

# 데이터 분석 prompt 정리

## 1. 기술적 데이터 분석

- GPT4에 데이터 파일(beer\_sales\_data.csv) 입력하기
- prompt: 당신은 데이터분석의 최고 전문가입니다. 위 데이터를 잘 읽고 제가 전달하는 데이터 분석 절차에 따라 데이터분석을 수행해 주면 되겠습니다. 준비되면 'ok'라고 답변해주세요.
- prompt: 위 csv 파일에 있는 데이터의 구조와 데이터형을 확인하고 말해주세요.
- prompt: 데이터에 결측치가 존재하는지 확인하고, 존재한다면 어느 위치에 있는지 확인해주세요.
- prompt: 변수 "날씨"의 기술통계량을 계산하고 보여주세요.
- prompt: 범주의 수가 3개이고, 가장 빈도가 높은 "맑음"이 10번이면, 다른 범주의 빈도도 10번인가요? 그렇다면 "맑음"이 가장 빈도가 높은 게 아니라, 모든 범주의 수가 10개씩 동일한 거 아닌가요?
- prompt: 변수 "맥주매출량"의 기술통계량을 계산하고 알려주세요.
- prompt: 이번에는 "날씨"에 따른 "맥주매출량"의 기술통계량을 계산하고 알려주세요.

## 2. 탐색적 데이터 분석

- prompt: "날씨"별 "맥주매출량"의 합계를 보기에 가장 적절한 그래프로 표현해 주세요.
- prompt: 그래프에서 한글이 깨져서 나오네요. 그래프에서 한글은 모두 영어로 번역한 후 표현해 주세요. 이 사항을 적용하여 다시 한번 "날씨"별 "맥주매출량"의 합계를 보기에 가장 적절한 그래프로 표현해 주세요.
- prompt: 이번에는 "날씨"별 "맥주매출량"의 분포를 박스플롯으로 표현해 주세요.
- prompt: 위 박스플롯에서 평균값을 붉은색 라인으로 표시해서 다시 그래프를 그려주세요.
- prompt: 이번에는 "날씨"별 "맥주매출량"의 분포를 바이올린플롯으로 그리고, 그래프를 해석해 주세요.

## 3. 확증적 데이터 분석

- prompt: 각 "날씨"별 "맥주매출량"의 분포를 보니 정규분포처럼 보이기도 하지만, 표본 수가 너무 적어서 확신이 서질 않아요. 이 표본을 이용하여 모수를 추정하는데 무리가 없을까요? 무리가 없다면, 통계적추론을 통해 각 범주별 맥주매출량의 모수는 어떻게 되는지 계산해 주세요.
- prompt: z-값을 이용하여 모수의 95% 신뢰구간을 계산해 주세요.
- prompt: t-값을 이용하여 모수의 95% 신뢰구간을 계산해 주세요.
- prompt: z-값과 t-값을 이용하여 구한 모수의 차이점이 있다면 설명해 주세요. 그리고 지금 사용하고 있는 데이터의 경우 어떤 검정통계량을 사용하는게 좋은지 설명해 주세요.
- prompt: (만약 GPT4의 이전 기억이 흐려진다면, 데이터를 다시 입력 하기)
- prompt: 위에 첨부한 csv 파일을 읽고, 통계적가설검정을 통해 "날씨"별 "맥주매출량"의

유의미한 차이가 있는지 확인해 주세요.

- prompt: 위에서 수행한 통계적가설검정의 절차를 구체적으로 보여주세요. 어떤 가설 검정 기법을 사용했는지, 가설수립은 어떻게 했는지, F-값은 어떻게 도출했는지 천천히 생각하면서 작성해 주세요.

#### 4. 데이터분석 결과 정리

- prompt: 지금까지 수행했던 데이터분석을 토대로 결과를 정리해 보아요. 제가 질문했던 순서대로 prompt, 결과(내용 또는 그래프), 분석 인사이트를 묶어서 천천히 생각하고 정리해서 제게 보여주세요.

#### 5. 마케팅 전략 수립

- prompt: 다음 내용은 "날씨에 따른 맥주 매출량의 유의미한 차이"에 대한 데이터 분석 결과예요. 이 분석 결과를 토대로 마케팅 부서에서 맥주 매출량을 늘리기 위한 마케팅 전략을 적어도 3가지 이상 제시해 주면 좋겠어요. 최대한 데이터 분석 결과에 근거해서 도출하고 천천히 생각하면서 자세히 작성해 주세요.

[데이터분석 결과 요약본]