

In [1]:

```
#!/pip install IPython
from IPython.display import Image
import pandas as pd
import warnings
warnings.filterwarnings(action='ignore')
```

In [2]:

```
df= pd.read_csv('지역별 통계목록_2019.csv', header=None)
df1= pd.read_csv('지역별 통계목록_2020.csv', header=None)
```

In [3]:

```
# 2019년 자료 전처리
df = df.drop(index=[0,1,2,4])
df = df.T
df.columns = df.iloc[0]
df = df.drop(index=[0])
df.loc[5, '지역']='충객석수'
df.dropna(axis=0, inplace=True)
df.index = df['지역']
df = df.drop('지역', axis=1)

for i in df.columns :
    df[i]=df[i].str.replace(',','')
    df[i]=df[i].astype(int)

df = df.T
```

In [4]:

```
# 2020년 자료 전처리
df1 = df1.drop(index=[0,1,2,4])
df1 = df1.T
df1.columns = df1.iloc[0]
df1 = df1.drop(index=[0])

df1.loc[5, '지역']='충객석수'
df1.dropna(axis=0, inplace=True)

df1.index = df1['지역']
df1 = df1.drop('지역', axis=1)

for i in df1.columns :
    df1[i]=df1[i].str.replace(',','')
    df1[i]=df1[i].astype(int)

df1 = df1.T

df1.columns
```

Out[4]:

```
Index(['공연시설', '공연장', '충객석수', '공연건수', '개막편수', '상연횟수', '예매수', '매출액'], dtype='object',
      name='지역')
```

In [5]:

```
df.index
```

Out[5]:

```
Index(['서울', '경상', '부산', '대구', '경남', '경북', '울산', '경기/인천', '경기', '인천', '충청',
      '대전', '충남', '충북', '세종', '전라', '전북', '전남', '광주', '강원', '제주', '합계'],
      dtype='object', name=0)
```

In [6]:

```
data = df.loc[['서울', '경상', '경기/인천', '충청', '전라', '강원', '제주'],:]
data1 = df1.loc[['서울', '경상', '경기/인천', '충청', '전라', '강원', '제주'],:]
```

In [7]:

```
import pandas as pd
import matplotlib.pyplot as plt
import os
import matplotlib as mat
```

In [8]:

```
# 설치된 한글 폰트로 설정
mat.rcParams['font.family'] = 'HYGothic-Medium'
```

In [9]:

```
df = df.loc[['서울', '부산', '대구', '인천', '광주', '대전', '울산', '세종', '경기', '강원', '충북', '충남',
            '전북', '전남', '경북', '경남', '제주'],:]
df1 = df1.loc[['서울', '부산', '대구', '인천', '광주', '대전', '울산', '세종', '경기', '강원', '충북', '충남',
            '전북', '전남', '경북', '경남', '제주'],:]
```

In [11]:

```
import seaborn as sns

cost = pd.read_csv('적절하다고_생각하는_여가비용_한_달_평균.csv')

cost = cost.rename(columns=cost.iloc[0])
cost = cost.drop(cost.index[0])

cost_2018 = cost.iloc[:, :10]
cost_2019 = cost.iloc[:, [0,1,10,11,12,13,14,15,16,17]]

cost_2018.loc[:, '기준연도'] = 2018
cost_2019.loc[:, '기준연도'] = 2019

cost = pd.concat([cost_2018, cost_2019])
```

In [12]:

```
cost['통계분류(1)'].unique()
```

Out[12]:

```
array(['전체', '성별', '연령별', '학력별', '동거가구원수별', '혼인상태별', '가구주여부별', '종사자지위별',
      '가구소득별', '지역규모별', '권역별', '17개 시도별', '장애여부별'], dtype=object)
```

In [13]:

```
sex = cost.loc[cost['통계분류(1)']=='성별',:]
sex = sex[['통계분류(2)', '평균', '기준연도']]
```

In [14]:

```
sex['평균'] = sex['평균'].astype(float)
sex
```

Out[14]:

	통계분류(2)	평균	기준연도
2	남성	205.0	2018
3	여성	179.4	2018
2	남성	239.0	2019
3	여성	182.0	2019

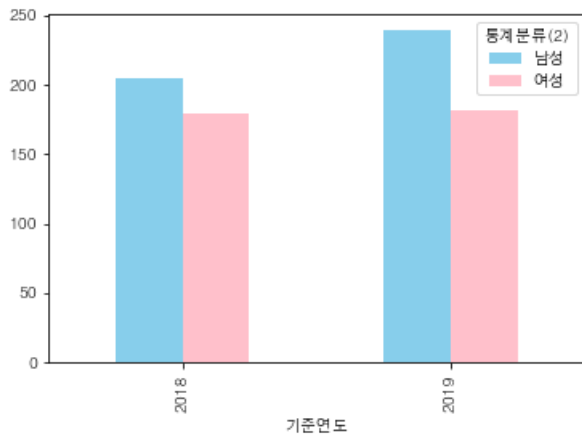
In [15]:

```
sex = pd.pivot_table(data=sex, index='통계분류(2)', columns='기준연도', values='평균')
```

In [16]:

```
sex.T.plot(kind='bar', color=['skyblue', 'pink'])
```

<AxesSubplot:xlabel='기준연도'>



Out[16]:

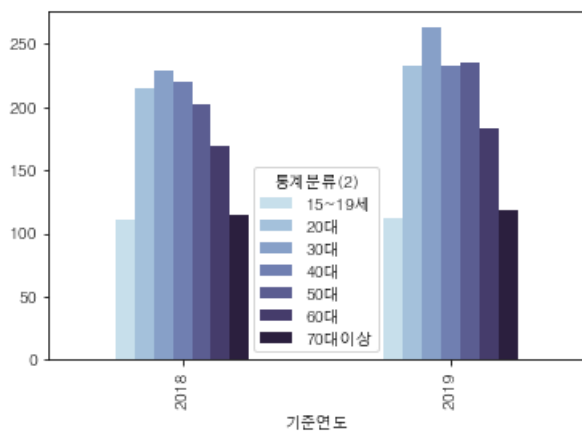
```
age = cost.loc[cost['통계분류(1)']=='연령별',:]
age = age[['통계분류(2)', '평균', '기준연도']]
age['평균'] = age['평균'].astype(float)
age = pd.pivot_table(data=age, columns='통계분류(2)', index='기준연도', values='평균')
```

In [17]:

```
colors = sns.color_palette('ch:s=.25,rot=-.25',len(age.columns))
age.plot(kind='bar', color=colors)
```

In [18]:

<AxesSubplot:xlabel='기준연도'>

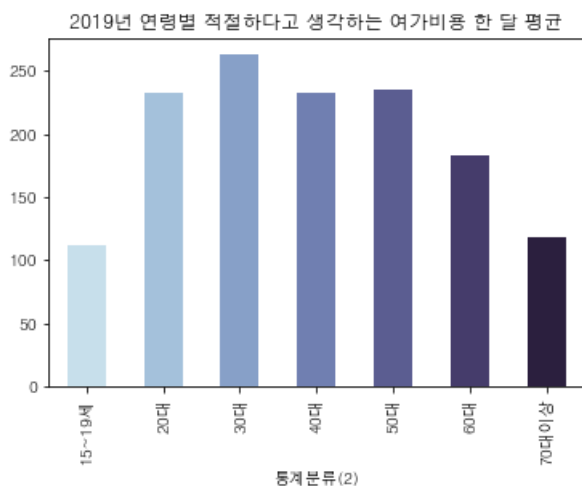


Out[18]:

```
plt.title('2019년 연령별 적절하다고 생각하는 여가비용 한 달 평균')
colors = sns.color_palette('ch:s=.25,rot=-.25',len(age.columns))
age.iloc[1].plot(kind='bar', color=colors)
```

In [19]:

<AxesSubplot:title={'center':'2019년 연령별 적절하다고 생각하는 여가비용 한 달 평균'}, xlabel='통계분류(2)'>



Out[19]:

In []:

In [20]:

```
# 적절하다고 생각하는 여가비용 한 달 평균

sido = cost.loc[(cost['통계분류(1)']=='17개 시도별') & (cost['기준연도']==2019),:]

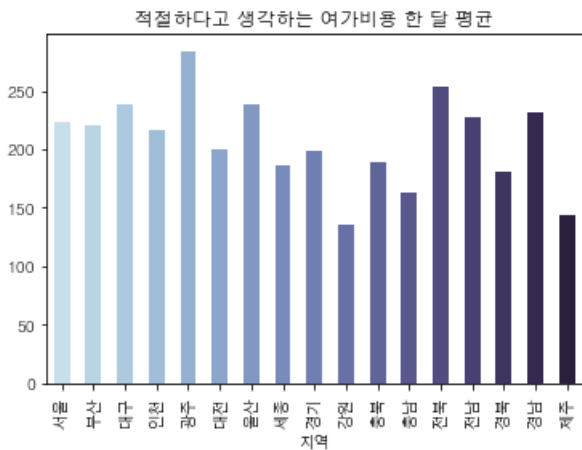
sido = sido.rename(columns={"통계분류(2)": "지역"})
sido.index = sido['지역']
sido = sido.drop(['통계분류(1)', '지역', '기준연도'], axis=1)

colors = sns.color_palette('ch:s=.25,rot=-.25',len(sido['평균']))

plt.title('적절하다고 생각하는 여가비용 한 달 평균')
sido['평균'].astype(int).plot(kind='bar', color=colors)
```

Out[20]:

```
<AxesSubplot:title={'center': '적절하다고 생각하는 여가비용 한 달 평균'}, xlabel='지역'>
```



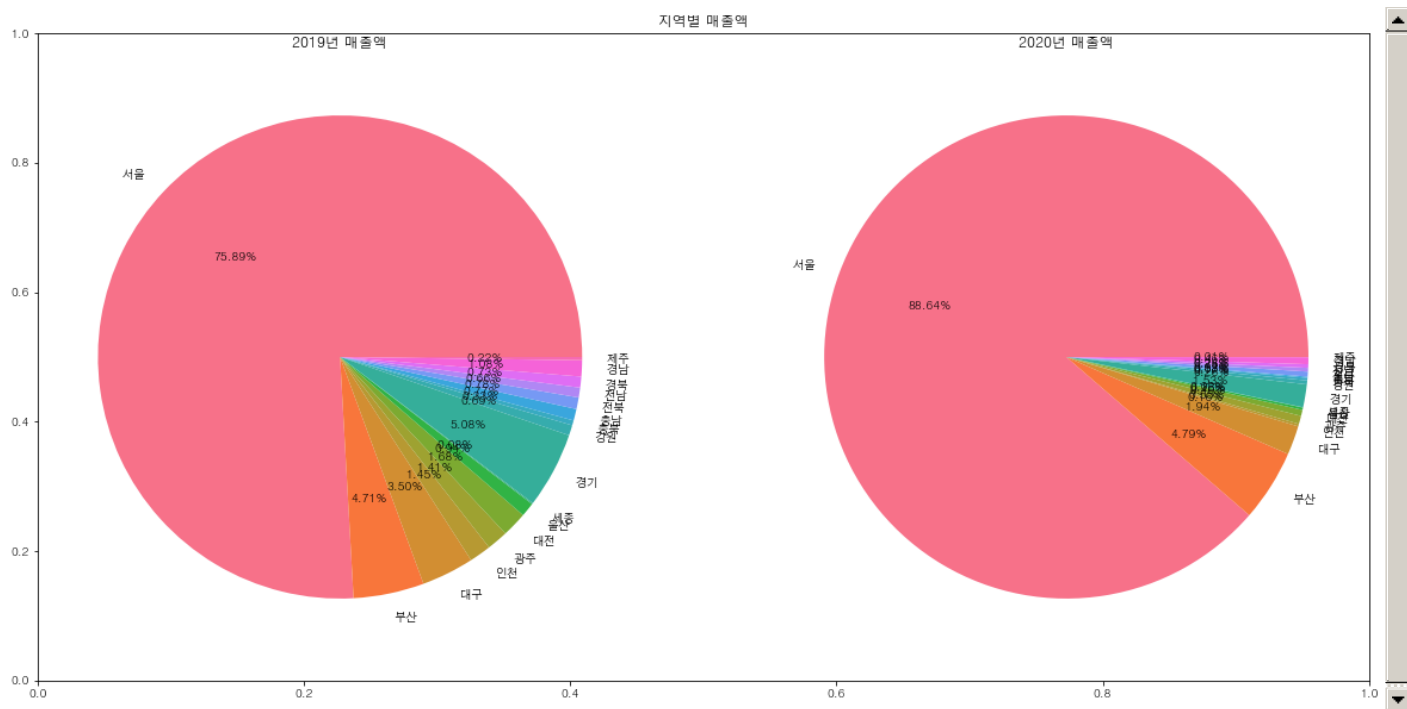
광주 사람들이 문화향유에 대한 욕구가 많음을 알 수 있고 타 지역에도 공연에 대한 수요가 높음을 유추 할 수 있다.

In [22]:

```
fig = plt.figure(figsize = (20, 10))

colors = sns.color_palette('husl',len(df['매출액']))
plt.title('지역별 매출액')
ax1 = fig.add_subplot(1, 2, 1)
plt.title('2019년 매출액')
ax2 = fig.add_subplot(1, 2, 2)
plt.title('2020년 매출액')

ax1.pie(x=list(df['매출액']),labels=df.index, autopct=lambda p : '{:.2f}%'.format(p), colors=colors)
ax2.pie(x=list(df1['매출액']),labels=df1.index, autopct=lambda p : '{:.2f}%'.format(p), colors=colors)
plt.show()
plt.savefig('savefig_default.png')
```



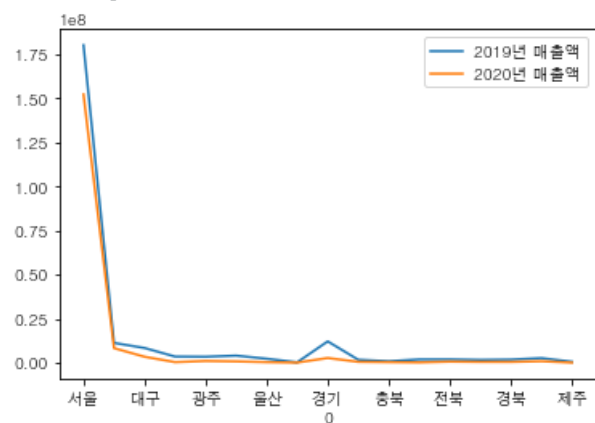
<Figure size 432x288 with 0 Axes>

In [29]:

```
co = pd.concat([df['매출액'],df1['매출액']], axis=1)
co.columns = ['2019년 매출액','2020년 매출액']
co.plot()
```

Out[29]:

```
<AxesSubplot:xlabel='0'>
```

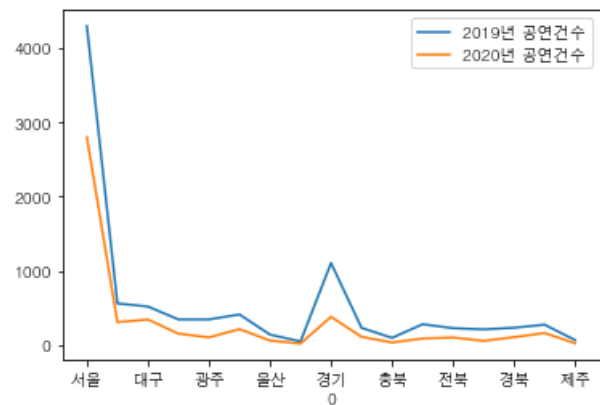


In [24]:

```
perform = pd.concat([df['공연건수'], df1['공연건수']], axis=1)
perform.columns = ['2019년 공연건수', '2020년 공연건수']
perform.plot()
```

Out[24]:

```
<AxesSubplot:xlabel='0'>
```



In [25]:

```
fig = plt.figure(figsize=(20, 10))
```

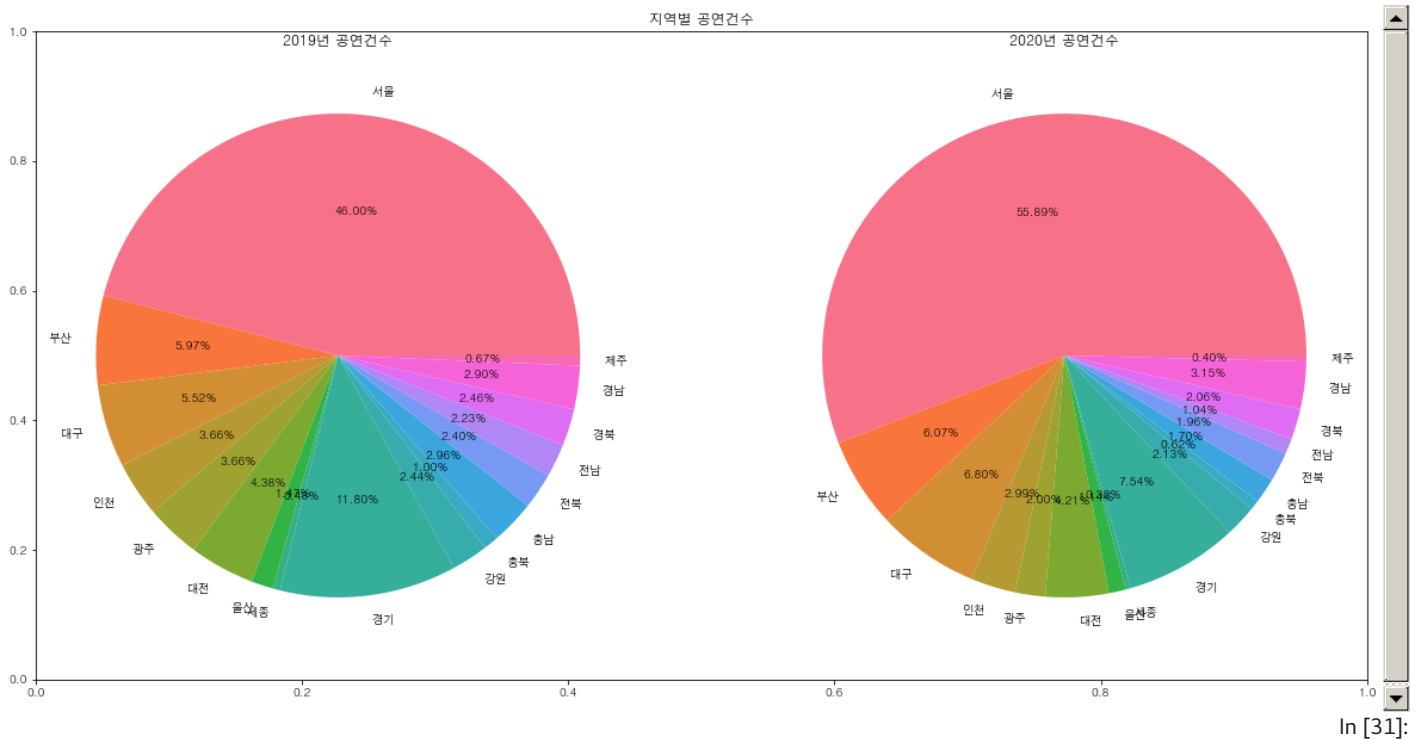
```
fig = plt.figure(figsize = (20, 10))

colors = sns.color_palette('husl',len(df['매출액']))

plt.title('지역별 공연건수')

ax1 = fig.add_subplot(1, 2, 1)
plt.title('2019년 공연건수')
ax2 = fig.add_subplot(1, 2, 2)
plt.title('2020년 공연건수')

ax1.pie(x=list(df['공연건수']),labels=df.index, autopct=lambda p : '{:.2f}%'.format(p), colors=colors)
ax2.pie(x=list(df1['공연건수']),labels=df1.index, autopct=lambda p : '{:.2f}%'.format(p), colors=colors)
plt.show()
```



```
# 매출액 높은 순으로 정리
pd.concat([sido['평균'], df['매출액']], axis=1).sort_values(by='매출액', ascending=False).T
```

Out[31]:

	서울	경기	부산	대구	대전	인천	광주	경남	울산	전북	충남	경북	강원	전남
평균	223	199	221	238	200	217	284	231	238	254	163	181	135	221
매출액	180286137	12078893	11199605	8316336	3982462	3443294	3355096	2561904	2221899	1842649	1837572	1741361	1642379	156136

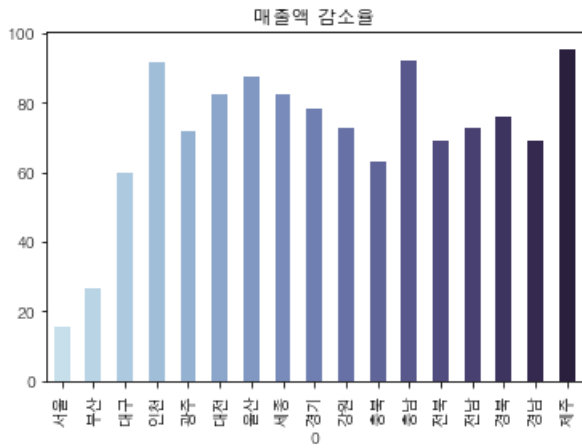
```
# 감소율
dec = (df - df1) / df * 100
colors = sns.color_palette('ch:s=.25,rot=-.25',len(dec['매출액']))

plt.title('매출액 감소율')
dec['매출액'].plot(kind='bar', color=colors)
```

In [27]:

Out[27]:

```
<AxesSubplot:title={'center':'매출액 감소율'}, xlabel='0'>
```



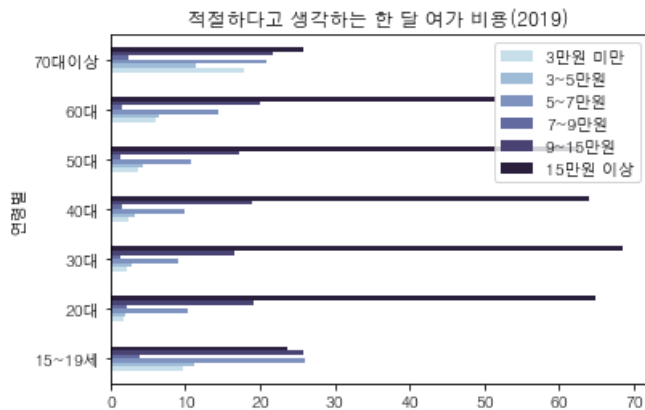
In [28]:

```
cost = pd.read_csv('적절하다고 생각하는 여가비용_한 달_평균.csv')
age = cost.loc[(cost['통계분류(1)']=='연령별'),:]
age = age.drop(['통계분류(1)', '2018', '2018.1', '2018.2', '2018.3', '2018.4', '2018.5', '2018.6', '2018.7', '2019',
age.columns=['연령별', '3만원 미만', '3~5만원', '5~7만원', '7~9만원', '9~15만원', '15만원 이상']
age.set_index('연령별', inplace=True)
age=age.astype(float)

colors = sns.color_palette('ch:s=.25,rot=-.25',6)
age.plot(kind='barh', color=colors)
plt.title('적절하다고 생각하는 한 달 여가 비용 (2019)')
```

Out[28]:

Text(0.5, 1.0, '적절하다고 생각하는 한 달 여가 비용 (2019)')



여가생활에 15만원 이상의 높은 비용을 지불할 의향이 있는 연령층은 주로 30대에서 특히 높게 나타났으며, 50대, 20대/40대가 뒤를 이었다. 10대와 70대 이상 연령층에서는 타 연령층과 비교했을 때 높은 여가비용을 지불할 의향이 현저히 낮음을 확인할 수 있었다.