

* 2020 봄학기 파이썬 크롤링 커리큘럼

해당 커리큘럼은 인프런 잔재미 코딩의 “파이썬입문과 크롤링기초 부트캠프 [쉽게! 하지만, 견고한 자료로!]” 강의와 민형기의 책 “파이썬으로 데이터 주무르기”를 기반으로 합니다. 이 커리큘럼은 봄학기 동안 진행될 커리큘럼입니다.

| 일자 | 프로젝트 | 주제 | 배우는 내용 |
|------|-------------------|---|---|
| 3/16 | 서울시 구별 CCTV 현황 분석 | <ul style="list-style-type: none"> - 데이터 분석 끝판왕 Pandas 기초 익히기 - matplotlib로 시각화 하기 | <ul style="list-style-type: none"> - CCTV 현황과 인구 현황 데이터 구하기 - 파이썬에서 텍스트 파일과 엑셀 파일을 읽기 (pandas) - Pandas 기초 익히기 - pandas를 이용해서 CCTC와 인구 현황 데이터 파악하기 - Pandas 고급 기능 - 두 DataFrame 병합하기 - CCTV 데이터와 인구 현황 데이터를 합치고 분석 하기 - 파이썬의 대표 시각화 도구 - Matplotlib - CCTV 현황 그래프로 분석하기 |
| 3/23 | 서울시 범죄 현황 분석 | <ul style="list-style-type: none"> - pandas와 numpy를 이용한 데이터 전처리 - 더 좋은 시각화 도구 Seaborn 익히기 - 지도 시각화 도구 Folium으로 지도 만들기 | <ul style="list-style-type: none"> - 데이터 획득하기 - pandas를 이용하여 데이터 정리하기 - 지도 정보를 얻을 수 있는 Google Maps - Google Maps를 이용해서 주소와 위도, 경도 정보 얻기 - pandas의 pivot_table 학습하기 - pivot_table을 이용해서 데이터 정리하기 - 데이터 표현을 위해 다듬기 - 좀 더 편리한 시각화 도구 - Seaborn - 범죄 데이터 시각화하기 - 지도 시각화 도구 - Folium - 서울시 범죄율에 대한 지도 시각화 - 서울시 경찰서별 검거율과 구별 범죄 발생율을 동시에 시각화하기 |
| 3/30 | 공공데이터 미세먼지 정보 크롤링 | <ul style="list-style-type: none"> - Beautiful Soup 이해 하기 - 웹구조와 HTML/CSS 이해하기 - 정부 제공 공공데이터 Open API 활용하기 - Open API 이해하기 - JSON 포맷 이해하기 - 정규표현식 | <ul style="list-style-type: none"> - Beautiful Soup으로 크롤링하기 - HTML/CSS 이해를 바탕으로 크롤링하기 - CSS selector 사용해서 크롤링하기 - 네이버 쇼핑, 네이버 주식 사이트 크롤링하기 - CSS selector 사용해서 크롤링하기 - urllib 라이브러리 사용법 - 크롤링 데이터로 엑셀 파일 만들기 - Open API/JSON 포맷 이해하기 - XML 포맷 이해와 공공데이터 Open API로 미세먼지 정보 크롤링하기 - 로그인 필요한 웹페이지 크롤링을 위한 쿠키와 세션 이해하기 - 정규표현식: 크롤링과 프로그래밍 고급 문자열 기술 |

| | | | |
|------|-------------------------|---|---|
| 4/6 | 시카고 샌드위치 맛집 분석 | <ul style="list-style-type: none"> - Beautiful Soup 활용하기 - 데이터 전처리 - 전처리된 데이터를 엑셀에 담아 시각화하기 | <ul style="list-style-type: none"> - 웹 데이터를 가져오는 Beautiful Soup 익히기 - 크롬 개발자 도구를 이용해서 원하는 태그 찾기 - 실전: 시카고 샌드위치 맛집 소개 사이트에 접근하기 - 접근한 웹 페이지에서 원하는 데이터 추출하고 정리하기 - 다수의 웹 페이지에 자동으로 접근해서 원하는 정보 가져오기 - Jupiter notebook에서 상태 진행바를 쉽게 만들어주는 tqdm 모듈 - 상태 진행바까지 적용하고 다시 샌드위치 페이지 50개에 접근하기 - 50개 웹 페이지에 대한 정보 가져오기 - 맛집 위치를 지도에 표기하기 - 네이버 영화 평점 기준 영화의 평점 변화 확인하기 - 영화별 날짜 변화에 따른 평점 변화 확인하기 |
| 4/13 | 셀프 주유소는 정말 저렴할까 | <ul style="list-style-type: none"> - Selenium으로 크롤링하기 - 데이터 전처리 - boxplot으로 시각화하기 - 지도로 시각화하기 | <ul style="list-style-type: none"> - Selenium 사용하기 - 서울시 구별 주유소 가격 정보 얻기 - 구별 주유 가격에 대한 데이터의 정리 - 셀프 주유소는 정말 저렴한지 boxplot으로 확인하기 - 서울시 구별 주유 가격 확인하기 - 서울시 주유 가격 상하위 10개 주유소 지도에 표기하기 |
| 4/20 | Q&A 세션 | | |
| 4/27 | 휴강 | | |
| 5/4 | 중간고사 기간 | | |
| 5/11 | 우리나라 인구 소멸 위기 지역 분석 | <ul style="list-style-type: none"> - 국가통계포털 KOSIS의 통계자료 활용하기 - 데이터 전처리 - JSON 활용하기 - Folium으로 시각화하기 | <ul style="list-style-type: none"> - 목표 명확히 하기 - 인구 데이터 확보하고 정리하기 - 인구 소멸 위기 지역 계산하고 데이터 정리하기 - 대한민국 지도 그리는 방법에 대한 소개 - 지도 시각화를 위해 지역별 고유 ID 만들기 - Cartogram으로 우리나라 지도 만들기 - 인구 현황 및 인구 소멸 지역 확인하기 - 인구 현황에서 여성 인구 비율 확인하기 - Folium에서 인구 소멸 위기 지역 표현하기 |
| 5/18 | 19대 대선 결과 분석 | <ul style="list-style-type: none"> - Selenium과 Beautiful Soup으로 크롤링해 데이터 만들기 - 데이터 전처리 - 득표율 시각화 | <ul style="list-style-type: none"> - Selenium과 Beautiful Soup을 이용한 데이터 획득 준비 작업 - 19대 대선 개표 결과 데이터 획득하기 - 각 후보의 득표율과 지역 ID 정리 - 19대 대선 결과 득표율 시각화하기 |
| 5/25 | 시계열 데이터 kosp이주가 정보 분석하기 | <ul style="list-style-type: none"> - 회귀 분석하기 - 페이스북의 Fbprophet 모듈로 데이터 예측하기 - 시계열 데이터 분석하기 | <ul style="list-style-type: none"> - Numpy의 polyfit으로 회귀(regression) 분석하기 - Prophet 모듈을 이용한 forecast 예측 - Seasonal 시계열 분석으로 주식 데이터 분석하기 - Growth Model과 Holiday Forecast |

| | | | |
|------|-------------------------------|--|--|
| 6/1 | 육아휴직 관련 법안에 대한 분석 (자연어 처리 기초) | <ul style="list-style-type: none"> - KoNLPy를 이용한 한글 자연어 처리 기초 익히기 - 워드 클라우드로 시각화하기 - 나이브 베이즈 분류기 이해 | <ul style="list-style-type: none"> - 한글 자연어 처리 기초 - KoNLPy 및 필요 모듈의 설치 - 한글 자연어 처리 기초 - 워드 클라우드 - 육아휴직 관련 법안에 대한 분석 - Naive Bayes Classifier의 이해 - 문장의 유사도 측정하기 - 애인 선물 고르기 |
| 6/8 | 코사다마 공모전 프로젝트 기획 | | |
| 6/15 | 휴강 | | |
| 6/22 | 기말고사 기간 | | |