

6장. 다중변수 자료의 분석

학번	32233421	이름	이은지
----	----------	----	-----

5장에서 제공된 user_behavior_dataset.csv 는 스마트폰 모델별 사용자 이용 패턴을 정리한 데이터셋이다. 다음과 같이 컬럼 추출 및 컬럼 이름을 변경한 후 문제를 해결 하시오

```
import pandas as pd

df = pd.read_csv('d:/data/user_behavior_dataset.csv')

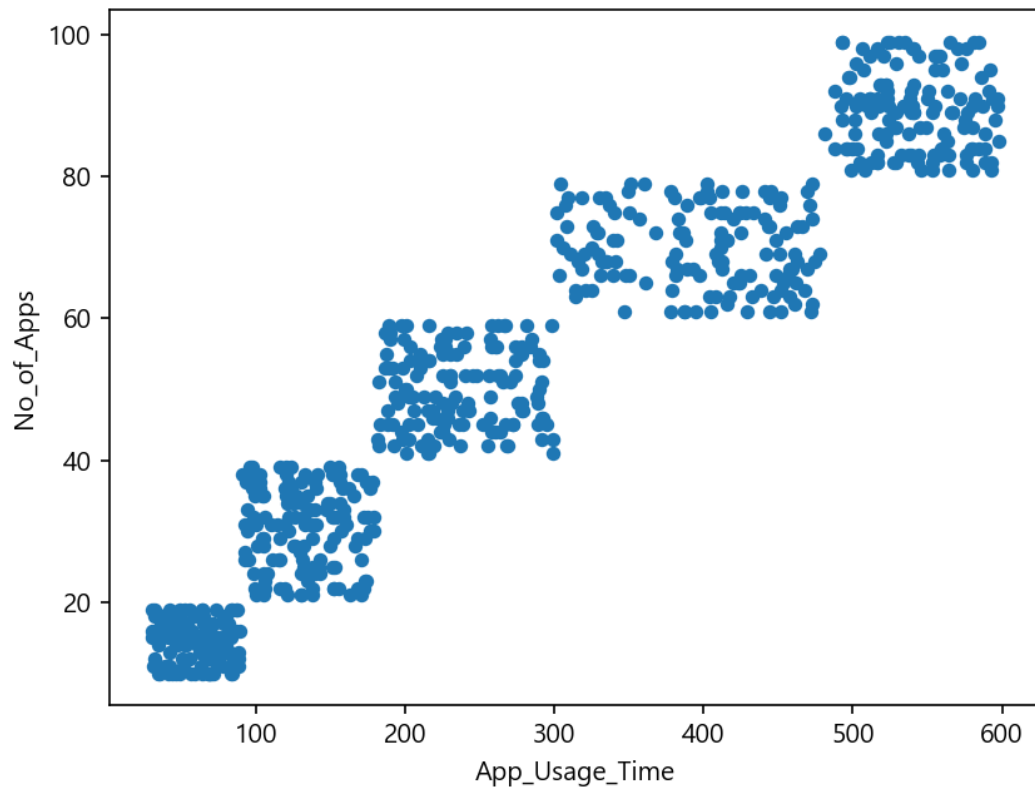
df = df.iloc[:,3:10]

df.columns = [ 'App_Usage_Time', 'Screen_On_Time', 'Battery_Drain',
               'No_of_Apps', 'Data_Usage', 'Age', 'Gender']
```

(각 문제별로 코드 & 실행결과 화면을 캡처를 제출. 실행 결과가 길 경우에는 결과의 앞쪽 일부분만 자라서 캡처해도 됨)

*** 그래프 제목, x축레이블, y축 레이블은 각자 알아서 정하여 표시한다.**

1. 'App_Usage_Time'을 x축으로, 'No_of_Apps' 을 y축으로 하는 산점도를 작성하시오

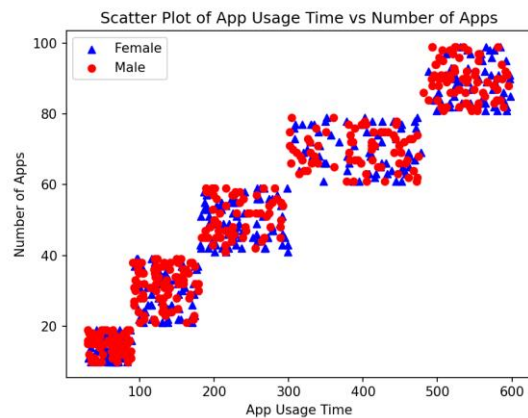


2. 'App_Usage_Time'을 x축으로, 'No_of_Apps' 을 y축으로 하는 산점도를 작성하되

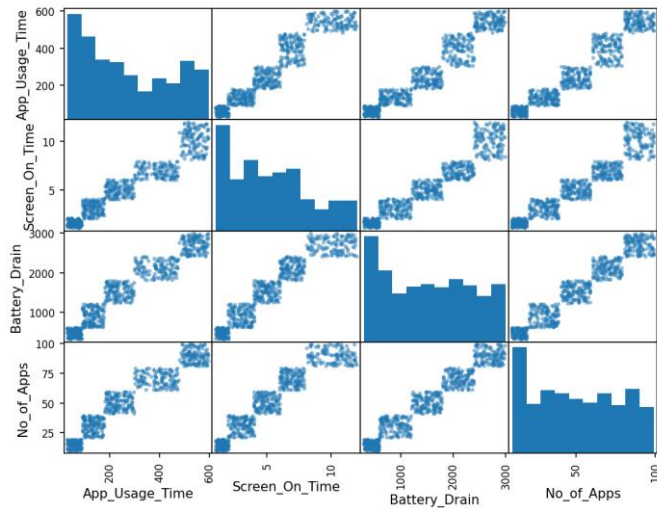
'Gender' 별로 점의 모양과 색을 달리하여 표시하시오

Male - 점의 모양: 'o', 점의 색:빨강

Female - 점의 모양: '^', 점의 색:파랑



3. 'App_Usage_Time', 'Screen_On_Time', 'Battery_Drain', 'No_of_Apps'에 대해 여러 변수들간 산점도를 작성하여 보이시오



4. 두 변수 'App_Usage_Time', 'No_of_Apps' 의 상관 계수를 구하시오

```
>>> df['App_Usage_Time'].corr(df['No_of_Apps'])
np.float64(0.955252834422547)
```

5. 'Gender'를 제외한 변수들간의 상관계수를 보이시오. 가장 상관도가 높은 변수의 조합과 가장 상관도가 낮은 변수의 조합을 보이시오

```
>>> df2=df.loc[:, ~df.columns.isin(['Gender'])]
>>> df2_corr=df2.corr()
>>> max=-1
>>> min =1
>>> max_pair=None
>>> min_pair=None
>>> for i in range(len(df2_corr.columns)):
...     for j in range(i+1, len(df2_corr.columns)):
...         var1=df2_corr.columns[i]
...         var2=df2_corr.columns[j]
...         value=df2_corr.iloc[i,j]
...         if value > max:
...             max=value
...             max_pair=(var1, var2)
...         if value<min:
...             min=value
...             min_pair=(var1, var2)
...
>>> print("가장 상관도 높은 변수 조합",max_pair,"=", max)
가장 상관도 높은 변수 조합 ('Battery_Drain', 'No_of_Apps') = 0.9618532619047168
>>> print("가장 상관도 낮은 변수 조합",min_pair,"=", min)
가장 상관도 낮은 변수 조합 ('Battery_Drain', 'Age') = -0.002722411868857505
```

6. 제공된 students.csv 파일에는 연도별로 초중고 학생수가 저장되어 있다.

데이터사이언스

	A	B	C	D
1	연도	초등학교	중학교	고등학교
2	1999	3,935,537	1,896,956	2,251,140
3	2000	4,019,991	1,860,539	2,071,468
4	2001	4,089,429	1,831,152	1,911,173
5	2002	4,138,366	1,841,030	1,795,509

연도별로 초중고 학생수를 선 그래프로 작성하시오. (초중고를 각각 하나의 선으로 표시하고 색을 달리하시오. 그래프에 한글이 표시되도록 하시오)

