

دوره دیتا ساینس کاربردی

Index in columns

—● dataroadmap ●—

مدرس: مونا حاتمی

جلسه تکمیلی

```
▶ import pandas as pd
```

```
▶ df=pd.read_csv('dataset_falcon9.csv')
```

```
▶ df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 90 entries, 0 to 89
```

```
Data columns (total 18 columns):
```

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	FlightNumber	90 non-null	int64
1	Date	90 non-null	object
2	BoosterVersion	90 non-null	object
3	PayloadMass	90 non-null	float64
4	Orbit	90 non-null	object
5	LaunchSite	90 non-null	object
6	Outcome	90 non-null	object
7	Flights	90 non-null	int64
8	GridFins	90 non-null	bool
9	Reused	90 non-null	bool
10	Legs	90 non-null	bool
11	LandingPad	64 non-null	object

Columns in order

```
In [4]: df[['PayloadMass', 'Date', 'BoosterVersion', 'FlightNumber']]
```

Out[4]:

	PayloadMass	Date	BoosterVersion	FlightNumber
0	6104.959412	2010-06-04	Falcon 9	1
1	525.000000	2012-05-22	Falcon 9	2
2	677.000000	2013-03-01	Falcon 9	3
3	500.000000	2013-09-29	Falcon 9	4
4	3170.000000	2013-12-03	Falcon 9	5
...
85	15400.000000	2020-09-03	Falcon 9	86
86	15400.000000	2020-10-06	Falcon 9	87
87	15400.000000	2020-10-18	Falcon 9	88
88	15400.000000	2020-10-24	Falcon 9	89

```
In [5]: df.columns
```

```
Out[5]: Index(['FlightNumber', 'Date', 'BoosterVersion', 'PayloadMass', 'Orbit',  
              'LaunchSite', 'Outcome', 'Flights', 'GridFins', 'Reused', 'Legs',  
              'LandingPad', 'Block', 'ReusedCount', 'Serial', 'Longitude', 'Latitude',  
              'Class'],  
             dtype='object')
```

```
In [6]: df.columns[[2,6,0]]
```

```
Out[6]: Index(['BoosterVersion', 'Outcome', 'FlightNumber'], dtype='object')
```

```
In [7]: df[df.columns[[2,6,0]]]
```

```
Out[7]:
```

	BoosterVersion	Outcome	FlightNumber
0	Falcon 9	None None	1
1	Falcon 9	None None	2
2	Falcon 9	None None	3

```
In [9]: df.iloc[0:3, 1:5]
```

Out[9]:

	Date	BoosterVersion	PayloadMass	Orbit
0	2010-06-04	Falcon 9	6104.959412	LEO
1	2012-05-22	Falcon 9	525.000000	LEO
2	2013-03-01	Falcon 9	677.000000	ISS

```
In [10]: df.loc[:, 'Outcome':'Legs']
```

Out[10]:

	Outcome	Flights	GridFins	Reused	Legs
0	None None	1	False	False	False
1	None None	1	False	False	False
2	None None	1	False	False	False
3	False Ocean	1	False	False	False
4	None None	1	False	False	False
...
85	True ASDS	2	True	True	True

```
In [9]: df.iloc[0:3, 1:5]
```

Out[9]:

	Date	BoosterVersion	PayloadMass	Orbit
0	2010-06-04	Falcon 9	6104.959412	LEO
1	2012-05-22	Falcon 9	525.000000	LEO
2	2013-03-01	Falcon 9	677.000000	ISS

```
In [10]: df.loc[:, 'Outcome':'Legs']
```

Out[10]:

	Outcome	Flights	GridFins	Reused	Legs
0	None None	1	False	False	False
1	None None	1	False	False	False
2	None None	1	False	False	False
3	False Ocean	1	False	False	False
4	None None	1	False	False	False
...
85	True ASDS	2	True	True	True

```
In [15]: ► # Drop Multiple Columns by Labels.  
df1 = df.drop(['Longitude', 'BoosterVersion', 'Block'],axis = 1)  
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 90 entries, 0 to 89
```

```
Data columns (total 15 columns):
```

#	Column	Non-Null Count	Dtype
0	FlightNumber	90 non-null	int64
1	Date	90 non-null	object
2	PayloadMass	90 non-null	float64
3	Orbit	90 non-null	object
4	LaunchSite	90 non-null	object
5	Outcome	90 non-null	object
6	Flights	90 non-null	int64
7	GridFins	90 non-null	bool
8	Reused	90 non-null	bool
9	Legs	90 non-null	bool
10	LandingPad	64 non-null	object
11	ReusedCount	90 non-null	int64
12	Serial	90 non-null	object
13	Latitude	90 non-null	float64
14	Class	90 non-null	int64

```
dtypes: bool(3), float64(2), int64(4), object(6)
```

```
memory usage: 8.8+ KB
```

```
[11]: ► # Drop columns based on column index.
df2 = df.drop(df.columns[[3, 1, 2]],axis = 1)
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 90 entries, 0 to 89
```

```
Data columns (total 15 columns):
```

#	Column	Non-Null Count	Dtype
0	FlightNumber	90 non-null	int64
1	Orbit	90 non-null	object
2	LaunchSite	90 non-null	object
3	Outcome	90 non-null	object
4	Flights	90 non-null	int64
5	GridFins	90 non-null	bool
6	Reused	90 non-null	bool
7	Legs	90 non-null	bool
8	LandingPad	64 non-null	object
9	Block	90 non-null	float64

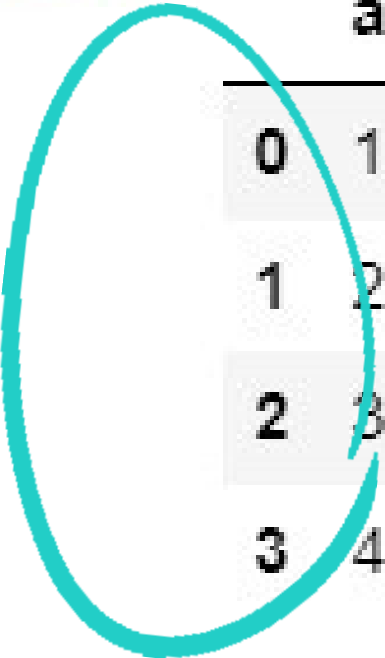
About ignore_index=True ¶

```
n [18]: ▶ # Creating the first Dataframe using dictionary
df1 = df = pd.DataFrame({"a": [1, 2, 3, 4],
                          "b": [5, 6, 7, 8]})
```

```
n [19]: ▶ # Creating the Second Dataframe using dictionary
df2 = pd.DataFrame({"a": [1, 2, 3],
                    "b": [5, 6, 7]})
```

In [20]: ▶ df1

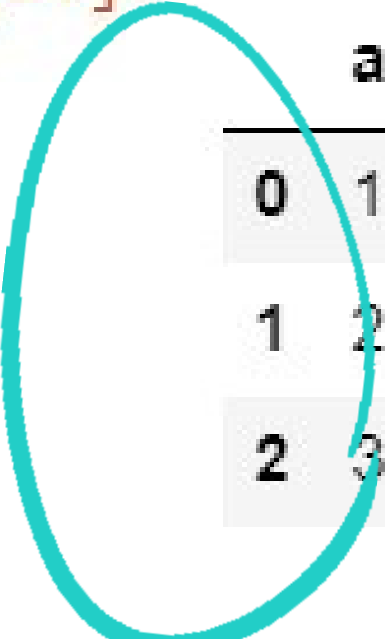
Out[20]:



	a	b
0	1	5
1	2	6
2	3	7
3	4	8

In [21]: ▶ df2

Out[21]:



	a	b
0	1	5
1	2	6
2	3	7

[22]: ▶ df1.append(df2)

Out[22]:

	a	b
0	1	5
1	2	6
2	3	7
3	4	8
0	1	5
1	2	6
2	3	7

[23]: ▶ df1.append(df2, ignore_index=True)

Out[23]:

	a	b
0	1	5
1	2	6
2	3	7
3	4	8
4	1	5
5	2	6
6	3	7

```
df=pd.read_csv('dataset_falcon9.csv')

new_row = {'FlightNumber':1, 'Date':2, 'BoosterVersion':3, 'PayloadMass':4, 'Orbit':5,
           'LaunchSite':6, 'Outcome':7, 'Flights':8, 'GridFins':9, 'Reused':10, 'Legs':11,
           'LandingPad':12, 'Block':13, 'ReusedCount':14, 'Serial':15, 'Longitude':16, 'Latitude':17,
           'Class':18, 'new':19}
#append row to the dataframe
df = df.append(new_row, ignore_index=True)

df.tail(2)
```

26]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	I
89	90	2020-11-05	Falcon 9	3681.0	MEO	CCAFS SLC 40	True ASDS	1	1	0	1	5e9e3032383ecb6bb234e7ca	5.0	
90	1	2	3	4.0	5	6	7	8	9	10	11	12	13.0	

```
] : ▶ new_row = {'FlightNumber':1, 'Date':2, 'BoosterVersion':3, 'PayloadMass':4, 'Orbit':5,  
               'LaunchSite':6, 'Outcome':7, 'Flights':8, 'GridFins':9, 'Reused':10, 'Legs':11,  
               'LandingPad':12, 'Block':13, 'ReusedCount':14, 'Serial':15, 'Longitude':16, 'Latitude':17,  
               'Class':18, 'new':19}  
#append row to the dataframe  
df = df.append(new_row)
```

```
-----  
TypeError                                Traceback (most recent call last)
```

```
~\AppData\Local\Temp\ipykernel_46240\3600306850.py in <module>
```

```
      4         'Class':18, 'new':19}  
      5 #append row to the dataframe  
----> 6 df = df.append(new_row)
```

```
~\AppData\Roaming\Python\Python38\site-packages\pandas\core\frame.py in append(self, other, ignore_index, verify_integrity,  
sort)
```

```
    8930         if isinstance(other, dict):  
    8931             if not ignore_index:  
-> 8932                 raise TypeError("Can only append a dict if ignore_index=True")  
    8933             other = Series(other)  
    8934             if other.name is None and not ignore_index:
```

```
TypeError: Can only append a dict if ignore_index=True
```
