

## 마케팅 데이터 분석 기말 과제

# Sales Data Analysis

본 과제는 주어진 데이터(sales.csv)에 대한 분석을 토대로 아래 1~5번 문항에 답하는 과제입니다.

- 제출 기한: 12월 10일 (일) 오후 3시 사이버캠퍼스 과제 란
- 양식: 1) 답안 파일(MSWord 또는 PDF형식), 2) R스크립트(본인학번.R) 파일을 함께 제출
  - 1) 답안 파일 포함 내용
    - 과제 상단에 학번과 이름 명시
    - 각 문항의 답안: R code 및 실행 결과 output 캡처와 함께 본인의 해석
  - 2) R스크립트 파일 포함 내용: 1~5번까 답안을 작성하는 데에 사용한 코드를 문항 순서대로 정리

### ● 데이터 설명

“Sales.csv” 파일은 온라인 쇼핑몰을 운영하는 한 의류업체의 고객 835명에 대한 방문 및 구매 정보와 함께 만족도 평가 결과를 담고 있다. 변수명(variable)과 변수의 정의(description)는 다음 표와 같다.

Variable	Description
1. acctAge	거래기간: 회원 가입 후 경과 시간(단위: 개월)
2. visitsMonth	최근 한달 간 쇼핑몰 방문 횟수(단위: 회)
3. spendToDate	누적 총 구매 금액(단위: USD)
4. spendMonth	최근 한 달간 구매 금액 (단위: USD)
5. satSite	우리 쇼핑몰 웹사이트 디자인에 대한 만족도(단위: 1~10점)
6. satQuality	우리 쇼핑몰 판매 제품 품질에 대한 만족도 (단위: 1~10점)
7. satPrice	우리 쇼핑몰 판매 제품의 가격에 대한 만족도 (단위: 점)
8. satOverall	우리 쇼핑몰에 대한 전반적인 만족도 (단위: 점)
9. coupon	타깃 쿠폰 발송(수신) 여부
10. purchase	타깃 쿠폰 반응(=쿠폰을 사용하여 구매) 여부

1. summary()함수를 이용하여 acctAge, spendToDate의 평균값을 구하고 그 결과를 해석하여라.
  
2. hist ()함수를 이용하여 visitsMonth의 히스토그램을 그리고 그 결과를 해석하여라.
  
3. kmeans()함수를 이용하여 주어진 변수 중 1~8번까지의 8개 변수를 사용하여 k=4로 하는 k평균 군집분석을 수행한 뒤, 아래 질문에 답하세요.
  - 3-1) 각각의 군집에 몇 명의 고객이 배정되었는가? (hint: table()함수를 이용)
  - 3-2) 4개 군집에 속한 고객들의 spendToDate값 평균은 각각 어떻게 되는가? (hint: aggregate()함수 이용)
  - 3-3) 4개 군집에 속한 고객들의 spendToDate값의 평균이 차이가 있는 지를 확인하기 위한 분산분석을 실행하고, 그 결과를 해석하여라.
  
4. 분류나무 분석: 835명의 데이터를 학습데이터(80%)와 검증데이터(20%)로 분류한 뒤, purchase를 예측하기 위한 분류나무 기본형을 그리는 분석을 실시하여라.
  - 4-1) 학습데이터와 검증데이터에는 각각 몇 명의 고객이 배정되었는가?
  - 4-2) 분류나무(기본형)를 그린 결과그림을 아래에 첨부하고, purchase가 1(구매)로 판정되기 위한 의사결정 규칙을 나열하여라.
  - 4-3) 위에서 그린 분류나무(기본형)로 학습데이터를 예측할 경우 정확도(accuracy)는 얼마인가? (confusion matrix 활용)
  - 4-4) 위에서 그린 분류나무(기본형)로 검증데이터를 예측할 경우 정확도(accuracy)는 얼마인가? (confusion matrix 활용)