

Week 3 (후반부) Homework 김윤서

<Stage 2>

미션\$1

```
# Fare 빈칸을 평균으로 채우기
df['Fare'] = df['Fare'].fillna(df['Fare'].mean())
df_test['Fare'] = df_test['Fare'].fillna(df_test['Fare'].mean())
```

df

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	2.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	3.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	2.0	0	0	STON/O2, 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	3.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	3.0	0	0	373450	8.0500
...
886	887	0	2	Montvila, Rev. Juozas	male	2.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	1.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	2.0	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	2.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	3.0	0	0	370376	7.7500

미션\$2

```
df['Embarked'].value_counts()
```

```
S    646
C    168
Q     77
Name: Embarked, dtype: int64
```

```
df['Pclass'].value_counts()
```

```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

```
df['Embarked'].isnull().sum()
```

```
2
```

```
df['Embarked'] = df['Embarked'].fillna("S")
```

Challenge 1

```
df.loc[df['Sex'] == 'male', 'Sex'] = 0
df.loc[df['Sex'] == 'female', 'Sex'] = 1

df_test.loc[df_test['Sex'] == 'male', 'Sex'] = 0
df_test.loc[df_test['Sex'] == 'female', 'Sex'] = 1

df.head()
```

```
df.loc[df['Embarked'] == 'S', 'Embarked'] = 0
df.loc[df['Embarked'] == 'C', 'Embarked'] = 1
df.loc[df['Embarked'] == 'Q', 'Embarked'] = 2

df_test.loc[df_test['Embarked'] == 'S', 'Embarked'] = 0
df_test.loc[df_test['Embarked'] == 'C', 'Embarked'] = 1
df_test.loc[df_test['Embarked'] == 'Q', 'Embarked'] = 2

df.head()
```

```
train = df[['Survived', 'Sex', 'Age', 'Fare', 'Embarked']]
test = df_test[['Sex', 'Age', 'Fare', 'Embarked']]

train.head()
```

* 결과

	Survived	Sex	Age	Fare	Embarked
0	0	0	22.0	7.2500	0
1	1	1	38.0	71.2833	1
2	1	1	26.0	7.9250	0
3	1	1	35.0	53.1000	0
4	0	0	35.0	8.0500	0

Challenge 2

```
import pandas as pd
import numpy as np

df = pd.read_csv('data/train.csv')
df_test = pd.read_csv('data/test.csv')
```

In [67]:

```
import matplotlib.pyplot as plt
%matplotlib inline

import seaborn as sns
sns.set()
```

In [68]:

```
sns.catplot(data=df, x='Embarked', y='Age', hue='Sex', kind='box', aspect=4)
```

● 결과

