

Chapter 03. 금융 AI 모델의 사용데이터

- 핀테크 AI 모델에서 사용하는 데이터에는 무엇이 있나?



3-1. 핀테크 AI에서 사용하는 데이터

■ 변수와 자료의 개념

- 변수(Variable) : 연구자의 관심대상이 되는 성격 or 속성
- 자료(Data) : 이러한 변수를 관찰하여 기록한 결과

한 집단의 특성을 쉽게 알아보고 분석하기 위해서는

수집된 자료를 의미있는 모양으로 분류·정리하는 것이 중요하다

3-2. 자료의 종류

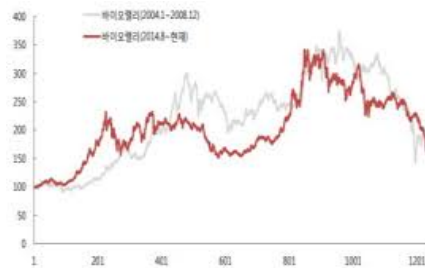
- 범주형 자료(질적, 비계량)
 - 명목변수(Nominal Variable) : 개체나 사람의 특성만 (성별, 종교 등)
 - 순서(서열)변수(Ordinal Variable) : 측정대상 간 선호를 부여 (선호도, 만족도)
- 양적인 자료(양적, 계량)
 - 등간변수(Interval Variable) : 측정대상 간의 순서 + 값 사이의 간격이 일정
IQ(지능지수), 온도 등
*절대 영점이 없음(100도는 50도의 2배 뜨겁다?)
 - 비율변수(Ratio Variable) : 측정대상 간에 비율 계산이 가능함
(연령, 무게, 거리, 시간)
 - 이산형 변수(Discrete Variable) : 점수, 빈도수
 - 연속형 변수(Continuous Variable) : 실수, 키, 몸무게

3-2. 자료의 종류

■ 연속형 변수와 이산형 변수

- 연속형 변수 : 시간에 대해서 행동과 상태가 유한차원 벡터공간
: 주가 수익률, 자동차의 운행거리
- 이산형 변수 : 시간에 대해서 행동과 상태가 원소형태의 값(Value)를 가짐
: {라이트를 켜다, 라이트를 끄다, 매도한다, 매수한다}

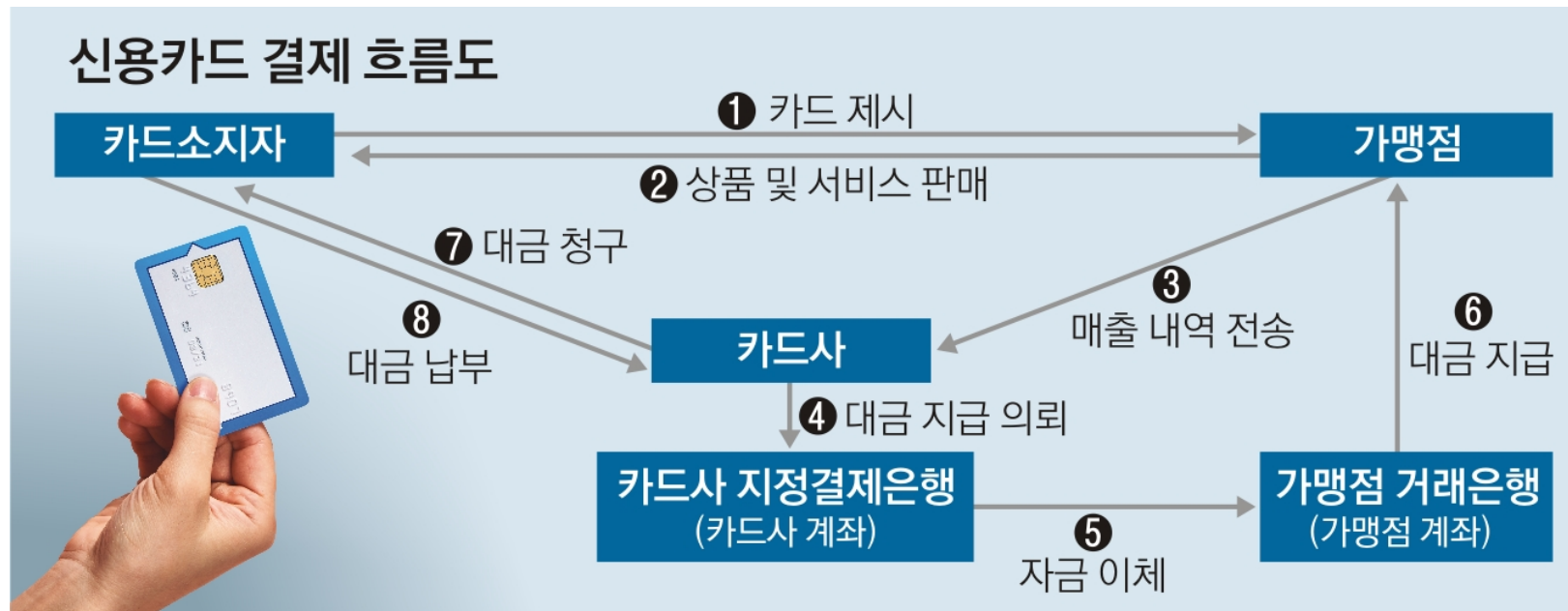
<연속형 변수 : 수익률>



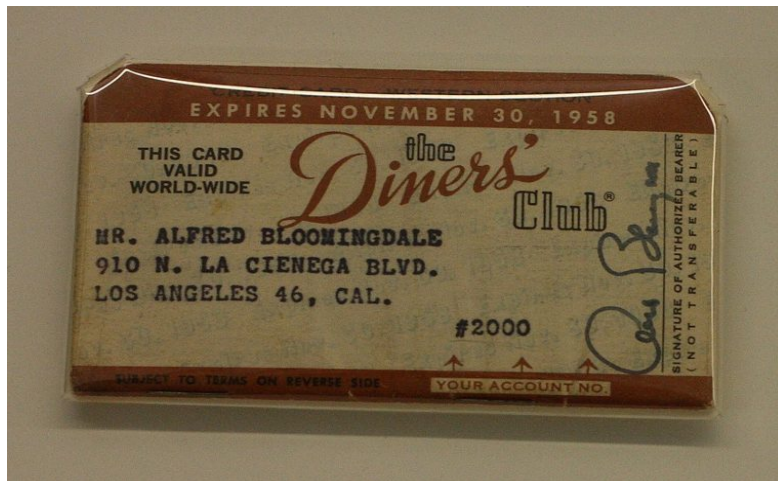
<이산형 변수 : 전원 상태 >



3-3. 사례 : 신용카드 데이터



3-3. 사례 : 신용카드 데이터



3-3. 사례 : 신용카드 데이터

- <https://www.kaggle.com/mlg-ulb/creditcardfraud>



학습정리 #3차

■ 데이터(자료)의 종류

- 인공지능 모델에서 사용하는 데이터의 종류는 크게 질적 데이터와 양적 데이터로 구분
- 질적 데이터는 성별, 직업, 국가 등 대상의 타입(TYPE)과 관련된 것이 많으며, 양적 데이터는 계산이 가능한 값(Value)의 형태를 가지고 있음

■ 이산형 데이터와 연속형 데이터

- 이산형 데이터 : 점수, 빈도 수, (금융 분야) 신용 점수 등
- 연속형 데이터 : 실수, 키, 몸무게, (금융 분야) 수익률, 재무비율 등

■ 신용카드 데이터 사례

- 신용카드 결제 데이터 : 금액, 장소, 시간, 카드번호 등
- 신용카드 데이터 분석 : 시간 + 장소 + 품목의 조합으로 어떤 패턴을 만들어 낼 수 있음
- 시간, 장소, 품목을 “패턴”을 특성(Feature)으로 만들 수 있음