

**2024**

한 권으로 마무리하는 직무 면접

# 제로베이스 면접 실전 기출문제

5,000제 이상의 기출문제 중, 과정 별 300문제 엄선  
실무에서 물어보는 빈출 문제+합격자 답변과 함께 선별

DATA SCIENCE

**zero-base**

# 제로베이스 소개

## 안녕하세요. 제로베이스 스쿨입니다.

제로베이스는 대학 교육과 실무의 괴리를 해결하기 위해 바로 실무에 투입이 가능한 인재를 양성하고 있습니다. 취업·이직·전직을 희망하는 분들을 대상으로 각 과정의 기초부터 결과물을 만드는 것까지 진행하는 취업 집중 교육과정입니다.

현재 제로베이스에선 개발, 데이터, 기획, 마케팅, 디자인 등 다양한 분야에서 취업에 직접 영향을 주는 실무 교육 과정을 운영하고 있으며, 전에 없던 새로운 분야의 취업 교육을 계속해서 선보이고 있습니다.

## 취업보장 혜택을 자신 있게 내놓을 수 있는 이유

제로베이스에선 과정 졸업 후에도 취업이 되지 않을 경우, 수강료 전액을 환급해 드리는 '취업보장' 혜택을 제공합니다. 어떻게 이런 혜택이 가능할까요?

### 1. 누적 수강생 10,000명 이상, 데이터로 쌓아 올린 커리큘럼

제로베이스 스쿨은 2016년 패스트캠퍼스 스쿨을 시작으로, 1만명에 달하는 수강생과 함께 국내 최고 수준의 취업 결과를 만들고 있습니다. IT 기업 취업에 필요한 어디에서 찾기 어려운 규모의 온라인 강의 콘텐츠, 과제, 테스트, 최대 규모의 기출문제까지 따라올 수 없는 초격차를 만들어내고 있습니다.

### 2. 실무 영역을 정의하는 5,000개 이상의 기출문제

입시, 자격증, 공무원 등등. 모두 '기출문제'를 기반으로 공부하는데, 과연 '취업'은 어떨까요? 제로베이스에선 면접 및 채용 현장에서 나온 5,000개 이상의 문제를 분석해, 기업에서 신입 취준생에게 원하는 지식과 역량을 정확하게 정의했습니다.

### 3. 50만 기업 회원과 함께, 사람인 취업 연계 서비스

서류와 면접만으로는 모든 역량을 검증하기 어려운 시대, 취업을 준비하는 과정까지 들여다 볼 수 있는 '사람인 채용연계' 서비스가 2024년부터 시작되었습니다.

사람인의 50만 기업회원에게 학습 리포트와 포트폴리오, 데모영상을 통해 차별화된 지원자로 먼저 제안을 받을 수 있습니다.

이외에도, 제로베이스의 자세한 취업 서비스가 궁금하다면? 제로베이스 홈페이지를 확인해 보세요!

<https://zero-base.co.kr/>

# 기출문제집 활용법

**[제로베이스 면접 실전기출문제]**는 제로베이스 스쿨에서 다양한 경로를 통해 쌓아 올린 면접 기출 문제 은행 자료 중, 일부를 추려 구성한 문제집입니다. 아직 면접을 경험해본 적이 없거나, 당장 눈앞에 면접을 앞둔 분들께는 더욱 큰 도움이 될 수 있을 것 같습니다. 이 자료는 세 가지 파트로 구성되어 있으며, 각 파트별로 용도에 맞게 활용할 수 있습니다.

## 1. 직무 기출문제 300제

내 직무 면접의 카테고리별 유형을 확인하고, 어떤 질문들이 주로 나오는지 확인해 보세요. 이 면접 질문들은 실제 채용 현장에서 적어도 한 번 이상은 출제된 적 있는 문제들입니다. 양이 너무 많다면? 뒤에 등장하는 빈출문제부터 차근차근 준비해 보세요.

## 2. 빈출 문제 / 합격 답변 30제

방대한 분량의 면접 질문 중에서도, 특히 높은 빈도로 출제되는 질문입니다. 물론 300가지 질문의 답변을 모두 준비할 수 있다면 좋겠지만, 우리의 시간은 한정되어 있고 때론 선택과 집중을 해야 할 때가 있습니다. 모든 질문에 대비하진 못하더라도, 이 30가지 질문만큼은 꼭 숙지해서 면접장에 들어가세요.

## 3. 예상문제 10제

당장 면접장에 가야 하는데 준비할 시간이 충분치 않다면? 적어도 이 질문만큼은 대비해 가세요. 기업 입장에선 꼭 확인하고 싶지만, 미리 준비하지 않으면 대답하기 어려운 10가지 문제를 엄선했습니다. 해당 질문들을 출력한 뒤, 아래 빈칸에 자신의 답변 키워드를 메모하여 함께 가지고 가는 것도 추천드려요!

제로베이스는 언제나  
언제나 여러분의 취업 성공을 위해 함께 달리겠습니다.

“  
본 자료는 아래 기업들을 포함해  
총 232개의 기업 면접 문제를 바탕으로  
제작되었습니다.

”

kakao

kakaoenterprise

wadiz

IBM

 kakao bank

 imbc

 HYUNDAI  
AutoEver

 woongjin

 coupang

 SERVICE

 COMMERCE

 당근마켓

kakaohealthcare

 KEB 하나은행

 롯데홈쇼핑

 weverse  
Official for All Fans

 Danal

 11D

 대학내일

 신한카드

 HYUNDAI  
TRANSYS

 CJ ENM

 해커스

 toss

 NEXON

 LG CNS

 CJ OLA  
올리브영

 HYPER  
CONNECT

 megastudy

 NHN

 여기어디.

 배달의민족

 CATCHTABLE

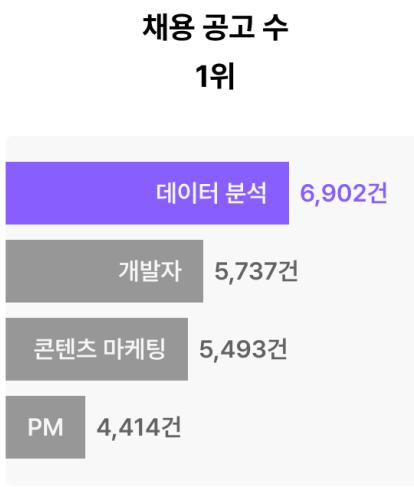
 NAVER  
WORKS

 HD

## ▶ Why : 데이터 직무를 왜 해야 할까요?

데이터 관련 직업이 핫하다는 말은 들어 보셨을 텐데요. 요즘 데이터 직무는 세분화가 진행되었고, 지원자도 많이 증가했다고 합니다. 이미 지원자는 많은 것 같은데 데이터 직무를 공부하면 안되는 것 아니냐고요? 밑에서 확인해 보세요!

### 1. 채용 공고 수 1위 직무



### 2. 세계경제포럼이 뽑은 미래 유망 직업 (1,3,6 위)

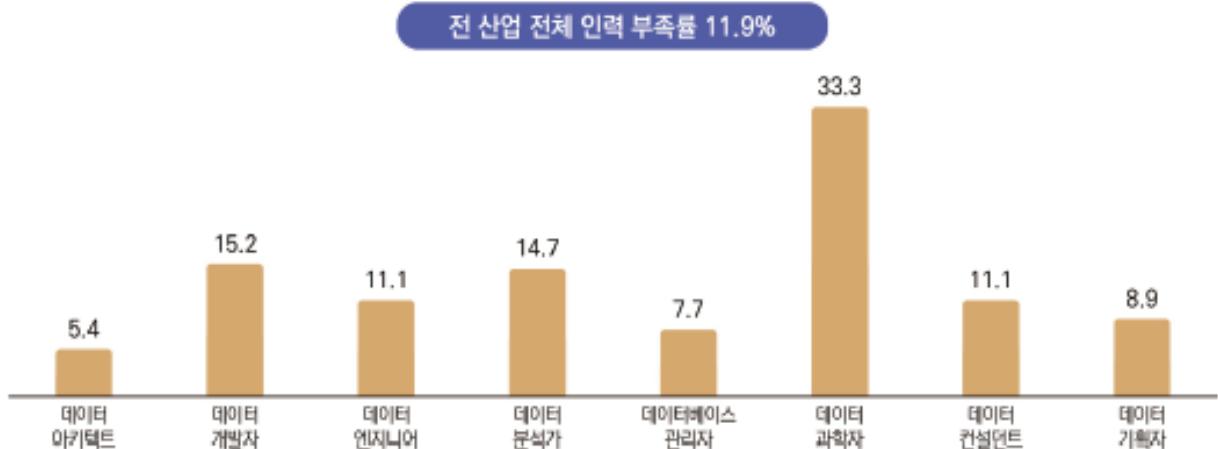
#### Top 10 fastest growing jobs

1. AI and Machine Learning Specialists
2. Sustainability Specialists
3. Business Intelligence Analysts
4. Information Security Analysts
5. Fintech Engineers
6. Data Analysts and Scientists

### 3. 산업 전체 데이터 인력 부족률 평균 12%

그림 3-1-17 국내 전 산업의 향후 5년 내 데이터 직무 인력 부족률

(단위: %)



\* 통계결과는 반올림되어 부분의 합계가 전체와 일치하지 않을 수 있음

## ▶ What : 데이터 직무는 무엇을 할까요?

데이터 직무가 뜬다고 하지만 막상 어떤 일을 하는지, 어떤 직무가 있는지 알기 어려웠을 거예요.  
제로베이스가 데이터 직무를 파헤쳐 드리겠습니다!

직무를 소개하자면 [데이터 수집, 저장, 분석, 해석] 등 데이터와 관련된 업무를 맡습니다.  
대표적으로 ① 분석가와 ② 사이언티스트가 있어요. 두 직무를 중심으로 밑에서 더 자세하게 다뤄보도록 하겠습니다.

	데이터 분석가	데이터 사이언티스트
업무 내용	<ul style="list-style-type: none"><li>- 주로 데이터를 시각적으로 이해하기 쉽게 추출, 변형을 통해 정리하고 표현합니다.</li><li>- 다른 부서와 협업을 통해 논리적으로 설득하고 인사이트를 제공합니다.</li><li>- ex : 기상예보 그림, 지하철 혼잡도 등</li></ul>	<ul style="list-style-type: none"><li>- 데이터를 활용하여 미래를 예측하거나 추천해주는 모델(서비스)을 만듭니다.</li><li>- ex : 유튜브 알고리즘, 인스타그램 추천, 구글 제품 광고</li></ul>
공통 요구 역량		
Python	기본 수준 요구 : 기본 이해, 다른 분석 툴(R) 가능	상 수준 필요
SQL	중~상 수준 요구	중~상 수준 요구
시각화	상 수준 필요 : 잘 다루는 시각화 툴 필요 (ex : Tableau, Power BI)	중 수준 필요
기초 수학(통계)	기본 수준 요구	중~상 수준 요구 : 통계 데이터 활용 가능 (ex : 사설 검정)
직무별 요구 역량		
	커뮤니케이션 역량, 도메인에 대한 이해 (서비스, 상품 등)	데이터 모델 구축, 모델 평가, 모델 고도화 및 운영

## ▶ How : 직무 역량은 어떻게 기를 수 있을까요?

데이터 현직자 인터뷰 100개, 2023년 하반기 채용 공고 100건을 모두 분석한 결과!

아래에서 공유드립니다! 과연 기업이 데이터 직무에게 원하는 역량은 무엇일까요?

기업이 원하는 데이터 분석가와 데이터 사이언티스트의 요구 역량을 정리해드립니다.

+ 제로베이스 스쿨에서 그 역량을 어떻게 기를 수 있는지도 함께 전달 드립니다.

### 1. 데이터 분석가

요구 역량	데이터 분석가	제로베이스 스쿨	현직자 한마디
공동 1위	유관 경험	실무 프로젝트	- 야놀자 데이터 분석가 L님 데이터를 추출한 후 가공하는 역량은 모든 데이터 직무의 기초, SQL 역량이 중요합니다.
공동 1위	SQL	SQL 중심 커리큘럼	- 데이터 분석가는 데이터를 활용하여 현재 상황에 대해 설명하는 보고서, 시각화 자료를 통해 여러 연관 부서와 협업을 진행합니다.
2위	포트폴리오 (프로젝트, 공모전)	현직자 포트폴리오 첨삭, 프로젝트 멘토 피드백	- 따라서 도메인(서비스, 상품 등) 이해, 커뮤니케이션 역량이 중요합니다.
3위	도메인 지식 (서비스, 상품)	20개 도메인별 프로젝트 보유	
4위	기초 수학	필수 수학 통계 지식	

### 2. 데이터 사이언티스트

요구 역량	데이터 사이언티스트	제로베이스 스쿨	현직자 한마디
1위	유관 경험	실무 프로젝트	- 네이버 데이터사이언티스트 K님
2위	유관 전공	대학 교수진(한양대), 필요한 전공 과목만 학습	- 데이터 사이언티스트 직무의 특징은 머신러닝, 딥러닝을 활용하여 모델(서비스)을 구축한다는 점입니다.
3위	포트폴리오 (프로젝트, 공모전, 논문)	현직자 포트폴리오 첨삭, 프로젝트 멘토 피드백	- 모델은 비즈니스 속 문제들을 해결하며 예측하는데 목적을 둡니다. 따라서 모델을 운영하고, 평가하며, 성능을 고도화 시키는 것에 집중합니다.
4위	ML/DL 개념 (알고리즘)	ML/DL 3개월 이상 학습, 알고리즘 강의 부록 제공	

## ▶ And : 데이터 직무 취업? 제로베이스!

제로베이스에서는 데이터 관련된 모든 역량을 기를 수 있습니다! AI 직무 취업을 원하신다면 [AI 취업 스쿨], 데이터 분석가 취업을 원하신다면 [데이터 분석 취업 스쿨]을 추천드립니다! 2가지 커리큘럼을 비교하며, 나는 어떤 스쿨과 맞을지 확인해보세요.

제로베이스 스쿨 교육 과정						
스쿨명	AI 스쿨			데이터 분석 스쿨		
수강 시간	6개월			4개월		
목표 직무	분석가, 사이언티스트			분석가		
학습 내용						
커리큘럼	Part 1	Python	공통	Part 1	Python	공통
		기초수학			기초 수학	
		자료구조/알고리즘			SQL	
	Part 2	EDA/SQL/Git		Part 2	SQL/시각화	
		통계/태블로			(태블로/파워BI/루커)	
	Part 3	머신러닝		Part 3	SQL + Python 고급	
	Part 4	Tensorflow/PyTorch			EDA 고급	
		딥러닝/openCV/GPT				
프로젝트	프로젝트1	머신러닝 프로젝트		프로젝트1	산업군별 20가지 샘플 프로젝트	
	프로젝트2	딥러닝 프로젝트		프로젝트2	산업군별 20가지 개인 프로젝트	
	프로젝트3	파이널 프로젝트		프로젝트3	팀 프로젝트	
	프로젝트4	코딩테스트 강의		프로젝트4	파이널 프로젝트	

**PART**

**4**

---

**기출문제**

---



No	카테고리	문제
문제1	MLOps	운영 아키텍처를 MSA로 가져갔을 때 장점에 대해 설명해 주세요.
문제2	MLOps	(프로젝트 기반) 머신러닝 파이프라인을 어떻게 구성하셨는지 설명 해주세요.
문제3	MLOps	MLOps와 DevOps의 차이에 대해 설명해 주세요.
문제4	MLOps	데이터 드리프트에 대해 설명해 주세요.
문제5	MLOps	데이터 드리프트를 해결하기 위한 방법에 대해 설명해 주세요.
문제6	MLOps	모델 레지스트리에 대해 설명해 주세요.
문제7	MLOps	모델 거버넌스를 위한 요소들이 어떤게 있을지 설명해주세요.
문제8	프로젝트 경험	데이터 분석 프로젝트를 다양하게 했는데, 어떤 식으로 수행했는지가 궁금합니다.
문제9	프로젝트 경험	외부 데이터를 프로젝트에 통합할 때 고려해야 할 사항은 무엇인가요?
문제10	프로젝트 경험	데이터 분석 프로젝트의 우선순위와 자원 할당 방식에 대해 어떻게 결정하나요?
문제11	프로젝트 경험	대표 프로젝트의 성과를 설명해 주세요.
문제12	프로젝트 경험	프로젝트에서 사용한 샘플 개수는 몇개나 되나요?
문제13	프로젝트 경험	대표 프로젝트에서 맡은 역할은 무엇인가요?
문제14	프로젝트 경험	이미지 데이터를 다룰 때, 어떤 전처리를 해 보았나요?
문제15	프로젝트 경험	음성 데이터를 다룰 때, 어떤 전처리를 해 보았나요?
문제16	프로젝트 경험	NH투자증권 빅데이터 경진대회에서 사용한 모델을 설명해주세요
문제17	프로젝트 경험	그 모델을 사용해서 원하는 결과를 얻었나요?
문제18	프로젝트 경험	공공 빅데이터 인턴 당시 진행한 프로젝트에 대해 설명해주세요
문제19	프로젝트 경험	경진대회에서 사용한 모델 구조는 어떤 이유로 그렇게 구성하셨나요?
문제20	프로젝트 경험	팀프로젝트를 하면서 힘들었던 경험이 있나요?



No	카테고리	문제
문제21	프로젝트 경험	(프로젝트 기반) 사용한 추천모델과 왜 사용했는지 이유를 설명해주세요.
문제22	프로젝트 경험	(프로젝트 기반) 추천모델을 평가하기 위해 어떤 메트릭을 사용했고, 그 이유를 설명해주세요.
문제23	프로젝트 경험	프로젝트 진행시 어려웠던 점과 해결 방법에 대해 설명해 주세요.
문제24	프로젝트 경험	프로젝트에 사용된 모델의 성능을 개선하기 위해 어떤 후속 작업이 가능할 것이라 생각하나요?
문제25	프로젝트 경험	진행했던 프로젝트 중 가장 임팩트가 있었던 프로젝트에 대해 이야기해 주세요. 이유는 무엇인가요?
문제26	프로젝트 경험	고객 수익 예측 및 vip 분류 프로젝트에서 파생 피쳐로 상품 관심도를 생성했던데, 해당 피쳐는 어떻게 생성하였나요?
문제27	프로젝트 경험	고객 수익 예측 및 vip 분류 프로젝트의 수익 예측 모델링에서 구매고객을 먼저 분류하고, 구매고객의 수익을 예측하는 식으로 프로젝트를 진행하면서 하나로 합쳐서 미구매고객의 수익을 0으로 예측하는 대신 구매고객과 미구매고객을 분리하여 예측한 이유가 있나요?
문제28	프로젝트 경험	프로젝트에서 본인이 직접 발굴했던 피쳐 중 가장 성과가 좋았던 피쳐를 하나 꼽아서 피쳐 발굴 과정부터 data wrangling 과정과 이유를 설명한다면?
문제29	프로젝트 경험	LSTM 모델의 특장점에 대해 설명해 주세요.
문제30	프로젝트 경험	고객 수익 예측 및 vip 분류 프로젝트 결과 수익 예측 모델의 RMSE는 높았지만 vip 모델 분류 결과의 한계점으로 인해 프로젝트 결과물을 실서비스 적용하기에는 어려워 보이는데요, 한계점들을 극복하기 위해 next step으로 어떤 시도를 해볼 수 있을까요?
문제31	프로젝트 경험	고객 수익 예측 및 vip 분류 프로젝트 결과를 실서비스에 적용하기 위해 RFM 모델 결과를 가지고 VIP를 재분류한다면, 어떤 시도를 해볼 수 있을까요?
문제32	프로젝트 경험	음원스트리밍 서비스 사용자 이탈 여부 예측 프로젝트에서 나이, 성별의 이상치 비중이 높아 모델링에 사용하지 않았다고 하던데, 해당 피쳐가 예측에서 중요하지 않은 피쳐인지 확인하는 과정을 거쳤나요? 일반적으로는 고객의 데모 정보는 주요 피쳐 중 하나인데 이를 제외함으로써 모델의 성능에 영향이 있지는 않았을까요?
문제33	프로젝트 경험	세상에서 가장 가벼운 단기 전력 수요 예측 모델 프로젝트의 진행 배경과 과정, 결과를 요약해 설명해 주세요.
문제34	프로젝트 경험	세상에서 가장 가벼운 단기 전력 수요 예측 모델 프로젝트에 사용한 BiLSTM 모델의 특장점에 대해 설명해 주세요.
문제35	프로젝트 경험	각 프로젝트 별로 본인이 맡았던 역할은 무엇이었나요?



No	카테고리	문제
문제35	프로젝트 경험	각 프로젝트 별로 본인이 맡았던 역할은 무엇이었나요?
문제36	프로젝트 경험	C2C (개인간거래) 플랫폼을 위한 불량 제품 검수 모델 개발 프로젝트에 사용한 YOLO 모델의 특장점에 대해 설명해주세요
문제37	프로젝트 경험	C2C(개인간거래) 플랫폼을 위한 불량 제품 검수 모델 개발 프로젝트의 최종 모델을 실서비스에 적용한다면, 판매자가 사진을 제대로 촬영해야 할텐데 이 점은 데이터적으로 어떻게 보완할 수 있을까요?
문제38	프로젝트 경험	학부 전공은 데이터직무와 무관하던데, 데이터직무로 지원한 계기나 배경이 있을까요?
문제39	프로젝트 경험	기업 데이터 분석을 통한 매출 예측 프로젝트의 매출 예측 정확도가 높았던 부분은 이해했는데, 이로 인해 얻을 수 있는 분석 결과가 무엇이었나요?
문제40	프로젝트 경험	기업 데이터 분석을 통한 매출 예측 프로젝트의 목적이 매출 분석을 통한 효과적인 마케팅 전략 수립이라고 되어 있던데, 분석한 내용을 기반으로 어떤 마케팅 전략을 수립할 수 있을까요?
문제41	프로젝트 경험	분석 목적에 맞게 분석 결과를 기반으로 인사이트를 도출하여 가장 큰 성과를 낼 수 있을 거라 생각하는 프로젝트는 무엇이고, 어떤 성과를 낼 수 있을 것이라 기대하며, 왜 그렇게 생각하나요?
문제42	프로젝트 경험	기업 데이터 분석을 통한 매출 예측 프로젝트에서 주/월별 예측보다 분기별 예측의 정확도가 더 높았다고 하셨는데, 그렇다고 분기별로 예측을 하게 되면 텀이 너무 길진 않을까요? 분기별로 예측을 해도 비즈니스적으로 효용 가치가 있을까요?
문제43	프로젝트 경험	CCTV 차종 분류를 통한 교통량 분석 프로젝트의 목적이 차종, 교통량 파악하여 도로관리, 어린이보호구역 안전조치, 미세먼지 저감조치라고 하셨는데, 분석 결과를 통해 이 분석 목적을 어떻게 달성할 수 있을까요?
문제44	프로젝트 경험	고객 분석을 통한 고객 이탈 예측 프로젝트 결과 20000명 고객중 25%가 이탈 할 것으로 예측하셨는데요, 정확도가 어느 정도 되는 결과일까요? 또 이 이탈예측 결과를 비즈니스에 활용하려면 어떻게 해석해야 할까요? (언제쯤 이탈한다는 의미인지 등)
문제45	프로젝트 경험	본인의 SQL 역량을 상이라 기재해주셨던데, 그렇게 판단한 이유와 기준이 무엇인가요?
문제46	프로젝트 경험	스타벅스 & 이디야 커피 매장 위치적 상관관계 분석 결과, 이디야 커피 매장은 전략적으로 스타벅스 매장 근처에 위치시킨다는 가설은 성립되지 않는다는 점을 파악한 것 같은데 이로 인해 얻을 수 있는 비즈니스 인사이트는 무엇인가요?
문제47	프로젝트 경험	셀프 주유소 & 일반 주유소 간의 가격 분석 프로젝트에서 가격을 어떤 통계 검정 과정을 거쳐 비교하였나요?



## No 카테고리

## 문제

문제48	<b>프로젝트 경험</b>	셀프 주유소는 일반 주유소에 비해 저렴하다는 가설이 성립되었다는 점을 통해 어떤 비즈니스 인사이트를 얻을 수 있을까요?
문제49	<b>프로젝트 경험</b>	의류 실측 데이터 기반 프리사이즈 분석 및 개인화 의류 추천 시스템을 구현 하셨던데, 해당 시스템을 통해 어떤 비즈니스적 효과가 있을 것이라 예상하고, 해당 효과의 성과는 어떤 지표를 통해 측정할 수 있을까요?
문제50	<b>프로젝트 경험</b>	일반적으로 당연하다고 여겨지는 가설도 검증할 필요가 있다고 생각하시나요? 답변의 이유는 무엇인가요?
문제51	<b>프로젝트 경험</b>	개인적으로 진행한 프로젝트들 중 비즈니스적으로 임팩트가 가장 클 거라 생각하는 프로젝트는 무엇이고, 그 이유는 무엇인가요? (비즈니스 성과를 기대 했다기 보다 호기심 충족을 위한 프로젝트로 보여져서)
문제52	<b>프로젝트 경험</b>	정확도로는 매우 높은 성과를 보였지만 비즈니스 성과를 내기 어려운 모델과 정확도는 다소 떨어지더라도 비즈니스 활용도가 높은 모델 중 어떤 모델이 더 효용가치가 높다고 생각하는지? (답: 회사 입장에서는 당연히 후자임)
문제53	<b>프로젝트 경험</b>	의류 실측 데이터 기반 프리사이즈 분석 및 개인화 의류 추천 시스템 프로젝트가 비즈니스 성과를 내려면 배경으로 고객에게 맞는 사이즈의 옷일수록 구매전환율이 높고 반품율이 낮다는 것이 입증되어야 할 것 같은데, 이러한 가설 검증 없이 고객에게 맞는 사이즈의 추천 아이템을 보여주는 것만으로 어떠한 임팩트가 있을 것이라 예상하는지?
문제54	<b>프로젝트 경험</b>	명품 커머스의 VoC 분석과 분류 모델 개발 프로젝트를 통해 구현한 VoC 분류 서비스를 활용하여 회사 입장에서 어떤 비즈니스 효과를 어떻게 얻을 수 있을까요?
문제55	<b>프로젝트 경험</b>	명품 커머스의 VoC 분석과 분류 모델 개발 프로젝트에 사용한 BERT 모델에 대해 간단히 설명해주실 수 있을까요?
문제56	<b>프로젝트 경험</b>	실내(편의점, 무인점포) 절도행위 탐지 모델 개발 프로젝트에서 선정된 최종 모델은 무엇이었고, 정확도가 어느 정도였나요?
문제57	<b>프로젝트 경험</b>	강남구 위치 기반 시설 밀집도 분석 프로젝트를 통해 밀집도 레벨을 분류하고, 레벨별 특성을 시각화를 통해 정리하신 것 같은데요, 시각화 자료를 통해 얻은 분석 결과를 좀더 자세히 설명해주실 수 있을까요?
문제58	<b>프로젝트 경험</b>	강남구 위치 기반 시설 밀집도 분석 프로젝트를 통해 부동산과 도시개발에 대한 인사이트를 제공하고자 했다고 하셨는데, 결론적으로 어떤 인사이트를 얻으셨나요?
문제59	<b>프로젝트 경험</b>	전자제품 온라인 스토어 매출 분석 프로젝트에서 이탈 고객의 특징으로 86%의 이탈고객이 1개 상품만 조회하고 이탈한 부분을 정의하고, 다양한 상품을 조회할 수 있도록 해야한다고 해주셨는데요, 데이터적으로 어떤 프로젝트를 통해 유저가 다양한 상품을 조회할 수 있도록 할 수 있을까요?



No	카테고리	문제
문제60	프로젝트 경험	전자제품 온라인 스토어 매출 분석 프로젝트에서 장바구니 전환율 개선을 위해 여러 상품을 조회하고 장바구니에 담을 수 있도록 추천 시스템을 강화해야 한다고 하셨는데, 장바구니 전환 관련해서는 기간적인 이슈만 파악하신 게 아니었나요? 다양한 상품을 조회하도록 하는 것이 장바구니 전환율을 높이는 데에 높은 상관관계를 보인 근거가 있나요?
문제61	프로젝트 경험	패션 리뷰 이미지 등급 분류 모델 개발 프로젝트에서 리뷰가 있는 제품이 없는 제품보다 구매전환율과 장바구니비율 모두 높다고 되어있던데, 이 부분이 통계적으로 유의성 검정을 거친 결과일까요?
문제62	프로젝트 경험	패션 리뷰 이미지 등급 분류 모델 개발 프로젝트에서 리뷰등급이 없는 29CM 와 등급이 있는 무신사의 데이터를 비교하셨던데, 가설 검증을 위해 구체적으로 어떤 과정을 거치셨나요?
문제63	프로젝트 경험	패션 리뷰 이미지 등급 분류 모델 개발 프로젝트에 ResNet-50 사전 학습 모델을 사용하셨던데, 그 이유와 해당 모델의 특장점을 설명해주실 수 있나요?
문제64	통계	비모수검정이 무엇이고 어떤 것들이 있나요?
문제65	통계	베이즈 통계가 뭔지 아나요? 아시면 간단하게 설명해 보실 수 있나요?
문제66	통계	회귀 분석과 분류 분석의 차이점을 설명해주세요.
문제67	통계	이상치 감지 방법과 그 중요성을 어떻게 생각하시나요?
문제68	커뮤니케이션	데이터분석에서 분석결과가 특정부서의 성과에 대해 부정적인 피드백을 내포할 때, 그 부서와의 갈등을 어떻게 풀 수 있을까요?
문제69	커뮤니케이션	분석 후 내용을 어떻게 효과적으로 상급자에게 전달할 수 있을까요?
문제70	커뮤니케이션	최종 만들어진 모델의 설명력을 어떻게 확보할 수 있는지 설명해 주세요.
문제71	커리어 전환	이전 직장의 산업군에서 데이터분석을 기획한다면 어떤 주제로 어떤 모델을 써서 문제를 해결하면 좋을까요?
문제72	커리어 전환	데이터 분석가로 전환하려는 이유는 무엇인가요?
문제73	커리어 전환	데이터 분석가로 전환하기 위해 어떤 학습이나 준비를 했나요?
문제74	커리어 전환	사업개발 직무에서의 경험이 데이터 분석 업무와 어떻게 연관되는지 설명해 주세요.
문제75	커리어 전환	본인이 생각할 때, 이전에 재직했던 회사의 데이터를 보면 어떤 문제가 있다고 생각하나요?



No	카테고리	문제
문제76	커리어 전환	만약 당신이 데이터 분석가, 컨설턴트라면 그 문제 해결을 위해 어떤 액션 아이템을 도출할 수 있을까요?
문제77	커리어 전환	방금 말씀하신 답변에 대해서 회사가 왜 그렇게 하지 않는다고 생각하시나요?
문제78	커리어 전환	마케터로 일했던 경험이 데이터 분석가로 일하는데 어떻게 적용할 수 있을까요?
문제79	커리어 전환	웹 관련 지식이 많은데 왜 데이터 쪽으로 직무를 선택했나요?
문제80	자연어처리	(프로젝트 기반) 왜 해당 토크나이저를 사용했나요?
문제81	자연어처리	(프로젝트 기반) 왜 해당 임베딩 기법을 사용했나요?
문제82	자연어처리	(프로젝트 기반) 텍스트 데이터는 어떻게 수집하셨나요?
문제83	자연어처리	텍스트 데이터가 부족하여 학습이 어려운 경우 어떻게 해결할 수 있나요?
문제84	자연어처리	text augmentation에서 사용할 수 있는 방법을 설명해주시겠어요?
문제85	자연어처리	BERT와 현재 LLM 모델의 차이점을 설명해주세요.
문제86	자연어처리	BERT의 임베딩 방식에 대해 설명해주세요.
문제87	자연어처리	LLM에서 Hallucination(환각) 문제를 해결하기 위해 어떤 기법을 사용할 수 있는지 설명해주세요.
문제88	소프트 스킬	만약 데이터로 낸 인사이트를 상대방이 받아들이지 못한 경우에, 프로젝트 진행을 위해서 어떤 식으로 설득하고 커뮤니케이션했나요?
문제89	소프트 스킬	인사이트를 내고, 구성원들에게 전달하는 과정에서 중요한 것은 어떤 것인가요?
문제90	소프트 스킬	프로젝트나 업무를 추진할 때 발생한 어려움과 어떻게 그 어려움을 극복했는지에 대해 말씀해주세요.
문제91	소프트 스킬	데이터 분석가로 향후 커리어 및 계획은 어떻게 되시는지 궁금합니다.
문제92	소프트 스킬	어떤 조직 문화에서 가장 효과적으로 일할 수 있다고 생각하시나요?
문제93	소프트 스킬	분석 과정에서 구성원과 갈등이 생겼을 때, 어떻게 긍정적인 태도를 유지할 수 있었나요?
문제94	소프트 스킬	팀 내에서 의견 충돌이 있을 때 어떻게 대처하나요?
문제95	소프트 스킬	다양한 출신의 사람들이 모인 문화에 잘 적응할 수 있나요?



No	카테고리	문제
문제96	머신러닝	pandas와 pyspark의 차이를 설명해주세요.
문제97	머신러닝	배깅과 부스팅의 차이를 설명해주세요.
문제98	머신러닝	트리모델의 특징에 대해 설명해주세요
문제99	머신러닝	다중공선성에 대해 설명해주세요.
문제100	머신러닝	Feature Engineering을 통해 파생 변수를 만들 수 있는 방법은 어떤게 있을까요?
문제101	머신러닝	L1, L2 정규화에 대해 설명해주세요
문제102	데이터 툴	sas나 r을 다뤄본 적 있나요?
문제103	데이터 툴	현재 SQL을 어느정도 수준까지 활용할 수 있는지 궁금합니다.
문제104	데이터 툴	현재 Python을 어느정도 수준까지 활용할 수 있는지 궁금합니다.
문제105	데이터 툴	파이썬으로 데이터 전처리 했던 경험과 그 과정을 설명해주세요.
문제106	데이터 툴	본인의 R과 Python 활용 수준을 설명해 주세요.
문제107	데이터 툴	BI툴 중 태블로는 얼마나 다를 줄 아나요?
문제108	데이터 시각화	혹시 태블로는 사용해보지 않았나요?
문제109	데이터 시각화	대시보드를 구축할 때 가장 신경썼던 점이 있다면 어떤게 있을까요?
문제110	데이터 시각화	데이터 시각화의 중요성에 대해 설명해주세요.
문제111	데이터 시각화	어떤 파이썬 데이터 시각화 도구를 주로 사용하나요?
문제112	데이터 시각화	데이터 시각화를 설계할 때 고려해야 할 요소는 무엇인가요?
문제113	데이터 시각화	데이터 시각화에서 색상 선택은 왜 중요한가요?
문제114	데이터 시각화	신입의 경우 주로 시각화 업무를 맡게 되는데, 괜찮나요?
문제115	데이터 분석	챗gpt의 코딩실력을보면 현업의 데이터분석가보다 쿼리를 잘 짜는데, 이 직무에 어떤 영향을 미칠까요?



No	카테고리	문제
문제116	데이터 분석	데이터 분석 업무를 수행하면서 어떤 도구나 프로그래밍 언어를 사용해 본 경험이 있나요?
문제117	데이터 분석	코호트 분석의 개념과 어떤 식으로 진행했는지 설명해주세요.
문제118	데이터 분석	만약에 당신이 지금 회사의 거래액을 감소 원인을 분석한다면, 어떤 형태로 문제에 접근하고 분석을 할지 설명해주세요.
문제119	데이터 분석	데이터 품질이 좋지 않은 경우에, 관련한 업무를 어떻게 해결했는지가 궁금 합니다.
문제120	데이터 분석	데이터 모델링 작업을 했던 것 같은데 어떤 방식으로 진행했나요?
문제121	데이터 분석	데이터 전처리 과정에서 마주치는 문제는 무엇인가요? 이를 해결한 경험을 말해주세요.
문제122	데이터 분석	매출이 감소한 경우, 어떤 분석활동을 진행할 건가요?
문제123	데이터 분석	데이터 분석 프로세스에서 가장 강점인 영역에 대해 이야기해 주세요.
문제124	데이터 분석	평가지표인 Recall과 Precision의 Trade-off 문제를 어떻게 해결할 수 있나요?
문제125	데이터 분석	2~3천개의 Feature가 있을 경우, 어떻게 접근할지 설명해 주세요.
문제126	데이터 분석	클라우드 분석환경에서 분석업무를 수행한 경험이 있나요?
문제127	데이터 분석	전공자가 아니면 데이터 분석 관련 경험이 있나요?
문제128	데이터 분석	자기소개서(포트폴리오)에 기입한 아이디어에서 라벨링이 되어 있지 않은 신규 데이터는 어떻게 처리할건가요?
문제129	데이터 분석	pseudo labeling에서 가장 중요한 부분이 뭐라고 생각하시나요?
문제130	데이터 분석	PostgreSQL의 장점은 무엇일까요?
문제131	데이터 모델	사용한 알고리즘과 모델 선택에 대한 고려 사항은 무엇이었나요?
문제132	데이터 모델	모델의 성능을 평가하기 위해 어떤 지표를 사용했나요?
문제133	데이터 모델	모델의 성능을 더 향상시키기 위해 어떤 시도를 해보았나요?
문제134	데이터 모델	과적합(Overfitting)이 무엇인가요? 이를 방지하기 위한 방법은 무엇인가요?
문제135	데이터 모델	정확도와 재현율의 관계를 설명하고, 어떤 경우에 더 중요한지 설명해주세요.



## No 카테고리

## 문제

문제136	데이터 모델	모델의 예측 성능을 개선하기 위해 추가적으로 시도할 수 있는 방법은 무엇인가요?
문제137	데이터 모델	데이터 스케일링과 정규화의 차이점과 목적을 설명해주세요.
문제138	데이터 모델	피처 선택과 피처 엔지니어링의 차이를 설명해주세요.
문제139	데이터 모델	데이터 불균형 문제에 대처하기 위한 방법은 무엇인가요?
문제140	데이터 모델	multi-input 방식의 모델을 사용할 때 모델 관리를 어떻게 하면 좋을지 설명해주세요
문제141	직무이해도	과제전형에서 제출한 보고서를 외부에 판매한다면 얼마에 판매하고 싶고 그 이유는?
문제142	직무이해도	본인에게 데이터를 비즈니스와 연관지어 생각할 수 있는 역량이 있다고 생각하나요?
문제143	직무이해도	코딩을 하다보면 답답한 상황들이 있을텐데, 어떻게 처리하는 편인가요?
문제144	직무이해도	데이터 엔지니어로서 가져야 할 자질이 무엇이라고 생각하나요?
문제145	커리어 전환	어떻게 데이터 직군으로 취업준비를 시작하게 되었나요?
문제146	직무이해도	데이터는 계속 공부해야 하는 분야인데, 어떻게 생각하나요?
문제147	직무이해도	애널리스트가 되고 싶어하시는 것 같은데 엔지니어 직무도 괜찮나요?
문제148	도메인 이해도	지원한 도메인(AML)에 대해 알고 있는 부분을 말씀해주세요
문제149	도메인 이해도	본인의 역량으로 AML에 어떻게 활용 할 수 있나요?
문제150	기타	저희와 일하면 지금까지 했던 자연어처리와 다른 일을 할 수도 있는데 괜찮나요?
문제151	직무이해도	'실력 있는' 데이터 사이언티스트는 어떤 역량을 갖추는 게 가장 중요하다고 생각하나요?
문제152	직무이해도	(PT면접 기반) AI를 다루는 부분에 있어 '이루다' 이슈와 같은 문제를 해결하기 위한 방법을 설명해주세요.
문제153	직무이해도	데이터 사이언스 직무에서 본인만의 경쟁력은 뭐라고 생각하시나요?
문제154	직무이해도	직무와 회사 중 어느 것이 선택에 있어서 우선순위인가요?
문제155	도메인 이해도	많은 금융기업이 플랫폼화 되고 있는데, 어떤 서비스를 추가하면 좋을까요?



No	카테고리	문제
문제156	직무이해도	지원한 직무와 관련해서 읽은 논문 중 인상깊었던 내용을 설명해 주세요.
문제157	직무이해도	자사 홈페이지 안에서 개선할 부분이 있나요?
문제158	교육/학습	통계학, 데이터 시각화, 머신러닝 등 데이터 분석에 필요한 기술들을 얼마나 숙지하고 있나요?
문제159	교육/학습	그렇게 공부를 하면서 업무에 적용하는 과정에서, 비전공자로 겪는 어려움은 없었나요?
문제160	데이터 툴	사용하는 딥러닝 프레임워크와 왜 해당 프레임워크를 사용하는지 설명해 주세요.
문제161	기타	사용해본 GPU 스펙에 대해 말씀해주실 수 있나요?
문제162	도메인 이해도	CNS에 대해 아는 만큼 설명해 주세요.
문제163	도메인 이해도	CNS가 주력해야 하는 방향은 무엇이라고 생각하나요?
문제164	직무이해도	디지털 트윈을 위해 필요한 역량은 무엇이라 생각하나요?
문제165	직무이해도	디지털 트윈을 위해 본인이 기여할 수 있는 것은 무엇인가요?
문제166	커리어 전환	입사하면 데이터 입출력 실무에 투입되는데, 비전공자여도 할 수 있나요?
문제167	커리어 전환	SI와 SM 직무의 차이를 알고 있나요?
문제168	커리어 전환	SI와 SM, 두 직무 중 본인이 선호하는 분야는 무엇인가요?
문제169	소프트 스킬	데이터 관련 프로젝트 경험에서 본인이 한 기여도는 어느 정도인가요?
문제170	데이터 툴	사용할 수 있는 코딩 언어는 무엇이고, 활용도는 얼마나 되나요?
문제171	도메인 이해도	로봇을 다루는데 필요한 역량은 무엇이라고 생각하나요?
문제172	도메인 이해도	로봇과 AI의 연관점에 대해 설명해 주세요.
문제173	도메인 이해도	로봇 안에 AI를 어떤 식으로 활용할 수 있는지 아이디어가 있다면 이야기해 주세요.
문제174	도메인 이해도	협동 로봇을 만든다면 어떤 식으로 구현할 수 있을까요?
문제175	도메인 이해도	로봇과 관련되어 현장에도 갈 수 있는데, 현장에도 잘 적응할 자신 있나요?



No	카테고리	문제
문제176	데이터 분석	데이터를 다뤄본 경험이 얼마나 있나요? 본인의 숙련도는 어느 정도라고 생각하나요?
문제177	데이터 분석	데이터를 다룰 때 가장 중요하다고 생각하는 것은 무엇인가요?
문제178	도메인 이해도	왜 로봇 분야의 AI를 해보고 싶다고 생각했나요?
문제179	데이터분석	좋은 feature란 무엇인가요? 이 feature의 성능을 판단하기 위한 방법에는 어떤 것이 있나요?
문제180	프로젝트 경험	현재 직장의 업무를 개선한 경험이 있는데, 어떻게 해당 부분을 개선하게 되었나요?
문제181	프로젝트 경험	업무 기준은 기업의 여러 가지 사항들을 고려해서 세운 것인데, 바꾸면 다른 부분에 문제가 생길 수 있지 않을까요?
문제182	소프트 스킬	회사에서 건의사항을 통과시켜주지 않는다면 어떻게 할 것 같나요?
문제183	도메인 이해도	고객을 위한 서비스가 다양한데, 은행이 놓치고 있는 부분은 무엇이 있을까요?
문제184	도메인 이해도	고객들을 위한 맞춤형 서비스를 기획한다면 어떤 걸 떠올릴 수 있을까요?
문제185	도메인 이해도	메타버스가 굉장히 불이었는데, 지금은 분위기가 많이 사그라들었어요. 그 이유가 뭐라고 생각하나요?
문제186	도메인 이해도	미래의 메타버스를 활용한 금융 서비스는 무엇이 있을까요?
문제187	통계	확률 모형과 확률 변수에 대해 설명해 주세요.
문제188	데이터 분석	게임 안에는 무수히 많은 데이터가 있는데, 이를 이용해 어떤 업무를 하고 싶나요?
문제189	도메인 이해도	모든 게이머의 요구사항을 들어줄 수 없다고 적었는데, 이를 잘 결정하는 게임은 어떤 게 있을까요?
문제190	소프트 스킬	만약 일의 기한이 촉박할 때, 상사와 의견이 다르다면 어떻게 대처할 것 같나요?
문제191	도메인 이해도	게임사의 고객으로 오랜 시간 즐긴 베테랑들과 뉴비들이 있는데요. 누구를 우선시 우선하여 서비스를 운영해야 할까요?
문제192	데이터 분석	당사의 게임은 여러 플랫폼에서 게임이 운영되는데요. 플랫폼 별로 유저들의 분석 기법을 다르게 사용해야 할까요?
문제193	도메인 이해도	게임 커뮤니티 안에서 유행하는 게임의 히점을 공략하는 부분은 어떻게 처리하면 좋을까요?
문제194	데이터 분석	RDB의 char와 varchar의 차이는 무엇일까요?
문제195	소프트 스킬	본인이 만든 모델링을 통해 억울하게 제제를 당한 유저가 생긴다면 어떻게 대처해야 할까요?



## No 카테고리

## 문제

문제196	데이터 분석	동작하는 MySQL에 Alter table을 하면 안되는 이유를 설명해주세요.
문제197	도메인 이해도	당사의 주 타겟층은 어디라고 생각하나요?
문제198	직무이해도	여러 데이터 직무 중 왜 데이터 마케팅을 선택했나요?
문제199	기타	왜 많은 커머스 중 당사에 지원했나요?
문제200	도메인 이해도	본인은 어떤 화장품을 주로 소비하나요? 그리고, 최근 트렌드에 대해서도 이야기해 주세요.
문제201	도메인 이해도	최근 다른 코스메틱 유통 업계들은 지점을 줄이고 있는데, 당사는 어떻게 전략을 가져가야 할까요?
문제202	데이터 분석	옴니체널을 구축해서 온오프라인을 연결했다면, 통합된 데이터를 어떻게 활용해야 할까요?
문제203	도메인 이해도	코스메틱은 트렌드 분석이 중요한데, 어떻게 분석하면 좋을까요?
문제204	도메인 이해도	당사의 지점에서 근무한 경험이 있는데, 새로운 마케팅 방안을 제안해 줄 수 있나요?
문제205	데이터 분석	사내 모든 데이터를 활용할 수 있다면 먼저 어떤 데이터를 활용하고 싶나요?
문제206	도메인 이해도	화성에서 카드결제시스템을 만든다면 어떤 것부터 진행하는 게 좋을까요?
문제207	프로젝트 경험	(자소서기반) 0000 프로젝트를 진행하면서 어려웠다고 했는데 어떤점이 어려웠나요?
문제208	도메인 이해도	회사에 들어온다면 어떤 데이터를 다루고 싶나요?
문제209	데이터 분석	Cross Validation은 무엇이고 어떻게 해야하나요?
문제210	데이터 모델	LSA, LDA, SVD가 각각 어떤 뜻이고 어떤 관계를 가지는지 설명해 주세요.
문제211	프로젝트 경험	본인이 했던 프로젝트 중, 가장 열정적으로 완성한 프로젝트를 골라 설명해 주세요.
문제212	도메인 이해도	20년 뒤에 당사(카드사)의 경쟁사는 어디라고 생각하는지?
문제213	도메인 이해도	해당 프로젝트를 우리회사에서 진행했을 때 어떤 점에 기여할 수 있을까요?
문제214	데이터 분석	p-value 값에 대해 설명해 주세요.
문제215	통계	귀무가설에 대해 설명해 주세요.



## No 카테고리

## 문제

문제216	데이터 툴	Tableau같은 BI Tool은 어느 경우 도입하면 좋을까요?
문제217	직무이해도	그로스 프로덕트 데이터분석 직무로 지원한 이유는 무엇인가요?
문제218	데이터 분석	군집분석의 장단점에 대해 설명해 주세요.
문제219	데이터 모델	프로젝트를 진행할 때, 군집분석의 어떤 모델을 활용했나요?
문제220	데이터 모델	비정형 데이터 분석을 진행할 때, 어떤 모델을 사용했나요?
문제221	데이터 모델	워드클라우드 말고도 계속적으로 진행했던 모델이 있었나요?
문제222	자연어처리	TD-IDF를 활용해본적이 있나요?
문제223	데이터 툴	xgb부스트의 특징에 대해 설명해 주세요.
문제224	데이터 모델	랜덤포레스트의 특징에 대해 설명해 주세요.
문제225	데이터 모델	의사결정나무와 회귀모델의 특징은 무엇인가요?
문제226	데이터 모델	DBscan 모델까지 진행해본 경험이 있나요?
문제227	데이터 툴	AWS를 사용해 본 적 있나요?
문제228	데이터 분석	계층적 군집분석과 비계층적 군집분석의 차이는 무엇인가요?
문제229	프로젝트 경험	과제를 진행하면서 추가적으로 더 발견한 데이터가 있나요?
문제230	데이터 분석	고객 분류 중, 우대고객의 경우 어떤 방법으로 사로잡으면 좋을까요?
문제231	데이터 모델	머신러닝 모델의 경우 어떤 모델을 사용해 봤나요?
문제232	데이터 시각화	히스토그램의 가장 큰 문제는 무엇인가요?
문제233	직무이해도	커리어로 이루고 싶은 비전에 대해 이야기해 주세요.
문제234	소프트 스킬	협업할 때, 가장 일하기 어려웠던 팀원이 있었나요?
문제235	소프트 스킬	본인이 맡은 업무와 다른 일을 할 때에는 어떻게 할 것 같나요?



No	카테고리	문제
문제236	데이터 모델	랜덤 포레스트에 대해서 알고 있나요?
문제237	직무이해도	가지고 있는 자격증을 해당 직무와 연관되서 할 수 있는 부분이 있을까요?
문제238	데이터 툴	R의 결과물을 python으로 만든 대시보드에 넣고 싶다면 어떤 방법들이 가능할까요?
문제239	데이터 분석	동시에 여러개의 A/B테스트를 실행하기에 모수가 부족할 때, 어떻게 대처할 수 있을까요?
문제240	데이터 툴	R, 파이썬, GA을 모두 사용할 수 있나요?
문제241	데이터 분석	인터뷰나 강의에 활용하기 위한 백데이터는 어느 수준까지 일반화 해서 사용해야 할까요?
문제242	직무이해도	본인이 데이터 직무에서 갖고 있는 강점은 무엇이라고 생각하나요?
문제243	직무이해도	본인이 지원한 업무 이외의 업무를 할 수 있다면 어떤 업무를 맡아 보고 싶나요?
문제244	직무이해도	지원하신 직무가 무슨 일을 하는지 정확히 알고 있나요?
문제245	데이터 분석	인덱스가 무엇인지 설명해 주세요.
문제246	데이터 분석	인덱스를 생성할 때 고려하는 기준이 있나요? (인덱스 순서)
문제247	데이터 분석	인덱스 컬럼들은 무조건 WHERE조건절에 모두 명시되어있는 것이 좋을까요?
문제248	데이터 분석	WHERE 조건절에 A,B,C,D컬럼이 나와서 컬럼별로 단일 인덱스를 생성(4개) 해놓다고 가정할 때, 쿼리문 실행하면 어떤 일이 일어날까요?
문제249	데이터 분석	UNDO, REDO 로그의 차이를 설명해 주세요.
문제250	데이터 분석	TEMP스페이스가 다 찬(full) 경우 어떻게 조치해야 할까요?
문제251	데이터 분석	인덱스를 사용하면 무조건 조회 성능이 빨라지나요?
문제252	데이터 분석	RDBMS와 NoSQL의 차이에 대해 설명해 주세요.
문제253	데이터 분석	MVCC가 무엇인지 설명해주세요.
문제254	데이터 분석	트랜잭션의 4가지 특성에 대해 설명해주세요.
문제255	데이터 분석	SERIALIZABLE' 격리수준에 대해 설명해주세요.



## No 카테고리

## 문제

문제256	데이터분석	격리수준 4단계는 어떤 것들이 있나요?
문제257	CS	가비지컬렉터에 대해 설명해주세요.
문제258	CS	프로세스와 스레드 차이에 대해 설명해주세요.
문제259	CS	레이즈 컨디션에 대해 설명해주세요.
문제260	CS	HTTP 코드를 아는대로 설명해 주세요.
문제261	CS	HTTPS 환경에서 보안 취약점은 어떤 부분이 있나요?
문제262	CS	HTTPS가 동작하는 방식에 대해서 설명해 주세요.
문제263	CS	HTTP 프로토콜에 대해서 아는 대로 설명해주세요.
문제264	CS	Storage 저장 방식 중 Block storage, file storage, object storage에 대해 설명해 주세요.
문제265	CS	사용자가 www.amazon.com에 접속했을 때 어떤 일이 발생되는지 (DNS resolve 하는 방식) 설명해 주세요.
문제266	CS	로드밸런서에 대해 설명해주세요.
문제267	데이터 툴	맵리듀스에 대해 설명해 주세요.
문제268	데이터 툴	맵리듀스와 스파크의 차이점에 대해 설명해 주세요.
문제269	데이터 분석	HDFS에서 네임노드와 데이터노드에 대해 설명해 주세요.
문제270	데이터 분석	네임노드에서 메타데이터가 적재되는 과정을 설명해 주세요.
문제271	데이터 툴	스파크는 인메모리 기반인데 항상 메모리만 사용하나요?
문제272	데이터 툴	스파크의 트랜스포메이션과 액션에 대해 설명해 주세요.
문제273	데이터 툴	스파크로 실행 가능한 2가지 모드, 클러스터모드와 클라이언트모드 차이점을 설명해 주세요.
문제274	데이터 분석	Row형태와 Columnar형태의 차이와 예시를 들어주세요.
문제275	데이터 분석	Yarn 방식에서 Application Master, Node Manager, Resource Manager 각각 어떤 역할을 하는지 설명해 주세요.



No	카테고리	문제
문제276	데이터분석	Hive의 External테이블에 대해 설명해 주세요.
문제277	데이터 분석	리눅스 파일에 755 권한 부여하면 어떻게 되나요?
문제278	데이터 분석	데이터 샘플링에 대해 설명해 주세요.
문제279	데이터 분석	정규화와 표준화의 차이점은 무엇인가요?
문제280	데이터 모델	Overfitting과 Underfitting의 차이는 무엇인가요?
문제281	데이터 분석	A/B 테스트에 대해 설명해주세요.
문제282	데이터 분석	EDA가 중요한 이유에 대해 설명해주세요.
문제283	데이터 분석	PCA(주성분 분석)가 무엇이고, 어떤 상황에서 사용하나요?
문제284	데이터 분석	데이터 레이크(DL)와 데이터 웨어하우스(DW)의 차이는 무엇인가요?
문제285	데이터 분석	데이터 파이프라인 구축해봤던 경험 또는 프로젝트에 대해 이야기해주세요.
문제286	데이터 분석	데이터 파이프라인 구성 단계에 대해 설명해주세요.
문제287	데이터 툴	EC2와 RDS에서 DB 구축할 때 각각 어떤 차이가 있는지 설명해주세요.
문제288	데이터 분석	DB 성능 개선을 쿼리단에서 말고 dbms기능 단에서 한다면, 어떻게 처리할 것인가?
문제289	데이터 분석	인덱스는 크게 Hash 인덱스와 B+Tree 인덱스가 있습니다. 이것은 무엇일까요?
문제290	데이터 분석	Nested Loop 조인은 무엇일까요?
문제291	데이터 분석	Windows 함수는 무엇이고 어떻게 작성할까요?
문제292	데이터 분석	샘플링(Sampling)과 리샘플링(Resampling)에 대해 설명해주세요. 리샘플링은 무슨 장점이 있을까요?
문제293	커뮤니케이션	고객에게 이야기할 때, p-value를 어떻게 설명하는게 이해하기 편할까요?
문제294	프로젝트 경험	한글 감성 사전 제작 시 KcELECTRA 모델이 다른 딥러닝 모델이나 TF-IDF 모델보다 더 정확도가 높을 거라 생각한 이유가 무엇인가요?
문제295	프로젝트 경험	고객 이탈율 예측 프로젝트에서 성별과 연령을 그냥 feature로 넣지 않고 해당 기준을 가지고 데이터를 분리해서 사용한 이유가 무엇인가요?



## No 카테고리

## 문제

문제296	프로젝트 경험	데이터 분리 시 연령 기준을 잡은 기준이 무엇인가요?
문제297	데이터 모델	NLP 와 예측 모델링 중 어떤 쪽이 본인의 성향이나 흥미에 맞는 것 같은지? 그 이유를 설명해 주세요.
문제298	프로젝트 경험	RFM 분석을 결과내 등급별 고객 비중을 4집단 모두 비슷하게 두었던데 이유를 설명해 주세요.
문제299	프로젝트 경험	RFM 분석 내용을 바탕으로 각 등급별 조건을 제안하고, 그 이유를 설명해 주세요.
문제300	프로젝트 경험	Feature Selection 후 feature importance 도 별도로 확인해보았는지? 확인했다면 어떤 방법으로 확인했고, 어떤 feature 가 가장 중요도가 높게 나타났나요?

PART

2

---

인간문제

---



자연어처리

**문제 1**

▶ (프로젝트 기반) 왜 해당 토크나이저를 사용했나요?

**합격 답변**

Mecab으로 쪼갠 단어 토큰과 자음모음 단위 토큰을 함께 사용한 이유는, 해당 텍스트에 오탈자와 실제 사용하지 않는 줄임말 등 여러 노이즈들이 많았기 때문입니다. 자음모음 단위 토큰으로 Mecab 단어 토큰을 보완하여 더 높은 성능을 만들 수 있었습니다.



자연어처리

**문제2**

▶ (프로젝트 기반) 왜 해당 임베딩 기법을 사용했나요?

**합격 답변**

fasttext의 경우 word2vec과 같은 효과를 얻을 수 있으면서, 빠르고 특히 OoV(Out of Vocab)에 있어 subword를 활용할 수 있기 때문에 실제 신조어 등 고객이 사용할 수 있는 단어들에 대응할 수 있기 때문에 사용하였습니다.



자연어처리

**문제3** ▶ (프로젝트 기반) 텍스트 데이터는 어떻게 수집하셨나요?**합격 답변**

BeautifulSoup을 주로 활용하여 웹 크롤링을 진행했으며, 동적 크롤링이 필요한 경우 Selenium을 활용하였습니다.



## 자연어처리

## 문제4

- ▶ 텍스트 데이터가 부족하여 학습이 어려운 경우 어떻게 해결할 수 있나요?

## 합격 답변

text augmentation 기법을 활용할 수 있습니다. Back Translation 등 다양한 방법으로 한 데이터를 다양한 단어와 문장 구조를 갖는 데이터로 증가시켜 데이터를 늘릴 수 있기 때문입니다.



## 자연어처리

## 문제5

▶ text augmentation에서 사용할 수 있는 방법을 설명해주시겠어요?

## 합격 답변

먼저 기본적으로 SMOTE 등 머신러닝에서 주로 사용되는 기법을 사용할 수 있으며, 문장 구조를 변경하는 EDA: Easy Data Augmentation과 외국어로 번역 후 다시 번역하여 다양한 단어를 얻어낼 수 있는 Back Translation 등이 있습니다.



자연어처리

**문제6**

▶ BERT와 현재 LLM 모델의 차이점을 설명해주세요

**합격 답변**

BERT는 Encoding 모델 계열이며, 현재 주로 사용되는 LLM 모델은 Decoding 계열 모델입니다. BERT 계열은 Masked LM로써 중간 단어를 맞추는 방식으로 학습되며, LLM 계열은 시퀀셜하게 다음 문장이나 단어를 맞추는 방식으로 학습됩니다.



## 자연어처리

## 문제7

▶ BERT의 임베딩 방식에 대해 설명해주세요

## 합격 답변

BERT는 총 3가지의 인코딩 방식을 활용하여 문장단위로 임베딩을 진행합니다. 단어 벡터를 활용하는 토큰임베딩, 문장에서의 위치를 표현하는 세그먼트 임베딩, 단어 자체의 위치를 표현하는 포지션 임베딩으로 이루어져 있습니다.



## 자연어처리

### 문제8

- ▶ LLM에서 Hallucination(환각) 문제를 해결하기 위해 어떤 기법을 사용할 수 있는지 설명해주세요

### 합격 답변

Hallucination은 실제 있지 않은 이야기를 모델이 생성해낸 정보로 왜곡시키는 문제를 의미합니다. 이는 RAG라는 방법으로 해결할 수 있습니다. 자연어 정보를 DB에서 retrieval한 후 해당 텍스트로 응답을 generation하는 방식입니다. 기존 데이터를 기반으로 생성하기 때문에 환각 문제를 효과적으로 해결할 수 있습니다.



## 프로젝트 경험

## 문제9

- ▶ (프로젝트 기반) 사용한 추천모델과 왜 사용했는지 이유를 설명해주세요

## 합격 답변

LightFM 모델을 활용하여 추천모델을 구성하였습니다. 가장 큰 이유는 한정된 cpu/gpu자원을 통해 모델을 학습시켜야 했기 때문에 가벼운 모델을 선택하였습니다.

**프로젝트 경험**

**문제10** ▶ (프로젝트 기반) 추천모델을 평가하기 위해 어떤 메트릭을 사용했고, 그 이유를 설명해주세요

**합격 답변**

Recall\_k를 이용하였습니다. 실제로 유저가 관심이 있을만한 아이템 중 모델이 얼마나 잘 골라내어 추천하는가를 확인하고자 Recall을 사용하였습니다.



머신러닝

**문제 1.1** ► pandas와 pyspark의 차이를 설명해주세요**합격 답변**

가장 기본적으로 데이터의 규모 차이를 들 수 있습니다. pandas는 단일 머신에서 작은 데이터를 주로 다루는데 사용되며, pyspark는 분산 데이터 처리를 통해 대규모 데이터에 활용할 수 있습니다. 또한 pyspark의 경우 클러스터에서 동작하므로 대용량 데이터를 효과적으로 처리할 수 있습니다.



머신러닝

**문제12** ▶ 배깅과 부스팅의 차이를 설명해주세요**합격 답변**

배깅과 부스팅은 머신러닝에서 사용하는 양상을 기법입니다. 먼저 배깅은 여러 모델의 예측값을 평균하여 모델 자체의 분산을 줄이는 것이 목적입니다. 반면 부스팅은 순차적으로 모델을 구축하여 이전 모델에서 발생한 오차를 다음 모델의 학습에 감안하는 방법을 사용합니다.



머신러닝

## 문제13 ▶ 트리모델의 특징에 대해 설명해주세요

### 합격 답변

트리모델의 가장 큰 특징은 해석이 가능하다는 점입니다. 트리 모델의 분기를 시각화하여 확인하거나, Feature Importance 등을 활용하여 어떤 피처가 영향을 가장 많이 끼쳤는지도 확인이 가능합니다. 반면 트리모델은 모델을 어떻게 구성하느냐에 따라 과적합 위험이 있기 때문에 depth를 조절하는 것이 중요합니다.



머신러닝

**문제 14** ▶ 다중공선성에 대해 설명해주세요**합격 답변**

다중공선성은 독립변수들 사이에 강한 상관관계를 보이는 문제입니다. 트리 모델이 아닌 회귀분석에서 고려해야하는 요소로, VIF로 진단합니다. 다중공선성을 해결하기 위해 상관관계가 있는 변수를 제거하거나, 릿지 회귀분석 등 변형된 회귀모델을 사용합니다.



머신러닝

## 문제15 ▶ Feature Engineering을 통해 파생 변수를 만들 수 있는 방법은 어떤게 있을까요?

### 합격 답변

가장 좋은 방법은 전문적인 도메인 지식을 활용하는 것이라 생각합니다. 특히 데이터 자체가 해당 산업군에 국한되어 수집되는 데이터라면 더더욱 도메인 지식이 중요합니다. 일반적으로 파생시킬 수 있는 방법은 굉장히 여러가지가 있지만, 다항변수나 통계값을 활용하는 방법이 대표적입니다.



MLOps

## 문제16 ▶ 운영 아키텍처를 MSA로 가져갔을 때 장점에 대해 설명해주세요

### 합격 답변

가장 큰 장점은 장애 상황에서 확인할 수 있다고 생각합니다. MSA는 전체 구성 아키텍처 중 하나의 서비스에서 장애가 나도, 전체 시스템에 미치는 영향을 최소화할 수 있기 때문입니다. MLOps에 특정한 상황이라면 모델 배포 및 재사용성이 MSA 아키텍처의 장점이라고 생각합니다.

**MLOps****문제 17** ▶ (프로젝트 기반) 머신러닝 파이프라인을 어떻게 구성하셨는지 설명해주세요**합격 답변**

Sagemaker의 Pipeline을 활용하여 train, inference, monitoring 3가지의 파이프라인을 구성하고, 해당 파이프라인을 스케줄링하기 위해 Airflow를 활용하였습니다.

**MLOps****문제18** ▶ MLOps와 DevOps의 차이에 대해 설명해주세요**합격 답변**

MLOps는 전통적인 DevOps의 방법론을 머신러닝 체계에 접목한 방법론입니다. DevOps의 경우 CI/CD 자동화에 초점을 두는 반면, MLOps는 더 나아가서 CT(Continuous Training)까지 자동화하는 것을 목적으로 두고 있습니다.

**MLOps****문제19** ▶ 데이터 드리프트에 대해 설명해주세요**합격 답변**

데이터 드리프트는 학습데이터의 분포와 실제 수집되는 데이터의 분포가 달라져 발생하는 문제로, 모델 성능 저하를 불러올 수 있습니다.



MLOps

## 문제20 ▶ 데이터 드리프트를 해결하기 위한 방법에 대해 설명해주세요

### 합격 답변

가장 먼저 효과적인 모니터링 방식이 중요하다고 생각합니다. 모델의 성능과 데이터의 분포를 지속적으로 모니터링하여 해당 문제를 빠르게 발견하고 실무자들에게 alert을 줘야합니다. 추가로 재학습을 자동화하여 이런 문제가 발생했을 때 새로운 분포의 데이터로 빠르게 대처할 수 있어야합니다.

**MLOps****문제21** ▶ 모델 레지스트리에 대해 설명해주세요**합격 답변**

모델 레지스트리는 MLOps의 구성요소 중 하나로, 운영중인 머신러닝 모델의 메타 정보, 버전, 성능 등 모델 관리를 중앙화하여 관리하는 시스템입니다.



MLOps

**문제22** ▶ 모델 거버넌스를 위한 요소들이 어떤게 있을지 설명해주세요**합격 답변**

모델 거버넌스를 위한 요소로는, 모델 작성자, 사용된 모델, 버저닝 등 다양한 요소가 있습니다. 개인적으로 모델이 윤리적 책임을 가지고 만들어졌는지, 법적인 문제는 없는지 잘 확인하여 기록하는 것이 거버넌스적 측면에서 중요하다고 생각합니다.



## 직무이해도

**문제23** ▶ 지원한 직무와 관련해서 읽은 논문 중 인상깊었던 내용을 설명해 주세요**합격 답변**

제가 읽었던 가장 인상깊었던 논문은 Are These Birds Similar라는 논문입니다. 해당 논문을 읽으며 처음으로 Multimodal Neural Network라는 방법론에 대해 배울 수 있었고, CV는 CV, NLP는 NLP로 처리하는게 당연하다고 여겼던 당시 저에게 이미지와 텍스트 정보를 한번에 학습시켜 성능을 높인다는 방법론이 굉장히 인상 깊었습니다.



## 데이터 툴

**문제24** ▶ 사용하는 딥러닝 프레임워크와 왜 해당 프레임워크를 사용하는지 설명해주세요

**합격 답변**

여러 프레임워크가 있지만 주로 Pytorch를 사용하였습니다. HuggingFace등 쉽게 접할 수 있는 코드가 있으며, 현재 딥러닝 시장에서 표준으로 사용되고 있다고 생각했기 때문입니다.



기타

**문제25** ▶ 사용해본 GPU 스펙에 대해 말씀해주실 수 있나요?**합격 답변**

개인 프로젝트를 위해 주로 Colab Pro 등 쉽게 접할 수 있는 gpu 스펙을 사용하였습니다. (연구생의 경우 연구실 스펙)



## 직무이해도

**문제26** ▶ 데이터 사이언스 직무에서 본인만의 경쟁력은 뭐라고 생각하시나요?**합격 답변**

학부생이 쉽게 경험할 수 없는 자연어처리 프로젝트와 성과라고 생각합니다. 자연어처리 논문 스터디 등 전문적인 지식을 쌓은 후 해당 내용을 바탕으로 대회에 나가 수상한 경험들로 이를 입증할 수 있다고 생각합니다.



## 데이터 모델

**문제27** ▶ multi-input 방식의 모델을 사용할 때 모델 관리를 어떻게 하면 좋을지 설명해주세요

## 합격 답변

딥러닝 아키텍처를 마이크로하게 잡아 모델 별로 관리하는 것이 효율적이라고 생각합니다. 프로젝트에서 사용한 모델의 경우 3가지 input 중 하나의 모델 성능이 떨어지면 전체 성능이 떨어질 수 있기 때문에 각각 관리하여 성능을 측정하는 것이 중요하다고 생각합니다.



## 직무이해도

**문제28** ▶ 실력 있는 데이터 사이언티스트는 어떤 역량을 갖추는게 가장 중요하다고 생각하나요?

**합격 답변**

데이터 사이언티스트는 코딩, 통계, 시각화 등 다양한 능력을 필요로하지만, 저는 문제해결 능력이 가장 중요하다고 생각합니다. 해당 역량들은 결국 문제를 해결하기 위한 도구들이라고 생각하기 때문입니다.



## 데이터 분석

**문제29** ▶ pseudo labeling에서 가장 중요한 부분이 뭐라고 생각하시나요?**합격 답변**

pseudo labeling에서 중요한 것은 3가지라고 생각합니다. 그 3가지는 데이터의 기본적인 품질, 라벨링 룰의 논리성, 학습 모델입니다. 그 중 라벨링 룰의 논리성이 가장 중요한 요소라고 생각합니다. 소규모 데이터로 대규모 데이터셋을 만드는 과정이기 때문에, 룰이 어긋나면 전체 데이터셋을 망칠 수 있는 위험성이 있기 때문입니다.



## 데이터 분석

**문제30** ▶ 트랜잭션의 4가지 특성에 대해 설명해주세요.

**합격 답변**

ACID라고 부르는 4가지 특성으로는 원자성(Atomicity), 일관성(Consistency), 고립성(Isolation), 지속성(Durability)이라고 합니다.

간단하게 특성별로 설명드리자면 원자성은 트랜잭션 단위로 변경 또는 롤백이 되어야 하고,  
일관성은 트랜잭션 시작부터 종료까지 데이터베이스 상태가 일관되어야 하고,  
고립성은 트랜잭션 실행 중엔 다른 트랜잭션에 영향을 줘서도 받아서도 안되고,  
지속성은 트랜잭션이 커밋된 이후엔 변하지 않아야 한다 라고 설명할 수 있습니다.

PART

3

---

예상문제

---



## 커리어 전환

문제1 이전 직장의 산업군에서 데이터분석을 기획한다면 어떤 주제로 어떤 모델을  
써서 문제를 해결하면 좋을까요?

---

---

---

---

---

---

---

---

---

---

## 소프트 스킬

문제2 프로젝트나 업무를 추진할 때 발생한 어려움과 어떻게 그 어려움을 극복했는  
지에 대해 말씀해주세요.

---

---

---

---

---

---

---

---

---

---



## 데이터 분석

문제3 데이터 분석 프로세스에서 가장 강점인 영역에 대해 이야기해 주세요.

---

---

---

---

---

---

---

---

---

---

---

---

## 데이터 시각화

문제4 어떤 파이썬 데이터 시각화 도구를 주로 사용하나요?

---

---

---

---

---

---

---

---

---

---

---

---

**소프트 스킬**

**문제5 인사이트를 내고, 구성원들에게 전달하는 과정에서 중요한 것은 어떤 것이 있을까요?**

---

---

---

---

---

---

---

---

---

---

**커뮤니케이션**

**문제6 데이터분석에서 분석결과가 특정부서의 성과에 대해 부정적인 피드백을 내포할 때, 그 부서와의 갈등을 어떻게 풀 수 있을까요?**

---

---

---

---

---

---

---

---

---

---



## 교육/학습

문제7 통계학, 데이터 시각화, 머신러닝 등 데이터 분석에 필요한 기술들을 얼마나 숙지하고 있나요?

---

---

---

---

---

---

---

---

---

---

---

---

---

## 커리어 전환

문제8 어떻게 데이터 직군으로 취업 준비를 시작하게 되었나요?

---

---

---

---

---

---

---

---

---

---

---

---

---



## 데이터 툴

**문제9** 본인의 R과 Python 활용 수준을 설명해 주세요.

---

---

---

---

---

---

---

---

---

---

## 직무이해도

**문제10** 데이터 분석가로 향후 커리어 및 계획은 어떻게 되시는지 궁금합니다.

---

---

---

---

---

---

---

---

---

---

제로베이스에서 드리는  
**100% 취업 보장 혜택!**  
취업 실패 시 수강료는 전액 돌려드려요.

**EVENT**

제로베이스 스쿨 전과정 10만원 할인 쿠폰

**100,000원**

쿠폰코드 : 2024합격



사용 방법 : 수강신청 → 수강료 선택, 결제 페이지에서  
쿠폰 코드 **2024합격** 를 입력하세요.

\* 쿠폰 수량 소진시 조기 종료될 수 있음

**100%**  
**현업 강사진**

누구나 끝까지  
취업 성공할 수 있도록  
100% 현업 강사진 구성

**100%**  
**포트폴리오 무한첨삭**

취업할 때까지 무제한  
피드백으로 나만의  
포트폴리오 준비 가능

**100%**  
**취업 보장**

졸업 후 1년 이내  
미취업시 전액 환불