

1강 : 데이터 리터러시

강사: 김진용



강사 소개

본인 소개

- ❑ 이름 : 김진용
- ❑ 직무 : 3D Vision Engineer
- ❑ 관심 분야 : 3D AI, SLAM.. 등등

회사에서 하는 업무

- ❑ 음식의 부피를 측정하는 알고리즘 개발
- ❑ 음식 도메인에서 겪는 여러 엣지케이스 분석하고 처리할 수 있는 방법 연구

경력

- ❑ Food Tech 스타트업 2년차 개발자
- ❑ 네이버 부스트 캠프 AI Tech 5기 조교
- ❑ 위즈라이브 Python Tutor 약 1년

TMI

- ❑ MBTI : ESFJ
- ❑ 취미 : 클라이밍, 등산, 맛집 가기

커리큘럼 소개

WEEK 1-1

1. 데이터 리터러시
2. 데이터 시각화 기초 이론
3. 사전 과제
4. 아하!모먼트 : 데이터 과학자 이해하기

WEEK 1-2

1. matplotlib을 활용한 차트 그리기
2. 차트의 디테일한 요소 표현하기
3. 아하!모먼트 : 현업에서 겪었던 데이터 분석 경험

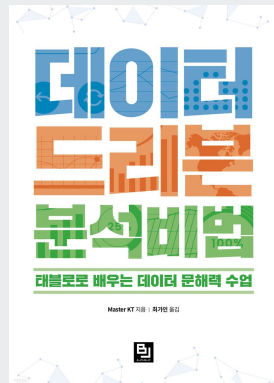
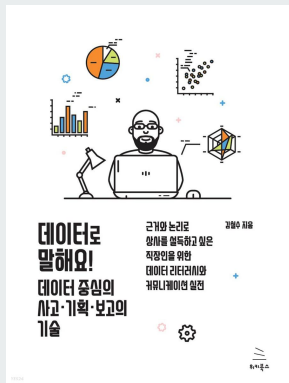
WEEK 2-1

1. seaborn을 활용한 차트 그리기
2. plotly를 활용한 **interactive**한 차트 그리기
3. 아하!모먼트 : **3D** 데이터 시각화해보기

WEEK 2-2

1. 실전 프로젝트 : 서울시 범죄 현황 분석
2. 아하!모먼트 : 프리온보딩 챌린지 과정 최대한 써먹기

참고한 책 소개



목차



1. 데이터 리터러시

- a. 데이터 리터러시란?
- b. 데이터 리터러시의 중요성
- c. 데이터 리터러시의 핵심 능력

2. 데이터 중심의 비즈니스

- a. 데이터와 비즈니스는 어떤 관계가 있을까?
- b. 일 잘하는 데이터 과학자의 사고 방식

3. 상사에게 보고하는 방법

- a. 효과적인 보고
- b. 보고 시 주의할 점



1. 데이터 리터러시

1. 데이터 리터러시



1.1 데이터 리터러시란?

- ❑ 데이터 리터러시는 데이터와 관련된 개념, 기술 및 도구를 이해하고 사용할 수 있는 능력입니다.
- ❑ 개인이나 조직이 데이터를 수집, 분석, 해석하고 의사결정을 내리는 과정에서 필요한 기술과 지식을 포함합니다.
- ❑ 현대 사회에서 데이터 리터러시는 경쟁력과 혁신을 이끌 수 있는 역량으로 간주됩니다.

1. 데이터 리터러시



1.2 데이터 리터러시의 중요성

- ❑ 데이터의 양이 증가하고 데이터 기반의 의사결정이 중요해지면서 데이터 리터러시는 필수적인 역량이 되었습니다.
- ❑ 데이터 리터러시를 갖춘 개인과 조직은 데이터를 효과적으로 활용하여 경쟁 우위를 얻을 수 있습니다.
- ❑ 데이터를 올바르게 이해하고 활용하는 능력은 조직의 성과 향상과 혁신을 이끌어냅니다.

1. 데이터 리터러시



1.3 데이터 리터러시의 핵심 능력

- ❑ **데이터 수집:** 다양한 소스에서 데이터를 수집하고 필요한 데이터를 식별합니다.
- ❑ **데이터 정제 및 준비:** 데이터를 정제하고 전처리하여 데이터의 품질을 향상시킵니다.
- ❑ **데이터 분석:** 통계적 개념과 분석 기법을 사용하여 데이터를 분석하고 해석합니다.
- ❑ **데이터 시각화:** 데이터를 시각화하여 정보를 명확하게 전달합니다.
- ❑ **데이터 기반 의사결정:** 데이터를 기반으로 전략을 개발하고 문제를 해결하는 의사결정을 내립니다.



2. 데이터 중심의 비즈니스

2. 데이터 중심의 비즈니스



퀴즈 : 비즈니스는 어디에서 시작할까요?

- 1번 : 제품
- 2번 : 고객의 니즈
- 3번 : 새로운 아이디어
- 4번 : 재능있는 사람들
- 5번 : 문제

2. 데이터 중심의 비즈니스



퀴즈 : 비즈니스는 어디에서 시작할까요?

1번 : 제품

2번 : 고객의 니즈

3번 : 새로운 아이디어

4번 : 재능있는 사람들

5번 : 문제

2. 데이터 중심의 비즈니스

2.1 데이터와 비즈니스는 어떤 관계가 있을까?

“비즈니스 = 문제 해결 과정”

문제

- 사업비가 없다.
- 고객은 필요로 하는데 시장에 제품이 없다.

2. 데이터 중심의 비즈니스

-

2.1 데이터와 비즈니스는 어떤 관계가 있을까?

“비즈니스 = 문제 해결 과정”

문제

- 사업비가 없다.
- 고객은 필요로 하는데 시장에 제품이 없다.

비즈니스 제약

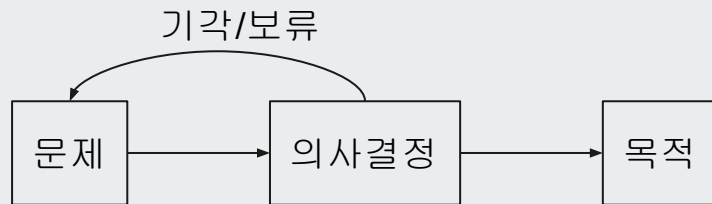
- 돈, 사람, 시간
- 이 문제를 풀지 말지

2. 데이터 중심의 비즈니스

2.1 데이터와 비즈니스는 어떤 관계가 있을까?

문제를 풀기로 의사결정이 난 다음 어떻게 해야할까?

- 이 문제를 왜 풀어야 하는가?
- 어떤 **impact**를 얻을 수 있는가?
- 무슨 목적을 달성할 수 있는가?
- 이 문제의 배경은 무엇인가?

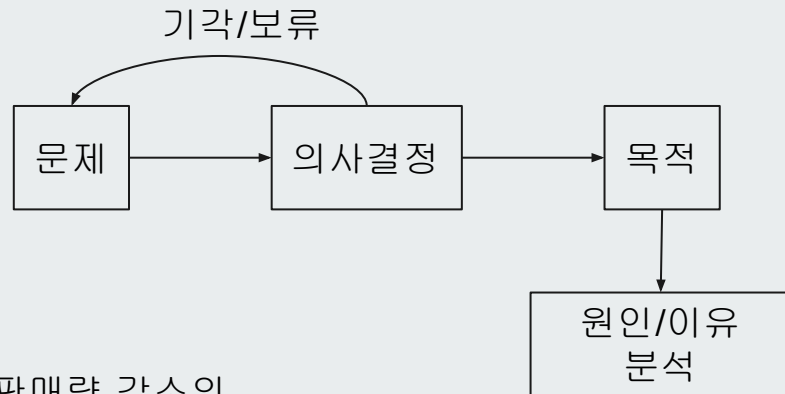


2. 데이터 중심의 비즈니스

2.1 데이터와 비즈니스는 어떤 관계가 있을까?

목적의식을 분명히 한 다음 어떻게 해야할까?

- 이러한 문제가 발생한 원인이 뭐지?
- **문제** : 온라인 소매업체의 판매량 감소
- **목적** : 판매량 감소 개선
- **분석** : 판매 관련 **데이터 수집 및 분석**을 통해 판매량 감소의 원인 파악
- 마케팅 캠페인의 효과 부족, 제품의 경쟁력 저하, 웹사이트의 사용자 경험

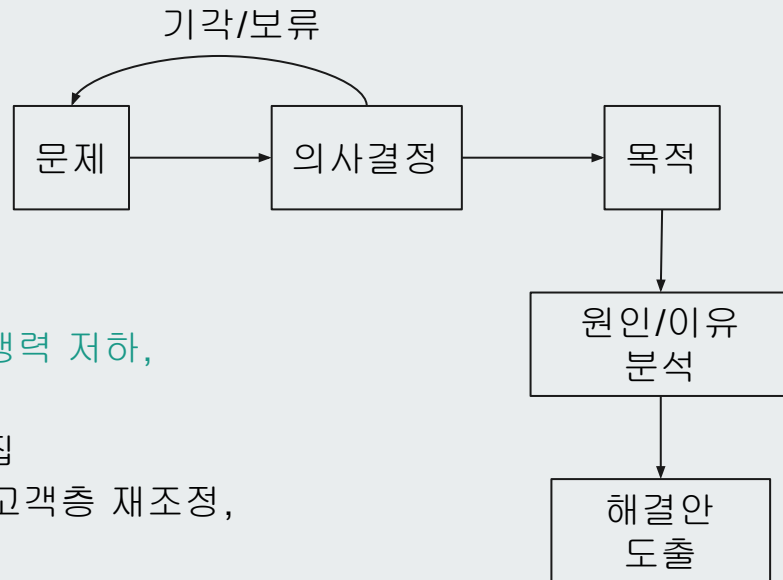


2. 데이터 중심의 비즈니스

2.1 데이터와 비즈니스는 어떤 관계가 있을까?

문제의 원인을 파악한 다음 어떻게 해야할까?

- 문제 : 온라인 소매업체의 판매량 감소
- 목적 : 판매량 감소 개선
- 분석 : 마케팅 캠페인의 효과 부족, 제품의 경쟁력 저하, 웹사이트의 사용자 경험
- 해결안 도출 : 원인을 파악한 후 개선 전략 수집
- 마케팅 캠페인의 효과를 개선하기 위해 타겟 고객층 재조정, 제품의 기능 개선이나 가격 조정

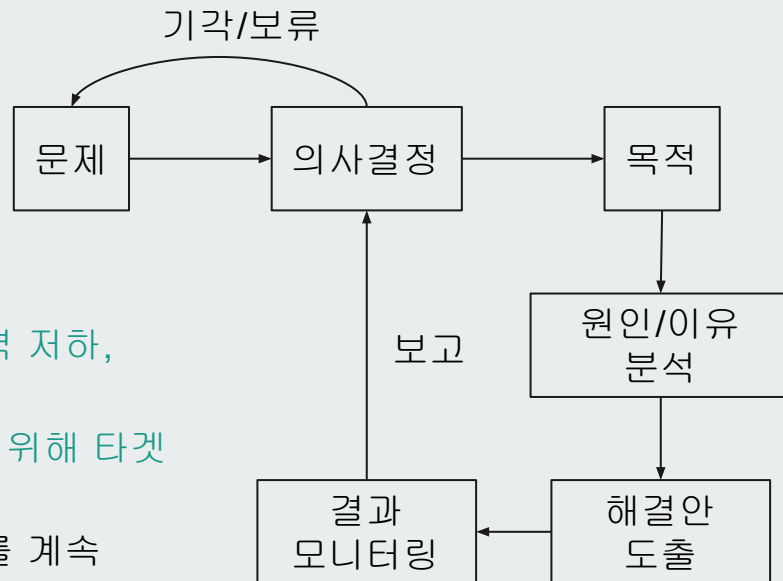


2. 데이터 중심의 비즈니스

2.1 데이터와 비즈니스는 어떤 관계가 있을까?

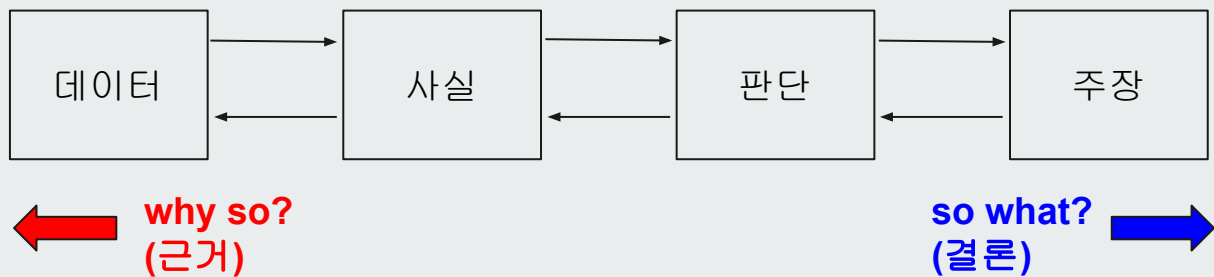
문제의 원인을 파악한 다음 어떻게 해야할까?

- 문제 : 온라인 소매업체의 판매량 감소
- 목적 : 판매량 감소 개선
- 분석 : 마케팅 캠페인의 효과 부족, 제품의 경쟁력 저하, 웹사이트의 사용자 경험
- 해결안 도출 : 마케팅 캠페인의 효과를 개선하기 위해 타겟 고객층 재조정, 제품의 기능 개선이나 가격 조정
- 결과 모니터링 : 개선 전략을 실행한 후, 데이터를 계속 모니터링하여 결과를 추적. 판매량의 변화와 관련된 데이터를 지속적으로 분석하여 개선 전략의 효과를 평가하고 필요한 조치 취하기.



2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식



2. 데이터 중심의 비즈니스



2.2 일 잘하는 데이터 과학자의 사고방식

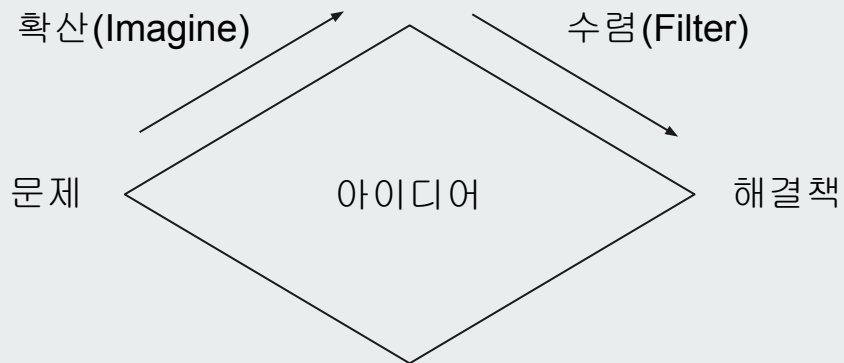
데이터 역량

- **데이터 수집** : 적절한 데이터 수집을 통해 정확하고 신뢰할 수 있는 자료를 확보할 수 있습니다.
- **사실 파악** : 데이터 분석 기술과 도구를 활용하여 유용한 통찰력을 얻을 수 있습니다.
- **판단** : 데이터를 신중하게 평가하고 판단 기준을 확립하여 올바른 결정을 할 수 있습니다.
- **주장** : 데이터를 이해하고 분석하여 강력한 주장을 구성할 수 있습니다.

2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식

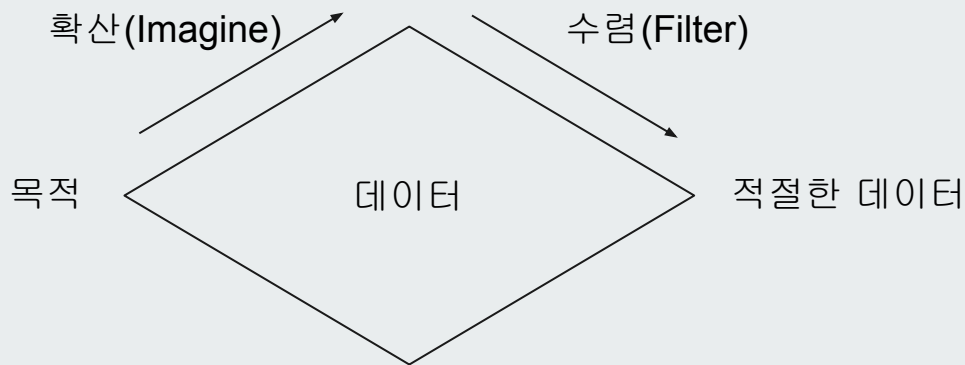
확산적 사고와 수렴적 사고



2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식

적절한 데이터 수집

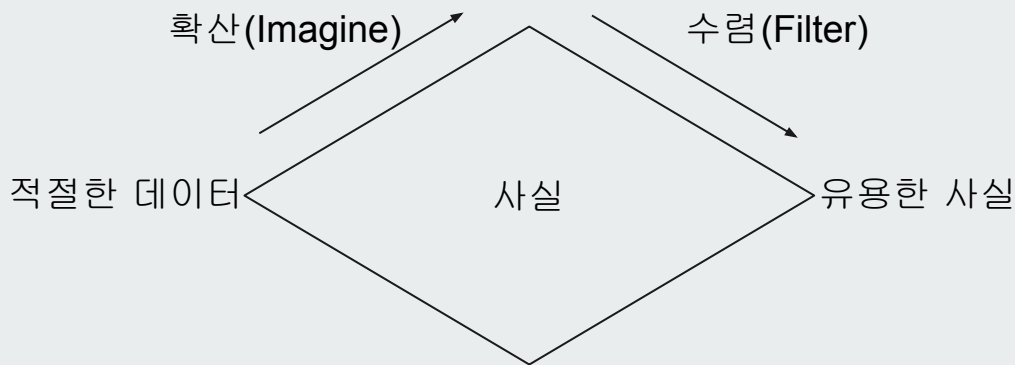


- ❑ **확산적 사고:** 확산적 사고에서는 다양한 소스에서 데이터를 수집하는 방법을 고민할 수 있습니다. 예를 들어, 고객의 구매 데이터 뿐만 아니라 소셜 미디어, 온라인 리뷰, 설문 조사 등 다양한 소스에서 데이터를 수집하여 더 많은 정보를 얻을 수 있습니다.
- ❑ **수렴적 사고:** 수렴적 사고에서는 확산적 사고에서 고민한 여러 소스 중 목적에 맞는 데이터를 추려냅니다.

2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식

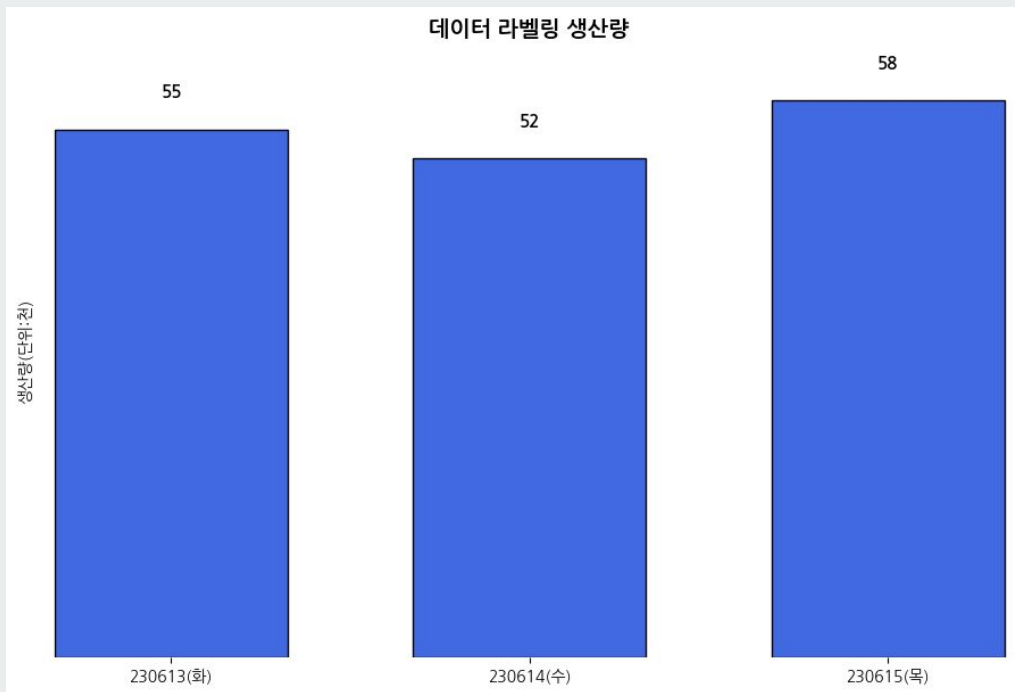
유용한 사실 선정하기



- ❑ **확산적 사고:** 확산적 사고에서는 다양한 데이터 분석 기법과 시각화를 활용하여 사실을 파악할 수 있습니다. 예를 들어, 고객의 구매 데이터를 분석하여 특정 상품의 인기도, 고객의 선호도 등을 파악할 수 있습니다.
- ❑ **수렴적 사고:** 수렴적 사고에서는 기존에 알고 있던 사실과 데이터 분석 결과를 비교하고, 새로운 사실을 발견할 수 있습니다. 예를 들어, 기존에는 단순히 A 제품의 매출이 줄어들었다고 생각했지만, 데이터 분석을 통해 경쟁사 제품이 할인 이벤트로 가격 경쟁력을 높인 것을 발견할 수 있습니다.

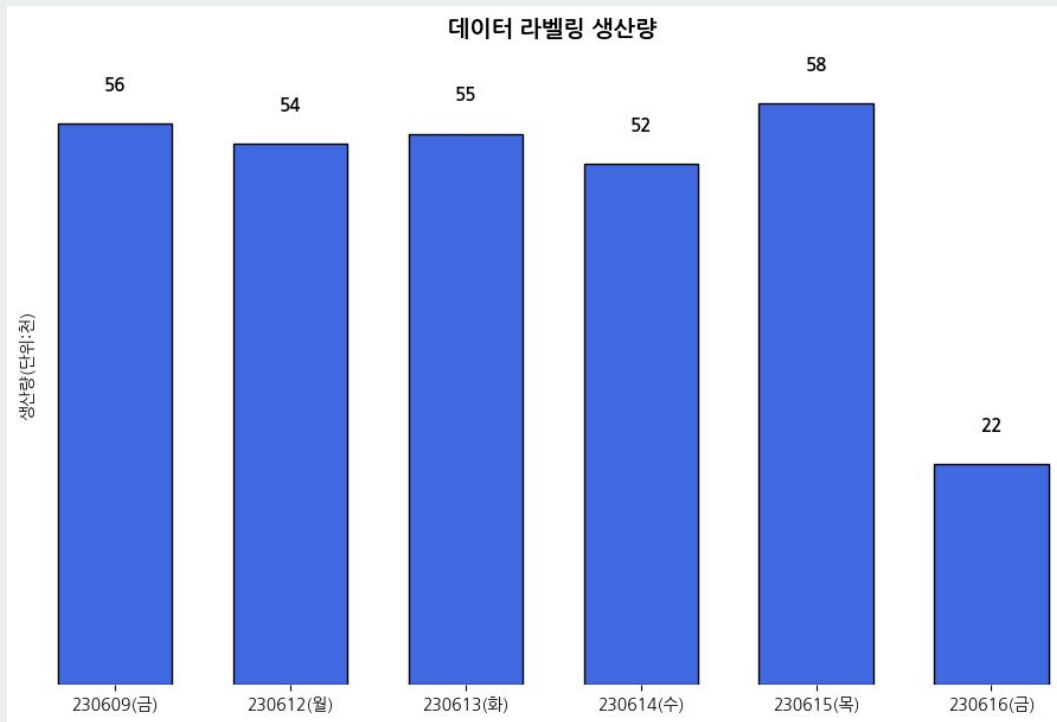
2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식



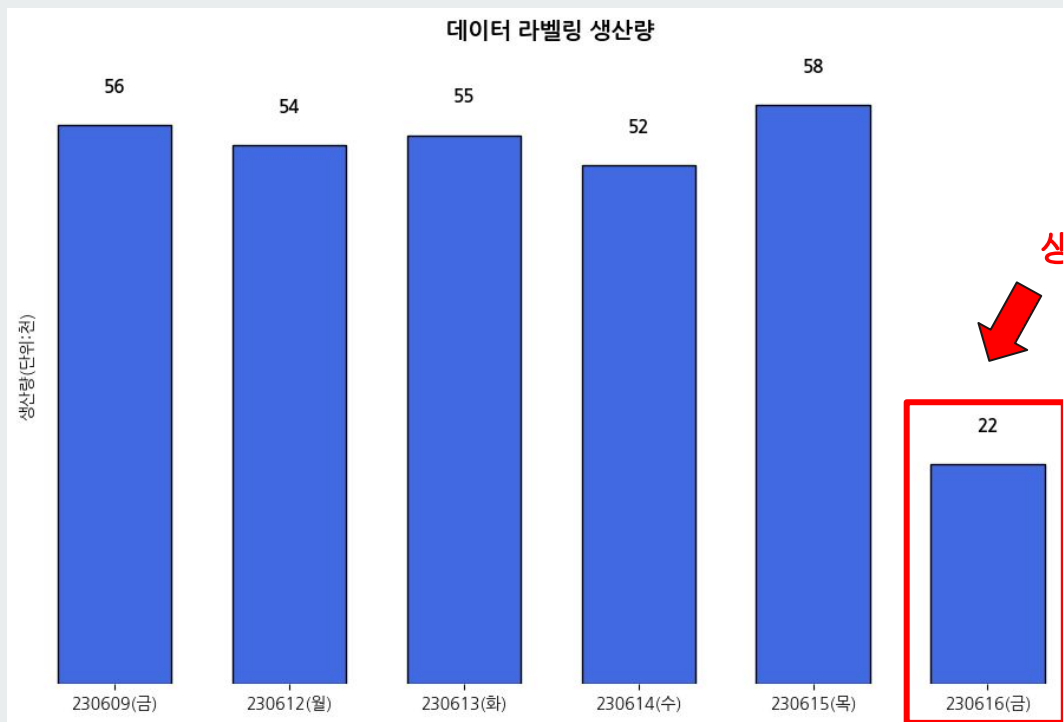
2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식



2. 데이터 중심의 비즈니스

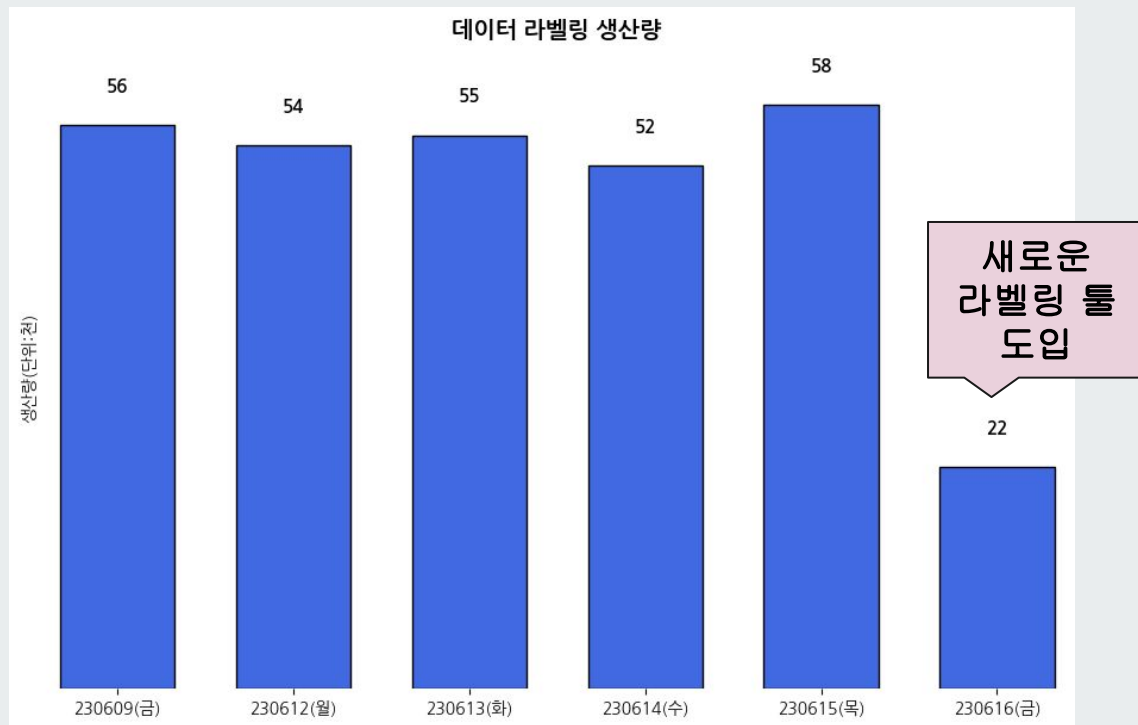
2.2 일 잘하는 데이터 과학자의 사고방식



생산량 반토막!

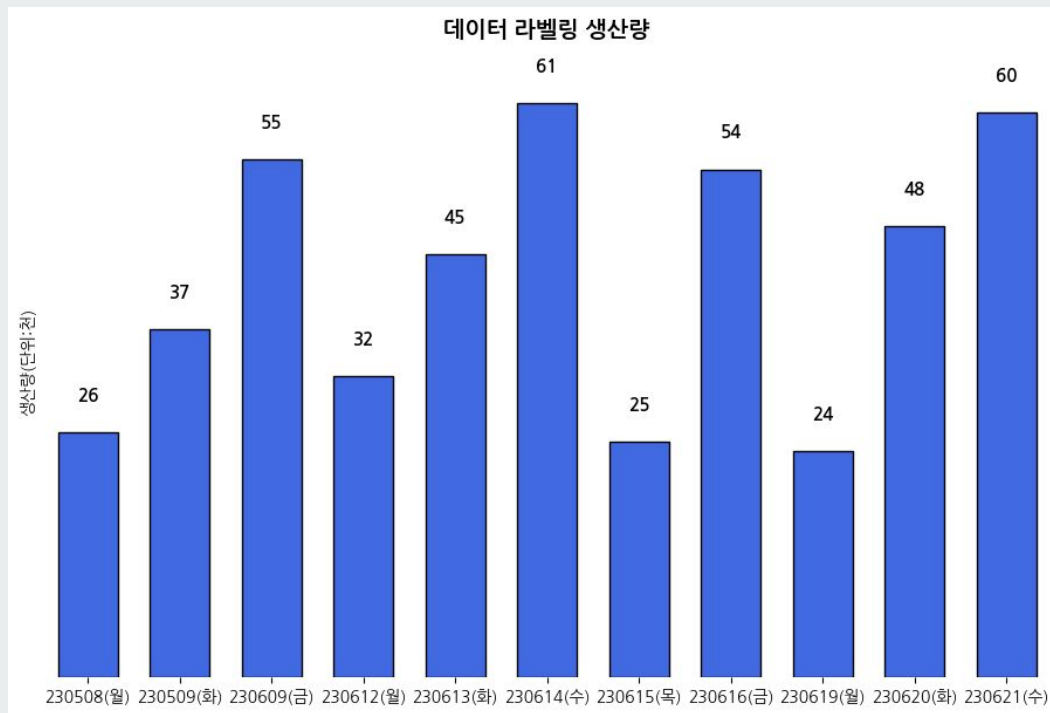
2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식



2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식



2. 데이터 중심의 비즈니스



2.2 일 잘하는 데이터 과학자의 사고방식

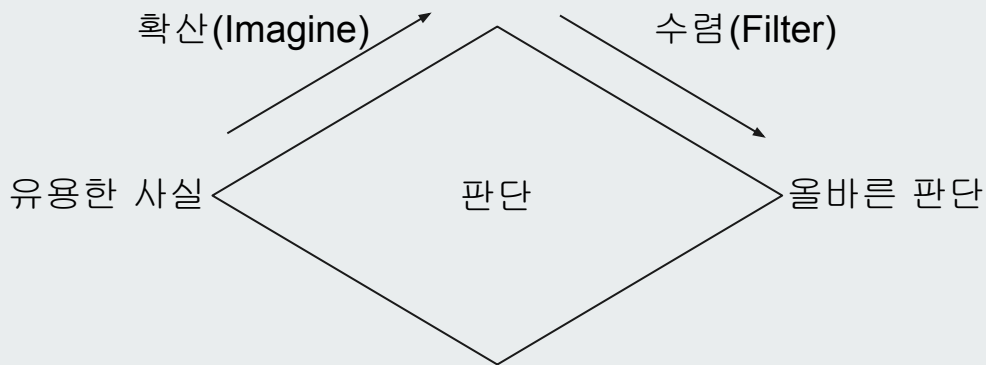
내가 무엇을 모르는지 모를 때 무지를 극복하는 방법

- 데이터 분석(인과 관계 분석, 상관 관계 분석)
- 새로운 데이터를 만들어 분석
- 현장 담당자의 노하우나 고객 경험으로부터 힌트 얻기(패턴 찾기)

2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식

올바른 판단 선정하기



- ❑ **확산적 사고:** 확산적 사고에서는 **다양한 가능성을 고려하여 판단**을 내릴 수 있습니다. 예를 들어, 다양한 변수와 지표를 고려하여 고객 이탈 가능성을 예측하고, 이를 기반으로 예방적인 조치를 취할지를 판단합니다.
- ❑ **수렴적 사고:** 수렴적 사고에서는 **기존의 지식과 경험을 바탕으로 판단**을 내릴 수 있습니다. 예를 들어, 과거에 유사한 상황에서 적용된 전략이 효과가 있었다면, 비슷한 전략을 선택하여 판단을 내릴 수 있습니다. 또한, 수렴적 사고에서는 특정한 기준이나 가설을 설정하고 그에 따라 데이터를 분석하여 판단을 내릴 수 있습니다. 혹은, 특정 지표가 특정 임계값을 초과하면 문제가 발생한다는 가설을 설정하고, 해당 지표를 모니터링하여 판단을 내릴 수 있습니다.

2. 데이터 중심의 비즈니스



2.2 일 잘하는 데이터 과학자의 사고방식

판단의 종류

- ❖ 사실 판단 : 과학적 사실에 기반한 판단
- ❖ 가치 판단 : 어떤 사실을 기준으로 판단

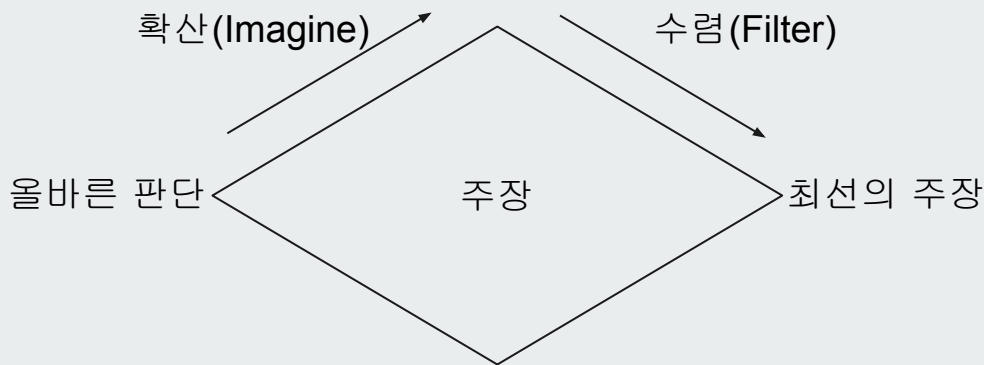
판단 예시

- ❖ 사실 판단 : A 라벨링 작업자의 작업량은 평균적인 작업자의 작업량보다 **40%** 높다.
- ❖ 가치 판단 : A 라벨링 작업자는 생산성이 높다. 또는, 다른 작업자들의 생산성이 낮다.

2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식

최선의 주장 선정하기



❑ **확산적 사고:** 확산적 사고에서는 다양한 가능성을 고려하고, 새로운 아이디어나 주장을 제시할 수 있습니다. 예를 들어, 고객 데이터를 분석하여 새로운 시장을 발견하고, 새로운 마케팅 전략을 주장할 수 있습니다.

❑ **수렴적 사고:** 수렴적 사고에서는 기존의 지식과 데이터를 토대로 가장 타당한 주장을 만들 수 있습니다. 예를 들어, 특정 제품의 특징이나 마케팅 전략의 효과가 있었으므로 비슷한 다른 제품에도 동일하게 적용하겠다는 주장을 할 수 있습니다.

2. 데이터 중심의 비즈니스



2.2 일 잘하는 데이터 과학자의 사고방식

판단

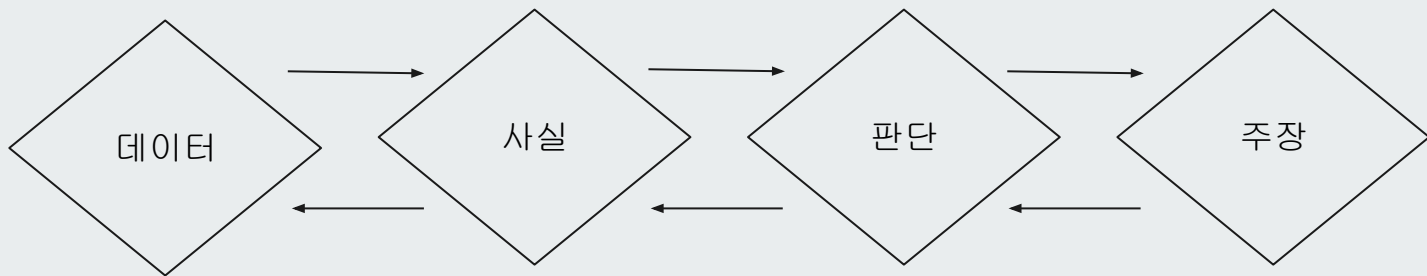
A 라벨링 작업자의 생산량이 높다.

주장 예시

- A 라벨링 작업자에게 평균 작업량 초과분만큼 성과금을 지급한다.
- A 라벨링 작업자에게 일회성 격려금을 지급한다.
- A 라벨링 작업자의 작업 정확도를 검증해야 한다.

2. 데이터 중심의 비즈니스

2.2 일 잘하는 데이터 과학자의 사고방식

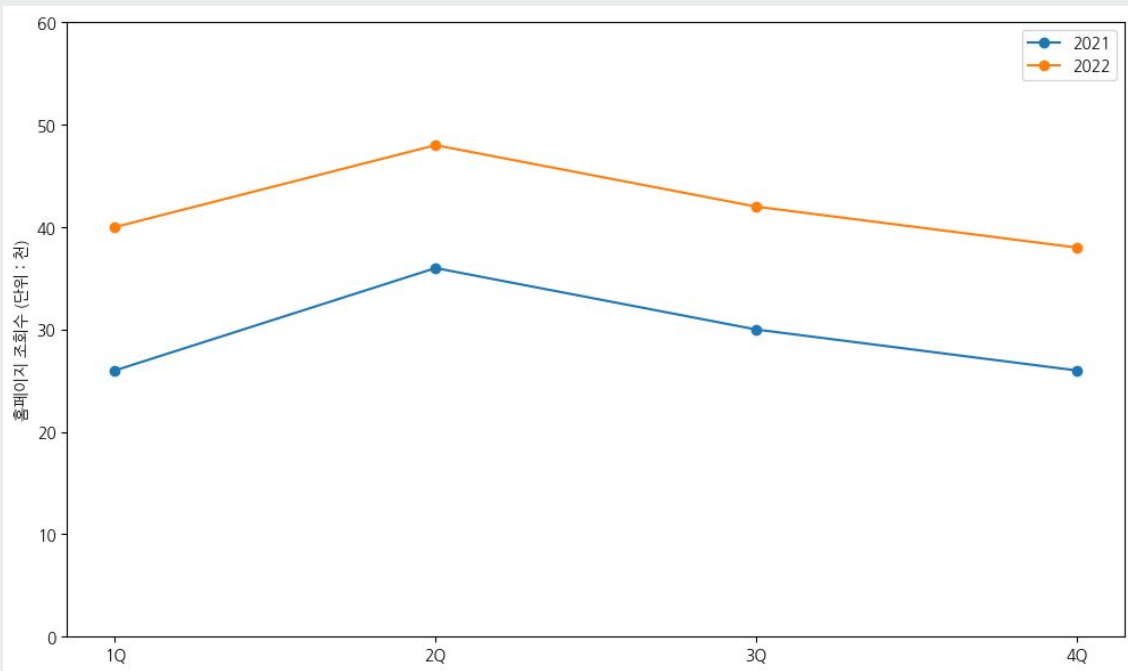




3. 상사에게 보고하는 방법

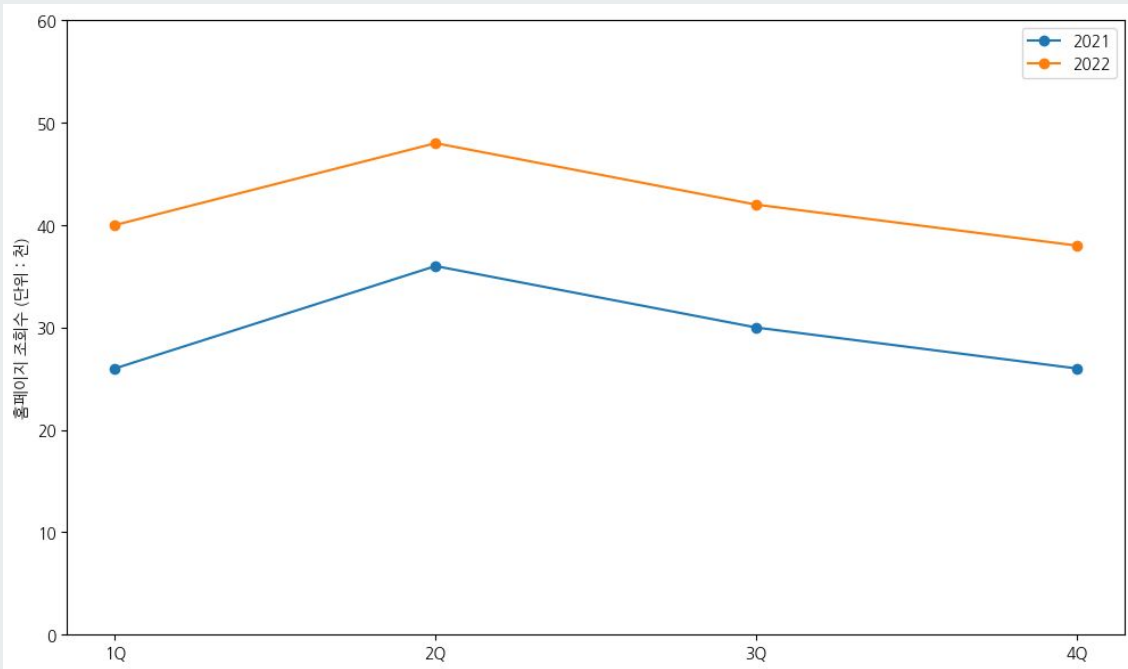
3. 상사에게 보고하는 방법

3.1 효과적인 보고



3. 상사에게 보고하는 방법

3.1 효과적인 보고

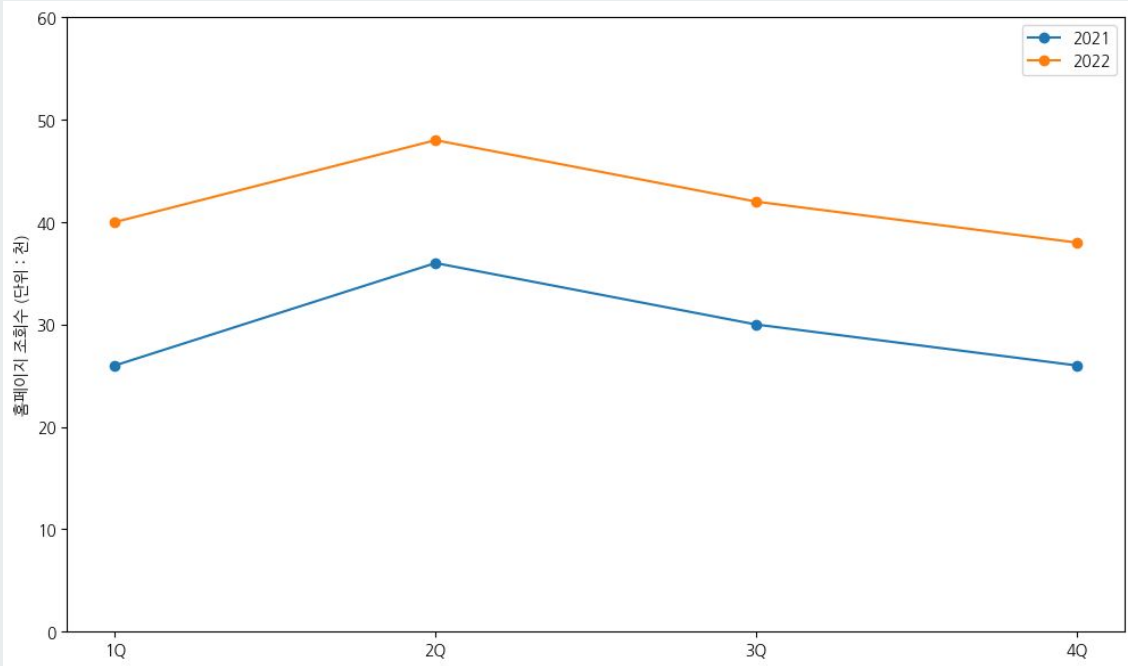


단순 차트 읽기

- 2022년 1Q 조회수 40000회
- 2021년보다 2022년 조회수 증가
- 2Q 조회수가 높음

3. 상사에게 보고하는 방법

3.1 효과적인 보고



통찰과 시사점

- 홈페이지 리뉴얼로 사용자 편의성 증가
- 매해 2분기마다 진행되는 이벤트로 인해 접속자 증가

3. 상사에게 보고하는 방법



3.2 보고 시 주의할 점

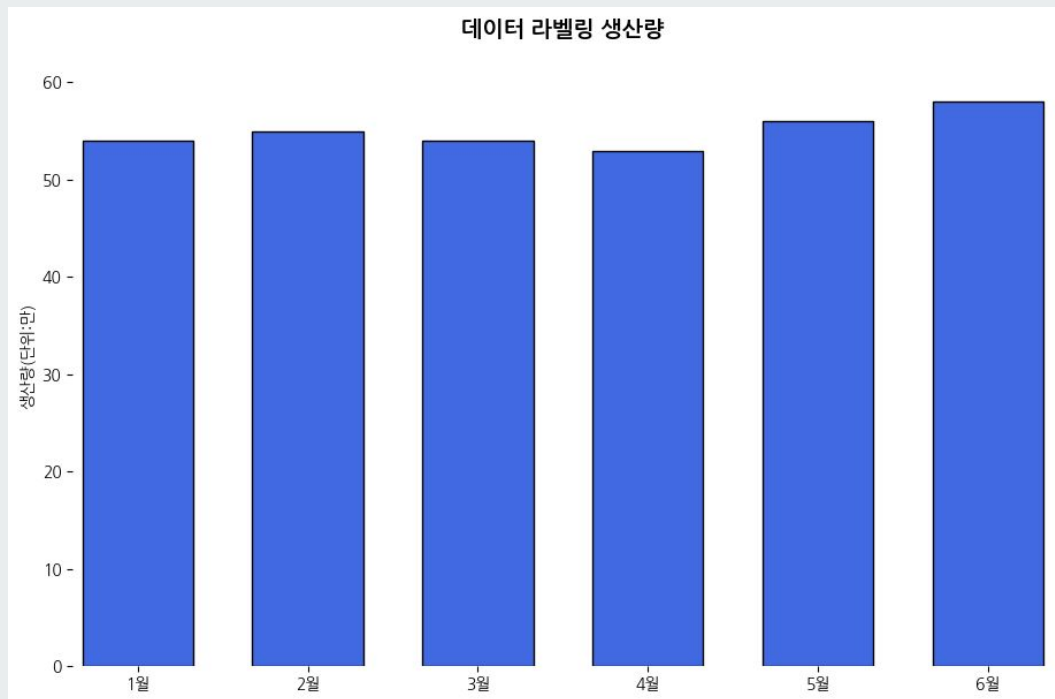
강조 : 시각화를 통해 중요한 정보를 강조하여 전달하는 것

왜곡 : 시각화를 통해 데이터를 잘못 표현하여 특정 목적을 달성하려는 것입니다.

데이터 시각화에서는 강조가 효과적인 정보 전달을 돕는 반면, 왜곡은 부적절하고 부정확한 정보 전달을 초래할 수 있으므로 신중하게 사용해야 합니다.

3. 상사에게 보고하는 방법

3.2 보고 시 주의할 점



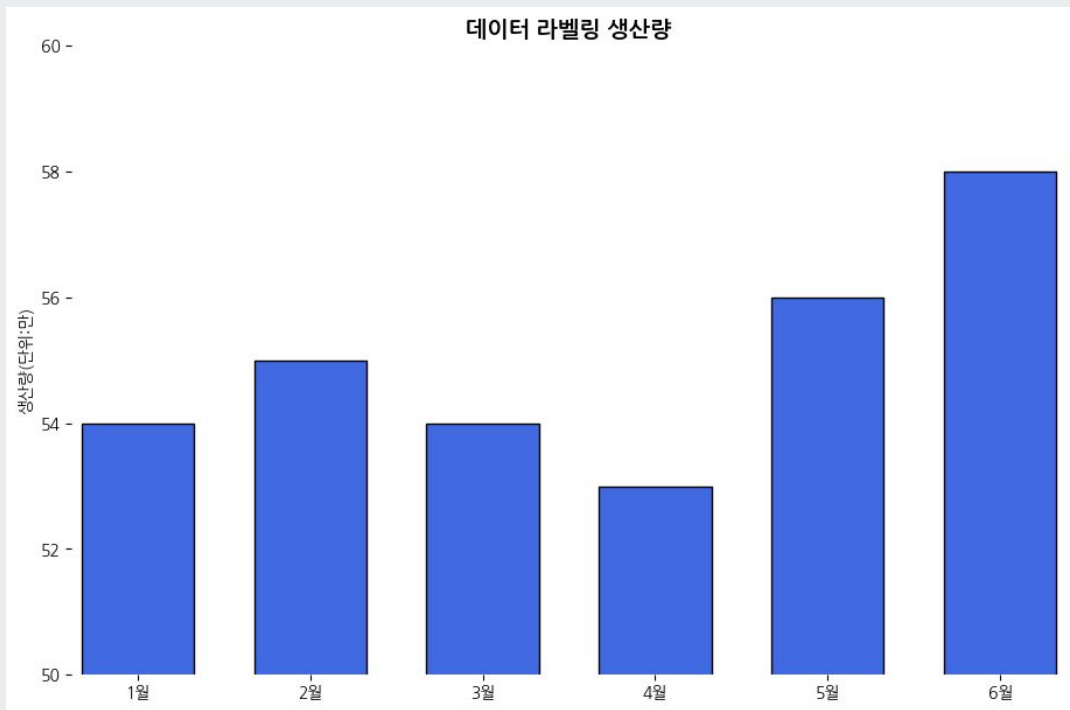
보고하고자 하는 내용 :

“4월부터 꾸준히 생산량이
증가하고 있다.”

→ 변화량이 미미해보이므로
성과가 뚜렷하게 보이지 않아 보고
받는 상사 입장에서 설득력 있는
시각화가 아니다. 변화하는 부분을
강조해서 보여주면 어떨까?

3. 상사에게 보고하는 방법

3.2 보고 시 주의할 점



“4월부터 생산량이 늘어난다는 것을 확실히 보여줄 수 있게 되었다!”

특정 부분 강조로 인해 발생한 왜곡

- y축의 시작점을 0이 아닌 50으로 변경
- 보고받는 사람의 입장에서 봤을 때 처음 이 그래프를 인지한 순간 4월과 6월의 생산량 차이가 2배 이상 나보임
- 수치가 잘 표현되어 있지 않아 데이터를 한 눈에 파악하기 어렵고 피로감을 줄 수 있음

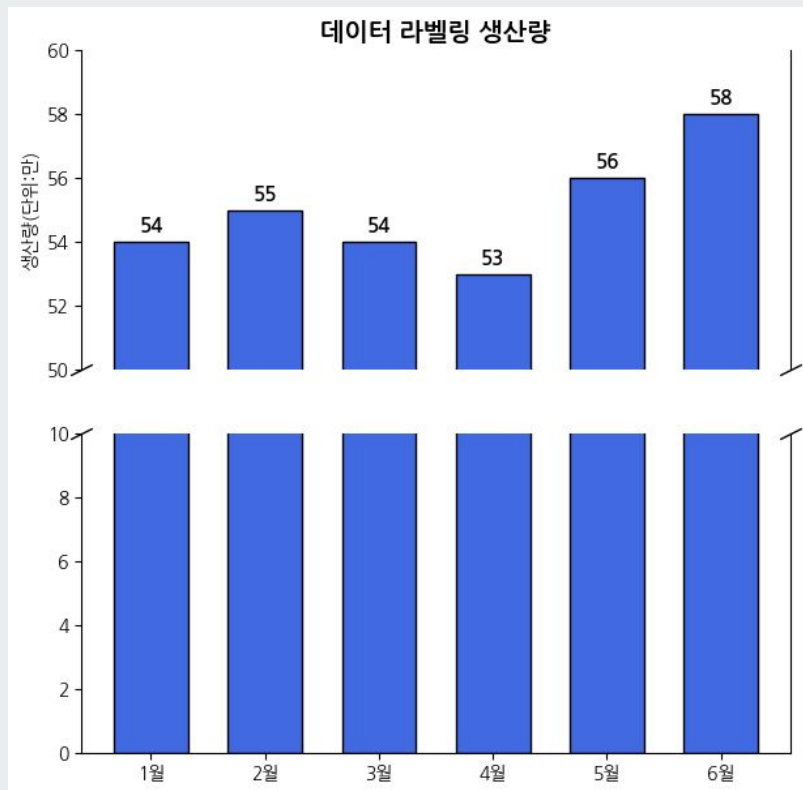
3. 상사에게 보고하는 방법

3.2 보고 시 주의할 점

왜곡을 줄이는 방법

- text로 수치 표시
- 물결 표시로 생략 부분 표현

→ 보고자의 입장에서 생산량의 변화를 뚜렷하게 보여줄 수 있으면서 보고 받는 입장의 상사는 한 눈에 차트를 보고 오해 없이 최소한의 비용으로 데이터를 파악할 수 있다.



3. 상사에게 보고하는 방법



3.2 보고 시 주의할 점

상사에게 보고하는 법 TIP

- 보고하기 전 무엇을 보고할 것인지 명확히 한다.
- 데이터를 알맞게 선정하고 그를 분석하여 유용한 정보를 뽑아낸다.
- 상사가 인지하길 원하는 부분을 한 눈에 알아보기 쉽도록 강조 등의 방법으로 데이터를 시각화한다.

정리하기



1. 데이터 리터러시

- a. 데이터 리터러시란?
- b. 데이터 리터러시의 중요성
- c. 데이터 리터러시의 핵심 능력

2. 데이터 중심의 비즈니스

- a. 데이터와 비즈니스는 어떤 관계가 있을까?
- b. 일 잘하는 데이터 과학자의 사고 방식

3. 상사에게 보고하는 방법

- a. 효과적인 보고
- b. 보고 시 주의할 점



감사합니다 :)



Q&A