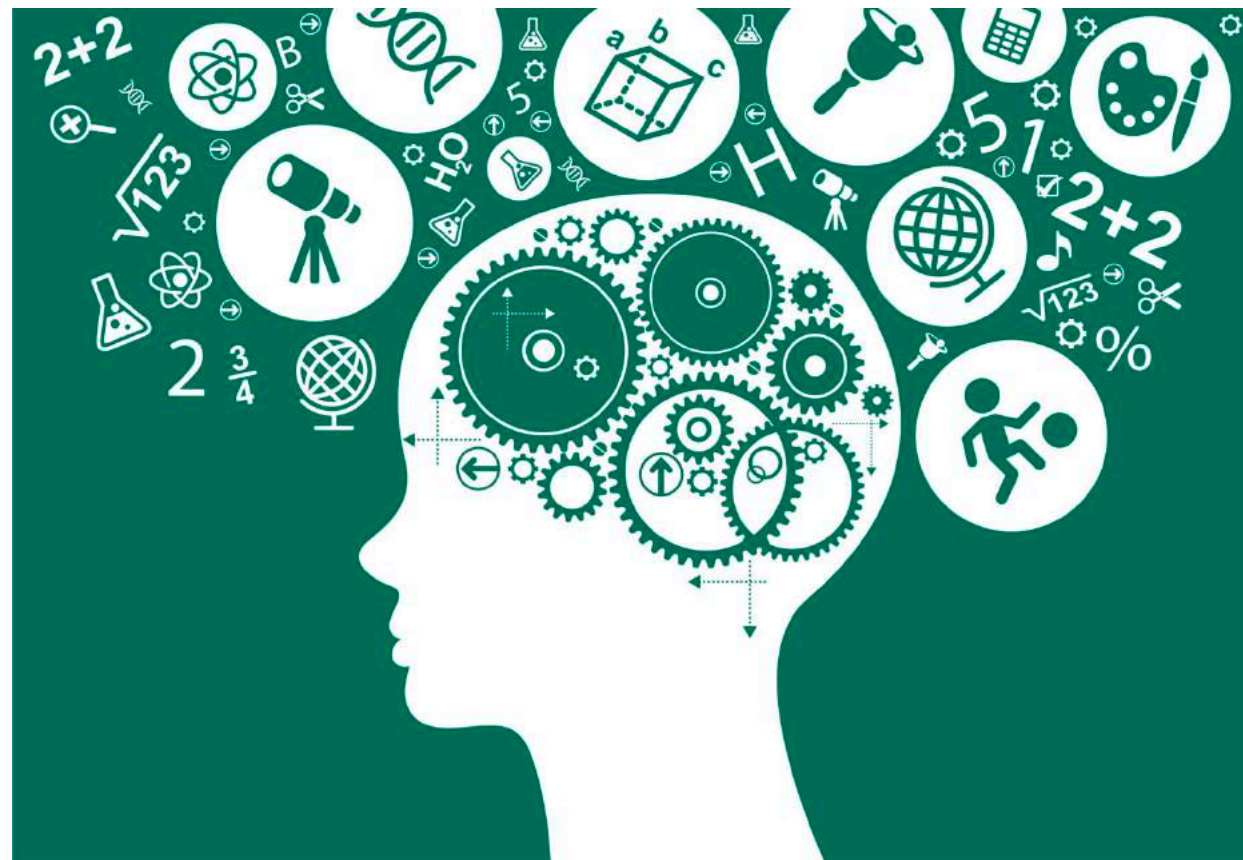


목차

1. 머신러닝의 정의와 분류
2. 캐글소개
3. 데이터 분석 Flow
4. 데이터 시각화 및 분석 실습

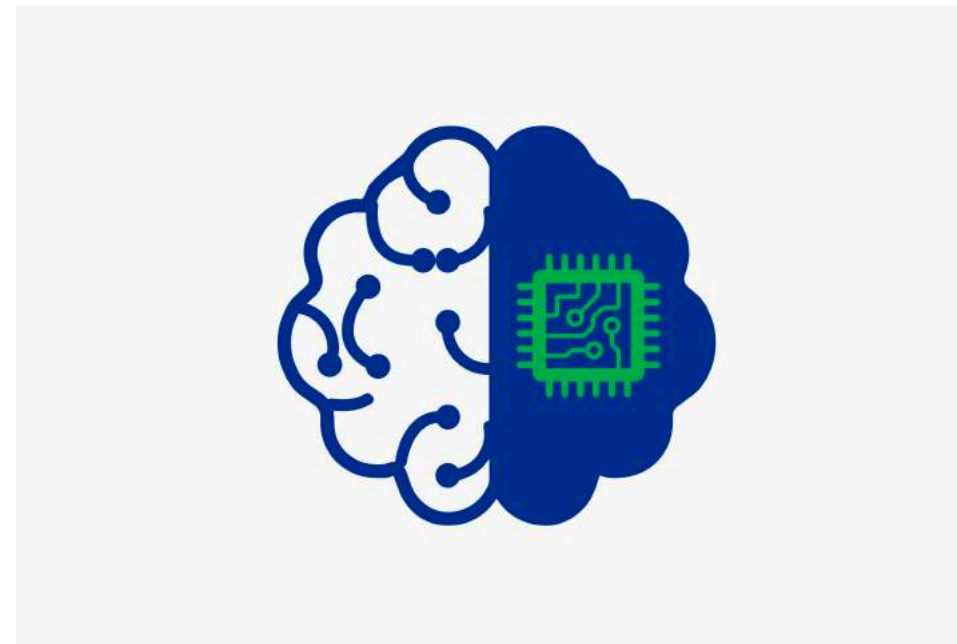
머신러닝 정의와 분류

머신러닝이란?



머신러닝은 프로그램의 한 종류로 데이터를 학습할 수 있는 알고리즘을 주어서 컴퓨터가 어떤 경우에 대해서 예측할 수 있도록 하는 프로그래밍 방법

머신러닝이란?



자율주행 프로그램, 스팸메일을 분류하는 프로그램은
경우의 수가 너무 많기 때문에 Explicit programming이
불가능하여 프로그래머가 컴퓨터에게 데이터와
데이터를 학습할 수 있는 알고리즘을 주어서 컴퓨터가
어떤 경우에 대해서 예측할 수 있도록 하는 프로그래밍 방법

머신러닝의 분류

지도 학습

(Supervised learning)

비지도 학습

(Unsupervised learning)

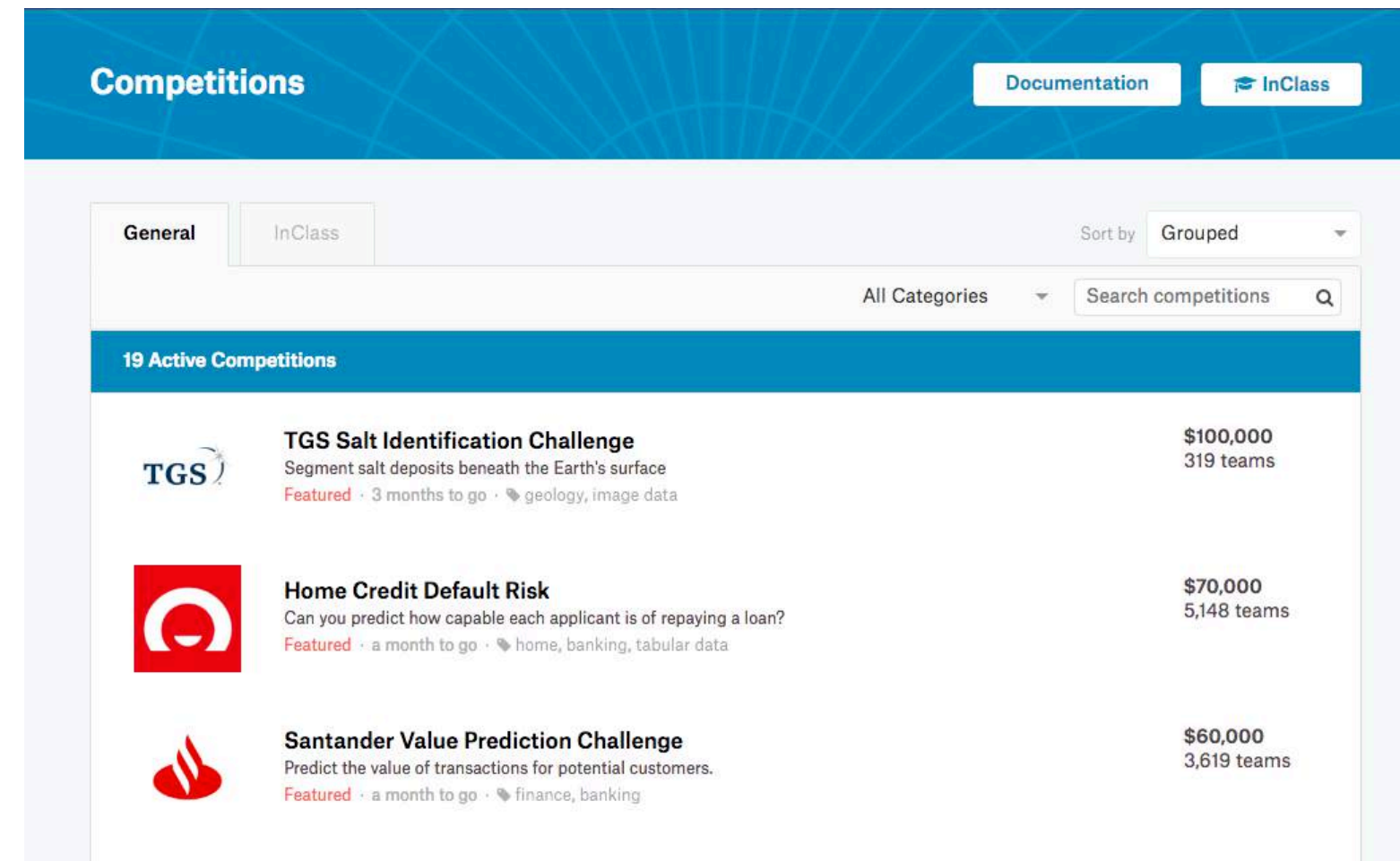
강화 학습

(Reinforcement learning)

데이터를 보고 어떤 학습에 맞는 데이터인지 판단하고
그에 맞는 알고리즘을 적용시키는 능력이 필요

Kaggle 이란?

Kaggle 이란?



예측모델 및 분석 대회 플랫폼

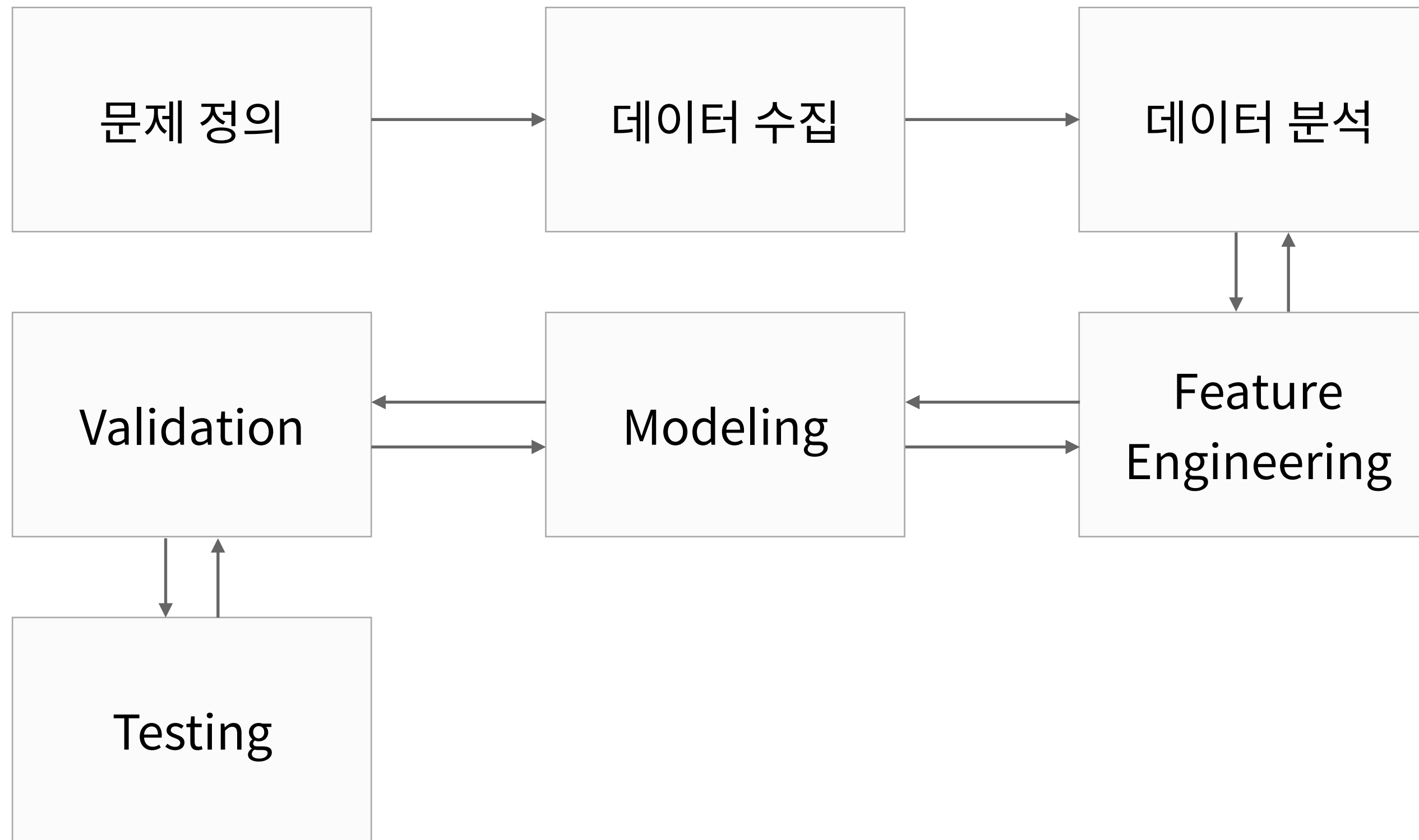
Kaggle 이란?



- 예측모델 및 분석 대회 플랫폼
 - 기업이나 단체에서 데이터와 해결문제를 등록
 - 데이터 과학자들 이를 해결하는 모델을 개발하고 경쟁
-
- 큰 대회는 상금이 수억원
 - 입상자는 구글, 페이스북 등 글로벌 기업에서 앞장서서 채용

데이터 분석 Flow

데이터 분석 Flow



데이터 시각화 분석 및 실습

csv 파일 읽고, 데이터 모양 확인하기

```
import pandas as pd

train_data = pd.read_csv("csv 파일 경로")

// 데이터의 모양 확인하기
train_data.shape

// 앞의 5개 데이터만 확인해보기, n을 input으로 주어 원하는 개수만큼 볼 수 있음.
train_data.head()

// data의 datatype, null 여부 확인하기
train_data.info()

// data의 null 값 체크
train_data.isnull().sum()
```

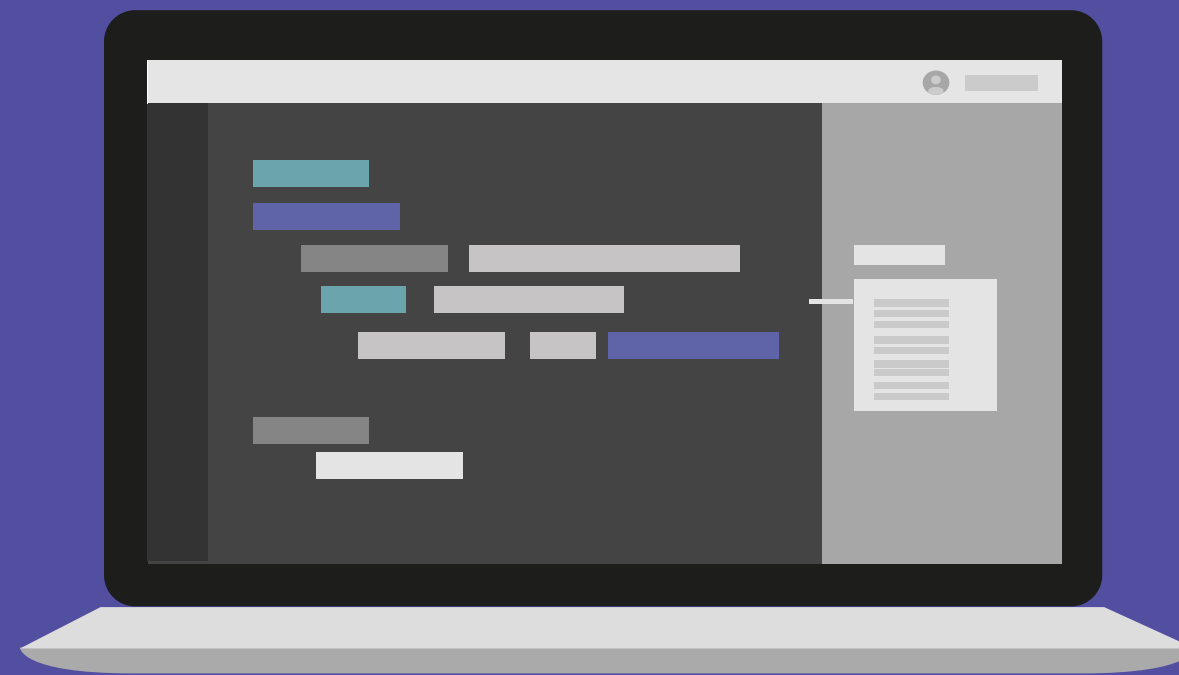
데이터 정보 확인하기

Data 정보 확인하기

Data 정보

- Survived: 0 = No, 1 = Yes
- pclass: Ticket class 1 = 1st, 2 = 2nd, 3 = 3rd
- sibsp: # of siblings / spouses aboard the Titanic
- parch: # of parents / children aboard the Titanic
- ticket: Ticket number
- cabin: Cabin number
- embarked: Port of Embarkation C = Cherbourg, Q = Queenstown, S = Southampton

[실습1] pandas 로 데이터 읽기



`/* elice */`

시각화를 통한 데이터 분석1

```
# 데이터 분석에 필요한 라이브러리들을 import
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from matplotlib import rc
```


시각화를 통한 데이터 분석1

```
# os x(mac os) 한글 font
rc('font', family='AppleGothic')
# 노트북 안에 그래프를 그리기 위해 셋팅
%matplotlib inline
# 그래프를 격자 스타일로
plt.style.use("ggplot")

# 그래프에서 마이너스 폰트 깨지는 문제 해결을 위해
mpl.rcParams["axes.unicode_minus"] = False
```

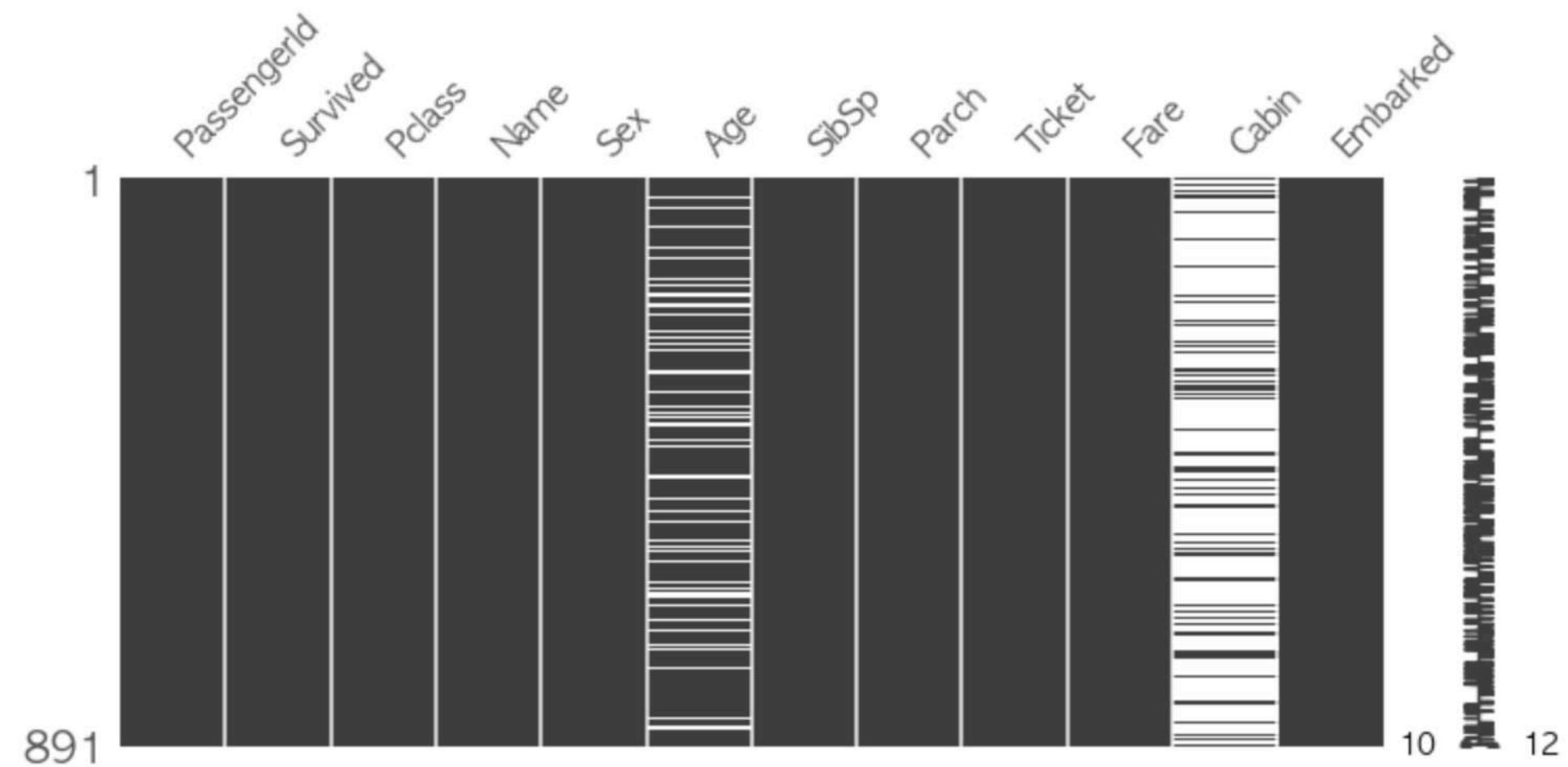
시각화를 통한 데이터 분석2

null인 데이터 시각화 하기

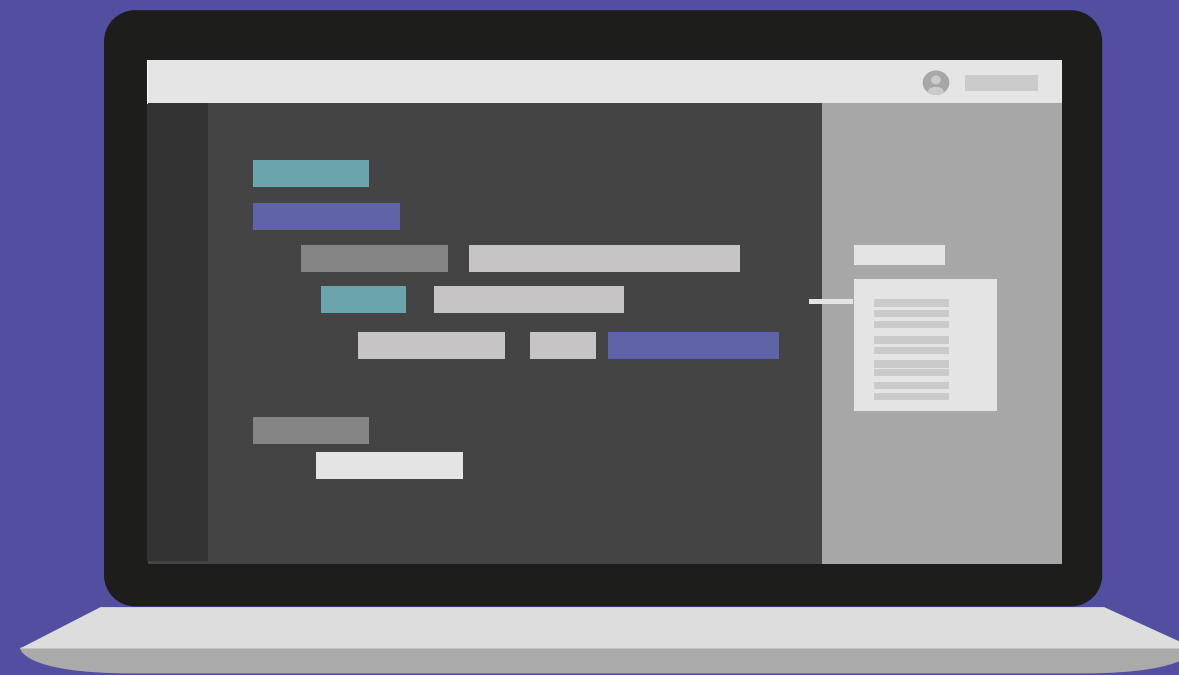
```
In [9]: import missingno as msno
```

```
In [10]: msno.matrix(train, figsize=(12,5))
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x10cd5f080>
```



[실습2] Null 데이터 시각화



```
/* elice */
```

시각화를 통한 데이터 분석3

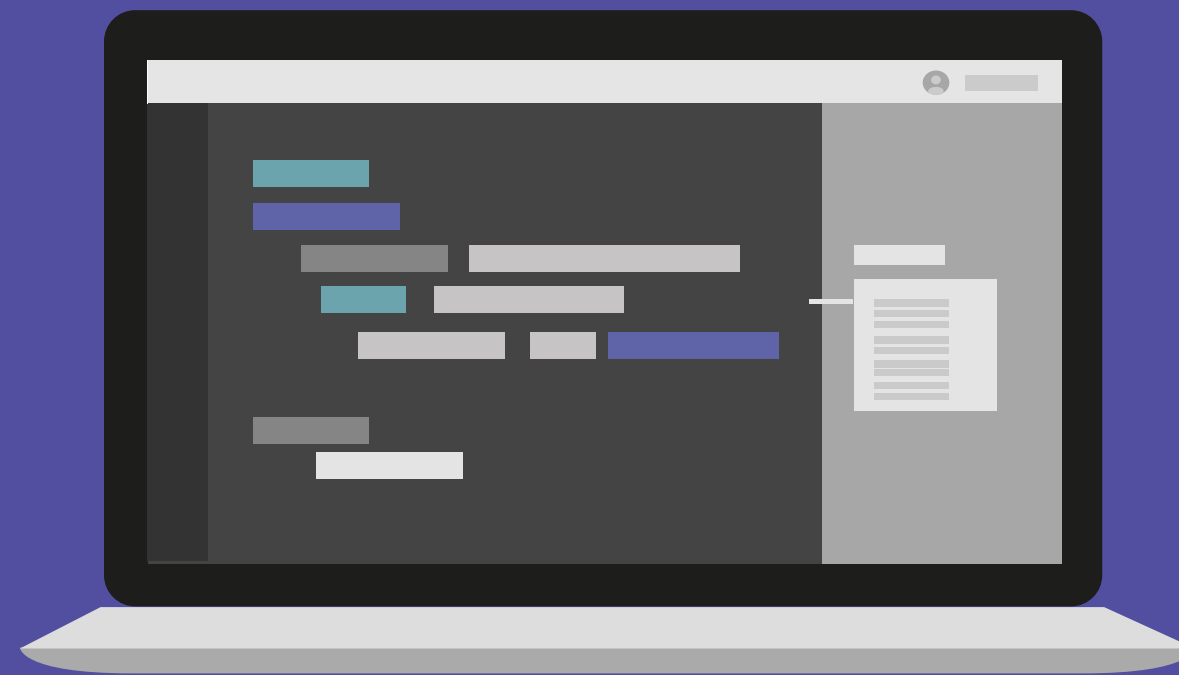
```
# discrete 데이터를 bar chart로 시각화하기
def bar_chart(feature, ax=None):
    survived = train[train['Survived']==1]
    [feature].value_counts()
    dead = train[train['Survived']==0]
    [feature].value_counts()
    df = pd.DataFrame([survived, dead])
    df.index = ['Survived', 'Dead']
    df.plot(kind='bar', stacked=True, ax=ax)

figure, ((ax1,ax2,ax3), (ax4,ax5,ax6)) =
plt.subplots(nrows=2, ncols=3)
figure.set_size_inches(18,12)
```

시각화를 통한 데이터 분석3

```
bar_chart('Sex', ax1)
bar_chart('Pclass', ax2)
bar_chart('SibSp', ax3)
bar_chart('Parch', ax4)
bar_chart('Embarked', ax5)
ax1.set(title="성별 생사정보")
ax2.set(title="티켓 class")
ax3.set(title="형제 수")
ax4.set(title="부모 자식의 수")
ax5.set(title="승선 장소")
```

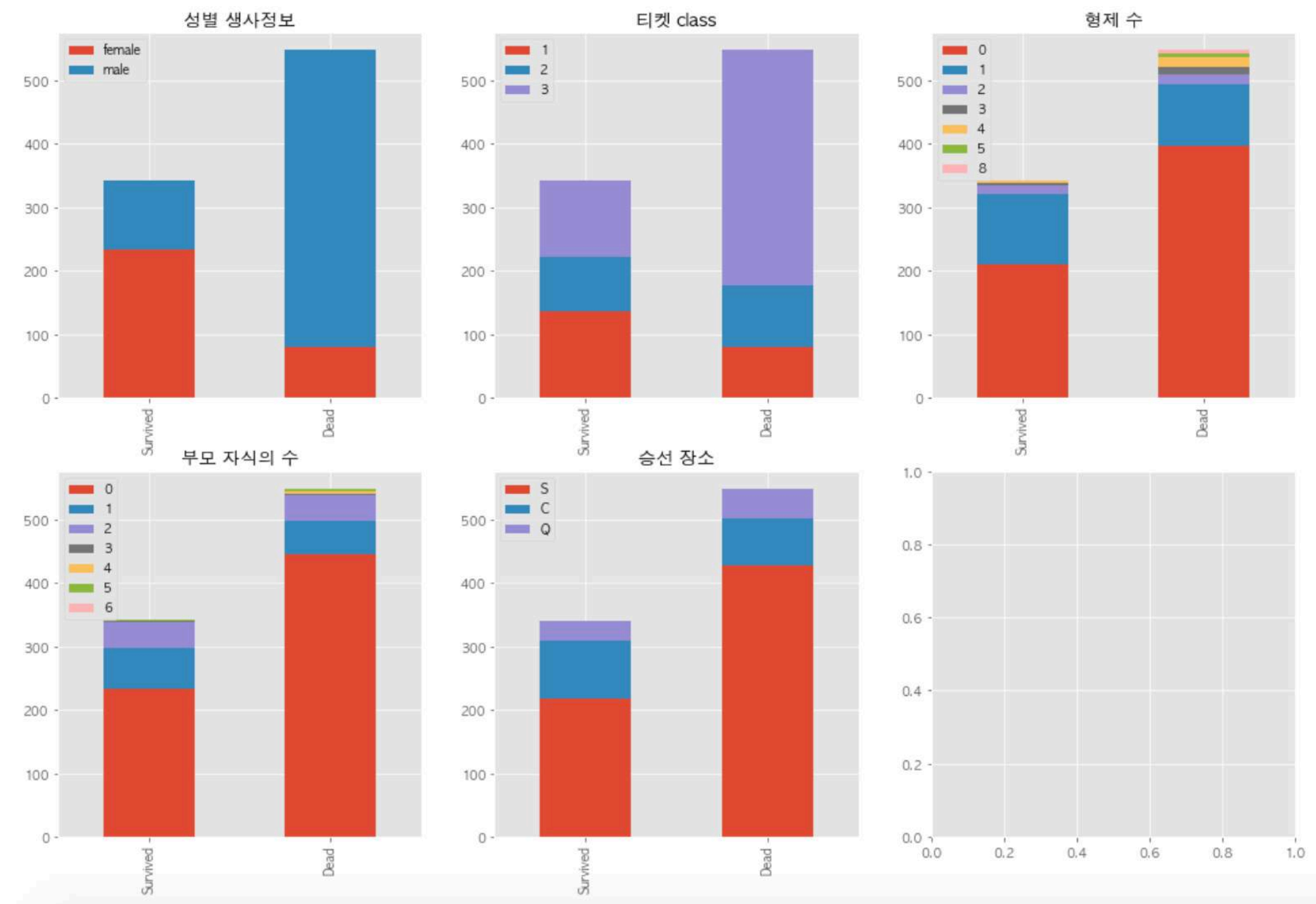
[실습3] Discrete 데이터 시각화



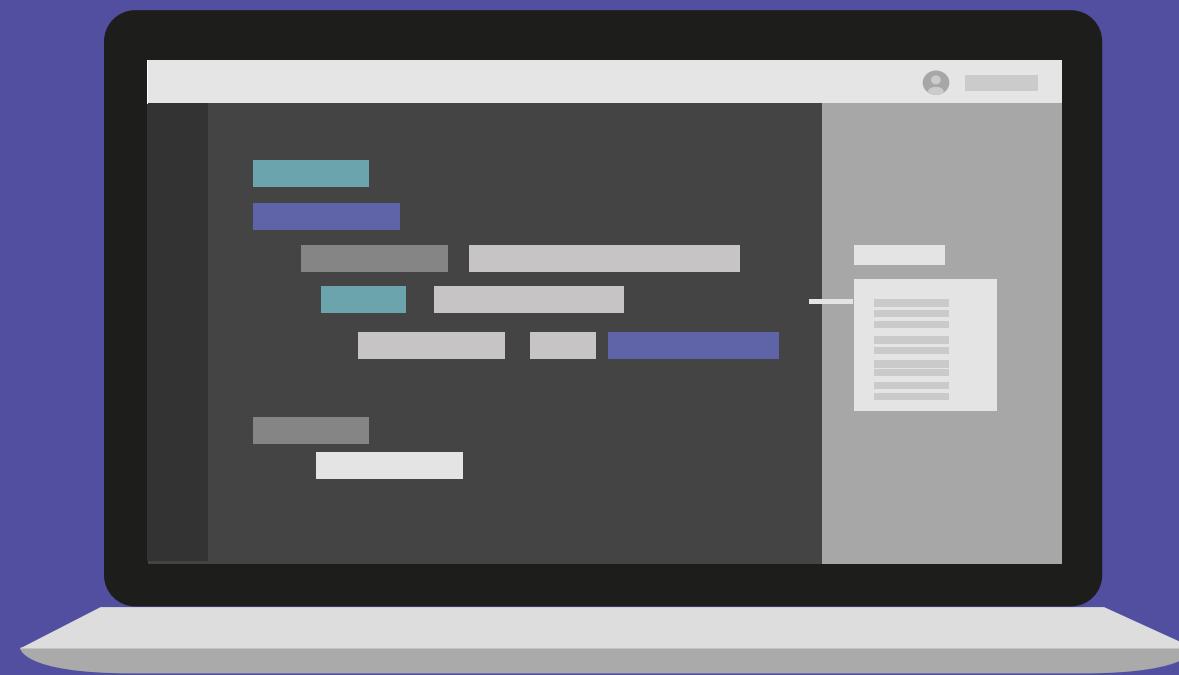
```
/* elice */
```

시각화를 통한 데이터 분석3

전 페이지 코드 실행 output



[실습4] Continuous 데이터 시각화



`/* elice */`

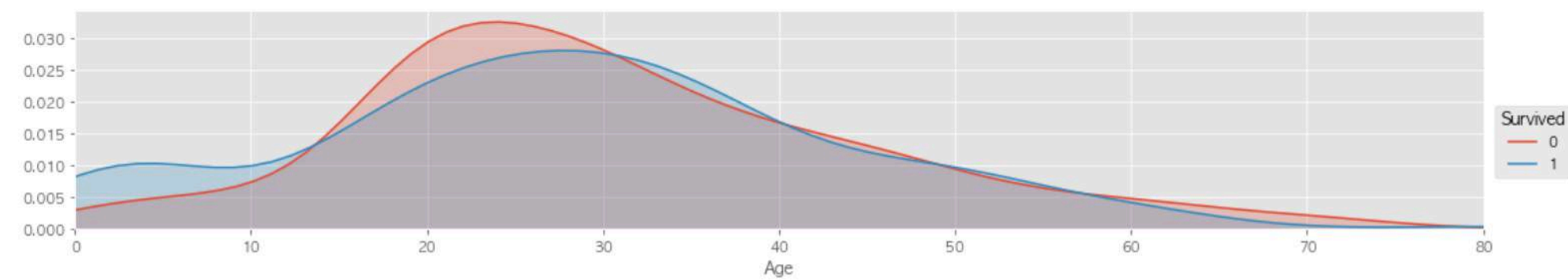
시각화를 통한 데이터 분석4

```
# Continuous 데이터를 facet 으로 시각화하기1
def draw_facetgrid(feature):
    facet = sns.FacetGrid(train, hue="Survived",
aspect=5)
    facet.map(sns.kdeplot, feature, shade=True)
    facet.set(xlim=(0, train[feature].max()))
    # survived 라벨을 표시.
    facet.add_legend()
    plt.show()
```

시각화를 통한 데이터 분석4

코드 실행 output

```
In [44]: draw_facetgrid("Age")
```

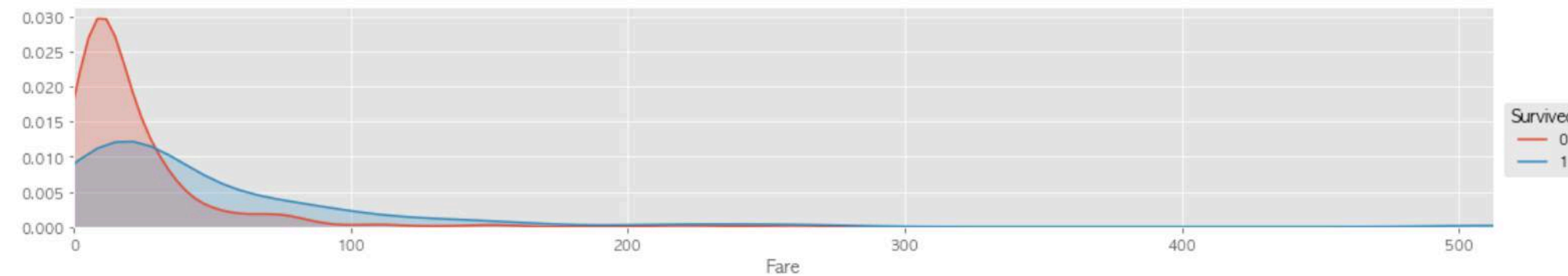


10 이하의 경우 살 확률이 크고, 10대 중반에서 30살까지는
죽을 확률이 더 크며, 나머지 데이터는 큰 의미가 없음

시각화를 통한 데이터 분석4

코드 실행 output

```
In [45]: draw_facetgrid("Fare")
```



전에 pclass 데이터로 확인한 것처럼 비싸게
표를 산 사람들이 살 확률이 더 높음

[미션]

타이타닉 Pclass 별 cabin 시각화



/* elice */

코드 분석 및 실습

코드 분석 및 실습



코드 주소:

[https://github.com/baidoosik/kaggle-solving/blob/master/Titanic/
EP01.KaggleTitanic_data_analyzer.ipynb](https://github.com/baidoosik/kaggle-solving/blob/master/Titanic/EP01.KaggleTitanic_data_analyzer.ipynb)

`/* elice */`

문의 및 연락처

academy.elice.io

contact@elice.io

facebook.com/elice.io

medium.com/elice