

```
In [1]: import pandas as pd
from scipy import stats
def corr_analysis(feature_list):
    get_feature = []
    for feature in feature_list:
        corr, pval = stats.pearsonr(df_raw['환자통증정도'], df_raw[feature])
        print("corr : {:.3f}, p-value : {:.3f}".format(corr, pval))
    if pval > 0.05:
        print("상관분석 결과 p-value가 0.05보다 크므로 PM10과 {}간의 상관성이 없다고 할 수 있다.")
    else:
        print("상관분석 결과 p-value가 0.05보다 작으므로 PM10과 {}간의 상관성이 있다고 할 수 있다.")
    get_feature.append(feature)
    return get_feature
```

```
In [2]: df_raw = pd.read_csv("연습_total.csv")
```

```
In [7]: df_raw.columns
```

```
Out[7]: Index(['index', '환자ID', '전방디스크높이(mm)', '후방디스크높이(mm)', '지방축적도', 'Instability',
              'MF + ES', 'Modic change', 'PI', 'PT', 'Seg Angle(raw)', 'Vaccum disc',
              '골밀도', '디스크단면적', '디스크위치', '척추이동척도', '척추전방위증', 'Large Lymphocyte',
              'Location of herniation', 'ODI', '가죽력', '간질성폐질환', '고혈압여부', '과거수술횟수',
              '당뇨여부', '말초동맥질환여부', '빈혈여부', '성별', '스테로이드치료', '신부전여부', '신장',
              '심혈관질환',
              '암발병여부', '연령', '우울증여부', '입원기간', '입원일자', '종양진행여부', '직업', '체중',
              '퇴원일자',
              '헤모글로빈수치', '혈전합병증여부', '환자통증정도', '흡연여부', '통증기간(월)', '수술기법',
              '수술시간',
              '수술실패여부', '수술일자', '재발여부', '혈액형'],
              dtype='object')
```

```
In [3]: corr_analysis(['환자통증정도', 'Large Lymphocyte'])
```

```
corr : 1.000, p-value : 0.000
corr : -0.020, p-value : 0.414
상관분석 결과 p-value가 0.05보다 크므로 PM10과 Large Lymphocyte간의 상관성이 없다고 할 수 있다.
```

```
Out[3]: []
```

```
In [4]: corr_analysis(['환자통증정도', 'Location of herniation'])
```

```
corr : 1.000, p-value : 0.000
corr : 0.159, p-value : 0.000
상관분석 결과 p-value가 0.05보다 작으므로 PM10과 Location of herniation간의 상관성이 있다고 할 수 있다.
```

```
Out[4]: ['Location of herniation']
```

```
In [5]: corr_analysis(['환자통증정도', 'ODI'])
```

```
corr : 1.000, p-value : 0.000
```

UFuncTypeError

Traceback (most recent call last)

```

<ipython-input-5-edbf66d14734> in <module>
----> 1 corr_analysis(['환자통증정도', 'ODI'])

<ipython-input-1-6596fa0f0017> in corr_analysis(feature_list)
      4     get_feature = []
      5     for feature in feature_list:
----> 6         corr, pval = stats.pearsonr(df_raw['환자통증정도'], df_raw[feature])
      7         print("corr : {:.3f}, p-value : {:.3f}".format(corr, pval))
      8         if pval > 0.05:

~\anaconda3\lib\site-packages\scipy\stats\stats.py in pearsonr(x, y)
    3917     # that the data type is at least 64 bit floating point. It might have
    3918     # more precision if the input is, for example, np.longdouble.
-> 3919     dtype = type(1.0 + x[0] + y[0])
    3920
    3921     if n == 2:

```

UFuncTypeError: ufunc 'add' did not contain a loop with signature matching types (dtype('<U32'), dtype('<U32')) -> dtype('<U32')

In [6]: corr_analysis(['환자통증정도', '과거수술횟수'])

corr : 1.000, p-value : 0.000

corr : 0.007, p-value : 0.785

상관분석 결과 p-value가 0.05보다 크므로 PM10과 과거수술횟수간의 상관성이 없다고 할 수 있다.

Out[6]: []

In [8]: corr_analysis(['환자통증정도', '당뇨여부'])

corr : 1.000, p-value : 0.000

corr : 0.040, p-value : 0.098

상관분석 결과 p-value가 0.05보다 크므로 PM10과 당뇨여부간의 상관성이 없다고 할 수 있다.

Out[8]: []

In [9]: corr_analysis(['환자통증정도', '신장'])

corr : 1.000, p-value : 0.000

corr : 0.045, p-value : 0.062

상관분석 결과 p-value가 0.05보다 크므로 PM10과 신장간의 상관성이 없다고 할 수 있다.

Out[9]: []

In [10]: corr_analysis(['환자통증정도', '체중'])

corr : 1.000, p-value : 0.000

corr : 0.040, p-value : 0.097

상관분석 결과 p-value가 0.05보다 크므로 PM10과 체중간의 상관성이 없다고 할 수 있다.

Out[10]: []

In [11]: corr_analysis(['환자통증정도', '연령'])

corr : 1.000, p-value : 0.000

corr : 0.060, p-value : 0.013

상관분석 결과 p-value가 0.05보다 작으므로 PM10과 연령간의 상관성이 있다고 할 수 있다.

Out[11]: ['연령']

1) 의료서비스 불편 - 개선하자 -- 요일별 2) 타 경쟁사 -> 놓치고 있는 고객들 3) 타겟팅 20대

데이터 인사이트 뽑기 날짜, 연령, 흡연여부, 20대 - 병 많아 골밀도 / 이때 많이

EDA / 데이터로 인사이트 뽑기 / 해결방안

단계

- 인사이트 도출 경향성을 가지고 주제로 연결

어떻게 할지

- 거꾸로 가서 주제를 잡자.

데이터 돌려서 확인.

가설세우기
