



데이터 사이언스

비고	현재 있는 과목
예상 출시일	

코스 요약

수강 기간

4주

수업 구성

동영상 + 노트 + 퀴즈 + 실습 과제

선수 과목

- 프로그래밍 기초 (필수)
- 데이터 분석 (필수)

사용 기술

- 파이썬
- 주피터 노트북
- numpy, pandas 등 파이썬 라이브러리

수업 목표

- 전반적인 데이터 사이언스 프로세스에 대한 이해
- 데이터 수집, 데이터 클리닝, 데이터 시각화, 데이터 분석 등 모든 데이터 사이언스 과정에 대한 기본기
- 인공지능/머신러닝 등 심화 분야를 공부할 수 있는 기본 역량

수강 대상

- 데이터 사이언티스트/분석가로 성장하고 싶은 분
- 데이터 홍수의 시대에서, 데이터에 대한 탄탄한 이해로 미래를 이끌어내고 싶은 분

강사



강영훈

Content Producer

- Dartmouth College Computer Science 전공
 - 코드잇 CEO, Co-Founder
 - 더벤처스 Software Engineer
 - Samsung HumanTech Paper Award 인공지능 연구를 통해 수학/전산 부문 동상
 - CISSE International Conference 논문 발표 및 출판
 - ARPN Journal of Engineering and Applied Sciences 논문 발표
-



손동욱

Content Director

- 고려대학교 전기전자 공학부/심화 전공
- 상명대학교 뮤직테크놀로지학과 석사 과정
- 코드잇 Content Director
- LG전자 Software Engineer

커리큘럼 요약

토픽 1: 시각화와 데이터 분석, 한 걸음 더

- ▼ Chapter 1: 데이터 사이언스란?
 - 데이터 분석가 vs. 데이터 사이언티스트
 - 데이터 사이언스의 범위와 분야
- ▼ Chapter 2: 컴퓨터가 하는 데이터 분석
 - 머신러닝이란?

- 선형 회귀 초간단 설명
- kNN 초간단 설명
- ▼ Chapter 3: 상세한 시각화와 리포팅
 - pyplot 사용해서 그래프 상세하게 그리기
- ▼ Chapter 4: 여러 그래프 그리기
 - subplot 사용해서 한 번에 여러 그래프 그리기

토픽 2: 데이터 품질 높이기

- ▼ Chapter 1: 좋은 데이터의 기준
 - 데이터 품질의 중요성
 - 좋은 데이터의 기준: 완결성
 - 좋은 데이터의 기준: 유일성
 - 좋은 데이터의 기준: 통일성
 - 좋은 데이터의 기준: 정확성
- ▼ Chapter 2: 데이터 클리닝
 - 데이터 클리닝: 완결성
 - 데이터 클리닝: 유일성
 - 데이터 클리닝: 정확성
 - 프로젝트: 스팀 게임 데이터 정리하기
 - 프로젝트: 영화 평점 분석하기
- ▼ Chapter 3: 좋은 데이터의 형태 (Tidy Data)
 - Tidy Data: 좋은 데이터의 기준
 - Tidy Data의 세 가지 규칙

토픽 3: 데이터 만들기

- ▼ Chapter 1: 데이터를 만드는 방법
 - 데이터 만들기의 중요성
 - 센서 사용하기

- 웹에서 모으기
- ▼ Chapter 2: 웹에서 긁어오기
 - 웹 페이지의 구성
 - 서버와 클라이언트
 - 파이썬으로 서버에 요청 보내기
 - 웹사이트 주소 이해하기
- ▼ Chapter 3: 필요한 것 골라내기 (Beautiful Soup)
 - 웹 페이지 스타일링 원리 (CSS)
 - BeautifulSoup으로 정보 추리기
 - 프로젝트: 검색어 순위 받아오기
 - 프로젝트: TV 시청률 크롤링
- ▼ Chapter 4: 셀레니움 (Selenium)
 - 셀레니움: 사람인 척 웹 사이트 방문하기
 - 프로젝트: 플레이리스트 가져오기
 - 프로젝트: 인스타그램 인기 게시물 찾아내기