거래내역(매출내역) 데이터 사용

select \* from bbp\_prd\_tabown.\_\_ where substring(strd\_date,1,6)='202109'

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

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

rows\_df[rows\_df['cstmr\_nw\_zp']==8866]

# 서울만 추출

rows\_df = rows\_df[:253605]

**[연령대&성별]**

rows\_df['mer\_tpbuz\_cd'].nunique() #83개 업종코드

df = rows\_df[['cstmr\_nw\_zp','gn\_cd','sex\_cd', 'sale\_dv\_cd', 'mer\_tpbuz\_cd','sale\_amt']]

df['sale\_amt'] = list(map(lambda x: int(x), df['sale\_amt']))

**[업종명 데이터와 merge]**

14번 데이터(비씨카드 플랫폼 업종별코드 정보) 컬럼 사용

topic = rows\_df[['tpbuz\_cd','tpbuz\_nm']]

topic.rename(columns={'tpbuz\_cd':'mer\_tpbuz\_cd'}, inplace=True

df = pd.merge(df, topic, on='mer\_tpbuz\_cd')

# 각 케이스별 개수 카운트(추후 평균 등 계산 위해)  
a=pd.DataFrame(df[['cstmr\_nw\_zp','gn\_cd','sex\_cd','sale\_dv\_cd','mer\_tpbuz\_cd','tpbuz\_nm']].value\_counts().reset\_index())

+ 'sale\_dv\_cd' 매출구분코드는 할부인지 등등 구분

04: DEBIT/ 05: 일시불/ 06: 현금서비스/ 07: 회전결제/ 08: 할부/ 09: 즉시블

# 최종 사용할 데이터 프레임 생성

sales= pd.merge(df, a, on=['cstmr\_nw\_zp','gn\_cd','sex\_cd', 'sale\_dv\_cd', 'mer\_tpbuz\_cd','tpbuz\_nm'])

sales.rename(columns = { 0 : 'counts'}, inplace = True)

sales.loc[(sales.sex\_cd == 'M'),'sex\_cd']= 0

sales.loc[(sales.sex\_cd == 'F'), 'sex\_cd']=1

**[연령대 별]**

sales\_age = sales\_age.drop(['sex\_cd'], axis=1)

sales\_age =sales\_age.groupby(['cstmr\_nw\_zp','gn\_cd','sale\_dv\_cd','mer\_tpbuz\_cd','tpbuz\_nm']).sum().reset\_index()

**[성별]**

sales\_sex = sales.drop(['gn\_cd'], axis=1)

sales\_sex=sales\_sex.groupby(['cstmr\_nw\_zp','sex\_cd','sale\_dv\_cd', 'mer\_tpbuz\_cd','tpbuz\_nm']).sum().reset\_index()

(성별 기준 PLOT 한번 그려보기)

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

plt.rcParams['figure.figsize'] = [5,5] # size 지정

b = sales\_sex.groupby(['sex\_cd','mer\_tpbuz\_cd']).sum().reset\_index()

for i in sorted(set(b['mer\_tpbuz\_cd'])):

print(i)

tmp= b[b['mer\_tpbuz\_cd']==i]

g = sns.barplot(data= tmp ,x= "sex\_cd" ,y= "sale\_amt", ci=None)

ax = g

#annotate axis = seaborn axis

for p in ax.patches: ax.annotate("%.0f" % p.get\_height(), (p.get\_x() + p.get\_width()/2., p.get\_height() - 30), ha='center', va='center', fontsize=10, color='black', xytext=(0, 10), textcoords='offset points')

plt.xticks(fontsize=10)

plt.show()