Untitled11

April 21, 2024

1 Python Data Cleaning and File Handling

[1]:
```
import pandas as pd

# Define the data
data = {
    'A': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
    'B': ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j'],
    'C': [True, False, True, False, True, False, True, False, True, False],
    'D': [10.5, 20.5, 30.5, 40.5, 50.5, 60.5, 70.5, 80.5, 90.5, 100.5],
    'E': ['apple', 'banana', 'cherry', 'date', 'elderberry', 'fig', 'grape',␣
 ↪'honeydew', 'kiwi', 'lemon'],
    'F': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000],
    'G': [1.1, 2.2, 3.3, 4.4, 5.5, 6.6, 7.7, 8.8, 9.9, 10.1],
    'H': ['red', 'green', 'blue', 'yellow', 'orange', 'purple', 'pink', 'cyan',␣
 ↪'magenta', 'black'],
    'I': [True, True, False, False, True, True, False, False, True, True],
    'J': ['dog', 'cat', 'fish', 'bird', 'rabbit', 'hamster', 'turtle', 'snake',␣
 ↪'frog', 'lizard']
}

# Create the DataFrame
df = pd.DataFrame(data)

# Display the DataFrame
print(df)
```

```
   A  B      C     D           E    F    G        H      I        J
0  1  a   True  10.5       apple  100  1.1      red   True      dog
1  2  b  False  20.5      banana  200  2.2    green   True      cat
2  3  c   True  30.5      cherry  300  3.3     blue  False     fish
3  4  d  False  40.5        date  400  4.4   yellow  False     bird
4  5  e   True  50.5  elderberry  500  5.5   orange   True   rabbit
5  6  f  False  60.5         fig  600  6.6   purple   True  hamster
6  7  g   True  70.5       grape  700  7.7     pink  False   turtle
7  8  h  False  80.5    honeydew  800  8.8     cyan  False    snake
8  9  i   True  90.5        kiwi  900  9.9  magenta   True     frog
```

```
9  10  j  False  100.5       lemon  1000  10.1    black   True   lizard
```

1.0.1 How many rows and columns are there in the DataFrame?

[2]:
```
print("number of rows is ",df.shape[0])
print("number of columns is ",df.shape[1])
```

```
number of rows is  10
number of columns is  10
```

1.0.2 What are the column names of the DataFrame?

[3]:
```
df.columns
```

[3]:
```
Index(['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J'], dtype='object')
```

1.0.3 What are the data types of each column?

[13]:
```
df.dtypes
```

[13]:
```
A      int64
B     object
C       bool
D    float64
E     object
F      int64
G    float64
H     object
I       bool
J     object
K      int64
dtype: object
```

1.0.4 How can you access the first 5 rows of the DataFrame?

[4]:
```
df.head(5)
```

[4]:

	A	B	C	D	E	F	G	H	I	J
0	1	a	True	10.5	apple	100	1.1	red	True	dog
1	2	b	False	20.5	banana	200	2.2	green	True	cat
2	3	c	True	30.5	cherry	300	3.3	blue	False	fish
3	4	d	False	40.5	date	400	4.4	yellow	False	bird
4	5	e	True	50.5	elderberry	500	5.5	orange	True	rabbit

1.0.5 How can you access the 'C' column of the DataFrame?

[5]:
```
df['C']
```

[5]:
```
0     True
1    False
2     True
3    False
4     True
5    False
6     True
7    False
8     True
9    False
Name: C, dtype: bool
```

1.0.6 What is the average value of column 'F'?

[6]:
```
df["F"].mean()
```

[6]:
```
550.0
```

1.0.7 How can you filter the DataFrame to show only rows where column 'I' is True?

[7]:
```
df[df["I"]==True]
```

[7]:

	A	B	C	D	E	F	G	H	I	J
0	1	a	True	10.5	apple	100	1.1	red	True	dog
1	2	b	False	20.5	banana	200	2.2	green	True	cat
4	5	e	True	50.5	elderberry	500	5.5	orange	True	rabbit
5	6	f	False	60.5	fig	600	6.6	purple	True	hamster
8	9	i	True	90.5	kiwi	900	9.9	magenta	True	frog
9	10	j	False	100.5	lemon	1000	10.1	black	True	lizard

1.0.8 How can you sort the DataFrame based on column 'D' in descending order?

[8]:
```
df.sort_values("D",ascending=False)
```

[8]:

	A	B	C	D	E	F	G	H	I	J
9	10	j	False	100.5	lemon	1000	10.1	black	True	lizard
8	9	i	True	90.5	kiwi	900	9.9	magenta	True	frog
7	8	h	False	80.5	honeydew	800	8.8	cyan	False	snake
6	7	g	True	70.5	grape	700	7.7	pink	False	turtle
5	6	f	False	60.5	fig	600	6.6	purple	True	hamster
4	5	e	True	50.5	elderberry	500	5.5	orange	True	rabbit
3	4	d	False	40.5	date	400	4.4	yellow	False	bird

2	3	c	True	30.5	cherry	300	3.3	blue	False	fish
1	2	b	False	20.5	banana	200	2.2	green	True	cat
0	1	a	True	10.5	apple	100	1.1	red	True	dog

1.0.9 How can you add a new column 'K' to the DataFrame with values [1, 2, 3, ..., 15]?

[9]:
```
df["K"]=[1,2,3,4,5,6,7,8,9,10]
df
```

[9]:

	A	B	C	D	E	F	G	H	I	J	K
0	1	a	True	10.5	apple	100	1.1	red	True	dog	1
1	2	b	False	20.5	banana	200	2.2	green	True	cat	2
2	3	c	True	30.5	cherry	300	3.3	blue	False	fish	3
3	4	d	False	40.5	date	400	4.4	yellow	False	bird	4
4	5	e	True	50.5	elderberry	500	5.5	orange	True	rabbit	5
5	6	f	False	60.5	fig	600	6.6	purple	True	hamster	6
6	7	g	True	70.5	grape	700	7.7	pink	False	turtle	7
7	8	h	False	80.5	honeydew	800	8.8	cyan	False	snake	8
8	9	i	True	90.5	kiwi	900	9.9	magenta	True	frog	9
9	10	j	False	100.5	lemon	1000	10.1	black	True	lizard	10

1.0.10 How can you drop the 'E' column from the DataFrame?

[10]:
```
df.drop(columns="E")
```

[10]:

	A	B	C	D	F	G	H	I	J	K
0	1	a	True	10.5	100	1.1	red	True	dog	1
1	2	b	False	20.5	200	2.2	green	True	cat	2
2	3	c	True	30.5	300	3.3	blue	False	fish	3
3	4	d	False	40.5	400	4.4	yellow	False	bird	4
4	5	e	True	50.5	500	5.5	orange	True	rabbit	5
5	6	f	False	60.5	600	6.6	purple	True	hamster	6
6	7	g	True	70.5	700	7.7	pink	False	turtle	7
7	8	h	False	80.5	800	8.8	cyan	False	snake	8
8	9	i	True	90.5	900	9.9	magenta	True	frog	9
9	10	j	False	100.5	1000	10.1	black	True	lizard	10

[ ]:
```

```