



● 범주형 : 몇 개의 범주로 나뉘어진 자료를 의미

● 명목형 : 성별, 성공여부, 혈액형 등
단순히 분류된 자료

● 순서형 : 개개의 값들이 이산적이며 그들 사이에
순서 관계가 존재하는 자료

● 수치형 : 이산형과 연속형으로 이루어진 자료를 의미

● 이산형 : 이산적인 값을 갖는 데이터로
출산횟수 등을 의미

● 연속형 : 연속적인 값을 갖는 데이터로 신장,
체중 등을 의미

단일변수(DDA분석)

하나의 컬럼(feature)변수를 사용 ~> 데이터 신뢰성 확인(이상치/불순물)

데이터 타입	유형
수치형	확률분포 or 히스토그램
범주형	빈도 수 그래프(Bar / Pie chart)

다변수(EDA분석)

여러 개의 컬럼(feature)변수 사용 ~> 변수간 분석

데이터 타입	유형
Y(Target : 목표변수) : 연속 / X : 연속	산점도 / 회귀 그래프 / 히트맵
Y(Target : 목표변수) : 연속 / X : 범주	막대 그래프 / 상자그래프
Y(Target : 목표변수) : 연속 / X : 시간(순서)	선 그래프 / 포인트 그래프

데이터 분석 절차

1. DDA : 묘사적 데이터 분석

-가져온 데이터의 통계량을 확인하고 분석하고자 하는 타겟값 찾기(Y[종속변수]) ,
다른 변수(X[독립변수])의 특징 파악(시각화[단일변수분석]를 통해 데이터의 신뢰성 확보)하는 절차

2. EDA : 탐색적 데이터 분석

-Y라는 목표 변수와 X라는 설명 변수들 간 관계를 파악하는 단계 ~> 데이터간 관계성(경향성, Trend) 파악
시각화!!(다 변수 분석)

시각화 라이브러리

Matplot : MATLAB기반 파이썬 라이브러리

Seaborn : Matplot 라이브러리를 기반, 통계 전용 시각화 툴, Pandas 라이브러리와 연계가 잘됨