

SESSION 01 · INTRODUCTION

데이터 분석 입문

데이터 분석의 기초부터 실전 활용까지

SESSION

데이터 분석이란 무엇인가

OUTCOME

분석의 6단계 흐름과 AI 활용법 익히기

01

CONTENTS

목차

-
- 01 데이터 분석이란? →

 - 02 데이터 분석의 6단계 →

 - 03 데이터의 종류 →

 - 04 AI와 함께 데이터 분석하기 →

 - 05 AI 사용 시 주의사항 →
-

01

SECTION 1

데이터 분석이란?

DEFINITION · 정의

데이터에서 의미 있는 정보를 도출하는 것

KEY "단순한 숫자"를 "의미 있는 정보"로 바꾸는 작업 — 이것이 데이터 분석의 본질입니다.

수집된 데이터를 정리 · 가공 · 해석하는 일

데이터 속의 패턴 · 경향 · 인사이트를 찾는 과정

관찰을 의사결정 가능한 형태로 정제하는 작업

WHY THIS MATTERS

모든 분석은 "데이터 → 의미"
라는 한 줄로 시작합니다.

01

SECTION 1

데이터 분석이란?

WHY · 왜 배울까

데이터 분석의 4가지 핵심 목적

01

사실 확인

"우리 학교 학생들의 평균 수면 시간은 얼마일까?"

02

원인 파악

"작년보다 동아리 가입률이 왜 떨어졌을까?"

03

미래 예측

"다음 시험 점수를 예측해보자."

04

합리적 의사결정

"부족한 과목에 시간을 얼마나 더 쓸까?"

4 CORE PURPOSES

사실을 보고, 원인을 찾고, 미래를 예측하고, 결정을 내린다 — 네 가지 목적이 분석의 모든 일을 설명합니다.

02

SECTION 2

데이터 분석의 6단계

WORKFLOW · 분석의 흐름

분석에는 흐름이 있다 — 6단계



⑤번 모델링은 이번 수업의 범위 밖 — 나머지 5단계를 학습합니다.

PROCESS

데이터 분석은 우연이 아닌 정해진 흐름을 따릅니다. 6단계를 한 번에 외우지 말고, 이름을 익히는 것부터 시작하세요.

02

SECTION 2

분석은 반복이다

MINDSET · 반복 가능성

기억할 것 — 분석은 한 방향이 아니다.



LOOP 분석에서 이상한 값을 발견하면 → 다시 전처리로 돌아갑니다. 이 "왕복"이 분석의 정상적인 모습입니다.

분석 → 새 문제 발견 → 다시 전처리

다시 분석 → 또 새 문제 발견 → 다시 전처리

한 번에 끝나는 분석은 거의 없다

KEY INSIGHT

한 번에 끝나는 분석은 없습니다. 전처리와 탐색적 분석 사이를 왕복하는 것이 "정상"입니다.

03

SECTION 3

데이터의 종류

TAXONOMY · 형태에 따른 분류

정형 · 반정형 · 비정형

TYPE 01 · 우리가 다룰 것

정형 데이터

STRUCTURED

행과 열로 깔끔하게 정리된 데이터.
표 형태로 한눈에 들어옵니다.

예시 · 엑셀, CSV, DB 테이블

TYPE 02

반정형

SEMI-STRUCTURED

구조는 있지만 자유로운 형식. 키와 값
으로 표현됩니다.

```
{ "name": "학생A",
  "score": 87, "tags":
  ["수학", "과학"] }
```

예시 · JSON, XML, 로그 파일

TYPE 03

비정형

UNSTRUCTURED

규칙적인 구조가 없는 데이터. 의미를
뽑아내려면 별도의 처리가 필요
합니다.



예시 · 사진, 영상, SNS 글

THREE FORMS

데이터는 "형태"에 따라 세 가지로 나뉩니다. 우리가 다룰 것은 가장 직관적인 정형 데이터입니다.

03

SECTION 3

수업에서 다룰 데이터

STARTING POINT · 우리의 출발

정형 데이터(CSV)부터 시작한다.

#	STUDENT	SLEEP_HR	STUDY_HR	SCORE
1	학생 A	6.5	3.2	78
2	학생 B	7.8	2.0	72
3	학생 C	5.2	4.5	85
4	학생 D	8.0	1.5	69
5	학생 E	6.0	3.8	82

sample.csv · 5 rows × 4 columns

엑셀과 비슷한 **행/열 구조**

통계와 시각화의 **기본기를 익히기에 최적**

비정형 데이터는 정형에 익숙해진 후 도전

WHY CSV

가장 직관적이고, 가장 많이 쓰이고, 가장 많은 도구에서 다룰 수 있는 형태입니다.

04

SECTION 4

AI를 분석 파트너로

AI AS PARTNER · 함께 분석하기

데이터를 첨부하고, 단계별로 대화한다.

✘ 한 번에 모든 걸 시키지 마세요. "이 파일 분석해 줘"는 결과가 모호하고 검증이 어렵습니다.

✓ 단계를 나눠 차근차근 묻기. 아래의 5단계 프롬프트 가이드를 따라가 봅니다.



5-STEP PROMPT GUIDE

한 번에 "분석해줘"라고 하지 말고, 다섯 단계로 나눠 차근차근 묻습니다.

04

SECTION 4

단계 1 · 2

PROMPT GUIDE · STEP 01 — 02

데이터를 파악하고, 문제부터 찾는다.

STEP 01 01 데이터 이해

우선 데이터셋이 무엇인지부터 묻습니다. 컬럼의 의미와 데이터 양을 파악해야 다음 질문이 만들어집니다.

"이 데이터셋 설명해줘. 각 컬럼이 뭘 의미하는지, 행 수는 몇 개인지."

STEP 02 02 전처리 점검

데이터에 문제가 있는지 확인합니다. 빠진 값(결측), 튀는 값(이상), 잘못된 형식을 미리 점검합니다.

"전처리해야 할 항목 있을까? 결측치 · 이상치 · 형식 문제를 확인해줘."

UNDERSTAND & INSPECT

먼저 데이터를 파악하고, 문제를 찾는 두 단계입니다.

04

SECTION 4

단계 3 · 4

PROMPT GUIDE · STEP 03 — 04

"무엇을" 분석할지 명확히 지시한다.

STEP 03 03 범위 지정

분석의 경계선을 그어줍니다. AI가 엉뚱한 방향으로 가지 않게.

"모델링 · 머신러닝은 진행하지 않을 거야. 전처리 후 EDA를 부탁해."

STEP 04 04 목적별 분석

한 번에 하나의 질문씩, 구체적으로 묻습니다.

"공부 시간과 시험 점수의 관계를 분석해줘."

"우리 반 수면 시간과 집중도의 상관관계를 알려줘."

SCOPE & QUESTION

AI에게 무엇을 분석할지 명확히 지시합니다. 범위가 좁아질수록 답은 또렷해집니다.

04

SECTION 4

단계 5

PROMPT GUIDE · STEP 05

의사결정 가능한 리포트로 정리한다.

FINAL PROMPT · STEP 05

"지금까지 분석한 내용을 바탕으로 데이터 분석 보고서를 작성해줘. 단계별로 꼼꼼하게, 시각화가 필요한 요소도 명시하고, 마지막에 인사이트와 제안을 포함해줘."

BEST PRACTICE · 보고서 요청

STEP 05 / 05

FINAL STEP

분석 결과를 의사결정 가능한 형태로 정리합니다. 시각화와 인사이트까지 함께 요청하세요.

TIP

한 번에 다 시키지 말고 — 단계별로 대화하기. 결과의 질이 달라집니다.

05

SECTION 5

① 보안 · 프라이버시

CAUTION 01 · 데이터 보안

01

민감한 정보는 절대 업로드하지 않는다.

AI는 입력한 데이터를 학습에 사용할 수도 있습니다. 무엇을 올리는지 항상 의식하세요.

RULE 01

민감한 정보는 절대 업로드하지 않습니다. 익명화·공개 샘플을 사용하세요.



업로드 금지 이름 · 주민번호 · 연락처 · 학교 내부 자료



익명화 또는 공개 샘플 데이터를 사용한다.



본인 데이터여도 다른 사람 정보가 섞여있다면 신중히.

REMEMBER

AI는 입력한 데이터를 학습에 사용할 수도 있다.

05

SECTION 5

② 결과 검증

CAUTION 02 · 검증의 필요성

02

AI도 틀린다 환각을 의심하라.

AI는 통계 수치를 그럴듯하게 "지어낼" 수 있습니다.
AI의 답은 출발점이지 정답이 아닙니다.

RULE 02

AI도 틀립니다. 환각 (Hallucination)을 늘 의심하고, 다른 도구로 재검증하세요.



환각(Hallucination) 주의 AI는 통계 수치를 그럴듯하게 지어낼 수 있다.



의심스러우면 실제 도구(ABCD Jupyter, ML Lab)로 재검증.



"이 숫자 **어떻게 계산했어?**"라고 다시 물어보기.

REMEMBER

AI의 답은 출발점이지 정답이 아니다.

05

SECTION 5

③ 프롬프트 설계

CAUTION 03 · 프롬프트가 답을 만든다

"어떻게 묻느냐"가 결과를 결정한다.

✘ "분석해줘"

✔ "수면 시간과 시험 점수의 상관관계를 산점도 로 보여주고, 상관계수도 알려 줘."

구체적으로 — 무엇을, 어떻게, 어떤 형태로?

단계적으로 — 한 번에 다 시키지 않기

명확하게 — 모호한 표현은 피하기

RULE 03

같은 데이터라도 "묻는 방식"에 따라 결과가 완전히 달라집니다.



SUMMARY

꼭 기억할 핵심 3가지

TODAY'S TAKEAWAYS

꼭 기억할 핵심 3가지.

01

데이터 분석은 데이터
에서 의미를 찾는 과정.

"6단계 흐름을 따른다. 정형
→ 정리 → 의미."

02

데이터를 수집, 전처리,
시각화 해보기.

"마지막은 개인 리포트로 마
무리한다."

03

AI는 강력한 도구이지만
검증 · 보안 · 프롬프트가
핵심.

"답은 출발점이지 정답이 아
니다."

TAKEAWAYS

세 가지만 들고 갑니다.

Q&A

궁금한 점이 있다면 자유롭게
질문해 주세요.

오늘 배운 내용 중 막힌 부분이 있으면 바로 물어보세요.

데이터 모델링

