목차

<Step1.> : 문제정의 및 가설설정

[문제정의]

유가에 영향을 끼치는 요인들이 무엇인지 알아보고자 한다.

[가설설정]

유가에 영향을 미치는 요인을 3가지로 설정하고 이 3가지요인이 각각 정말 유가에 영향을 미치는지 확인한다.

요인 1) 지역구
요인 2) 인건비 (셀프여부)
요인 3) 대리점 공급가

[참고자료]

전략 : 같은 지역인데 주유소마다 기름값이 다른 이유는 무엇일까?
시각화 : 데이터 사이언스 스쿨 Seaborn gallery

<Step2.> : 데이터 수집

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
import platform

from matplotlib import font_manager, rc

plt.rcParams['axes.unicode_minus'] = False

if platform.system() == 'Darwin':
    rc('font', family='AppleGothic')
elif platform.system() == 'Windows':
    path = "c:/Windows/Fonts/malgun.ttf"
    font_name = font_manager.FontProperties(fname=path).get_name()
    rc('font', family=font_name)
else:
    print('Unknown system... sorry~~~~')
```

[데이터 불러오기]

In [3]:
```
# 상반기 주유소 판매 데이터 불러오기
first = pd.read_csv('../data/2020년 상반기 주유소 판매가격.csv', encoding='cp949')
first
```

Out[3]:

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유
0	기준 : 월간 (202001~202006)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 01월	SK에너지	셀프	1802.00	1657.00	1495.00	0.0
2	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 02월	SK에너지	셀프	1795.38	1637.97	1483.97	0.0
3	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 03월	SK에너지	셀프	1741.26	1570.61	1431.13	0.0
4	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 04월	SK에너지	셀프	1617.00	1425.33	1290.00	0.0
...	...	...	...	...	...	...	...	...	...	...	...

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유
2990	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 02월	SK에너지	셀프	1872.93	1542.14	1365.55	0.0
2991	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 03월	SK에너지	셀프	1798.10	1479.32	1278.87	0.0
2992	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 04월	SK에너지	셀프	1652.93	1322.80	1128.67	0.0
2993	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 05월	SK에너지	셀프	1593.19	1252.55	1058.48	0.0
2994	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 06월	현대오일뱅크	셀프	1618.07	1328.73	1137.67	0.0

2995 rows × 11 columns

In [4]:
```
second = pd.read_csv('../data/2020년 하반기 주유소 판매가격.csv', encoding='cp949')
second
```

Out[4]:

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유
0	기준 : 월간 (202007~202012)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 07월	SK에너지	셀프	1635.0	1435.00	1265.00	0.0

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유
2	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 08월	SK에너지	셀프	1635.0	1435.00	1265.00	0.0
3	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 09월	SK에너지	셀프	1635.0	1435.00	1265.00	0.0
4	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 10월	SK에너지	셀프	1635.0	1435.00	1265.00	0.0
...	...	...	...	...	...	...	...	...	...	...	...
2946	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 08월	현대오일뱅크	셀프	1598.0	1402.65	1211.10	0.0
2947	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 09월	현대오일뱅크	셀프	1598.0	1393.00	1203.00	0.0
2948	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 10월	현대오일뱅크	셀프	1598.0	1338.42	1148.42	0.0
2949	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 11월	현대오일뱅크	셀프	1598.0	1292.00	1092.37	0.0
2950	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 12월	현대오일뱅크	셀프	1598.0	1367.58	1167.58	0.0

2951 rows × 11 columns

[데이터 결합]

In [5]:
```
# 상반기와 하반기 데이터를 상하로 결합
```

```
df = pd.concat([first, second])
df
```

Out[5]:

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유
0	기준 : 월간 (202001~202006)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 01월	SK에너지	셀프	1802.00	1657.00	1495.00	0.0
2	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 02월	SK에너지	셀프	1795.38	1637.97	1483.97	0.0
3	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 03월	SK에너지	셀프	1741.26	1570.61	1431.13	0.0
4	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 04월	SK에너지	셀프	1617.00	1425.33	1290.00	0.0
...	...	...	...	...	...	...	...	...	...	...	...
2946	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 08월	현대오일뱅크	셀프	1598.00	1402.65	1211.10	0.0
2947	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 09월	현대오일뱅크	셀프	1598.00	1393.00	1203.00	0.0
2948	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 10월	현대오일뱅크	셀프	1598.00	1338.42	1148.42	0.0

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유
2949	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 11월	현대오일뱅크	셀프	1598.00	1292.00	1092.37	0.0
2950	A0009197	서울 중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 12월	현대오일뱅크	셀프	1598.00	1367.58	1167.58	0.0

5946 rows × 11 columns

In [6]:
```
# 지역 컬럼의 값을 이용하여 시와 구를 분리

df['시'] = df['지역'].str[:2]
df['구'] = df['지역'].str[3:]
df.head()
```

Out[6]:

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유	시	구
0	기준 : 월간 (202001~202006)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 01월	SK에너지	셀프	1802.00	1657.00	1495.00	0.0	서울	강남구
2	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 02월	SK에너지	셀프	1795.38	1637.97	1483.97	0.0	서울	강남구
3	A0006039	서울 강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 03월	SK에너지	셀프	1741.26	1570.61	1431.13	0.0	서울	강남구

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유	시	구
4	A0006039	서울강남구	(유)동하석유힐탑셀프주유소	서울강남구논현로640	2020년04월	SK에너지	셀프	1617.00	1425.33	1290.00	0.0	서울	강남구

<Step3.> : 데이터 전처리

[결측값 제거]

In [7]:
```
df.dropna(inplace=True)
df
```

Out[7]:

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유	시	구
1	A0006039	서울강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년01월	SK에너지	셀프	1802.00	1657.00	1495.00	0.0	서울	강남구
2	A0006039	서울강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년02월	SK에너지	셀프	1795.38	1637.97	1483.97	0.0	서울	강남구
3	A0006039	서울강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년03월	SK에너지	셀프	1741.26	1570.61	1431.13	0.0	서울	강남구
4	A0006039	서울강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년04월	SK에너지	셀프	1617.00	1425.33	1290.00	0.0	서울	강남구
5	A0006039	서울강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년05월	SK에너지	셀프	1565.00	1350.81	1197.74	0.0	서울	강남구
...	...	...	...	...	...	...	...	...	...	...	...	...	...

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유	시	구
2946	A0009197	서울중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 08월	현대오일뱅크	셀프	1598.00	1402.65	1211.10	0.0	서울	중랑구
2947	A0009197	서울중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 09월	현대오일뱅크	셀프	1598.00	1393.00	1203.00	0.0	서울	중랑구
2948	A0009197	서울중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 10월	현대오일뱅크	셀프	1598.00	1338.42	1148.42	0.0	서울	중랑구
2949	A0009197	서울중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 11월	현대오일뱅크	셀프	1598.00	1292.00	1092.37	0.0	서울	중랑구
2950	A0009197	서울중랑구	현대오일뱅크㈜직영중랑교셀프주유소	서울 중랑구 망우로 229 (중화동)	2020년 12월	현대오일뱅크	셀프	1598.00	1367.58	1167.58	0.0	서울	중랑구

5944 rows × 13 columns

[데이터 구조 변경]

In [8]:
```
# 시 컬럼의 서울을 서울특별시로 변환

df['시'] += '특별시'
df.head()
```

Out[8]:

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유	시	구
1	A0006039	서울강남구	(유)동하석유힐탑셀프주유소	서울 강남구 논현로 640	2020년 01월	SK에너지	셀프	1802.00	1657.00	1495.00	0.0	서울특별시	강남구

	번호	지역	상호	주소	기간	상표	셀프여부	고급휘발유	휘발유	경유	실내등유	시	구
2	A0006039	서울강남구	(유)동하석유 힐탑셀프주유소	서울 강남구 논현로 640	2020년 02월	SK에너지	셀프	1795.38	1637.97	1483.97	0.0	서울특별시	강남구
3	A0006039	서울강남구	(유)동하석유 힐탑셀프주유소	서울 강남구 논현로 640	2020년 03월	SK에너지	셀프	1741.26	1570.61	1431.13	0.0	서울특별시	강남구
4	A0006039	서울강남구	(유)동하석유 힐탑셀프주유소	서울 강남구 논현로 640	2020년 04월	SK에너지	셀프	1617.00	1425.33	1290.00	0.0	서울특별시	강남구
5	A0006039	서울강남구	(유)동하석유 힐탑셀프주유소	서울 강남구 논현로 640	2020년 05월	SK에너지	셀프	1565.00	1350.81	1197.74	0.0	서울특별시	강남구

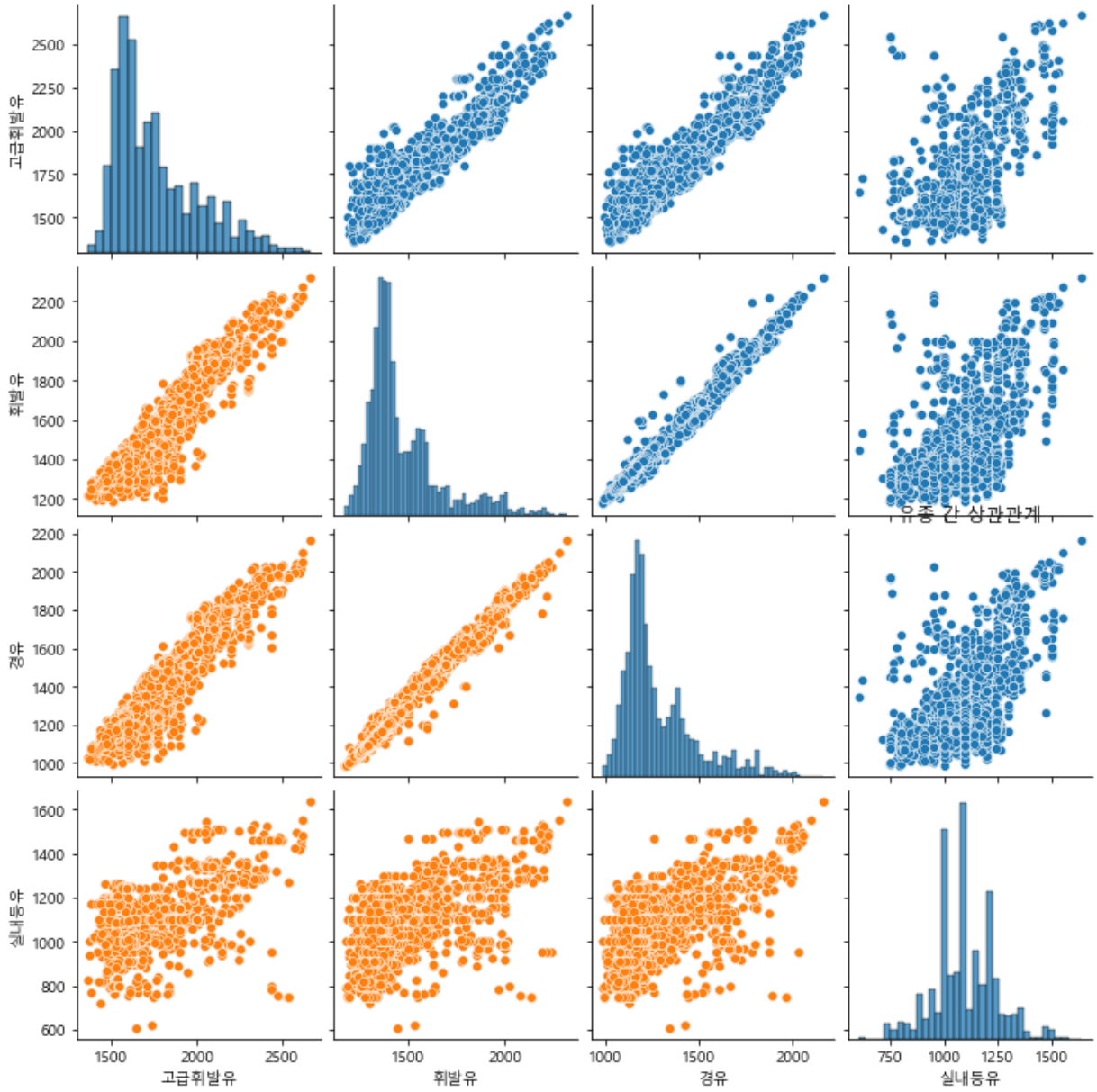
<Step4.> : 데이터 모델링 및 시각화

[요인분석1 : 지역구]

1.1 유종 간 상관관계 _ (Pairplot Chart)

In [9]:

```
df_nonull= df.replace(0, np.NaN)
df_nonull

import matplotlib.pyplot as plt
import matplotlib.font_manager as fm

fm.get_fontconfig_fonts()
font_location = 'C:₩Windows₩Fonts₩malgun.ttf'
font_name = fm.FontProperties(fname=font_location).get_name()
plt.rc('font', family=font_name)

graph = sns.pairplot(df_nonull)
graph.map_lower(sns.scatterplot)
graph.map_upper(sns.regplot, scatter=False, truncate=False, ci=False)
plt.title("유종 간 상관관계")
plt.show()
```



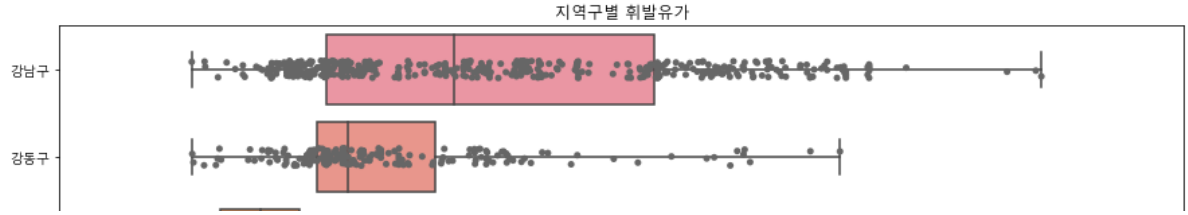
분석결과

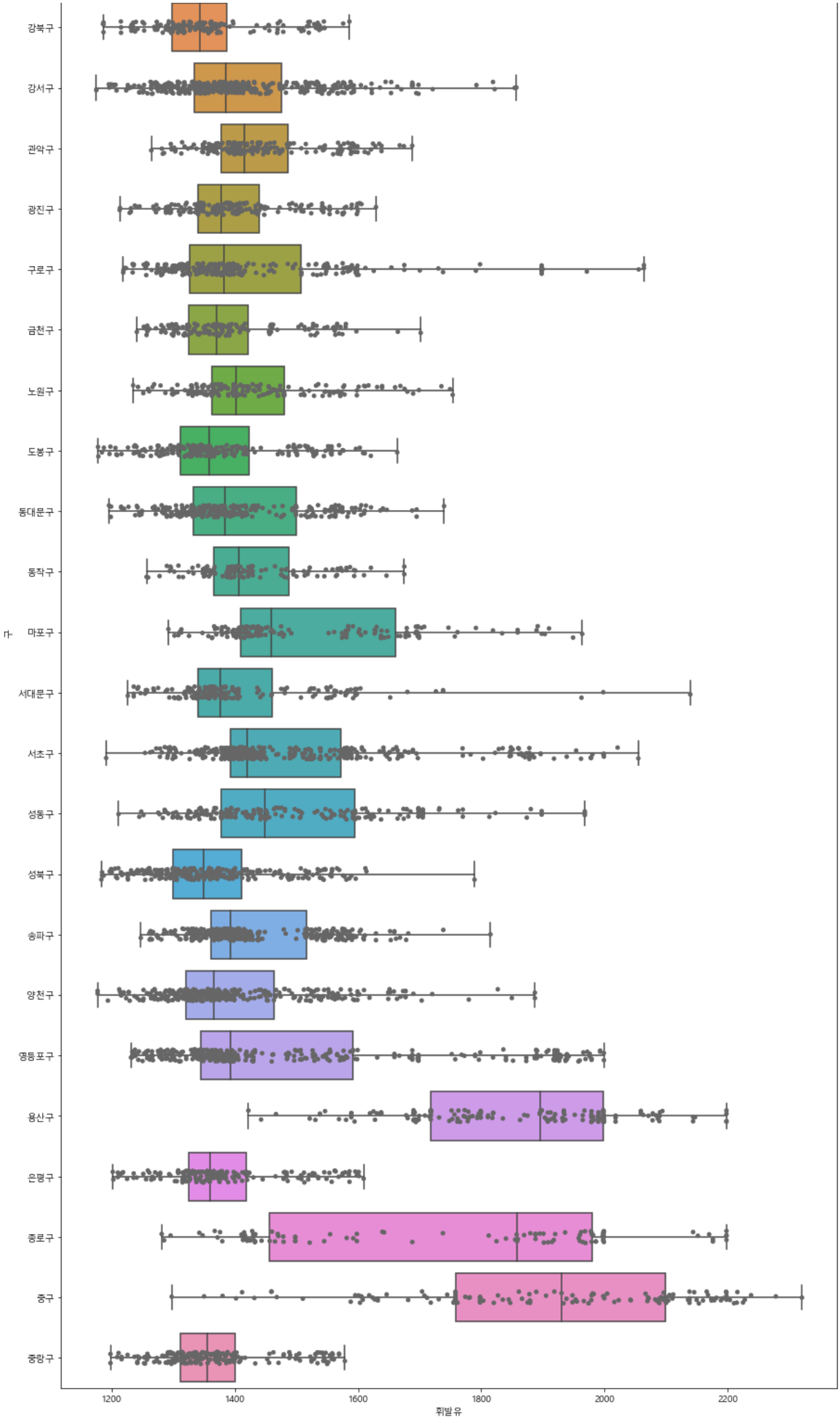
지역구 별 유가를 비교하기 전에 유가별 상관관계를 파악하여 가장 대표적인 유종이 무엇인지 알아보았다.

> 분석결과 : 모든 유종은 양의 상관관계를 가지고있으며 그 중에서도 휘발유와 경유의 상관관계가 제일 높다.
  따라서, 편의상 휘발유를 대표적 유종으로 선정하였다.

1.2 지역구별 유가 _ (Boxplot)

In [10]:
```
plt.figure(figsize=(15, 30))
plt.title("지역구별 휘발유가")
sns.boxplot(x="휘발유", y="구", data=df, whis=np.inf)
sns.stripplot(x="휘발유", y="구", data=df, jitter=True, color="0.4")
plt.show()
```




강북구
강서구
관악구
광진구
구로구
금천구
노원구
도봉구
동대문구
동작구
마포구
서대문구
서초구
성동구
성북구
송파구
양천구
영등포구
용산구
은평구
종로구
중구
중랑구
구
1200
1400
1600
1800
2000
2200
휘발유

분석결과

대표 유종인 휘발유의 지역구별 가격을 알아보았다.

> 분석결과 : Boxplot으로 알 수 있는 것은 중앙값, 산포도, 최대값, 최소값, 이상치 등이 있다.
위 그래프를 해석하면, 휘발유의 평균가격이 가장 높은 지역구는 중구, 용산구, 종로구 순이고 가장 낮은 지역구는 강북구,성북구, 중랑구
임을 알 수 있다.

특히, 중구의 그래프를 보면 2,3,4분위수 범위의 간격이 비교적 일정한 것으로보아 중구의 높은 휘발유 평균가격은 이상치에 의해 좌우된 결과가 아님을 알 수 있다.

따라서, 해당 그래프는 유의미하다.

1.3 지역구별 유가 _ (Barplot)

In [11]:
```
df_a = df.replace(0, np.NaN) # 0을 제외하고 평균을 내기 위해 nan처리
df_a = df_a.groupby('구')[['고급휘발유', '휘발유', '경유', '실내등유']].mean() # df의
df_a
```

Out[11]:

구	고급휘발유	휘발유	경유	실내등유
강남구	1820.577020	1601.450148	1426.829409	1250.738244
강동구	1742.760476	1483.317188	1297.847500	1082.447983
강북구	1740.838125	1358.062500	1168.330962	942.135278
강서구	1616.750072	1407.981343	1218.795721	1068.254754
관악구	1678.794375	1440.764404	1266.050363	1087.529294
광진구	1633.337292	1397.081373	1202.645196	1065.067746
구로구	1588.155789	1429.092710	1226.033435	1070.564000
금천구	1533.020625	1390.482484	1196.913248	1065.275529
노원구	1661.301000	1434.355371	1246.633600	1066.436341
도봉구	1669.422292	1378.404167	1190.656389	993.692963
동대문구	1616.753056	1408.674144	1215.942824	1088.621518
동작구	1600.552453	1430.096083	1264.556583	1060.882778
마포구	1835.779714	1533.331408	1356.522676	1161.473958
서대문구	1708.995488	1413.302099	1223.311547	1140.450896
서초구	1712.043105	1490.589930	1324.340634	1155.363154
성동구	1758.545816	1492.690099	1307.588218	1165.003548
성북구	1666.202018	1362.384425	1173.994390	1034.432033
송파구	1595.915313	1423.959394	1242.611439	1125.365122
양천구	1637.812301	1401.761962	1209.962310	1083.694105
영등포구	1804.245648	1497.842619	1313.364603	1122.859420

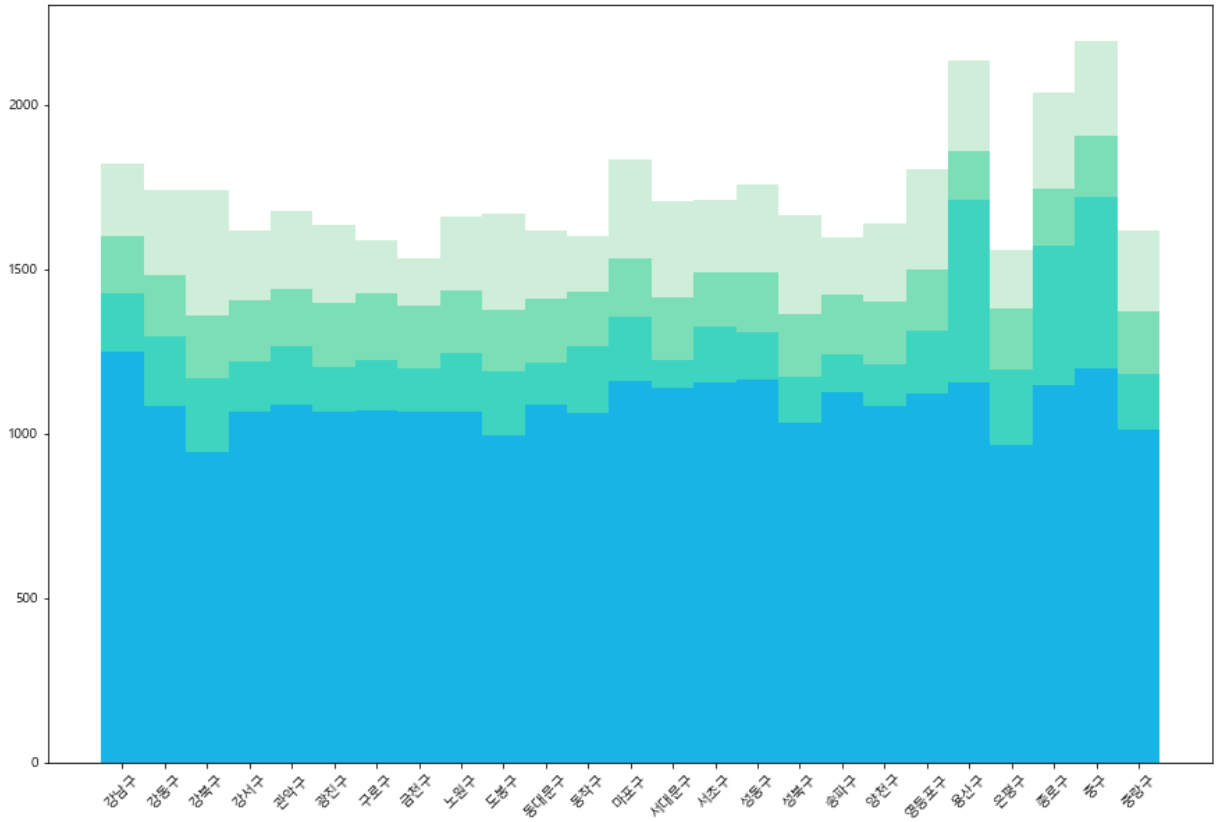
구	고급휘발유	휘발유	경유	실내등유
용산구	2133.165179	1858.926071	1709.826726	1154.484000
은평구	1560.113553	1382.617371	1192.846031	966.631096
종로구	2036.235000	1746.083853	1571.764495	1146.844706
중구	2194.344646	1904.910216	1718.084029	1197.343226
중랑구	1618.713793	1370.667396	1180.817396	1011.434426

In [12]:
```
df_a.reset_index(inplace=True)
df_a
```

Out[12]:

	구	고급휘발유	휘발유	경유	실내등유
0	강남구	1820.577020	1601.450148	1426.829409	1250.738244
1	강동구	1742.760476	1483.317188	1297.847500	1082.447983
2	강북구	1740.838125	1358.062500	1168.330962	942.135278
3	강서구	1616.750072	1407.981343	1218.795721	1068.254754
4	관악구	1678.794375	1440.764404	1266.050363	1087.529294
5	광진구	1633.337292	1397.081373	1202.645196	1065.067746
6	구로구	1588.155789	1429.092710	1226.033435	1070.564000
7	금천구	1533.020625	1390.482484	1196.913248	1065.275529
8	노원구	1661.301000	1434.355371	1246.633600	1066.436341
9	도봉구	1669.422292	1378.404167	1190.656389	993.692963
10	동대문구	1616.753056	1408.674144	1215.942824	1088.621518
11	동작구	1600.552453	1430.096083	1264.556583	1060.882778
12	마포구	1835.779714	1533.331408	1356.522676	1161.473958
13	서대문구	1708.995488	1413.302099	1223.311547	1140.450896
14	서초구	1712.043105	1490.589930	1324.340634	1155.363154
15	성동구	1758.545816	1492.690099	1307.588218	1165.003548
16	성북구	1666.202018	1362.384425	1173.994390	1034.432033
17	송파구	1595.915313	1423.959394	1242.611439	1125.365122
18	양천구	1637.812301	1401.761962	1209.962310	1083.694105
19	영등포구	1804.245648	1497.842619	1313.364603	1122.859420
20	용산구	2133.165179	1858.926071	1709.826726	1154.484000
21	은평구	1560.113553	1382.617371	1192.846031	966.631096
22	종로구	2036.235000	1746.083853	1571.764495	1146.844706
23	중구	2194.344646	1904.910216	1718.084029	1197.343226
24	중랑구	1618.713793	1370.667396	1180.817396	1011.434426

In [13]:
```
plt.figure(figsize=(15, 10))
plt.xticks(rotation=45)

plt.bar(df_a['구'], df_a['고급휘발유'], alpha = 0.3, width=1, color='#62c983')
plt.bar(df_a['구'], df_a['휘발유'], alpha = 0.4, width=1, color="#00c786")
plt.bar(df_a['구'], df_a['경유'], alpha = 0.5, width=1, color="#00cec9")
plt.bar(df_a['구'], df_a['실내등유'], alpha = 0.6, width=1, color="#009fff")

plt.show()
```



분석결과

지역별 전체 유가 평균을 살펴보았다.

> 분석결과 : Boxplot으로 봤을 때와 같이 평균가격이 가장 높은 지역구는 중구, 용산구, 종로구 순이고 가장 낮은 지역구는 강북구,성북구, 중랑구임을 알 수 있다.또한 지역구별로 유종간에 상관관계가 있음도 확인할 수 있었다.

1.4 지역구별 월별 유가 _ (Heatmap)

In [21]:
```
df1=df.replace(0,np.NaN)
df1=df1.groupby(['구','기간'])[['고급휘발유','휘발유','경유','실내등유']].mean()
df1
```

Out[21]:

구	기간	고급휘발유	휘발유	경유	실내등유
강남구	2020년 01월	1983.367059	1771.252750	1620.019000	1332.228462
	2020년 02월	1968.100294	1754.086750	1600.688750	1332.799167
	2020년 03월	1928.760588	1703.356500	1545.263750	1313.131818
	2020년 04월	1815.537941	1580.068250	1418.166250	1230.497692

구	기간	고급휘발유	휘발유	경유	실내등유
	2020년 05월	1740.169118	1504.057250	1335.514000	1161.004545
...	...	...	...	...	...
중랑구	2020년 08월	1570.570000	1364.416875	1165.096875	995.800000
	2020년 09월	1564.000000	1349.686250	1151.018125	995.000000
	2020년 10월	1560.130000	1326.126250	1126.568750	982.742000
	2020년 11월	1555.600000	1306.229375	1107.904375	966.500000
	2020년 12월	1575.270000	1371.651875	1171.588750	987.452000

300 rows × 4 columns

In [22]:

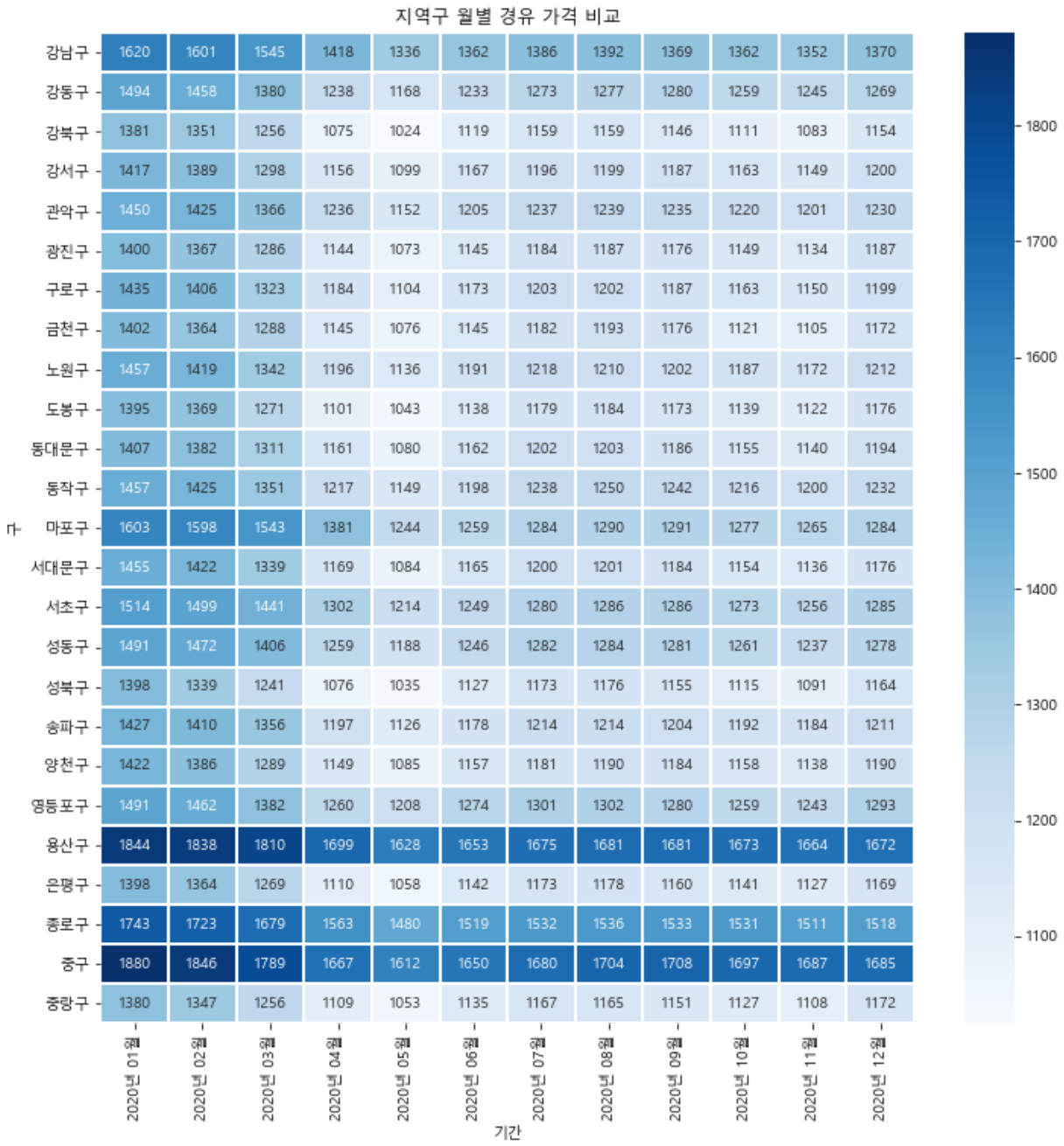
```
df2=pd.pivot_table(df1,index="구",columns='기간', values="고급휘발유")
df3=pd.pivot_table(df1,index="구",columns='기간', values="휘발유")
df4=pd.pivot_table(df1,index="구",columns='기간', values="경유")
df5=pd.pivot_table(df1,index="구",columns='기간', values="실내등유")
plt.figure(figsize=(12, 12))
sns.heatmap(df2,annot=True,fmt=".0f",linewidths=1,cmap='Blues')
plt.title('지역구 월별 고급휘발유 가격 비교')
plt.show()
```

In [23]:

```
plt.figure(figsize=(12, 12))
sns.heatmap(df3,annot=True,fmt=".0f",linewidths=1,cmap='Blues')
plt.title('지역구 월별 휘발유 가격 비교')
plt.show()
```

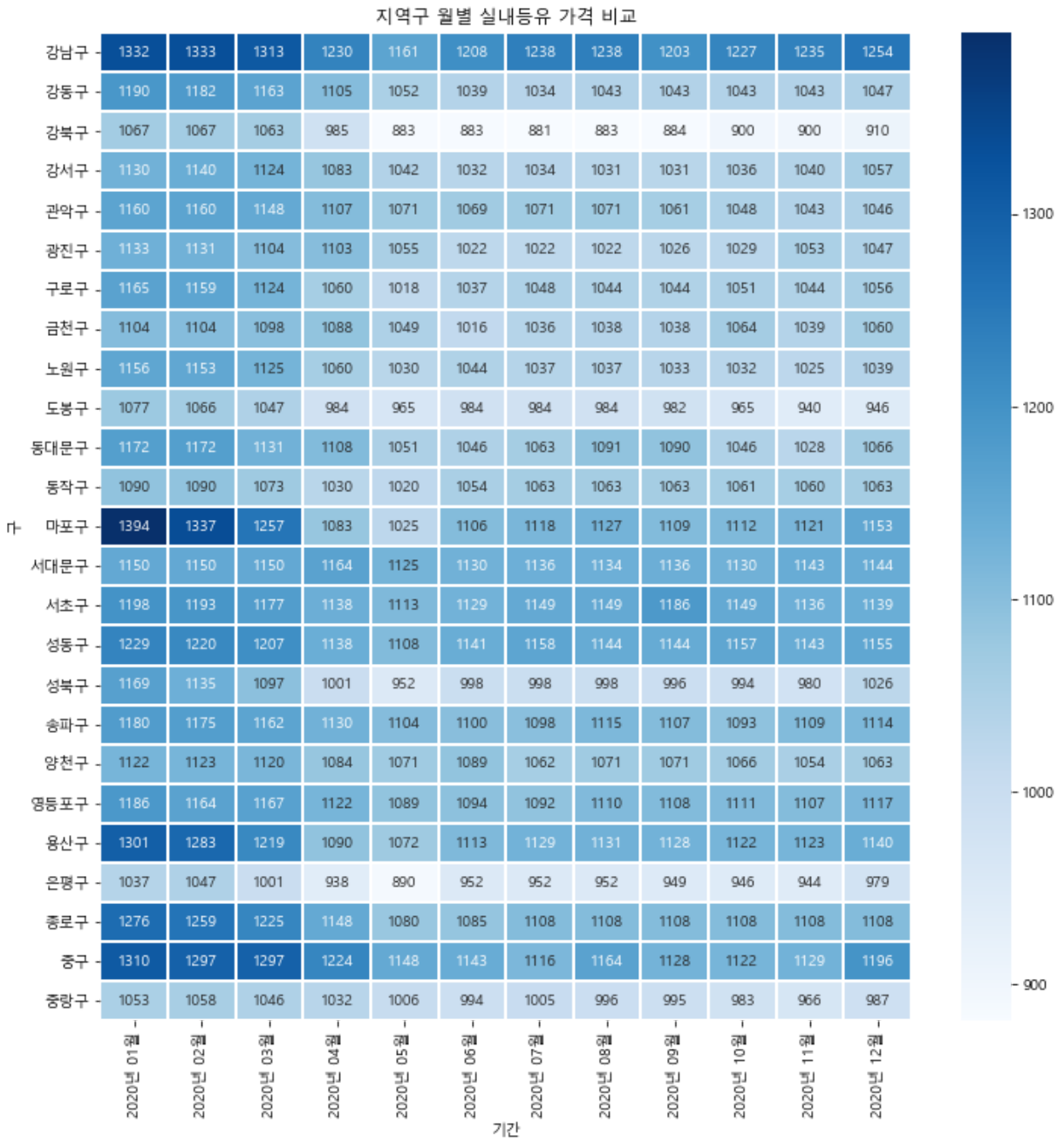
In [24]:

```
plt.figure(figsize=(12, 12))
sns.heatmap(df4,annot=True,fmt=".0f",linewidths=1,cmap='Blues')
plt.title('지역구 월별 경유 가격 비교')
plt.show()
```



In [25]:

```
plt.figure(figsize=(12, 12))
sns.heatmap(df5,annot=True,fmt=".0f",linewidths=1,cmap='Blues')
plt.title('지역구 월별 실내등유 가격 비교')
plt.show()
```



분석결과

지역과 유종, 유가간의 관계를 알아보았으니 유가에 월별 영향이 미치는지 알아보았다.

> 분석결과 : 월별 유가를 비교해보니 각 월이 유가에 미치는 영향은 적은 것을 확인 할 수 있었다.
따라서 위 그래프는 무의미하다고 봤다.

[요인분석2 : 인건비]

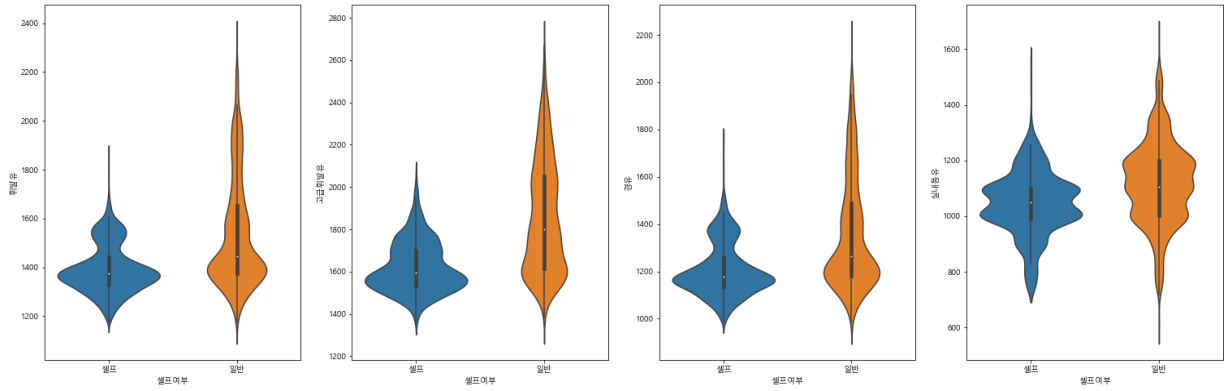
2.1 인건비와 유가의 상관관계 _ (Violinplot)

In [26]:
```
# 한글처리
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm

fm.get_fontconfig_fonts()
font_location = 'C:WWindowsWFontsWmalgun.ttf'
```

```
font_name = fm.FontProperties(fname=font_location).get_name()
plt.rc('font', family=font_name)
```

In [28]:

```
df0 = df.replace(0, np.NaN)
fig, ax = plt.subplots(figsize=(26, 8), ncols=4)
sns.violinplot(data=df0,x='셀프여부', y='휘발유', ax=ax[0])
sns.violinplot(data=df0,x='셀프여부', y='고급휘발유',  ax=ax[1])
sns.violinplot(data=df0,x='셀프여부', y='경유',  ax=ax[2])
sns.violinplot(data=df0,x='셀프여부', y='실내등유', ax=ax[3])

plt.show()
```



분석결과

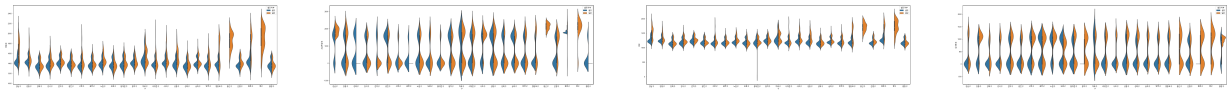
인건비가 유가에 미치는 영향을 알아보았다.

> 분석결과 : 대부분의 지역에서는 유종에 상관없이 셀프주유소의 가격이 일반 주유소의 가격보다 저렴하다는 것을 알 수 있다.

2.2 지역구별 인건비와 유가의 상관관계 _ (Violinplot)

In [27]:

```
fig, ax = plt.subplots(figsize=(120, 8),ncols=4)
sns.violinplot(data=df, hue='셀프여부',x='구', y='휘발유', split=True, ax=ax[0])
sns.violinplot(data=df, hue='셀프여부',x='구', y='고급휘발유', split=True, ax=ax[1])
sns.violinplot(data=df, hue='셀프여부',x='구', y='경유', split=True, ax=ax[2])
sns.violinplot(data=df, hue='셀프여부',x='구', y='실내등유', split=True, ax=ax[3])
plt.show()
```



분석결과

지역별로 셀프 주유소와 일반 주요소의 가격 차이를 알아보았다.

> 분석결과 : 대부분의 지역에서는 셀프 주유소의 가격이 일반 주유소의 가격보다 저렴했다.
  하지만, 일부 지역 (ex: 강북구, 광진구, 금천구 등)에서는 인건비의 영향이 무의미하기도 했다.
  따라서, 인건비는 유가에 영향을 미치는 결정적인 요소는 아니라고 판단된다.

[요인분석3 : 대리점 공급가]

3.1 지역별 대리점의 수 _ (Wordcloud)

In [37]:
```
# 서울시 전체 상호 중복
part = df
part.dropna(inplace=True)
seoul=part.drop_duplicates(['상호'], keep='first')
```

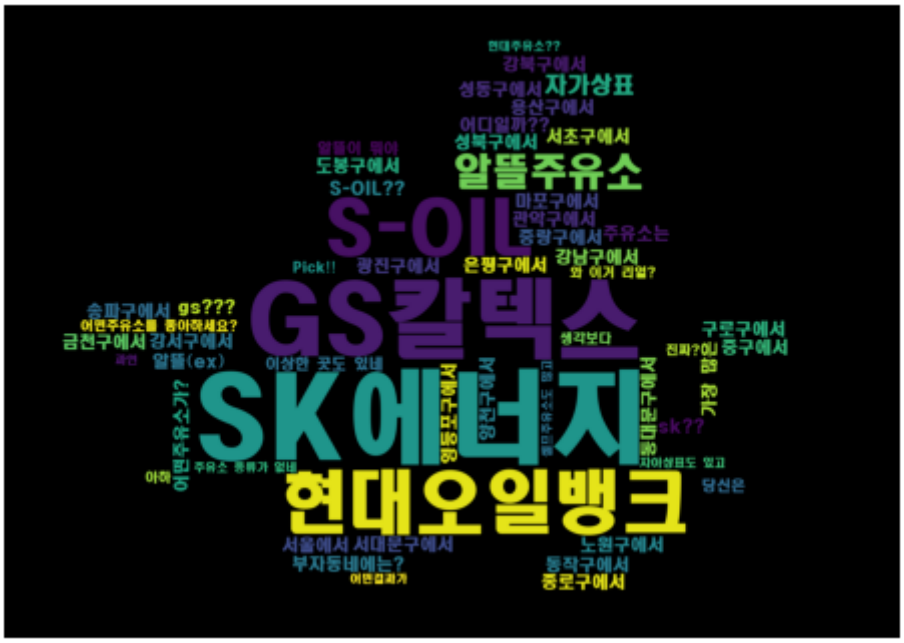
In [38]:
```
#서울 상표 카운트
s_freq = seoul.groupby(['상표']).count()
s_count = s_freq.loc[:,['지역']]
s_count
```

Out[38]:

상표	지역
GS칼텍스	133
S-OIL	76
SK에너지	190
알뜰(ex)	1
알뜰주유소	11
자가상표	2
현대오일뱅크	80

In [39]:
```
sdict0 = s_count['지역'].to_dict()
추가 = {"가장 많은":1, "주유소는":1, "어디일까??":1, "부자동네에는?":1, "어떤주유소가
        "gs???":1, "sk??":1, "S-OIL??":1, "서울에서":1, "강남구에서":1, "강북구에서":
      "강서구에서":1, "관악구에서":1, "광진구에서":1, "구로구에서":1, "금천구에서":1,
      "도봉구에서":1, "동대문구에서":1, "동작구에서":1, "마포구에서":1, "서대문구에서"
      "성동구에서":1, "성북구에서":1, "송파구에서":1, "양천구에서":1, "영등포구에서":1
      "은평구에서":1, "종로구에서":1, "중구에서":1, "중랑구에서":1, "이상한 곳도 있네"
        "당신은":1, "어떤주유소를 좋아하세요?":1, "Pick!!":1, "와 이거 리얼?":1, "진
        "생각보다 ":1, "주유소 종류가 없네":1, "현대주유소??":1, "셀프주유소도 많고":
        "과연":1, "어떤결과가":1}
sdict0.update(추가)
```

In [40]:
```
from wordcloud import WordCloud
from PIL import Image
import numpy as np

imgmask = np.array(Image.open("img₩seoul.png"))

keywords = sdict0

wc = WordCloud(font_path="C:₩Windows₩Fonts₩HMKMRHD.ttf",
                          width=800,
                          height=600,
                          background_color='black',
                          mask = imgmask
                          )


cloud=wc.generate_from_frequencies(keywords)
plt.figure(figsize=(8, 8))
plt.imshow(cloud)
plt.axis("off")
plt.show()
```



분석결과

지역별 대리점의 수를 알아보았다.

3.2 지역구별 대리점의 수와 유가 _ (Plot)

In [14]:
```
df_m = df.replace(0, np.NaN) # 0을 제외하고 평균을 내기 위해 nan처리
df_m = df_m.groupby('상표')[['휘발유']].mean() # df의 mean()은 nan 제외한 평균을 내줌
df_m
```

Out[14]:

상표	휘발유
GS칼텍스	1480.077478
S-OIL	1434.653561
SK에너지	1517.494476
알뜰(ex)	1330.236667
알뜰주유소	1355.216043
자가상표	1414.056800
현대오일뱅크	1427.582881

In [15]:
```
# 추가됨
df_m.sort_values(by='휘발유').drop(['알뜰(ex)', '알뜰주유소', '자가상표'])
```

Out[15]:

상표	휘발유
현대오일뱅크	1427.582881
S-OIL	1434.653561
GS칼텍스	1480.077478

	휘발유
상표	
SK에너지	1517.494476

In [16]:
```
df_k = df.pivot_table('휘발유', index='구', columns='상표', aggfunc='count').dropna(t
df_k.fillna(0, inplace=True)
df_k
```

Out[16]:

상표	GS칼텍스	S-OIL	SK에너지	현대오일뱅크
구				
강남구	143.0	71.0	186.0	74.0
강동구	60.0	24.0	63.0	45.0
강북구	36.0	24.0	53.0	43.0
강서구	84.0	72.0	142.0	68.0
관악구	45.0	24.0	60.0	48.0
광진구	60.0	36.0	46.0	50.0
구로구	48.0	72.0	53.0	77.0
금천구	34.0	36.0	41.0	31.0
노원구	72.0	36.0	48.0	19.0
도봉구	36.0	72.0	46.0	62.0
동대문구	68.0	33.0	86.0	76.0
동작구	36.0	12.0	22.0	50.0
마포구	12.0	34.0	70.0	26.0
서대문구	35.0	36.0	77.0	33.0
서초구	167.0	60.0	130.0	57.0
성동구	36.0	24.0	82.0	48.0
성북구	84.0	58.0	57.0	76.0
송파구	120.0	48.0	135.0	93.0
양천구	60.0	12.0	132.0	100.0
영등포구	119.0	31.0	140.0	76.0
용산구	48.0	0.0	87.0	33.0
은평구	72.0	24.0	53.0	32.0
종로구	24.0	24.0	42.0	19.0
중구	43.0	12.0	60.0	12.0
중랑구	60.0	60.0	53.0	19.0

In [17]:
```
# df_k = (df_k['SK에너지'] + df_k['GS칼텍스']) / (df_k['S-OIL'] + df_k['현대오일뱅크']
df_k = df_k['SK에너지'] / df_k['현대오일뱅크']
df_k
```

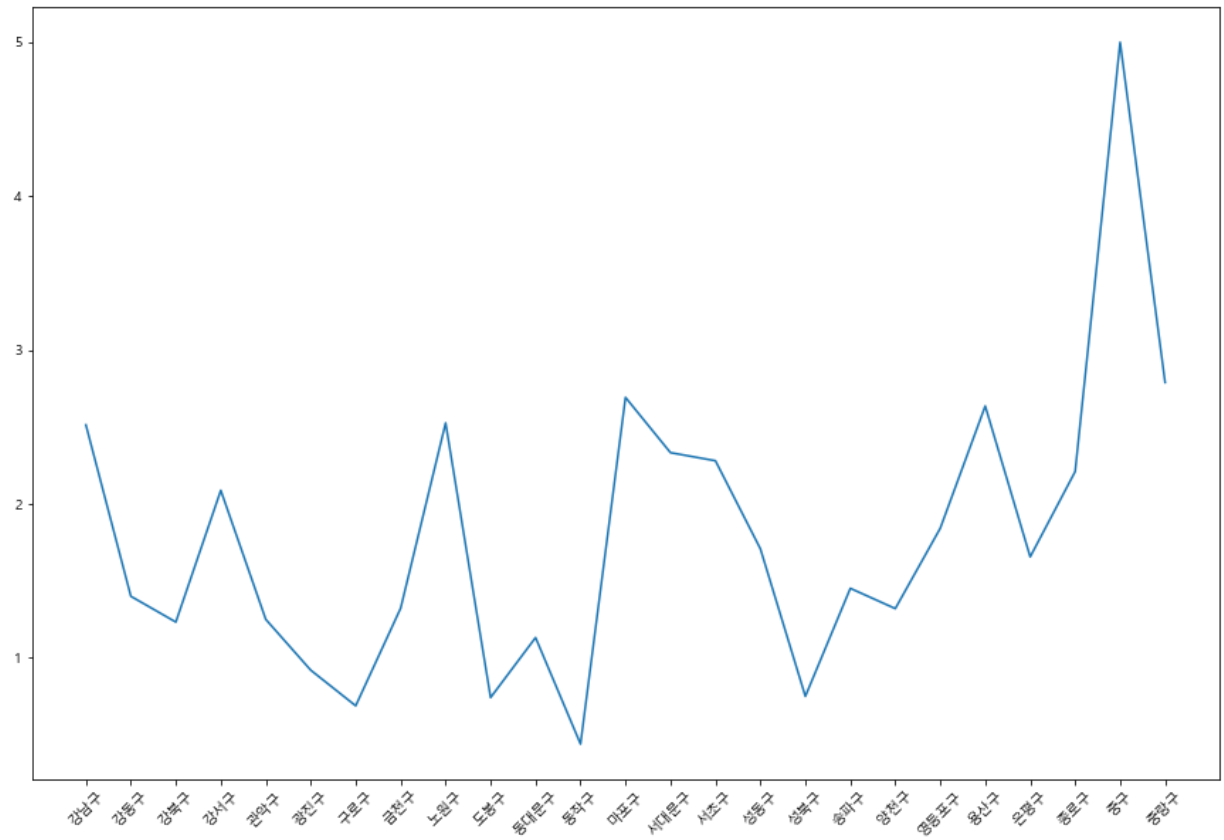
Out[17]:
```
구
강남구     2.513514
강동구     1.400000
강북구     1.232558
강서구     2.088235
관악구     1.250000
광진구     0.920000
구로구     0.688312
금천구     1.322581
노원구     2.526316
도봉구     0.741935
동대문구    1.131579
동작구     0.440000
마포구     2.692308
서대문구    2.333333
서초구     2.280702
성동구     1.708333
성북구     0.750000
송파구     1.451613
양천구     1.320000
영등포구    1.842105
용산구     2.636364
은평구     1.656250
종로구     2.210526
중구      5.000000
중랑구     2.789474
dtype: float64
```

In [18]:
```
df_k.sort_values()
```

Out[18]:
```
구
동작구     0.440000
구로구     0.688312
도봉구     0.741935
성북구     0.750000
광진구     0.920000
동대문구    1.131579
강북구     1.232558
관악구     1.250000
양천구     1.320000
금천구     1.322581
강동구     1.400000
송파구     1.451613
은평구     1.656250
성동구     1.708333
영등포구    1.842105
강서구     2.088235
종로구     2.210526
서초구     2.280702
서대문구    2.333333
강남구     2.513514
노원구     2.526316
용산구     2.636364
마포구     2.692308
중랑구     2.789474
중구      5.000000
dtype: float64
```

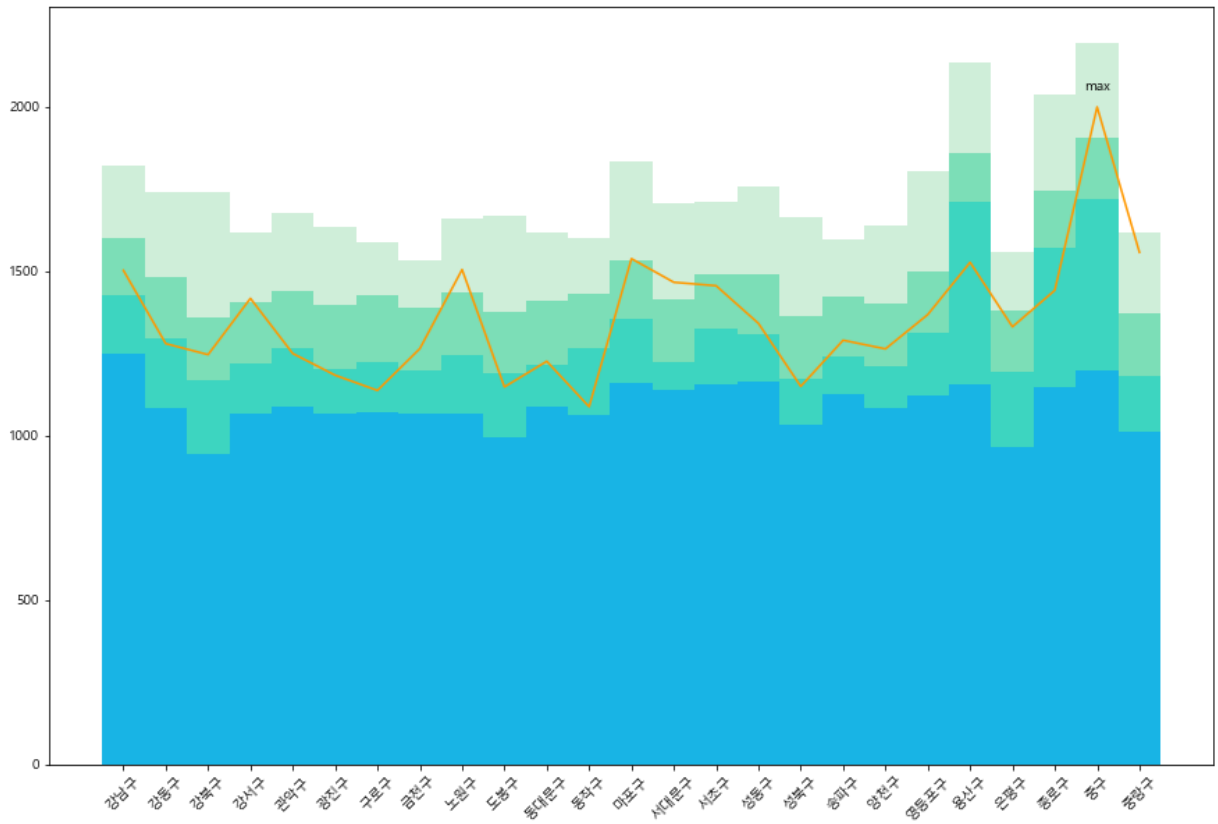
In [19]:
```
plt.figure(figsize=(15, 10))
plt.xticks(rotation=45)

plt.plot(df_k)
```

Out[19]:
```
[<matplotlib.lines.Line2D at 0x16e3967dee0>]
```



In [30]:
```
df_k = df_k * 200 + 1000
```

In [31]:
```
plt.figure(figsize=(15, 10))
plt.xticks(rotation=45)
plt.text(22.7, 2050, f"max")

plt.bar(df_a['구'], df_a['고급휘발유'], alpha = 0.3, width=1, color='#62c983')
plt.bar(df_a['구'], df_a['휘발유'], alpha = 0.4, width=1, color="#00c786")
plt.bar(df_a['구'], df_a['경유'], alpha = 0.5, width=1, color="#00cec9")
plt.bar(df_a['구'], df_a['실내등유'], alpha = 0.6, width=1, color="#009fff")
plt.plot(df_k, color='#ff9900')

plt.show()
```



분석결과

지역별로 공급업체가 유가에 영향을 끼치는지 확인했다.

분석결과> 휘발유 가격이 가장 높은 업체의 수와 가장 낮은 업체의 수로 비율을 구해서 지역별 전체 유가 평균 그래프와 함께 봤다.
유가 평균이 높은 지역구에는 해당 비율이 높게 나오고,유가 평균이 낮은 지역구에는 해당 비율이 낮게 나온다.

따라서, 공급업체가 유가에 영향을 끼친다고 할 수 있다.

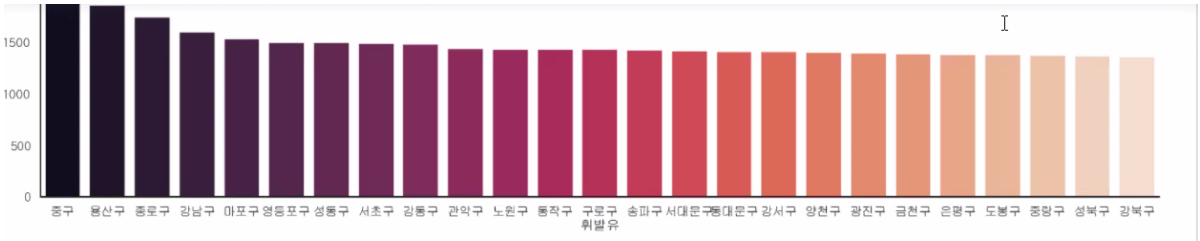
Feedback

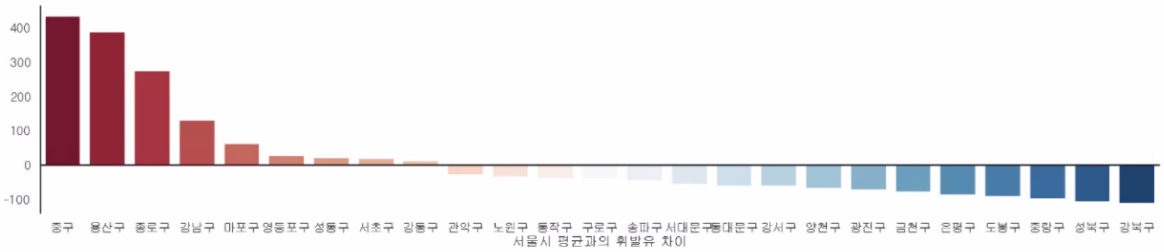
- 데이터 전략 잘짰다.
- 데이터 수집 : 기업, 공공, 크롤링
- 데이터 전략은 전처리에 많은 영향을 끼친다.
- 강남구가 있는데 중구가 왜 비쌀까 같은 거? 아쉽
- 시각화 여러 그래프로 해서 좋았다.

아쉬운 점

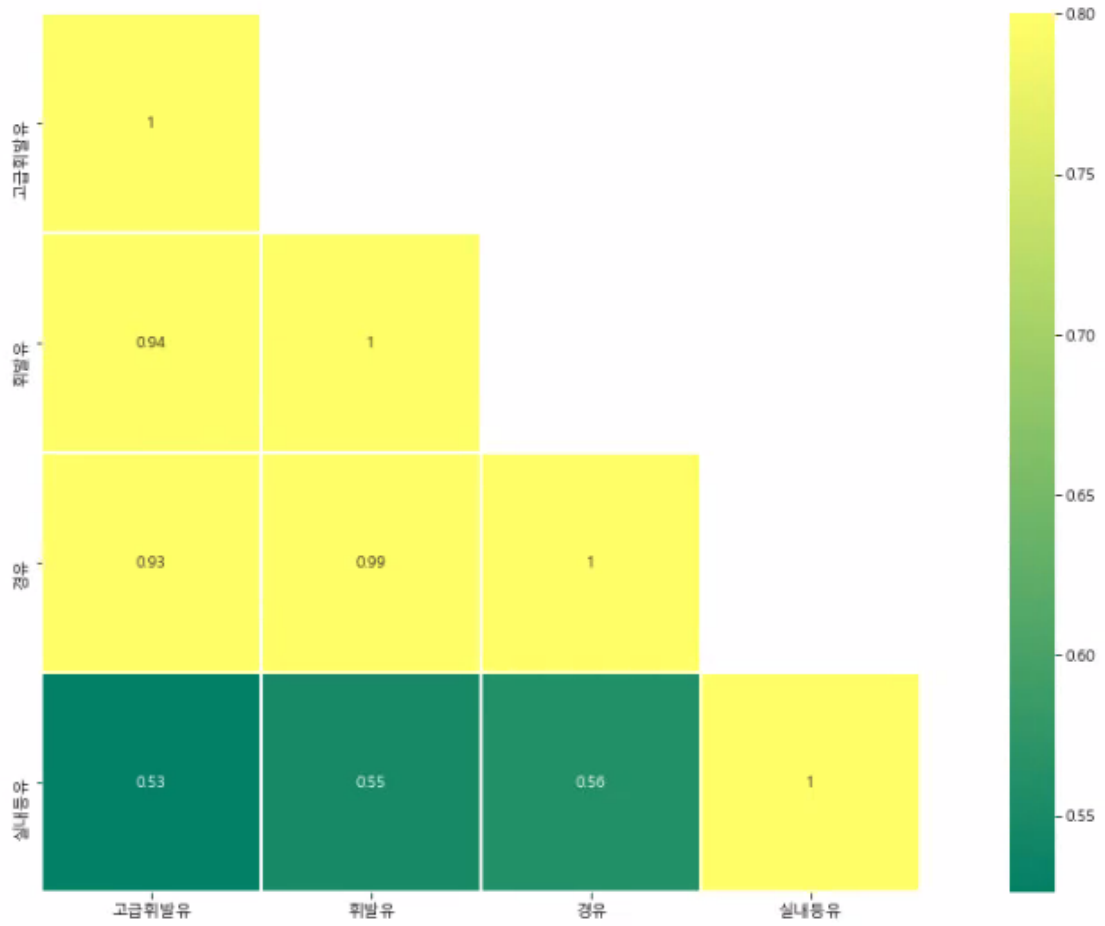
- 우리는 주어진 데이터에 국한되어 분석하는 걸 당연하게 생각했다. 하지만 외부 데이터를 가지고 와서 사용한 팀을 보고, 갇힌 사고를 한 게 아닌가 생각했다.
- 다른 팀 작업방식 중 질문에 대한 답을 찾아낼 때까지 과정을 담아낸 게 좋았던 것 같다.
- 다른 팀들의 발표를 보고 좋았던 시각화 자료들

다른 팀 잘한 거


1500
1000
500
0
중구 용산구 종로구 강남구 마포구 영등포구 성동구 서초구 강동구 관악구 노원구 동작구 구로구 송파구 서대문구 동대문구 광서구 양천구 광진구 금천구 은평구 도봉구 중랑구 성북구 강북구
휘발유


400
300
200
100
0
-100
중구 용산구 종로구 강남구 마포구 영등포구 성동구 서초구 강동구 관악구 노원구 동작구 구로구 송파구 서대문구 동대문구 광서구 양천구 광진구 금천구 은평구 도봉구 중랑구 성북구 강북구
서울시 평균과의 휘발유 차이


1750
1500
1250
1000
750
500
250


고급휘발유
휘발유
경유
실내등유
1
0.94
1
0.93
0.99
1
0.53
0.55
0.56
1
고급휘발유
휘발유
경유
실내등유
0.80
0.75
0.70
0.65
0.60
0.55