

데이터 분석가

1. 간단하게 자기소개를 해 달라.
2. 본인의 강점이 가장 잘 드러나는 프로젝트는 무엇이라고 생각하는지.
3. 당사에 들어와서 하고싶은 역할이 무엇인가.
4. 머신러닝 프로젝트에서 유기견의 관련된 프로젝트를 진행하며 입양률 예측을 했었는데, 전체 유기견 중 입양률이 원래 어느정도인지(30%정도였음), 그렇다면 입양률을 예측할 때 사용한 metric과 그걸 사용한 이유는 무엇인지?
5. 딥러닝 프로젝트에서 본인이 한 역할은 무엇이었는지 자세히 설명해달라.
6. 음성분석이 본인에게 생소한 분야였을 것 같은데(전공이 경제랑 국제통상입니다), 이런 생소한 분야를 분석할 때 어떻게 접근하는지?
7. 데이터 분석가로 지원했는데 프로젝트를 보면 엔지니어 쪽 프로젝트 이력이 보인다. 그 이유는 무엇인지?
8. 본사에 들어와서 하고 싶은 역할과 본인의 프로젝트 경험 사이의 연관성을 말해달라.

** 위와 같은 질문이 들어왔던 이유는 제가 데이터 분석가보다는 데이터 엔지니어스러운 이력서를 들고 갔기 때문입니다.

데이터 엔지니어

회사 1

1. 자기소개 간단하게 해 달라.
2. 쉘 스크립트를 써보았는지, 그리고 어느 수준까지 써 보았는가?
3. C, C++을 기반으로 훈련을 받았다고 했는데 pipe가 무엇이고 STDIN, STDOUT, STDERR의 의미에 대해서 말해달라.
4. 도커를 써 보았는지, 그랬다면 도커를 쓰는 이유는 무엇인가.
5. Python을 활용해서 데이터 분석을 해 보았다고 했는데, DB를 연동할 때 어떤 DB를 썼고, 라이브러리는 무엇을 썼는가?
6. R을 써보았는지, 써 봤다면 어느정도로 써 봤는지?
7. SQL을 어느 정도로 써 보았는가?
8. Window함수가 무엇인가? ROW_NUMBER를 어떻게 활용할 수 있는가?
9. Join이 무엇이고 어떤 종류가 있는가?
10. DB 설계 경험이 있는지?
11. Elastic Stack 구축 경험이 있다고 했는데, 왜 했고 어떤 점이 어려웠는지?
12. 데이터 엔지니어가 하고 싶은 이유는?

** 비즈니스 영어가 가능하다고 써 두어서 어느정도 가능하냐고 물어보시길래 지금까지 취득했던 공인어학성적과 해외경험을 말씀드렸습니다. 그랬더니 왜 안쓰셨냐고 하시더라고요.. 그게 의미가 있는지 몰랐습니다..

** linux, 쉘스크립트, Docker 사용 경험에 대해서 특히나 많이 물어보았습니다. 전부 42Seoul에서 배웠던 내용만 말씀드렸습니다. 추가로 Docker에 대해서 더 자세히 이해하고 싶으시다면 Docker compose를 이해하시면 좋습니다만 이부분까지는 면접에서 질문하진 않았네요.

회사2

1. SQL 어느 정도로 쓸 수 있는지, window 함수를 쓸 수 있다면 예제를 들어달라.
2. SQL 튜닝을 해 본적이 있는가.
3. 프로젝트 할 때 크롤링 어떻게 했는지, 19만건 모았다고 했는데 시간은 얼마나 걸렸고 컴퓨터 몇 대 돌렸는지, 컴퓨터 한 대당 크롤링 할 범위를 어떻게 정했는지, 멀티 프로세싱과 같은 방법을 고민했는지 말해달라.
4. Oracle이랑 MySQL을 써봤다고 했는데 셋팅은 어떻게 했는가. 어느 정도로 두 DB를 써 보았는지 경험을 말해달라.
5. 실무 경험은 없고 교육사항만 있는데 프로젝트는 모두 교육받은 내용에 기반한 것인지, 모르는게 있을 때 보통 어떻게 답을 찾는가.
6. 엔지니어로서 본인의 강점은 무엇이라고 생각하는가.
7. Docker랑 kubernetes를 써봤다고 했는데 어느정도로 쓸 수 있는가.
8. 42Seoul에서 교육받은 내용이 어떤 게 있는가.
9. 여러 언어를 써 봤는데 쓸 수 있는 언어중에서 본인이 제일 강점이 있는 언어는 무엇이고 그 이유는 무엇인지?
10. 팀워크 관련 질문 : 팀프로젝트에서 팀장을 했었다고 했는데 어려운 점은 무엇이고 어떻게 해결했는지, 팀프로젝트에서 역할을 나눌 때 어떻게 했는지 기타등등...

+본 전공이 경제인데 이 길을 선택한 이유, 기획에는 관심이 있는지, 데이터 분석가와 데이터 엔지니어중에 본인이 더 적성에 맞는 건 무엇인지 등 다양한 질문이 나왔습니다. 거의 1시간 정도 이런 저런 대화를 4명의 면접관님들과 나눴습니다.

+이 기업의 경우 지금 당장 다룰 수 있는 기술도 중요하지만, 저라는 인간이 어떤 인간인지에 대해 궁금해하신다는 인상을 많이 받았습니다. 그래서 최대한 (적절한 필터링 전제 하에)솔직하게 말씀드렸더니 좋은 인상을 드린 것 같습니다.
+질문이 더 다양했던 것 같은데 하나하나 기억나지는 않습니다~

회사3

** 이 회사는 이미지, 동영상, 위치정보 등 다양한 데이터를 가공하는 엔지니어를 뽑는 기업으로, 일반적인 데이터엔지니어와는 성격이 다를 수 있습니다.

1. 객체지향 프로그래밍에서 은닉화, 캡슐화의 개념을 어떻게 활용하는지 예를 들어달라.
 2. 객체지향 프로그래밍에서 인터페이스가 무엇이고, 언제 인터페이스를 쓰면 좋은지 예시를 들어 달라.
 3. 스레드와 프로세스의 차이가 무엇인가?
 4. 알고리즘에서 빅오(O) 표기법 중 시간복잡도와 공간복잡도가 있는데, 각 의미를 설명해달라.
 5. 자료구조에서 스택과 큐의 차이가 무엇이고, 각 자료구조가 필요한 경우를 예시를 들어달라.
 6. Python의 dictionary와 map의 차이점, list와 set의 차이점을 말하고 각각의 자료구조를 사용하는 경우의 예시를 들어달라.
 7. kubernetes를 사용해 본 적 있다고 했는데 그 경험을 설명해달라.
 8. 딥러닝 - 드롭아웃이 무엇이고 왜 필요한가? 활성화 함수는 어떤 것을 썼는가? 어떤 종류의 데이터를 다뤄봤는가? 음성분석을 했을 때 딥러닝을 통해 학습시킬 데이터를 어떻게 변환하여 활용했는가?
 9. 42Seoul에서 C, C++을 기반으로 훈련했다고 했는데, 두 언어와 python의 장단점을 비교해달라.
- 1번 질문 : 은닉화를 하는 이유는 특정 변수의 셋팅을 한정된 범위에서만 수행해서 충돌 및 권한을 제한하기 위함이라고 답했고, 캡슐화는 재사용성을 위해서라고 답했습니다.
 - 3번 질문 : philosophers 경험을 기반삼아 프로세스 안에서 스레드가 돌아가고, 한 프로세스 안의 스레드끼리는 스택의 내용을 공유할 수 있는데 프로세스끼리는 공유하지 않기 때문에 semaphore같은 별도의 방법이 필요하다는 방식으로 대답했습니다. 그랬더니 어느 때 프로세스를 사용하고, 어느 때 스레드를 사용하는지 context switching과 CPU 관련해서 설명해 주셨는데 정확히는 기억이 안납니다;
 - 5번 질문 : libasm과제를 떠올리면서 콜스택의 예시와, BFS 알고리즘의 예시를 들었습니다.
 - 6번 질문 : ft_services 과제 경험을 설명드렸습니다.

42과제를 성실히 수행하셨다면 대답할 수 있는 질문들이 나와서 신기했습니다. ft_services에서 kubernetes, CPP module에서 배운 객체지향 프로그래밍의 개념, push_swap 과제를 수행하면서 배운 빅오 표기법과 시간복잡도 개념 등 다양하게 활용할 수 있어서 좋았습니다.