

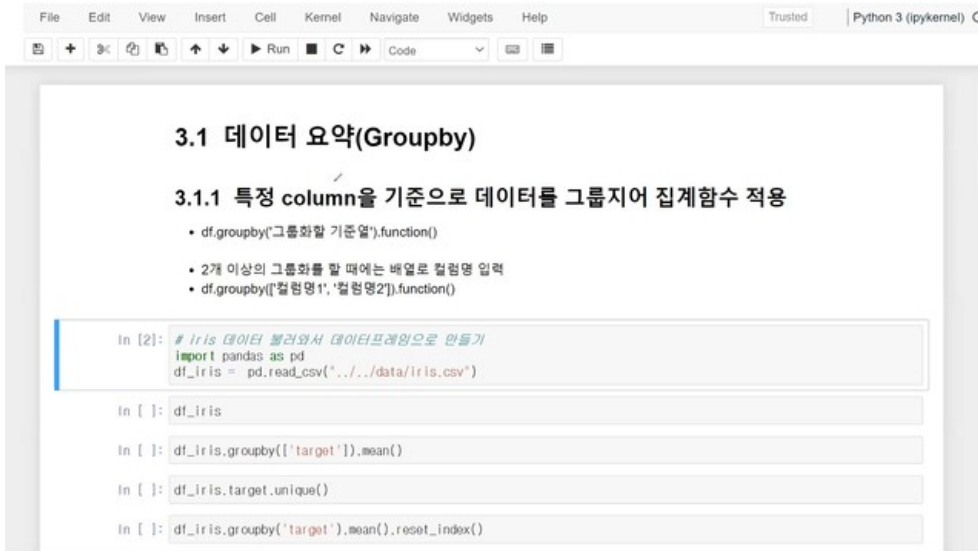
03

데이터 요약 (groupby)

[데이터 요약 groupby]

특정 컬럼의 unique값을 기준으로 데이터를 그룹화하여 집계하는 방법

00:49



[범주형 변수의 도수분포표, 이원분할표]

value_counts, pd.crosstab 을 사용하여 도수 분포표를 만들 수

있습니다. 도수분포표는 통계분석이나 EDA에서 자주 사용되니

익숙해져야 합니다~^^

06:57

The screenshot shows a Jupyter Notebook interface. The top bar includes menus: File, Edit, View, Insert, Cell, Kernel, Navigate, Widgets, Help. On the right, it says 'Trusted' and 'Python 3 (ipykernel)'. Below the menu bar is a toolbar with icons for undo, redo, run, and other notebook functions. The main area displays the output of a cell (Out[10]:) which is a table with columns 'sepal width', 'petal width', and 'target'. The table contains three rows of data for Iris species. Below the table, there is a section header '3.1.2 범주형변수의 도수분포표, 이원분할표' and a sub-header '3.1.2.1 기준이 하나일 때 도수분포표'. This is followed by a list of two code snippets: `np.unique(df['기준열'], return_counts=True)` and `pd.Series(df['기준열']).value_counts()`. Below this, there are four code cells. The first cell loads the Titanic dataset: `# Titanic 데이터 불러오기`
`df_t = pd.read_csv('../data/titanic.csv')`. The second cell shows `df_t['survived'].value_counts()`. The third cell shows `df_t.info()`. The fourth cell shows `# pandas를 이용한 도수분포표`
`two = df_t['class'].value_counts()`.

| | sepal width | petal width | target |
|-----------------|-------------|-------------|--------|
| Iris-setosa | 3.418 | 0.244 | |
| Iris-versicolor | 2.770 | 1.326 | |
| Iris-virginica | 2.974 | 2.026 | |

3.1.2 범주형변수의 도수분포표, 이원분할표

3.1.2.1 기준이 하나일 때 도수분포표

- `np.unique(df['기준열'], return_counts=True)`
- `pd.Series(df['기준열']).value_counts()`

```
In [ ]: # Titanic 데이터 불러오기
df_t = pd.read_csv('../data/titanic.csv')

In [ ]: df_t['survived'].value_counts()

In [ ]: df_t.info()

In [ ]: # pandas를 이용한 도수분포표
two = df_t['class'].value_counts()
```

다음 강의에서는 groupby 와 비슷하지만 더 활용 범위가 넓은 melt 와 pivot_table을 이용하여 데이터를 재구조화 하는 방법을 배워볼 거예요

다음 강의에서 보요~