

# 신용카드 거래 시각화

DACON뱅크샐러드

이은진

# 데이터

- 소개 : 소상공인 신용카드거래 내역 데이터  
(2016.08.01 ~ 2018.07.31)
- 목적 : 금융 쉽게 알기, 인사이트 발굴 및 시각화 대회
- 목표 : 가게 별 매출액 예측하기 (가게 0~199)

변수명	변수 설명
Store_id	상점 고유 번호
Date	거래 일자
Time	거래 시간
Card_id	카드 번호의 hash 값
Amount	매출액, 0보다 작은 음수는 거래 취소(환불)
Installments	할부개월수. 일시불은 빈 문자열
Days_of_week	요일, 월요일이 0, 일요일은 6
holyday	공휴일 여부, 1이면 공휴일

# EDA 해볼 것

- 상점별 amount range에 따라 어떤 업종인지 파악할 수 있을 것
  - > 오른쪽 빈도표를 만들어보자
  - > 편의점이나 슈퍼 같은 상점이면 range도 다양할 것이고, 많은 고객들이 이용할 것 ㄱ  
이라 예상. 반면, 고가의 물건이나 서비스를 판매하는 곳은 이용하는 사람도 적고(카드 id  
unique 개수), amount가 높고, 할부 가능성이 높다고 예상됨
- 추가적으로 상점별로 환불을 자주했는지도 같이 파악하면 좋을 듯하다.

\*\* 0번 가게는 어떤 걸 팔길래 웬만한 사람들이 (자주) 이용했지??

Store_id	카드 id unique 개수	range
0		
1		
...		
200		

# EDA 해볼 것

- 시계열 파악해보자 -> 요일, 월, 시간대, 평일/주말 별로  
-> 주기적으로 방문하는 가게인지? (카드 id 기준으로)  
-> 예를 들어, 평일에 이용빈도가 높는데, 주말에 확 줄어들었으면 직장 근처. 반대의 경우에는 주택가 근처로 예상
- 요일 특성, 공휴일 특성별 소비패턴 파악하기

\*\* 카드 id '2b76ad7f30'은 할부 자주하며, 큰 돈 결제(1번에서), 어쩌면 개인카드가 아니라 법인카드일 수도 있지 않을까? 더 파악해봐야 할 듯

- EDA를 할 때, 1)전체적인 흐름을 보여주는 그래프와 2)특정 상점이나 특정 카드 id를 그래프로 그려 특이한 소비패턴을 비교하는 그래프 둘다 필요해보임

# 분석 방향?

- 우선은 파일 크기 때문에 엑셀에서 피벗테이블 실행속도도 느렸다ㅠ -> 무조건 파이썬!
- 어떻게 상점별로 매출액(총합)을 예측할 수 있을까?  
-> 일단, 최종모델에는 카드 id 정보를 포함하는 설명변수가 없어야 하지 않을까??
- 총합을 예측하는 것이니, 1) 혹은 특정 기간별 상점의 매출액을 예측 후 합산하는 방법,  
2) 처음부터 총합을 예측하는 모델을 만든다...-> 가능할지???

	A	B
1	store_id	total_sales
2	0	0
3	1	0
4	2	0
5	3	0
6	4	0
7	5	0
8	6	0

<submission 데이터>