

03

문자열 데이터 변환하기 (str)

[문자열 변환]

00:08

데이터 핸들링에서 문자열을 잘 다루는 것이 중요합니다!

파이썬 기초 언어를 몰라도 pandas 내에 str 객체를 활용하시면

문자열을 쉽게 변환할 수 있습니다.

```

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipyk...
+ 3< 3> ↺ ↻ ⬆ ⬇ ▶ Run ■ C ▶ Markdown ▾ 🗨 📄

3.3 문자열 변환

3.3.1 인덱싱
df.str[:n] : 문자열 인덱스로 추출

In [ ]: import pandas as pd
import numpy as np

In [ ]: ##데이터 출처 : https://compas.lh.or.kr/subj/competition/data?subjNo=SBJ_2102_001
df_sejong = pd.read_csv('../data/7. 세종시_아파트(전월세)_실거래가.csv')

In [ ]: df_sejong.drop(columns='Unnamed: 0', inplace=True)

In [ ]: df_sejong

In [ ]: df_sejong.시군구.unique()
  
```

[문자열 인덱싱]

02:06

```

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipyk...
+ 3< 3> ↺ ↻ ⬆ ⬇ ▶ Run ■ C ▶ Code ▾ 🗨 📄

3.3.1 인덱싱
시리즈.str[:n] : 문자열 인덱스로 추출

In [27]: import pandas as pd
import numpy as np

In [ ]: ##데이터 출처 : https://compas.lh.or.kr/subj/competition/data?subjNo=SBJ_2102_001
df_sejong = pd.read_csv('../data/7. 세종시_아파트(전월세)_실거래가.csv')

In [ ]: df_sejong.drop(columns='Unnamed: 0', inplace=True)

In [ ]: df_sejong

In [ ]: df_sejong.시군구.unique()

In [ ]: df_sejong['시군구'].str[0:7]

In [ ]: ## 앞자리만 추출
df_sejong['시'] = df_sejong['시군구'].str[0:7]
df_sejong['시']
  
```

[문자열 분할]

05:33

```

38890 세종특별자치시
38891 세종특별자치시
38892 세종특별자치시
38893 세종특별자치시
38894 세종특별자치시
Name: 시, Length: 38895, dtype: object

3.3.2 분할 📄

In [ ]: df_sejong['시군구'].str.split(" ", expand=True)

In [ ]: temp = df_sejong['시군구'].str.split(" ", expand=True)

In [ ]: df_sejong['시군구']

In [ ]: df_sejong['시군구'].unique()

In [ ]: df_sejong_split = df_sejong['시군구'].str.split(" ", expand=True)

In [ ]: df_sejong_split.columns = ['시', '군', '구']

In [ ]: df_sejong_split

```

[시작, 끝, 포함 문자열 index 추출]

15:50

```

'가락마을8단지(고운돌아파트)', '가락마을9단지(신동아 파밀리에)', '가락마을10단지(이지더원)',
'가락마을4단지(이지더원)', '가락마을 19단지', '가락마을 2단지', '가락마을18단지',
'가락마을1단지(힐데스하임)', '가락마을14단지(우남퍼스트빌)'], dtype=object)

3.3.3 시작, 끝, 포함 글자 인식

df.str.startswith("시작글자")
df.str.endswith("끝글자")
df.str.contains("포함글자")

In [ ]: df_sejong['단지명'].unique()[:10]

In [ ]: df_sejong['단지명'].str.startswith('가락마을')

In [ ]: df_sejong[df_sejong['단지명'].str.endswith('골드클래스')].head()

In [ ]: df_sejong[df_sejong['시군구'].str.contains('고운동')].head()

```

다음 강의에서는 데이터에 함수를 적용하는 방법에 대해서 배워볼거예요

다음 강의에서 만나요~