

[DIG 실전 프로젝트]

부제: 핵심 인재 대이탈 원인 규명 및 방어 전략 수립

1. 프로젝트 개요

올해 상반기, 우리 회사는 업계 최고 수준의 연봉 인상을 단행했습니다. 인사팀은 "연봉 인상으로 직원 만족도가 4.8점(5점 만점)을 기록했으며 퇴사율이 급감할 것"이라고 경영진에 보고했습니다.

하지만 현장의 상황은 다릅니다. 핵심 부서인 R&D팀의 에이스들이 줄줄이 사표를 던지고 있습니다.

CEO는 이 기이한 현상(연봉은 올랐는데 직원은 떠나는)을 해결하기 위해 당신을 [조직문화 데이터 감사관]으로 임명했습니다.

당신의 목표는 인사팀이 숨기고 있는(혹은 보지 못한) 원본 데이터를 분석하여 인재 유출의 진짜 원인을 밝히고, 조직을 구할 전략을 제언하는 것입니다.

2. 증거물 확보

"인사 데이터는 그 어떤 데이터보다 민감하다. 안전하게 복원하라."

인사팀으로부터 퇴사자 및 재직자 설문조사 원본 파일(Raw_HR_Survey.csv)을 확보했습니다. 하지만 현재 상태로는 통계분석 시스템에 데이터를 업로드 할 수 없습니다. 데이터 내에 민감한 개인정보가 포함되어 있고, 분석을 방해하는 오염된 데이터가 섞여있기 때문입니다.

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

⚠ 결과물 표준

당신이 제출할 데이터는 다음 두 가지 조건을 반드시 만족해야 합니다.

Condition 1. 보안 및 컴플라이언스 준수

- 기준:** 인사 데이터에는 직원의 주민번호, 연봉 계약서 등 절대 유출되면 안 되는 PII(개인식별정보)가 포함되어 있습니다.
- 요구사항:** 데이터 내에 사번(Employee_ID)을 제외한 이름, 주민등록번호, 개인 휴대전화 번호가 포함되어 있는지 확인하고, 발견 즉시 마스킹 혹은 삭제(Drop) 처리하십시오. 식별 정보가 하나라도 남아있을 경우 보안 사고로 간주하여 0점 처리하십시오.

리됩니다.

Condition 2. 데이터 무결성 확보

- **기준:** 연봉과 만족도, 퇴사 여부에 대한 평균, 합계, 추세 분석을 수행할 예정입니다.
- **요구사항:**
 - **형식 통일:** '5,000만원', '\$50k', '50000000' 등으로 제각각인 연봉 데이터를 원화(KRW) 기준의 산술 연산이 가능한 순수 숫자형으로 통일하십시오.
 - **결측치 처리:** 만족도 점수가 비어있는(Null) 퇴사자는 분석에서 제외(Drop)하고, 부서명(Dept)이 누락된 경우 직원을 삭제하면 퇴사를 통계가 왜곡되니, Unknown 값으로 채워 데이터를 보존하십시오.

3. 검증 리포트 작성

데이터 정제가 완료되었다면 [기초 인사 통계 요약표]를 출력하여 인사팀 보고의 진위를 1차 판별하십시오.

[입력 프롬프트 예시]

```
# Role 너는 꼼꼼한 인사 데이터 감사관이야.
```

```
# Task 전처리가 완료된 Cleaned_HR_Data.csv를 검증하기 위해 [기초 통계 요약 리포트]를 표형태로 작성해 줘.
```

```
# Requirements (필수 포함 항목)
```

1. **[기초 통계량]:** 'Satisfaction'(만족도)와 'Salary'(연봉) 컬럼의 데이터 개수, 평균 (Mean), 중앙값(Median), 최솟값, 최댓값, 결측치 개수를 계산해 줘. (※ 점수와 연봉이 숫자로 변환되었는지, 이상치는 없는지, 결측치가 0개인지 확인)
2. **[인구 통계 분포]:** 부서와 퇴사 여부별 응답자 수와 비율을 보여줘 (※ 데이터가 특정 부서에 쓸려 있는지, 퇴사자 데이터가 충분한지 확인)
3. **[논리적 검증]:** '위 결과를 바탕으로, 데이터가 전 부서에 고르게 분포되었는지 아니면 편향되었는지 한 줄로 평가해줘.'

4. 제출물

다음 파일들을 DIG_Phase1_학번_이름.zip으로 압축하여 제출하십시오.

1. **전처리 완료 데이터 (Cleaned_HR_Data.csv)**: 위 표준에 맞춰 정제된 CSV 파일.
2. **기초 통계 요약표 (Audit_Summary.pdf)**: AI가 출력한 요약표를 캡처하거나 복사한 파일. (* 필수포함 : 만족도/연봉 통계표, 부서/회사여부 분포표)
3. **감사 로그 (Prompt_Log)**: AI에게 어떤 문제를 발견하고 어떻게 처리를 지시했는지 보여주는 대화 내역 전문.

5. 평가 기준

평가 분류	세부 평가요소	상세 채점 가이드 (Log 확인)	배점
ACI 평가요소	①-1. 보안/법률 처리 (Compliance)	<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	10점
	①-2. 데이터 클리닝 (Cleaning)	<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?	30점
	①-3. 기초 검증 (Verification)	<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?	30점
AES 평가요소	① [보안] PII 유출	<ul style="list-style-type: none">- 결과물 파일에 식별 가능한 개인정보가 1 건이라도 남아있으면 감점	(위반시 감점)-30
	② [전략] 편향 탐지	<ul style="list-style-type: none">- 검증 단계에서 데이터의 인구통계학적 쓸림 (Bias) 현상을 식별하지 못하면 감점	(위반 시 감점)-20

(※ Phase 1을 성공적으로 수행한 감사관에게만, 다음 단계인 [Phase 2. 현장 조사]를 위한 접속 권한이 부여됩니다.)

[Phase 2] 데이터 분석 및 통계표 작성

"말뿐인 주장을 데이터로 검증하고, 시각적 증거를 확보하라."

1. 현황

[Phase 1]에서 정제된 데이터(Cleaned_HR_Data.csv)를 확보했습니다.

이제 본격적인 시각화를 그리기 전, 인사팀의 주장("연봉 인상으로 만족도 4.8점 달성, 퇴사율 감소 예상")이 사실인지 통계적 수치로 먼저 검증해야 합니다.

차트는 숫자를 그림으로 바꾼 것에 불과합니다. 따라서 '누가', '왜' 나가는지 보여주는 정확한 집계표가 없다면 올바른 시각화도 불가능합니다.

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

2. 결과물 표준

당신이 생성할 통계표는 다음 두 가지 조건을 반드시 만족해야 합니다.

Condition 1. 집계 프롬프트 설계

- **기준:** 단순히 "표 만들어줘"가 아닌, 데이터를 어떻게 묶고, 무엇을 계산할지 구체적으로 지시해야 합니다.
- **요구사항:**
 - **[파생 변수 생성]** 현재 데이터에는 '퇴사 여부' 컬럼이 명시되어 있지 않습니다. Exit_Date(퇴사일) 데이터가 있으면 '퇴사(Resigned)', 비어있으면 '재직(Employed)'으로 분류하는 새로운 기준(Column)을 만들도록 지시하십시오.
 - **[구조화 지시]** 데이터를 "부서별로 묶어서" 또는 "직급을 행으로 배치해서" 와 같이 표의 구조를 말로 풀어서 설명하십시오.
 - **[요약 수치]** 단순 개수뿐만 아니라 '평균 연봉', '평균 만족도', '구성 비율 (%)'을 계산하도록 지시하여 분석의 깊이를 더하십시오.

Condition 2. 테이블 완성도

- **기준:** 생성된 표는 행과 열이 논리적으로 구성되어야 하며, 환각 없는 정확한 수치를 담아야 합니다.
- **요구사항:**
 - **교차 분석표:** 부서(Department) × 퇴사여부(Resignation) 구조로 구성되어

- 야 하며, 구성 비율(%)이 포함되어야 합니다.
- **핵심 지표 집계표:** 직급(Position)별로 평균 연봉과 평균 만족도를 비교할 수 있어야 합니다.
-

3. 수행 미션 : 분석 테이블 요구사항 정의

결과물 표준을 달성하기 위해 아래의 [분석 요건]을 읽고, AI에게 입력할 최적의 프롬프트를 직접 작성하여 아래 두 가지 표를 출력해 내십시오.

Mission A. 부서별 인력 유출 현황

- **목적:** 인사팀은 "전사적으로 퇴사율이 줄었다"고 주장합니다. 이 주장이 맞는지, 아니면 특정 부서만 피를 흘리고 있는지 확인하기 위해 부서와 퇴사 현황을 교차 비교해야 합니다.
- **[테이블 구성 요건]**
 - **형태:** 교차 분석표 (Pivot Table 권장)
 - **행 (Row/Index):** 부서명
 - **열 (Column):** 퇴사 여부 (※ Exit_Date를 활용해 만든 파생 변수 사용)
 - **값 (Values):** 각 그룹의 인원 수 및 **부서 내 퇴사자 비율(%)**
- **Tip:** 프롬프트에 'Exit_Date가 비어있지 않으면 퇴사로 간주해줘', '각 부서별로 퇴사자 비율을 계산해줘' 등의 지시가 반드시 포함되어야 합니다.

Mission B. 부서별 연봉 대비 만족도

- **목적:** 연봉을 올렸으니 만족할 것이다라는 주장을 검증해야 합니다. 부서 별로 실제 연봉과 만족도가 비례하는지 확인하십시오.
- **[테이블 구성 요건]**
 - **형태:** 집계표 (Groupby Table 권장)
 - **기준 (Group):** 부서명(Edpt)
 - **값 (Values):**
 1. 평균 연봉 (원 단위, 천 단위 구분 기호 포함)
 2. 평균 만족도 점수 (소수점 2자리까지)

- 정렬: 평균 연봉이 높은 순서대로 내림차순 정렬
 - Tip: 프롬프트에 '직급을 기준으로 묶어서', '연봉과 만족도의 평균을 구하고 연봉 순으로 정렬해줘' 등의 지시가 포함되어야 합니다.
-

4. 제출물

다음 두 가지 파일을 DIG_실습_Phase2_학번_이름.zip으로 압축하여 제출하십시오.

1. **통계표 결과 (Analysis_Tables.jpg)**: AI가 생성한 [부서별 교차 분석표]와 [직급별 처우 집계표]가 모두 포함된 화면 캡처 이미지 및 문서
 2. **분석 로그 (Prompt Log)**: 위 테이블을 만들기 위해 본인이 직접 작성한 프롬프트 와 AI의 답변 전문.
 - 주의: '퇴사 여부' 변수를 생성하는 과정이 로그에 반드시 포함되어야 함.
-

5. 평가 기준

평가 분류	세부 평가요소	상세 채점 가이드 (Log 확인)	배점
ACI	② -1. 집계 프롬프트 설계 (Aggregation Prompt)	<ul style="list-style-type: none">- [변수 생성] Exit_Date를 활용하여 '퇴사/재직'을 구분하는 논리를 AI에게 제시했는가?- [지시 명확성] Pivot, Groupby, 비율(%) 계산 등 원하는 형태를 만들기 위한 명확한 키워드를 사용했는가?	30점
평가요소	② -2. 테이블 완성도 (Table Quality)	<ul style="list-style-type: none">- [구조 적합성] 생성된 표가 분석 목적(부서별 비교, 직급별 연봉 확인)에 부합하는가?- [검증] R&D 부서의 높은 퇴사율이나, 직급별 연봉/만족도의 괴리가 숫자로 드러나는가? (환각 없는 데이터)	30점

"숫자를 시각적 증거로 변환하여, 숨길 수 없는 진실을 경영진 앞에 드러내십시오."

1. 현황

[Phase 2]에서 생성한 통계표를 통해 우리는 "평균 연봉 1위인 R&D 부서의 퇴사율이 가장 높다"는 충격적인 사실을 숫자로 확인했습니다.

하지만 엑셀 표만으로는 경영진에게 위기감을 전달하기 어렵습니다. "고작 몇 명 나간 것 아니냐"는 안일한 반응을 깨트려야 합니다.

이제 이 숫자들을 가장 강력한 시각적 증거로 변환하여 보고서에 실어야 합니다.

생성형 AI(Canvas 기능)를 활용하여, 아래 [결과물 표준]을 완벽하게 충족하는 [핵심 증거 차트] 3 종을 생성하십시오.

2. 결과물 표준

당신이 생성할 차트는 다음 두 가지 조건을 반드시 만족해야 합니다.

Condition 1. 차트 적합성

- **기준:** 데이터의 성격(비중, 추세, 비교)에 가장 적합한 차트 유형을 선택해야 합니다.
- **요구사항:**
 - **비중 분석:** 전체 퇴사자 중 특정 부서가 차지하는 비율을 보여줄 때는 **Pie Chart** 또는 **Donut Chart**를 사용하십시오.
 - **추세 분석:** 시간의 흐름(월별)에 따른 퇴사 규모 변화를 보여줄 때는 **Line Chart**를 사용하십시오.
 - **비교/순위 분석:** 불만 키워드 빈도 등을 비교할 때는 **Bar Chart**를 사용하십시오.

Condition 2. 시각화 완성도

- **기준:** 차트만 덩그러니 있는 그림은 보고서에 쓸 수 없습니다. 가독성을 높이는 다음 3 가지 요소를 챙겨야 합니다.
- **요구사항:**
 - **제목:** 그래프가 무엇을 나타내는지 명확한 제목을 달아주세요. (예: 부서별 퇴사자 비중)

- **라벨 및 범례:** X 축/Y 축의 의미를 표시하고, 파이 차트에는 비율(%)이 숫자로 표시되어야 합니다.
 - **디테일:** 막대 그래프 끝에 건수(Count)를 표시하거나, 선 그래프에 마커(Marker)를 찍어 가독성을 높이십시오.
-

3. 수행 미션: 시각화 요구사항 정의

[Phase 2]에서 만든 통계분석표와 원본 데이터를 바탕으로, AI (Canvas 기능)에게 아래 차트를 그리도록 지시하십시오.

Mission A. 퇴사자 구성비 시각화 (Targeting)

- **목표:** 경영진은 "퇴사자가 전사적으로 발생한다"고 이해하고 있습니다. 퇴사자가 특정 부서(R&D)에 집중되어 있음을 시각적으로 증명하십시오.
- **[시각화 요건]**
 - **입력 데이터:** 부서별 퇴사자 수 (Phase 2 결과물 활용)
 - **차트 유형:** **Pie Chart** (또는 Donut Chart)
 - **필수 데코레이션:**
 - 각 부서 영역에 비율(%)을 숫자로 반드시 표시
 - 차트 제목: "부서별 퇴사자 점유율"
- **Tip:** 차트가 완성되면, R&D 부서가 전체 퇴사의 과반을 차지하는지 눈으로 확인하세요.

Mission B. 퇴사 급증 시점 포착

- **목표:** "언제부터 사람들이 떠났는가?"를 밝혀야 합니다. 퇴사일(Exit_Date) 데이터를 분석하여 월별 퇴사자 수의 변화 추이를 시각화하십시오.
- **[시각화 요건]**
 - **입력 데이터:** 전체 데이터셋 내 Exit_Date 컬럼
 - **차트 유형:** **Line Chart**
 - **필수 데코레이션:**
 - X 축: 월, Y 축: 퇴사자 수

- 퇴사자가 급증하는 달이 잘 보이게 마커 표시
- 차트 제목: "월별 퇴사 추이 분석"
- **Tip:** Exit_Date 를 '월' 단위로 묶어서 카운트하라고 지시해야 합니다.

Mission C. R&D 팀 불만 원인 규명

- **목표:** 가장 중요합니다. 도대체 왜 나가는 걸까요? R&D 부서 퇴사자들의 리뷰에서 가장 많이 등장한 **부정 키워드**를 추출하여 시각화하십시오.
- **[시각화 요건]**
 - **입력 데이터:** 전체 데이터셋 내 Review 컬럼 (단, **R&D 부서 & 퇴사자** 데이터만 필터링하여 사용)
 - **차트 유형:** **Bar Chart** (가로형/세로형 무관)
 - **필수 데코레이션:**
 - 가장 많이 언급된 **상위 5 개 단어** 추출
 - 각 막대 끝에 빈도수 표시
 - 차트 제목: "R&D 핵심 인재 퇴사 사유 TOP 5"
- **Tip:** 프롬프트에 *'R&D 부서이면서 퇴사한 사람들의 리뷰만 골라서', '가장 많이 나온 단어 5 개를 뽑아줘'*라고 명확히 지시해야 합니다.

4. 제출물

다음 시각화 파일을 DIG_실습_Phase3_학번_이름.zip 으로 제출하십시오.

1. **시각화 결과 캡처 (Visualization_Charts.jpg):** AI 가 생성한 **[부서 점유율 파이 차트]**, **[월별 퇴사 라인 차트]**, **[퇴사 사유 바 차트]** 이미지 3 종
2. **시각화 로그 (Prompt Log):** 차트의 종류와 디테일(제목, 색상 등)을 설정하기 위해 작성한 프롬프트 전문.

5. 평가 기준

평가 분류	세부 평가요소	상세 채점 가이드 (Log 확인)	배점
-------	---------	--------------------	----

ACI 평가요소	③ -1. 차트 적합성 (Chart Fitness)	<p>- [유형 선택] 데이터 성격에 맞는 차트를 지시했는가?</p> <ul style="list-style-type: none"> • 비중(Share) → Pie (O) • 추세(Time) → Line (O) • 비교(Rank) → Bar (O) 	30 점
	③ -2. 시각화 완성도 (Visualization Detail)	<p>- [가독성] 차트의 정보를 읽을 수 있도록 요소를 챙겼는가?</p> <ul style="list-style-type: none"> • [Mission C 검증] 전체 리뷰가 아닌 'R&D 퇴사자'로 대상을 좁혀서 키워드를 추출했는가? (전체 리뷰 분석 시 감점) 	30 점

[Phase 4] 경영 제언: 조직 붕괴를 막을 위기 탈출 전략

"차트와 텍스트 속에 숨겨진 '떠나는 이유'를 규명하고, R&D 조직을 구할 결단을 내리십시오."

1. 현황 및 목표

[Phase 3]를

[Phase 3]를 통해 우리는 R&D 부서가 전사 퇴사의 진원지임을 확인했고, 퇴사자들의 리뷰 데이터를 통해 그 원인이 '돈'이 아닌 '성장과 문화의 부재'임을 파악했습니다. 지금 경영진에게 필요한 것은 막연한 위로가 아니라, 떠나는 핵심 인재들의 발길을 돌릴 수 있는 구체적이고 실질적인 해결책입니다.

당신의 역할은 전략기획팀장입니다. 데이터에 기반하여 [R&D 조직 문화 혁신 및 인재 리텐션(Retention) 전략]을 수립하고, 이를 경영진에게 보고하십시오.

2. 분석 자료

당신은 앞선 단계에서 다음 세 가지 확실한 증거를 확보했습니다.

- 통계적 팩트:** R&D 부서의 평균 연봉은 전사 1 위이나, 퇴사율은 90%에 육박함 (고비용 저효율 구조).
- 시각적 증거:** 연봉 인상 시점 이후에도 퇴사 추세가 꺾이지 않음.
- 정성적 증거(Review):** "기술적 성장 정체", "노후화된 장비", "비효율적 보고 문화" 등 업무 환경 및 시스템에 대한 불만이 지배적임.

3. 수행 과제

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

Mission 4-1. 원인 규명 및 핵심 전략 수립 (3-Step Logic)

데이터에서 발견된 문제점(Fact)을 근거로 원인(Reason)을 지목하고, 이를 해결할 전략(Proposal)을 도출하십시오.

[보고서 필수 구조 (Fact - Reason - Proposal)]

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

- **Fact (진단):**

- 추상적인 표현 대신 구체적인 데이터를 인용하십시오. (예: "R&D 부서의 퇴사율이 타 부서 대비 O 배 높으며, 퇴사 사유 키워드 Top 5에 '성장', '시스템' 등이 포함됨")
- **Reason (원인):**
 - 단순히 "일이 힘들어서"가 아니라, 리뷰 데이터에 기반하여 **구조적인 원인**을 명시하십시오.
 - **분석 포인트:** 기술 부채, 멘토링 시스템 부재, 비효율적 의사결정 구조 등.
- **Proposal (제언):**
 - 확인된 원인을 제거할 수 있는 **즉각적인 개선 액션 플랜**을 제시하십시오.
 - 예시: 최신 개발 장비 도입 예산 편성, 코드 리뷰 문화 정착을 위한 시간 보장, 불필요한 문서 작업 간소화 등.

Mission 4-2. 경영 제언: 미래 비전 및 조직 문화 혁신

기술적인 해결책을 넘어, 회사가 개발자들의 성장을 지원한다는 **미래 비전**을 제시하여 조직의 신뢰를 회복해야 합니다.

[제언 요건]

- **성장 지원 시스템 구축:** "여기선 도태될 것 같다"는 불안감을 해소하기 위해, 사수-부사수 멘토링 제도, 기술 컨퍼런스 지원, 학습 휴가 등 구체적인 성장 지원책을 제안하십시오.
- **소통 및 리더십 혁신:** "소통 부재"와 "구식 마인드"를 해결하기 위해, 경영진이 정기적으로 비전을 공유하고 현장의 목소리를 듣는 **타운홀 미팅** 또는 정기 1on1 미팅 정례화를 제안하십시오.

4. 제출물

다음 내용이 포함된 DIG_실습_Phase4_학번_이름.pdf를 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

평가 분류	세부 평가요소	상세 채점 가이드 (Log 확인)	배점
ACI 평가요소	④-1. 사실 기반 해석 (Fact Finding)	<ul style="list-style-type: none"> - [데이터 인용] 리뷰 데이터의 구체적 내용(장비 노후, 사수 부재, 보고 문화 등)을 근거로 원인을 분석했는가? - [오진단 방지] 돈(연봉) 문제가 아님을 명확히 인지하고 비금전적 요인을 지적했는가? 	30 점
	④-2. 논리적 제언 (Logical Proposal)	<ul style="list-style-type: none"> - [전략의 구체성] 막연한 "문화 개선"이 아니라, 멘토링, 장비 교체, 문서 간소화 등 실행 가능한 대책을 제시했는가? - [구조] Fact → Reason → Proposal 흐름이 명확한가? 	30 점