|  |
| --- |
| **이력서** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
| 박성훈 |  | 성명 | 박성훈 | 영문 | Seonghoon Park |
|  | 지원분야 | 데이터 분석가 | 희망연봉 | 회사 내규에 따름 |
| 생년월일 | 1994.05.05 | 휴대전화 | 010 6309 6202 |
| E - Mail | shmi1234@naver.com | 주소 | 경기도 남양주시 퇴계원읍  퇴계원로 91번길 11 대림하이츠 2동 203호 |
| 포트폴리오 주소 |  | | |
|  | 깃허브 주소 | https://github.com/shmi1234 | | |

**학력사항**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 기간 | 학교명 | 학과(전공) | 학점 | 졸업여부 |
| 2010 ~ 2012 | 경기기계공업고등학교 | 로봇과 |  | 졸업 |

**교육이수**

|  |  |  |  |
| --- | --- | --- | --- |
| 기간 | 교육기관 | 과정명 | 시수(H) |
| 2023-09-11 ~ 2023-02-07 | KG 아이티뱅크 | 퍼블릭 클라우드 서비스를 활용한 파이썬 기반 AL-ML 인재 양성 과정 | 800H |

**교육세부내용**

|  |  |
| --- | --- |
| 과정명 | 교육 내용 |
| 퍼블릭 클라우드 서비스를 활용한 파이썬 기반 AL-ML 인재 양성 과정 | - 파이썬 이해  - 웹 크롤러  - 데이터 사이언스 이해  - 회귀 모델 분석  - 분류 모델 분석  - 비지도 학습  - 인공 신경망  - 심층 신경망  - Deep Learning on AWS  - The Machine Learning Pipeline on AWS  - 프로젝트  - 파이썬 웹 제작 |

**자격증**

|  |  |  |  |
| --- | --- | --- | --- |
| 자격증 & 장학금 | 발행처 | 취득년월 | 비고 |
| SQL 개발자(SQLD) | 한국 데이터 산업 진흥원 | 2023.12.15 |  |

**병역**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 복무기간 | 군별 | 병과 | 계급 | 병역구분 | 비고 |
| 2016~2018 | 육군 | 총기수리병 | 병장 | 군필 |  |

**경력**

|  |  |  |  |
| --- | --- | --- | --- |
| 기간 | 근무처 | 직책 | 업무내용 |
| 2022 ~ 2023 | 이모네 김치찜 | 주방 | 주방 근무 |
| 2018 ~ 2020 | TGIF 잠실점 | 코치 | 주방 코치 및 홀 코치 |
| 2014 ~ 2016 | 잠실 롯데월드 | 비정규 | 재고 관리 및 직원 스케줄 관리 |
| 2012 ~ 2013 | ㈜ 글로텍 | 생산직 | 필름 생산직 |

**자기소개서**

|  |  |
| --- | --- |
| **관심분야** | **[배움]**  데이터 분석 및 기계 학습에 대한 관심으로 학원을 등록하고 Web Crawling, Data Analyze, Machine Learning, Deep Learning 등 다양한 프로젝트를 경험하며 직무 역량을 강화했습니다. Machine Learning을 통해 각자의 데이터에 다양한 분석 방법과 선형 회귀, 로지스틱 회귀, 결정 트리, 랜덤 포레스트, 시계 열 예측 모델 등 다양한 학습 방법을 적용하여 결과를 도출하는 것이 흥미로웠습니다. 이러한 경험을 토대로 공부를 함께 병행하며, SQLD 자격증을 취득하였습니다., |
| **성격** | **[소통 능력 강화]**  롯데월드에서 고객응대 당시, 중국인 손님들과의 소통이 어려운 상황이 많았습니다. 문화적 차이와 언어적 어려움으로 인해 다양한 어려움에 직면했습니다. 이러한 상황에 맞춰서 기본적인 중국어 단어를 공부하고, 표정과 제스처를 통해 외국인 고객들과 소통하는 방법을 찾게 되었습니다. 그렇게 의사소통과 공감을 중요시하며, 커뮤니케이션 능력을 향상시키는 데에 주력했습니다.  **[책임감]**  TGIF에서 코치로 일할 때, 전문 코칭 교육을 수료했습니다. 교육 중 '책임감'에 대한 중요성을 강조 받아, 이를 일에 적용하게 되었습니다. 교육을 마친 후, 일에 대한 책임감을 높이고 동료들과의 소통을 강화하기 위해 노력했습니다. 경험을 바탕으로 동료들에게 지식을 전하며 작은 일에도 신중함을 기울이고, 인간관계와 신뢰를 쌓아가며 책임감을 키워나갔습니다. |
| **직무역량** | **[능력]**  Web Crawling 프로젝트에서는 네이버 부동산 사이트와 부동산 실 거래가 사이트에서 Python의 BeautifulSoup과 Selenium을 활용하여 Crawling을 수행했습니다. 주로 Selenium의 Xpath를 사용하여 경로를 지정하는 Crawling을 진행했지만, 중간에 Xpath가 정상적으로 작동하지 않는 문제가 발생했습니다. 이 문제를 해결하기 위해 나중에 iframe이 존재한다는 것을 확인하고, Element 식별을 통해 iframe을 지정하여 문제를 해결하였습니다. 이를 통해 프로젝트의 완성도를 향상시켰습니다. Data Analyze 프로젝트에서는 서울시의 대기오염과 호흡기질환 환자 데이터를 Open API를 활용해 CSV 파일로 수집하고, 이를 기반으로 데이터 프레임을 생성하여 전처리 및 결측 값 처리, 이상치 변경/삭제 등의 작업을 수행했습니다. 이후 머신 러닝 기법을 활용하여 선형 회귀, 로지스틱 회귀, 결정 트리, 랜덤 포레스트 등을 활용하여 대기오염 수치를 기반으로 한 질병 발병률, 미세먼지 수치에 따른 초 미세먼지 수치 등을 예측하였습니다. 이러한 다양한 학습 프로그램을 사용하면서 데이터에 따라 특정 학습 프로그램을 선택하는 효과적인 방법을 배우게 되었습니다. |
| **입사 후 포부** | **[업무 능력 강화]**  업무 이외의 시간을 활용하여 ADSP 자격증 취득을 목표로 삼고, 주제별로 매주 목표를 설정하고 학습과 실습을 진행할 계획입니다. 이를 위해 데이터 분석 이론서부터 실전 예제까지 체계적인 학습 계획을 수립했습니다. 온라인 강의, 교재 학습, 실전 프로젝트 등 다양한 학습 자료를 활용하여 이론과 실무를 함께 강화할 것이며, 주기적인 퀴즈와 모의고사를 통해 복습하고 실력을 향상시킬 것입니다.  ADSP 공부를 통해 데이터 분석 이론 뿐만 아니라 데이터 수집, 전처리, 시각화, 분석, 결과 해석 등 데이터 전 과정을 이해하고 실무에 적용할 수 있는 능력을 키우고, 데이터베이스 관리 및 쿼리 작성 능력을 향상시켜 구조화된 데이터를 효과적으로 다루고 관리할 수 있도록 노력할 것입니다. 또한, 데이터 품질 관리 기법, 통계적 분석 방법, 머신 러닝 기법 등을 활용하여 예측 모델링과 인사이트 도출 능력을 강화할 계획입니다.  ADSP에서 습득한 데이터 관리 및 분석 기술을 업무에 적용하여 데이터를 효율적으로 수집, 관리하고 분석에 활용함으로써 회사의 전략 수립과 실행에 기여할 수 있습니다. 데이터 기반의 의사결정을 지원하고 비즈니스 인사이트를 도출하여 팀원들과의 원활한 업무 협업을 이끌어낼 것입니다. 지속적인 학습을 통해 ADP와 SQLP 등의 전문가 자격증을 취득하여 회사에서 요구하는 역량을 충족시키고, 전문성을 더욱 강화하여 회사의 핵심 인재로서 역할을 수행할 것입니다. |
| **지원동기** | **[성장]**  롯데월드 에서의 식음료팀에서 근무하면서 사람들과의 소통 및 상대 대응 업무를 수행하며 커뮤니케이션 스킬을 향상시켰습니다. 이외에도 IT 업계에서의 데이터베이스 관리와 유사한 경험으로 재고 관리와 스케줄 관리를 경험하였습니다. 그리고 이 경험을 했었던 때를 생각해봤을 때, 재고 관리나 스케줄 관리를 혹시 더 편하게 할 수 있는 방법이 있었을까 생각해본적이 있었습니다. 그 편한 방법 중에 하나가 코딩을 이용한 방식이었고, 그런 큰 데이터베이스 방식으로 관리 했었다면 어땠을까 라는 마음이 있었습니다. 이를 계기로 처음 IT업계에 관심이 가기 시작했고, 학원에서 수행한 다양한 프로젝트, 그리고 여러 Machine Learning 기법을 활용한 경험들을 통해 제가 쌓은 직무 역량과 지식을 결합하여 효과적인 분석을 수행하는 데이터 분석가로 성장하고자 합니다. 이러한 다양한 경험들을 토대로 실무에서의 도전에 대비하며, 지속적인 학습과 성장을 추구하고 있습니다. |

**Technology (보유기술)**

|  |  |
| --- | --- |
| **기술분류** | **보유기술** |
| **Language** | **[Python]**  - Pandas, NumPy 등의 데이터 분석 계산 라이브러리 활용  - BeautifulSoup, Selenium과 같은 Crawling 라이브러리 활용  - Matplotlib, Seaborn, Plotly, Folium과 같은 시각화 라이브러리 활용  - Scikit-learn, TensorFlow의 머신 러닝 라이브러리 활용  **[JAVA]**  - Java Bean을 활용, Set/Get 메소드를 이용해 DB와 연동하여 DB 입출력  - Model1 / Model2 기반 개발  **[JSP]**  - 자바를 이용하여 웹 어플리케이션을 개발  - EJB의 개념과 WAS 이용 EJB기반 코드 작성  **[JavaScrpt]**  - IE/Firefox 기반의 HTML/DHTML 코딩  - JavaScript 내장객체, 사용자 정의 객체의 사용법과 활용법  **[HTML]**  - HTML과 CSS를 사용하여 웹 페이지의 구조를 설계  **[Ajax / Flex]**  - Ajax / Flex 기반 application 프로그램 작성 |
| **DBMS** | **[Oracle / MySQL / SQLite]**  - DB 설치 및 기본 SQL 작성  - DB 모델링 기법 및 정규화  - 트랜잭션의 개념과 처리 방법과 PL/SQL 작성 |
| **TOOL** | **[IDE / DB Tool / VSCode / Jupyter]**  - Eclipse를 이용한 자바 코드 작성  - ERWin을 이용한 DB설계  - Rational Rose를 이용한 UML 구조 작성  - VSCode를 이용한 파이썬 코드 작성  - Jupyter를 이용한 파이썬 코드 작성 |
| **Framework** | **[Struts2/iBATIS]**  - Struts2 로직처리, 화면처리, 기타기능  - Struts2 MVC 패턴 및 iBATIS 연동 이용 WebApp 제작  **[Spring]**  - Spring AOP 개념과 활용  - Spring MVC패턴 이용 WebApp 제작  **[Django]**  - 템플릿 언어를 활용한 동적 웹 페이지 개발  - ORM(Object-Relational Mapping)을 사용하여 데이터베이스 모델을 정의  - 보안 기능 및 인증 시스템을 활용 |

**Project (프로젝트)**

|  |  |
| --- | --- |
| **프로젝트** | 웹 크롤링을 활용한 부동산 데이터 수집 |
| **개발기간** | 2023.11.27 ~ 2023.12.01 (총 5일) |
| **참여인원** | 4명 |
| **담당업무** | 윤상하(조장) : 기획, PPT  **박성훈 : 기획, 제안, 기능 개발**  정선우 : 기획  김동규 : 기획, 제안 |

|  |  |
| --- | --- |
| **개발환경** | Windows 10 |
| **사용도구** | Jupyter,, VSCode |
| **사용기술** | PYTHON, HTML |

|  |
| --- |
| **개요** |

최근 수행된 설문 조사에 따르면, 20대의 1,094명을 대상으로 한 조사 결과에 따르면, 20대의 주요 목표로는 내 집을 마련하는 것이 가장 큰 비중을 차지합니다. 주식이나 코인을 포함한 재테크(20%), 대출 상환(27%), 여행(38%), 저축(52%) 등이 있지만, 내 집 마련이 75%의 응답자가 선택한 답안으로 나타났습니다. 이는 서울시 내에서 20대에게 내 집 소유가 큰 꿈이며, 인생의 중요한 목표임을 시사합니다.

이에 따라, 우리는 이러한 결과를 바탕으로 서울시 내의 집값을 조사하여, 현재의 계획에 맞는 최적의 지역을 찾아보고 집값을 빠르게 비교해보고자 합니다. 이를 통해 20대의 내 집 마련 꿈을 실현하기 위한 최상의 전략을 마련할 수 있을 것으로 기대됩니다.

|  |
| --- |
| **기획 의도(동기)** |

부동산은 많은 사람들에게 주요한 자산으로 작용하기 때문에, 장기적인 자산 관리를 위해 부동산 가격 변동 추이를 조사하고자 하는 경우가 많습니다.

부동산 시장에 대한 장기적인 데이터 수집과 분석은 부동산 시장의 전반적인 흐름을 파악하고, 미래의 전망을 예측하는 데 도움이 됩니다. 이를 통해 부동산 시장의 동향을 이해하고, 다양한 산업 및 경제 지표와의 상호 연관성을 파악할 수 있습니다.

요약하면, 부동산 가격 변동 추이를 장기적으로 조사하고 분석하는 것은 부동산 시장의 향후 전망을 예측하고 다른 산업 및 경제 지표와의 관계를 파악하는 데 도움이 됩니다.

|  |
| --- |
| **목표 및 설계** |

**[데이터 수집]**

아실 <https://asil.kr/> 네이버 부동산 <https://new.land.naver.com/>

서울시를 대상으로 25개구에 위치한 467개 동에 대해 총 3,846개의 아파트 별 실 거래가를 수집 하였습니다. 이를 통해 각 아파트의 실 거래가 정보를 상세히 파악할 수 있으며, 서울시 내의 다양한 지역에서 발생하는 부동산 시장의 동향을 분석할 수 있습니다.

**[아실 사이트의 Iframe]**

Inline Frame(iframe)은 웹 페이지 안에 또 다른 웹 페이지를 삽입하는 기능을 의미합니다. 현재 브라우저에서 렌더링 되는 문서 안에 다른 HTML 페이지를 삽입하여 표시할 수 있도록 합니다. 그러나 최근에는 보안 문제, 모바일 호환성 및 대체 명령어 등장 등의 이유로 사용이 줄어 들고 있는 추세입니다.

XPath가 /html/body/div[2]/div[1]/div[1]/div[3]/a[1]인 경우 브라우저가 탐색 실패 오류를 반환할 때, 이는 해당 소스가 id가 'sub1'인 iframe 안에 존재하기 때문입니다. 이를 해결하기 위해서는 먼저 해당 iframe으로 진입해야 합니다. 이를 위해 진입할 iframe을 변수로 지정하고, Browser.switch\_to.frame() 메서드를 사용하여 해당 iframe으로 진입합니다. 데이터를 추출한 후에는 Browser.switch\_to.default\_ content() 메서드를 사용하여 상위 HTML로 돌아갑니다. 이 작업을 수행해야 다른 iframe에 있는 데이터를 추출할 수 있습니다.

테이블의 각 row에서 데이터를 추출하기 위해 for 루프를 사용하는데, 해당 코드가 Selenium이 데이터를 한 번만 추출한 후에는 이후의 for 루프에서 경로를 찾지 못하고 오류가 발생하는 문제가 있습니다. 이를 해결하기 위해서는 데이터를 추출할 때마다 iframe으로 진입하여 작업을 수행해야 합니다. 테이블의 전체 텍스트를 불러올 수 있지만 줄바꿈 삽입으로 데이터를 가공하는 것이 어려울 수 있습니다.

**[네이버 부동산 사이트]**

브라우저 진입 및 데이터 웹 크롤링을 위한 사이트 링크 파도타기를 진행하고, 서울시의 모든 '구'를 리스트화 하여 추가합니다. 그 후 각 '구'에 포함된 '동' 리스트도 생성하고, 각 '동'에 포함된 모든 '아파트' 리스트를 만듭니다. 데이터를 웹 크롤링 하는 과정에서 '평수'와 '매물' 정보가 없는 경우에는 오류가 발생할 수 있습니다. 이를 해결하기 위해 try ~ except 구문을 활용하여 예외 처리를 수행하였습니다. 각 아파트의 평수를 구분하기 위해 리스트 처리를 하고, 각 아파트를 해당 평수 별로 분류한 후 '상한가' 데이터를 찾습니다. 마지막으로, CSV 파일에 입력을 위해 결과값이 나오는 데이터를 추가할 final\_list를 만들고, 각 구에 대한 동 별로 데이터를 정리합니다. 면적 별로 '상한가' 데이터를 입력 받고, 해당 매물이 없는 경우에는 'None' 데이터를 삽입하여 구분합니다. 최종적으로 약 2000개의 상한가 데이터를 CSV 파일로 정렬하여 출력합니다.

**[시각화]**

자치구별로 아파트의 수를 비교하고, 각 자치구에서의 최고가와 최저가를 비교해보며, 마지막으로 자치구별로 평균가를 비교하고자 합니다. 먼저, 서울시의 각 자치구에서 아파트의 수를 측정하고 비교합니다. 그런 다음, 각 자치구에서의 아파트 중에서 최고가와 최저가를 확인하여 비교합니다. 마지막으로, 각 자치구에서의 아파트의 평균가를 계산하여 비교합니다. 이러한 분석을 통해 서울시의 각 자치구가 아파트 시장에서 어떻게 위치하고 있는지를 이해할 수 있습니다.

|  |
| --- |
| **후기** |

이 데이터 수집과 시각화를 통해 우리는 서울시 부동산 시장의 현황을 더 잘 이해할 수 있게 되었습니다. 먼저, 아실과 네이버 부동산 사이트를 통해 수집한 정보를 통해 서울시 25개구에 위치한 467개 동에 총 3,846개의 아파트에 대한 실 거래가를 상세히 파악할 수 있었습니다. 이를 통해 서울시 내 다양한 지역에서 발생하는 부동산 시장의 동향을 분석할 수 있었습니다.

특히, 아실 사이트의 iframe을 통해 데이터를 수집하는 과정에서 Selenium을 사용하면서 발생한 문제들을 해결하는 과정은 새로운 도전이었습니다. 또한, 네이버 부동산 사이트에서도 데이터를 웹 크롤링 하는 과정에서 예외 처리를 통해 문제를 극복하는 등 다양한 어려움을 만났지만, 이를 해결하고 데이터를 성공적으로 수집할 수 있었습니다.

시각화를 통해 자치구별로 아파트의 수, 최고가, 최저가, 그리고 평균가를 비교함으로써, 서울시 각 자치구의 부동산 시장에서의 위치와 특징을 파악할 수 있었습니다. 이러한 분석을 통해 부동산 투자나 거주지 선정 등의 의사 결정에 도움을 줄 수 있을 것으로 기대됩니다.

팀별 단위로 진행한 첫 프로젝트였습니다. 처음에는 Iframe과 같이 웹 크롤링하기 어려운 과제에 직면했지만, 결국 데이터 수집에는 성공했습니다. 그러나 데이터 가공에 어려움을 겪게 되었고, 팀원들과 의견을 조율한 끝에 네이버 부동산 사이트로 다시 웹 크롤링을 시작하게 되었습니다. 이후에는 프로젝트를 성공적으로 완료하여, 팀원들과의 신뢰를 쌓고 서로 돈독한 관계를 형성하게 되었습니다.

**Project (프로젝트)**

|  |  |
| --- | --- |
| **프로젝트** | 대기 오염 Air Pollution |
| **개발기간** | 2024.12.27 ~ 2024.01.03 (총 5일) 기간 내에 자소서 코칭 기간 포함 |
| **참여인원** | 4명 |
| **담당업무** | 윤상하(조장) : 기획, EDA, Machine Learning, PPT  **박성훈 : 기획, 제안, EDA, Machine Learning,**  정선우 : 기획, EDA  김동규 : 기획, 제안 |

|  |  |
| --- | --- |
| **개발환경** | Windows 10 |
| **사용도구** | Jupyter, VSCode |
| **사용기술** | PYTHON |

|  |
| --- |
| **개요** |

대기 중의 미세먼지, 화학물질 및 다양한 오염 물질들은 우리의 건강과 더불어 지구 전체의 생태계에도 심각한 영향을 미칩니다. 이러한 대기오염은 우리가 살아가는 환경에서 심각한 문제로 대두되고 있으며, 이는 우리의 삶과 미래에 큰 영향을 미칠 수 있습니다. 지구 환경을 위협하는 대기 오염은 근본적인 해결이 필요한 과제 중 하나로 인식되고 있으며, 개인 및 사회적 차원에서의 노력과 대책 마련이 필수적입니다. 이를 통해 우리는 더 건강하고 지속 가능한 미래를 위한 발전을 이루어 나갈 수 있을 것입니다.

|  |
| --- |
| **기획 의도(동기)** |

비염은 비루(콧물), 재채기, 가려움증, 그리고 코 막힘과 같은 비점막의 염증성 질환으로, 한 가지 이상의 증상이 동반됩니다.

천식은 특정한 유발 원인 물질에 노출되었을 때 기관지의 염증으로 인해 기관지가 심하게 좁아져 기침, 천명(숨쉴 때 쌕쌕 거리는 소리), 호흡곤란, 그리고 가슴 답답함이 반복적으로 발생하는 질환입니다.

아토피는 유아기 혹은 소아기에 시작되는 만성 재발성의 염증성 피부질환으로, 소양증(가려움증)과 피부건조증, 그리고 특징적인 습진을 동반합니다.

가설:

지역별 가설: 특정 서울시 지역에서는 다른 지역에 비해 대기오염도가 높거나 낮을 것이다.

질병과의 연관성: 대기 오염도가 높거나 낮을 때, 발생하는 질병이 다르고, 환자의 숫자가 다를 것이다.

미세먼지: 미세먼지는 계절마다 다를 것으로 보이고, 이에 따라 계절 예측 및 미래의 오염도 예측이 가능할 것으로 보인다.

|  |
| --- |
| **목표 및 설계** |

**[데이터 수집]**

대기오염도 데이터는 에어코리아, 날씨 데이터는 기상청, 질병 데이터는 국민 건강 보험 공단 및 건강 보험 심사 평가원의 웹사이트를 참고하여 데이터 수집을 진행하였습니다.

**[데이터 전처리]**

데이터를 월별로 취합하고, 질병 월별 데이터를 합쳐서 3월부터 5월까지는 봄, 6월부터 8월까지는 여름, 9월부터 11월까지는 가을, 12월부터 2월까지는 겨울로 구분하여 계절 컬럼을 생성 합니다.

**[데이터 모델링]**

미세먼지와 초 미세 먼지의 예측을 수행하고, 이 정보를 활용하여 계절을 예측할 수 있는 모델을 개발합니다. 그리고 대기오염도를 기반으로 질병에 따른 유병률을 예측합니다.

**[시각화]**

Pandas, Matplotlib, Seaborn, 등을 활용하여 분석 및 예측한 데이터를 시각적으로 표현합니다. 이를 통해 결과를 명확하게 전달하고 이해하기 쉽게 시각화 합니다.

|  |
| --- |
| **후기** |

대기오염 수치와 미세먼지 데이터만으로는 유의미한 분석을 수행하는 데에 한계가 있었습니다. 초기에는 지역별로의 차이가 있을 것으로 예상했지만, 실제로는 큰 차이를 발견하지 못했습니다. 아마도 전국적인 데이터를 활용한다면 서울과 다른 지방 간의 차이가 뚜렷하게 드러날 것으로 예상됩니다. 이에 따라 다른 지역의 데이터를 활용하면 예측이 가능할 것으로 보입니다. 그러나 서울에 한정된 데이터로는 지역별 예측을 수행하는 데에는 어려움이 있었습니다.

또한, 질병과 대기오염의 연관성을 예상했지만, 실제로는 그 영향이 미미했습니다. 온도와 미세먼지 수치로 초 미세 먼지를 분석할 수 있었으며, 이러한 분석에서는 계절과 미세먼지 수치가 상당히 중요한 역할을 한다는 것을 확인했습니다.

추후에는 다른 프로젝트를 진행하거나 더 많은 시간을 투자한다면 더 유의미한 결과를 얻을 수 있을 것으로 예상 됩니다. 그러나 현재의 분석 단계에서는 이 정도의 결론을 내는 것이 적절 하다고 판단 됩니다.

여러 가지 분석과 머신 러닝 모델을 활용하면서 각각의 모델이 어떤 분석에 유리하고 어떤 데이터에 적합한지를 배웠습니다. 처음에 예상한 결과와는 다르게 예상하지 못한 상관관계를 확인하면서 데이터 분석이 예상보다 쉽지 않다는 것을 깨달았습니다. 이러한 경험들을 통해 자신감을 많이 얻게 되었는데, 전처리 작업과 분석을 진행할 때마다 예상치 못한 결과가 나오고 추가적인 데이터가 필요한 경우에는 상황에 맞게 대처하면서 더 나은 학습을 진행하기 위해 노력했습니다.

데이터 분석가로서 정답이 없고 본인의 노력과 능력이 중요하다는 것을 느꼈습니다. 앞으로의 마지막 프로젝트가 더욱 재미있고 도전적으로 진행될 것으로 기대되며, 현재 프로젝트를 통해 쌓은 경험이 많은 도움이 될 것으로 기대합니다.

**Project (프로젝트)**

|  |  |
| --- | --- |
| **프로젝트** | Criminal Anatomy |
| **개발기간** | 2024.01.08 ~ 2024.02.06 (총 30일) |
| **참여인원** | 4명 |
| **담당업무** | 윤상하(조장) : 기획, EDA, Machine Learning, Framework 설계, 디자인, PPT  **박성훈 : 기획, EDA, Machine Learning, DB 설계, Framework 설계, 기능 개발,**  **디자인, PPT**  정선우 : 기획, EDA, 제안  김동규 : 기획, EDA, Machine Learning |

|  |  |
| --- | --- |
| **개발환경** | Windows 10 |
| **사용도구** | Jupyter, VSCode, QGIS, SQLite |
| **사용기술** | HTML, CSS, JS, PYTHON, DJANGO |

|  |
| --- |
| **개요** |

서울특별시의 5대 범죄 발생률과 음주운전 범죄를 기반으로 각 범죄 유형 및 서울시 행정동 별 범죄율을 분석하고, 서울의 지역별 사회적 특성과 범죄 발생 간의 상관관계를 조사합니다. 또한, 범죄 예방을 위해 시행된 제도 및 법령의 효과를 확인하고 실효성을 검증하여 향후 예측에 기여합니다. 이를 토대로 안전에 관한 자체 지표를 수립하여 지역 사회의 안전성을 측정하고 개선에 기여하는데 목표를 두고 있습니다.

|  |
| --- |
| **기획 의도(동기)** |

작년 발표된 경찰청의 '2022년 국민 체감 안전도 조사 결과'에 따르면 전체적인 국민의 범죄 체감안전도는 전년보다 0.4점 하락한 83점으로 조사되었습니다. 여성 상대 범죄는 81.4점으로 0.3점, 강도·살인은 84.6점으로 0.6점 하락하는 등 전 분야에서 체감안전도가 낮아진 것으로 나타났습니다. 이어서 지난해 8월부터 이상동기범죄의 증가, 범행 예고 및 협박 글의 증가 등의 현상이 나타나면서 전문가들은 시민들이 재난 수준의 불안을 느낄 것으로 진단했습니다. 이러한 상황에서 우리는 특히 서울의 범죄 데이터를 분석하여 사회 안전 문제에 대한 심각성을 파악하고자 합니다.

가설:

지역별 가설: 특정 서울시 지역에서는 다른 지역에 비해 범죄 발생률이 높을 것으로 예상됩니다.

인구밀도 별 가설: 생활 인구가 범죄 발생률과 높은 상관관계를 갖을 것으로 예상되며, 거주인구보다는 더 큰 영향을 미칠 것으로 가정합니다.

범죄 유형 간 상관 가설: 특정 범죄 유형이 다른 범죄 유형과 상관 관계가 있을 것으로 예상하며, 이를 통해 범죄 간의 패턴을 확인하고자 합니다.

|  |
| --- |
| **목표 및 설계** |

**[데이터 수집]**

경찰청, 도로교통공단, 공공데이터포탈, 서울 열린데이터광장 등에서 서울시 5대 범죄 발생 현황, 서울시 구별 인구밀도, 생활 인구 밀도, 음주운전 추이와 같은 범죄 데이터 및 사회 특성 데이터를 획득합니다.

**[데이터 전처리]**

결측값 및 이상치를 제거할 수 있는 데이터에 대해 처리를 진행하고, 동일한 특성을 가진 열에 대해서는 칼럼 명을 통일화 합니다.

**[데이터 모델링]**

전처리된 데이터를 기반으로 시나리오에서 세운 가설을 확인하기 위해 다양한 데이터 셋을 조합합니다. 또한, 앞으로의 범죄율 및 음주운전 다발 지역을 예측하는 모델링을 수행합니다.

**[시각화]**

Pandas, Matplotlib, Seaborn, QGIS 등을 활용하여 분석 및 예측한 데이터를 시각적으로 표현합니다. 이를 통해 결과를 명확하게 전달하고 이해하기 쉽게 시각화 합니다.

**[안전 지수 서비스 제공]**

사용자의 실시간 위치 정보를 활용하여 주변의 안전 관련 시설과 위험 시설을 분류하고, 실시간으로 사용자에게 안내합니다. 사용자는 이를 통해 안전한 경로 선택이 가능하며, 자체적인 안전 지표를 통해 사용자의 안전 수준을 시각적으로 파악할 수 있습니다. 이로써 사용자의 불안함을 낮추고, 지역의 안전 인프라에 대한 인식을 강화할 것으로 기대됩니다.

|  |
| --- |
| **후기** |

어려운 작업이었지만, 생각보다 순조롭게 진행되어 문제들을 성공적으로 해결할 수 있었습니다. 범죄 데이터의 난해한 구조와 연단위로 정리된 데이터의 한계로 인해 몇 가지 어려움이 있었지만, 여러 가지 컬럼을 추가하고 새로운 데이터를 만들어내며 상관관계를 찾아내는 작업은 새로운 도전이었습니다. 범죄 분석은 데이터뿐만 아니라 각 개인의 심리적인 요소도 고려해야 하기 때문에 예측이 쉽지 않았습니다.

CCTV와 검거율 간의 상관관계를 찾아보려 했으나 예상과는 달리 큰 연관성이 없었습니다. 그러나 서울시의 CCTV 개수를 살펴보면 우리나라의 치안 수준이 나쁘지 않다는 인식을 얻을 수 있었습니다. 상관관계가 크게 나타나지 않는 데이터로 예측 모델을 구축하려는 시도에서 여러 의문과 고민이 생기게 되었습니다.

장고 서비스를 구현하는 과정에서는 처음에 어려움을 느낄 수 있을 것이라 생각했지만, 여러 함수를 조합하여 좌표 중심으로 반경을 설정하고 마킹 하는 작업을 성공적으로 마무리할 수 있었습니다. 앞으로도 이와 관련된 코딩 작업은 더 수월하게 해낼 수 있을 것으로 기대됩니다.

팀원들과 이전에 다양한 프로젝트를 수행했을 때, 팀 내 의사소통은 매우 원활하게 이루어졌습니다. 하지만 의견이 충돌하는 상황도 가끔 발생했습니다. 이러한 경우에는 정기적인 회의를 통해 의견을 조율하고 문제를 해결하는 방식으로 대응했습니다. 프로젝트 도중 발생한 문제나 의견 차이를 주기적으로 피드백하면서 의사소통을 강화했습니다. 의견 충돌을 조율하고 해결해 나가는 과정을 통해 프로젝트의 진행이 원활하게 이뤄지고, 마지막 단계에 도달하여 현재의 결과를 도출할 수 있었습니다. 이 경험은 효과적인 팀 협업과 의사소통의 중요성을 배울 수 있는 기회가 되었습니다.

이제 학원의 마지막 프로젝트를 통해 초보자에서 코드를 읽을 수 있는 수준으로 성장한 것 같아 뿌듯합니다. 그러나 이 분야에서는 지속적인 공부와 노력이 필요하며 계속해서 스스로 발전해 나가야 함을 깨닫게 되었습니다. 학습의 마무리가 아니라 시작에 불과하며, 더 나은 개발자가 되기 위해 끊임없이 노력하고자 합니다.