

추세선 그리는 명령어: np.polyfit()

```
import matplotlib.pyplot as plt
import numpy as np

# 예시 데이터
height = [160, 162, 165, 168, 170, 172, 175, 178, 180, 182]
weight = [55, 57, 70, 72, 65, 68, 70, 73, 83, 88]

# 추세선을 위한 다항식 회귀
degree = 1 # 1차 다항식 (직선)
coefficients = np.polyfit(height, weight, degree) # ax + b
polynomial = np.poly1d(coefficients) # coefficients[0]x + coefficients[1]

# 추세선을 위한 x값 생성
height_range = np.linspace(min(height), max(height), 100)

# 한글 처리
plt.rc('font', family='Malgun Gothic')
plt.rc('axes', unicode_minus=False)

# 그래프 그리기
plt.scatter(height, weight, label='데이터 포인트')
plt.plot(height_range, polynomial(height_range), color='red', label='추세선')

# 그래프에 레이블 추가
plt.xlabel('키 (cm)')
plt.ylabel('몸무게 (kg)')
plt.title('키와 몸무게의 추세선')
plt.legend()

# 그래프 표시
plt.show()
```

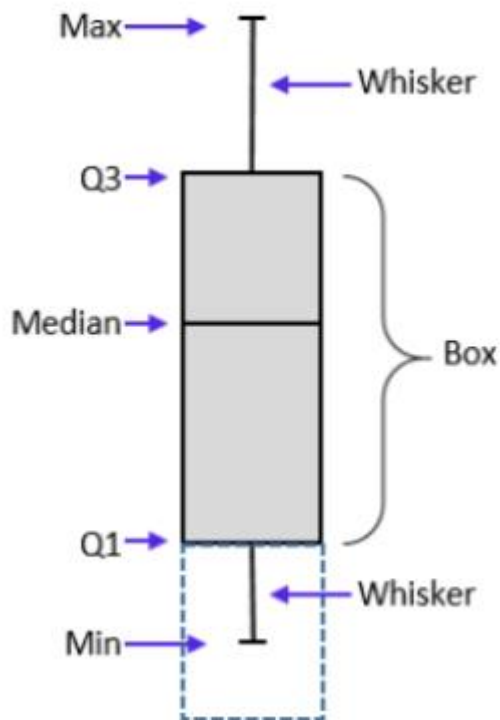
추세선은 (1)차 함수이다.

Polyfit 은 $y = ax+b$ 에서 무엇을 만들어 주나? A, b

```
import csv
import matplotlib.pyplot as plt
f = open('seoul.csv', encoding='cp949')
data = csv.reader(f)
next(data)

day = [[] for i in range(31)]
for row in data :
    if row[-1] != '' :
        if row[0].split('-')[1] == '01':
            day[int(row[0].split('-')[2])-1].append(float(row[-1]))

plt.figure(figsize=(10, 5), dpi=300)
plt.boxplot(day, showfliers=False)
plt.show()
```



```
import matplotlib.pyplot as plt

explode = (0, 0.1, 0, 0) # only "explode" the 2nd slice (i.e. 'Hogs')

labels = 'Frogs', 'Hogs', 'Dogs', 'Logs'
sizes = [15, 30, 45, 10]

fig, ax = plt.subplots()
ax.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%',
      shadow=True, startangle=90)
plt.show()
```

```
import random
import matplotlib.pyplot as plt

x = []
y = []
size = []

for i in range(100):
    x.append(random.randint(50,100))
    y.append(random.randint(50,100))
    size.append(random.randint(10,100))

plt.scatter(x, y, s=size, c=x, cmap='jet')
plt.colorbar()
plt.grid(True)
plt.show()
```

해당 코드에서 색상은 무엇을 기준으로 바뀌나요? X값

```

import csv
import matplotlib.pyplot as plt

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next(data)

male = []
female = []

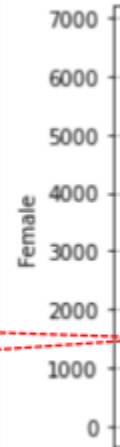
name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            male.append(int(row[i].replace(',', '')))
            female.append(int(row[i + 103].replace(',', '')))
        break

f.close()

plt.scatter(male, female, c=range(len(male)), cmap='jet', alpha=0.7)
plt.colorbar()
plt.plot(range(max(male)), range(max(male)), 'g')
plt.xlabel('Male')
plt.ylabel('Female')
plt.grid(True)
plt.show()

```

궁금한 동네



위 코드에서 c는 무엇을 뜻하나요? 나이

```

import csv
import matplotlib.pyplot as plt
import math

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next(data)

male = []
female = []
size = []

name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            male.append(int(row[i].replace(',', '')))
            female.append(int(row[i + 103].replace(',', '')))
            size.append(math.sqrt(int(row[i].replace(',', '')) + int(row[i + 103].replace(',', ''))))
        break

f.close()

plt.rc('font', family='Malgun Gothic')
plt.figure(figsize=(10, 5))
plt.title(name + ' 지역의 성별 인구 그래프')
plt.scatter(male, female, c=range(len(male)), cmap='jet', alpha=0.7, s=size)
plt.colorbar()
plt.plot(range(max(male)), range(max(male)), 'g')
plt.xlabel('남성 인구수')
plt.ylabel('여성 인구수')
plt.grid(True)
plt.show()

```

```
from numpy import random

print(random.rand(2, 3))
print(random.random((2, 3)))
```

✓ 0.0s

```
[[0.73816928 0.99684857 0.60786187]
 [0.14365811 0.66123217 0.09070631]]
[[0.11688829 0.70637377 0.4439654 ]
 [0.29883582 0.69668008 0.80457289]]
```

0에서 1미만의 실수인 난수 배열 생성하는 코드:

`np.random.rand(n)`

```
print(random.random(size=(2,)))
```

정규 분포에서 난수 또는 배열을 생성하는 코드:

`np.random.randn(n)`

일정 범위 내의 정수 생성하는 코드

`np.random.randint()`

`np.random.randint(low=0, high=None, size=None)`

주어진 1차원 배열에서 정해진 요소의 개수를 선택: `np.random.choice()`

주어진 범위 내의 실수인 난수를 생성: `np.random.uniform()`

배열의 합을 가로 방향 →으로 구하려면: `arr.sum(axis=1)`

배열의 합을 세로 방향 으로 구하려면: `arr.sum(axis=0)`

```
rand(d0, d1, ..., dn)
random(size=None)
randn(d0, d1, ..., dn)
randint(low=0, high=None, size=None)
choice(a, size=None, replace=True, p=None)
```

하나의 값의 명칭: 스칼라

1차원 배열의 명칭: 벡터

2차원 배열의 명칭: 행렬

데이터 프레임 sum 명령어: `df.sum(axis=1)`

Numpy sum 명령어: `np.sum(df, axis=1)`

판다스에서 1차원 배열의 명칭: Series

판다스에서 2차원 배열의 명칭: Dataframe

```
df.describe()
```

✓ 0.0s

	A	C	D
count	8.0	8.000000	8.0
mean	1.0	1.750000	3.0
std	0.0	1.224745	0.0
min	1.0	0.000000	3.0
25%	1.0	0.875000	3.0
50%	1.0	1.750000	3.0
75%	1.0	2.625000	3.0
max	1.0	3.500000	3.0

판다스 시리즈의 속성 대표적 3가지: name, values, index