



# 데이터 과학 기초

- 2021년 여름학기 -

## 2.3 Python으로 해보는 데이터 과학



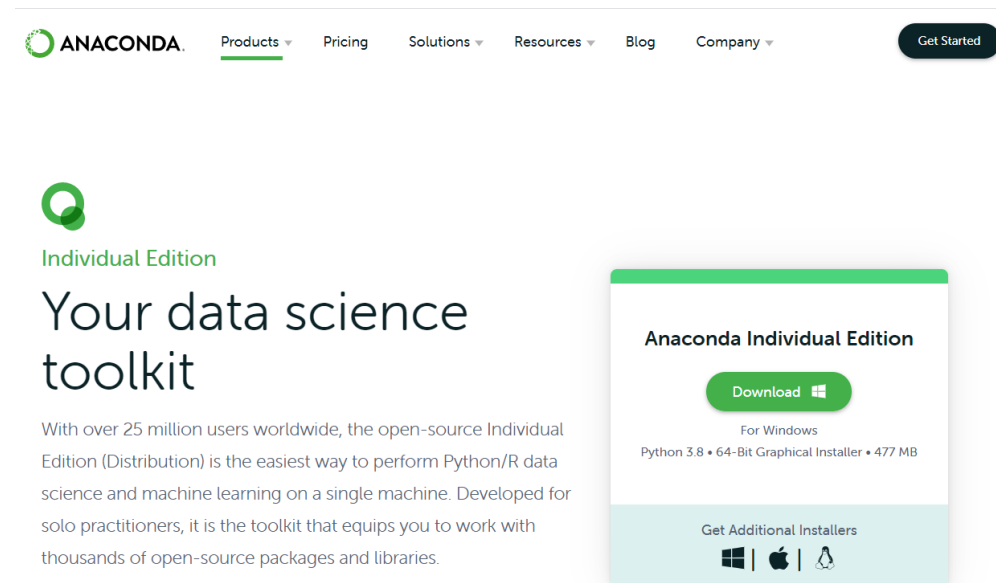
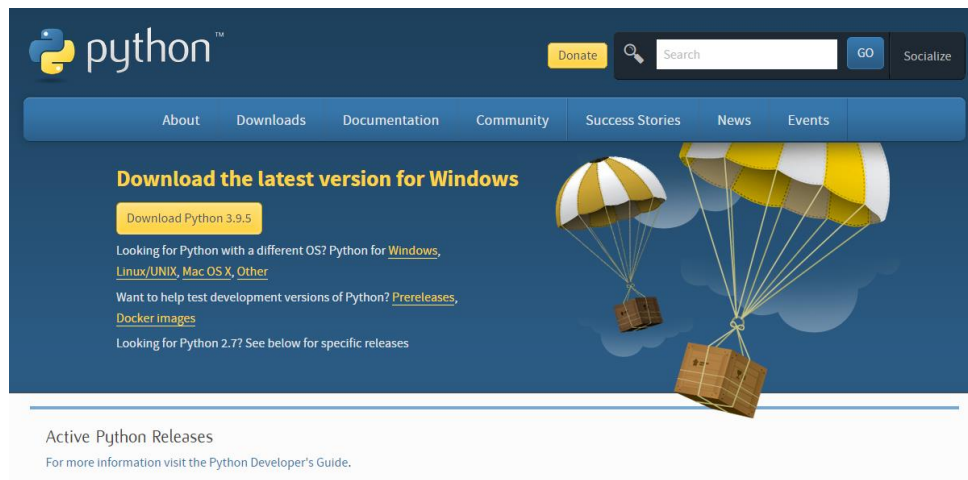


## 2.3 Python으로 해보는 데이터 과학

2

### ■ Python + Jupyter Notebook

- Python: <https://www.python.org/downloads/>
- Anaconda: <https://www.anaconda.com/products/individual>





### ■ IRIS dataset 준비:

- ScikitLearn에서 디폴트로 IRIS 데이터셋이 포함되어 있음.

```
# Load the IRIS dataset
from sklearn.datasets import load_iris
iris = load_iris()
# iris
```



### ■ 훈련용 데이터셋 준비:

```
# Prepare train data
train_features = iris.data
# train_features
train_targets = iris.target
# train_targets
```



### ■ kNN을 이용한 기계학습:

```
from sklearn.neighbors import KNeighborsClassifier

clf = KNeighborsClassifier()
clf.fit(train_features, train_targets)
# clf
```



## 2.3 Python으로 해보는 데이터 과학

### ■ 고희의 붓꽃 데이터로 테스트용 데이터 준비:

```
# Prepare test data
import pandas as pd
import numpy as np

data = np.array([[2.7, 2.4, 1.65, 0.67],
                 [5.84, 5.48, 3, 2.16],
                 [3.97, 4.01, 1.7, 0.67]])
df = pd.DataFrame(data, columns=iris.feature_names)
df
```



### ■ 분류기 적용 및 결과 확인:

```
# Classification with the k-Nearest-Neighbor  
pred = clf.predict(df)  
df['species'] = pred  
df  
  
print(iris['target_names'])  
iris['target_names'][0]
```

*Any Questions?*

