

897_scratchpad_20191015

October 15, 2019

```
[212]: import pandas as pd
```

```
[213]: alfa = pd.read_csv('scq_20191015.csv'); alfa.head(2)
```

```
[213]:
```

| | size | color | quantity |
|---|------|-------|----------|
| 0 | S | green | 40.0 |
| 1 | L | red | 20.0 |

```
[214]: # _ = alfa.head(2).to_csv("test.csv")
```

```
[215]: # -----  
# this creates a copy of "alfa"  
# later down in a lower cell when i modify "alfa"  
# "beta" remains the same, it does not change  
# -----  
beta = alfa[:]; beta
```

```
[215]:
```

| | size | color | quantity |
|----|------|--------|----------|
| 0 | S | green | 40.0 |
| 1 | L | red | 20.0 |
| 2 | M | yellow | 9.0 |
| 3 | M | black | 17.0 |
| 4 | L | white | 53.0 |
| 5 | S | grey | 84.0 |
| 6 | S | NaN | 48.0 |
| 7 | NaN | gray | 48.0 |
| 8 | S | yellow | 9.0 |
| 9 | NaN | black | 88.0 |
| 10 | S | red | NaN |

```
[216]: # -----  
# returns a 1 (single) record DATA FRAME  
# in this case, just 1 record meets the filtering criteria  
# the "color" column IS NA (null)  
# -----  
color_is_na = alfa[alfa.color.isna()]; color_is_na  
# type(color_is_na)
```

```
[216]:  size color  quantity
        6    S   NaN      48.0
```

```
[217]: # -----
# after this, "alfa" will not include the record at index 6
# -----
alfa = alfa[(alfa.color.notna())]; alfa
```

```
[217]:  size  color  quantity
0     S   green    40.0
1     L    red    20.0
2     M  yellow     9.0
3     M   black   17.0
4     L   white   53.0
5     S    grey   84.0
7    NaN   gray   48.0
8     S  yellow     9.0
9    NaN   black  88.0
10    S    red    NaN
```

```
[218]: # -----
# returns the record at index 6 (position 7) AS A SERIES
# "beta" still has the record at index 6
# which was removed from "alfa" above
# -----
beta.iloc[6]
```

```
[218]: size      S
color      NaN
quantity   48
Name: 6, dtype: object
```

```
[219]: beta_before_3 = beta.iloc[:3]; beta_before_3
```

```
[219]:  size  color  quantity
0     S   green    40.0
1     L    red    20.0
2     M  yellow     9.0
```

```
[220]: beta_starting_at_3 = beta.iloc[3:]; beta_starting_at_3
```

```
[220]:  size  color  quantity
3     M   black    17.0
4     L   white    53.0
5     S    grey    84.0
6     S    NaN     48.0
7    NaN   gray    48.0
```

| | | | |
|----|-----|--------|------|
| 8 | S | yellow | 9.0 |
| 9 | NaN | black | 88.0 |
| 10 | S | red | NaN |

```
[221]: # -----
# this returns A SERIES
# "size" is an attribute/property of the data frame
# CANNOT USE "beta.size"
# -----
beta_size = beta['size']; beta_size
# type(beta_size)
```

```
[221]: 0      S
1      L
2      M
3      M
4      L
5      S
6      S
7      NaN
8      S
9      NaN
10     S
Name: size, dtype: object
```

```
[225]: distinct_size = pd.Series(beta['size'].unique()); distinct_size
```

```
[225]: 0      S
1      L
2      M
3      NaN
dtype: object
```