평가방법: 필답형 – 캐글 데이터 분석

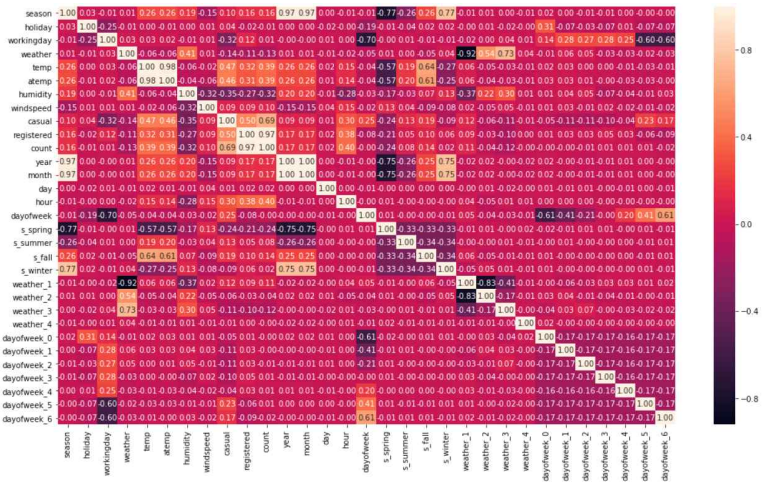
평가일시	2019년 10월 15일		
과정명	SBA 빅데이터 사이언스 실무자 양성과정		
교과목	캐글 데이터 분석	수강생	
단원명	캐글 데이터 분석	문제 출제 및 확인	황석현(임동조)

구분 (능력단위요소)	문 항	점 수
	(문항 1~ 문항 12) 5점 1. 데이터프레임에서 하나의 컬럼을 선택하는 방법이 아닌 것은? **(1) df['Embarked']** **(2) df.Embarked** **(3) df[:,0]** **(4) df.column** (정답) _____ 2. 다음의 datetime object 컬럼을 year, month, day만을 갖는 각각의 컬럼으로 만드시오. |	datetime
---	---
0	2011-01-01 00:00:00
1	2011-01-01 01:00:00
2	2011-01-01 02:00:00
3	2011-01-01 03:00:00
4	2011-01-01 04:00:00
...	... | new_test['year'] = new_test['datetime'].___.___ # (A) new_test['month'] = new_test['datetime'].___.___ # (B) new_test['day'] = new_test['datetime'].___.___ # (C)	**60/ 60 점**

(정답) _(A)_____
(정답) _(B)_____
(정답) _(C)_____

3. 데이터 프레임의 상관계수를 heatmap으로 확인하고자 한다.
다음에 들어갈 함수들을 순서대로 쓰시오.

```
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(18,10))
g = sns._____(new_tr._____, annot=True, fmt=".2f")
```



(정답) _____

4. 'Fare'라는 컬럼을 train에서 빼고자 한다.
이 때 사용할 수 있는 메서드는?

```
train = train.____(['Fare'], axis = 1)
```

(정답) _____

5. train 데이터의 'Age' 컬럼에는 결측치가 존재한다.
이 때, Age컬럼의 결측치를 평균값으로 채우려 할 때 들어갈 것을 순서대로

쓰시오.

train['Age'] = train['Age'].______(train['Age'].______)

(정답) _____

6. 컬럼이 갖는 값들에 대한 빈도(횟수)를 확인하고자 한다. 이때 사용할 수 있는 메서드는?

train['Embarked'].________

```
Out[54]: S    644
         C    168
         Q     77
         Name: Embarked, dtype: int64
```

(정답) _____

7. Embarked 컬럼의 유일한 값들이 무엇이 있는지 확인하고자 한다.
이 때, 사용할 수 있는 메서드는?

train['Embarked']._____

```
Out[53]: array(['S', 'C', 'Q', nan], dtype=object)
```

(정답) _____

8. 성별('Sex') 컬럼을 그룹화 시켜 이를 기준으로 Survived의 평균을 확인하고 싶다. 빈칸에 들어갈 것들을 순서대로 쓰시오.

train._____(____)[_____].mean()

```
Out[39]: Sex
         female    0.742038
         male      0.188908
         Name: Survived, dtype: float64
```

(정답) _____

9. train 데이터에서 'Age'가 40 이하인 사람들의 'Survived'의 평균을 확인하고자 한다. 빈칸에 들어갈 메서드는?

```
train._____[train['Age'] <=40 , :]['Survived'].mean()
```

```
Out[61]: 0.4166666666666667
```

(정답) _____

10. 아래와 같이 인덱스로 원하는 행과 열을 찾고자 할 때 쓰이는 메서드는?

```
train._____[2:10 , 2:5]
```

Out[63]:

	Pclass	Name	Sex
2	3	Heikkinen, Miss. Laina	female
3	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female
4	3	Allen, Mr. William Henry	male
5	3	Moran, Mr. James	male
6	1	McCarthy, Mr. Timothy J	male
7	3	Palsson, Master. Gosta Leonard	male
8	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female
9	2	Nasser, Mrs. Nicholas (Adele Achem)	female

(정답) _____

11. 문자로 된 범주형 변수를 one-hot encoding 해주어 더미 변수로 만들어주는 메서드는?

```
new_train = pd._____(train)
new_test = pd._____(test)
```

Out[25]:

	Pclass	Age	SibSp	Parch	Fare	Sex_female	Sex_male	Embarked_C	Embarked_Q	Embarked_S
0	3	22.0	1	0	7.2500	0	1	0	0	1
1	1	38.0	1	0	71.2833	1	0	1	0	0
2	3	26.0	0	0	7.9250	1	0	0	0	1
3	1	35.0	1	0	53.1000	1	0	0	0	1
4	3	35.0	0	0	8.0500	0	1	0	0	1

(정답) _____

12. python으로 데이터프레임을 합치고자 할 때 사용할 수 있는 함수는?

(1) rbind( )

(2) cbind( )

(3) concat( )

(4) bind_rows( )

(정답) _____

총점	