

## 데이터 이해 먼저 해야 함

- info보고
- describe보고
- 결측치 보고
- 데이터의 분포를 봐야 함 - 분포를 보면 데이터가 어떻게 생겼는지 볼 수 있다
- 중간값과 평균
- 사분위
- 50% 분위가 중간값
- 왼쪽으로 치우친 종 중간값
- 오른쪽으로 치우친 종의 중간값
- 도메인 정보 조사

In [ ]:

- 1 목적
- 2 머신러닝 : 분석용 데이터셋 -> 학습 -> 결과 예측
- 3
- 4 그래서 분석용 데이터셋을 잘 만들어야 함 -> 파생변수 -> EDA

In [ ]:

- 1 분석용 데이터셋 잘 만들기
- 2
- 3 결측치나 이상치 처리
- 4 피쳐 엔지니어링 - 파생 변수 만드는 것 (분석 모델이 더 잘 학습)
- 5 탐색적 데이터 분석(EDA) - 통계나 시각화를 이용해서 보고 분석

In [ ]:

- 1 확률

In [ ]:

- 1 편차
- 2
- 3 표준편차
- 4 평균으로부터 데이터가 얼마나 떨어져 있는가
- 5 편차를 제공해서 절대값으로 볼 수 있게 만든 것
- 6

In [ ]:

```
1 분포
2 확률을 나타내는 하나의 방법
3
4 균일 분포
5 모든 확률이 같음
6 히스토그램 키가 같은 모습
7
8 정규 분포
9
10 표준 정규 분포
11 평균이 0
12 정규분포가 1인 정규분포
13 표준 정규 분포에 가까울수록 좋다
14 퍼지너가 뽕족하다고 좋은게 아님
15
16 샘플링수가 많을수록 (표준?)정규분초에 가깝게 나온다
17
18
```

In [ ]:

```
1 벡터는 방향, 화살표
2
3 행렬
4
5 머신러닝 기본에서 코드가 다 계산을 해주는데
6 계산을 할 때 그 코드를 까보면 다 행렬을 포함한 수학적 공식으
7
8 이미지 처리에 주로 쓰이는게 행렬이다
9
10 주성분 분석에 쓰인다
11 이미지 회전이나
12 이미지에 쓰인다
```

In [ ]:

```
1 회귀
2 선형 회귀
3
```

In [1]:

```
1 정리할 것
2
3
4 분포는 파생변수와 연관
5
6 배열
7 알고리즘은 수학
8 수학적 계산
9
10 2차원
11 3차원 - 실무적으로는 상관x, 이미지분석에서는 쓰이는데, 거의
12
13 deep러닝을 위한 수학
14
```

```
File "C:\Users\Wh\AppData\Local\Temp\ipykernel_6212\2045648064.py", line 1
```

```
    정리할 것
```

```
    ^
```

```
SyntaxError: invalid syntax
```

In [ ]:

```
1
```