



# Chapter.03 탐색적 데이터 분석-01.

## 탐색적 데이터 분석이란?



EDA(Exploratory Data Analysis) : 데이터에서 분석에 필요한 여러가지 통계량을 계산하고, 시각화를 통해서 이를 확인하는 작업.



### RECAP

- 탐색적 데이터 분석(Exploratory Data Analysis, EDA)는 데이터와 친해지는 과정입니다.
- 분석을 하면서 데이터에서 확인하고 싶은 정보들을 찾아가는 과정입니다.
- 정답처럼 규칙처럼 정해진 프로세스는 따로 없고, 완전히 분석가들의 직관과 분석 프로세스에 따라 갑니다.
- 어떤 데이터를 사용하느냐에 따라서도 천차만별로 갑니다.
- 나만의 EDA process를 구축하는 것이 Data Scientist로서의 역량이라고 볼 수 있습니다.

- 이제부터 하게 될 3가지 케이스에서 주요하게 생각해야 하는건 다음과 같습니다.

- 해당 데이터를 보고 어떤 인사이트를 이끌어내고 싶은지? (분석의 목표)

2. 데이터에 대한 여러가지 특성을 어떻게 파악하는지? (분석 방법론)
3. 이 때까지 배운 오픈소스 라이브러리들을 사용하여 EDA를 어떻게 수행하는지? (분석을 위한 프로그래밍)

## Hands-on

1. 첫번째 분석 케이스의 데이터입니다. 데이터의 설명을 읽고, 다운로드 받아서 구글드라이브에 업로드해보세요.

### Starbucks Customer Data

Starbucks customer dataset

[k https://www.kaggle.com/ihormuliar/starbucks-customer-data](https://www.kaggle.com/ihormuliar/starbucks-customer-data)



2. 두번째 분석 케이스의 데이터입니다. 데이터의 설명을 읽고, 다운로드 받아서 구글드라이브에 업로드해보세요.

### 2021 Kaggle Machine Learning & Data Science Survey

The most comprehensive dataset available on the state of ML and data science

[k https://www.kaggle.com/c/kaggle-survey-2021/data](https://www.kaggle.com/c/kaggle-survey-2021/data)



3. 세번째 분석 케이스의 데이터입니다. 데이터의 설명을 읽고, 다운로드 받아서 구글드라이브에 업로드해보세요.

### 소상공인시장진흥공단\_상가(상권)정보\_20211231

영업 중인 전국 상가업소 데이터를 제공합니다. (상호명, 업종코드, 업종명, 지번주소, 도로명주소, 경도, 위도 등)

[D https://www.data.go.kr/data/15083033/fileData.do](https://www.data.go.kr/data/15083033/fileData.do)

