여름방학 Python 스터디

- 스터디 소개 -

데이터 분석을 위한 기초 파이썬 문법 습득
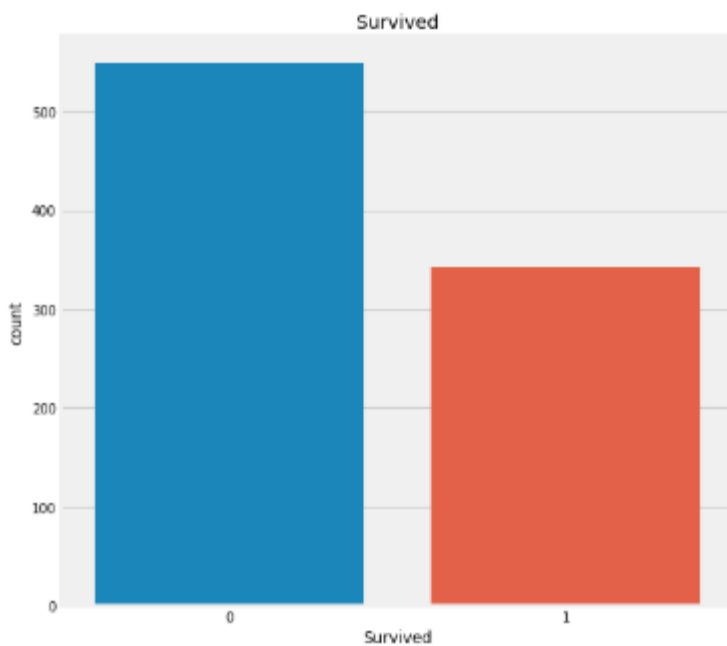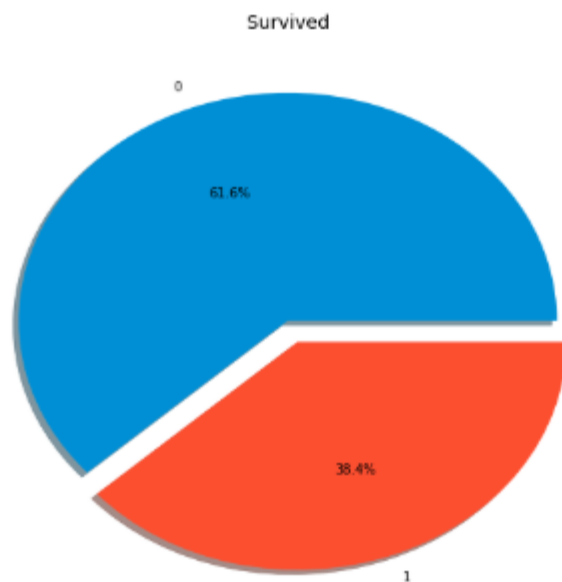
-데이터 분석
-파이썬 문법

데이터 분석

|	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

데이터 분석

```
f,ax=plt.subplots(1,2,figsize=(18,8))
data['Survived'].value_counts().plot.pie(explode=[0,0.1],autopct='%1.1f%%',ax=ax[0],shadow=True
)
ax[0].set_title('Survived')
ax[0].set_ylabel('')
sns.countplot('Survived',data=data,ax=ax[1])
ax[1].set_title('Survived')
plt.show()
```

데이터 분석

```python
f,ax=plt.subplots(1,2,figsize=(20,10))
data[data['Survived']==0].Age.plot.hist(ax=ax[0],bins=20,edgecolor='black',color='red')
ax[0].set_title('Survived= 0')
x1=list(range(0,85,5))
ax[0].set_xticks(x1)
data[data['Survived']==1].Age.plot.hist(ax=ax[1],color='green',bins=20,edgecolor='black')
ax[1].set_title('Survived= 1')
x2=list(range(0,85,5))
ax[1].set_xticks(x2)
plt.show()
```
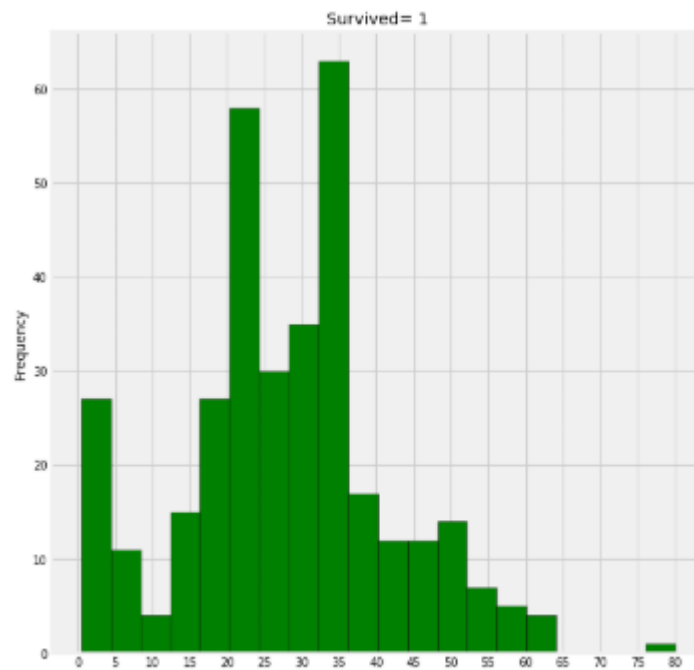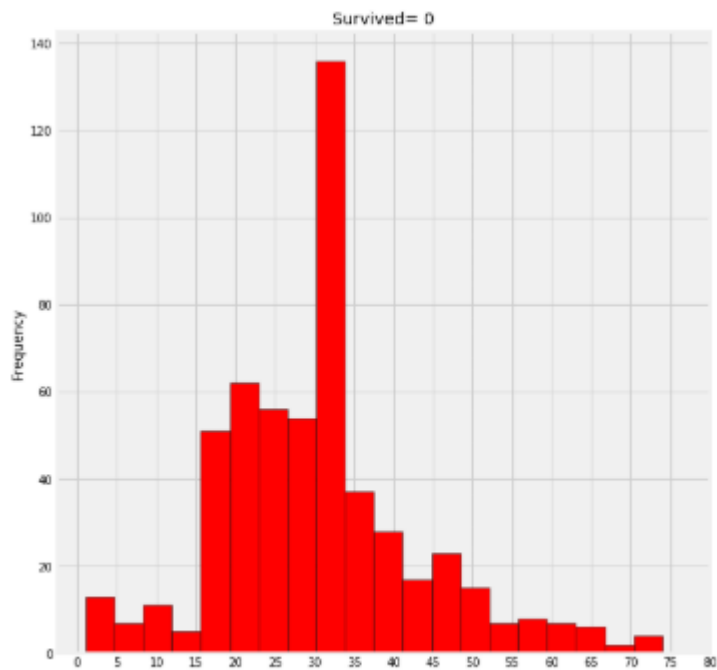
Python

```python
level = 'gds_grp_nm'

df_all = pd.concat([df_train, df_test])
train1 = pd.pivot_table(df_all, index='cust_id', columns=level, values='amount',
                        aggfunc=lambda x: np.where(len(x) >=1, 1, 0), fill_value=0). \
                        reset_index(). \
                        query('cust_id not in @IDtest'). \
                        drop(columns=['cust_id']).values
test1 = pd.pivot_table(df_all, index='cust_id', columns=level, values='amount',
                       aggfunc=lambda x: np.where(len(x) >=1, 1, 0), fill_value=0). \
                       reset_index(). \
                       query('cust_id in @IDtest'). \
                       drop(columns=['cust_id']).values


level = 'gds_grp_mclas_nm'

df_all = pd.concat([df_train, df_test])
train2 = pd.pivot_table(df_all, index='cust_id', columns=level, values='amount',
                        aggfunc=lambda x: np.where(len(x) >=1, 1, 0), fill_value=0). \
                        reset_index(). \
                        query('cust_id not in @IDtest'). \
                        drop(columns=['cust_id']).values
test2 = pd.pivot_table(df_all, index='cust_id', columns=level, values='amount',
                       aggfunc=lambda x: np.where(len(x) >=1, 1, 0), fill_value=0). \
                       reset_index(). \
                       query('cust_id in @IDtest'). \
                       drop(columns=['cust_id']).values


level = 'goods_id'

df_all = pd.concat([df_train, df_test])
train3 = pd.pivot_table(df_all, index='cust_id', columns=level, values='amount',
                        aggfunc=lambda x: np.where(len(x) >=1, 1, 0), fill_value=0). \
                        reset_index(). \
                        query('cust_id not in @IDtest'). \
                        drop(columns=['cust_id']).values
test3 = pd.pivot_table(df_all, index='cust_id', columns=level, values='amount',
                       aggfunc=lambda x: np.where(len(x) >=1, 1, 0), fill_value=0). \
                       reset_index(). \
```

커리큘럼

1주차 : 작업환경 설정 / 변수할당 / 자료구조 / 조건문

2주차 : Pandas,Numpy를 활용한 데이터 분석

3주차 : Pandas,Numpy를 활용한 데이터 분석 (심화)

4주차 : 시각화

➢ 멘토 – 멘티

➢ 2회차 수업 & 1주차 과제

➢ 실습위주 / 코딩이 처음

지각 , 결석