

한국데이터정보과학회 & SAS KOREA 데이터 분석 경진대회: 데이터 안내

- SAS®Viya®와 함께 데이터 기반의 문제를 해결하다 -

□ 데이터 소개

제공된 데이터는 R기관에서 조사한 상권별 전기 사용량에 대한 데이터이다. 주어진 학습용 데이터 (TRAIN_DATA.csv)를 활용하여 각 상권의 전기 총사용량을 예측하는 모델을 개발하고, 평가용 데이터 (TEST_DATA.csv)에서 상권 내 모든 건물에 대한 전기 총사용량을 예측한 결과를 y_pred 열에 생성하여 csv 또는 xlsx 형식으로 저장하여 제출하시오.

단, 예측 결과는 RMSE (Root mean square error)를 바탕으로 평가한다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- 데이터 내 변수 설명

	변수명	설명	타입	형태(예시)
1	DATA_YM	데이터가 작성된 월	numeric	YYYYMM(202303)
2	AREA_ID	상권 코드	numeric	9626
3	AREA_NM	상권 명칭	character	중앙로역_4
4	DIST_CD	상권 위치한 지역 코드	numeric	27110
5	DIST_NM	상권 위치한 지역 명칭	character	중구
6	TOTAL_BIDG	상권 내 총 건물 수	numeric	538
7	FAC_NEIGH_1	상권 내 1종 생활편의시설 수	numeric	266
8	FAC_NEIGH_2	상권 내 2종 생활편의시설 수	numeric	88
9	FAC_CULT_MTG	상권 내 문화 및 집회시설 수	numeric	6
10	FAC_RELG	상권 내 종교시설 수	numeric	3
11	FAC_RETAIL	상권 내 판매시설 수	numeric	2
12	FAC_MEDI	상권 내 의료시설 수	numeric	3
13	FAC_YOSE	상권 내 아동 및 노인복지시설 수	numeric	0
14	FAC_TRAIN	상권 내 수련원 시설 수	numeric	0
15	FAC_SPORT	상권 내 운동시설 수	numeric	0
16	FAC_STAY	상권 내 숙박시설 수	numeric	16
17	FAC_LEISURE	상권 내 여가 및 휴식시설 수	numeric	3
18	TOTAL_GAS	상권 내 모든 건물의 가스 총 사용량	numeric	517520
19	CMRC_GAS	상권 내 상업용 가스 사용량	numeric	466070
20	TOTAL_ELEC	상권 내 모든 건물의 전기 총 사용량	numeric	1174.57

※ 모든 건물이란 상업용 건물을 비롯하여 기타 용도의 건물을 모두 지칭함

붙임 1

예측값 제출 양식

	F	G	H
1	y_pred		
2	0.1235		
3	0.1906		
4	0.5507		
5	0.8618		
6	0.9624		
7	0.3858		
8	0.9403		
9	0.0207		
10	⋮		
11	0.6374		

※ 위 작성된 'y_pred'는 임의의 난수를 생성한 값임

붙임 2

분석 경진대회 일정

① 참가 신청 접수	'25.3.10.(월) ~ 3.24.(월)	제출서류: 참가신청서, 개인정보 수집·이용 동의서, 참가 서약서
↓		
② 사전교육	'25.3.28.(금)	SAS®Viya® 관련 온라인교육 (참가 팀 내 1명이상 이수)
↓		
③ 대회 진행	'25.3.28.(금) ~ 4.30.(수)	주 1회 채점 후 중간순위 공지
↓		
④ 예선심사 발표	'25.5.9.(금)	상위 5개 팀(개인) 본선 진출 발표
↓		
⑤ 본선 발표평가 및 시상식	'25.5.16.(금)	조선대학교에서 진행

※ 운영 상황에 따라 세부 일정이 변경될 수 있음