데이터 사이언스에 통계가 필요한 이유는 무엇일까요. 통계학이 왜 중요하게 여겨질까요.(300자)

데이터 분석의 목표는 관찰된 데이터를 넘어서 그 근간에 있는 현상에 대한 더 나은 이해를 하는 것에 초점을 맞추고 있는데, 모집된 데이터를 통해 결론을 유도하려는 것이기 때문에 본질적으로 불확실성을 수반하게 된다. 그러므로 데이터 사이언티스트들은 통계학의 도구와 방법(예를 들면 회귀분석, 분류 등)을 사용하여 적절한 실험을 하고 그 결과를 해석하여 데이터 분석이 가지고 있는 불확실성을 계량화하는데 사용한다. 또한 통계학이 데이터 구조의 기반이 되기 때문에 많은 양의 데이터가 쌓일 경우 수치들을 효율적으로 다루고 인사이트를 뽑아내기 위해서는 통계학이 꼭 필요하다 생각합니다.

여러분 생활에서 찾을 수 있는 중요한 요소 중 통계적 해석/활동이 필요한 것은 무엇인가요. 왜 필요하다고 생각하나요. 올해 실천할 수 있다면 무엇을 할 수 있을까요?

생활에서 찾을 수 있는 중요한 요소 중 선택이 관련된 모든 활동이 통계적 해석이 필요하다고 생각합니다. 예를들면 어떠한 물건을 살 때 여러 브랜드의 여러 종류의 물건의 성능을 비교분석하고, 특정한 제품을 사용하는 소비자의 만족도를 통계적으로 해석 함으로써 어떠한 물건을 골라야 합리적인 소비인지 알 수 있을 것입니다. 이렇게 합리적인 소비를 함으로써 불필요한 지출을 줄임으로서 삶의 질을 높일 수 있다고 생각합니다. 또한 규칙적인 생활, 예를 들면 식습관생활, 수면습관 등 규칙적인 생활이 필요한 부분들도 통계적 해석을 통해 많이 개선될 수 있다고 생각합니다.

데이터를 이용한 작업에서 오류가 발생하는 가장 큰 원인은 무엇이라고 생각하시나요. 예시를 들어 설명하시오

데이터를 이용한 작업에서 오류가 발생하는 가장 큰 원인은 데이터의 불완전성 때문이라고 생각합니다. 예를 들면 여러 소스 시스템에서 파생되어 중복되는 데이터, 여러 형식으로 저장되어 일관되지 않은 데이터, 검증되지 않은 데이터의 입력에 따른 불완전한 데이터, 정제되지 않아 오류가 많은 데이터, 최신 업데이트 여부가 불분명한 신뢰할 수 없는 데이터 등 오류가 많고 신뢰할 수 없는 데이터가 분석에 사용되었을 시 분석에 큰 오류가 난다고 생각합니다.

이 책을 읽기 전과 후가 가장 다른 점은 무엇인가요? 나에게 적용할 수 있는 것은 어떤 것이 있을까요?(200자)

이 책을 읽기 전에는 막연하게 크게 데이터 사이언티스트의 꿈을 꾸고 있었다면, 이 책을 읽고 나서는 도메인 전문가가 될 것인가 데이터 전문가가 될 것인가부터 데이터 비즈니스맨 / 데이터 창조자 / 데이터 연구자 / 데이터 개발자 중 어떠한 역량을 가진 데이터 분석가가 될 것인지 까지 좋은 데이터 분석가가 되기 위하여 어떠한 것을 목표로 하야하는 지 알게 되었습니다.

통계적 사고는 무엇이라고 생각하시나요. 예시를 들어 기술해주세요.

통계적 사고란 어떠한 문제를 해결하고자 할 때 그 해답의 출발점을 데이터의 수집 및 해석으로부터 찾는 사고방식이라고 생각합니다. 예를 들면, 평소 피곤함을 많이 느껴 체력 개선을 하기 위해서 전날 수면의 질, 수면시간, 운동여부, 스트레스 여부 등 여러가지 요인들을 기록하고 분석하여 피로도에 어떠한 요인이 큰 영향을 주는지 알아내는 것처럼 문제를 해결하고자 할 때 데이터를 통하여 원인과 결과를 분석하는 것을 통계적 사고라고 생각합니다.