소비성향 프로파일 데이터 이용

(앞과 동일) # 우편번호 앞 세자리 추출하여 새 컬럼 만들기

rows\_df['nw\_zp\_3'] = list(map(lambda x: x[:3], rows\_df['nw\_zp']))

rows\_df['nw\_zp'] = list(map(lambda x: int(x), rows\_df['nw\_zp']))

rows\_df = rows\_df.sort\_values(by=['nw\_zp'], axis=0).reset\_index(drop=True)

# 서울만 추출

rows\_df = rows\_df[:600838]

# 개인소득금액 chr -> int 변환

rows\_df['indv\_incm\_amt'] = list(map(lambda x: int(x), rows\_df['indv\_incm\_amt']))

# 우편번호 앞 세자리(구 단위)로 grouping

grouped = rows\_df.groupby(rows\_df['nw\_zp\_3'])

rows\_df = rows\_df.drop(columns = ["cent\_cd","load\_dtm"],axis = 1)

rows\_df.columns = ["기준년월","신우편번호","생년","연령대코드","성별코드","세분화코드","개인소득구간코드","가구소득금액","개인소득금액","nw\_zp\_3"]

# 우편번호, 성별에 따른 가구소득금액 합계

groupBynw\_sex = rows\_df.groupby(['nw\_zp\_3','성별코드'])['개인소득금액'].agg(['count','mean','sum']).reset\_index()

groupBynw\_sex.columns = ["nw\_zp\_3","성별","count","개인소득금액평균","개인소득금액 합계"]

# 우편번호, 연령대에 따른 가구소득금액 합계

groupBynw\_age = rows\_df.groupby(['nw\_zp\_3','연령대코드'])['개인소득금액'].agg(['count','mean','sum']).reset\_index()

groupBynw\_age.columns = ["nw\_zp\_3","연령대","count","개인소득금액평균","개인소득금액 합계"]

# 우편번호, 소득구간에 따른 가구소득금액 합계

groupBynw\_int = rows\_df.groupby(['nw\_zp\_3','개인소득구간코드'])['개인소득금액'].agg(['count','mean','sum']).reset\_index()

groupBynw\_int.columns = ["nw\_zp\_3","소득구간","count","개인소득금액평균","개인소득금액 합계"]

강남구 vs 중구 연령대별 소득 비교

def func(data) :

if data['연령대코드'] == '010' or data['연령대코드'] =='015':

return "10대"

elif data['연령대코드'] == '020' or data['연령대코드'] =='025':

return "20대"

elif data['연령대코드'] == '030' or data['연령대코드'] =='035':

return "30대"

elif data['연령대코드'] == '040' or data['연령대코드'] =='045':

return "40대"

elif data['연령대코드'] == '050' or data['연령대코드'] =='055':

return "50대"

elif data['연령대코드'] == '060' or data['연령대코드'] =='065':

return "60대"

else :

return "70대 이상"

# 강남구

gangnam = rows\_df[rows\_df['nw\_zp\_3'] == "063"]

# 연령대, 소득구간, 소득금액 합계

gangnam['연령대코드'] = gangnam['연령대코드'].astype('category')

gangnam['개인소득구간코드'] = gangnam['개인소득구간코드'].astype('category')

gangnam['연령대'] = gangnam.apply(func,axis = 1)

gangnam\_group = gangnam.groupby(['연령대','개인소득구간코드'])['개인소득금액'].agg(['sum']).reset\_index()

# 중구

jung = rows\_df[rows\_df['nw\_zp\_3'] == "045"]

# 연령대, 소득구간, 소득금액 합계

jung['연령대코드'] = jung['연령대코드'].astype('category')

jung['개인소득구간코드'] = jung['개인소득구간코드'].astype('category')

jung['연령대'] = jung.apply(func,axis = 1)

jung\_group = jung.groupby(['연령대','개인소득구간코드'])['개인소득금액'].agg(['sum']).reset\_index()

result = pd.merge(gangnam\_group,jung\_group,on = ["연령대","개인소득구간코드"])

result.columns = ["연령대","개인소득구간코드","강남구소득합계","중구소득합계"]

result

import matplotlib.pyplot as plt

fig = plt.figure(figsize=(12, 8))

ax = fig.add\_subplot()

ax.plot(result.index, result.강남구소득합계, marker='s', color='blue')

ax.plot(result.index, result.중구소득합계, marker='o', color='red')

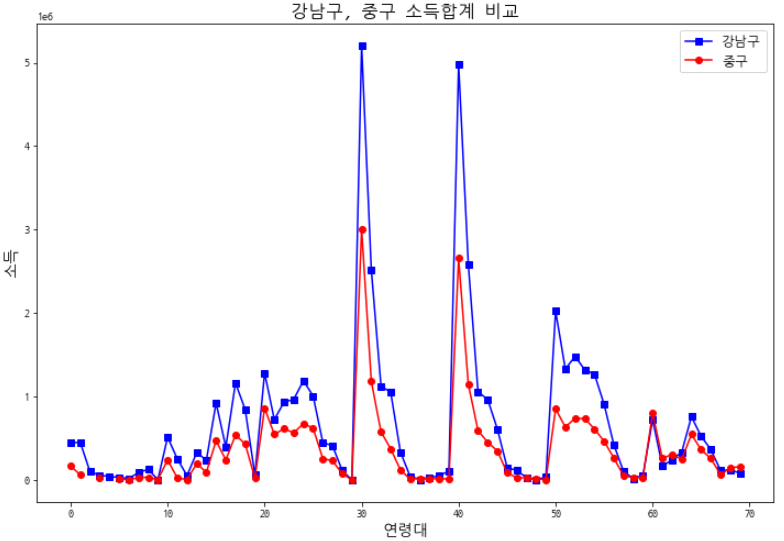
plt.title('강남구, 중구 소득합계 비교', fontsize=16)

plt.xlabel('연령대', fontsize=14)

plt.ylabel('소득', fontsize=14)

plt.legend(['강남구', '중구'], fontsize=12, loc='best')

plt.show()



**지역별 업종별 매출집계**

2021년 9월 기준 지역별 업종코드 표시 & 매출 집계

rows\_df = rows\_df.drop(columns = ["cent\_cd","load\_dtm"],axis = 1)

rows\_df.columns = ["기준일자","신우편번호","업종코드","평균건수","평균금액"]

# 우편번호 앞 세자리 추출하여 새 컬럼 만들기

rows\_df['nw\_zp\_3'] = list(map(lambda x: x[:3], rows\_df['신우편번호']))

rows\_df['신우편번호'] = list(map(lambda x: int(x), rows\_df['신우편번호']))

rows\_df = rows\_df.sort\_values(by=['신우편번호'], axis=0).reset\_index(drop=True)

# 서울시 추출

rows\_df['nw\_zp\_1'] = rows\_df['nw\_zp\_3'].astype(int)

seoul = (rows\_df['nw\_zp\_1']>=10) & (rows\_df['nw\_zp\_1'] <= 90)

rows\_df = rows\_df.loc[seoul]

rows\_df = rows\_df.drop(columns = ["nw\_zp\_1"],axis = 1)

# 우편번호 앞 세자리(구 단위)로 grouping

grouped = rows\_df.groupby(rows\_df['nw\_zp\_3'])

rows\_df['기준일자'] = pd.to\_datetime(rows\_df['기준일자'])

rows\_df.describe(include="object")

rows\_df.groupby('nw\_zp\_3')['업종코드'].unique()

rows\_df = rows\_df[["nw\_zp\_3","기준일자","업종코드","평균건수","평균금액"]]

rows\_df = rows\_df.sort\_values(['nw\_zp\_3','기준일자']).reset\_index(drop = True)

rows\_df['평균건수'] = rows\_df['평균건수'].astype(int)

rows\_df['평균금액'] = rows\_df['평균금액'].astype(int)

# 2021년 9월 업종코드별 매출건수 합계

grouped\_nw\_tp\_cnt\_sum = rows\_df.groupby(['nw\_zp\_3','업종코드'])['평균건수'].agg(\*\*{'9월 매출건수 합계':'sum'}).reset\_index()

# 2021년 9월 업종코드별 매출금액 합계

grouped\_nw\_tp\_amtsum = rows\_df.groupby(['nw\_zp\_3','업종코드'])['평균금액'].agg(\*\*{'9월 매출금액 합계':'sum'}).reset\_index()

# 2021년 9월 업종코드별 평균매출금액

grouped\_nw\_tp\_amtmean = rows\_df.groupby(['nw\_zp\_3','업종코드'])['평균금액'].agg(\*\*{'9월 평균매출금액':'mean'}).reset\_index()

result = pd.merge(grouped\_nw\_tp\_cnt\_sum,grouped\_nw\_tp\_amtsum,on = ["nw\_zp\_3","업종코드"])

result = pd.merge(result,grouped\_nw\_tp\_amtmean,on = ["nw\_zp\_3","업종코드"])

result