

[붙임 : 유형별 예시문제]

○ 단답형

단답형 예시문제

여러 명의 사용자들이 컴퓨터에 저장된 많은 자료들을 쉽고 빠르게 조회, 추가, 수정, 삭제할 수 있도록 해주는 소프트웨어는 무엇인가?

○ 작업형 제1유형 : 데이터 처리 영역

작업형 제1유형
예시문제

mtcars 데이터셋(mtcars.csv)의 qsec 컬럼을 최소최대 척도(Min-Max Scale)로 변환한 후 0.5보다 큰 값을 가지는 레코드 수를 구하시오.

○ 작업형 제2유형 : 모형 구축 및 평가 영역

**작업형 제2유형
예시문제**

아래는 백화점 고객의 1년 간 구매 데이터이다.

아 래

(가) 제공 데이터 목록

- ① **y_train.csv** : 고객의 성별 데이터 (학습용), CSV 형식의 파일
- ② **X_train.csv, X_test.csv** : 고객의 상품구매 속성 (학습용 및 평가용), CSV 형식의 파일

(나) 데이터 형식 및 내용

- ① **y_train.csv** (3,500명 데이터)

cust_id	gender
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9

- * custid: 고객 ID
- * gender: 고객의 성별 (0: 여자, 1: 남자)

- ② **X_train.csv** (3,500명 데이터), **X_test.csv** (2,482명 데이터)

cust_id	총구매액	최대구매액	환불금액	주구매상품	주구매지점	내점일수	내점당구매건수	주말방문비율	구매주기
0	68282840	11264000	6860000.0	기타	강남점	19	3.894737	0.527027	17
1	2136000	2136000	300000.0	스포츠	잠실점	2	1.500000	0.000000	1
2	3197000	1639000	NaN	기타	관악점	2	2.000000	0.000000	1
3	16077620	4935000	NaN	기타	광주점	18	2.444444	0.318182	16
4	29050000	24000000	NaN	기타	본점	2	1.500000	0.000000	85
5	11379000	9552000	462000.0	디자이너	일산점	3	1.666667	0.200000	42
6	10056000	7612000	4582000.0	시티웨어	강남점	5	2.400000	0.333333	42
7	514570080	27104000	29524000.0	명품	본점	63	2.634921	0.222892	5
8	688243360	173088000	NaN	기타	본점	18	5.944444	0.411215	15
9	26640850	13728000	NaN	농산물	대전점	1	12.000000	0.000000	0

고객 3,500명에 대한 학습용 데이터(y_train.csv, X_train.csv)를 이용하여 성별예측 모형을 만든 후, 이를 평가용 데이터(X_test.csv)에 적용하여 얻은 2,482명 고객의 성별 예측값(남자일 확률)을 다음과 같은 형식의 CSV 파일로 생성하시오.(제출한 모델의 성능은 ROC-AUC 평가지표에 따라 채점)

<제출형식>

```
custid,gender
3500,0.267
3501,0.578
3502,0.885
...
```

<유의사항>

성능이 우수한 예측모형을 구축하기 위해서는 적절한 데이터 전처리, Feature Engineering, 분류 알고리즘 사용, 초매개변수 최적화, 모형 앙상블 등이 수반되어야 한다.