

제2장 현황 분석 및 교육과정 가이드라인 원칙

- 통계학 전공에 대한 현황 분석을 위해 산업계에서 바라는 통계학의 역할과 통계학 전공 학생들의 니즈 등을 확인하기 위해 산업계 전문가 인터뷰를 실시하였으며 그 결과를 정리하였음.
- 설문조사를 통해 통계학 전공의 현황 및 현 주소를 정리하였으며, 각 대학의 학부 및 대학원 교육과정을 정리하여 타 학교에서도 교육과정 등을 참고할 수 있도록 하였음.
- 해외 대학 및 기관의 교육과정에 대해 정리하여 참고할 수 있도록 하였음.

2.1 산업계 니즈

- 데이터 분석과 관련한 기업들에서 필요로 하는 통계학 또는 데이터사이언스 전공자에 대한 인재상을 파악하여 그들의 니즈를 충족시킬 수 있는 전문인력을 배출하기 위한 통계교육의 혁신 방향성을 도출하고자 하였음
- 이를 위해 네이버클라우드의 김00 대표, LG CNS의 인사팀장 고00 상무와 이00 담당, (주)베가스 진00 부사장과의 인터뷰²⁾를 진행하였고 이를 통해 각 기업에서 필요로 하는 데이터과학 혹은 분석과 관련한 직무, 전공자들의 역량 및 자질 등을 파악하고 기존 통계학 교육에 바라는 요청 및 개선 사항을 파악하였음

2.1.1 네이버클라우드

- (김00 대표 인터뷰 내용 정리) 본인에 소속되었던 조직은 9:1의 비율로 데이터 엔지니어가 데이터과학자보다 많았음. 그럼에도 조직의 색깔과 방향, 그리고 가치를 만드는 것은 데이터사이언스이기 때문에 데이터 엔지니어 출신이 아님에도 불구하고 조직을 이끄는 역할을 할 수 있었음
- 데이터과학자와 데이터 엔지니어 역할 비교

2) 인터뷰에 앞서 인터뷰 취지와 내용을 설문지 형태로 준비하여 사전에 배포하였고, 인터뷰는 대면으로 진행함. 설문지는 부록 A를 참고할 수 있음

○ 데이터과학자

- 축적된 데이터로부터 회사의 제품이나 고객들에게 높은 가치를 만들 수 있는 일, 예로, 어떠한 지표와 직관을 뽑아낼 것이냐 등은 정하는 역할
- 데이터 엔지니어들의 방향을 결정
- 데이터를 통합하고 해석하는 역할이 많이 필요
- 회사 전체를 통합하는 역할까지 확장됨

○ 데이터 엔지니어

- 구체적 구현
- 고객에 제공되는 서비스에 대한 지속적 품질 관리 프로세스는 엔지니어의 높은 역할 및 비중

□ 데이터 사이언티스트의 역량

- 여러 부서와 협업을 할 수 있어야 함
- 전체 분석 범위(scope_)을 통합하는 역할들이기 때문에 문제 해결 능력에 대해서 많이 준비 필요

□ 통계학 커뮤니티의 다양성 및 문제 해결에 대한 가치 제고

- 경영학이나 산업공학 전공자들은 나름대로 업무가 주어지면 이를 해결하는 방법이나 엔지니어 고용 등의 그들만의 해결 프로세스에 대한 문제 해결 생태계를 갖고 있음
- 반면, 통계학 기초의 일들을 하는 사람들은 고립되는 경우가 많음. 방법론에 대한 순수과학적인 통계학 측면에서의 옳지 않음만을 판단한 후 문제 해결 노력을 회피하는 경향이 있음. 이는 통계학 전공자가 문제 해결의 과정에서 고립되는 결과를 낳게 됨
- 통계학 전공자는 통계학적 원리(principle)를 깨닫고 과학적인 깊이가 더함은 물론, 당면한 문제에 대한 통계학적 통찰력을 기반으로 문제를 해결을 위한 노력 및 능력(혹은 생태계)을 만든 것이 필요함.
- 이러한 문제 해결 능력의 생태계를 마련하는 것은 통계학계 전체의 노력 필요함
- 세상에 여러 가지 가치를 만들어 내는 방법 중에는 좋은 논문을 쓰는 것도 있지만 상업적 또는 산업적인 가치를 창출하는 것에 대해서도 통계학의 가치를 부여하는 통계학 커뮤니티의 분위기나 문화를 만들어 갈 필요가 있음

- 잘 정제된 교과서 안의 예시가 아닌 세상 속 문제들을 해석해 보고 해결해 보고 자 하는 도전적인 노력이 필요. 이를 통해 통계학이 굉장히 유용한 도구이고 이것을 통해 통계학 전공자가 제일 잘할 수 있다는 자신감 등이 생기면 좋겠음
- 예로, 문제 해결을 위한 데이터분석 상의 논리 비약 혹은 허점(jump라고 표현함)이 있는지를 생각하고 어떻게 판단할 수 있는지도 (문제 밑에 있는 구조를 상상하고 다양한 가설 설정(예, 다중 공선성) 후 검정을 통해) 파악하고, 토론을 통해 문제 해결을 할 수 있도록 훈련하여야 협업 가운데 긍정적 임무를 수행할 수 있음

□ 통계학 커뮤니티에 대한 바람

- 세상의 혁신을 만들고 이끄는 역할이 통계학 인재를 통해 가능할 것이라는 웅장한 생각
- 세상에서 누구보다 우리(통계학 전공자)가 제일 잘할 수 있다는 자신감이 있으면 좋겠음
- 통계학 전공자는 세상에 대한 관심이 끊임없이 있어야 되는 것 같음
- 교육과정이 기업 측면에서 보면 (MBA 사례분석과 같이) 구체적이었으면 좋겠음

2.1.2 LG CNS

(가) LG CNS 설문 답변

- LG CNS의 경우 설문에 대한 답변을 다음과 같이 3 그룹으로 구분하여 의견을 전달 하였음

- A그룹 : 데이터 조직내 통계학과 출신들의 의견 + LG CNS 직무역량 정의서 등
- B그룹 : 통계학 박사 출신으로 자사에서 데이터 전문위원 (LG CNS 근속 10년)
- C그룹 : 자사 재직자들의 역량 평가 항목과 요구 수준

Q1. 귀사의 데이터 분석과 관련된 직무는 어떠한 것들이 있으며, 해당 직무를 위해 수행하기에 필요한 역량은 무엇인지 기술하여 주십시오.

(직무예: 분석기획, 고급데이터분석, 데이터 수집 및 전처리, 시각화 등

역량예: 분석의 통찰력, 의사소통 능력, 통계지식, 분석기술의 숙련도, 프로그래밍 능력, 해당분야에 대한 도메인 지식 등)

□ A그룹

○ 직무관련 답변

- 비즈니스 문제를 해결하기 위한 분석 요구사항을 정의하고 문제를 해결하기 위한 데이터 분석 방법론과 최적 알고리즘 설계
- 통계적 지식에 기반하여 데이터를 탐색하고 머신러닝 알고리즘 등을 활용한 모델 구현을 통해 적합한 평가 지표를 도출하고 비즈니스 개선방안을 제시
- 고객의 업무와 문제점을 확실히 이해하고 이를 해결하기 위한 문제 식별 및 데이터 모델링을 위한 데이터 저장, 처리, 분석, 시각화 그리고 활용까지 데이터와 관련된 전 업무상에서 전문 지식과 역량을 보유
- 고객의 데이터를 분석하고 비즈니스 문제를 정의하여 창의적 대안을 제시하는 업무를 수행 및 리딩. 다양한 분야/industry에 대한 Domain Knowledge를 보유함과 동시에 데이터에 적용되는 기술을 이해하고 있어 데이터 중 가치 있는 정보를 찾아 활용기회를 포착하여 사업기획에 필요한 Data를 정의하고 내/외부의 다양한 형태의 Data를 결합하여 전방위적인 신규 사업기회를 만들 수 있는 사람

○ 역량관련 답변

- 데이터 분석 지식 : 사업 현황, 업무, 테스트 및 과제수행 등 다양한 상황에서 데이터를 통한 통계적 의미를 산출하고 시사점을 얻기 위하여, 적절한 데이터를 선정하고 목적에 부합하게 가공하고 분석하는 지식
- 분석 모델 구현/평가/배포 전문성 : 분석 모델에 적합한 변수선택, 변수추출, 변수변환, 샘플링 방법을 설계하여 적용하고 데이터 탐색 결과를 바탕으로 적합한 알고리즘을 선정하고 모델 구현(학습)방안 설계 및 분석 모델을 적합한 프로그램 언어로 구현하고 최고 성능을 도출하기 모델 튜닝 과정 수행할 수 있으며 분석 모델의 구현 결과를 평가 지표로 변환하고 결과의 비교 평가를 통해 최적 알고리즘과 모델을 선정할 수 있는 능력
- 데이터 거버넌스 수립 전문성 : 비즈메터와 테크니컬 메타 데이터에 대한 정의를 수행하고 메타데이터를 수행한 비즈니스 활용 사례를 개발 / 데이터 플랫폼 내의 데이터 품질에 대한 정의를 수행하고 시스템내의 품질에 대하여 모니터링 체계를 구축 / 데이터 Life Cycle에 대하여 정의하고 데이터 플랫폼내에서 Life Cycle에 맞추어 시스템을 구축할 수 있는 능력

□ B그룹

○ 직무관련 답변

- 분석기획, 고급데이터분석, 데이터 수집 및 전처리, 시각화와 더불어, 고급데이터분석 영역은 통계분석과 머신러닝 모델 개발 및 운영으로 나누어 볼 수 있고, 분석컨설팅도 포함됨
- 다만, 현장에서의 업무 구분은 주로 리더, 데이터파트, 모델파트로 나누어 진행하는 경우가 많고, 혹은 Senior/Junior로 구분하여 업무가 진행되므로 CNS의 분석가의 업무는 위 내용을 모두 포함한다고 볼 수 있고, 분석가 개인에게 세부작무가 부여되는 방식은 아님

○ 역량관련 답변

- 분석의 통찰력, 의사소통능력, 통계지식, 분석기술의 숙련도, 프로그래밍 능력, 도메인 지식 중에서 도메인 지식은 필수는 아니나 현장에 빠르게 적응하기 위해 사전에 갖추고 있다면 매우 바람직함.
- 그 외에 컨설팅 역량, 문제 해결능력, 현장 적응력(문제상황을 빠르게 파악하는 능력)등을 추가로 필요함.
- 주로 현실의 문제상황을 수리적인 문제상황으로 해석하는 능력이고, 주어진 데이터에 대해서 random variable 과 fixed value를 정의하는 능력, random variable의 분포를 정의하는 능력 등을 들 수 있으며, 정리하면 상황파악 능력이라고 보면 될 것 같음

□ C그룹

○ 데이터과학자(Data Scientist)

1) 분석 프로세스 설계 및 비즈니스 인사이트 도출

- Level1 : 단순한 데이터 분석결과를 해석할 수 있으며 구체적 분석방법론과 단계별 결과물을 정의할 수 있어야 함
- Level2 : 단순한 문제상황에 대해 적용 가능한 분석 방법론을 정의할 수 있으며 비즈니스 문제 상황을 이해하고, 이미 수행된 분석 결과를 활용하여 해결 방안을 제시할 수 있어야 함

2) 프로그래밍 활용능력 및 데이터 핸들링

- Level1 : Python으로 분석목적에 부합하게 데이터를 가공할 수 있어야 함
- Level2 : Python으로 분석 모델에 적합한 변수선택, 변수추출, 변수변환, 샘플

플링 방법을 구현할 수 있어야 함

3) 통계분석 및 EDA

- Level1 : Python으로 분석목적에 부합하게 데이터를 가공할 수 있어야 함
- Level2 : 데이터의 분포를 통해 현황을 파악할 수 있어야 함

4) 모델개발 및 검증

- Level1 : 단순한 분석 모델을 적합한 프로그램 언어로 구현할 수 있어야 함
- Level2 : 모델의 성능을 높이기 위한 튜닝방법을 알고 있으며 데이터 탐색 결과를 바탕으로 적합한 알고리즘을 선정할 수 있어야 함

5) 결과 배포 및 시스템 적용

- Level1 : 모델의 서빙에 대해 이해하고 있어야 함
- Level2 : 정해진 시스템 규약에 따라 모델 및 소스코드를 배포할 수 있고, 서빙되고 있는 모델을 호출한 경험이 있어야 함

○ 데이터 엔지니어(Data Engineering)

1) 빅데이터 플랫폼 구성

- Level1 : 빅데이터 플랫폼의 정의 및 기본 개념, 컨테츠 및 기능 구성요소를 이해하고 있어야 함
- Level2 : 프로그램 설계서를 이해하고, 데이터 수집/저장 및 빅데이터 플랫폼 상의 프로그램을 표준에 따라 개발할 수 있어야 함

2) 데이터 거버넌스 관리역량

- Level1 : 데이터 모델링 수행 및 데이터 품질관리를 위한 데이터 거버넌스의 필요성을 이해하고 있어야 함
- Level2 : 데이터 모델링에 필요한 항목, 용어사전, 도메인 등 데이터 표준 항목을 설계에 활용하는 데이터 표준 프로세스를 준수할 수 있어야 함

3) 데이터 유관 문제해결 역량

- Level1 : 데이터 파이프라인 설계 및 구축을 수행함에 있어서, 효율적인 처리의 필요성을 이해하고 있어야 함
- Level2 : 효율적인 Data 저장 및 처리, 리포팅을 위한 솔루션을 이해하고 있어

야 함

4) 데이터 파이프라인 구축 & 모델링 역량

- Level1 : SQL 기반으로 타겟 테이블별 적재 프로그램을 생성할 수 있어야 함
- Level2 : 적재 프로그램 파싱을 통한 전체 선후행 관계(리니지)를 파악할 수 있어야 함. 특히, 모델링 표준에 따라서, L0 (Data Collection) / L1 (Data Integration & Merging) / L2 (Data Analysis & Summary) 영역별 모델링을 수행할 수 있어야 함

5) 프로젝트 혹은 내부과제 리딩 경험

- Level1 : 프로젝트 및 내부과제에서 해당 업무 경험 미보유
- Level2 : (최근 3년 내) 1건 이상의 프로젝트 혹은 내부과제에 참여하여 설계/구축 업무를 수행한 경험이 있어야 함

Q2. 데이터 과학자가 갖추어야 할 핵심 자질은 무엇이라고 생각하시는지 기술하여 주십시오. (자질: 고급기법 적용을 위한 통계 전문성, 공감능력, 리더십, 문제해결 능력, 오픈마인드(비전공자를 위한 포용력), 창의력, 새로운 통계 기법과 도구에 대한 학습 의지와 적응력 등)

□ A그룹

○ 전문성

- Data 또는 AI 영역에서 기업고객 대상 프로젝트를 성공시킨 경험: 본인이 가진 전문성을 잘 이해하고, 본인이 수행한 프로젝트에서 어떤 식으로 활용되었는지 명확히 설명할 수 있는 사람

○ 성장

- 조직, 회사와 더불어 성공하려는 사람. 자사 직원에게 가르쳐 줄 것이 있거나 도전하려는 사람. 끊임없이 노력하여 고객과 기술을 리딩 하려는 사람

○ 책임

- 자사의 업, 즉 고객을 대상으로 Professional Service를 제공하여 성공으로 이끄는 일을 핵심으로 여기고, 자부심을 느끼는 사람, 또한 끝까지 해결해내는 사람. 프로젝트 성과 창출 과정에서 보이는 빈틈(Gray Area)을 본인이 직접 채

우려고 하는 사람

○ 존중

- 혼자가 아닌 동료와 함께 성과를 낼 수 있는 사람

○ 협업

- 본인의 성과와 조직의 성과를 동시에 중요시하고, 본인의 지식이 혼자만의 것이 아니라 조직에 내재화되도록 하려는 사람

□ B그룹

○ 핵심자질

- 앞에서 언급한 역량에서 상황파악능력과 문제해결 능력이 가장 중요하고 육성이 필요한 역량입니다. 이런 사고를 위해서는 기본적으로 높은 통계전문성이 필요하고, 다른 BP사례들에서 핵심내용을 이해/추출해서 현장에 적용할 수 있는 창의력과 적응력도 중요한 역량임
- 그 외 머신러닝, 딥러닝 모델들을 개발하는 능력, 파이썬 프로그램 능력은 기본적으로 필요한 필수 역량이며, 이 역량과 새로운 기법에 대한 학습의지가 맞물려서 점차적으로 스스로 기술을 개선할 수 있는 성향이 가장 중요한 자질이라고 생각됨.
- 여러 역량 및 배경을 가진 인력들, 서로 다른 직무의 인력들이 같이 업무를 하는 경우가 많기 때문에 상호간 이해력, 포용력도 필요한 자질입니다.

Q3. 현재 통계학과와 전공 교과목 중에서 기업 수요에 맞게 변경되어야 하거나 추가 또는 확장되어야 할 것이 있다면 기술하여 주십시오.

□ A그룹

- Python교육: 직무와 기업에 따른 차이는 있겠지만 최근 들어 데이터 분석업무의 결과가 보고서를 작성하는 것으로 끝나는 것이 아닌 반복사용을 위해 기업시스템과 연계되는 경우가 많기에 Dashboard, MLOps등을 개발할 때에는 통계분석을 위해 만들어진 SAS, R보다는 태생이 개발언어인 Python이 압도적으로 우위에 있

음

- SQL, 데이터베이스: 기업데이터는 데이터베이스에 적재 되어있고 이를 분석하기 위해서는 데이터를 끌어와서 분석가능한 형태로 가공하는 것이 가장 먼저 진행됩니다. SQL 문법 뿐만 아니라 관계형 데이터베이스의 데이터 정규화 등 대용량 데이터가 실제로 적재 되어 있는 기본적인 형태를 이해하면 도움이 될 것임.
- 결측치 처리: 현장에서 접하는 많은 데이터에는 다양한 사유로 결측이 존재할 가능성이 높기에 데이터의 성격에 맞춰 적절한 방식으로 결측 대체(Missing imputation)를 하는 스킬이 현장에서 중요함.

□ B그룹

- 파이썬 프로그래밍 역량은 기초통계역량 만큼 기본이 되는 역량이므로 학교에서도 강조되었으면 좋겠음
- SQL역량은 데이터 엔지니어의 역량으로 CNS에서는 데이터 분석가와 직무가 구분되어 필수 역량은 아니지만, 현장의 DB에서 분석용 데이터를 추출하는 과정에서 SQL이 사용되므로 기본적인 ETL역량은 있어야 함
- LG CNS가 IT기업인 만큼 프로그래밍을 강조하고 기본으로 여기고 있지만 데이터 분석가라면 수리통계학에 대한 지식은 필수적이며 절대로 축소되어서는 안 되고, 오히려 우도관려 이론(Likelihood theory)를 강조하여 현장의 데이터에 대해 분포론적 시각에서 바라볼 수 있어야 함 (고급 분석가 양성을 위한 의견임)
- 사업적으로는 현장에서는 결국 분석가가 문제 해결을 위해 개발한 머신러닝 모델이 MLOps환경에서 운영되어야 실질적인 가치가 도출되므로 MLOps에 대한 교육도 필요함
- 최근에 집중되고 있는 LLM과 관련해서 학습과 추론에 대한 내용은 GPU의 활용이 필수인 관계로 컴퓨터 공학적 지식을 많이 필요로 하고 있어서 통계학 전공자가 접근하기 어려운 면이 많습니지만, 이 부분도 빠르게 적응할 수 있도록 기초적인 내용을 학교에서 다룰 수 있으면 좋을 것으로 생각되고, LLM 활용을 위한 제반 지식(LangChain, Vector DB, Prompt Engineering)에 대한 교육도 필수가 되고 있는 상황임.

Q4. 기존 대학 교육과정의 내용이 현재의 업무에서 요구하는 방향성과 달라 아쉬웠던 점들이 있으시면 기술하여 주십시오.

□ A그룹

- 문제해결을 위한 방법론을 선정하는 경험이 부족: 개별 방법론의 이론적인 이해는 있지만 문제상황이 주어졌을 때 어떤 방법론을 통해 문제를 해결해야 하는지를 학습하는 과정이 부족한 것으로 생각되며 대부분의 학생들이 이런 역량을 공모전, 경연대회 등을 통해 얻고 있어 학교에서부터 정리된 과정으로 학습할 수 있다면 도움이 될 것 같다고 생각됨
- 분석을 위한 데이터를 정의하는 경험 부족: 많은 수업에서 이론적인 내용을 전달하기 위해 특정 상황에 맞게 준비된 데이터 전처리 등이 필요없는 깔끔한 예제 데이터를 활용하기 때문에 실제로 정제되지 않은 데이터 (여러 개의 테이블로 분할된 데이터, 센서에서 수집된 로그 데이터, 등간격으로 수집되지 않은 데이터 등)를 만났을 때 어떤 정보가 필요하고 어떤 처리를 거쳐야 하는지 경험이 부족한 것 같음
- 최근 통계학과 학부를 졸업한 인턴 및 신입 사원의 파이썬 활용 역량의 편차가 큰 케이스를 경험한 바로는 이 같은 편차가 단순히 개인의 역량 차이일 수 있을 듯 하지만 교육 과정에서 현재 현장에서 중요한 몇 가지 언어(파이썬, SQL) 등을 자주 사용할 수 있는 기회가 있다면 좋을 것 같다고 생각됨
- 이 외에도 잘 만들어진 데이터로 모델을 만들고 해석하는 것 외에 현재 주어진 상황에 맞는 시점의 데이터로 학습 데이터를 구성하고, 모델링을 하고 평가를 하는 역량은 다소 부족한 경우가 많다고 생각됨

□ B그룹

- 현장에 요구되는 방향성
- 현장에서 통계학 전공자에게 요구되는 내용은 (당연히 대부분의 업무는 머신러닝 모델을 만들고 운영하는 것이지만) 범주형 변수들 간의 상관관계, 두 요약문 사이의 유사성, 5점 척도 변수들 간의 유사도와 같은 교과서에서 다루지 않는 상황에 대해 적절한 통계량을 정의하고 산출하는 문제들을 수반함
- 이런 문제들에 대해 깊은 고민 없이 기계적으로 접근하는 경우가 많고, 이렇게 기계적인 접근으로 인해 발생하는 문제들에 대해 해석하는 역량들이 많이 부족함 (예를 들어 5점척도 변수들로 pearson-correlation을 계산하는 경우)
- 머신러닝 모델을 만들거나 활용할 때, 현실적인 문제에 대해 고민이 부족함 (예를 들어, 모델의 성능을 개선하기 위해 여러 파생변수들을 만드는 과정에서 교호작용 변수를 만들어 RandomForest모델에 넣거나, One-hot Encoding된 컬럼들

(binary변수들)의 변수 중요도(Feature importance)를 산출하거나, Train/Test 데이터 셋을 같은 parameter로 표준화해서 데이터 누수(data leakage)가 발생하는 문제들을 종종 볼 수 있었음.

- 의외로 “실험계획법”을 수강하지 않는 학생들이 있으며, 실험계획법에 사용되는 ANOVA, ANCOVA모델들을 자주 다루지는 않지만 실험 설계하는 역량은 많이 필요함
- 기본적인 RCBD에대한 개념이 잡혀 있어야 실험을 적절하게 설계하고 실험 결과에 대해 해석할 수 있는데, 이런 개념없이 진행된 편향된 설계를 자주 목격하였음
- 실험설계 역량은 모델을 만드는 상황보다 모델을 적절히 평가하기 위한 도구로 사용되고, 이런 실험설계는 보고하고자 하는 내용에 대한 근거자료가 되므로 분석 컨설턴트가 갖추기 위한 기본 기술이라고 볼 수 있음.

Q5. 데이터 분석 직무를 희망하는 통계학 전공자들이 입사 전에 갖추어야 할 실질적 역량은 무엇인지 기술하여 주십시오.

□ A그룹

- 통계지식, 프로그래밍 능력 등은 기본적인 요소이고 추가적으로 의사소통 능력이 중요한 역량임
- 데이터분석 직무로 입사하게 되어 협업하는 대부분의 사람들이 설비 엔지니어, 마케팅담당자 등 통계, 머신러닝 등에 대한 지식이 없는 사람들이 대부분이며 이런 분들과 분석결과에 대해 이야기할 때 어려움을 겪는 경우를 종종 목격하였음
- 설계된 분석 방안을 구현할 수 있는 프로그래밍 역량 및 기초통계, 머신러닝 모델 관련 기초적인 지식은 필수적이라고 생각하며 이 외에도 딥러닝 알고리즘에 대한 기본적인 이해가 필요함.

□ B그룹

○ 실질적 필요역량

- 파이썬 등 프로그래밍 역량과 기초통계학, 머신러닝 모델링 지식은 기본으로 갖추어야 하며 작은 규모의 MLOps를 구현한 경험이 있어야 함
- 실제 데이터의 모호한 상황을 보고 판단할 수 있는 눈썰미가 필요하므로 다양한 데이터의 EDA 경험이 필요함
- LLM과 같은 신기술에 빠르게 적응할 수 있는 코딩역량, 관심과 호기심이 필요함

며. AutoEncoder, RNN, LSTM과 같은 딥러닝 모델을 만들고 실행해 본 경험이 있어야 함.

Q6. 통계학 전공자들에게 일반적으로 나타나는 아쉬운 직무능력이 있다면 어떠한 것이 있는지 기술하여 주십시오.

□ A그룹

- 데이터 분석에 있어서 데이터 전처리 및 모델링이 매우 중요한 업무이지만, 모델링 결과에 대한 해석이 무엇보다 중요함
- 특히 현업의 고객을 대상으로 분석을 진행하는 경우 현업의 도메인 지식을 기반으로 결과를 해석해서 현업 실무자가 이해할 수 있을 정도로 설명할 수 있는 역량이 절실히 요구됨
- 이를 위해 기본적인 PPT 작성 및 프리젠테이션 발표 스킬도 필요한데 일부 주니어 인력은 모델링에만 집중하는 경향을 많이 목격하였음
- 프로그래밍 스킬이 부족한 경우가 간혹 있으며, 파이썬 클래스를 이해하고, 이를 기반으로 모듈을 만드는 것은 현장에서 필수적임
- 또한, 간혹 데이터에 몰입하는 경향이 있음. 좋은 모델을 만들기 위해서는 데이터에 대한 명확한 이해 기반의 데이터 처리가 필수적이며 이를 위해서는 도메인 전문가와 지속적인 소통이 중요하나 몇몇 통계학 전공자들의 경우 주어진 데이터에만 몰입하는 경향이 있는 것 같음
- 이와 같은 아쉬운 점을 해소하기 위해서는 실제 현장의 사례들을 많이 접하는 것이 필요할 것으로 생각됨.

□ B그룹

- 학교에서 배우는 기법들을 (특히 머신러닝 모델들을) 너무 기계적으로 사용하는 경향이 있음.(예를 들어, “차이를 비교”하는 상황이 나오면 무조건 T-test를 적용한다거나, classification상황이 나오면 무조건 cross-entropy를 사용하는 경우)
- 데이터의 분포를 보지 않고, 현재 수집되지 않은 데이터에 대해 생각해 보지 않고 기계적으로 train/test를 나누고 모델을 적용하는 경우를 자주 목격함
- 사실 이런 현상은 통계학 전공자에게만 나타나는 모습은 아니고, 대부분의 신입 분석가에게서 보이는 현상으로 EDA를 기계적으로 차트를 그리는 행위로 보는 경우도 많고, EDA에서 무엇을 도출해야 하는지 모르는 경우도 많음
- 대용량 데이터에서 차근차근 정보를 요약해 가는, insight를 도출해 가는 역량이

많이 부족하다고 생각됨. 이 역량은 통계학 전공자가 강점을 가질 수 있는 부분이나 실제로는 그런 경향이 보이지 않음 (통계전공자가 강점을 가지지 못하고 있음)

- 과거에는 통계학 지식 외에 프로그래밍 역량은 강점을 받을 수 있는 옵션이었지만, 현재는 필수역량이 되었으며 여전히 R을 사용하는 분석가도 가끔이지만 보임니다만 파이썬으로 현장이 바뀌는 상당히 오래 되었으므로 파이썬 프로그래밍 역량을 키우는데 많은 시간을 할애해야 함.

DS(Data Scientist) 직무 기술역량 정의(1/2)

■ 데이터 분석 영역

기술역량 항목	lv1 “상급자의 지시에 따라 데이터 핸들링, 간단한 시각화, ML모델 개발을 수행할 수 있다.”	lv2 “상급자의 지시에 따라 분석 방법을 정의할 수 있고, 성능 개선을 위한 방안을 도출할 수 있다.”	lv3 “상급자의 지시 없이 고객 요구사항을 기반으로 비즈니스 상황에 맞는 방법을 설계할 수 있다.”	lv4 “고객의 요구사항에 부합하는 BP사례를 찾아, 현장에 적용할 수 있다.”	“최신 이들을 해결할 수 있는 방법을 찾아 적용할 수 있다.”
분석 프로세스 설계 및 비즈니스 인사이트 도출	<ul style="list-style-type: none"> • 단순한 데이터 분석결과를 해석할 수 있다. • 구체적인 분석방법론과 단계별 결과물을 정의할 수 있다. 	<ul style="list-style-type: none"> • 단순한 문제상황에 대해 적용 가능한 분석 방법론을 정의할 수 있다. • 비즈니스 문제 상황을 이해하고 이미 수행된 분석 결과를 활용하여 해결 방안을 제시할 수 있다. 	<ul style="list-style-type: none"> • 데이터 및 분석 결과를 비즈니스 상황에 맞게 해석하고 실행 가능한 전략을 수립할 수 있다. • 비즈니스 목표 달성을 위한 복잡한 분석프로세스를 정의할 수 있다. 	<ul style="list-style-type: none"> • 전략적 비즈니스 의사결정을 위한 데이터 주도적 인사이트를 제공하며 주도적 예산 관리를 수립할 수 있다. • 문제해결을 위한 분석모델 설계/개발/검정/배포/모델 관리 과정을 담당할 수 있으며 복잡한 문제해결이 가능하다. 	<ul style="list-style-type: none"> • 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다. • 분석 결과의 신뢰성을 높일 수 있다. • 분석 결과의 신뢰성을 높일 수 있다.
프로그래밍 활용능력 및 데이터 핸들링	<ul style="list-style-type: none"> • Python으로 분석목적에 부합하게 데이터를 가공할 수 있다. 	<ul style="list-style-type: none"> • Python으로 분석 모델에 적합한 변수선택, 변수수축, 변수변환, 변수생성 방법을 구현할 수 있다. 	<ul style="list-style-type: none"> • Python 혹은 R을 자유롭게 사용할 수 있고, SQL문으로 하상/추론에 필요한 데이터베이스 생성할 수 있다. • 분석 모델에 적합한 변수선택, 변수수축, 변수변환, 변수생성 방법을 설계할 수 있다. 	<ul style="list-style-type: none"> • 다양한 언어를 활용한 경험이 있고 환경에 관계없이 데이터를 핸들링 할 수 있다. • NLP등 비정형 데이터를 자유롭게 핸들링 할 수 있다. 	<ul style="list-style-type: none"> • 데이터 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다. • 데이터 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다. • 데이터 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다.
통계분석 및 EDA	<ul style="list-style-type: none"> • 정제된 데이터를 이용하여 간단한 차트나 틀릿을 산출할 수 있고, 기초 통계량을 산출할 수 있다. 	<ul style="list-style-type: none"> • 데이터의 본질을 통해 현상을 파악할 수 있다. 	<ul style="list-style-type: none"> • 시장 현황, 정부, 테스트 및 과제 수행 등 다양한 상황에서 데이터를 통한 통계적 의미를 도출할 수 있다. 	<ul style="list-style-type: none"> • 고객의 요구사항에 부합하는 평가지표를 설계할 수 있다(통계량 도출). 	<ul style="list-style-type: none"> • 통계 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다. • 통계 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다. • 통계 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다.
모델개발 및 검증	<ul style="list-style-type: none"> • 단순한 분석 모델을 적용할 수 있고, 프로그램 언어로 구현할 수 있다. 	<ul style="list-style-type: none"> • 모델의 성능을 높이기 위한 튜닝방법을 알고 있다. • 데이터 탐색 결과를 바탕으로 적절한 알고리즘을 선정할 수 있다. 	<ul style="list-style-type: none"> • 다양한 ML모델들의 구동원리를 이해하고 성능을 진단할 수 있다. • 성능 개선을 위한 모델 튜닝 방법을 정의할 수 있다. 	<ul style="list-style-type: none"> • Blackbox모델의 상터를 진단하고 현 상황에 필요한 개선활동을 진행할 수 있다. 	<ul style="list-style-type: none"> • 다양한 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다. • 다양한 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다. • 다양한 분석 방법론을 최신 트렌드와 비교하여 분석 결과의 신뢰성을 높일 수 있다.
결과 배포 및 시스템 적용	<ul style="list-style-type: none"> • 모델의 서빙에 대해 이해하고 있다. 	<ul style="list-style-type: none"> • 정제된 시스템 규약에 따라 모델 및 소스코드를 배포할 수 있고, 서빙 되고 있는 모델을 조율한 경험에 있다. 	<ul style="list-style-type: none"> • MLOps시스템의 특성을 이해하고, MLOps시스템 구축 과제를 수행한 경험이 있다. 	<ul style="list-style-type: none"> • MLOps시스템 구축에 필요한 다양한 기법과 방법론을 알고 있으며, 현장상황에 맞게 조율하여 적용할 수 있다. 	<ul style="list-style-type: none"> • MLOps시스템 구축에 필요한 다양한 기법과 방법론을 알고 있으며, 현장상황에 맞게 조율하여 적용할 수 있다. • MLOps시스템 구축에 필요한 다양한 기법과 방법론을 알고 있으며, 현장상황에 맞게 조율하여 적용할 수 있다. • MLOps시스템 구축에 필요한 다양한 기법과 방법론을 알고 있으며, 현장상황에 맞게 조율하여 적용할 수 있다.

[그림 3] 데이터 분석 영역에서 직무에 따른 기술 역량 정의

DS(Data Scientist) 직무 기술역량 정의(2/2)

■ 데이터 엔지니어링 영역

기술역량 항목	lv.1 “데이터 엔지니어링의 기본 개념을 익히며 구현을 시작하는 Starter	lv.2 “표준 방법론에 따라 설계/개발을 수행하는 요소 기술 전문가	lv.3 “주도적으로 문제 해결을 하며 요구사항을 구현할 수 있는 Specialist”	lv.4 “설계 개발 표준을 수립할 수 있으며 프로젝트를 리딩할 수 있는 전문가
빅데이터 플랫폼 구성	• 빅데이터 플랫폼의 정의 및 기본 개념/컨셉을 이해하고 있다 • 주요 요소를 이해하고 있다	• 표준 프레임워크를 이해하고 있다 • 수집/저장 및 빅데이터 플랫폼 상의 플랫폼을 표현에 따라 개발할 수 있다	• 빅데이터 플랫폼상의 데이터 수집/저장 방안과 데이터 처리방안, 시각화 방안, 분석모델에 대한 전문/화성의 필요내용을 제시하고, 요구사항에 부합하는 설계를 작성할 수 있다	• 빅데이터 플랫폼의 설계/개발 표준을 제시하고 설계자/개발자를 리딩하여 구축을 수행할 수 있다.
데이터 거버넌스 관리역량	• 데이터 모델링 수행 및 데이터 플랫폼을 위한 데이터 거버넌스의 필요성을 이해하고 있다	• 데이터 모델링에 필요한 항목, 용어사전 도메인 등을 데이터 표준 항목을 설계에 활용하는 데이터 표준 프로세스를 준수할 수 있다.	• Data 활용 및 품질을 위한 Meta Data 관리의 중요성과 주요 구성 요소들을 이해하고 있으며 관련 프레임워크를 구현할 수 있다.	• 수립한 Meta Data 관리 체계에 따라 Meta Data 표준화, 수집 및 설계 관리 효율성을 적용하고 관련 프로세스를 설계할 수 있다.
데이터 유관 문제해결 역량	• 데이터 인프라인 설계 및 구축을 수행함에 있어서 효율적인 처리의 필요성을 이해하고 있다.	• 효율적인 Data 저장 및 처리/리포팅을 위한 솔루션을 이해하고 있다.	• 고객의 PainPoint를 해결하는 DX과제에 대한 정의할 수 있다.	• 고객의 PainPoint를 해결하는 DX과제에 대하여 이를 지칭하고 이를 분석할 수 있는 데이터 수집/저장 방안, 데이터 처리방안, 분석모델에 대한 주력화정에 대하여 정의할 수 있다.
데이터 인프라 구축 & 모델링 역량	• SQL 기반으로 타겟 테이블 생성할 수 있다.	• 적정 프레임 파싱을 통한 적정 선출할 수 있다. • 모델링 표준에 따라서, LO (Data Collection) / LI (Data Integration & Merging) / L2 (Data Analysis & Summary) 모델링을 수행할 수 있다	• 요구사항 기반의 버스 아키텍처를 구성하여 디메진/쿼리/관계를 정의하고 해당 기준으로 주제영역별 설계 리딩할 수 있다.	• 개념/논리/물리 데이터 모델링 방향성을 수립할 수 있다. • 디메진 / 쿼리/관계를 모델링 표준 가이드 수립 후, 설계자를 리딩할 수 있다.
프로젝트 혹은 내부과제 리딩 경험	• 프로젝트 및 내부과제에서 해당 업무 경험 미보유	• (최근 3년 내) 1건 이상의 프로젝트 혹은 내부과제에 참여하여 설계/구축 업무를 수행하였다.	• (최근 3년 내) 1건 이상의 프로젝트 혹은 내부과제에 참여하여 설계/구축 업무를 수행하며 주도적으로 문제를 해결하였다.	• (최근 3년 내) 1건 이상의 프로젝트 혹은 내부과제의 설계/구축을 주도적으로 리딩한 사례를 보유하고 있다.

[그림 4] 데이터엔지니어 영역에서 직무에 따른 기술 역량 정의

관련 교육과정

■ AI/데이터 영역 교육

방식 Offline/Online

구분		기본 과정	심화 과정	고급 과정
AI	기본	[ABC-On] AI기초	데이터 실무 기법/심화	TDC 핵심
		AI기초	인공지능 실무 기법/심화	AI 전문인력 4 선발
		One-day로 배우는 딥러닝	사각지능 실무 기법/심화	데이터 핵심 1
		경이로운 AI Coding Overview	AI/OD AI Coding Practice	- 응용 과정 적용 (응용 및 중급자)
		ChatGPT 활용 AI 애플리케이션 개발	Junior AI/OD AI Coding	
	정상시	GenAI 사업개발	AI/OD AI Coding 실무	
			AI/OD 플랫폼으로 데이터어싱	
			LangChain 활용	
		[ABC-On] 데이터	분석가가 고작각 일하는 방법	
		데이터 분석 기본	AI기초 이상형지 데이터 분석	MLOps
데이터		1+1 역량강화를 위한 데이터 분석 입문	ML 데이터어싱	
		1+1 역량강화를 위한 데이터 분석 실무	데이터 분석 심화	
		One-day로 배우는 머신러닝	데이터 데이터어싱 심화	
		데이터 데이터어싱 기본	DAP MLOD 파악하기	
		[교재보기] 데이터 데이터어싱		

■ DS직무 교육

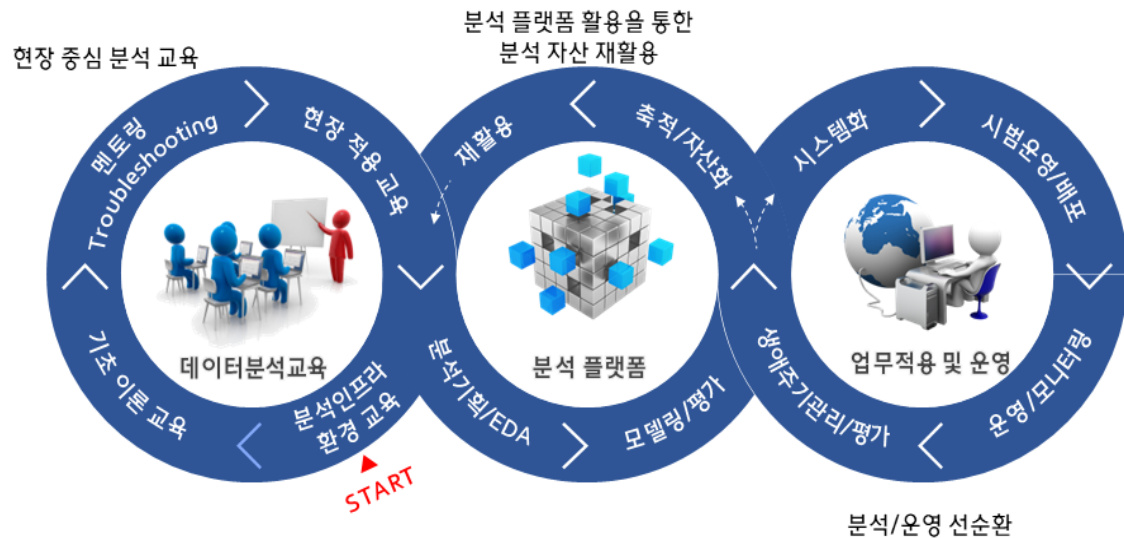
Data Science	데이터 분석	데이터 분석 기본	AI기초 이상형지 데이터 분석	MLOps
		One-day로 배우는 머신러닝	ML 데이터어싱	
	분석가가 고작각 일하는 방법		데이터 분석 심화	
	데이터 데이터어싱	데이터 데이터어싱 기본	데이터 데이터어싱 심화	

[그림 5] LG CNS의 AI/데이터 영역 교육 과정 및 DS 직무 교육 과정 예시

2.1.3 베가스

Q1. 귀사의 데이터 분석과 관련된 직무(예로, 분석기획, 고급데이터분석, 데이터 수집 및 전처리 등)는 어떠한 것들이 있으며, 해당 직무를 위해 수행하기 위해 필요한 역량은 무엇인지 간략하게 기술하여 주십시오. (예로, 분석의 통찰력, 의사소통 능력, 통계 지식, 분석기술의 숙련도, 프로그래밍 능력, 해당 분야에 대한 도메인 지식 등)

- 데이터분석 컨설팅, 교육(멘토링) 서비스를 제공하고 있는 베가스는 다음과 같은 분석관련 직무를 갖고 있음
- 데이터 과학자 직무는 고객사의 다양한 비즈니스적 문제를 발굴하여 이에 대한 과제기획을 실행하고 최적의 데이터 분석을 통한 현실적인 해결책을 제시하는 것으로 세부 직무 단위로는
 - 1) 과제기획, Pilot(PoC), 데이터전처리 및 데이터마트구성, EDA, 고급 분석, 시각화, 분석 결과 리포팅 등의 업무가 포함됨
 - 2) 실제 프로젝트 방식은 문제해결형 프로젝트계약과 분석 CoE형태의 계약으로 나뉘질 수 있음
- 데이터 분석 교육 직무는 고객사 분석 조직의 정착과 확장을 위해서 관련된 역량 교육 및 멘토링을 통해 PBL 교육을 수행하며 세부 직무 단위로는 분석기획 강의, 분석 기법 강의, 미니프로젝트형 실습, 멘토링형태의 PBL 교육이 있음
- 데이터 분석관련 직무는 인공지능 기술, 인프라 발전으로 급변하고 있음 실제로 많은 업무들이 인공지능의 도움으로 많은 변화들이 생기고 있음
- 이렇게 볼때 현재 생성형 인공지능이 만드는 기업활동의 변화를 주목해 볼 필요가 있음
(앞으로 인공지능이 단순히, 검색(Search)을 대행해줄 것인가? 답(Answer)을 해 줄 것인가? 실행(Action)을 해줄 것인가?에 대한 고민이 필요한 시점이 되었음
향후 인공지능은 실행까지 되는 강인공지능의 시대에 살게 될 것이고 이때 데이터 분석가들이 가져야할 역량에 대해서는 기존과 다른 관점이 필요할 것임)



[그림 4] 기업 현장에서 데이터 분석관련 업무 선순환 프로세스 구조

- [그림 4]는 기업에서 데이터 분석을 활용하는 아주 기본적인 전체 프로세스에 대한 흐름임.
- 최근 데이터 관련 직무 수행자에게 새롭게 부각되고 있는 현장의 요구역량은 다음과 같음
 - 분석 IT 인프라에 대한 이해 : 데이터 분석관련 새로운 Tool들과 인공지능 활용사례가 계속 나오고 있고 오픈소스 기반의 Library는 하루가 다르게 업데이트 되고 있음. 따라서 이 분석 인프라에 대한 지식이 어느 정도 갖추고 있어야 함
 - 이론교육 및 멘토링 역량 : 분석을 담당하는 데이터과학자는 항상 공급이 부족한 상황으로 현장의 분석 수요를 감당하기 위해서는 교육 시스템만으로는 커버가 안되고 있음. 일종의 AI 전도사로서 통계전공자가 내부기술전파교육을 담당할 수 있어야 함.
 - 분석기획/EDA : 현장의 문제를 사업화하기 위해서는 커뮤니케이션 역량, 문제 해결기법과 같은 역량이 요구됨. 대기업에서도 DA, DE 이외에 BA와 같은 직군이 생기고 있음. 이때 핵심은 Data로 문제에 대해 Fact Based Approach를 하는 것임. 데이터를 통합적으로 상상해 보는 능력, 데이터에서 숨겨진 문제를 상상해 보는 능력을 배가시켜야 함
 - 모델링/평가 : 수많은 모델링 방법들이 쏟아져 나오고 있는 상황에서 분석가는 어떤 모델을 써야 할지 헷갈릴 수 밖에 없음. AutoML처럼 많은 모델을 시도하는 것도 답중 하나가 될수 있으나 결국 모델을 선택하는 능력이 매우 중요함.

모델을 선택하기 위해서는 다양한 유사 사례가 검색되어지고 이를 참고하여 일을 수행할 수 있는 환경이 있어야함.(분석자산화와 연계하여 고려해야 함) 한가지 첨언하면 EDA에서 얻은 상상력을 기반으로 파생변수를 만들어 내는 것도 매우 중요함

- 시스템화 : 기업의 분석업무 환경은 다양한 방식으로 개선되고 있음. 이때 시각화 기법역량은 점점 중요성이 높아지고 있음. 다양한 시각화툴을 사용하는 것 외에도 시각화 스토리텔링에 대한 역량도 매우 중요해지고 있음

- 운영/모니터링 : 기업에서 분석모델을 만들고 그 모델의 단기간 운영후 폐기되는 경우가 많음. 데이터 분석이 기하급수적으로 늘어나는 환경에서 Data Drift, Model Drift 관련 역량에 주목해볼 필요가 있음

- 축적/자산화 : 오픈소스 기반의 많은 분석 Library가 있음에도 자산화는 아직 기업이 완성하지 못한 초기 단계로 파악됨. 하나의 분석결과가 자산화되려면 보고서, 분석방법론 문서, 데이터, 소스코드가 모두 연계되어야 함. 또한 이러한 분석자산은 어느정도는 정형화된 포맷을 가질 때 검색과 빠른 이해가 가능해짐. 이렇게 하기 위한 자산화 축적 역량이 기업의 핵심역량으로 부각되고 있음 이는 데이터분석가의 분석전도사 역할전환과도 맞물려 있다고 볼 수 있음

Q2. 귀하께서 생각하시는 데이터과학자가 갖추어야 할 핵심 자질은 무엇인지 간략하게 기술하여 주십시오. (예로, 고급기법 적용을 위한 통계 전문성, 공감능력, 리더십, 문제해결 능력, 오픈 마인드(비전공자를 위한 포용력), 창의력 등)

- 데이터과학자가 가져야할 핵심 역량은 데이터 분석이론, 데이터 분석 IT 지식, 산업 지식이 핵심적으로 이야기되고 있으나 현실적으로 우수한 데이터과학자를 양성하는데 있어서는 상상력을 발휘하는 문제해결 능력을 가장 현실적인 핵심 역량이라 판단됨
- 실제 세계에서의 문제해결은 데이터로 현장의 문제를 다양한 모델을 사용하여 풀어내는 것임. 이때 모델은 매우 중요하고, 이론적으로도 많은 이해가 필요한 난이도 있는 부분으로 전문적인 교육이 필요하지만, 현실 문제에서는 모델의 문제보다는 데이터의 문제가 실질적인 해결에 도움이 됨. 데이터는 현실을 어떤 목적성에 의해 대표하는 것으로 그 목적성을 정확히 이해할 때 그 현실의 문제를 정확히 상상해 낼 수 있음.
- 즉, 데이터 과학자는 상상력이 풍부한 과학자이어야 하고 이는 공학의 목표와 다른 탐구정신이 필요함

Q3. 귀하께서 대학시절 학습하셨던 통계학과 관련된 전공 교과목 중 현재 상황에 맞게 변경되어야 할 것이나 또는 추가·확장되어야 할 교과목이 있다면 간략히 기술하여 주십시오. (예, 수리통계학 축소, 프로그래밍 과목 확대, SQL 등 필요성 등)

- 실제 업무에서 활용되는 기법 중심으로 변화가 필요함
- 분석역량 자산화 : 분석방법론 교육 및 이에 따른 문서작성 기법 교육이 전 교과과정 중 실습미니프로젝트에 반영될 필요가 있음 효율적 커뮤니케이션도구로써, 잘 정의된 일 방법으로써 분석 방법론 템플릿은 매우 중요함
- EDA 과정의 확대 : 단순 데이터 시각화에서 벗어나 데이터 스토리텔링이 가능한 커리큘럼으로 확장이 되면 매우 현장 중심적인 교과과정중 하나가 될 것임. EDA를 통해 파생 변수 생성 능력, 데이터 상상력 향상에 대한 교육이 만들어짐
- 추가적으로 현업에서 가볍게 사용할 수 있는 엑셀 기능 피벗, 차트, 통계모델, 생성형 AI 연계 등 VBA를 통한 코딩

Q4. 기존의 대학 교육과정에서 수강하셨던 교육과정과 내용이 현재의 업무에서 요구하는 방향성과 달라 아쉬웠던 점들이 있으시면 간략히 적어주세요. (예, 다양한 도메인 지식 획득 기회 부족, 실제 업무에선 이론 적용의 중요도가 낮음, 프로그래밍 언어 교육의 다양성 부재(실제 업무에서 사용하는 프로그램과 상이) 등)

- 현재 커리큘럼을 보면 이론에 치중된 커리큘럼이 아쉬움
- 실무형 인재 양성을 위해서는 맥킨지와 같은 컨설팅펌에서 교육용으로 많이 사용하는 문제해결방법론, MECE, 피라미드 원칙, 논리의 기술(바바라민토) 이런 류의 교육이 개설된다면 산업계입장에서는 매우 바람직하다고 생각됨
- 현재 모든 학교의 세부 커리큘럼의 내용은 파악할 수 없으나 많은 과정의 제목들이 실습을 포함하고 있는 것은 바람직하다고 봄. 다만, 이러한 실습을 할 때 교과과정의 교육이수 후 이에 대한 과제를 실습하는 방식이 아니라 실습을 통해 교과과정의 내용을 학습하게 되는 과정의 도입도 필요해보임.
- 일례로, MBA 과정들의 경우 CASE Study를 통해 가상의 문제해결을 위한 약식프로젝트를 많이 다루는데 이렇게 다양한 산업의 사례를 중심으로 실무적인 문제해결 프로젝트를 통해 Learning by Doing 이 가능토록 유도하는 것이 중요하다고 생각됨

Q5. 데이터 과학자의 직무를 희망하는 통계학 전공자들이 입사 전 갖추어야 할 실질적인 기본 역량이 있다면 무엇이라고 생각하십니까? (예, 분석의 통찰력, 의사소통 능력, 통계 지식, 분석기술의 숙련도, 프로그래밍 능력, 해당 분야에 대한 도메인 지식 등)

- 실제 현장의 데이터과학자가 필요한 역량으로 커뮤니케이션 역량, 의사소통 능력에 대해서는 사전에 기본역량으로 확보가 필요.
- 대부분의 전공자들이 이러한 역량을 기업에서 확보하는데 많은 시간이 소요되고 있음. 실제로 많은 Business Case를 통해 현장에서 사용되는 의사소통 능력을 확보한다면 기업에서 통계전공자에 대한 인식의 전환이 있을 것임
- 분석의 통찰력이나 분석기술의 숙련도 측면에서는 데이터를 통한 상상력을 충분히 발휘될 수 있도록 EDA역량을 확보하는 것이 중요할 것으로 보임. 왜냐하면 분석기술에 대한 생성형 AI가 프로그래밍 자체에 대한 지원은 나날이 좋아지고 있어서, 실제로 프로그래밍의 오류수정, 코드작성, 심지어 결과 문서 작성까지 생성형 AI가 대신 해줄 수 있는 시대가 오고 있음. 하지만 이러한 데이터에 대한 상상력은 대신해 줄 수가 없음.
- 또한 분석기술의 숙련도에 대한 정의도 좀 바뀌어야 하지 않을까 싶음. 여태까지는 분석기술 숙련도가 높은 사람은 분석코드를 효율적으로 짜고 좋은 모델링 결과를 내는 것이었으나. 이제 분석 기술의 숙련도는 다른 여러 사람들이 만들어낸 모델링 결과에 대해서 잘 Audit 해줄 수 있는 능력이 아닐까 싶음