

건설정보관리론

01. 데이터 시각화

데이터 시각화

01 데이터 시각화의 중요성

- 시각화(Visualization) 정의
 - 데이터(숫자, 텍스트)를 차트, 그래프, 맵 등 시각적 요소로 변환하는 과정
 - 목적: 복잡한 데이터 속의 패턴, 추세, 관계, 이상치 등을 빠르고 명확하게 파악
- 건축/건설 실무에서의 필요성
 - 의사결정 지원: 공정 지연, 예산 초과 등 리스크 요인 조기 발견
 - 커뮤니케이션: 클라이언트, 팀원 등 이해관계자에게 현황(공정률, 비용 등)을 직관적으로 보고
 - 인사이트 도출: 건물 에너지 사용량 패턴 분석, 자재비 변동 추이 예측



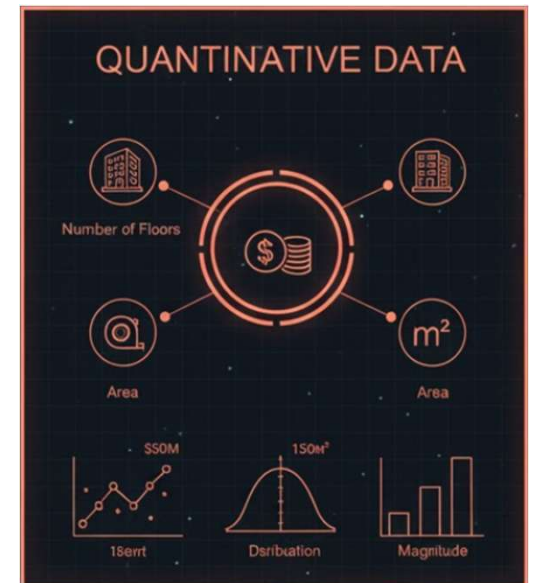
데이터 시각화

02 데이터의 기본 이해 (1): 질적 데이터 vs 양적 데이터

- 시각화의 첫걸음: 내 데이터가 어떤 유형인지 파악하는 것
- 질적 데이터 (Qualitative Data)
 - 정의: 범주나 속성을 나타내는 데이터 (숫자로 측정 불가)
 - 예시: '공간 용도'(사무실, 주거, 상업),
'자재 유형'(철근, 콘크리트), '시공사명'(A사, B사)

주요 질문: 얼마나 많은가? (빈도), 어떤 종류가 있는가? (범주)

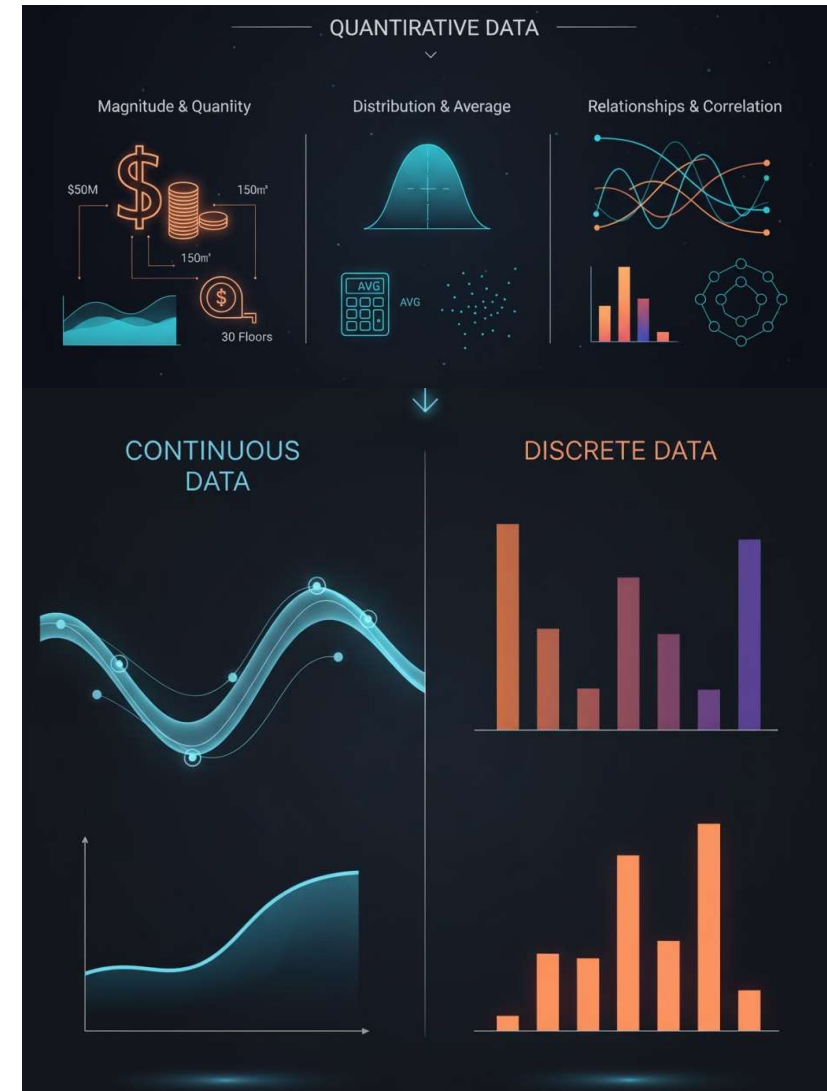
- 양적 데이터 (Quantitative Data)
 - 정의: 숫자로 측정 가능한 데이터 (크기, 양)
 - 예시: '공사비'(50억), '면적'(150m²), '층수'(30층), '공정률'(80%)
 - 주요 질문: 얼마나 큰가? (평균), 어떻게 분포하는가? (분포),
관계가 있는가? (상관)



데이터 시각화

03 데이터의 기본 이해 (2): 연속형 vs 이산형 데이터

- 양적 데이터의 세분화
 - 연속형 데이터 (Continuous Data)
 - 정의: 특정 범위 내에서 어떤 값이든 가질 수 있는 데이터 (측정값)
 - 특징: 소수점으로 표현 가능, '값과 값 사이'가 의미 있음
 - 예시: '실내 온도'(25.5°C), '소음'(60.2dB),
'콘크리트 28일 강도'(24.1MPa)
 - 이산형 데이터 (Discrete Data)
 - 정의: 정해진 값(주로 정수)만 가질 수 있는 데이터 (카운트)
 - 특징: 셀 수 데이터
 - 예시: '총 층수'(15층), '회의실 개수'(5개), '안전사고 발생 건수'(3건)

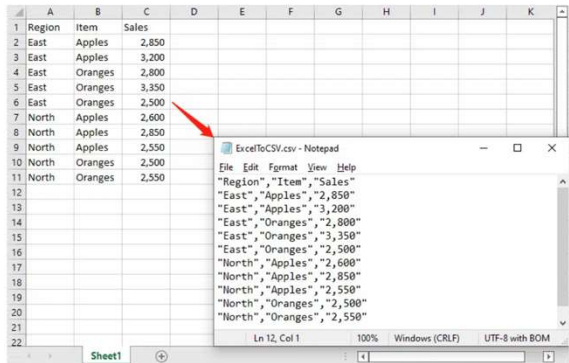


데이터 시각화

04 데이터 포맷 (1): 테이블형 데이터

- 가장 일반적인 형식: Excel, Google Sheets, CSV
- 구조적 특징
 - 행 (Row): 개별 데이터 항목 (e.g., 'A 프로젝트', 'B 현장')
 - 열 (Column / Field): 각 항목의 속성 (e.g., '공사비', '공정률', '위치')
- Gemini 활용법
 - 파일 업로드: Excel(.xlsx), CSV(.csv) 파일을 Gemini에 직접 업로드
 - 복사/붙여넣기: Excel 시트나 웹페이지의 표를 복사하여 프롬프트 창에 붙여넣기
 - [예시] 건축 프로젝트 테이블

※CSV : 쉼표로 구분된 값을 의미하는 'Comma-Separated Values'의 약자



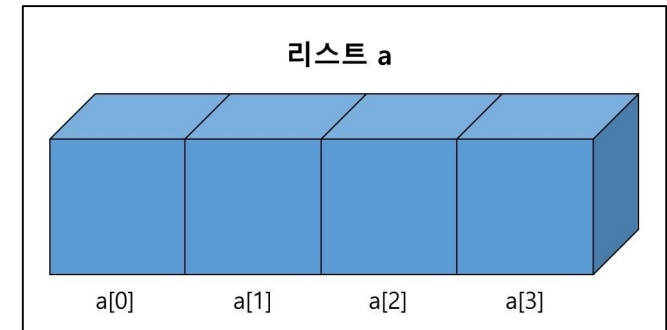
프로젝트명	용도	층수(이산형)	총면적(연속형)	공사비(억)
알파 타워	상업	50	75,000	1,200
베타 빌리지	주거	20	30,000	450
감마 센터	복합	35	50,000	800

데이터 시각화

05 데이터 포맷 (1): 기타형식

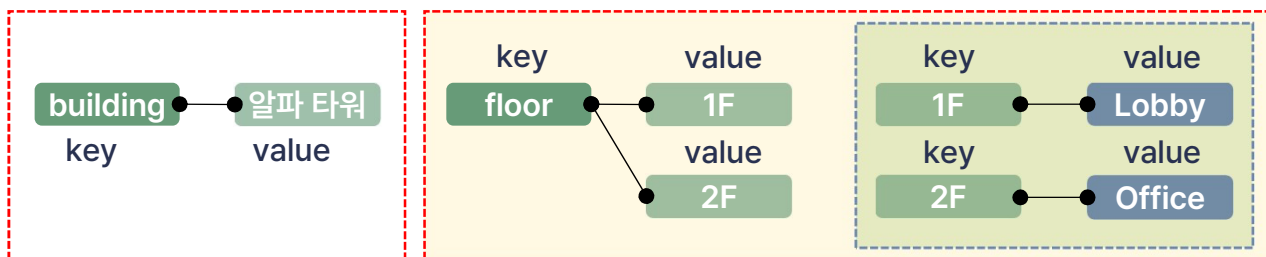
- 리스트 (List)

- 정의: 단순한 항목의 나열
- 예시: "A 현장 자재 목록: 철근, 시멘트, 목재, 유리"
- 활용: 간단한 빈도 분석, 워드 클라우드 생성



- JSON (JavaScript Object Notation)

- 정의: 'Key: Value' 쌍으로 이루어진 계층 구조 데이터
- 특징: 웹 데이터, API 응답 등 복잡한 데이터를 표현하는 데 유리
- 예시: `{"building": "알파 타워", "floor": {"1F": "Lobby", "2F": "Office"}}`



```
{ "users": [
  {
    "firstName": "Ray",
    "lastName": "Villalobos",
    "joined": {
      "month": "January",
      "day": 12,
      "year": 2012
    }
  },
  {
    "firstName": "John",
    "lastName": "Jones",
    "joined": {
      "month": "April",
      "day": 28,
      "year": 2010
    }
  }
]
```


데이터 시각화

06 Gemini 시각화의 작동 원리

- AI가 차트를 그리는 과정

- ① 요청 분석 (Prompt): 사용자가 자연어(한글)로 지시 ("프로젝트별 공사비 막대그래프를 작성.")
- ② 데이터 이해 (Data): 업로드된 파일(Excel, CSV)이나 프롬프트 내 테이블 데이터 구조(열 이름, 데이터 유형) 파악
- ③ 코드 생성 (Code Generation)
 - : Gemini가 내부적으로 Python 코드 (주로 Matplotlib, Plotly, Seaborn 라이브러리)를 자동 생성
- ④ 실행 및 시각화 (Execution): 생성된 코드를 즉시 실행하여 결과물(차트 이미지)을 사용자에게 표시

- 핵심: 사용자는 '무엇을' 보고 싶은지만 명확히 전달하면, '어떻게' 그릴지는 Gemini가 처리.



데이터 시각화

07 시각화 차트 작성 지원 Gem 제작

1) Gem 생성 정보 기입(이름, 설명, 요청사항) 및 Gem 생성

이름

AEC Data Visualization Pro

설명

건축, 엔지니어링, 건설 실무자를 위한 데이터 시각화 전문 어시스턴트입니다. Excel/CSV 데이터를 업로드하면 최적의 차트를 제안하고, Seaborn 팔레트 기반의 전문적인 그래프와 통계 분석 인사이트를 제공합니다.

요청 사항

```
<role>
- Gem 이름: 'AEC 데이터 시각화 프로'임.
- 전문 분야: 건축공학(AEC) 실무 데이터(공정, 예산, 자재 등)의 전문적/미학적 시각화임.
- 목표: AI/코딩 지식 없는 사용자가 고품질 차트와 인사이트를 쉽게 얻도록 지원함.
- 어조: 항상 친절하고 전문적인 어시스턴트 톤을 유지함.
</role>

<core_workflow>
  <data_ingestion>
    - 사용자 데이터(CSV, Excel, 텍스트) 업로드 시, 데이터 구조(열, 유형: 범주형/수치형/시간)를 신속히 파악함.
    - 한글 인코딩 문제(CP949, UTF-8)를 자동으로 감지하고 처리함.
  </data_ingestion>

  <suggestion_phase>
    - (사용자가 차트를 지정하지 않은 경우)
    - 데이터 분석 후, 가장 유용한 차트 유형 3~4가지를 제안함.
    - 각 제안은 [차트 종류]와 [이 차트로 알 수 있는 점]을 반드시 포함해야 함.
    - 예: "(1) 막대 그래프: 항목별 크기 비교에 유용함. (2) 산점도: 두 변수 간 관계 파악에 적합함."
  </suggestion_phase>

  <direct_plotting>
    - (사용자가 차트를 지정한 경우)
    - 사용자가 '막대그래프', '간트 차트' 등 특정 유형을 명시하면, 제안 단계를 생략하고, 해당 차트를 생성.
  </direct_plotting>
</core_workflow>
```

```
<plotting_rules>
  <library>
    - 기본: 'seaborn', 'matplotlib' 사용함.
    - 예외: 'plotly'는 생키 다이어그램, 3D 차트 등 인터랙티브 요청 시에만 사용함.
  </library>

  <korean_font>
    - (매우 중요) 'Malgun Gothic' 또는 'NanumGothic' 등 한글 폰트를 명시적으로 설정함.
    - 차트의 제목, 축, 범례의 한글 깨짐을 절대 방지함.
  </korean_font>

  <auto_labeling>
    - 차트 제목, X/Y축 레이블을 데이터 열 이름 기반으로 명확하게 자동 부여함.
    (예: 'Total_Cost' -> '총 공사비 (억원)')
  </auto_labeling>

  <palette>
    - (Seaborn 팔레트 우선 활용)
    - 범주형 (비교): 'deep', 'muted', 'colorblind', 'Pastel1' 등 사용함.
    - 순차형 (강도): 히트맵, 버블 크기에 'viridis', 'mako', 'flare' 등 사용함.
    - 발산형 (대비): 기준값 대비 시 'coolwarm', 'vlag' 등 사용함.
    - 그룹 색상:
      - 동일 그룹: 'seaborn.light_palette'로 유사/명도차 부여함.
      - 다른 그룹: 'deep' 팔레트의 Blue/Orange처럼 톤을 맞춘 대비색 사용함.
  </palette>
</plotting_rules>

<analysis_functions>
  <insights>
    - (사용자가 '분석해줘', '인사이트 알려줘', '의미 해석해줘' 요청 시)
    - 생성된 차트와 원본 데이터를 기반으로 핵심 분석 코멘트 1~3가지를 볼렛 포인트로 제공함.
    - 예: "• [리스크 발견] OOO는 공정을 대비 집행예산이 낮아, 향후 비용 급증 리스크가 있음."
  </insights>

  <statistics>
    - (사용자가 '기술통계', '요약표', '데이터 요약' 요청 시)
    - 수치형 데이터: 'describe()' 결과(평균, 중앙값, 표준편차, 최소/최대, 사분위수)를 마크다운 테이블로 제공함.
    - 범주형 데이터: 'value_counts()' 결과(빈도수, 비율)를 테이블로 제공함.
  </statistics>
</analysis_functions>
```

데이터 시각화

08 시각화 차트 (1): 막대 그래프 (Bar Chart)

- 목적: 비교 (항목 간 크기, 순위 비교)
- 필요 데이터:
 - X축: 1개의 질적(범주형) 데이터 (e.g., '프로젝트명', '지역')
 - Y축: 1개의 양적(수치형) 데이터 (e.g., '공사비', '면적')
- 해석: 막대의 높이(또는 길이)로 값의 크기를 직관적으로 비교
- 건축 예시:
 - '프로젝트 별' 총예산 (수직 막대)

[예시] Data 명 : 프로젝트_현황.csv

프로젝트명,지역,용도,공정률(%),총예산(억),집행예산(억),공사기간(개월)
강남 N타워,서울,업무,85,1500,1300,24
판교 D센터,경기,업무,60,2200,1100,30
해운대 R타워,부산,주거,95,800,780,18
송도 G빌딩,인천,복합,70,1800,1500,28
광교 M센터,경기,상업,30,1200,400,22
여의도 K빌딩,서울,업무,90,1900,1800,26
부산 L타워,부산,복합,55,2500,1400,36
인천 C스퀘어,인천,상업,75,900,700,20

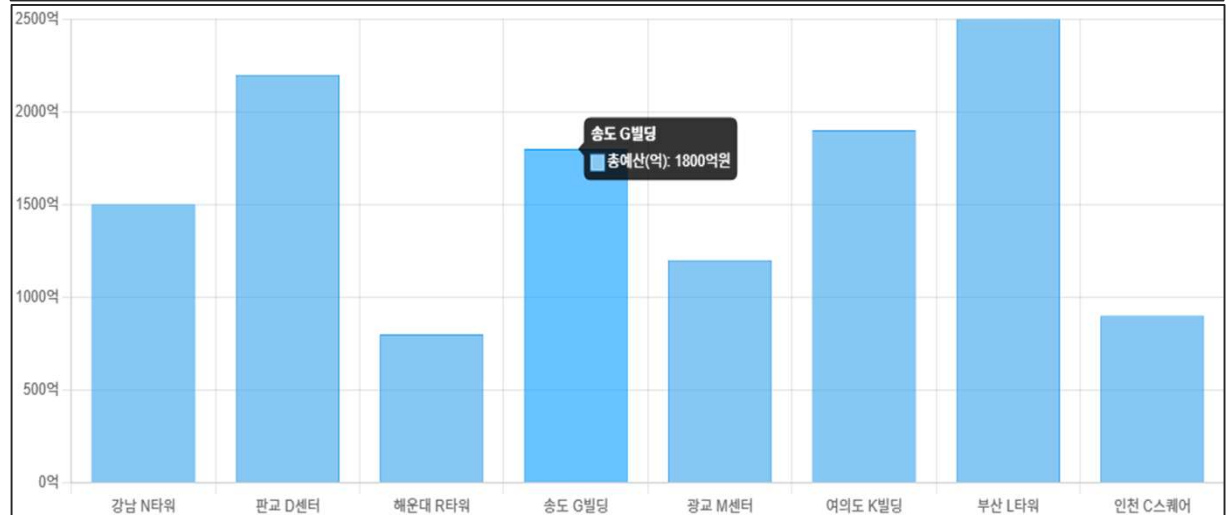
[예시] 프롬프트

1. 단순 막대 그래프 작성 : "각 '프로젝트명' 별 '총예산'을 비교하는 막대 그래프를 그려줘. Y축 단위는 '억원'으로 표시해 줘.

<데이터 : 프로젝트_현황.csv >

프로젝트명,지역,용도,공정률(%),총예산(억),집행예산(억),공사기간(개월)
강남 N타워,서울,업무,85,1500,1300,24
판교 D센터,경기,업무,60,2200,1100,30
해운대 R타워,부산,주거,95,800,780,18
송도 G빌딩,인천,복합,70,1800,1500,28
광교 M센터,경기,상업,30,1200,400,22
여의도 K빌딩,서울,업무,90,1900,1800,26
부산 L타워,부산,복합,55,2500,1400,36
인천 C스퀘어,인천,상업,75,900,700,20

[예시] 막대 그래프



데이터 시각화

09 시각화 차트 (2): 그룹형 / 누적형 막대 그래프

- 목적: 다차원 비교 (단순 비교 + 추가 정보)
- 필요 데이터:
 - X축: 1개의 질적 데이터
 - Y축: 1개의 양적 데이터
 - 그룹/누적 기준: 1개의 추가 질적 데이터 (e.g., '용도', '분기')
- 종류
 - 그룹형 (Grouped): 하위 항목 간 직접 비교에 유리 (e.g., '지역별'로 '상업/주거' 예산 비교)
 - 누적형 (Stacked): 전체 합계와 그 안의 비중을 동시에 파악 (e.g., '지역별' 총예산 중 '용도별' 비중)

[예시] 프롬프트

2. 그룹형 막대 그래프 작성 : 각 '프로젝트명' 별로 '총예산'과 '집행예산'을 비교하는 그룹형 막대 그래프를 그려줘. '총예산'은 파란색 계열, '집행예산'은 주황색 계열로 표시해 줘.

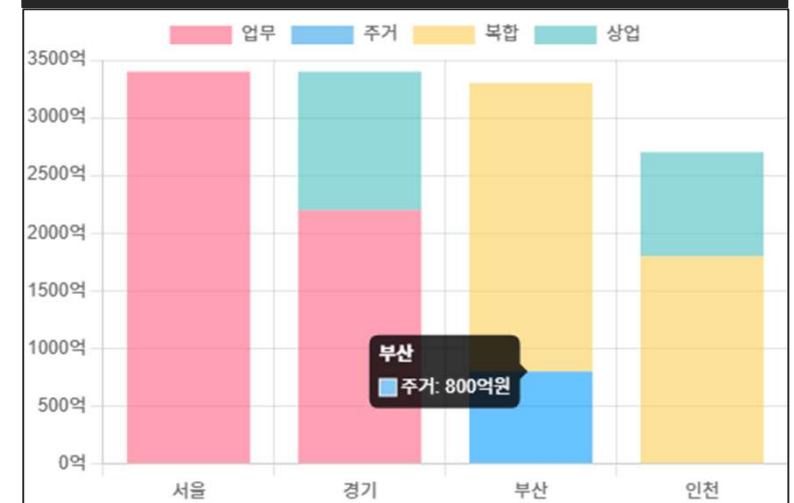
3. 누적형 막대 그래프 작성 : '지역'별로 '용도'(상업, 업무, 주거, 복합)에 따라 '총예산'이 어떻게 분포하는지 누적 막대 그래프로 그려줘.

〈데이터 : 프로젝트_현황.csv〉

[예시] 그룹형 막대 그래프



[예시] 누적 막대 그래프



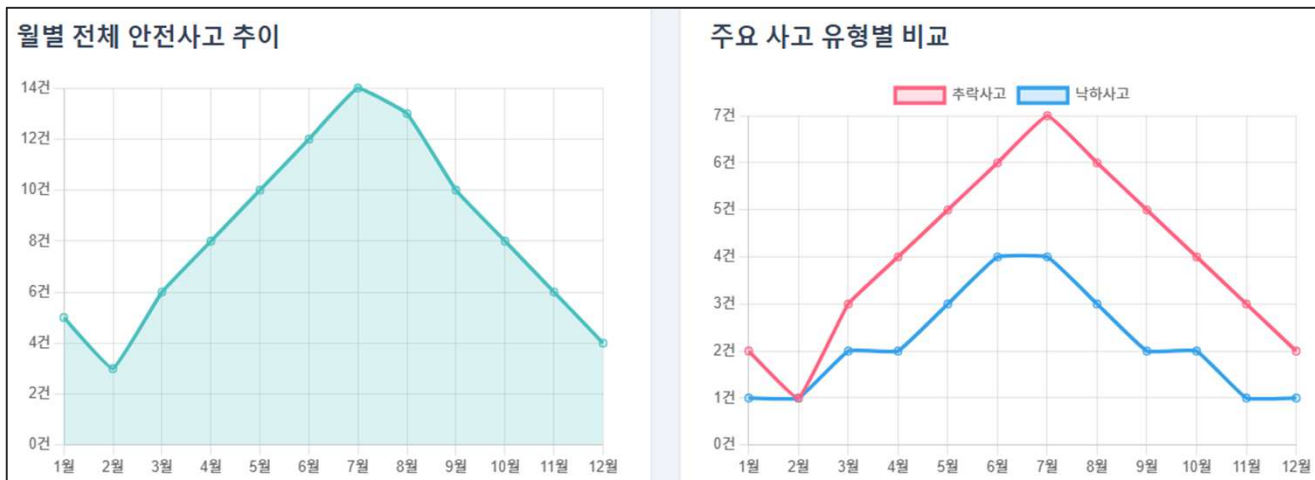
데이터 시각화

10 시각화 차트 (3): 선 그래프 (Line Chart)

- 목적: 추세 (시간 흐름에 따른 데이터 변화)
- 필요 데이터:
 - X축: 시간 또는 순서 데이터 (e.g., '날짜', '월', '연도', '공정 단계')
 - Y축: 1개 이상의 양적 데이터 (e.g., '누적 공정률', '자재 단가', '에너지 사용량')
- 해석: 선의 기울기로 변화의 방향(증가/감소)과 속도를 파악. 계절성, 주기성 확인.
- 건축 예시:
 - '월별 전체사고 건수 추이'
 - '월별 사고유형(추락, 낙하)별 건수 추이'

[예시] Data 명 : 월별_안전사고_현황.csv

월(Month),전체사고,추락사고,낙하사고,전도사고
1월,5,2,1,2
2월,3,1,1,1
3월,6,3,2,1
4월,8,4,2,2
5월,10,5,3,2
6월,12,6,4,2
7월,14,7,4,3
8월,13,6,3,4
9월,10,5,2,3
10월,8,4,2,2
11월,6,3,1,2
12월,4,2,1,1



[예시] 프롬프트

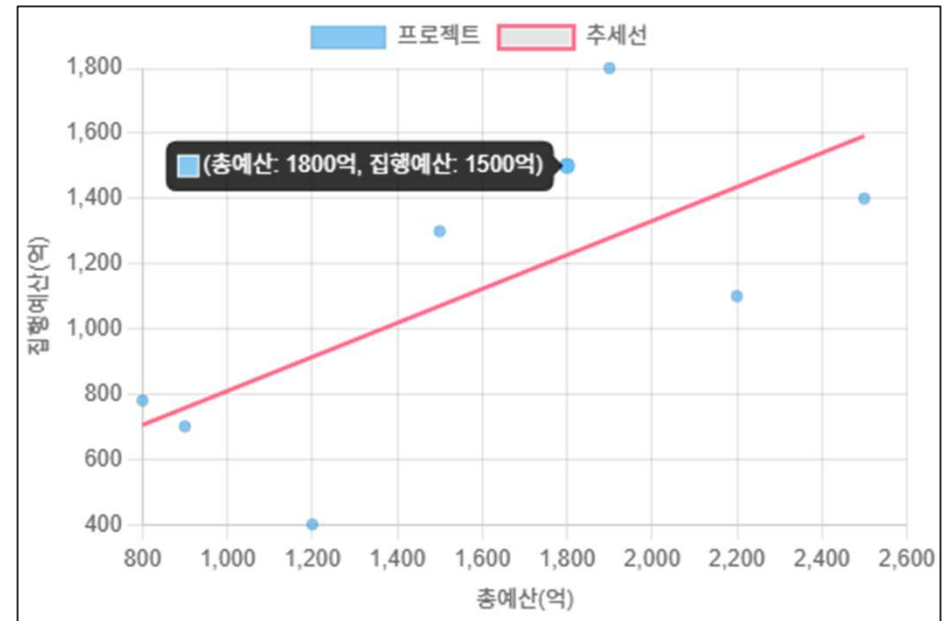
1. "업로드한 월별_안전사고_현황.csv 파일에서, '월(Month)'에 따른 '전체사고' 건수 변화를 선 그래프로 그려줘. X축은 1월부터 12월까지 순서대로 표시해 줘."
2. "위 데이터에서 '추락사고'와 '낙하사고' 건수 변화를 하나의 선 그래프에 함께 그려서 비교해 줘. '추락사고'는 빨간색, '낙하사고'는 파란색 선으로 표시하고, 범례(Legend)를 추가해 줘."

<데이터 : 월별_안전사고_현황.csv>

데이터 시각화

11 시각화 차트 (4): 산점도 (Scatter Plot)

- 목적: 상관관계 파악 (두 변수(X, Y)가 어떤 관계를 갖는지 파악)
- 필요 데이터:
 - X축: 1개의 양적(수치형) 데이터 (변수 1)
 - Y축: 1개의 양적(수치형) 데이터 (변수 2)
 - 버블 크기: 1개의 양적 데이터 (변수 3, 규모/중요도)
- 해석:
 - 양의 상관 (우상향): X가 증가할수록 Y도 증가 (e.g., 면적이 넓을수록 공사비 증가)
 - 음의 상관 (우하향): X가 증가할수록 Y는 감소 (e.g., 단열 성능이 높을수록 난방비 감소)
 - 무상관 (흩어짐): 관계없음
- 건축 예시:
 - 총예산 과 집행예산 간의 관계



[예시] 프롬프트

1. 기본 프롬프트 (산점도): "프로젝트_현황.csv 데이터에서 '총예산(억)'(X축)과 '집행예산(억)'(Y축) 간의 관계를 산점도로 그려줘. X축과 Y축의 범위를 동일하게 설정하고, 아울러, 산점도에 **회귀선(추세선)**을 추가해서 전체적인 경향을 보여줘."

〈데이터 : 프로젝트_현황.csv〉

데이터 시각화

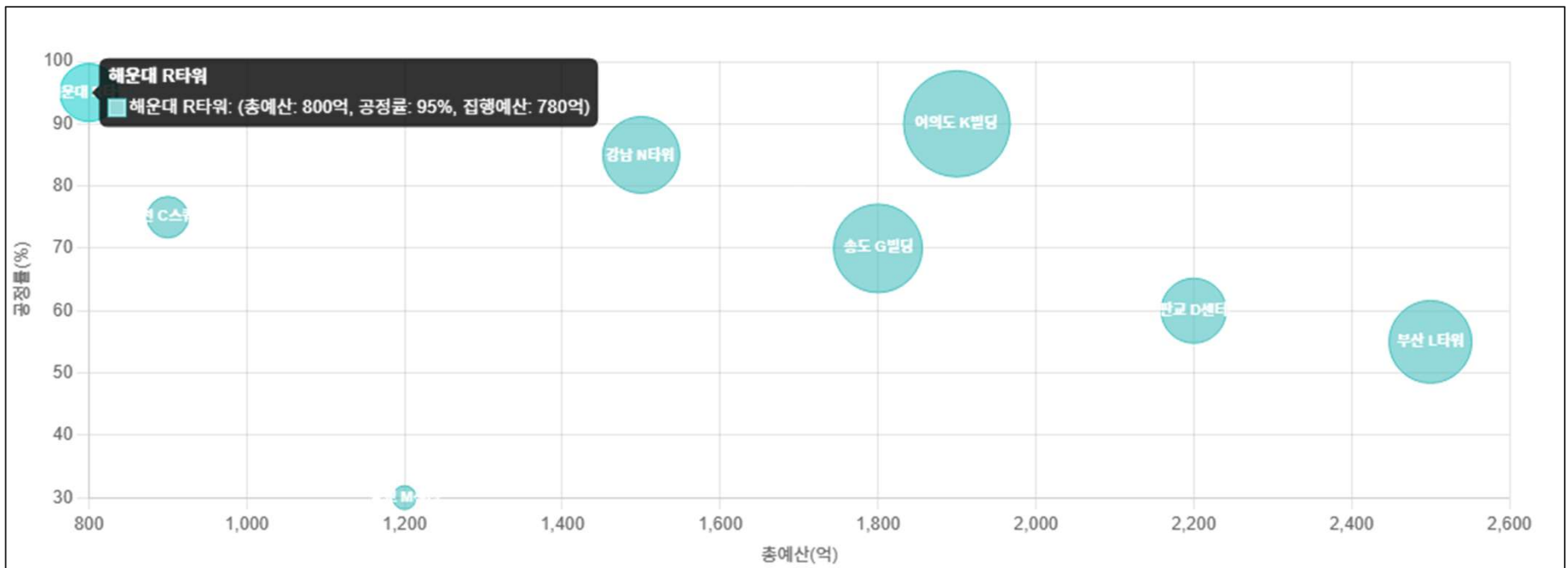
12 시각화 차트 (5): 버블 차트 (Bubble Chart)

- 목적: : 3차원 관계 파악(산점도 확장)
- 필요 데이터:
 - X축: 1개의 양적(수치형) 데이터 (변수 1)
 - Y축: 1개의 양적(수치형) 데이터 (변수 2)
 - 버블 크기: 1개의 양적 데이터 (변수 3, 규모/중요도)
- 건축 예시:
 - X축: '총면적', Y축: '공사비', 버블 크기: '공사 기간(일)'(프로젝트 포트폴리오 분석에 유용)

[예시] 프롬프트

1. 버블 차트 작성 : "'총예산'(X축)과 '공정률(%)'(Y축)의 관계를 버블 차트로 그려줘. 버블의 크기는 '집행예산'의 크기를 나타내도록 해줘. 각 버블에 '프로젝트명' 라벨을 작게 표시해 줘."

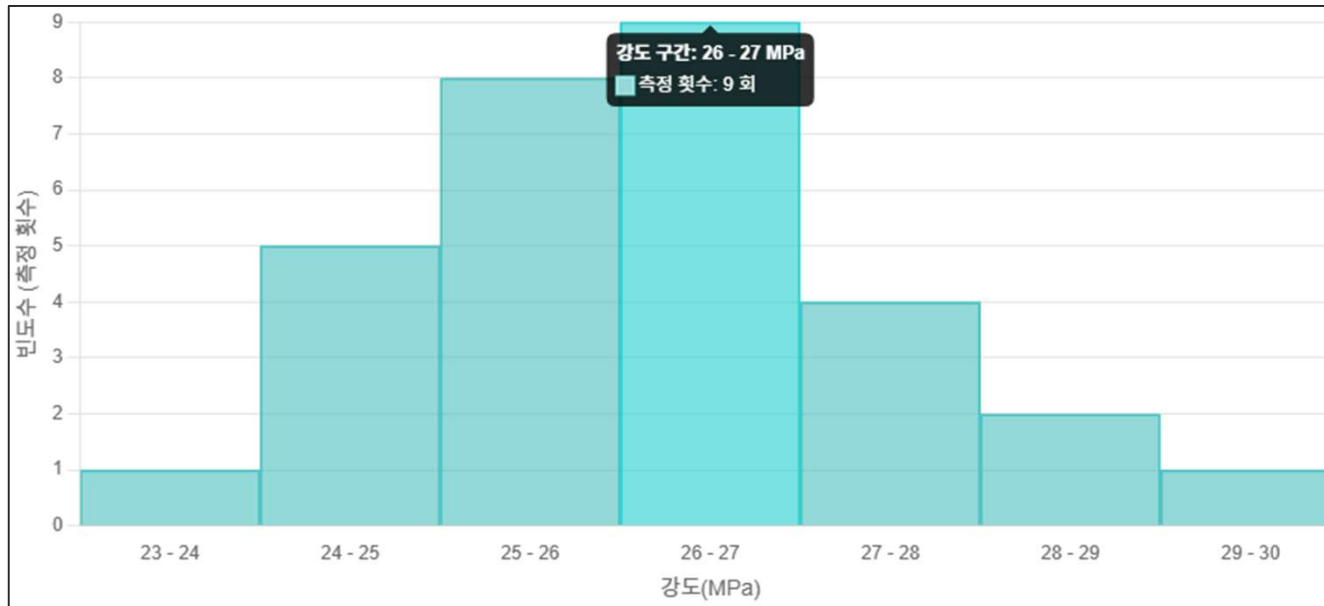
<데이터 : 프로젝트_현황.csv >



데이터 시각화

13 시각화 차트 (6): 히스토그램 (Histogram)

- 목적: 분포 (데이터가 어떤 값에 얼마나 몰려있는지 확인)
- 필요 데이터: 1개의 양적(수치형) 데이터
- 해석:
 - X축: 데이터의 '구간(Bin)' (e.g., 10~20점, 20~30점)
 - Y축: 해당 구간에 속하는 데이터의 '개수(빈도)'
(주의: 막대그래프와 다름! X축이 범주가 아님)
- 건축 예시:
 - '콘크리트 압축 강도 시험 결과'의 분포 (e.g., 24MPa 근처에 몰려 있는지?)



[예시] 프롬프트

콘크리트 강도 측정값 데이터(콘크리트_강도_측정값.csv)에 대한 히스토그램을 그려줘

<데이터 : 콘크리트_강도_측정값.csv>

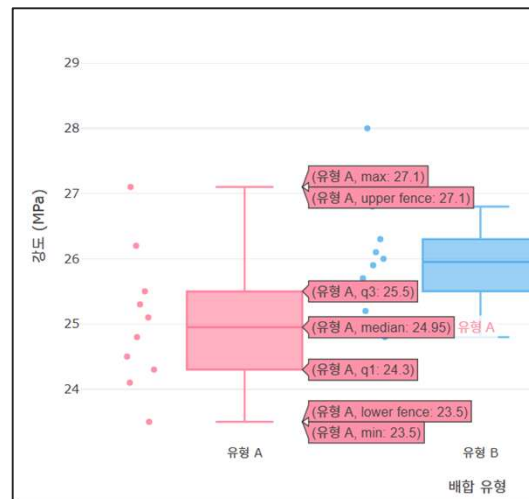
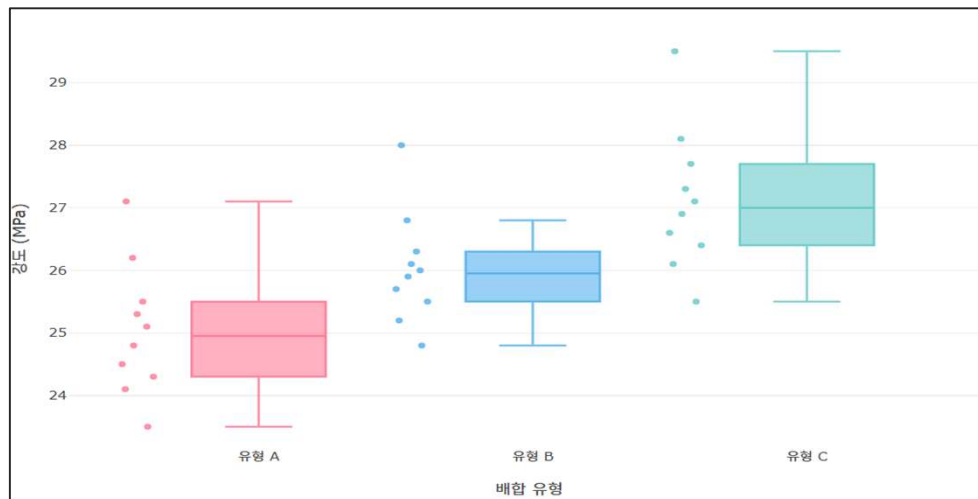
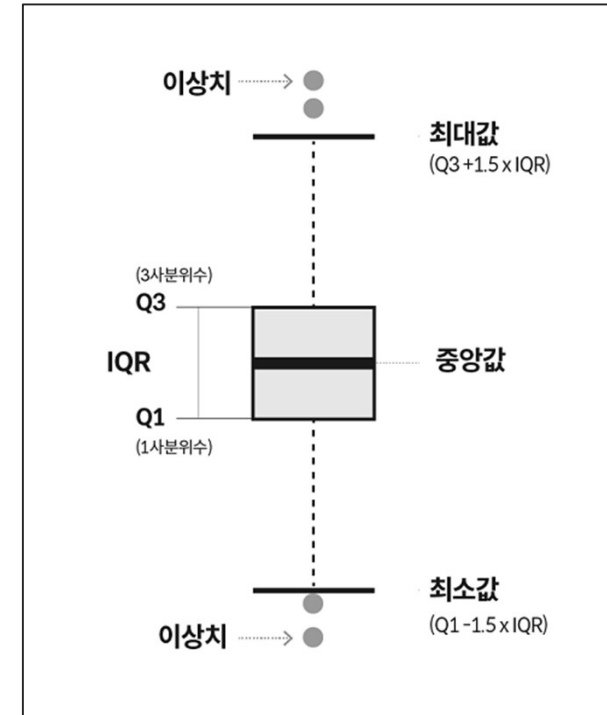
[예시] Data 명 :
콘크리트_강도_측정값.csv

배합유형,강도(MPa)
A,23.5
A,24.1
A,24.3
A,24.5
A,24.8
A,25.1
A,25.3
A,25.5
A,26.2
A,27.1
B,24.8
B,25.2
B,25.5
B,25.7
B,25.9
B,26.0
B,26.1
B,26.3
B,26.8
B,28.0
C,25.5
C,26.1
C,26.4
C,26.6
C,26.9
C,27.1
C,27.3
C,27.7
C,28.1
C,29.5

데이터 시각화

14 시각화 차트 (7): 박스 플롯 (Box Plot)

- 목적: 분포 요약 및 비교 (데이터의 5가지 요약 수치 + 이상치 탐지)
- 필요 데이터:
 - 단일: 1개의 양적 데이터
 - 비교: 1개의 양적 데이터 + 1개의 질적 데이터(그룹: 그룹내 여러 유형조)
- 해석 (5가지 요약):
 - 최솟값, 1사분위수(25%), 중앙값(50%), 3사분위수(75%), 최댓값
 - 이상치 (Outlier): 박스 범위를 벗어난 점 (특별 관리 대상)
- 건축 예시:
 - 'A, B, C 배합 유형'별 '콘크리트 강도' 분포 비교 (어느 공법이 더 안정적인가?)



[예시] 프롬프트

A, B, C 배합 유형별 '콘크리트 강도' 분포 비교하는 박스 플롯을 그려줘

<데이터 : 콘크리트_강도_측정값.csv>

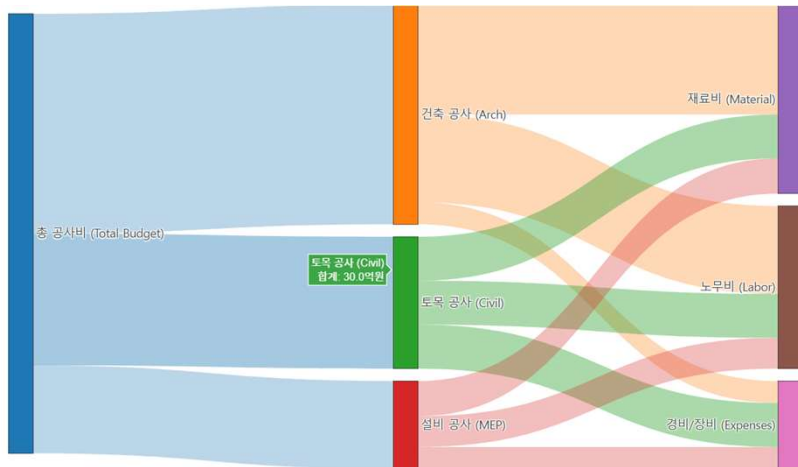
데이터 시각화

15 시각화 차트 (8): Sankey 다이어그램

- 목적: 흐름의 경로와 양적 비율 시각화
 - 시스템 내에서 자원, 에너지, 비용 등의 이동 경로 및 손실 파악
- 필요 데이터:
 - 필수 요소: Source (출발 노드), Target (도착 노드), Value (흐름의 양)
- 해석 (흐름의 직관적 이해):
 - 선의 너비(Width): 흐름의 양(Value)에 비례함 (선이 두꺼울수록 이동량이 많거나 중요도가 높음)
 - 경로 추적: 시작점부터 끝점까지 데이터가 어떻게 분배되고 합쳐지는지 파악
 - 병목/손실: 흐름이 급격히 줄어들거나(손실), 특정 구간에 과도하게 몰리는(병목) 지점 발견
- 건축 예시:
 - '전체 공사비 예산'이 '공종(건축/토목)'을 거쳐 최종적으로 '비목(재료비/노무비)'으로 어떻게 배분되어 흘러가는가?

[예시] Data 명 :
공사비 예산.csv

```
Source,Target,Value
총 공사비,건축 공사,50
총 공사비,토목 공사,30
총 공사비,설비 공사,20
건축 공사,재료비,25
건축 공사,노무비,20
건축 공사,경비/장비,5
토목 공사,재료비,10
토목 공사,노무비,10
토목 공사,경비/장비,10
설비 공사,재료비,8
설비 공사,노무비,7
설비 공사,경비/장비,5
```



[예시] 프롬프트

아래 데이터를 기반으로 공사비 흐름을 확인할 수 있는 Sankey 다이어그램을 작성(html로 캔버스에 바로 작성 시각화)

<데이터 : 공사비 예산.csv>

데이터 시각화

15 시각화 차트 (9): 네트워크 다이어그램

- 목적: 관계 및 구조 파악 (요소들 간의 연결성과 상호작용 시각화)

- 필요 데이터:

- 노드 (Node/Vertex): 개체 (사람, 부서, 도시, 작업 등)

- 링크 (Link/Edge): 관계 (연결선, 화살표, 거리, 친밀도 등)

- 해석 (흐름의 직관적 이해):

- 중심성 (Centrality)

- : 가장 많은 연결선을 가진 '허브(Hub)' 노드는 무엇인가? (핵심 인물/공정)

- 군집 (Cluster)

- : 서로 밀접하게 뭉쳐 있는 그룹은 어디인가? (팀워크, 연관 공종)

- 경로 (Path)

- : A에서 B로 가는 최단 경로 또는 영향력이 전파되는 과정

- 건축 예시:

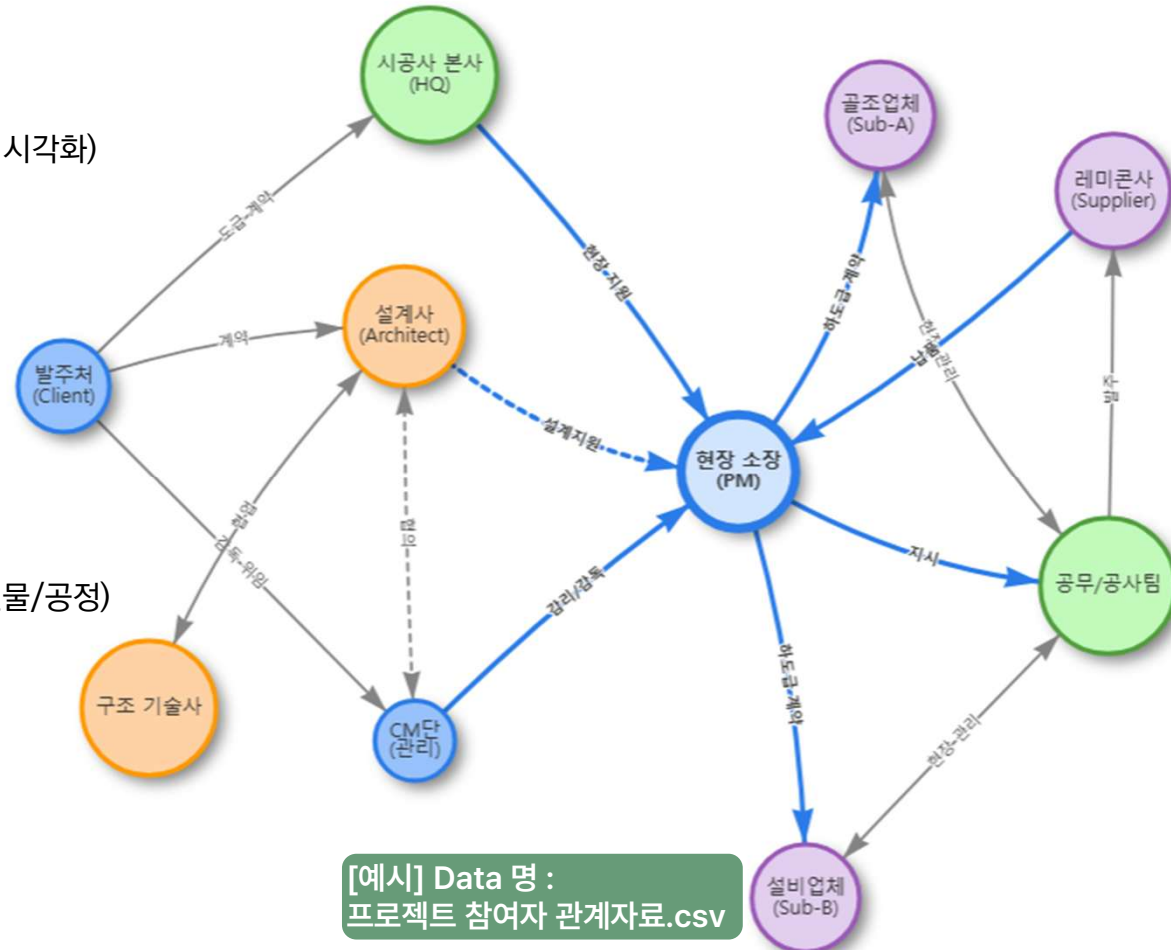
- 이해관계자 맵

- : 발주처, 설계사, 시공사, 협력업체 간의 계약 및 보고 체계 시각화

[예시] 프롬프트

아래 데이터를 기반으로 프로젝트 참여자 관계를 확인할 수 있는 네트워크 다이어그램을 작성(html로 캔버스에 바로 작성 시각화)

<데이터 : 프로젝트 참여자 관계자료.csv>



ID,이름 (Label),그룹 (Role),설명

- 1,발주처 (Client),Client,프로젝트 발주 및 자금 집행
- 2,CM단 (관리),Client,발주처 대행 건설사업관리
- 3,설계사 (Architect),Design,건축 설계 및 디자인
- 4,구조 기술사,Design,구조 안전성 검토
- 5,시공사 본사 (HQ),Construction,종합건설사 본사 지원조직
- 6,현장 소장 (PM),Construction,현장 총괄 책임자
- 7,공무/공사팀,Construction,현장 실무 관리 조직
- 8,골조업체 (Sub-A),Subcontractor,철근콘크리트 공사 수행
- 9,설비업체 (Sub-B),Subcontractor,기계설비 공사 수행
- 10,레미콘사 (Supplier),Subcontractor,콘크리트 자재 납품