

[분석 주제 및 심사 기준]

1. 분석주제: 지하철 역사 승하차자 수와 서울 기상상태 데이터를 활용하여 지하철 역사 내 실내 공기질 정도를 예측하는 모형을 개발하는 것으로, 공개된 공공 데이터를 추가적으로 매쉬업하는 것은 허용됩니다.

2. 데이터 설명/제공 데이터 및 교육자료

2.1 데이터 설명 및 제공 데이터: [☐ SAS 코리아 Hackathon 자료실](#)

2.2 교육 자료: <https://www.youtube.com/watch?v=67j2CkYxVcY> 접속 후 채널 방문([SAS 해커톤 Korea \(SAS/Viya 소개 \)](#))

3. 결과 제출 양식: 데이터 EDA에서 분석 결과까지 일련의 분석 과정에 대한 분석 내용에 대한 PPT 문서 1부를 제출합니다.

4. 심사 기준

4.1 모델정확성(40%)

; 공기질 데이터 중 22년도 Train Data는 모델링에 23년도 TEST Data는 홀드아웃 데이터로 사용하여 결과를 생성하고 해당 결과를 모델 정확성으로 평가합니다. 평가지표 중에서 AUROC(ROC아래면적) 값을 기준으로 우선으로 평가하며, 데이터 역할에 따른 결과표 반드시 첨부해 주시기 바랍니다(관측값 수, 오분류 비율, KS 통계량, AUROC 필수).

4.2 분석의 적정성(60%)

; 데이터 EDA, Target 설정 등의 분석 과정 내용(20%), 모형 생성에 대한 다양한 접근 및 적절한 내용(20%), 모형 결과에 올바른 해석 및 추가 제언 관련 내용(20%)을 포함해 주시고, 분석 단계에 따른 내용 및 결과 화면 등을 반드시 캡처 하셔서(필수) 문서에 포함해주시기 바랍니다.