



# Upstage AI Lab

Python 개발 프로젝트  
: 채용 정보 사이트 크롤링을 통한 채용 지도 만들기

## 목차

- 01. 팀원 소개
- 02. 프로젝트 주제 소개
- 03. 이론 및 개념 설명
- 04. 결과 및 인사이트 공유
- 05. 프로젝트 회고



01

# 팀원 소개

창의력을 제품으로 구현합니다.



성명기

Intereste in

- Data Analytics & Business Analytics
- LLM
- MLOps

Introduction

- AI소프트웨어전공.
- 대학 3학년때 차세대반도체 페스티벌 POLARIS SIF 2022에서 YOLOv5를 활용한 영유아 안전시스템을 제작 한 경험이 있습니다.

Role

- 팀장.
- Streamlit을 활용한 데이터 시각화.
- 원티드 채용플랫폼 데이터 크롤링.

In Upstage AI Lab

- Upstage AI Lab을 끝까지 수료 후 취업.
- Upstage AI Lab 파이널 프로젝트에서 좋은 성적을 거두기.



시작한 일은 반듯이 끝 마친다.



최제우

Intereste in

- 데이터 처리
- 머신러닝, 딥러닝

Introduction

- 소프트웨어전공.
- 경진대회 참여 이력은 없지만 학부시절 팀프로젝트 경험이 있습니다.

Role

- 사람인 채용플랫폼을 담당하여 데이터 크롤링 진행
- 지도 위에 시각화 된 마커에 이벤트 기능.

In Upstage AI Lab

- 다양한 경진대회와 프로젝트 경험을 쌓고 나가는 것이 목표이며 더 나아가 이후에도 데이콘이나 캐글 같은 대회에 스스로 참여 할 수 있는 스킬을 갖추고 싶습니다.
- 다양한 경진대회, 프로젝트 경험과 이를 끝까지 진행할 수 있다는 자신감.

코딩으로 상상을 실현합니다.



유정수

Intereste in

- LLM
- data-centric ai
- service development

Introduction

- Language Model
- 가격 데이터를 분석하여 저렴해진 제품 추천 서비스
- 개인 블로그 웹사이트 제작

Role

- 크롤링

In Upstage AI Lab

- 원하는 언어 모델 개발 성공
- 다양한 만남을 통한 인사이트

02

# 프로젝트 주제 소개



Upstage AI Lab

# 프로젝트 소개

: Python 개발 프로젝트

주제

채용 정보 사이트 크롤링을 통한 채용 지도 만들기

목표

목표

채용 정보사이트(사람인, 잡코리아, 원티드) 크롤링을 통해 현재 구인하고 있는 검색어에 해당하는 회사들의 위치를 한눈에 볼 수 있도록 지도에 표시한다.

개요

소개 및 배경 설명

취업 시 회사의 위치는 출퇴근 시간과 주변 환경을 고려할 때 매우 중요한 요소입니다. 그럼에도 불구하고 채용 정보 사이트들은 개별 회사의 위치만을 지도상에서 보여주어, 여러 회사의 위치를 비교하기 어렵습니다.

이러한 문제를 해결하기 위해, 저희 서비스는 검색어에 해당하는 모든 구인 회사들의 위치를 한 눈에 볼 수 있도록 지도에 표시합니다. 이를 통해 사용자는 위치를 기반으로 한 보다 효과적인 취업 결정을 내릴 수 있습니다.

기간

2024. 04. 11 ~ 2024. 04.17



# 프로젝트 진행 방법

: Python 개발 프로젝트

**팀원 소개** : 성명기(팀장), 유정수, 최제우

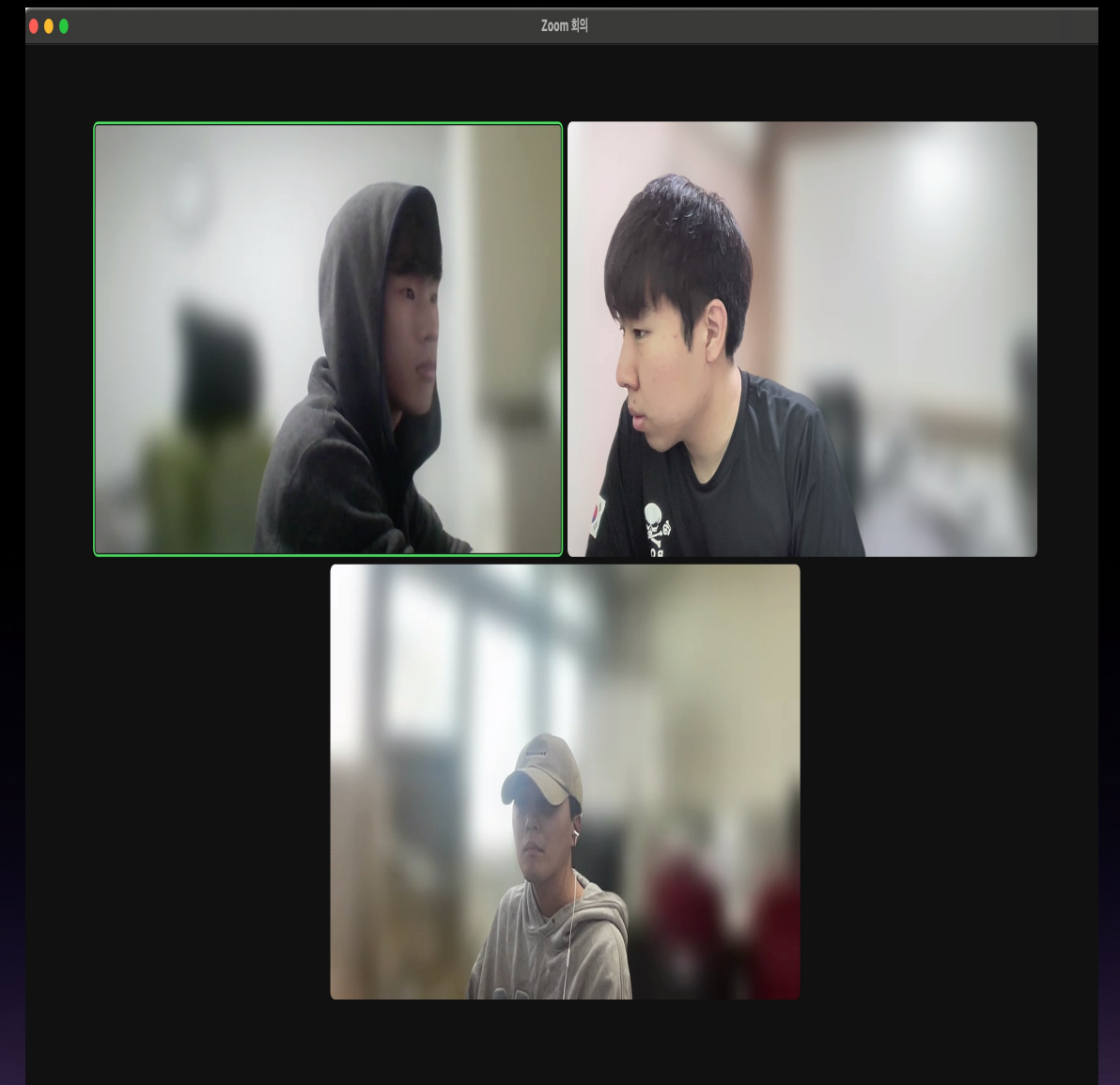
**스크럼 진행 횟수 및 일정**: 매일 오전 ZOOM 실시간 실습시간 마다 스크럼 회의 진행

**프로젝트 진행 장소** : 실시간 비대면 (ZOOM)

**프로젝트 진행 방법** : 우리가 활용 가능한 데이터를 먼저 수집하고 수집한 데이터를 파이썬 데이터시각화 웹 프레임워크인 Streamlit를 활용하여 데이터를 지도, 워드클라우드로 시각화한다.

**프로젝트 진행 시 생긴 문제점** : 구현 하고자 하였던 마커 클릭 이벤트 기능이 라이브러리의 한계로 구현에 어려움 및 채용플랫폼을 통한 크롤링 데이터 활용방안 부족

**문제 해결 방법** : 마커 클릭 이벤트를 마우스 오버 이벤트로 대체 후 모달창으로 회사명 보이게 문제해결





03

# 이론 및 개념 설명



# 이론 및 개념 설명

: Python 개발 프로젝트 (아래 내용은 예시입니다.)

STEP 01.

크롤링  
Crawling

웹 페이지를 그대로 가져와서 거기서 데이터를 추출해 내는 행위이다.  
단순히 웹페이지에서 데이터를 추출하는 스크래핑 이라는 행위와는  
달리 페이지를 동적으로 이동하며 정보를 취득한다.

STEP 02.

개방형 API  
Open API

API란 특정 프로그램이 운영 체제나 프로그래밍언어가 제공하는  
기능을 제어할 수 있게 만든 인터페이스 이다.  
Open API는 이러한 API를 외부의 사이트와 자유롭게 활용하고 공유  
가능하도록 설계된 API 이다.

STEP 03.

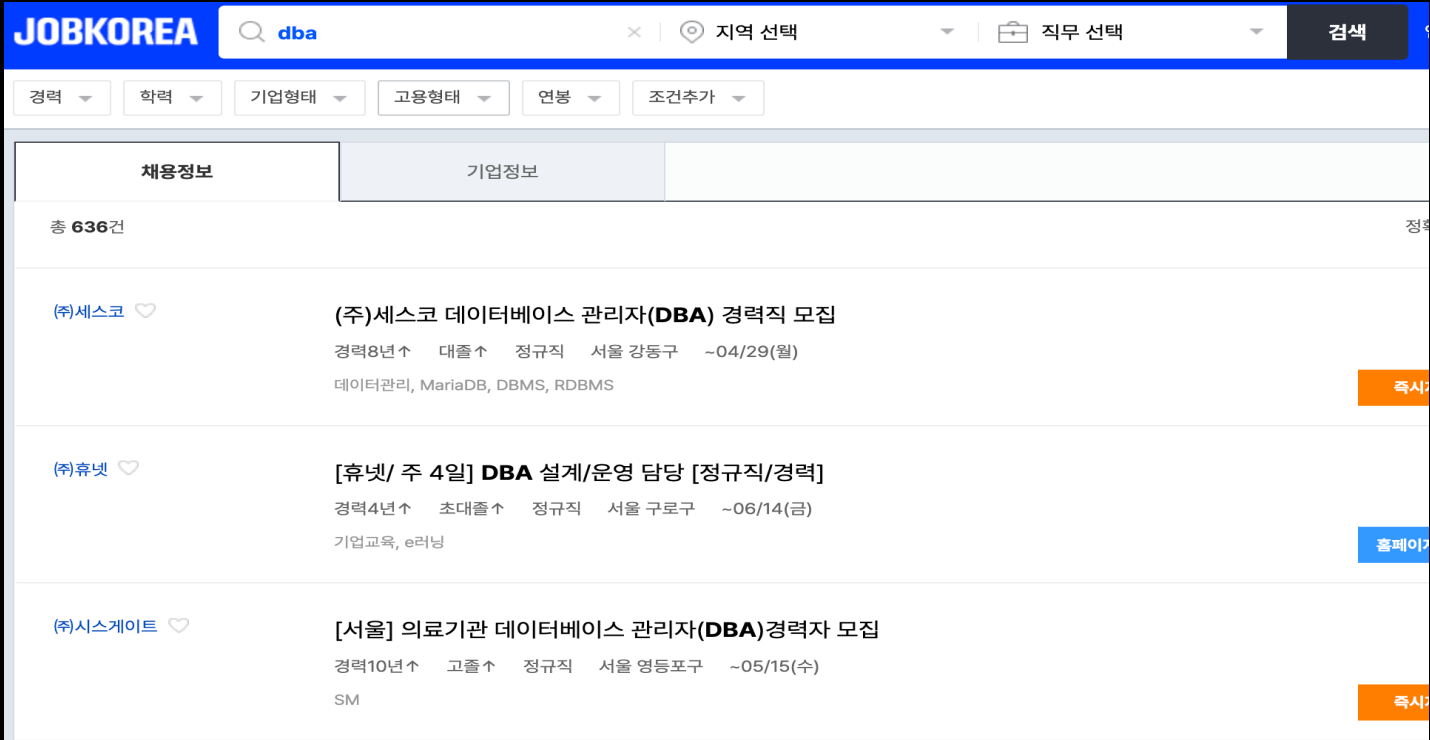
데이터 시각화  
Data visualization

데이터 시각화란 데이터를 시각적으로 표현하여 의사소통 하는  
것으로 많은 양의 데이터를 한 눈에 이해하기 쉽도록 표나  
차트로 정리한다.

Upstage AI Lab

실습 및 예시

: Python 개발 프로젝트



채용 플랫폼 크롤링.

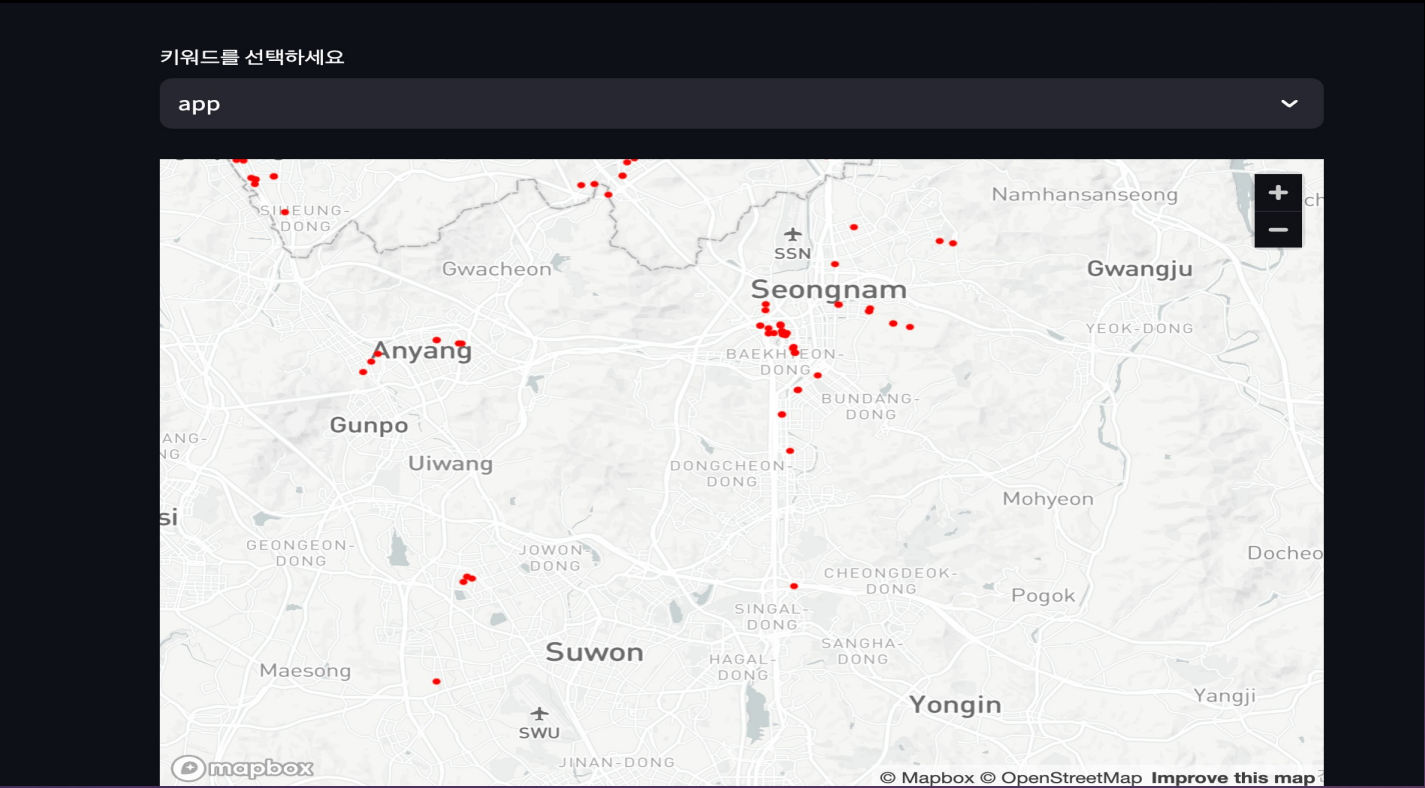
진행 : 사람인, 잡코리아, 원티드의 세가지 플랫폼에서의 IT관련 직종 채용공고 크롤링 진행.

어려웠던 점 : 각 회사별로 상세 주소를 가져오는 과정에서 채용 사이트별로 상세 채용페이지에 있는 플랫폼이 있었고 그것이 아닌 플랫폼이 있는 상황에 회사 상세주소 크롤링하는 작업이 어려웠다.

streamlit 프레임워크 활용.

진행 : 채용플랫폼에서 크롤링한 데이터를 지도와 워드클라우드로 시각화하기 위해 streamlit이라는 파이썬 웹 시각화 프레임워크를 사용

어려웠던 점 : streamlit 자체에서 map 메서드를 제공하는데 자체 map기능에서는 너무 제한적인 것들이 있어 마커를 표시할 때 클릭이벤트를 처리하려고 기획을 짰었는데 제한적인 부분 때문에 마우스 오버 이벤트로 해결방안을 우회했다.







# Google Maps Platform API.

**진행** : 세 개의 채용 플랫폼에서 크롤링해온 회사 상세주소를 Google Map API 중 Geocoding API인 위도, 경도를 가져오는 API를 사용하여 streamlit에서 회사 위치를 마커로 표시할 수 있도록 해주는 기능으로 사용하였다.

**어려웠던 점 :** 팀원들이 API를 전체적으로 익숙지 않아 사용 방법 및 API 요금부과에 대한 두려움으로 사용하기까지 시간과 코드 구현에 어려움이 있었다.

# 워드클라우드.

**진행** : 세 개의 채용 플랫폼에서 크롤링된 데이터 중 채용정보를 취합하여 각 분야별(APP, WEB, DBA, HARDWARE) 등으로 streamlit에 셀렉트 박스로 나누어 워드클라운을 활용하여 단어 빈도 시각화를 하였다.

**어려웠던 점** :처음에 워드클라우드가 Pandas나 Matplotlib에 내장되어 있는 기능인 줄 알고 잘못된 코딩을 계속해서 삽질하고 있었는데 구글링을 통해 빠른 해결을 하게 되었다.





04

# 결과 및 인사이트 공유



# 프로젝트 인사이트 공유

: Python 개발 프로젝트

1



## 인사이트 1

### 실시간 크롤링의 시간적 문제

처음 구현하고자 하는 방식은 검색어와 관련된 채용 정보를 바로 크롤링해서 가져오는 방식으로 하려고 했지만, 데이터가 상당히 많다보니 크롤링하는데 많은 시간이 소요되었습니다.

그래서 실시간 크롤링에서는 시간적인 이슈를 잘 고려하여야 한다는 점을 깨달았습니다.

2



## 인사이트 2

### 무차별적인 크롤링의 해결

프로젝트를 통해 크롤링 과정에서 대상 사이트를 잘 속이는 것이 얼마나 중요한지 알 수 있습니다.

또한 이를 위해 랜덤 한 시간으로 sleep을 줄 시 물리적으로 많은 시간이 들거나 너무 빠른 크롤링으로 차단을 당하는 등 시간 단축과 차단 사이에서 상당한 고민을 해야 됨을 알 수 있었습니다.

3



## 인사이트 3

### 시각화 및 사용자 편리성

아무리 데이터가 많고 잘 가공된 데이터 이더라도 시각화가 잘 되어있지 않고 데이터를 필요로 하는 제3자가 이용하기 불편하면 안 된다 라는 것을 프로젝트 중에 깨달았습니다.

streamlit에서 map 메서드를 활용하여 지도로 시각화도 하였고 워드클라우드를 사용하여 시각화를 해 보았지만 어딘가 한눈에 안 들어오고 사용자 편리성 및 가독성 측면에서 좋지 않다는 점을 알게 되었습니다.

시각화를 할 때는 고민을 많이 하여 어떻게 많은 데이터들을 한눈에 잘 들어오도록 시각화 할 지를 알 수 있었습니다.

05

# 프로젝트 회고



# 프로젝트 진행 느낀점

: Python 개발 프로젝트

## Point 1

개발은 아이디어와 데이터로 이루어진 산출물이다.

이유 : 채용 플랫폼의 데이터를 크롤링하면서 최근 채용시장의 트렌드를 알 수있었으며, 부트캠프 첫 팀프로젝트를 하면서 팀플은 힘들지만 배워가는게 많다라는 것을 또 다시 알게 되었다.

향후 계획 : 이번 프로젝트로 부족한게 많다는 사실을 알게 되었고 개인적으로 공부를 더 많이 해야겠습니다.

## Point 1

streamlit의 편리함과 한계

이유 : streamlit이 얼마나 편리한지 느꼈지만, 한계도 느끼게 되었습니다

향후 계획 : 프로젝트에 맞게 적절한 효율적인 프레임워크를 선택하여 활용할 것이다.

## Point 1

개발에서 제일 힘든 부분이 기획인 것 같다.

이유 : 내가 무엇을 할 수 있는지, 만들고자 하는 것이 실현 가능한지를 가늠하지 못하여 어려워 했던 것 같다. 이러한 문제때문에 프로젝트 도중 기획이 수정되며 진행되었다.

향후 계획 : 내가 무엇이 가능한지 먼저 작성해 나갈 생각이며, 무엇이 간으한지가 확신이 들면 할 수 있는 것과 없는 것이 확실해 질 것이라 생각합니다.

# 프로젝트 진행 소감

: Python 개발 프로젝트



성명기

파이썬을 활용해 데이터크롤링과 FastAPI를 배우면서 파이썬으로 다양한 경험을 해볼 수 있다는 점이 흥미로웠고 앞으로 EDA, ML/DL등 배울 것이 많은데 열심히 개인공부하면서 남은 프로젝트에서 팀프로젝트에서 누가 되지 않도록 노력하겠습니다.

유정수

새로운 경험을 할 수 있어서 보람이 있습니다.

최제우

이번 아직 많이 부족하다는 생각을 하게 되었습니다. 그렇기에 더욱 기반을 튼튼히 다지고 공부할 계기로 삼아야겠습니다.



Life-Changing Education

감사합니다.

---