

Chap. - 4 Machine Learning with Python

In [1]:

```
1 import pandas as pd
2 df = pd.DataFrame(
3     {'age':[18,20,23,19,18,22],
4     'city':['city-A','city-B','city-B','city-A','city-C','city-B']})
5 n_df = pd.get_dummies(df)
6 print(df)
7 print(n_df)
```

	age	city		
0	18	city-A		
1	20	city-B		
2	23	city-B		
3	19	city-A		
4	18	city-C		
5	22	city-B		
	age	city_city-A	city_city-B	city_city-C
0	18	1	0	0
1	20	0	1	0
2	23	0	1	0
3	19	1	0	0
4	18	0	0	1
5	22	0	1	0

. drops first column

In [2]:

```
1 import pandas as pd
2 df = pd.DataFrame(
3     {'age':[18,20,23,19,18,22],
4     'city':['city-A','city-B','city-B','city-A','city-C','city-B']})
5 n_df = pd.get_dummies(df, drop_first=True)
6 print(df)
7 print(n_df)
```

	age	city	
0	18	city-A	
1	20	city-B	
2	23	city-B	
3	19	city-A	
4	18	city-C	
5	22	city-B	
	age	city_city-B	city_city-C
0	18	0	0
1	20	1	0
2	23	1	0
3	19	0	0
4	18	0	1
5	22	1	0

In [4]:

```
1 import pandas as pd
2 df = pd.DataFrame(
3     {'age':[18,20,23,19,18,22]})
4
5 df['n_df']=df['age'] + 5
6 print(df)
```

	age	n_df
0	18	23
1	20	25
2	23	28
3	19	24
4	18	23
5	22	27

• Birth year is given and add age

```
In [9]: 1 df = pd.DataFrame({
2         'by': [1990, 1993, 2000, 2006, 2010, 2022]
3     })
4 df['age'] = 2025 - df['by']
5 print(df)
```

	by	age
0	1990	35
1	1993	32
2	2000	25
3	2006	19
4	2010	15
5	2022	3

```
In [13]: 1 df = pd.DataFrame({
2         'length': [18, 20, 10, 12, 18, 11],
3         'breadth': [20, 20, 10, 11, 19, 10]
4     })
5 df['area'] = df['length'] * df['breadth']
6 print(df)
```

	length	breadth	area
0	18	20	360
1	20	20	400
2	10	10	100
3	12	11	132
4	18	19	342
5	11	10	110

P.b.-170

```
In [22]: 1 import pandas as pd
2
3 df = pd.DataFrame({
4     'area': [4720, 2430, 4368, 3969, 6142, 7912],
5     'price': [2360000, 1215000, 1984500, 3071000, 3956000, 1500000]
6 })
7
8 # Initialize an empty list to store categories
9 categories = []
10
11 for price in df['price']:
12     if price > 3000000:
13         categories.append('High')
14     elif price < 1500000:
15         categories.append('Low')
16     else:
17         categories.append('Medium')
18
19 df['Range'] = categories
20 print(df)
```

	area	price	Range
0	4720	2360000	Medium
1	2430	1215000	Low
2	4368	1984500	Medium
3	3969	3071000	High
4	6142	3956000	High
5	7912	1500000	Medium

```
In [17]: 1 import pandas as pd
2 import numpy as np
3
4 df = pd.DataFrame({
5     'area': [4720, 2430, 4368, 3969, 6142, 7912],
6     'price': [2360000, 1215000, 1984500, 3071000, 3956000, 1500000]
7 })
8
9 df['cat'] = np.where(df['price'] > 3000000, 'High',
10                    np.where(df['price'] < 1500000, 'Low', 'Medium'))
11
12 print(df)
```

	area	price	cat
0	4720	2360000	Medium
1	2430	1215000	Low
2	4368	1984500	Medium
3	3969	3071000	High
4	6142	3956000	High
5	7912	1500000	Medium

In [21]:

```
1 import pandas as pd
2
3 df = pd.DataFrame({
4     'area': [4720, 2430, 4368, 3969, 6142, 7912],
5     'price': [2360000, 1215000, 1984500, 3071000, 3956000, 1500000]
6 })
7
8 def categorize(price):
9     if price > 3000000:
10         return 'High'
11     elif price < 1500000:
12         return 'Low'
13     else:
14         return 'Medium'
15
16 df['cat'] = df['price'].apply(categorize)
17
18 print(df)
```

	area	price	cat
0	4720	2360000	Medium
1	2430	1215000	Low
2	4368	1984500	Medium
3	3969	3071000	High
4	6142	3956000	High
5	7912	1500000	Medium

P.b.-171

In [3]:

```
1 import pandas as pd
2 import numpy as np
3
4 df = pd.DataFrame({
5     'age': [25, 32, 45, 61, 18, 55, 70, 29, 30, 60]
6 })
7
8 df['age_group'] = np.where(df['age'] < 30, 'Young',
9                             np.where(df['age'] <= 60, 'Middle-aged',
10                                     'Elderly'))
11
12 print(df)
```

	age	age_group
0	25	Young
1	32	Middle-aged
2	45	Middle-aged
3	61	Elderly
4	18	Young
5	55	Middle-aged
6	70	Elderly
7	29	Young
8	30	Middle-aged
9	60	Middle-aged

P.b.-173

In [4]:

```
1 df = pd.DataFrame({
2     'income': [25000, 42000, 85000, 19000, 70000, 30000, 71000]
3 })
4
5 df['income_category'] = np.where(df['income'] < 30000, 'Low',
6                                 np.where(df['income'] <= 70000, 'Medium',
7                                     'High'))
8
9 print(df)
```

	income	income_category
0	25000	Low
1	42000	Medium
2	85000	High
3	19000	Low
4	70000	Medium
5	30000	Medium
6	71000	High

In [9]:

```
1 import pandas as pd
2 import numpy as np
3
4 df = pd.DataFrame({
5     'age': [10,12,15,19,22,30,35,55,56,92,70,25, 32, 45, 61, 18, 55, 70, 29, 30, 60,97]
6 })
7
8 df['age_group'] = np.where(df['age'] < 15, 'Kid',
9                             np.where(df['age'] < 60, 'Young',
10                                     np.where(df['age'] < 60, 'Middle-aged',
11                                             'Elderly'))))
12 print(df)
```

	age	age_group
0	10	Kid
1	12	Kid
2	15	Young
3	19	Young
4	22	Young
5	30	Young
6	35	Young
7	55	Young
8	56	Young
9	92	Elderly
10	70	Elderly
11	25	Young
12	32	Young
13	45	Young
14	61	Elderly
15	18	Young
16	55	Young
17	70	Elderly
18	29	Young
19	30	Young
20	60	Elderly
21	97	Elderly