



[LAB-06] 5. 데이터 분포 시각화 (3)



#01. 준비작업



[1] 패키지 참조

```
from hossam import load_data
from matplotlib import pyplot as plt
from matplotlib import font_manager as fm
from pandas import pivot_table
import seaborn as sb
```



[2] 그래프 초기화

```
my_dpi = 200 # 이미지 선명도(100~300)
fpath = "./NotoSansKR-Regular.ttf" # 한글을 지원하는 폰트 파일의 경로
fm.fontManager.addfont(fpath) # 폰트의 글꼴을 시스템에 등록함
fprop = fm.FontProperties(fname=fpath) # 폰트의 속성을 읽어옴
fname = fprop.get_name() # 읽어온 속성에서 폰트의 이름만 추출
plt.rcParams['font.family'] = fname # 그래프에 한글 폰트 적용
plt.rcParams['font.size'] = 6 # 기본 폰트 크기
plt.rcParams['axes.unicode_minus'] = False # 그래프에 마이너스 깨짐 방지
```



[3] 데이터 가져오기

```
origin = load_data('flights')
origin
```

```
[94m[data] [0m https://data.hossam.kr/data/lab06/flights.xlsx
[94m[desc] [0m 어느 항공사의 년/월별 국제선 탑승객 수(출처: seaborn 내장 데이터)
```

field	description
-----	-----
year	항공 승객 수가 집계된 연도
month	항공 승객 수가 집계된 월
passengers	해당 연도/월의 국제선 항공 승객 수

	year	month	passengers
0	1949	January	112
1	1949	February	118
2	1949	March	132
3	1949	April	129
4	1949	May	121
...
139	1960	August	606
140	1960	September	508
141	1960	October	461
142	1960	November	390
143	1960	December	432

144 rows × 3 columns



[4] 데이터 전처리

```
df = origin.copy()
df['month'] = df['month'].map({
    "January": 1, "February": 2, "March": 3, "April": 4,
    "May": 5, "June": 6, "July": 7, "August": 8,
    "September": 9, "October": 10, "November": 11, "December": 12
})

df.head()
```

	year	month	passengers
0	1949	1	112
1	1949	2	118
2	1949	3	132
3	1949	4	129
4	1949	5	121

```
df2 = pivot_table(df, index="year", columns="month",
                  values="passengers")
df2
```

month	1	2	3	4	5	6	7	8	9	10	11	12
year												
1949	112.0	118.0	132.0	129.0	121.0	135.0	148.0	148.0	136.0	119.0	104.0	118.0
1950	115.0	126.0	141.0	135.0	125.0	149.0	170.0	170.0	158.0	133.0	114.0	140.0
1951	145.0	150.0	178.0	163.0	172.0	178.0	199.0	199.0	184.0	162.0	146.0	166.0
1952	171.0	180.0	193.0	181.0	183.0	218.0	230.0	242.0	209.0	191.0	172.0	194.0
1953	196.0	196.0	236.0	235.0	229.0	243.0	264.0	272.0	237.0	211.0	180.0	201.0
1954	204.0	188.0	235.0	227.0	234.0	264.0	302.0	293.0	259.0	229.0	203.0	229.0
1955	242.0	233.0	267.0	269.0	270.0	315.0	364.0	347.0	312.0	274.0	237.0	278.0
1956	284.0	277.0	317.0	313.0	318.0	374.0	413.0	405.0	355.0	306.0	271.0	306.0
1957	315.0	301.0	356.0	348.0	355.0	422.0	465.0	467.0	404.0	347.0	305.0	336.0
1958	340.0	318.0	362.0	348.0	363.0	435.0	491.0	505.0	404.0	359.0	310.0	337.0
1959	360.0	342.0	406.0	396.0	420.0	472.0	548.0	559.0	463.0	407.0	362.0	405.0
1960	417.0	391.0	419.0	461.0	472.0	535.0	622.0	606.0	508.0	461.0	390.0	432.0

#02. Heatmap 시각화

데이터의 패턴, 특히 밀도와 분포를 빠르게 파악할 수 있게 해주는 시각적 도구.

행과 열을 가진 행렬 형태의 데이터를 색상으로 나타내어 각 셀의 값에 따라 색상이 변한다.



[1] 기본 사용 방법

성적표 데이터에 대한 점수 분포 시각화

```
# 1) 그래프 초기화
width_px    = 1000                      # 그래프 가로 크기
height_px   = 1000                      # 그래프 세로 크기
rows        = 1                          # 그래프 행 수
cols        = 1                          # 그래프 열 수
figsize     = (width_px / my_dpi, height_px / my_dpi)
fig, ax     = plt.subplots(rows, cols, figsize=figsize, dpi=my_dpi)
```

2) heatmap 그리기

```
# → annot=True : 수치값을 함께 표시함
# → fmt: annot=True가 설정된 경우 표시되는 수치값의 형식 지정
# → linewidth: 각 셀 사이의 선 굵기
```

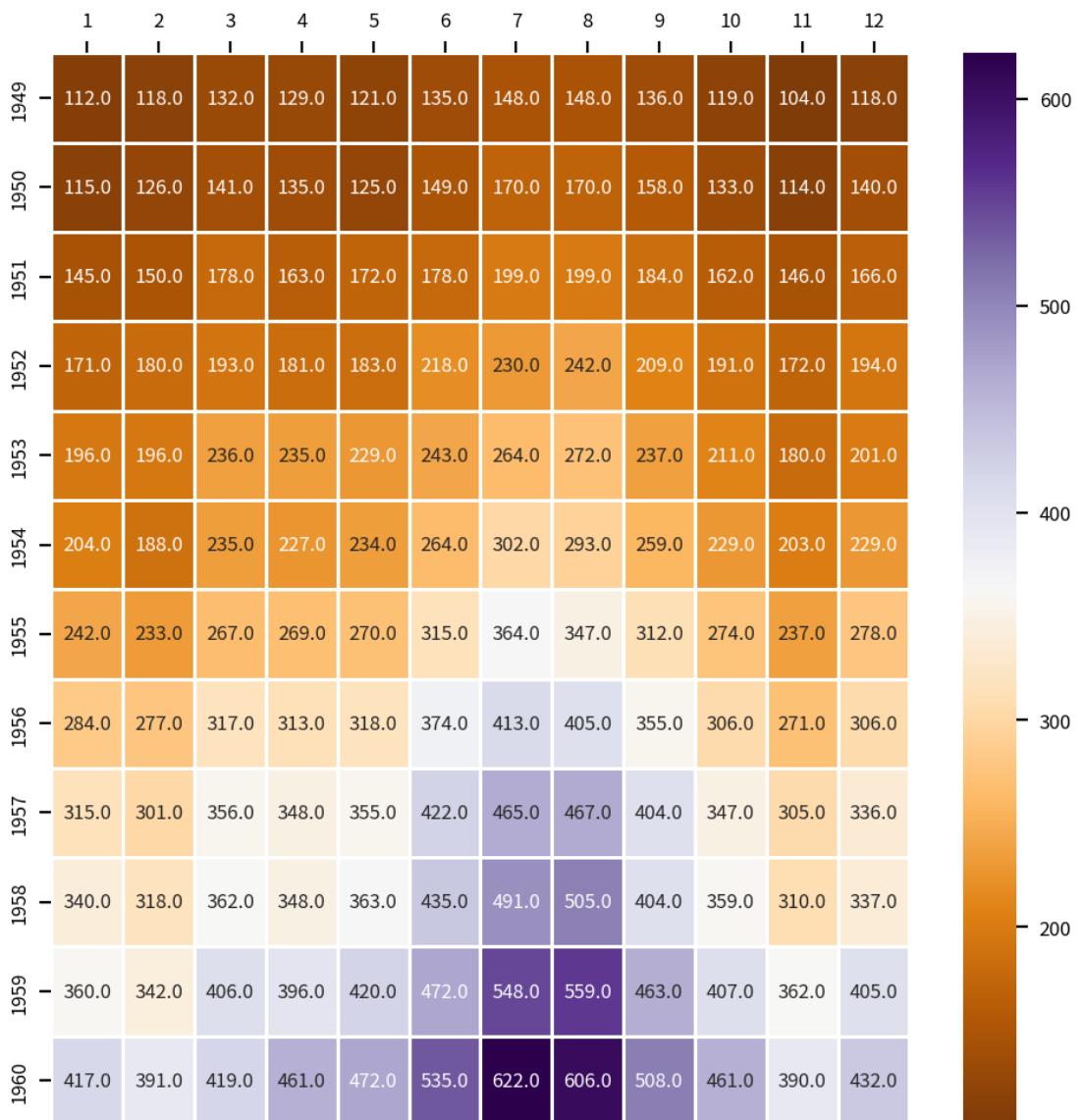
```

# → cmap: 칼라맵 ('Greys', 'Purples', 'Blues', 'Greens', 'Oranges',
#                   'Reds',
#                   'YlOrBr', 'YlOrRd', 'OrRd', 'PuRd', 'RdPu', 'BuPu',
#                   'GnBu', 'PuBu', 'YlGnBu', 'PuBuGn', 'BuGn', 'YlGn',
#                   'PiYG', 'PRGn', 'BrBG', 'PuOr', 'RdGy', 'RdBu',
#                   'RdYlBu',
#                   'RdYlGn', 'Spectral', 'coolwarm', 'bwr', 'seismic',
#                   'berlin', 'managua', 'viamo')
sb.heatmap(data=df2, annot=True, fmt="0.1f", linewidth=0.5,
            cmap="PuOr")

# 3) 그래프 꾸미기
ax.set_xlabel("")
ax.set_ylabel("")
ax.xaxis.tick_top()           # x축의 변수 이름을 상단으로 이동

# 4) 출력
plt.tight_layout()          # 여백 제거
plt.savefig("plot.png", dpi=my_dpi * 2)
plt.show()                  # 그래프 화면 출력
plt.close()                 # 그래프 작업 종료

```



png

- 여름(6-8월)에 진한 색 → 승객 수가 가장 많음
- 연도가 증가할수록 전체 색 농도가 짙어짐 → 항공 운송 수요 증가 추세
- 연도별 패턴은 거의 동일하지만 크기만 증가 → 계절성 + 추세 구조가 혼합된 전형적 시계열 패턴