does ...

Is there a "not equal" operator in Python? - Stack Overflow

Jun 16, 2012 · 1 You can use the != operator to check for inequality. Moreover in Python 2 there was <> operator which used to do the same thing, but it has been deprecated in Python 3.

Using or in if statement (Python) - Stack Overflow

Using or in if statement (Python) [duplicate] Asked 7 years, 6 months ago Modified 8 months ago Viewed 149k times

python - What is the purpose of the -m switch? - Stack Overflow

Python 2.4 adds the command line switch -m to allow modules to be located using the Python module namespace for execution as scripts. The motivating examples were standard library ...

What is Python's equivalent of && (logical-and) in an if-statement?

Mar 21, 2010 · There is no bitwise negation in Python (just the bitwise inverse operator ~ - but that is not equivalent to not). See also 6.6. Unary arithmetic and bitwise/binary operations and 6.7. ...

syntax - What do >> and <

Apr 3, 2014 · 15 The other case involving print >>obj, "Hello World" is the "print chevron" syntax for the print statement in Python 2 (removed in Python 3, replaced by the file argument of the print())
...

python - Is there a difference between "==" and "is"? - Stack ...

Since is for comparing objects and since in Python 3+ every variable such as string interpret as an object, let's see what happened in above paragraphs. In python there is id function that shows a ...

python - What does ** (double star/asterisk) and * (star/asterisk) do ...

Aug 31, 2008 · A Python dict, semantically used for keyword argument passing, is arbitrarily ordered. However, in Python 3.6+, keyword arguments are guaranteed to remember insertion ...

Explore our comprehensive Python for Data Science syllabus

[Back to Home](#)