

빅데이터

Ali

발표 소개

- 본 발표에서는 최근 5년 간의 교통사고 빅데이터를 한 눈에 이해할 수 있게 시각화
- 교통사고분석시스템 TASS에서 제공하는 연도별 교통사고정보 데이터를 수집해 시각화 및 분석
- 분석에는 2017년부터 2021년, 즉 최근 5년간의 데이터를 사용
- 분석하며 수업시간에 배운 다양한 그래프 그리기 라이브러리 및 시각화 방법들을 사용함

최근 5년간 요일별 교통사고 추세

- 금요일에 교통사고가 가장 많이 발생, 가장 적게 발생하는 요일은 일요일
- 각 연도마다 그 양상이 유사하게 나타나는 것을 볼 수 있음

요일별

```
dir_path = 'C:/python/교통사고_데이터.xlsx'  
df = pd.read_excel(dir_path, sheet_name='요일별', index_col=0)
```

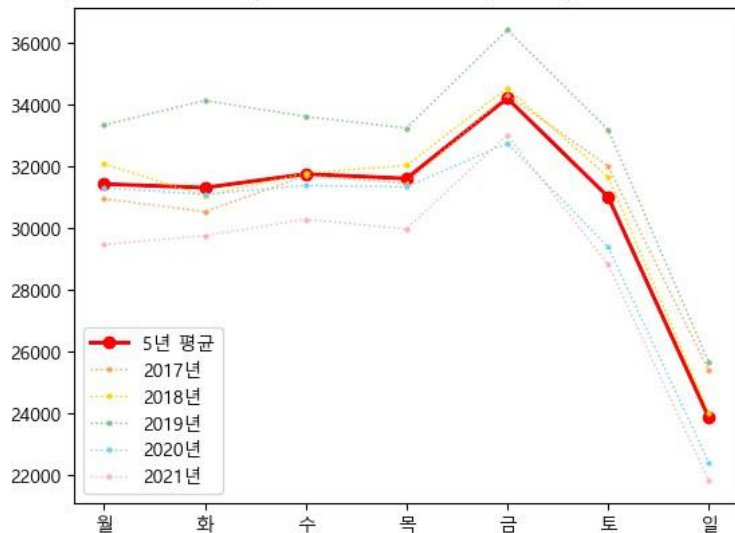
```
x = [1,2,3,4,5,6,7]
```

```
y_2017 = df.iloc[:, 0]  
y_2018 = df.iloc[:, 1]  
y_2019 = df.iloc[:, 2]  
y_2020 = df.iloc[:, 3]  
y_2021 = df.iloc[:, 4]  
y_avg = (y_2017 + y_2018 + y_2019 + y_2020 + y_2021) / 5
```

```
plt.plot(x, y_avg, marker='o', markersize=6, color='red', linewidth=2, label='5년 평균')  
plt.plot(x, y_2017, marker='o', ls=':', markersize=2, color='sandybrown', linewidth=1, label='2017년')  
plt.plot(x, y_2018, marker='o', ls=':', markersize=2, color='gold', linewidth=1, label='2018년')  
plt.plot(x, y_2019, marker='o', ls=':', markersize=2, color='darkseagreen', linewidth=1, label='2019년')  
plt.plot(x, y_2020, marker='o', ls=':', markersize=2, color='lightskyblue', linewidth=1, label='2020년')  
plt.plot(x, y_2021, marker='o', ls=':', markersize=2, color='lightpink', linewidth=1, label='2021년')
```

```
plt.title('최근 5년간 요일별 교통사고 추세')  
plt.xticks(x, ['월', '화', '수', '목', '금', '토', '일'])  
plt.legend()  
plt.show()
```

최근 5년간 요일별 교통사고 추세



최근 5년간 시간대별 교통사고 추세

- 교통량이 많은 오후 시간대에는 사고 발생 수가 꾸준히 증가하며 혼잡한 퇴근 시간대인 18~20시에 가장 높은 수치를 보임
- 각 연도마다 그 양상이 유사하게 나타남

시간별

```
dir_path = 'C:/python/교통사고_데이터.xlsx'
df = pd.read_excel(dir_path, sheet_name='시간별', index_col=0)

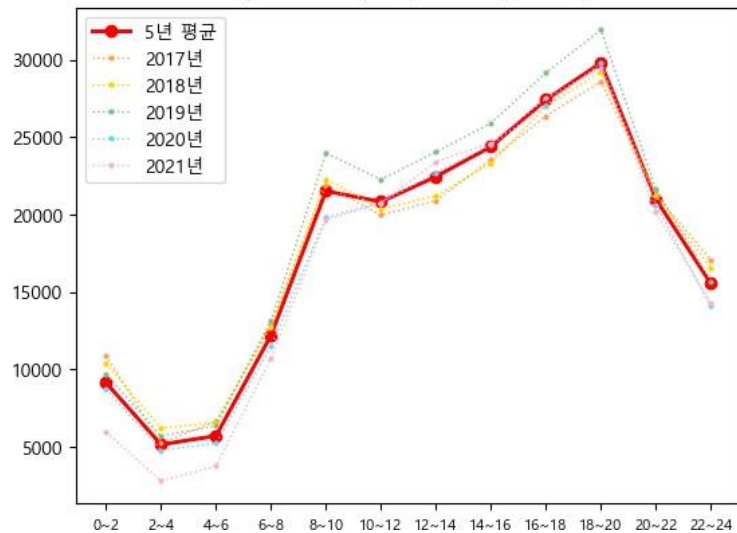
x = [1,2,3,4,5,6,7,8,9,10,11,12]

y_2017 = df.iloc[:, 0]
y_2018 = df.iloc[:, 1]
y_2019 = df.iloc[:, 2]
y_2020 = df.iloc[:, 3]
y_2021 = df.iloc[:, 4]
y_avg = (y_2017 + y_2018 + y_2019 + y_2020 + y_2021) / 5

plt.plot(x, y_avg, marker='o', markersize=6, color='red', linewidth=2, label='5년 평균')
plt.plot(x, y_2017, marker='o', ls=':', markersize=2, color='sandybrown', linewidth=1, label='2017년')
plt.plot(x, y_2018, marker='o', ls=':', markersize=2, color='gold', linewidth=1, label='2018년')
plt.plot(x, y_2019, marker='o', ls=':', markersize=2, color='darkseagreen', linewidth=1, label='2019년')
plt.plot(x, y_2020, marker='o', ls=':', markersize=2, color='lightskyblue', linewidth=1, label='2020년')
plt.plot(x, y_2021, marker='o', ls=':', markersize=2, color='lightpink', linewidth=1, label='2021년')

plt.title('최근 5년간 시간대별 교통사고 추세')
plt.xticks(x, ['0~2', '2~4', '4~6', '6~8', '8~10', '10~12', '12~14', '14~16', '16~18', '18~20', '20~22', '22~24'], fontsize=8)
plt.legend()
plt.show()
```

최근 5년간 시간대별 교통사고 추세



최근 5년간 법규 위반 사유별 교통사고 추세

- 안전운전 의무 불이행으로 인한 사고 발생 수가 현저히 많고, 그 다음으로는 신호위반으로 인한 사고 수가 많이 나타남
- 각 연도마다 그래프의 양상이 유사하게 나타남

법규 위반 사유별

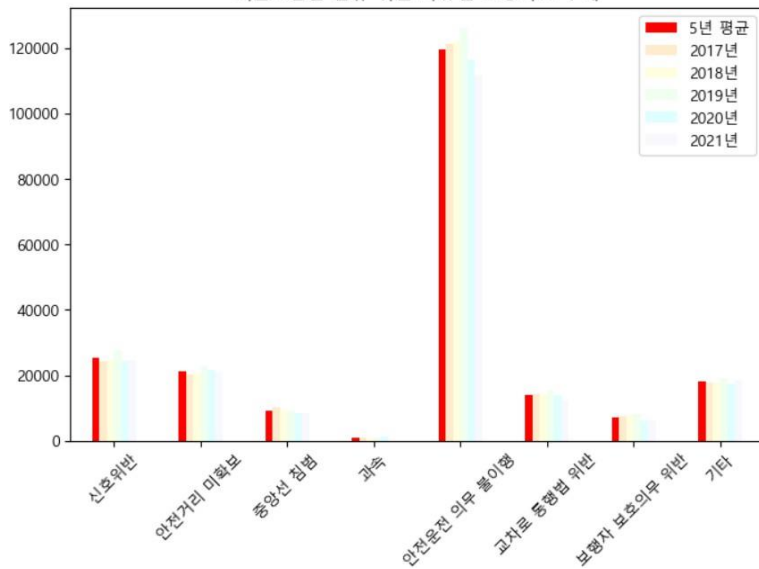
```
dir_path = 'C:/python/교통사고_데이터.xlsx'
df = pd.read_excel(dir_path, sheet_name='법규 위반 사유별', index_col=0)

x = [0, 1, 2, 3, 4, 5, 6, 7]

y_2017 = df.iloc[:, 0]
y_2018 = df.iloc[:, 1]
y_2019 = df.iloc[:, 2]
y_2020 = df.iloc[:, 3]
y_2021 = df.iloc[:, 4]
y_avg = (y_2017 + y_2018 + y_2019 + y_2020 + y_2021) / 5

y_dict = {'5년 평균': y_avg, '2017년': y_2017, '2018년': y_2018, '2019년': y_2019, '2020년': y_2020, '2021년': y_2021}
df_y = pd.DataFrame(y_dict)
df_y.plot(kind='bar', figsize=(8, 5), color=['red', 'blanchedalmond', 'lightyellow', 'honeydew', 'lightcyan', 'ghostwhite'])
plt.title('최근 5년간 법규 위반 사유별 교통사고 추세')
plt.xticks(x, ['신호위반', '안전거리 미확보', '중앙선 침범', '과속', '안전운전 의무 불이행', '교차로 통행법 위반', '보행자 보호의무 위반', '기타'], rotation=45)
plt.show()
```

최근 5년간 법규 위반 사유별 교통사고 추세

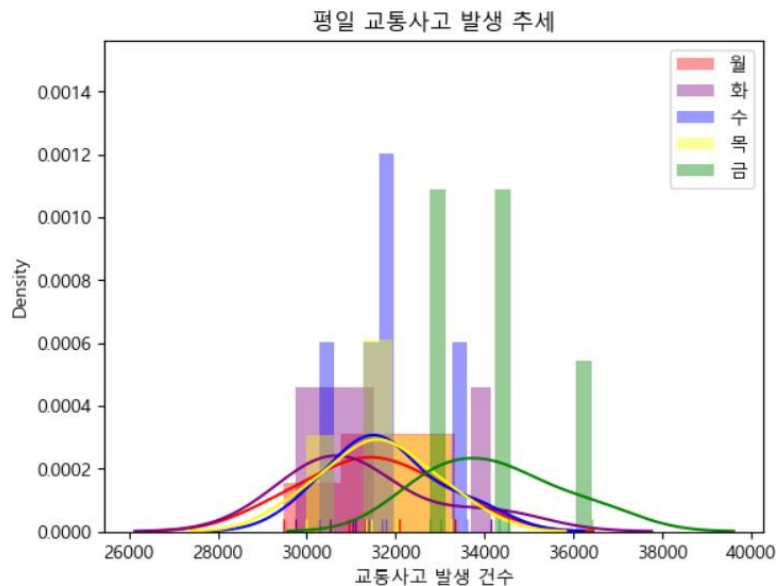


최근 5년간 평일 교통사고 발생 추세

- seaborn 라이브러리 활용
- 평일에 발생한 교통사고 발생 건수의 밀도를 측정
- 전반적으로 약 31,000건에서 가장 높은 밀도를 보임
- 금요일은 다른 요일과 밀도 분포가 다른 양상

```
dir_path = 'C:/python/교통사고_데이터.xlsx'
df = pd.read_excel(dir_path, sheet_name='요일별', index_col=0)
mon, tue, wed, thu, fri, _, _ = df.values.tolist()

plt.title('평일 교통사고 발생 추세')
plt.xlabel('교통사고 발생 건수')
sns.distplot(mon, color='red', rug=True, rug_kws={"color": "red"}, label='월')
sns.distplot(tue, color='purple', bins=10, rug=True, label='화')
sns.distplot(wed, rug=True, color='blue', bins=10, rug_kws={"color": "blue", "alpha": 0.4}, label='수')
sns.distplot(thu, color="yellow", rug=True, label='목')
sns.distplot(fri, rug=True, color='green', bins=10, rug_kws={"color": "green", "alpha": 0.4}, label='금')
plt.legend()
```



최근 5년간 주말 교통사고 발생 추세

- 주말에 발생한 교통사고 발생 건수의 밀도를 측정
- 일요일보다 토요일에 높은 발생량을 보임
- 토요일과 일요일의 밀도 분포가 다른 양상

```
dir_path = 'C:/python/교통사고_데이터.xlsx'
df = pd.read_excel(dir_path, sheet_name='요일별', index_col=0)
_, _, _, _, sat, sun = df.values.tolist()

plt.title('주말 교통사고 발생 추세')
plt.xlabel('교통사고 발생 건수')
sns.distplot(sat, color='pink', rug=True, label='토')
sns.distplot(sun, color='orange', bins=10, rug=True, label='일')
plt.legend()
```

