# 아하모멘트

프리온보딩 챌린지 과정 실전에서 써먹기

강사:김진용

# 목차

# 1. 채용 공고

- a. 데이터 엔지니어
- b. 데이터 사이언티스트
- c. 데이터 분석가

# 2. 포트폴리오 준비하기

- a. 지원하는 회사와 관련된 데이터셋 찾기
- b. 나만의 프로젝트

# 3. 이력서

a. 어떤 내용을 담아야 하는가?



### 1.1 데이터 엔지니어

# **Data Engineer**

### 담당업무

- 분석계 시스템 데이터의 ETL을 통해 data pipeline을 구축합니다.
- Hybrid Multi-Cloud (GCP, AWS, Azure)환경에서 배치, 준실시간, 실시간 Data Platform을 구축합니다.
- · Airflow를 통한 workflow orchestration을 구축합니다.
- · CI/CD pipeline을 구축합니다.
- 통합 모니터링 체계를 구축합니다.

### 자격요건

- 데이터 인프라 구축 및 운영 경험이 있으신 분을 찾습니다.
- · Java/Python/Scala 언어메 능숙하신 분을 찾습니다.
- · GCP 및 Big query에 대한 지식과 경험을 보유하신 분을 찾습니다.
- 워크 플로우 관리를 위해 Airflow 경험 보유하신 분을 찾습니다.

### 우대사항

- · Spark, Hadoop echo system 경험이 있으시면 좋습니다.
- · Azure Devops/Citlab 을 활용한 ci/cd pipeline 경험이 있으시면 좋습니다.
- · Cloud Infra 사용 경험이 (AWS, GCP, Azure 등)있으시면 좋습니다.
- 컴퓨터 공학 전공자를 우대합니다.

출처: https://emartdt.ai/54

### 1.2 데이터 사이언티스트

# **Data Scientist**

#### 합류하면 하게 될 업무

- · AI/ML 모델링 개발 및 상용화를 위해 Data Product Manager/Data Engineer/ML Engineer와 함께 Agile한 개발 사이클에 주도적 참여합니다.
- 프로모션/가격 최적화/수요예측 모델을 개발합니다.
- · 온/오프 통합 Personalization 모델 고도화 개발합니다.

#### 자격요건

- 학사 이상의 학력을 가지신 분을 찾습니다.
- · ML/AI에 대한 이해와 활용능력 보유한 분을 찾습니다.
- · Python/R/Spark 기반 모델링 개발 경험이 있으신 분을 찾습니다.
- · 중급이상 SQL skill이 있는 분을 찾습니다.

#### 우대사항

- 대고객 기반 ML/AI 모델링을 개발/운영 경험하신 분이면 좋습니다.
- 커머스/리테일 비지니스 분석/모델링 상용화를 경험하신 분이면 좋습니다.
- · ML/AI 논문을 기반으로 실 서비스에 적용 가능한 모델 개발 능력이 있으신 분이면 좋습니다.
- 비지니스 관점으로 문제를 해결하고 협업 파트너와 적극적인 협업하는 능력을 갖추신분이면 더 좋습니다.
- 수학/통계/컴퓨터공학/산업공학 등 관련 전공자를 우대합니다

출처 : https://emartdt.ai/ds

### 1.3 데이터 분석가

#### 주요업무

- 데이터 기반(Data Informed) 의사결정에 기여하기 위해 분석 과제를 정의하고 도출된 인사이트를 팀원 모두가 이해할 수 있도록 전달합니다.
- 서버 Database, BigQuery, S3 등에 수집된 Raw Data 를 SQL, Python을 이용해 처리하고 분석에 필요한 데이터로 가공합니다.
- 게임 운영 및 분석에 필요한 핵심 지표와, 데이터가 어떤 것인지 파악하고 필요한 데이터 구조를 설계하고 관리합니다.
- 분석된 내용을 바탕으로 실제 Action 으로 이어질 수 있는 리포트 및 대시보드를 제작합니다. 게임을 개선하기 위한 다양한 가설을 설정하고 이를 검증하기 위한 실험을 설계해 진행합니다.
- 위의 내용을 통하여 아래와 같은 질문의 해답을 찾습니다.
- -Q. "유저의 이탈 패턴은 어떤 것들이 있고, 각 패턴별 특이사항은 무엇인가?"
- -Q. "현재 게임 내에 부족한 자원은 무엇이고, 이를 기반으로 어떤 상품 또는 패키지 출시가 필요한가?"
- -Q. "유저의 플레이 경험을 향상시키기 위한 최적의 컨텐츠 등장 시기는 언제인가?"

#### 자격요건

- 게임회사 인하우스 데이터 분석 경험 1년 이상 보유하신 분
- Python, SQL, BI Tool(Superset, Data Studio, Tableau 등) 활용 역량을 가지신 분
- 데이터 기반 인사이트와 실험 결과를 누구나 이해할 수 있도록 쉽게 커뮤니케이션 할 수 있는 분
- 문제를 정의하고 논리적인 사고를 바탕으로 가설을 증명해 방향성을 이끌어 내실 수 있는 분
- 정확하고 유의미한 데이터 분석 결과를 도출하기 위한 수학적/통계적 지식을 보유하신 분

#### 우대사항

- 게임 좋아하시는 분
- A/B 테스트와 통계에 대한 이해를 바탕으로 목표에 맞는 실험 설계와 가설 검증이 가능하신 분
- 모바일 게임 데이터를 다루고 분석한 경험이 있으신 분
- 게임 데이터 분석 방법에 대한 이해도가 높으신 분 (ARPU, LTV, Retention, Cohort Analysis 등)
- 데이터 기반 조직을 만들기 위해 인프라와 환경을 개선하는데 관심 있으신 분
- BigQuery와 Airflow 를 다양하게 다뤄보신 분
- 데이터 분석 프로젝트 관련 이력서 혹은 포트폴리오를 첨부하신 분

출처: https://www.wanted.co.kr/wd/154954

# 1.3 데이터 분석가

#### 주요업무

- SQL을 사용한 데이터 추출 및 전처리를 통한 지표관리
- 데이터 시각화 및 대시보드(Tableau 등) 제작 및 운영
- 분석 요건 기획 및 데이터 수집 관리
- 데이터에 기반한 가설 수립 및 검증
- 각종 지표 정의 및 트래킹 자동화

#### [전형절차]

서류전형 - > 1차 인터뷰 - > 처우협의 및 최종합격

- \* 지원 포지션 별로 코딩테스트가 있을 수 있습니다. (인터뷰 일정 어레인지 시 별도안내)
- \* 전형과정 중 동의를 통해 레퍼런스체크를 진행할 수 있습니다.

#### 자격요건

- 석사이상의 학위수료자 必
- 데이터 추출, 분석을 위한 언어(SQL, Python)를 능숙하게 다루는 분
- 데이터 시각화 툴(Tableau)에 대한 기본 지식을 보유한 분
- Funnel, AARRR, Cohort Analysis 등 데이터 분석 방법에 대한 지식과 경험이 있으신 분
- 데이터에 기반한 가설을 세워 A/B Test를 설계하고 진행해보신 분
- 마케팅 지표(first open, retention, conversion rate)에 대한 이해도가 높으신 분

#### 우대사항

• AWS(DynamoDB, EC2, RDS), GCP(BigQuery), Google Analytics 사용 경험이 있으신 분

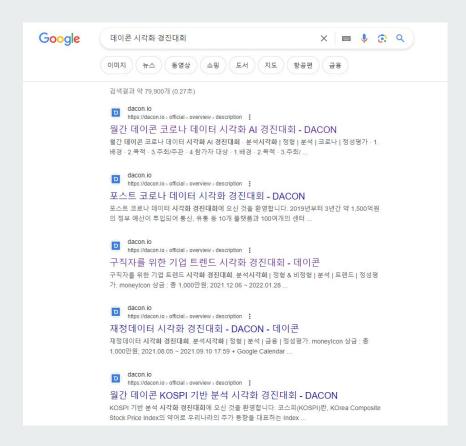
출처: https://www.wanted.co.kr/wd/157552

# 2.1 데이터셋 찾기

### 키워드

데이콘 시각화 경진대회

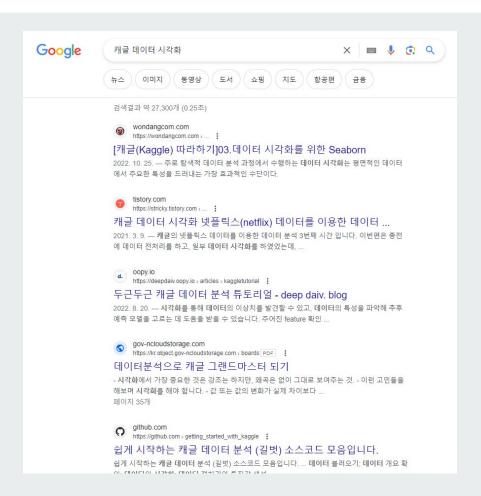
- □ 부동산 데이터 시각화 경진대회
- □ 코로나 데이터 시각화 경진대회
- □ <u>구직자를 위한 기업 트렌드 시각화</u> <u>경진대회</u>



2.1 데이터셋 찾기

키워드

캐글 데이터 시각화

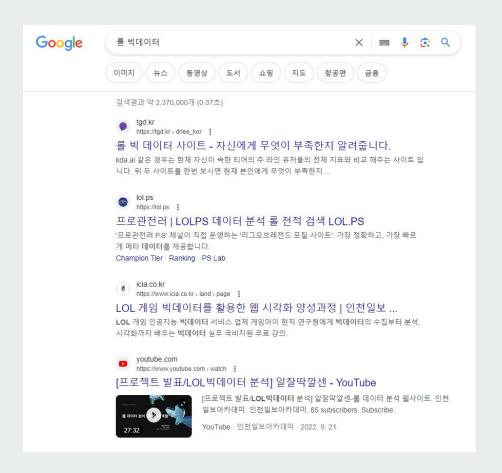


## 2.2 나만의 프로젝트

희망 직무 : 게임 회사의 데이터 분석가

#### 목표 프로젝트

- □ 현재 게임 내에 부족한 자원은 무엇이고, 이를 기반으로 어떤 상품 또는 패키지 출시가 필요한가?
- □ 유저의 플레이 경험을 향상시키기 위한 최적의 컨텐츠 등장 시기는 언제인가?

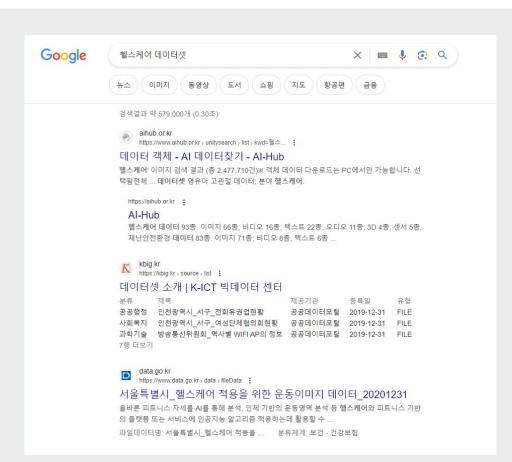


### 2.2 나만의 프로젝트

희망 직무 : 헬스케어 회사의 데이터 분석가

### 목표 프로젝트

- □ 유저의 행동을 변화시키기 위한 최적의 컨텐츠는 무엇인가?
- □ 당뇨병 환자가 혈당을 조절하기 위한 최적의 인슐린 주사양은 어느 정도일까?



### 3.1 어떤 내용을 담아야 하는가?

- 1. 소개
- 회사 입장에서 관심을 끌만한 사람인지를 나타내는 중요한 부분입니다.
- 채용담당자 입장에서 이력서 스크리닝을 할 때 이런 부분이 미흡하다면 뒷부분은 보지 않고 스킵할 수 있습니다.
- 어떤 분야에 관심이 있고 나의 성과가 무엇인지 간단하면서 임팩트 있게 나타내는 것이 바람직합니다. 주니어나 신입 분들은 어떤 기술 스택을 쌓았고 해당 산업 분야에 관심이 많다는 것을 어필하는 것이 좋습니다.(무슨 데이터셋으로 혼자 프로젝트를 해봤다.)

### 3.1 어떤 내용을 담아야 하는가?

1. 소개

### Introduction.

2년 차 개발자로 PyTorch, PyTorch Lightning, OpenCV, Numpy, Pandas, AWS, Git, Linux, FastAPI, Docker 등을 현업에서 활용하고 있습니다. 약 5억원 규모의 국방부 급식 빅데이터 사업에서 데이터 가공/모델 학습 & 추론/데이터 분석/DB 관리/API 서버 구축 등 전반적인 개발 업무를 단독으로 진행하였고, 프로젝트의 One Cycle을 마무리해본 경험이 있습니다.

빠른 성장을 통해 AI 기술 조직에 녹아들고자 다양한 기술에 끊임없이 파고드는 경험을 쌓아가고 있습니다. 이미지, 영상 처리, 최근에는 더 나아가 멀티 모달을 통한 이미지 생성 모델, 3D 데이터에 관심이 많습니다. 가짜 연구소 커뮤니티를 통해 멀티 모달 스터디를 진행했고, VQ-VAE/CLIP/Stable Diffusion/BEIT3 등의 다양한 논문을 리뷰하며 블로그에 기록하고 있습니다.

문제 해결을 위해서 자기 주도적으로 방법을 강구하며 해결 방안을 찾아냅니다. 빛 반사로 인해 노이즈가 있는 3D 이미지를 복원하기 위해서 논문 리서치를 진행했고 관련된 지식을 알고 있는 분들에게 질문하며 피드백을 받아 현장에서 획득된 망가진 3D 이미지를 좋은 성능으로 복원하는 작업을 수행할 수 있었습니다.

### 3.1 어떤 내용을 담아야 하는가?

### 2. Work Experience

- □ 지원하고자 하는 직무와 관련된 경험을 이야기하는 부분입니다.
- □ 직무와 직접적으로 관련된 경험이 바람직하지만 간접적인 경험이더라도 어떤 문제를 해결했는지를 잘 나타낸다면 좋습니다.
- □ 해당 경험에서 본인이 어떤 역할을 맡았는지 그로 인해 어떤 성과를 얼만큼 달성했는지 명시해주면 좋습니다.
- □ 신입, 주니어분들은 인턴 경험이나 학교에서 했던 캡스톤 디자인, 스스로 진행했던 프로젝트 등을 적어볼 수 있겠습니다.
- 해당 경험과 관련해 정리해둔 부분을 Git이나 ppt 형태로 정리하여 링크 형태로 바로가기 만들어주면 좋습니다.

### 3.1 어떤 내용을 담아야 하는가?

### 2. Work Experience

### Work Experience.

#### 누비랩

주방과 식당에서 생성되는 음식 데이터(칼로리 섭취량, 음식 잔반량)를 통해 헬스케어 서비스, 친환경 서비스를 개발하고 있는 푸드 테크 스타트업

#### Food 데이터 엔지니어

22.01.25~23.02.07

- 국방부 사업을 전담하고 있으며 API 서버 개발, DB 관리, 데이터 가공, 모델 학습, AI 및 부 피 분석 알고리즘을 활용한 음식 데이터 분석 등 사업에 관한 전반적인 기술 업무 수행
- Object Detection 모델을 활용한 Auto Labeling으로 Labeling 인력 40% 감축
- depth completion을 통해 및 반사로 망가진 데이터 복원 기술 개발

#### 3D Vision 엔지니어

• 음식의 부피 측정 알고리즘 고도화

23.02.08~now

AI를 활용한 3D Reconstruction 알고리즘 개발

#### 부스트 캠프 Al Tech 5기

AI 관련 지식을 익히고 실무를 경험할 수 있도록 강의 제작을 보조하고 Object Detection 대회를 구성하며 멘토링을 수행

Object Detection 조교

• 강의 자료 검수 및 수정, 베이스라인 코드 검수, 강의 관련 퀴즈 제작, 미션/실습코드 검수

23.03.06~23.05.19

• Real Time Object Detection Yolo v8 주제로 오피스아워(1H) 진행

### 3.1 어떤 내용을 담아야 하는가?

### 2. Work Experience

- □ 지원하고자 하는 직무와 관련된 경험을 이야기하는 부분입니다.
- □ 직무와 직접적으로 관련된 경험이 바람직하지만 간접적인 경험이더라도 어떤 문제를 해결했는지를 잘 나타낸다면 좋습니다.
- □ 해당 경험에서 본인이 어떤 역할을 맡았는지 그로 인해 어떤 성과를 얼만큼 달성했는지 명시해주면 좋습니다.
- □ 신입, 주니어분들은 인턴 경험이나 학교에서 했던 캡스톤 디자인, 스스로 진행했던 프로젝트 등을 적어볼 수 있겠습니다.
- 해당 경험과 관련해 정리해둔 부분을 Git이나 ppt 형태로 정리하여 링크 형태로 바로가기 만들어주면 좋습니다.

## 3.1 어떤 내용을 담아야 하는가?

### 3. Education

- □ 해당 직무와 관련해 역량을 기르기 위해 어떤 교육 과정을 수료했는지 무엇을 배웠는지를 적습니다.
- □ 이번 프리온보딩 데이터 챌리지 커리큘럼을 통해 학습했던 내용을 정리해보고 이력서에 어떻게 표현할지 고민해봅니다.

# 3.1 어떤 내용을 담아야 하는가?

### 3. Education

Naver Boost Camp Al TECH 27

| 인공지능, 딥러닝 활용 방법과                 | Al Production의 End-to-End 교육 프로그램   |
|----------------------------------|---|
| Image Classification Competition | <ul> <li>이미지 분류 모델을 활용한 마스크 착용 / 성별 / 나이에 대해 총 18개의 클래스로 나눠 판별<br/>하는 시스템을 구현하는 대회</li> </ul> |
| 21.08.23~21.09.02                | • 이미지에서 사람 얼굴을 탐지 및 크랍하여 모델의 성능 개선  |
| Object-Detection<br>Competition  | Object detection 모델을 활용해 쓰레기를 탐지하고 10종류의 쓰레기 class를 detection하는 시스템을 구현하는 대회                  |
| 21.09.27~21.10.15                | Heavy Augmentation, 앙상블 기법을 적용하여 모델의 성능 개선  |
|                                  |   |

| Semantic Segmentation<br>Competition | • Segmentation 모델을 활용해 쓰레기를 탐지하고 10종류의 쓰레기 class를 segment 단위<br>로분류하는 시스템을 구현하는 대회 |
|--------------------------------------|--|
| 21.10.18~21.11.04                    | 재활용 쓰레기를 copy paste하여 모델의 성능 개선  |
| 글자 검출 대회                             | • 글자 검출 모델을 활용해 글자 검출 시스템을 구현하는 대회   |
| 21.11.08~21.11.18                    | • 오라벨링된 데이터 검수 후 삭제하여 모델의 성능 개선  |
| 최적화 대회                               | <ul> <li>AutoML, Quantization 등을 활용한 모델 경량화 및 최적화 대회</li> </ul>                    |
| 21.11.22~21.12.02                    | • Optuna를 활용해 100번의 모델 실험하여 모델의 성능 개선  |
| 모델 서빙                                | • "간편한 식단 관리 Food Log 어플리케이션" 개발   |
| 21.11.08~21.12.27                    | • 데이터 전처리, Heavy Augmentation, fast API로 GCP에 모델 배포                                |

- 3.1 어떤 내용을 담아야 하는가?
- 4. 해당 직무와 직접적인 관련은 없지만 자기 계발과 관련한 경험

## Other Experience.

#### 자격증

#### English

TOEIC - 825/990 (2021.03.28 일 / 수험번호 - 168450)

OPIC - IM2 (2021.03.31(수) / 수험번호- A9686704168)

#### 인터뷰

- <청년꿈터> 미래를 향한 진지한 고민과 과감한 용기 김진용 2022.12.01
- 네이버 부스트캠프 온라인 설명회 「주저하고 계시다면 용기내서 지원해보세요」 2022.7.1
- <스브스뉴스> 누비랩의 음식 스캔 기술 2022.10.20

#### 스터디

- 알고리즘 스터디 다양한 알고리즘 문제를 같이 풀어보며 GitHub에 공유 2021.09~2021.12
- <u>가짜 연구소 논문 스터디</u> 2022.09 ~ 2022.12

#### ETC.

• 오픈 소스 기여 yolov5-fastapi-demo 버그 수정 2021.12

- 3.1 어떤 내용을 담아야 하는가?
- 5. 사용할 수 있는 기술
- □ 해당 스택과 숙련도는 어느 정도인지 표현해주면 바람직합니다.
- □ pandas : 중급, matplotlib : 상급, mysql : 초급

# Skill.

- AI: Pytorch, Pytorch Lightning, AWS(EC2, S3, RDS), Fast API, Docker(Docker-hub, CI/CD)
- Data: OpenCV, Pandas, Numpy, Matplotlib, MySQL, Pymysql, boto3
- Language: Python

Q&A