사전 포함 코드 :

```
import numpy as n
import pandas as p
import seaborn as s
ttn = s.load_dataset('titanic')
ttn.drop(['PassengerId', 'Cabin'], axis=1, inplace=True)
ttn.Name=ttn.Name.astype("string")
```

3.4 문자열 데이터 처리

3.4.1 문자열 분리하기

단어 묶음을 단어별로 분리하는 함수.
사용법 : ttn.Name.str.split()
쉼표도 없애고 싶다면 괄호 내에 pat="," 추가
문자가 분리되는 만큼 개별 컬럼 추가하려면 괄호 내에 expand=True 추가
각 단어는 배열 기호 []를 써서 그 번째에 해당하는 단어에 접근 가능

3.4.2 문잣값 교체하기

특정 문자를 다른 문자로 대체하는 함수. 만약 .을 공백으로 바꾸려면
사용법 : ttn.title.str.replace('.', '', regex=False)

3.4.3 정규 표현식 가이드

문자열에서 특정 단어를 검색하는 기능.
사용법 예시 몇 개 :
[] :문자 집합, 대괄호 내 어떤 문자던지 매치함
| :or
(?:) :검색만 하고 기억 X
() :괄호 내 문자를 그대로 하나의 그룹 취급
. : 하나의 임의 문자
{n} :앞의 문자 n번 반복
^ : 뒤의 문자로 문자열 시작
주로 count()나 match() 등에 넣어 사용한다.

3.4.4 문자 수 세기

문자의 수를 세는 함수.
모든 문자를 세는 경우 사용법 : ttn['Name'].str.count('')
단어 수를 세는 경우 : ttn['Name'].str.count(' ') + 1

3.5 카테고리 데이터 처리

3.5.1 숫자타입->카테고리타입

bins와 labels로 나눠서 범주를 만든다. 각 labels 값은 bins값들 사이사이를 의미.
예시:

```
bins=[0, 10, 20, 30, 40, 50]
labels=['0p', '1p', '2p', '3p', 4p']
ttn['Age_band_cut']=p.cut(ttn['Age'], bins=bins, labels=labels)
```

cut 대신 qcut() 쓰고 q에 원하는 수 넣으면 그 수만큼 각각의 데이터들을 적절한 수량으로

분배해준다.(q 값은 labels 수와 동일)

3.5.2 카테고리 데이터에 순서 만들기

카테고리 데이터는 크게 명목형과 순서형으로 나뉜다.
p.Categorical([1, 2, 3, 1, 2, 3, 4, 5, 6, 4, 5, 6, n.nan])
이러면 각각 카테고리가 0, 1, 2, 0, 1, 2, 3, 4, 5, 3, 4, 5, -1로 배분된다.
순서를 배분하려면 괄호 안에 ordered=True와 categories=[순서] 넣기.

자기 팀 멤버 문제 풀이

해답만 서술하겠습니다.

출제자 : 강지영

```
import numpy as n
import pandas as p
import seaborn as s
ttn = s.load_dataset('titanic')
ttn.Name=ttn.Name.astype(“string”)
ttn=ttn.Name.str.split(pat=“_”)
print(ttn)
```

출제자 : 박신영

```
k=[“Nothing","Bacon, Onion, Mushroom, Whipping Cream","Cooking Oil","Peanut Butter"] // 레시피 부분을 지정
ramens['Additions'] = ramens['Additions'].replace(k, "SIUUU", regex=False)
```

출제자 : 전유진

```
import numpy as n
import pandas as p
import seaborn as s
ttn = s.load_dataset('titanic')
ttn.Name=ttn.Name.astype(“string”)
k= ['어린이', '청소년', '20,30대', '40,50대', '60대 이상']
ttn.[‘Age_range’]=p.qcut(ttn[‘Age’], q=5, labels=k)
ttn.Age_range.value_counts()
```