

Python for Beginner

Master the latest data analysis with python

SYLLABUS

교육 목표

인생은 짧고 Python 은 배워야한다. 이 수업의 목적은 Python 을 처음 배우는 분과 배웠지만 기억이 안나고 거의 배우지 않은거나 다름없는 분들을 위한 수업입니다. Python 이라는 언어는 많은 분야에 활용할 수 있는 언어지만 이 강의에서는 데이터를 분석하기 위한 최적의 기본기를 공부하고 실습합니다. 1강과 2강에서 최소한의 문법을 공부하고 실제 현장에서 Python이 사용(데이터분석)되어지는 파트별로 즉시 적용해보는 수업이므로 Python 자체에 대한 공부보다는 Python 으로 특정 데이터를 가져오거나 분석하는데 초점을 맞춘 강의 입니다.

사전 필수 역량: 이 프로그램을 성공적으로 마치기 위해서는 먼저 엑셀과 기초 통계 능력을 요구합니다.

프로그램 일정: 이 교육 프로그램은 총 12시간의 교육 과정으로 구성되어 있으며, 한 강의당 3시간, 총 4주에 걸쳐 진행하게 됩니다.

강의 특징: 오프라인 face to face 교육, 퀴즈 및 테스트 포함, 개인별 프로젝트 리뷰 포함

커리큘럼

1강 Python 과의 만남

1) Python 배우는 이유

- Python 은 전공자가 아닌 사람들이 배우기가 쉬우며 다른 프로그래밍 언어에 비해 사람이 사용하는 언어와 비슷합니다. 또한 누구나 어디서든 무료로 설치해서 사용할 수 있으며 방대한 라이브러리를 이용 할 수 있습니다. 특정 기능을 수행하는 코드의 집합을 가져다 사용할 수 있다는 것은 실제로 이용만 잘 하면 작성해야 할 코드는 굉장히 적어 진다는 것을 의미합니다. 또한 최근에 데이터 분석을 하기 위해서 많은 라이브러리들이 이용 되어 지고 있고 속도 또한 타 언어에 비해 느리지 않습니다. 여러분들의 데이터를 손쉽게 가공 할 수 있게 하기 위해서 Python 을 배워봅니다.

2) Python 개발환경(Anaconda & Jupyter notebook) 구축 및 첫 번째 코드

작성

- Anaconda(아나콘다) : Python 기반의 데이터 분석에 필요한 오픈소스를 모아놓은 개발 플랫폼.

- Jupyter notebook : 주피터 노트북은 코드 작성 및 실행뿐만 아니라 코드 설명을 위한문서작성을 편리하게 할 수 있는 웹 응용 프로그램

아나콘다와 주피터 노트북을 설치 하여 간단하게 코드를 작성하여 결과값을 출력해낸 다면 Python 개발환경 설치를 완료.

3) Python 연산자 및 숫자 표현

- 사칙연산을 위한 7가지 연산자 (+, -, *, **, /, //, %)

- 과학적 표기법(표현하기 불편한 숫자를 편하게 입력)

- 논리 연산(and , or, not) 및 비교 연산(==, !=, <, >, <=, >=)

4) 변수, 문자열, 리스트, 튜플, 세트, 딕셔너리

- 변수명 지음 규칙을 알아본다. 실제 프로그래밍에서 가장 많이 사용하는 변수와 각 자료형들의 개념과 특징들을 알아보고 실제 코딩 해 봅니다.

2강 Python 문법 다루기

1) 제어문 : 조건문(if문)과 반복문(while문,for문)

- 제어문은 프로그래밍에서 아주 중요합니다. 제어문이 없다면 프로그램은 그저 실행문을 순서대로 실행할 뿐입니다. 제어문 때문에 특정 코드 블록을 조건에 따라서 또는 반복적으로 실행시킬 수 있습니다. 이 기본적인 구성 요소가 연결되면 놀라울 정도로 복잡한 프로그램을 만들 수 있다는 것을 경험해 봅니다.

2) 입력과 출력

- 화면으로 입력과 출력, 파일을 읽어보고 써보기

3) 함수에 대한 설명과 실제 함수 만들기 / 내장함수 / lambda 함수

- 지금까지는 간단하고 일회성 코드 블록을 만들었습니다. **Python** 코드를 조직하여 가독성과 재 사용성을 높이는 한가지 방법은 유용한 코드 일부를 재사용 가능한 함수로 뽑아내는 것입니다. 여기서는 함수를 만드는 두 가지 방법을 다룹니다. 어떤 종류의 함수도 만들 수 있는 def 문과 짧고 익명의 함수를 만드는 lambda 문입니다.

4) 객체와 클래스

5) 에러와 예외처리

- 문법 에러: **Python** 문법에 맞지 않는 코드로 인한 에러

- 실행 에러: 문법적으로는 옳지만 실행시에 나는 에러

- 논리 에러: 문제없이 코드가 실행되지만 원치않는 결과 발생.

위 에러들이 발생 했을 때 실제 어떻게 코딩 하는지 다뤄 봅니다.

3강 문자열과 정규 표현 식

1) 문자열 다루기(분리, 삭제, 연결, 찾기, 바꾸기)

- **Python**에서는 큰따옴표(“)나 작은따옴표(‘) 안에 들어 있는 문자의 집합을 문자열이라고 합니다. 또한 텍스트 파일의 내용을 읽어 오면 그것도 문자열이 됩니다. 텍스트 파일을 읽어서 대부분 문자열 처리를 통해 원하는 형태의 데이터로 변환해서 이용하는데 그 과정에 필요한 문자열 분리, 삭제, 연결, 찾기, 바꾸기 등 문자열 변환을 직접 코딩 해보겠습니다.

2) 정규식을 활용해보기 (데이터 읽어오기)

- Python 에서의 정규식의 문법을 한번에 다 배우기는 역부족입니다. 실제 현장에서 사용 되어지는 사례를 통해서 정규식의 활용성에 대해서 사고하고 코딩하며 그와 관련된 정규식 문법들에 대해서 공부합니다. 또한 바로 정규식 관련 로직을 만들고 바로 코딩하는 시간을 가져 보겠습니다.

4강 모듈 사용 이유와 모듈 사용

1) 모듈 사용 이유 & 모듈 생성 및 호출

- Python 코드를 작성한 후 파일로 저장하면 다른 코드에서도 이 파일의 변수, 함수, 클래스를 불러와 이용할 수가 있습니다. 이렇게 코드가 저장된 파일을 모듈(Module)이라고 합니다. 코드를 작성할 때 이미 만들어진 모듈을 활용하면 코드를 효과적으로 작성할 수 있습니다. 직접 자신만의 모듈을 만들어서 그 모듈을 이용하여 코드를 작성하는 시간을 가져 보겠습니다.

2) 내장 모듈

- Python에서 제공하는 내장 모듈 중 임의로 숫자(난수)를 발생시키는 random 모듈, 날짜 및 시간 관련 처리할 수 있는 datetime 모듈, 그리고 연도/월/주 등 달력과 관련된 처리를 할 수 있는 calendar 모듈을 이용하는 방법에 대해서 알아보면서 사용.

3) 엑셀파일 관련 모듈 이용하여 엑셀파일 만들기

- 외장 모듈(xlsxwriter)을 이용하여 실제 형식이 정해진 엑셀파일을 시스템상에서 파일로 다운로드 받는 과정에 대해서 공부하고 실제 적용된 엑셀 폼을 직접 만들어 보며 응용하는 시간을 가져 봅니다.

5강 데이터 분석을 위한 패키지 1

1) NumPy를 활용한 배열 데이터 처리

- NumPy는 Python으로 과학 연산을 쉽고 빠르게 할 수 있게 만든 패키지입니다. NumPy를 이용하면 Python의 기본 데이터 형식과 내장 함수를 이용하는 것보다 다차원 배열 데이터를 효과적으로 처리 할 수 있습니다. NumPy의 기본 사용법을 익히고 응용해 보는 시간을 갖습니다.

2) Pandas를 이용한 구조적 데이터 처리(Series, DataFrame)

- Python에서 데이터 분석과 처리를 쉽게 할 수 있게 도와주는 것이 바로 pandas 라이브러리입니다. NumPy가 같은 데이터 타입의 배열만 처리할 수 있는 데 반해 pandas는 데이터 타입이 다양하게 섞여 있을 때도 처리할 수 있습니다. 엑셀형식의 모든 데이터를 pandas로 컨트롤 할 수 있습니다. 여기서는 기본적인 pandas의 사용법을 익히고 pandas가 갖고 있는 다양한 함수들을 통해서 여러 가지 계산 값들을 가져오는 시간을 갖습니다.

3) matplotlib로 그래프 그리기

- matplotlib는 Python에서 데이터를 효과적으로 시각화하기 위해 만든 라이브러리입니다. 이것을 이용하면 몇줄의 코드로 간단하게 2차원 선 그래프(plot), 산점도(scatter plot), 막대그래프(bar chart), 히스토그램(histogram), 파이 그래프(pie chart)등을 그려보고 Pandas로 처리한 데이터를 matplotlib를 이용해서 그래프로 출력하는 과정을 연습해 봅니다.

6강 데이터 분석을 위한 패키지 2

1) NumPy를 활용한 배열 데이터 처리

- NumPy는 Python으로 과학 연산을 쉽고 빠르게 할 수 있게 만든 패키지입니다. NumPy를 이용하면 Python의 기본 데이터 형식과 내장 함수를 이용하는 것보다 다차원 배열 데이터를 효과적으로 처리 할 수 있습니다. NumPy의 기본 사용법을 익히고 응용해 보는 시간을 갖습니다.

2) Pandas를 이용한 구조적 데이터 처리(Series, DataFrame)

- Python에서 데이터 분석과 처리를 쉽게 할 수 있게 도와주는 것이 바로 Pandas 라이브러리입니다. NumPy가 같은 데이터 타입의 배열만 처리할 수 있는 데 반해 Pandas는 데이터 타입이 다양하게 섞여 있을 때도 처리할 수 있습니다. 엑셀형식의 모든 데이터를 Pandas로 컨트롤 할 수 있습니다. 여기서는 기본적인 Pandas의 사용법을 익히고 Pandas가 갖고 있는 다양한 함수들을 통해서 여러 가지 계산 값들을 가져오는 시간을 갖습니다.

3) Matplotlib로 그래프 그리기

- matplotlib는 Python에서 데이터를 효과적으로 시각화하기 위해 만든 라이브러리입니다. 이것을 이용하면 몇줄의 코드로 간단하게 2차원 선 그래프(plot), 산점도(scatter plot), 막대그래프(bar chart), 히스토그램(histogram), 파이 그래프(pie chart)등을 그려보고 Pandas로 처리한 데이터를 matplotlib를 이용해서 그래프로 출력하는 과정을 연습해 봅니다.

7강 실제 업무에서 사용되는 데이터 분석

1) Pandas를 이용하여 엑셀파일 다루기

- 기업체나 학교에서도 거의 모든 사람들이 쓰는 소프트웨어가 Excel입니다. 간단한 통계나 보고서용 표 정리를 위해 엑셀의 계산 기능을 많이들 사용합니다. Pandas를 이용하시면 다양하게 Excel의 데이터를 가공하고 다시 Excel로 저장이 가능합니다. 여러 개의 엑셀 파일을 본인의 조건에 따라서 통합하여 하나의 엑셀파일로 저장하는 과정까지 익혀 봅니다.

2) Pandas가 사용된 실제 업체 사례

- 실제 RPA를 제공하는 소프트웨어 업체에서 사용하고 있는 로직 소개. 두 개의 엑셀파일에서 실제 필요 정보 추출하여 필요한 정보와 실제 조건에 맞는 행과 열을 찾아내어(두 파일간의 교집합, 차집합, 합집합, 여집합의 row들을 분류) 하나의 엑셀파일로 만드는 실제 사용 사례를 이해해 보고 익혀봅니다.

8강 웹스크레이핑

1) 웹브라우저로 웹사이트 접속하기

- 일반적으로 마우스로 클릭해서 웹사이트를 접속하는 방식이 아닌 Python으로 웹브라우저를 열고 원하는 웹사이트에 접속하는 방법에 대해서 살펴봅니다. Python에서는 내장 모듈인 webbrowser를 이용해보고 다른 라이브러리들도 이용하여 웹사이트에 접속하는 방법들을 배워봅니다.

2) 데이터 요청 응답과 HTML의 기본구조

- 웹브라우저가 웹사이트에 접속해 데이터를 가져오는 과정과 웹페이지를 구성하는 언어인 HTML(HyperText Markup Language)의 구조에 대해서 간단하게 설명하고 그 구조를 이용하여 실제 데이터를 가져오는 방법에 대해서 알아보겠습니다.

3) 접속한 웹사이트 내에서 정보 가져오기 (실제적용)

- 네이버에서 실시간 검색어 스크레이핑 / 크롤링
- 504lab에 접속해서 강의 내용 정보 및 이미지 스크레이핑 / 크롤링

9강 웹 API 이해와 웹 API 적용해보기

- 8강에서는 HTML소스를 가져와 필요한 부분만 추출하는 웹스크레이핑에 대해 알아보았습니다. 8강의 방식에서는 HTML 소스를 일일이 분석해야 하는 과

정을 통해서 데이터를 가져왔다면 이번 강의에서는 웹API를 이용하여 웹서버에서 직접적으로 데이터를 요청하고 받는 과정에 대해서 공부하겠습니다. 즉 Python을 이용해 웹API에 데이터를 요청하고 응답 받은 데이터를 처리하는 방법을 알아보겠습니다.

1) API와 웹API의 이해

2) 웹API를 이용한 사례 (업비트 웹API를 활용하여 자동 매수/매도)

- 실제 업비트 계정을 만들어서 API 인증과정을 통해 API access key를 발급받고 원하는 순간에 자동으로 매수하고 매도하는 시스템을 함께 만들어 봅니다.

10강 수강생들의 실제 업무에 필요한 모듈을 함께 만드는 실전 적용

- 수강생들이 필요로 하는 모듈을 사전 수요조사를 통해 몇 가지의 후보를 정해서 모듈을 개발해 보는 시간을 갖습니다. 수업시간에 배운 내용 안에서 만들어 보는 시간을 가질 계획입니다. 본인이 필요한 모듈이 있다면 미리 생각하고 이 수업을 들으신다면 큰 도움이 될 거라 생각합니다.

Contact Info

프로그램과 관련된 문의는 support@504lab.com 또는 070-8840-5040으로 전달주시기 바랍니다.