

[DIG 실전 프로젝트] Project: Truth Finder

부제: 신제품 '제로쿠키' 런칭 성과 진실 규명 감사(Audit)

1. 프로젝트 개요 (Overview)

지난달 런칭한 신제품 '제로쿠키'에 대해 마케팅팀은 "고객 평점 4.9 점의 대박 상품"이라고 보고했습니다. 하지만 재무팀의 데이터는 **반품 급증과 매출 하락을** 가리키고 있습니다.

경영진은 이 모순된 상황을 해결하기 위해 당신을 특별 감사관(Auditor)으로 임명했습니다. 당신의 목표는 마케팅팀이 제출한 원본 데이터를 직접 분석하여 진실을 밝혀내고 경영 전략을 제언하는 것입니다.

[Phase 1] 증거물 확보 (Data Preprocessing)

"오염된 데이터를 정제하고, 기초 통계표를 통해 무결성을 입증하십시오."

마케팅팀으로부터 설문조사 원본 파일(Raw_Feedback.csv)을 확보했습니다. 하지만 현재 상태로는 통계 분석 시스템에 데이터를 업로드할 수 없습니다. 데이터가 오염되어 있고, 보안 규정에 위배되는 요소가 포함되어 있기 때문입니다.

생성형 AI를 활용하여, 아래 [결과물 표준]을 완벽하게 충족하는 정제된 데이터 파일(Cleaned Data)을 만드십시오.

⚠️ 결과물 표준 (Output Standard)

당신이 제출할 데이터는 다음 두 가지 조건을 반드시 만족해야 합니다.

Condition 1. 보안 및 컴플라이언스 준수 (AES)

- 기준:** 본 프로젝트는 외부 분석 툴(AI)을 사용하므로, **개인정보보호법(PII)** 및 개인정보 보안을 엄격히 준수해야 합니다.
- 요구사항:** 데이터 내에 개인을 식별할 수 있는 민감 정보가 포함되어 있는지 확인하고, 발견 즉시 해당 정보가 AI 서버에 그대로 올라가지 않게 처리하십시오. 식별 정보가 하나라도 남아있을 경우 보안 위반으로 감점으로 처리됩니다.

Condition 2. 데이터 무결성 확보 (Performance)

- 기준:** 평균(Mean), 합계(Sum), 추세(Trend) 분석을 수행할 예정입니다.
- 요구사항:**

1. 모든 정량적 지표(만족도 등)는 **산술 연산이 가능한 형태**여야 합니다. 모든 데이터는 숫자 형태로 통일하십시오.
2. 분석의 신뢰도를 떨어뜨리는 결측값이나 **불완전한 데이터**가 있다면 분석 대상에서 제외하십시오.

3. 검증 리포트 작성

데이터 정제가 완료되었다면 [기초 통계 요약표]를 출력하여 데이터가 분석 가능한 상태인지 즉시 검증해야 합니다. 이 표는 당신이 전처리를 성공적으로 수행했음을 증명하는 증거자료입니다.

- "전처리가 끝난 데이터가 시각화에 적합한지 검증하기 위해, 아래 프롬프트를 복사하여 AI에게 입력하세요."
- 입력 프롬프트

Role 너는 꼼꼼한 데이터 감사관이야.

Task 내가 전처리를 완료한 데이터셋('Cleaned_Data.csv')이 시각화 분석에 적합한지 검증하려고 해. 현재 데이터의 상태를 파악할 수 있도록 [기초 통계 요약 리포트]를 표(Table) 형태로 작성해 줘.

Requirements (필수 포함 항목)

1. [기초 통계량]: 'Score'(만족도) 컬럼의 데이터 개수(Count), 평균(Mean), 중앙값(Median), 최솟값(Min), 최댓값(Max), 결측치 개수를 계산해 줘. (※ 목적 : 점수가 숫자로 변환되었는지, 1~5 점 범위를 벗어난 이상치는 없는지, 결측치가 0 개인지 확인)
2. [인구 통계 분포]: 'Gender'(성별)과 'Age_Group'(연령대)별 응답자 수와 비율을 보여줘. (※ 데이터가 특정 성별/연령대에 쓸려있는지 확인)
3. [논리적 검증]: 위 결과를 바탕으로, 데이터가 전 연령층에 고르게 분포되었는지 아니면 편향되었는지 한 줄로 평가해줘

4. 제출물

다음 두 가지 파일을 DIG_Phase1_학번_이름.zip 으로 압축하여 제출하십시오.

1. **전처리 완료 데이터 (Cleaned_Data.csv)**: 위 표준에 맞춰 정제된 CSV 파일

2. **감사 로그 (Prompt Log)**: AI에게 어떤 문제를 발견하고 어떻게 처리를 지시했는지 보여주는 대화 내역 전문.
3. **기초 통계 요약표(감사 요약서)** : AI가 출력한 [기초 통계 요약표] -> 필수 포함: Score 통계표(무결성 증명), 성별/연령 분포표(편향성 증명)

4. 평가 기준

| 구분 | 평가 요소 | 배점 | 상세 기준 (Checklist) |
|-----|-------------------------------|------------------|--|
| ACI | ①-1. 보안/법률 처리 (Compliance) | 10 점 | <ul style="list-style-type: none"> - [PII 처리] 이름, 전화번호 등 민감 정보를 식별하고 비식별화(마스킹) 또는 삭제를 지시했는가? - [Keyword] 프롬프트에 마스킹, 가명처리, 삭제 포함 여부 |
| ACI | ①-2. 데이터 클리닝 (Cleaning) | 30 점 | <ul style="list-style-type: none"> - [결함 제거] 결측치(Null)와 형식 오류(문자형 숫자)를 찾아내어 처리했는가? - [Action] 코드 실행 로그에 dropna,fillna, replace, astype 등이 확인되는가? |
| ACI | ①-3. 기초 검증 (Verification) | 30 점 | <ul style="list-style-type: none"> - [상태 확인] 전처리 직후 데이터가 깨끗한지 확인하는 과정을 거쳤는가? - [Log] df.info(), df.describe(), df.head() 등을 출력하여 결과를 확인했는가? |
| AES | ① [보안] PII 마스킹 | (위반 시 감점) -30 | <ul style="list-style-type: none"> - 결과물 파일에 식별 가능한 개인정보가 1 건이라도 남아있으면 감점 |
| AES | ② [전략] 편향 탐지 | (위반 시 감점) -20 | <ul style="list-style-type: none"> - 검증 단계에서 데이터의 인구통계학적 쏠림(Bias) 현상을 식별하지 못하면 감점 |

※ Phase 1을 성공적으로 수행한 감사관에게만, 다음 단계인 [Phase 2. 시작화 및 검증]을 위한 접속 권한이 부여됩니다.

[DIG 실전 프로젝트]

[Phase 2] 데이터 분석 및 통계표 작성 (Data Analysis)

"분석된 통계표를 통해 마케팅팀 주장의 허구성을 숫자로 확정 지으십시오."

1. 현황 (Context)

[Phase 1]에서 정제된 데이터(Cleaned_Data.csv)를 확보했습니다. 이제 본격적인 시각화(Chart)를 그리기 전, 마케팅팀의 주장("전 국민 인기", "평점 4.9 지속")이 사실인지 통계적 수치(Table)로 먼저 검증해야 합니다.

차트는 숫자를 그림으로 바꾼 것에 불과합니다. 따라서 정확한 집계표(Aggregation Table)가 없다면 올바른 시각화도 불가능합니다.

생성형 AI를 활용하여, 아래 [결과물 표준]을 완벽하게 충족하는 [핵심 분석 통계표] 2종을 만드십시오.

2. 결과물 표준 (Output Standard)

당신이 생성할 통계표는 다음 두 가지 조건을 반드시 만족해야 합니다.

Condition 1. 집계 프롬프트 설계 (Aggregation Prompt)

- 기준: 단순 나열이 아닌, 분석 목적에 맞게 데이터를 그룹화(Grouping)하고 요약(Summary)하는 프롬프트를 직접 설계해야 합니다.
- 요구사항:
 - 단순 빈도뿐만 아니라 '비율(%)'이나 '평균(Mean)'과 같은 통계적 요약 수치를 포함하도록 지시하십시오.
 - 데이터 분석 전용 함수인 groupby, pivot_table, value_counts 등의 키워드를 프롬프트에 포함하여 AI가 명확하게 연산하도록 유도하십시오.

Condition 2. 테이블 완성도 (Table Quality)

- 기준: 생성된 표는 행(Index)과 열(Column)이 논리적으로 구성되어야 하며, 환각(Hallucination) 없는 정확한 수치를 담아야 합니다.
- 요구사항:
 - 교차 분석표: 성별 × 연령대 구조로 구성되어야 합니다.
 - 시계열 집계표: 날짜 순으로 정렬되어야 하며, 점수 변화 추세를 읽을 수 있어야 합니다.

3. 수행 미션 (Mission): 분석 테이블 요구사항 정의

아래의 [분석 요건]을 읽고, AI가 해당 형태의 표를 출력하도록 최적의 프롬프트를 직접 작성하여 입력하십시오.

Mission A. 인구통계 교차 분석표 (Cross-Tabulation)

- 목적: 마케팅팀의 "전 연령층 인기" 주장을 검증하기 위해, 특정 계층 쓸림 현상을 확인해야 함.
- [테이블 구성 요건]
 - 형태: 교차 분석표 (Pivot Table 권장)
 - 행 (Row/Index): 성별 (Gender)
 - 열 (Column): 연령대 (Age_Group)
 - 값 (Values): 각 그룹의 응답자 수(Count) 및 전체 대비 비율(%)
 - Tip: 프롬프트에 '성별을 행으로, 연령대를 열로 배치해줘', '비율을 계산해줘' 등의 지시가 포함되어야 합니다.

Mission B. 만족도 추세 집계표 (Trend Aggregation)

- 목적: 마케팅팀의 "평점 4.9 점 지속" 주장을 검증하기 위해, 시간 흐름에 따른 점수 변화를 확인해야 함.
- [테이블 구성 요건]
 - 형태: 집계표 (Groupby Table 권장)
 - 기준 (Group): 날짜 (Date) - 일자별(Daily) 기준
 - 값 (Values): 만족도 점수(Score)의 평균(Mean) (소수점 2 자리까지 표기)
 - 정렬: 날짜 순서대로 오름차순 정렬 필수
 - Tip: 프롬프트에 '날짜를 기준으로 묶어서', '점수의 평균을 구해줘' 등의 지시가 포함되어야 합니다.

4. 제출물

다음 두 가지 파일을 DIG_Phase2_학번_이름.zip 으로 압축하여 제출하십시오.

- 통계표 결과 캡처 (Analysis_Tables.jpg): AI가 생성한 [교차 분석표]와 [추세
집계표]가 모두 포함된 화면 캡처 이미지. (※ 예시 이미지와 같이 행/열 구조가
갖춰져야 함)
 - 분석 로그 (Prompt Log): 위 테이블을 만들기 위해 본인이 직접 작성한
프롬프트와 AI의 답변 전문.
-

5. 평가 기준

| 구분 | 평가 요소 | 배점 | 상세 기준 (Checklist) |
|------|---|------|---|
| 성과도출 | ②-1. 집계 프롬프트 설계 (Aggregation Prompt) | 30 점 | <ul style="list-style-type: none"> - [지시 명확성] 5 점 척도 만족도, 성별 교차 분석 등 원하는 형태의 통합 테이블을 만들도록 정확히 지시했는가? - [Keyword] 프롬프트에 성별 별, 항목 별, 평균, 비율 등의 집계 키워드와 groupby, pivot_table 함수 유도 내용이 포함되었는가? |
| 성과도출 | ②-2. 테이블 완성도 (Table Quality) | 30 점 | <ul style="list-style-type: none"> - [구조 적합성] 생성된 표가 분석 목적에 부합하는 행/열 구조와 수치를 담고 있는가? (환각 없는 데이터) - [검증] 생성된 객체(df_result)의 Index(성별 등)와 Column(만족도 등)이 지시대로 구성되었는가? |

※ Phase 2 를 성공적으로 수행한 감사관에게만, 다음 단계인 [Phase 3. 시각화 및 인사이트]를 위한 접속 권한이 부여됩니다.

[DIG 실전 프로젝트]

[Phase 3] 시각화 및 증거 확보 (Visualization)

"숫자를 시각적 증거(Chart)로 변환하여, 숨길 수 없는 진실을 경영진 앞에 드러내십시오."

1. 현황 (Context)

[Phase 2]에서 생성한 통계표를 통해 우리는 데이터의 편향(Bias)과 실적 하락(Trend)을 숫자로 확인했습니다.

하지만 텍스트로 된 표는 경영진에게 직관적인 충격을 주기 어렵습니다. 이제 이 숫자를 가장 강력한 시각적 증거(Chart)로 변환하여 보고서에 실어야 합니다.

canvas 를 활용하여, 아래 [결과물 표준]을 완벽하게 충족하는 [핵심 증거 차트] 2 종을 생성하십시오.

2. 결과물 표준 (Output Standard)

당신이 생성할 차트는 다음 두 가지 조건을 반드시 만족해야 합니다.

Condition 1. 차트 적합성 (Chart Fitness)

- 기준: 데이터의 성격(비중, 추세, 비교)에 가장 적합한 차트 유형을 선택해야 합니다.
- 요구사항:
 - 비중(Ratio) 분석: 전체 대비 특정 그룹의 비율을 보여줄 때는 Pie Chart 또는 Donut Chart 를 사용하십시오. (※ Line Chart 사용 시 감점)
 - 추세(Trend) 분석: 시간의 흐름에 따른 변화를 보여줄 때는 Line Chart 또는 Area Chart 를 사용하십시오. (※ Pie Chart 사용 시 감점)

Condition 2. 시각화 완성도 (Visualization Detail)

- 기준: 차트만 덩그러니 있는 그림은 보고서에 쓸 수 없습니다. 가독성을 높이는 요소를 챙겨야 합니다.
- 요구사항:
 - Title: 차트의 주제를 명확히 적으십시오. (예: "성별/연령대별 고객 분포", "일별 만족도 변화 추이")

- Label & Legend: X 축/Y 축이 무엇인지(Label)와 색깔이 무엇을 의미하는지(Legend) 표시하십시오.
 - Detail: 파이 차트에는 비율(%)이, 꺾은선 그래프에는 데이터 포인트(Marker)가 잘 보여야 합니다.
-

3. 수행 미션 (Mission): 시각화 요구사항 정의

[Phase 2]에서 만든 표(df_demographics, df_trend)를 바탕으로, AI 가 아래 차트를 그리도록 최적의 프롬프트를 직접 작성하여 입력하십시오.

Mission A. 고객층 편향 시각화 (Bias Visualization)

- 목적: "전 국민 인기"라는 주장이 거짓임을 증명하기 위해, 20 대 여성에게 데이터가 쓸려 있음을 한눈에 보여줘야 함.
- [시각화 요건]
 - 입력 데이터: 인구통계 교차 분석표 (Phase 2 결과물)
 - 차트 유형: Pie Chart (또는 Donut Chart)
 - 필수 데코레이션:
 - 각 조각(Slice)에 퍼센트(autopct) 표시.
 - 차트 제목: "고객 인구 통계 분포"
 - Tip: 프롬프트에 '비중을 보고 싶으니 파이 차트로 그려줘', '비율(%)을 표시해줘' 등의 지시가 포함되어야 합니다.

Mission B. 만족도 추세 시각화 (Trend Visualization)

- 목적: "평점 4.9 점 유지"라는 주장이 거짓임을 증명하기 위해, 날짜가 갈수록 점수가 떨어지는 하락세를 보여줘야 함.
- [시각화 요건]
 - 입력 데이터: 만족도 추세 집계표 (Phase 2 결과물)
 - 차트 유형: Line Chart
 - 필수 데코레이션:
 - X 축: 날짜(Date), Y 축: 만족도(Score) 라벨링.

- 데이터 포인트에 마커(Marker) 표시해 변곡점을 강조할 것
 - 차트 제목: "일별 만족도 변화 추이"
 - Tip: 프롬프트에 '흐름을 봐야 하니 꺾은선 그래프로 그려줘', '축 제목을 달아줘' 등의 지시가 포함되어야 합니다.*
-

4. 제출물

다음 두 가지 파일을 DIG_Phase3_학번_이름.zip 으로 압축하여 제출하십시오.

- 시각화 결과 캡처 (Visualization_Charts.jpg): AI가 생성한 [인구통계 파일 차트]와 [만족도 라인 차트] 이미지.
 - 시각화 로그 (Prompt Log): 차트의 종류와 디테일(제목, 라벨 등)을 설정하기 위해 작성한 프롬프트 전문.
-

5. 평가 기준

| 구분 | 평가 요소 | 배점 | 상세 기준 (Checklist) |
|------|--|------|---|
| 성과도출 | ③-1. 차트 적합성 (Chart Fitness) | 30 점 | <ul style="list-style-type: none"> - [유형 선택] 데이터 성격에 맞는 차트를 지시했는가? <ul style="list-style-type: none"> 비중(Ratio) → Pie/Donut (O) 추세(Trend) → Line/Area (O) (※ 엉뚱한 차트 선택 시 감점) |
| 성과도출 | ③-2. 시각화 완성도 (Visualization Detail) | 30 점 | <ul style="list-style-type: none"> - [가독성] 차트의 정보를 읽을 수 있도록 요소를 챙겼는가? <ul style="list-style-type: none"> [Parameter] 코드 내 title, xlabel, ylabel, legend, autopct(비율) 설정 여부 확인 |

※ Phase 3를 성공적으로 수행한 감사관에게만, 마지막 단계인 [Phase 4. 경영 제언]을 위한 접속 권한이 부여됩니다.

[DIG 실전 프로젝트]

[Phase 4] 경영 제언: 위기 탈출 전략 수립

"완성된 차트가 가리키는 진실을 읽고, 회사를 구할 결단을 내리십시오."

1. 현황 및 목표 (Context)

Phase 3를 통해 우리는 마케팅팀의 주장을 검증하는 '고객 반응 시각화 대시보드'를 완성했습니다. 하지만 개별적으로 작성된 대시보드에 분석 오류가 있을 경우, 회사의 운명을 건 전략이 잘못된 방향으로 갈 수 있습니다. 이에 본부 감사팀은 정말 검증된 [공식 시각화 데이터]를 확정 배포했습니다.

이제 확정된 데이터를 바탕으로 '왜(Why)' 이런 위기가 왔는지 규명하고, '무엇(What)'을 해야 할지 결정해야 합니다.

- 당신의 역할: 전략기획팀장
- 당신의 임무: 사장님께 보고할 [위기 대응 경영 제언서]를 AI와 함께 작성하십시오.

2. 분석 자료 (Input Data)

강사가 배포한 [공식 시각화 데이터(Final_Chart_Report.pdf)]를 다운로드하여 확인하십시오.

이 자료에는 '인구통계 편향(Pie Chart)', '만족도 급락 추세(Line Chart)', '부정 키워드 분석(Word Cloud)' 결과가 포함되어 있습니다.

3. 수행 과제 (Missions)

위 분석 자료를 토대로, 다음 두 가지 핵심 과제를 수행하여 보고서를 완성하십시오.

Mission 4-1. 원인 규명 및 전략 수립

차트와 키워드 데이터에서 만족도 하락의 핵심 원인(Root Cause)을 찾으십시오. 그리고 이를 해결하기 위한 비즈니스 전략을 AI에게 작성시키십시오.

[보고서 필수 구조 (3-Step Logic)]

AI가 작성한 보고서는 반드시 아래 3 단 논리 구조를 갖춰야 합니다.

- Fact (진단):

- 추상적인 표현은 금지합니다. 차트에 나타난 구체적인 수치(최댓값, 최솟값, 급락한 날짜, 편향된 비율 %)를 정확히 인용하여 현재의 위기 상황을 진단하십시오.

2. Reason (원인):

- 데이터(부정 키워드 등)에서 발견된 문제의 구체적 원인을 명시하십시오.
(예: 배송, 맛, 가격 등)

3. Proposal (제언):

- 원인을 제거하고 실적을 반등시킬 즉각적인 개선 액션 플랜을 제시하십시오.

Mission 4-2. 경영 제언 : 고객 신뢰 회복

기술적인 해결책만으로는 떠나간 고객을 되돌릴 수 없습니다. 사장님께 보고할 최종 제언서에는 무너진 신뢰를 회복할 경영적 판단이 포함되어야 합니다.

[요구사항]

Mission 4-1 의 결과에 더해, 고객 케어 및 리스크 관리 방안을 추가하여 최종 제언서를 완성하십시오.

[필수 포함 요소]

1. 피해를 입은 기존 고객에 대한 구체적인 보상안 (환불, 재발송, 쿠폰 등)
2. 브랜드 이미지 회복을 위한 진정성 있는 사과 및 소통 계획

(Tip: 단순 변명이 아닌, 책임 있는 기업의 태도(CSR)를 보여주는 것이 이번 보고의 핵심입니다.)

4. 제출물 (Deliverables)

다음 내용이 포함된 DIG_Phase4_학번_이름.pdf 를 제출하십시오.

1. [최종 보고서] 경영 제언 초안: 보고서 필수 구조(Fact-Reason-Proposal)를 갖춘 1페이지 전략 기획서.
 2. [대화 로그] AI 협업 기록: 원인 분석 및 제언을 도출하기 위해 AI와 나눈 프롬프트 대화 내역 전문.
-

5. 평가 기준 (Evaluation Criteria)

| 구분 | 평가 요소 | 배점 | 상세 기준 (Checklist) |
|------|---|------|--|
| 성과도출 | ④-1. 사실 기반 해석 (Fact Finding) | 30 점 | <ul style="list-style-type: none"> - [수치 인용] 그래프가 보여주는 수치적 사실(최댓값, 급락 날짜, 하락폭 등)을 정확히 언급했는가? - [일치 여부] 리포트 내 숫자가 제공된 공식 데이터와 일치하는가? |
| 성과도출 | ④-2. 논리적 제언 (Logical Proposal) | 30 점 | <ul style="list-style-type: none"> - [논리 구조] 현상(Fact) → 원인(Reason) → 제언(Proposal)의 3 단 흐름이 명확한가? - [키워드] 문장에 따라서, 전략, 개선, 필요 등의 단어를 사용하여 구체적 실행 방안을 제안했는가? |