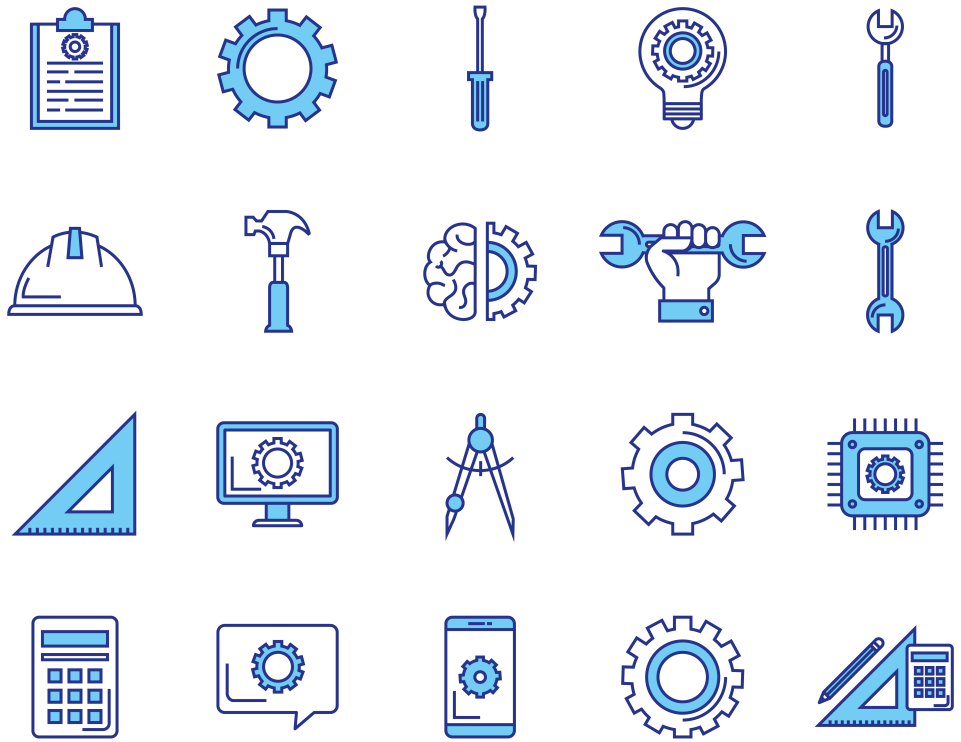
×

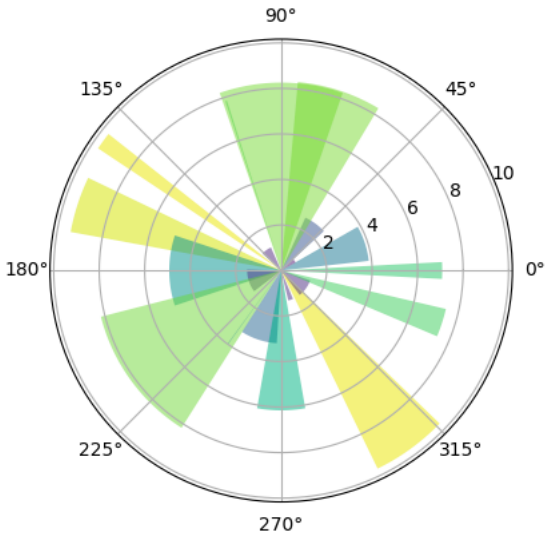
# 인슈어테크

# 파이썬 데이터 분석



이태일 강사

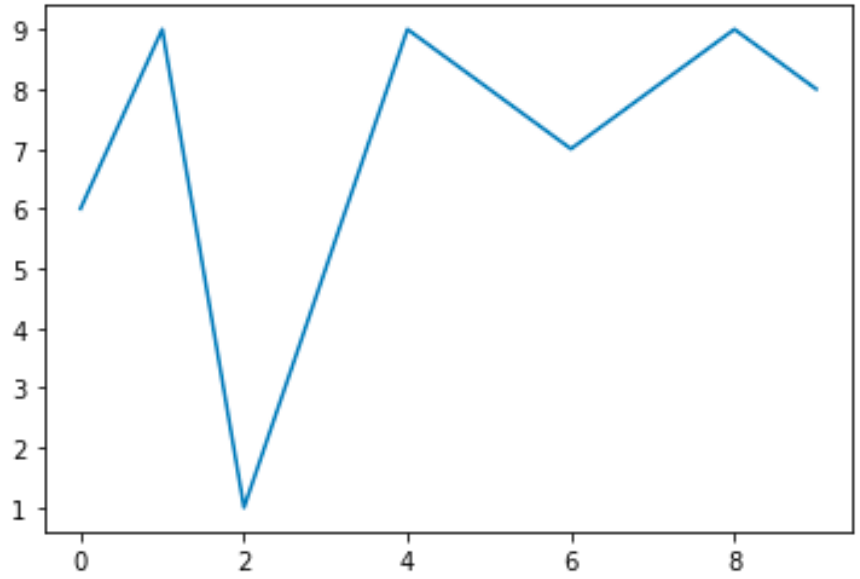
# Matplotlib




## 파이썬 데이터 시각화 라이브러리

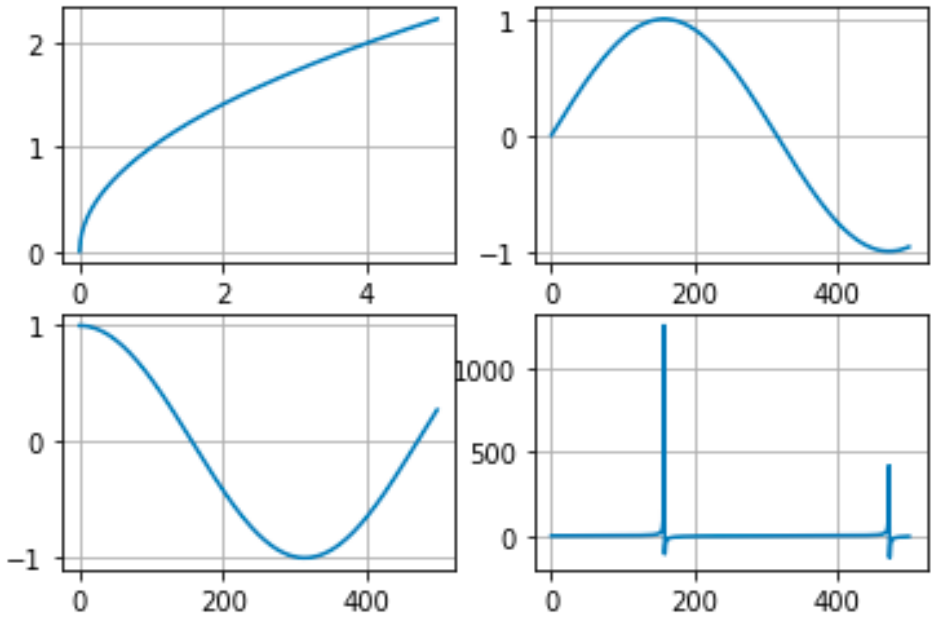
# Matplotlib

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

s1 = pd.Series([np.random.randint(1,10) for _ in range(10)])

plt.plot(s1)
```



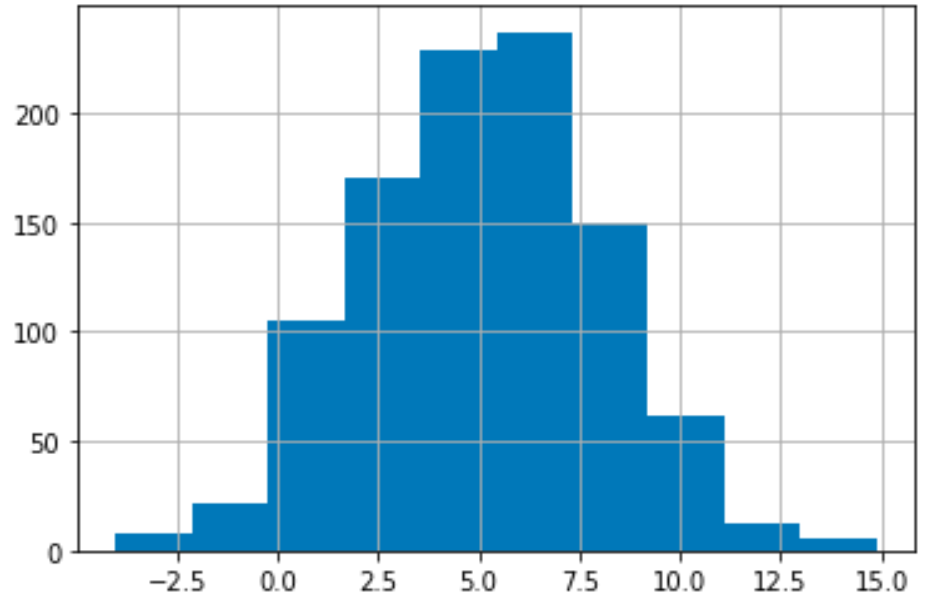
# Matplotlib

```
t=np.arange(0,5,0.01)
sqrt = np.sqrt(t)

ax1 = plt.subplot(2,2,1)
ax1.plot(t,sqrt)
ax1.grid()

ax2 = plt.subplot(2,2,2)
ax2.plot(np.sin(t))
ax2.grid()

ax3 = plt.subplot(2,2,3)
ax3.plot(np.cos(t))
ax3.grid()

ax4 = plt.subplot(2,2,4)
ax4.plot(np.tan(t))
ax4.grid()
```



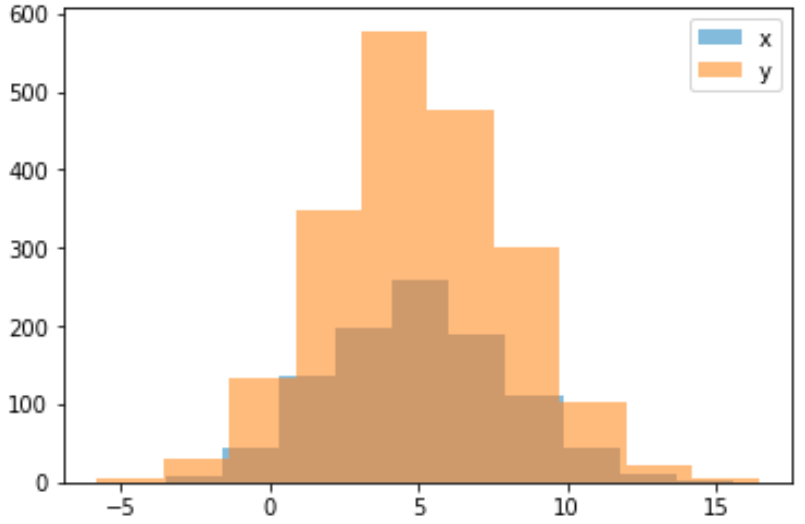
# Matplotlib

```
gauss = np.random.normal(5,3,1000)
plt.hist(gauss)
plt.grid()
```



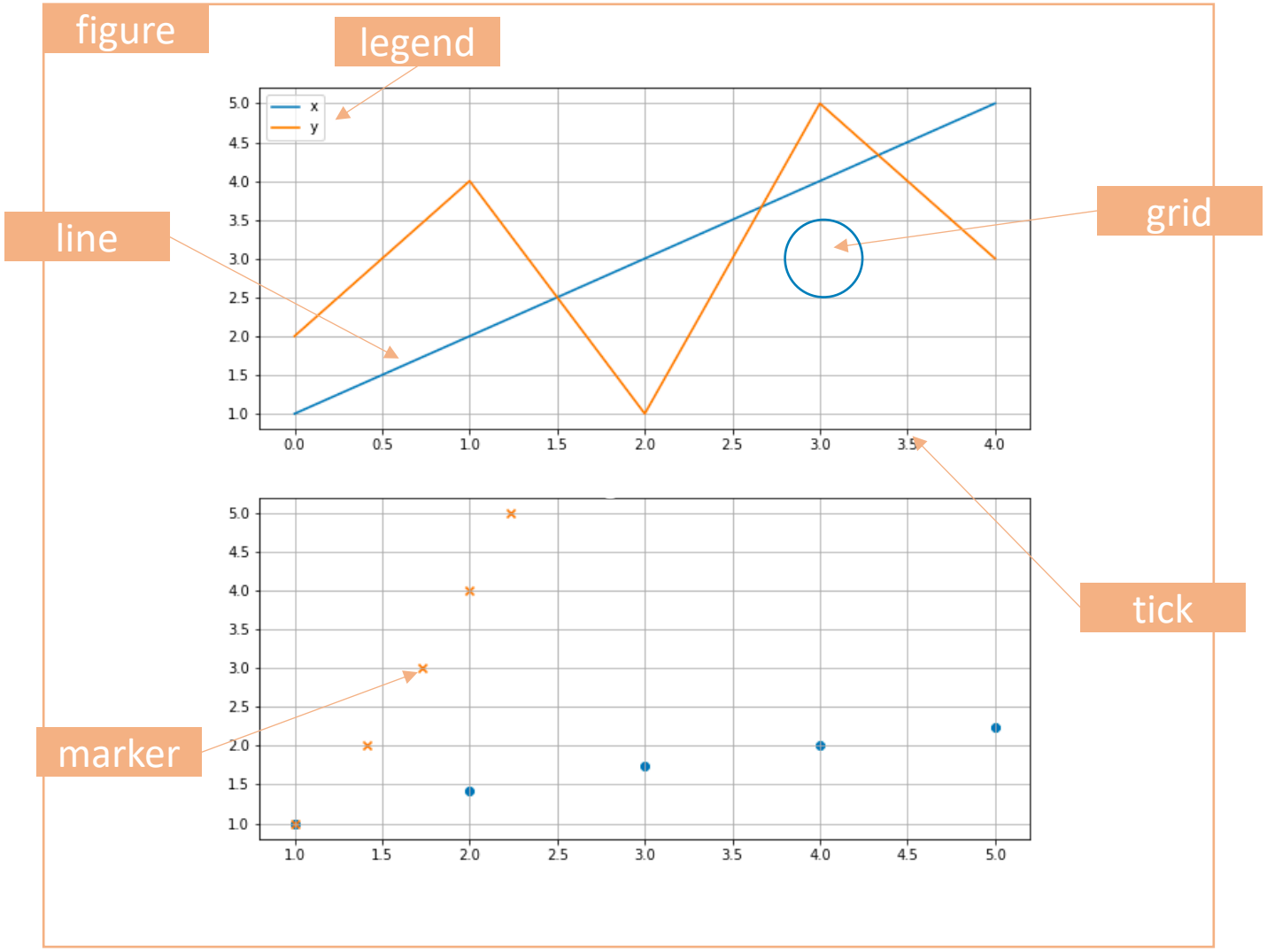
# Matplotlib

```
data1 = np.random.normal(5,3,1000)
data2 = np.random.normal(5,3,2000)
plt.hist(data1,alpha=0.5, label='x')
plt.hist(data2,alpha=0.5, label='y')
plt.legend(loc='upper right’)
```




가우스 분포 -> 히스토그램 비교

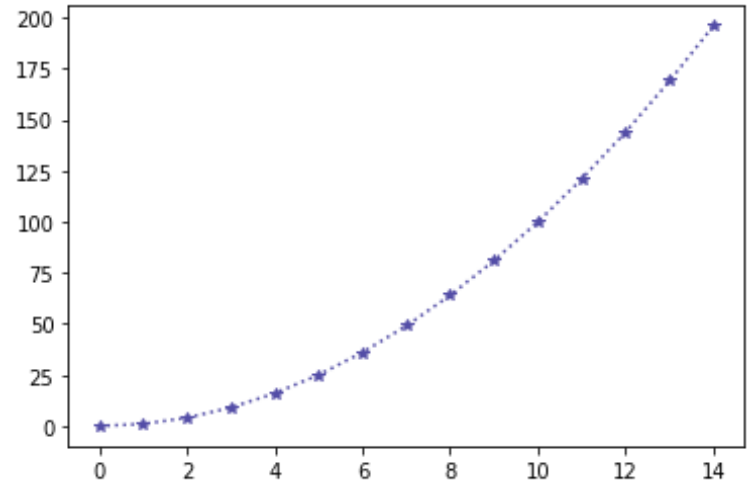
# Matplotlib



# Matplotlib

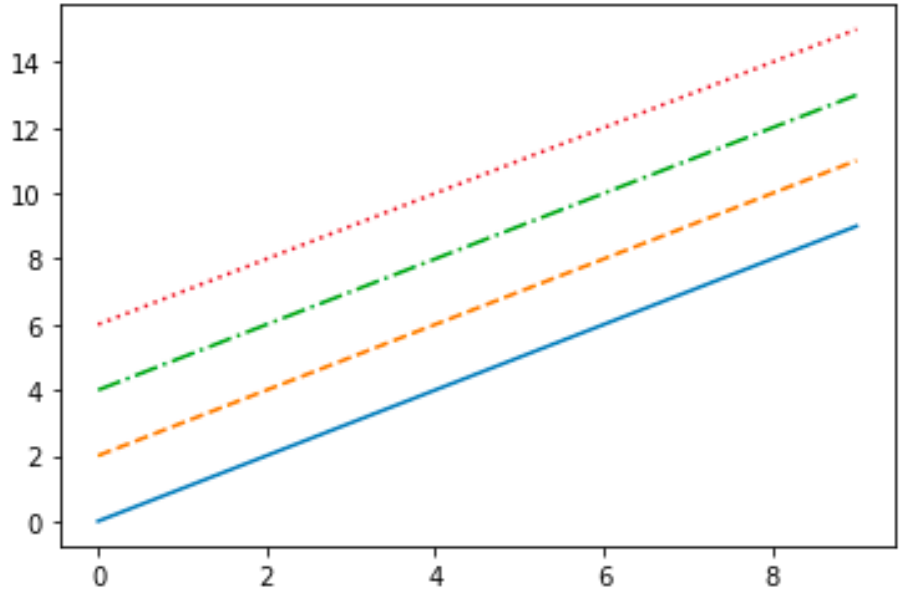
```
fig, ax = plt.subplots()
x = np.arange(15)
y = x ** 2
ax.plot(x, y, linestyle=":", marker="*", color="#524FA1")
```

라인 plot 스타일 변경



# Matplotlib

```
x = np.arange(10)
fig, ax = plt.subplots()
ax.plot(x, x, linestyle="-") # solid
ax.plot(x, x+2, linestyle="--") # dashed
ax.plot(x, x+4, linestyle="-.") # dashdot
ax.plot(x, x+6, linestyle=":") # dotted
```

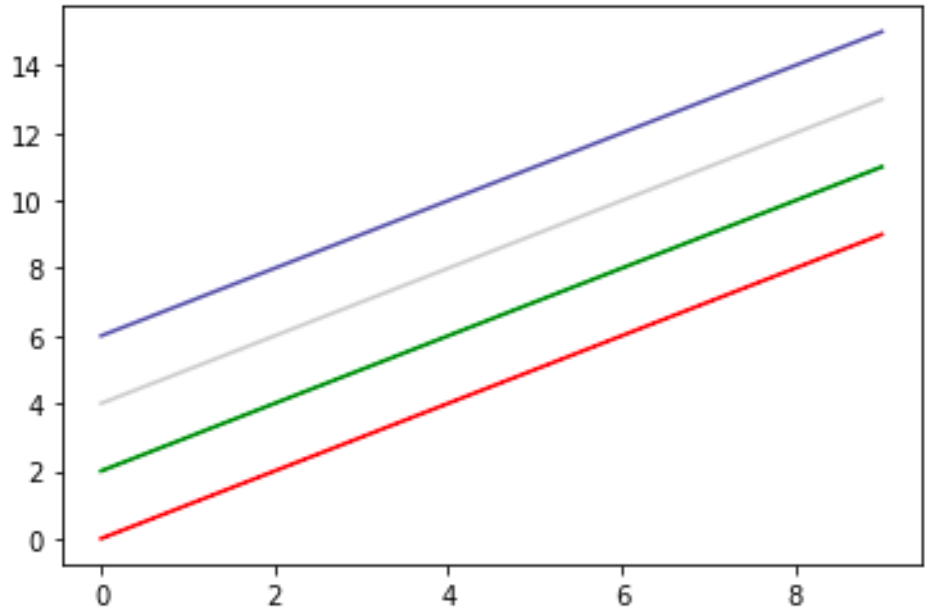



라인 plot 스타일 변경

# Matplotlib

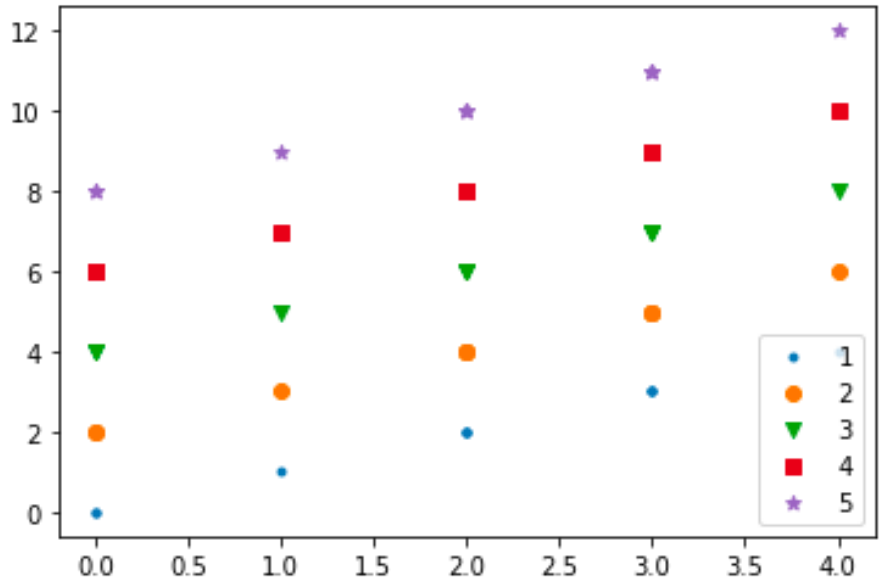
```
x = np.arange(10)
fig, ax = plt.subplots()
ax.plot(x, x, color="r")
ax.plot(x, x+2, color="green")
ax.plot(x, x+4, color='0.8')
ax.plot(x, x+6, color="#524FA1")
```

라인 plot 스타일 변경



# Matplotlib

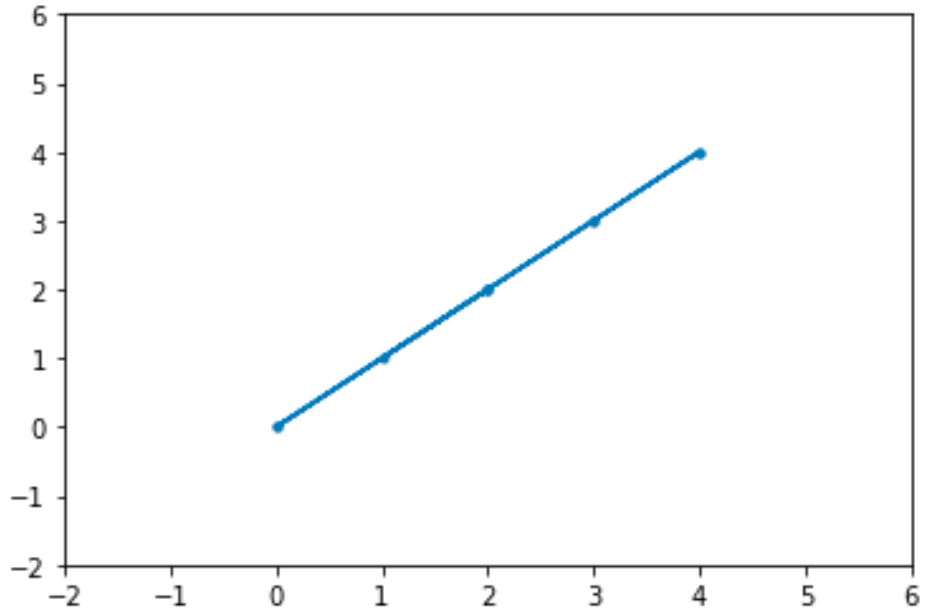
```
x = np.random.randint(0,5,10)
fig, ax = plt.subplots()
ax.scatter(x, x, marker=".",label="1")
ax.scatter(x, x+2, marker="o",label="2")
ax.scatter(x, x+4, marker='v',label="3")
ax.scatter(x, x+6, marker="s",label="4")
ax.scatter(x, x+8, marker="*",label="5")
ax.legend()
```




마커 스타일 변경 및 범례 추가

# Matplotlib

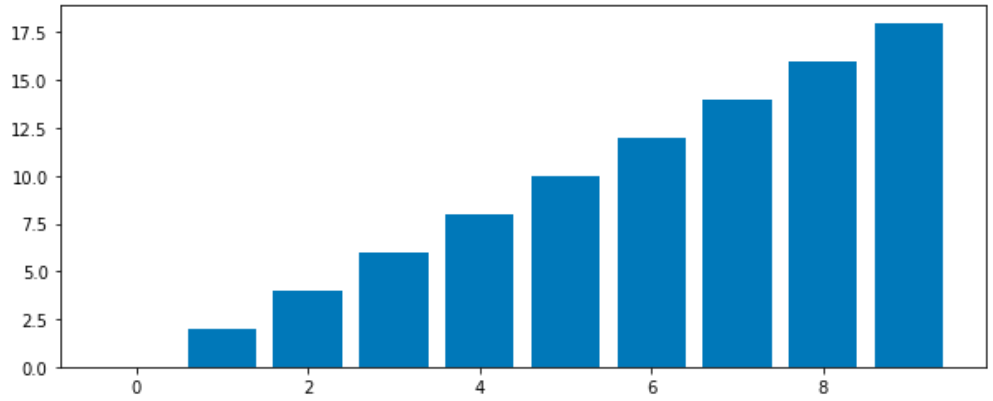
```
fig, ax = plt.subplots()
ax.plot(x, x, marker=".",label="1")
ax.set_xlim(-2,6)
ax.set_ylim(-2,6)
```




xlim == x범주 설정
ylim == y범주 설정

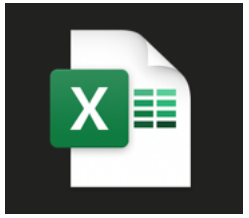
# Matplotlib

```
x = np.arange(10)
fig, ax = plt.subplots(figsize=(10, 4))
ax.bar(x, x*2)
```



# Matplotlib

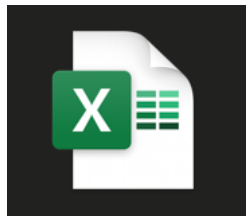
talks.csv



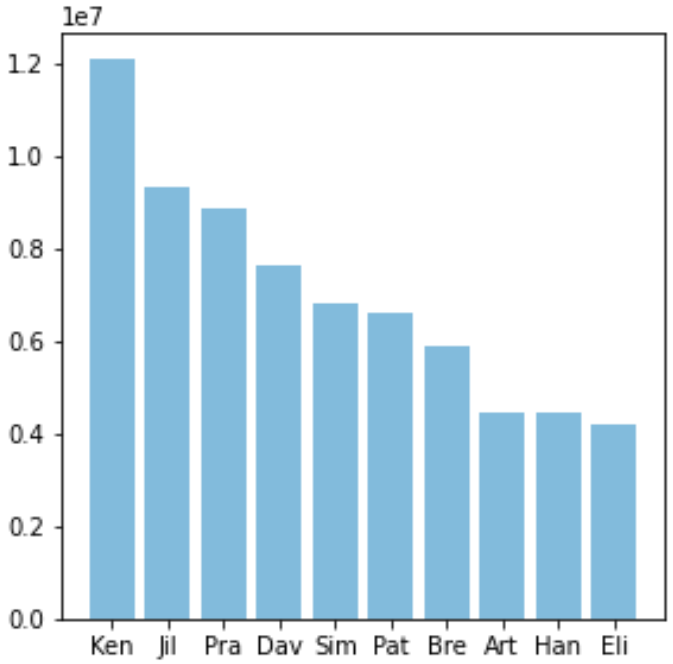
1.TED에서 가장 조회 수가 많은 스피커를 bar plot(단독 영상 기준)
2.TED에서 가장 조회 수가 많은 스피커를 bar plot(누적)
3. 비교 가능하도록 subplot을 이용해 나란히 출력

```
df = pd.read_csv("/content/drive/My Drive/TED/talks.csv")
new_df = df["speaker"].apply(lambda x : x[:3])
df["abr"] = new_df
df = df.sort_values("views",ascending=False)
plt.figure(figsize=(10,10))

summ = df.groupby("abr").sum()["views"].sort_values(ascending=False)

ax1 = plt.subplot(221)
ax1.bar(df['abr'].head(10),df['views'].head(10),alpha=0.5)

ax2 = plt.subplot(222)
ax2.bar(list(summ.index)[:10],summ.head(10),alpha=0.5, color="r")
```

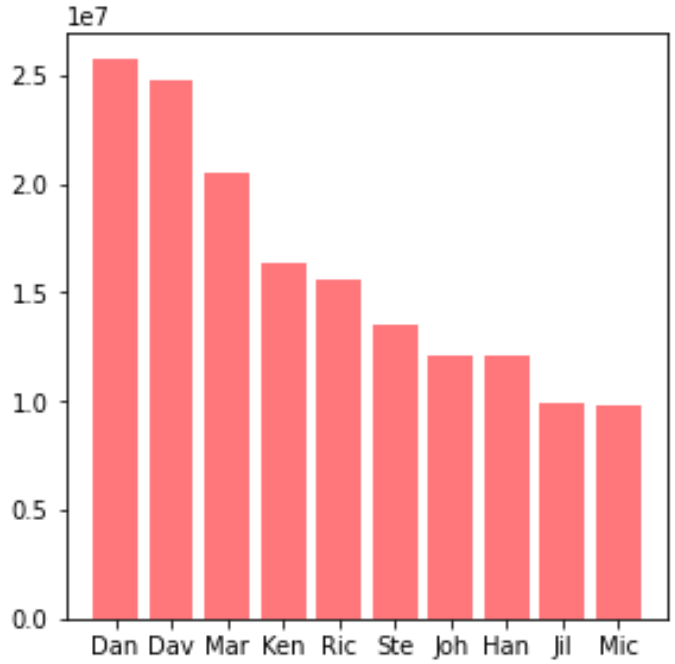
# Matplotlib



talks.csv




단독 영상 기준 조회 수




누적 영상 조회 수

# Q&A