20182791 권유진¶

20182806 서동혁¶

20182832 최준용¶

In [1]:

```
import pandas as pd
import numpy as np
import arrow
import seaborn as sns
import matplotlib.pylab as plt
%matplotlib inline
```

In [2]:

```
plt.rc('font', family='malgun gothic')
plt.rc('axes', unicode_minus=False)
```

데이터 정제 작업¶

In [3]:

```
demo = pd.read_csv('Demo.csv', encoding = 'cp949')
tran = pd.read_csv('구매내역정보.csv', encoding = 'cp949', engine = 'python')
```

In [4]:

```
mg = pd.merge(demo, tran, on = 'ID')
```

In [5]:

```
mg = mg.drop(mg.query('구매금액 == 0').reset_index()['index'],axis=0) # 구매금액이 0 원인 것 지움.
```

In [6]:

```
mg['평균금액'] = (mg['구매금액']/mg['구매수량']).astype(int)

# 개별 금액이 무의미해 보이는 수치들 제거
mg = mg.drop(mg.query('평균금액 < 100').reset_index()['index'],axis = 0)
mg = mg.drop(mg.query('평균금액 <= 200').reset_index()['index'], axis = 0)
mg = mg.drop(mg[mg.평균금액.apply(lambda x: str(x)[-1]) != "0"].reset_index()['index'], axis = 0)
mg = mg.reset_index().drop('index', axis=1)
```

```
mg.head()
```

Out[6]:

	ID	성별	연령	거주지역	상품대분류명	상품중분류명	구매지역	구매일자	구매시간	구매수량	구매금액	취소여부	평균금액
0	47820794 6	1	84	서울 성동구	가전제품	컴퓨터주변기기	서울 동대문구	20141219	13	1	59000	0	59000
1	47820794 6	1	84	서울 성동구	가전제품	TV/AV	서울 동대문구	20141031	14	1	106000	0	106000
2	47820794 6	1	84	서울 성동구	가전제품	주방가전	서울 중구	20140815	15	1	37000	0	37000
3	47820794 6	1	84	서울 성동구	의류잡화	여성용의류-이너웨어	서울 동대문구	20140322	17	1	118000	0	118000
4	47980698 4	1	84	서울 서초구	생활잡화	화장품	서울 중구	20140704	12	1	22000	0	22000

In [7]:

```
mg.to_csv('merging_data.csv')
```

데이터 분석¶

In [8]:

```
mg = pd.read_csv('merging_data.csv')
```

In [9]:

```
mg['구매월'] = mg.구매일자.apply(lambda x: int(str(x)[4:6]))
```

In [10]:

```
anm = pd.DataFrame()
anm['구매수량'] = mg.query('구매금액>0').groupby('상품대분류명')['구매수량'].sum()
anm['구매금액'] = mg.query('구매금액>0').groupby('상품대분류명')['구매금액'].sum()
anm['판매량 대비 매출'] = [round(anm['구매금액'][i]/anm['구매수량'][i],1) for i in range(len(anm['구매수량']))]
anm['판매금액비율'] = list(map(lambda x: round(anm.구매금액[x]/anm.구매금액.sum()*100,1),
range(len(anm.구매금액))))
anm = anm.reset_index()
anm
```

Out[10]:

	상품대분류명	구매수량	구매금액	판매량 대비 매출	판매금액비율

0	가구	686	189902000	276825.1	2.0
1	가전제품	11157	2733407000	244994.8	29.4
2	레포츠	7756	835010000	107659.9	9.0
3	명품	1046	577512000	552114.7	6.2
4	생활잡화	10138	719849000	71005.0	7.7
5	식품	43631	529767000	12142.0	5.7
6	의류잡화	29079	3714773000	127747.6	39.9

In [11]:

```
plt.pie(anm.구매수량, labels = anm.상품대분류명, shadow=True, autopct='%1.1f%%')
plt.title('상품대분류별 판매금액비율')
plt.show()
```

In [12]:

```
sns.catplot(y='판매량 대비 매출', x='상품대분류명', data = anm, kind = 'bar')
plt.title('대분류별 판매효율')

plt.show()
```

In [13]:

```
refund = pd.DataFrame()
refund['환불수량'] = mg.query('구매수량<0').groupby('상품대분류명')['구매수량'].sum().apply(lambda x: -int(x))
refund['환불금액'] = mg.query('구매수량<0').groupby('상품대분류명')['구매금액'].sum().apply(lambda x: -int(x))
refund['판매량 대비 환불량'] = [round(refund['환불수량'][i]/anm['구매수량'][i]*100,1) for i in
range(len(refund['환불수량']))]
```

```
refund['금액 비율'] = [round(refund['환불금액'][i]/anm['구매금액'][i]*100,1) for i in range(len(refund['환불수량']))]
refund=refund.reset_index()
refund
```

Out[13]:

	상품대분류명	환불수량	환불금액	판매량 대비 환불량	금액 비율
0	가구	57	26991000	8.3	14.2
1	가전제품	484	114741000	4.3	4.2
2	레포츠	1402	161992000	18.1	19.4
3	명품	211	124634000	20.2	21.6
4	생활잡화	876	109137000	8.6	15.2
5	식품	952	22103000	2.2	4.2
6	의류잡화	5192	1016649000	17.9	27.4

In [14]:

```
sns.catplot(x='상품대분류명', y='환불금액', kind = 'bar', data = refund)
plt.title('상품대분류별 판매량 대비 환불량')
plt.show()
```

In [15]:

```
result = pd.DataFrame()
result['최종판매량'] = mg.groupby('상품대분류명')['구매수량'].sum()
result['최종판매금액'] = mg.groupby('상품대분류명')['구매금액'].sum()
result['매출비율'] = list(map(lambda x: round(result.최종판매금액[x]/result.최종판매금액.sum()*100,1),
```

```
range(len(result.최종판매금액)))
result = result.reset_index()
result
```

Out[15]:

	상품대분류명	최종판매량	최종판매금액	매출비율
0	가구	629	162911000	2.1
1	가전제품	10673	2618666000	33.9
2	레포츠	6354	673018000	8.7
3	명품	835	452878000	5.9
4	생활잡화	9262	610712000	7.9
5	식품	42679	507664000	6.6
6	의류잡화	23887	2698124000	34.9

In [16]:

```
plt.pie(result.매출비율, labels = result.상품대분류명, shadow=True,autopct='%1.1f%%')
plt.title('총 매출 비율')
plt.show()
```

In [17]:

```
Mb1 = mg.groupby('구매월')['구매수량'].sum().reset_index()
sns.relplot(x='구매월', y='구매수량', kind='line', data = Mb1, color = 'black', marker='o', mfc='red')
plt.title('월별 총구매수량')
plt.xticks(range(1,13))
```

```
plt.ylim(5000,)
plt.grid()

plt.show()
```

In [18]:

```
Mb2 = mg.groupby(['구매월', '상품대분류명'])['구매수량'].sum().reset_index()
sns.relplot(x='구매월', y='구매수량', kind='line', data = Mb2, hue='상품대분류명', marker='o')
plt.title('상품대분류별 총구매수량 변화')
plt.xticks(range(1,13))
plt.grid()

plt.show()
```

In [19]:

```
Mb2 = mg.query('상품대분류명 in ["가전제품", "레포츠","명품","생활잡화","가구"]').groupby(['구매월',
'상품대분류명'])['구매수량'].sum().reset_index()
sns.relplot(x='구매월', y='구매수량', kind='line', data = Mb2, hue='상품대분류명', marker='o')
plt.title('상품대분류별 총구매수량 변화')
plt.xticks(range(1,13))
plt.grid()

plt.show()
```

In [20]:

```
Mb2 = mg.groupby(['구매월', '상품대분류명'])['구매금액'].sum().reset_index()
sns.relplot(x='구매월', y='구매금액', kind='line', data = Mb2, hue='상품대분류명', marker='o')
plt.title('상품대분류별 총구매금액 변화')
plt.xticks(range(1,13))
plt.grid()
```

```
plt.show()
```

In [21]:

```
Mb2 = mg.query('상품대분류명 in ["가구","레포츠","명품","생활잡화","식품"]').groupby(['구매월',
'상품대분류명'])['구매금액'].sum().reset_index()
sns.relplot(x='구매월', y='구매금액', kind='line', data = Mb2, hue='상품대분류명', marker='o')
plt.title('상품대분류별 총구매금액 변화')
plt.xticks(range(1,13))
plt.grid()

plt.show()
```

In [22]:

```
Mb3 = mg.query('상품대분류명=="가전제품"').groupby(['구매월', '상품중분류명'])['구매수량'].sum().reset_index()
sns.relplot(x='구매월', y='구매수량', kind='line', data = Mb3, hue='상품중분류명', marker = 'o')
plt.title('가전제품 중분류별 구매수량 변화')
plt.xticks(range(1,13))
plt.grid()

plt.show()
```

In [23]:

```
Mb3_1 = mg.query('상품중분류명=="휴대폰/태블릿" or 상품중분류명=="PC/노트북/프린터/카메라"').groupby(['구매월',
'상품중분류명'])['구매수량'].sum().reset_index()
sns.relplot(x='구매월', y='구매수량', kind='line', data = Mb3_1, hue='상품중분류명', marker = 'o')
plt.title('가전제품 중분류별 구매수량 변화')
plt.xticks(range(1,13))
plt.grid()

plt.show()
```

In [24]:

```
mb3 = pd.merge(Mb3.query('구매월 == 7'),Mb3.query('구매월 ==
8'),on='상품중분류명')[['상품중분류명','구매수량_x','구매수량_y']].rename(columns={'구매수량_x':'7 월
구매수량','구매수량_y':'8 월 구매수량'})
mb3['x 배 증가'] = round(mb3['8 월 구매수량']/mb3['7 월 구매수량'],1)
mb3
```

Out[24]:

	상품중분류명	7 월 구매수량	8 월 구매수량	x 배 증가
0	PC/노트북/프린터/카메라	17	54	3.2
1	TV/AV	19	173	9.1
2	생활가전	148	440	3.0
3	주방가전	228	555	2.4
4	컴퓨터주변기기	38	147	3.9
5	휴대폰/태블릿	21	83	4.0

In [25]:

```
Mb3 = mg.query('상품대분류명=="가전제품"').groupby(['구매월', '상품중분류명'])['구매금액'].sum().reset_index()
sns.relplot(x='구매월', y='구매금액', kind='line', data = Mb3, hue='상품중분류명', marker = 'o')
plt.title('가전제품 중분류별 구매금액 변화')
plt.xticks(range(1,13))
plt.grid()

plt.show()
```

In [26]:

```
Mb3 = mg.query('상품중분류명=="휴대폰/태블릿" or 상품중분류명=="PC/노트북/프린터/카메라" or
상품중분류명=="컴퓨터주변기기"').groupby(['구매월', '상품중분류명'])['구매금액'].sum().reset_index()
sns.relplot(x='구매월', y='구매금액', kind='line', data = Mb3, hue='상품중분류명', marker = 'o')
plt.title('가전제품 중분류별 구매금액 변화')
plt.xticks(range(1,13))
plt.grid()

plt.show()
```