

파이썬 프로그래밍

Matplotlib을 통해 데이터 시각화하기



광운대학교
KwangWoon University

목차

1. Matplot 복습하기

2. Matplot 활용하기

Matplotlib이란?

- Matplotlib은 파이썬에서 데이터 시각화를 위해 가장 널리 사용되는 라이브러리 중 하나
- 데이터를 시각적으로 표현하여 더 쉽게 이해할 수 있도록 도와준다

선 그래프(Line Plot)

```
import matplotlib.pyplot as plt

categories = ['A', 'B', 'C', 'D']
values = [10, 20, 15, 25]

plt.bar(categories, values)
plt.title('Bar Plot Example')
plt.xlabel('Category')
plt.ylabel('Value')
plt.show()
```

막대 그래프(Bar Plot)

```
import matplotlib.pyplot as plt

years = [2017, 2018, 2019, 2020, 2021]
values = [100, 200, 300, 400, 500]

plt.plot(years, values)
plt.title('Line Plot Example')
plt.xlabel('Year')
plt.ylabel('Value')
plt.show()
```

산점도(Scatter Plot)

```
import matplotlib.pyplot as plt

# 예제 데이터
x = [1, 2, 3, 4, 5]
y = [5, 7, 8, 5, 6]

plt.scatter(x, y)
plt.title('Scatter Plot Example')
plt.xlabel('X')
plt.ylabel('Y')
plt.show()
```

히스토그램(Histogram)

```
import matplotlib.pyplot as plt

# 예제 데이터
data = [1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5]

plt.hist(data, bins=5)
plt.title('Histogram Example')
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.show()
```

파이 차트(Pie Chart)

```
import matplotlib.pyplot as plt

# 예제 데이터
labels = ['A', 'B', 'C', 'D']
sizes = [20, 30, 25, 25]

plt.pie(sizes, labels=labels, autopct='%1.1f%%')
plt.title('Pie Chart Example')
plt.show()
```


Matplotlib 활용하기

- python seaborn 라이브러리에서 제공하는 교육용 데이터를 다운

```
pip install seaborn
```

타이타닉 데이터 불러오기

```
import seaborn as sns
import matplotlib.pyplot as plt

# Seaborn의 Titanic 데이터셋 불러오기
titanic = sns.load_dataset('titanic')

# 컬럼명을 한국어로 변경
titanic.columns = [
    '생존여부', '승객_등급', '성별', '나이', '형제_자매_배우자_수', '부모_자녀_수',
    '요금', '출발_항구', '승객_등급_명', '승객_구분', '성인_남성', '갑판', '출발_도시',
    '생존여부_명', '단독여부'
]

# 데이터셋의 첫 5행 출력
titanic.head()
```

한글 폰트 깨질 때 사용

```
import matplotlib.pyplot as plt
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus'] = False
```

나이에 따른 평균 요금을 선 그래프로 시각화

```
import matplotlib.pyplot as plt

# 나이별 평균 요금 계산
age_fare = titanic.groupby('나이')['요금'].mean().reset_index()

plt.plot(age_fare['나이'], age_fare['요금'], marker='o', linestyle='-')
plt.title('나이에 따른 평균 요금')
plt.xlabel('나이')
plt.ylabel('요금')
plt.grid(True)
plt.show()
```

각 승객 등급별 생존자 수 막대그래프 시각화

```
import matplotlib.pyplot as plt

# 승객 등급별 생존자 수 계산
class_survived = titanic.groupby('승객_등급')['생존여부'].sum().reset_index()

plt.bar(class_survived['승객_등급'], class_survived['생존여부'], color='skyblue')
plt.title('승객 등급별 생존자 수')
plt.xlabel('승객 등급')
plt.ylabel('생존자 수')
plt.show()
```

나이와 요금 간의 관계를 산점도로 시각화

```
import matplotlib.pyplot as plt
```

나이 분포를 히스토그램으로 시각화

```
import matplotlib.pyplot as plt

plt.hist(titanic['나이'].dropna(), bins=5, color='purple', edgecolor='black')
plt.title('나이 분포')
plt.xlabel('나이')
plt.ylabel('빈도')
plt.show()
```

생존 여부에 따른 비율을 파이 차트로 시각화

```
import matplotlib.pyplot as plt

# 생존 여부 비율 계산
survived_counts = titanic['생존여부'].value_counts()

plt.pie(survived_counts, labels=['생존하지 않음', '생존'], autopct='%1.1f%%', startangle=140)
plt.title('생존 비율')
plt.axis('equal')
plt.show()
```

산점도 심화(3D 산점도)

- 나이.요금 생존 여부에 대한 산점도

```
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
# 3D 산점도
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

# 생존 여부에 따른 색상 설정
colors = {0: 'red', 1: 'green'}
titanic['색상'] = titanic['생존여부'].map(colors)
ax.scatter(titanic['나이'], titanic['요금'], titanic['생존여부'], c=titanic['색상'])
ax.set_title('나이, 요금 및 생존 여부에 따른 3D 산점도')
ax.set_xlabel('나이')
ax.set_ylabel('요금')
ax.set_zlabel('생존 여부')
```


Thank you

Q / A