**\* 데이터 분석?**

- 유용한 정보를 발굴하고 결론 내용을 알리며, 의사결정을 지원하는 것을 목표.

- 이를 목표로 데이터를 정리, 변환, 모델링 하는 과정.

- 각기 다른 여러 분야에 사용됨.

- 비즈니스 부문에선 의사결정을 더 과학적으로 만들어주고 비즈니스를 더 효율적으로 운영할 수 있도록 도와주는 역할.

**\* 분석의 필요성?**

- 모델이 성숙도가 증가함에 따라 처리해야할 업무의 난이도와 복잡도가 증가하기 시작.

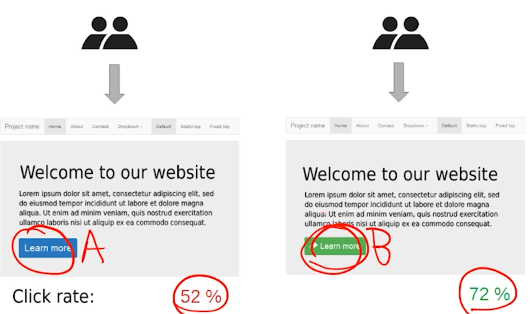
e.g.) 1개의 물품을 1곳에서 판매 🡪 1,000,000개의 물품을 10,000곳에서 판매

- 인간의 감에 의한 의사결정이 아닌, 데이터 분석에 기반한 합리적이고 효율적인 의사결정의 중요성이 대두.

- 대부분의 비즈니스가 IT기반이 되면서 분석을 위한 대량의 데이터를 손쉽게 구할 수 있는 환경이 조성됨.

Ex. A/B 테스트

- 마케팅, 웹 분석에서 전체 또는 일부 요소를 변형시킨 A안과 B안의 대조 실험을 통해 각각의 성능을 측정하는 방법론.



> 50%의 사용자에겐 파란버튼(A안)을, 50%의 사용자에겐 초록버튼(B안)을 노출.

> 각각의 클릭율을 분석

> 분석 결과 초록버튼의 클릭율이 높게 나왔다. 🡪 버튼UI는 초록색으로 하는 것이 좋을 것이다 라는 결론.

**\* Data Science 관련 직군 & Need Skills**

- 프로젝트 진행을 위해선 다음과 같은 팀 구성원들이 필요.

- **Data Scientist** : 데이터에 대한 가설 설정, 데이터 분석, 분석한 데이터로 파악한 특징을 시각화해서 공유.

> 비즈니스에 대한 이해력 (도메인 지식) : 어떤 데이터를 분석할지? 분석한 데이터에 기반한 정보가 비즈니스에 어떤 도움을 주는지? 등에 대한 고민 + 이를 반영하기 위해 회사에서 진행하는 비즈니스에 대한 이해력.

> 데이터 분석 능력 : 데이터의 중요 특징 파악, 가설 및 가설검정을 통해 유의미한 결론 도출.

> 데이터 시각화 능력 : 분석 결과를 그래프, 대시보드 등을 이용해 명확히 시각화. 이를 다른 팀원들과 공유.

🡺 데이터 분석을 통해 유의미한 결론 도출, 시각화를 통해 다른 구성원과 결과 공유.

- **Data Engineer** : 프로젝트 진행을 위한 데이터 수집/전처리, 학습된 모델의 안정적 배포를 위한 인프라 개발.

> 데이터 크롤링 및 데이터 정제를 위한 프로그래밍 스킬

> 모델배포 및 API 구축을 위한 인프라 구성을 위한 엔지니어링 능력

> DB및 SQL 활용능력

🡺 데이터 분석을 위한 환경 제공, 모델 배포에 도움.

- **ML Engineer** : 프로젝트 진행을 위한 새로운 머신러닝 모델 개발

> 최신 논문을 읽을 수 있는 능력 (새로운 모델은 보통 최신 논문을 통해 공개되기 때문.)

> ML 프레임워크(TensorFlow, PyTorch 등…)를 이용한 모델구현을 위한 프로그래밍 스킬

> 모델 성능측정 및 개선 능력

🡺 Research직군 느낌. 최신 모델을 학습해 현재 모델의 개선안을 찾아 감.

**\* 데이터 분석에 코딩이 필요한 이유?**

- 분석 업무를 수행하기 위해선 데이터의 정제/분석/시각화 과정이 필요.

- 각각의 기능을 지원하는 툴 or 서비스를 사용할 수 있음.

- But, 나만의 환경에 맞는, 내가 다루는 데이터에 꼭 맞는 분석을 하기 위해선 코딩을 통한 분석이 필요.

- 각 회사마다 데이터가 구성된 형태는 모두 다름. 코딩을 통한 전처리와 분석이 필수.