

■ 작업형 1 유형 - 실전 최종 문제

데이터 파일: 작업형 1 유형_리얼문제_데이터셋.csv

해당 데이터는 A 사 고객의 제품 불량 및 서비스 접수 이력을 담은 것이다.

다음 조건에 따라 분석을 수행하고, 최종 결과 값을 도출하시오.

문제 1. 결측치 처리

- '사용개월수'의 결측치는 전체 평균값으로 채워라.
- '서비스만족도'의 결측치는 '고객성별'과 '제품군'별 평균 만족도로 채워라.
- '고객연령' 결측치는 같은 '지역'과 '고객성별' 그룹의 중앙값으로 채워라.

문제 2. 파생변수 생성

- '서비스까지_소요일수': '생산일시'부터 '서비스접수일시'까지의 일수
- '고객연령대': '고객연령'을 기준으로 10 단위 연령대 구간 생성 (예: 20 대, 30 대...)
- '서비스시간대': '서비스접수일시' 기준, 시간(hour)을 추출하여
 - * 06~11 시는 '오전',
 - * 12~17 시는 '오후',
 - * 그 외는 '야간'으로 분류

문제 3. 서비스 이력 집계

- '고객성별', '제품군', '고객연령대' 별로
 - * 평균 '서비스만족도'
 - * 평균 '서비스까지_소요일수'
 - * 총 '재발횟수'
- 위 집계 결과에서 '서비스만족도' 평균이 가장 낮은 1 건을 출력하시오.

문제 4. 불량 집중 고객군 탐색

- 다음 조건을 만족하는 고객 수를 구하시오.
- * '재발횟수'가 2 회 이상
- * '서비스만족도'가 전체 평균보다 낮음

- * `리콜조치유무`가 'N'
- * `사용개월수`가 6 개월 이상

문제 5. 이상치 및 통계 분석

- `서비스까지_소요일수`가 3σ 이상인 값을 이상치로 판단하고, 해당 건수는 몇 건인가?
- `사용개월수`와 `서비스만족도` 간의 피어슨 상관계수를 소수점 셋째자리까지 구하시오.

문제 6. 최종 분석

- `고객성별`, `제품군`, `고객연령대` 별로 `서비스만족도`가 가장 낮은 조합을 구하고, 해당 조합의 다음 정보를 표 형태로 출력하시오:
- | 고객성별 | 제품군 | 연령대 | 평균만족도 | 평균사용개월수 | 평균소요일수 | 재발횟수합계 |

■ 최종 제출 항목

- * 문제 3: 서비스만족도 평균이 가장 낮은 조합 1 건
- * 문제 4: 조건을 만족하는 고객 수
- * 문제 5: 이상치 건수, 상관계수
- * 문제 6: 최종 조합 및 해당 조합의 통계 요약