

분석 데이터 설명



- 편성시간표

카테고리	No	컬럼명	데이터 형식	컬럼 레이블	Test set에 포함 여부	역할
PRGM_TIMETABLE 편성시간표	1	BDCT_DT	Date	방송일자	O	Input
	2	BDCT_PGM_ID	Numeric	방송프로그램ID	O	Input
	3	BDCT_PGM_NM	Character	방송프로그램명	O	Input
	4	BDCT_STRT_DTIME	Datetime	방송시작시간	O	Input
	5	BDCT_END_DTIME	Datetime	방송종료시간	O	Input

- PGM 편성정보

카테고리	No	컬럼명	데이터 형식	컬럼 레이블	Test set에 포함 여부	역할
PRGM_INFO PGM 편성정보	1	BDCT_PGM_ID	Numeric	방송프로그램ID	O	Input
	2	TIME_TERM	Numeric	노출시간	X	Input
	3	TV_GOODS_GRP_CD	Character	TV상품그룹코드	O	Input
	4	GOODS_CLSS_THN_CD	Character	상품세분류코드	O	Input
	5	TVGGRP_SALE_PRC	Numeric	TV상품그룹판매가	X	Input
	6	ORI_PD_CD	Character	PD코드	O	Input
	7	PD_1	Character	PD1코드	O	Input
	8	ORI_SH_CD	Character	SH코드	O	Input
	9	SH_1	Character	SH1코드	O	Input
	10	SH_2	Character	SH2코드	O	Input

- 상품정보

카테고리	No	컬럼명	데이터 형식	컬럼 레이블	Test set에 포함 여부	역할
GOODS_INFO 상품정보	1	TV_GOODS_GRP_CD	Character	TV상품그룹코드	O	Input
	2	BRND_MID_CD	Character	브랜드코드	O	Input
	3	GOODS_CLSS_LRG_CD	Character	상품대분류코드	O	Input
	4	GOODS_CLSS_LRG_NM	Character	상품대분류명	O	Input
	5	GOODS_CLSS_MID_CD	Character	상품중분류코드	O	Input
	6	GOODS_CLSS_MID_NM	Character	상품중분류명	O	Input
	7	GOODS_CLSS_SML_CD	Character	상품소분류코드	O	Input
	8	GOODS_CLSS_SML_NM	Character	상품소분류명	O	Input
	9	GOODS_CLSS_THN_CD	Character	상품세분류코드	O	Input
	10	GOODS_CLSS_THN_NM	Character	상품세분류명	O	Input

- PGM 실적

카테고리	No	컬럼명	데이터 형식	컬럼 레이블	Test set에 포함 여부	역할
PRGM_ACRS PGM 실적	1	BDCT_PGM_ID	Numeric	방송프로그램ID	O	Input
	2	TV_GOODS_GRP_CD	Character	TV상품그룹코드	O	Input
	3	GOODS_CLSS_THN_CD	Character	상품세분류코드	O	Input
	4	EXPC_SALE_AMT	Numeric	예상매출액	X	△
	5	PUR_SALE_AMT	Numeric	매출액	X	Target
	6	SALE_GOAL_AMT	Numeric	매출목표	X	△
	7	BDCT_PRFT_AMT	Numeric	매출기반이익금액	X	△
	8	CNTR_GOAL_AMT	Numeric	이익목표	X	△
	9	BDCT_PRFT_ACHV	Numeric	매출기반이익달성율	X	△
	10	PRFT_ACHV	Numeric	이익달성율	X	△
	11	SALE_ACHV_RATE	Numeric	매출달성율	X	△
	12	PUR_ORD_CNT	Numeric	순주문건수	X	Input/△
	13	ORD_QTY	Numeric	순주문수량	X	Input/△

- 데이터 셋 조인 정보

Join 대상 테이블		Key	비고
PRGM_TIMETABLE	PRGM_INFO	BDCT_PGM_ID	AND조건
PRGM_INFO	PRGM_ACRS	BDCT_PGM_ID	
		TV_GOODS_GRP_CD	
		GOODS_CLSS_THN_CD	
PRGM_ACRS	GOODS_INFO	TV_GOODS_GRP_CD	AND조건
		GOODS_CLSS_THN_CD	

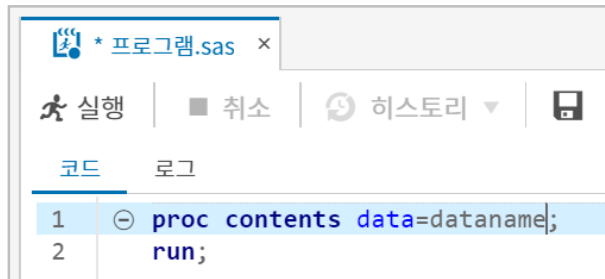
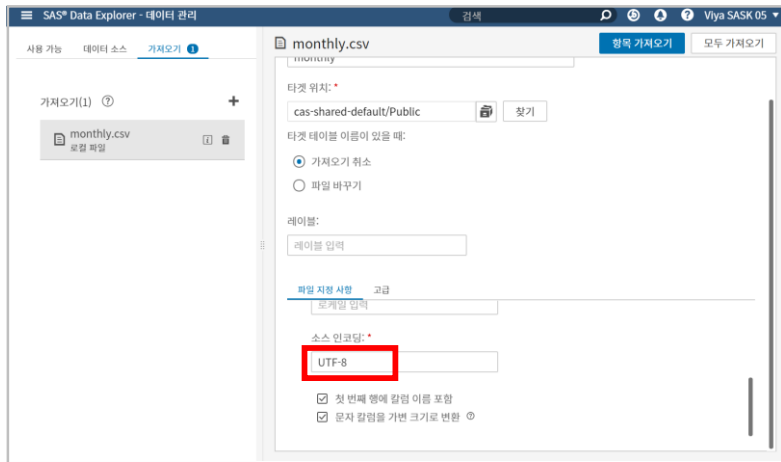
- 제출 데이터 형식

: 제공되는 Test 데이터에 대한 PUR_SALE_AMT(매출액) 컬럼을 예측한 값을 제출

BDCT_PGM_ID (방송프로그램ID)	TV_GOODS_GRP_CD (TV상품그룹코드)	PUR_SALE_AMT (매출액)
253963	02758	
253964	06928	
253966	00203	
...	...	

외부 데이터 활용

- 데이터 불러오기를 할 때 **인코딩**을 **UTF-8**로 설정하여 불러 올 것
(UTF-8이 아닌 EUC-KR로 불러오면 **오류**가 발생함)
- 로컬 환경에서 데이터를 불러오면 인코딩 문제가 있을 수 있으므로 **서버 환경에서 바로 불러올 것**
- PROC CONTENTS 프로시저를 활용하여 인코딩을 확인할 것



본 분석에 대한 가이드

- 주어진 4개 테이블을 조인하면 2018년 1월~12월의 PUR_SALE_AMT(매출액, Target) 컬럼은 결측으로 되어 있는데, 이 부분을 예측하여 제출해야 함
 - 훈련 데이터: 조인 후 2018년 이전의 데이터를 이용(파생변수, 외부변수 활용)
 - 테스트 데이터(제출): 2018년의 각 방송에서의 매출액을 예측하여 제출함(submission 데이터셋)
- 외부 데이터의 출처 표기 필수
- 홈쇼핑 방송 편성을 위한 다음 주의 매출액을 예측하는 것이기 때문에, 일주일 전에 알 수 없는 정보는 사용하지 않도록 주의. 단, 예측해서 사용하는 것은 가능.

SAS 분석 솔루션

1. 데이터 처리
 - ✓ SAS® Enterprise Guide®
2. 데이터 탐색 및 시각화: SAS® Drive
 - ✓ SAS® Visual Analytics
 - ✓ SAS® Visual Statistics
 - ✓ SAS® Visual Data Mining and Machine Learning
3. 예측 모델링 : SAS® Drive
 - ✓ SAS® Visual Statistics
 - ✓ SAS® Visual Data Mining and Machine Learning

감사합니다

