



# Round 3

**PRESS  
START**



## 《 Round 3 》

- 가상환경 구축
- 각종 라이브러리 소개
- 판다스 기초



New  
Assignment



## 《 Round 3 》


- 가상환경 구축 《
- 각종 라이브러리 소개
- 판다스 기초



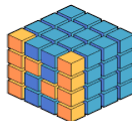
Let's  
Go




**Data Science Project\_1 =**




**+**




NumPy



**SciPy**

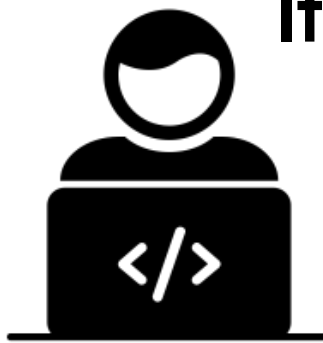


Pandas



Seaborn

**It's OK, but...**

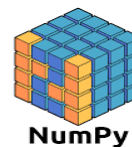


Data Science Project\_1

=



+



NumPy



SciPy

Pandas



Seaborn

Deep Learning Project\_1

=



+



TensorFlow

matplotlib

Data Science Project\_2

=



+

BeautifulSoup



learn



Keras

⋮

So sad...





- Anaconda는 파이썬의 배포판
- 특정 용도의 소프트웨어 패키지를 묶어서 제공
  - 각종 가상환경 구축 가능





  
ANACONDA

가상 환경 1



  
ANACONDA

가상 환경 2

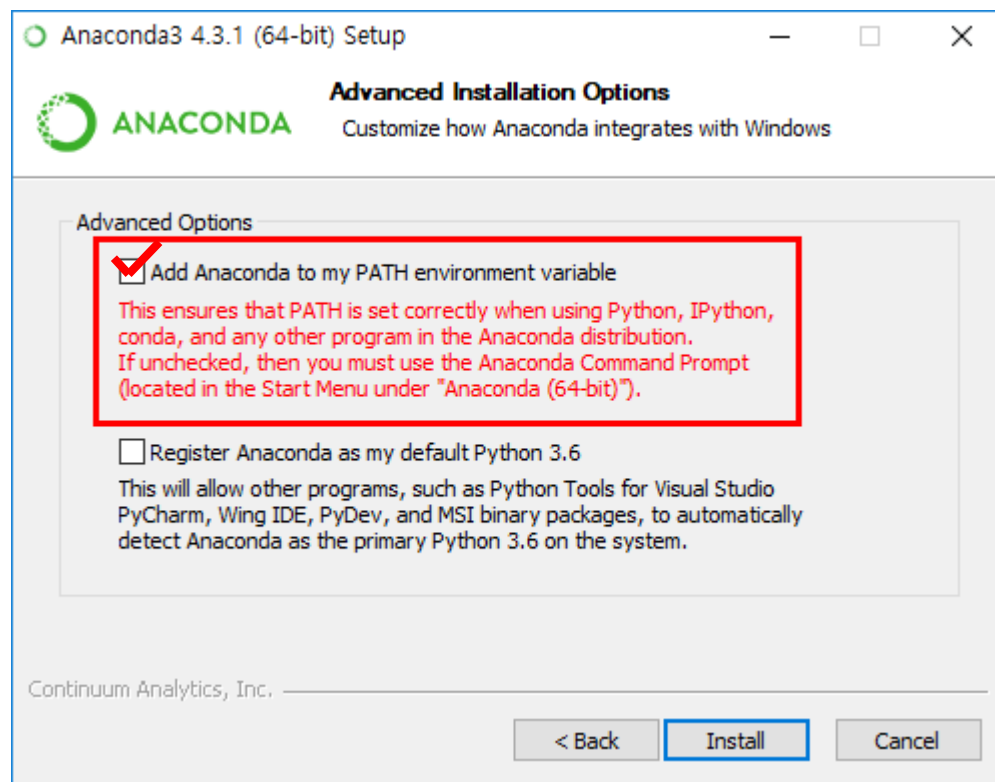


  
ANACONDA

가상 환경 3



...







## 4. Creating virtual environments with conda

```
conda create --name test python=3.7
```

## 5. Activating the virtual environments

## conda activate test / conda deactivate

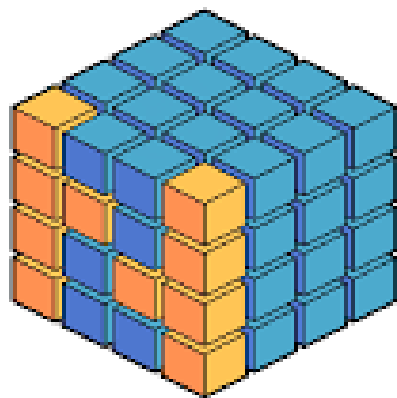
## 《 Round 3 》

- 가상환경 구축 - complete
- 각종 라이브러리 소개 《
- 판다스 기초



Let's  
Go





**NumPy**

- 대표적인 파이썬 기반 수치해석 라이브러리
  - 다양한 행렬과 배열 연산 함수 지원

# matplotlib

- 자료를 차트나 플롯으로 시각화(visualization)하는 패키지
- 판다스와 연계하여 다양한 방식으로 시각화하는 기능 제공

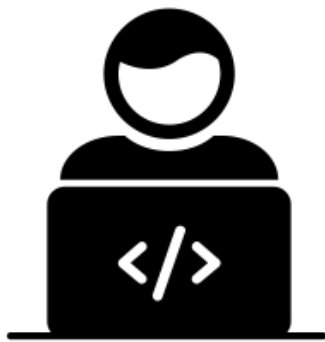


- 다양한 분석과 머신러닝 학습에 사용될 수 있는 파이썬 라이브러리
  - numpy와 Scipy와 상호 운영되도록 디자인 됨



- 데이터 분석 및 조작을 위한 라이브러리
- R과 비슷한 자료형을 사용함(Dataframe)

**이렇게 다양한 라이브러리를  
하나하나 관리해야 한다면...?**



**So sad...**



- 파이썬 라이브러리 패키지를 손쉽게 관리 할 수 있게 해주는 시스템
  - 파이썬 3.4이후 버전부터 기본 탑재!
  - 짧은 명령어 한 줄로 손쉽게 설치 가능



## 1. 패키지를 설치해줄 가상환경 활성화

```
C:\Users\user>conda activate test
```

## 2. pip install numpy로 numpy 설치

```
(test) C:\Users\user>pip install numpy
Collecting numpy
  Using cached https://files.pythonhosted.org/packages/a9/3f/numpy-1.18.1-cp37-cp37m-win_amd64.whl
Installing collected packages: numpy
Successfully installed numpy-1.18.1
```

## 3. pip freeze로 현재 설치된 패키지 확인

```
(test) C:\Users\user>pip freeze
certifi==2019.11.28
numpy==1.18.1
winertstore==0.2
```

## 《 Round 3 》

- 가상환경 구축 - complete
- 각종 라이브러리 소개 - complete
- 판다스 기초 《



Let's  
Go





**데이터 분석을 하기 위해서 필요한 것?**

**각종 통계적 기법??**

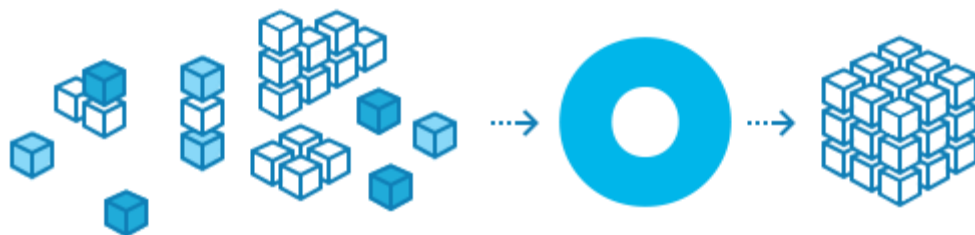
**다량의 데이터를 처리할 만한 풍부한 컴퓨팅 자원??**

**편리한 라이브러리와 시각화 도구??**

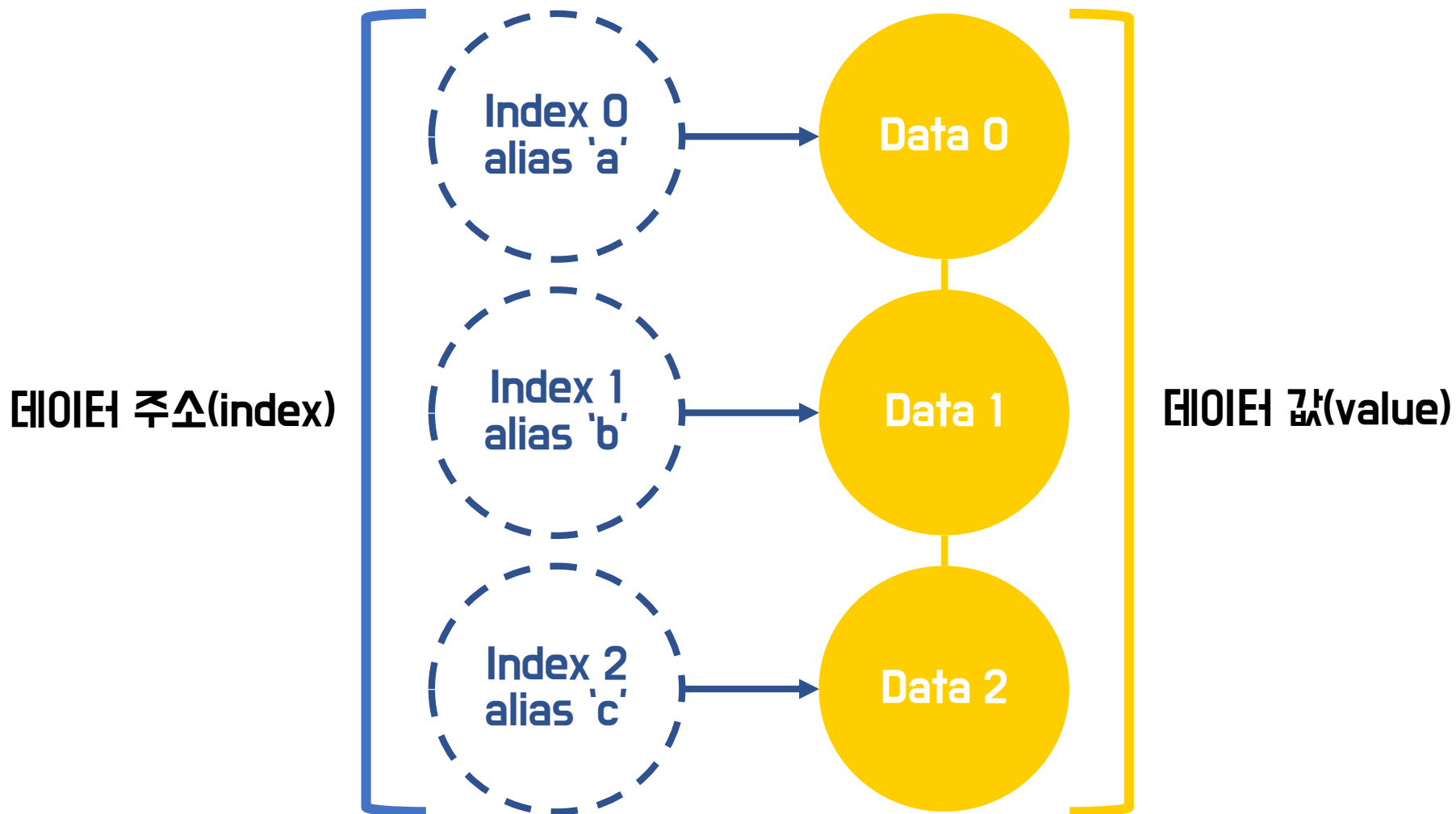


**데이터 그 자체가 가장 중요!**

**데이터분석 과정의 80%는 데이터를 다듬고  
분석에 적합한 형태로 만드는 것**



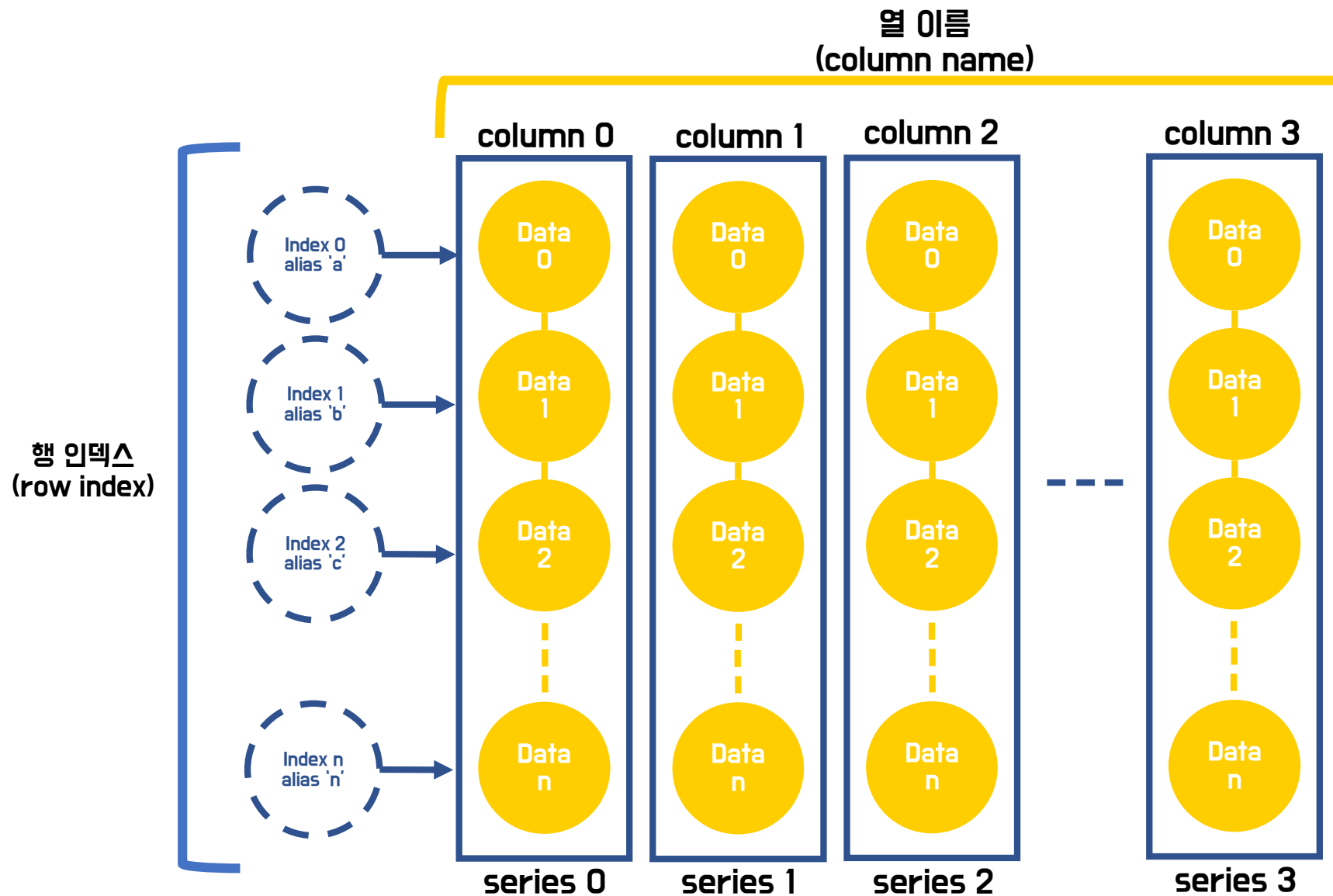
# Series



- index와 value가 일대일 대응(dictionary와 비슷한 구조)



# Dataframe



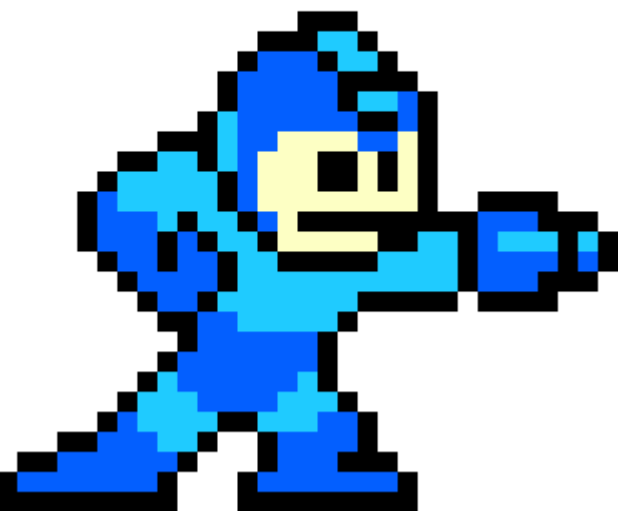


- dataframe을 만들기 위해선 같은 길이의 1차원 배열 여러 개가 필요

	나이	성별	학교
팡종	15	남	덕영중
습인	17	여	강남중
주년	19	남	상정고
선	14	여	산곡여중



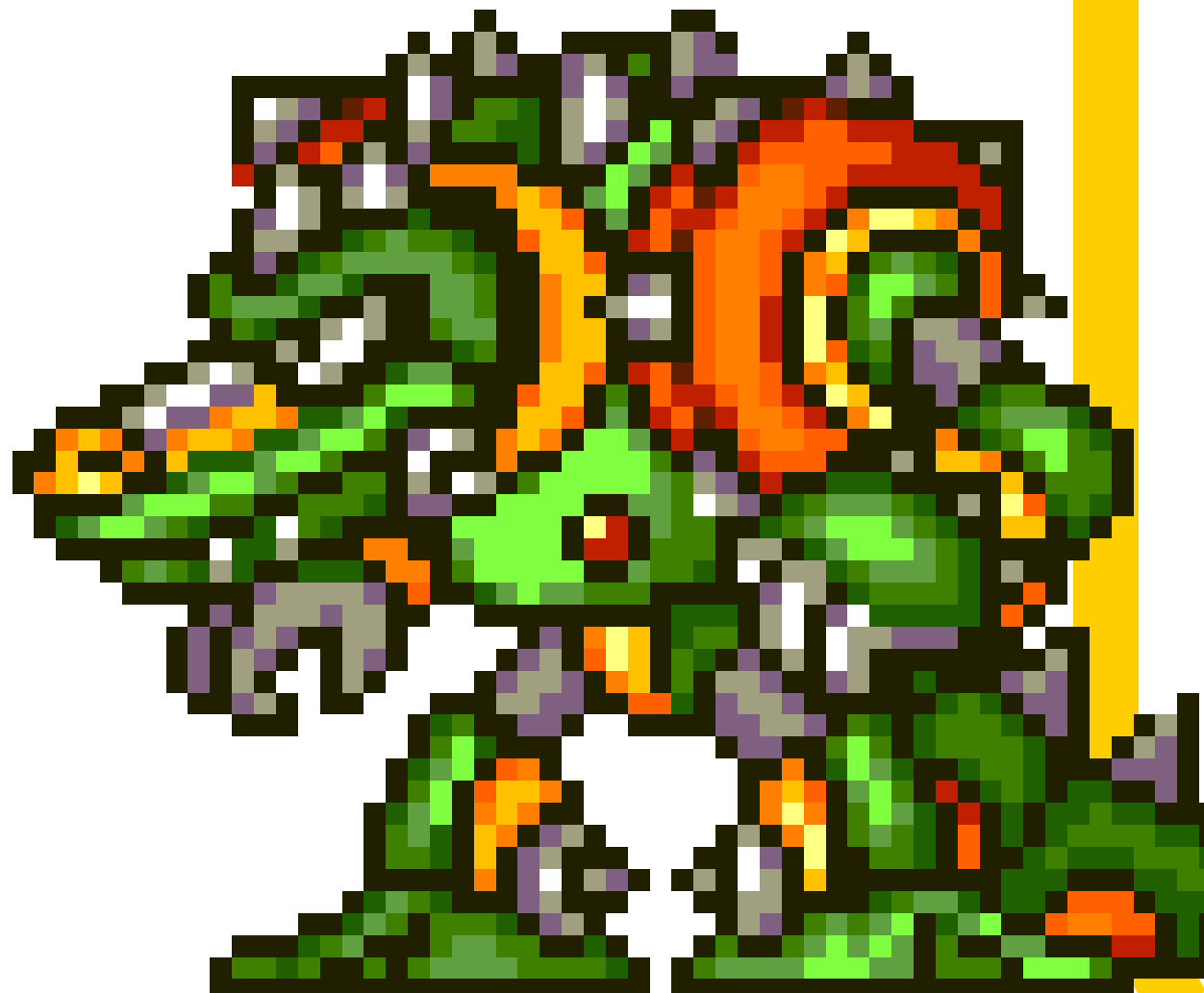
# WARNING



# WARNING

## 《 QUEST 》

첨부된 데이터셋을 가지고  
데이터 기초통계/조작 해보기



# NEXT STAGE

