

**pruning- to reduce, especially by eliminating(remove) superfluous(not needed) matter.**

In [1]:

```
1 import pandas as pd
2 import numpy as np
```

In [2]:

```
1 pd.__version__
```

Out[2]:

'1.0.1'

## Indexing with boolean masks

**step 1: generate sequence of booleans  
step 2: Use boolean sequence in [] or .loc[]**

In [3]:

```
1 players = pd.read_csv('soccer.csv')
```

In [4]:

```
1 players.head()
```

Out[4]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

**What are the players that have a market value exceeding 40M?**

In [5]:

```
1 players.market_value>40 #..... step1
```

Out[5]:

```
0      True
1      True
2     False
3     False
4     False
...
460    False
461    False
462    False
463    False
464    False
Name: market_value, Length: 465, dtype: bool
```

In [6]:

```
1 players[players.market_value>40] #.....step2
```

Out[6]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
96	Eden Hazard	Chelsea	26	LW	1	75.0	4220	10.5	2.30%
97	Diego Costa	Chelsea	28	CF	1	50.0	4454	10.0	3.00%
108	N%27Golo Kante	Chelsea	26	DM	2	50.0	4042	5.0	13.80%
218	Philippe Coutinho	Liverpool	25	AM	1	45.0	2958	9.0	30.80%
244	Kevin De Bruyne	Manchester City	26	AM	1	65.0	3252	10.0	17.50%

In [7]:

```
1 players[players.market_value>40].shape[0]
```

Out[7]:

13

In [8]:

```
1 #OR
2 val = players.market_value>40
```

In [9]:

```
1 players[val].shape[0]
```

Out[9]:

13

13 players have market value greater than 40M

## More Approaches To Boolean Masking

In [10]:

```
1 players.head()
```

Out[10]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

◀ ▶

In [11]:

```
1 players.age<25
```

Out[11]:

```
0      False
1      False
2      False
3      False
4      False
...
460     True
461     True
462     True
463     True
464    False
Name: age, Length: 465, dtype: bool
```

In [12]:

1 players[players.age&lt;25]

Out[12]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value
5		Hector Bellerin	Arsenal	22	RB	3	30.0	1675	6.0
9		Alex Iwobi	Arsenal	21	LW	1	10.0	1812	5.5
10		Granit Xhaka	Arsenal	24	DM	2	35.0	1815	5.5
11		Granit Xhaka	Arsenal	24	DM	2	35.0	1815	5.5
12		Alex Oxlade-Chamberlain	Arsenal	23	RM	2	22.0	1519	6.0
...	...	...	...	...	...	...	...	...	...
444		Manuel Lanzini	West+Ham	24	AM	1	15.0	493	7.0
460		Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5
461		Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5
462		Sam Byram	West+Ham	23	RB	3	4.5	198	4.5
463		Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5

130 rows × 17 columns



In [13]:

```
1 players[players.age<=25]
```

Out[13]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	f
5	Hector Bellerin	Arsenal	22	RB		3	30.0	1675	6.0	1
8	Shkodran Mustafi	Arsenal	25	CB		3	30.0	1877	5.5	
9	Alex Iwobi	Arsenal	21	LW		1	10.0	1812	5.5	
10	Granit Xhaka	Arsenal	24	DM		2	35.0	1815	5.5	
11	Granit Xhaka	Arsenal	24	DM		2	35.0	1815	5.5	
...	...	...	...	...		...	...	...	...	...
456	Pedro Obiang	West+Ham	25	CM		2	9.0	286	4.5	
460	Edimilson Fernandes	West+Ham	21	CM		2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB		3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB		3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF		1	1.0	412	4.5	

172 rows × 17 columns



In [14]:

```
1 players.loc[players.age<=25]
```

Out[14]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	f
5	Hector Bellerin	Arsenal	22	RB		3	30.0	1675	6.0	1
8	Shkodran Mustafi	Arsenal	25	CB		3	30.0	1877	5.5	1
9	Alex Iwobi	Arsenal	21	LW		1	10.0	1812	5.5	1
10	Granit Xhaka	Arsenal	24	DM		2	35.0	1815	5.5	1
11	Granit Xhaka	Arsenal	24	DM		2	35.0	1815	5.5	1
...	...	...	...	...		...	...	...	...	...
456	Pedro Obiang	West+Ham	25	CM		2	9.0	286	4.5	
460	Edimilson Fernandes	West+Ham	21	CM		2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB		3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB		3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF		1	1.0	412	4.5	

172 rows × 17 columns



In [15]:

```
1 players.loc[players.age.le(25)]
```

Out[15]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
5	Hector Bellerin	Arsenal	22	RB	3	30.0	1675	6.0	1
8	Shkodran Mustafi	Arsenal	25	CB	3	30.0	1877	5.5	1
9	Alex Iwobi	Arsenal	21	LW	1	10.0	1812	5.5	1
10	Granit Xhaka	Arsenal	24	DM	2	35.0	1815	5.5	1
11	Granit Xhaka	Arsenal	24	DM	2	35.0	1815	5.5	1
...	...	...	...	...	...	...	...	...	...
456	Pedro Obiang	West+Ham	25	CM	2	9.0	286	4.5	1
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	1
461	Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5	1
462	Sam Byram	West+Ham	23	RB	3	4.5	198	4.5	1
463	Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5	1

172 rows × 17 columns



In [16]:

```
1 players.loc[players.age<=25].equals(players.loc[players.age.le(25)])
```

Out[16]:

True

**Binary Operations with booleans**

In [17]:

```
1 # Binary OR -> | Always False unless there's a True expression
```

In [18]:

```
1 True | False
```

Out[18]:

True

In [19]:

```
1 True | True
```

Out[19]:

True

In [20]:

```
1 False | True
```

Out[20]:

True

In [21]:

```
1 False | False
```

Out[21]:

False

In [22]:

```
1 # Binary AND -> & Always True unless there's a False expression
```

In [23]:

```
1 True & False
```

Out[23]:

False

In [24]:

```
1 True & True
```

Out[24]:

True

In [25]:

```
1 False & True
```

Out[25]:

False

In [26]:

```
1 False & False
```

Out[26]:

False

In [27]:

```
1 f= pd.Series(False)
```

In [28]:

```
1 t = pd.Series(True)
```

In [29]:

```
1 t & f
```

Out[29]:

```
0    False
dtype: bool
```

In [30]:

```
1 t | f
```

Out[30]:

```
0    True
dtype: bool
```

In [31]:

```
1 t = pd.Series([True if i%2==0 else False for i in range(10)])
```

In [32]:

```
1 f = pd.Series([False for i in range(10)])
```

In [33]:

```
1 t
```

Out[33]:

```
0    True
1    False
2    True
3    False
4    True
5    False
6    True
7    False
8    True
9    False
dtype: bool
```

In [34]:

```
1 | f
```

Out[34]:

```
0    False
1    False
2    False
3    False
4    False
5    False
6    False
7    False
8    False
9    False
dtype: bool
```

In [35]:

```
1 | t & f
```

Out[35]:

```
0    False
1    False
2    False
3    False
4    False
5    False
6    False
7    False
8    False
9    False
dtype: bool
```

In [36]:

```
1 | t | f
```

Out[36]:

```
0    True
1    False
2    True
3    False
4    True
5    False
6    True
7    False
8    True
9    False
dtype: bool
```

In [37]:

```
1 | f = pd.Series(data=[False, True, True], index=['c', 'b', 'a'])
```

In [38]:

```
1 t = pd.Series(data=[True, False, False], index=['a', 'b', 'c'])
```

In [39]:

```
1 t
```

Out[39]:

```
a    True  
b    False  
c    False  
dtype: bool
```

In [40]:

```
1 f
```

Out[40]:

```
c    False  
b    True  
a    True  
dtype: bool
```

In [41]:

```
1 t|f
```

Out[41]:

```
a    True  
b    True  
c    False  
dtype: bool
```

In [42]:

```
1 t&f
```

Out[42]:

```
a    True  
b    False  
c    False  
dtype: bool
```

## XOR And Complement Binary Operations

In [43]:

```
1 #Binary XOR -> ^ inputs are different to give True  
2 #inputs are the same to give False
```

In [44]:

```
1 True ^ True
```

Out[44]:

False

In [45]:

```
1 True ^ False
```

Out[45]:

True

In [46]:

```
1 False ^ True
```

Out[46]:

True

In [47]:

```
1 False ^ False
```

Out[47]:

False

In [48]:

```
1 True^(False|False&True) | False
```

Out[48]:

True

In [49]:

```
1 #Two's Complement -> ~ tilde(-) negates
```

In [50]:

```
1 ~False
```

Out[50]:

-1

In [51]:

```
1 ~True
```

Out[51]:

-2

In [52]:

```
1 t = pd.Series([True, False, False])
```

In [53]:

```
1 t
```

Out[53]:

```
0    True
1   False
2   False
dtype: bool
```

## Combining Conditions

In [54]:

```
1 #Select all the left backs -> LB
```

In [55]:

```
1 players.head()
```

Out[55]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

◀ ▶

In [56]:

```
1 players.position == 'LB'
```

Out[56]:

```
0    False
1    False
2    False
3    False
4    False
...
460   False
461   True
462   False
463   False
464   False
Name: position, Length: 465, dtype: bool
```

In [57]:

```
1 players[players.position=='LB']
```

Out[57]:

		name	club	age	position	position_cat	market_value	page_views	fpl_rank
7	Nacho Monreal		Arsenal	31	LB	3	13.00	555	
18	Kieran Gibbs		Arsenal	27	LB	3	10.00	489	
29	Sead Kolasinac		Arsenal	24	LB	3	15.00	618	
34	Charlie Daniels		Bournemouth	30	LB	3	3.00	185	
54	Brad Smith		Bournemouth	23	LB	3	2.00	297	
62	Gaetan Bong	Brighton+and+Hove		29	LB	3	1.50	97	
65	Markus Suttner	Brighton+and+Hove		30	LB	3	2.00	23	
82	Stephen Ward		Burnley	31	LB	3	1.50	152	
99	Marcos Alonso Mendoza		Chelsea	26	LB	3	25.00	3069	
112	Kenedy		Chelsea	21	LB	3	7.00	566	
119	Patrick van Aanholt		Crystal+Palace	26	LB	3	9.00	524	
128	Jeffrey Schlupp		Crystal+Palace	24	LB	3	8.00	385	
134	Pape Souare		Crystal+Palace	27	LB	3	6.00	186	
138	Leighton Baines		Everton	32	LB	3	12.00	491	
168	Chris Lowe		Huddersfield	28	LB	3	1.50	84	
175	Scott Malone		Huddersfield	26	LB	3	0.75	81	
196	Christian Fuchs		Leicester+City	31	LB	3	6.00	417	
212	Ben Chilwell		Leicester+City	20	LB	3	2.50	288	
236	Alberto Moreno		Liverpool	25	LB	3	10.00	397	
251	Aleksandar Kolarov		Manchester+City	31	LB	3	10.00	633	
270	Daley Blind	Manchester+United		27	LB	3	22.00	867	
281	Luke Shaw	Manchester+United		22	LB	3	20.00	947	
294	Paul Dummett	Newcastle+United		25	LB	3	3.50	177	
298	Massadio Haidara	Newcastle+United		24	LB	3	1.50	114	

	name	club	age	position	position_cat	market_value	page_views	fpl_rank
312	Ryan Bertrand	Southampton	27	LB	3	20.00	578	
328	Matt Targett	Southampton	21	LB	3	3.00	110	
336	Erik Pieters	Stoke+City	28	LB	3	7.00	134	
354	Josh Tymon	Stoke+City	18	LB	3	1.00	120	
363	Martin Olsson	Swansea	29	LB	3	5.00	458	
369	Stephen Kingsley	Swansea	22	LB	3	0.75	168	
389	Ben Davies	Tottenham	24	LB	3	12.00	396	
390	Danny Rose	Tottenham	27	LB	3	28.00	848	
403	Jose Holebas	Watford	33	LB	3	2.00	345	
455	Aaron Cresswell	West+Ham	27	LB	3	12.00	380	
461	Arthur Masuaku	West+Ham	23	LB	3	7.00	199	

◀ ▶

In [58]:

```
1 players[players.position=='LB'].shape[0]
```

Out[58]:

35

In [59]:

```
1 #all left backs who are age 25 or younger
```

In [60]:

```
1 players[players.age <= 25]
```

Out[60]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	fpl_poi
5	Hector Bellerin	Arsenal	Arsenal	22	RB	3	30.0	1675	6.0	13.70%	
8	Shkodran Mustafi	Arsenal	Arsenal	25	CB	3	30.0	1877	5.5	4.00%	
9	Alex Iwobi	Arsenal	Arsenal	21	LW	1	10.0	1812	5.5	1.00%	
10	Granit Xhaka	Arsenal	Arsenal	24	DM	2	35.0	1815	5.5	2.00%	
11	Granit Xhaka	Arsenal	Arsenal	24	DM	2	35.0	1815	5.5	2.00%	
...	...	...	...	...	...	...	...	...	...	...	...
456	Pedro Obiang	West+Ham	West+Ham	25	CM	2	9.0	286	4.5	0.30%	

In [61]:

```
1 players[players.age <= 25].shape[0]
```

Out[61]:

172

In [62]:

```
1 players[(players.position=='LB') & (players.age<=25)]
```

Out[62]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value
29		Sead Kolasinac	Arsenal	24	LB	3	15.00	618	
54		Brad Smith	Bournemouth	23	LB	3	2.00	297	
112		Kenedy	Chelsea	21	LB	3	7.00	566	
128		Jeffrey Schlupp	Crystal+Palace	24	LB	3	8.00	385	
212		Ben Chilwell	Leicester+City	20	LB	3	2.50	288	
236		Alberto Moreno	Liverpool	25	LB	3	10.00	397	
281		Luke Shaw	Manchester+United	22	LB	3	20.00	947	
294		Paul Dummett	Newcastle+United	25	LB	3	3.50	177	
298		Massadio Haidara	Newcastle+United	24	LB	3	1.50	114	
328		Matt Targett	Southampton	21	LB	3	3.00	110	
354		Josh Tymon	Stoke+City	18	LB	3	1.00	120	
369		Stephen Kingsley	Swansea	22	LB	3	0.75	168	
389		Ben Davies	Tottenham	24	LB	3	12.00	396	
461		Arthur Masuaku	West+Ham	23	LB	3	7.00	199	

In [63]:

```
1 players[(players.position=='LB') & (players.age<=25)].shape[0]
```

Out[63]:

14

In [64]:

```
1 #3.conditions: Left backs, 25 or younger, market value 10M+
```

In [65]:

```

1 players[(players.position=='LB') &
2     (players.age<=25) &
3     (players.market_value>=10)
4 ]

```

Out[65]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
29	Sead Kolasinac	Arsenal	24	LB	3	15.0	618	4.
236	Alberto Moreno	Liverpool	25	LB	3	10.0	397	4.
281	Luke Shaw	Manchester+United	22	LB	3	20.0	947	5.
389	Ben Davies	Tottenham	24	LB	3	12.0	396	4.

In [66]:

```
1 #4.conditions: Left backs, 25 or younger, market value 10M+, and not from Arsenal or Tottenham
```

In [67]:

```

1 players[(players.position=='LB') &
2     (players.age<=25) &
3     (players.market_value>=10) &
4     ~(players.club.isin(['Arsenal', 'Tottenham']))]
5

```

Out[67]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
236	Alberto Moreno	Liverpool	25	LB	3	10.0	397	4.
281	Luke Shaw	Manchester+United	22	LB	3	20.0	947	5.

In [68]:

```

1 players[(players.position=='LB') |
2     (players.age<=25) ^
3     (players.market_value>=10) &
4     ~(players.club.isin(['Arsenal', 'Tottenham']))]
5

```

Out[68]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	fpl_poi
5	Hector Bellerin	Arsenal	22	RB	3	30.0	1675	6.0	13.70%	
7	Nacho Monreal	Arsenal	31	LB	3	13.0	555	5.5	4.70%	
8	Shkodran Mustafi	Arsenal	25	CB	3	30.0	1877	5.5	4.00%	
9	Alex Iwobi	Arsenal	21	LW	1	10.0	1812	5.5	1.00%	
10	Granit Xhaka	Arsenal	24	DM	2	35.0	1815	5.5	2.00%	
...	...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	0.40%	

## Conditions As Variables

In [69]:

```

1 players[(players.position=='LB') &
2     (players.age<=25) &
3     (players.market_value>=10) &
4     ~(players.club.isin(['Arsenal', 'Tottenham']))]
5

```

Out[69]:

	name	club	age	position	position_cat	market_value	page_views	fpl_valu
236	Alberto Moreno	Liverpool	25	LB	3	10.0	397	4.
281	Luke Shaw	Manchester+United	22	LB	3	20.0	947	5.

In [70]:

```
1 #new target: Arsenal rightbacks and chelsea goalkeepers
```

In [71]:

```
1 players[(players.club=='Arsenal') & (players.position=='RB')]
```

Out[71]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_se
5	Hector Bellerin	Arsenal	22		RB	3	30.0	1675	6.0	13.70%
27	Carl Jenkinson	Arsenal	25		RB	3	5.0	561	4.5	0.40%

In [72]:

```
1 arsenal_and_RB = (players.club=='Arsenal') & (players.position=='RB')
```

In [73]:

```
1 chelsea_and_GK = (players.club=='Chelsea') & (players.position=='GK')
```

In [74]:

```
1 players[arsenal_and_RB | chelsea_and_GK]
```

Out[74]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_se
5	Hector Bellerin	Arsenal	22		RB	3	30.0	1675	6.0	13.70%
27	Carl Jenkinson	Arsenal	25		RB	3	5.0	561	4.5	0.40%
102	Thibaut Courtois	Chelsea	25		GK	4	40.0	1260	5.5	18.50%
109	Willy Caballero	Chelsea	35		GK	4	1.5	542	5.0	0.20%

## Skill Challenge

In [75]:

```
1 players.nationality=='England'
```

Out[75]:

```
0    False
1    False
2    False
3    True
4    False
...
460   False
461   False
462   True
463   True
464   False
Name: nationality, Length: 465, dtype: bool
```

In [76]:

```
1 eng_nationality = players.nationality=='England'
```

In [77]:

```
1 players[players.nationality=='England'].shape[0]
```

Out[77]:

158

In [78]:

```
1 average = players.market_value.mean()
```

In [79]:

```
1 twice_mvalue = 2*average
```

In [80]:

```
1 twice_mvalue
```

Out[80]:

22.251298701298698

In [81]:

```
1 players.head()
```

Out[81]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [82]:

```
1 mvalue = players.market_value
```

In [83]:

```
1 views_and_signing = (players.page_views>4000) | (players.new_signing==1)
```

In [84]:

```
1 views_and_signing
```

Out[84]:

```
0      True
1      True
2     False
3     False
4     False
...
460    True
461    True
462   False
463    True
464   False
Length: 465, dtype: bool
```

In [85]:

```
1 players[(eng_nationality) & (mvalue>twice_mvalue) & (views_and_signing)]
```

Out[85]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
256	John Stones	Manchester+City	23	CB	3	35.0	1078	5.5
380	Dele Alli	Tottenham	21	CM	2	45.0	4626	9.5
381	Harry Kane	Tottenham	23	CF	1	60.0	4161	12.5

## 2d Indexing

In [86]:

```
1 players.head()
```

Out[86]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [87]:

```
1 #Chelsea players and 23 years old or younger
```

In [88]:

```
1 chelsea_23under = (players.club=='Chelsea') & (players.age.le(23))
```

In [89]:

1 chelsea\_23under

Out[89]:

```
0    False
1    False
2    False
3    False
4    False
...
460   False
461   False
462   False
463   False
464   False
Length: 465, dtype: bool
```

In [90]:

1 players.loc[chelsea\_23under]

Out[90]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_s
110	Michy Batshuayi	Chelsea	23	CF		1	25.0	1162	8.5	1.60
111	Kurt Zouma	Chelsea	22	CB		3	15.0	723	5.5	0.80
112	Kenedy	Chelsea	21	LB		3	7.0	566	5.0	0.10
115	Tiemoue Bakayoko	Chelsea	22	DM		2	16.0	1011	5.0	1.60

◀ ▶

In [91]:

1 players.loc[chelsea\_23under, ['name', 'position', 'market\_value', 'age']]

Out[91]:

		name	position	market_value	age
110	Michy Batshuayi		CF	25.0	23
111	Kurt Zouma		CB	15.0	22
112	Kenedy		LB	7.0	21
115	Tiemoue Bakayoko		DM	16.0	22

In [92]:

1 #Select all columns that begins with 'p'

In [93]:

1 #using startswith

In [94]:

```
1 p_cols = players.columns.str.startswith('p')
```

In [95]:

```
1 p_cols
```

Out[95]:

```
array([False, False, False, True, True, False, True, False, False,
       False, False, False, False, False, False, False])
```

In [96]:

```
1 players.loc[chelsea_23under, p_cols]
```

Out[96]:

	position	position_cat	page_views
110	CF	1	1162
111	CB	3	723
112	LB	3	566
115	DM	2	1011

In [97]:

```
1 chelsea_23under.shape
```

Out[97]:

```
(465,)
```

In [98]:

```
1 players.shape
```

Out[98]:

```
(465, 17)
```

In [99]:

```
1 p_cols.shape
```

Out[99]:

```
(17,)
```

In [100]:

```
1 # [] chaining
```

In [101]:

```
1 players[chelsea_23under]['position'] # this is slower
```

Out[101]:

```
110    CF
111    CB
112    LB
115    DM
Name: position, dtype: object
```

In [102]:

```
1 players.loc[chelsea_23under, 'position'] # this is faster
```

Out[102]:

```
110    CF
111    CB
112    LB
115    DM
Name: position, dtype: object
```

**Fancy Indexing with lookup()**

In [103]:

```
1 players.lookup([450], ['age'])
```

Out[103]:

```
array([30], dtype=int64)
```

In [104]:

```
1 players.loc[[0,132], ('name', 'market_value')] # fancy
```

Out[104]:

	name	market_value
<b>0</b>	Alexis Sanchez	65.0
<b>132</b>	Connor Wickham	6.0

In [105]:

```
1 players.lookup([0,132], ('name', 'market_value'))
```

Out[105]:

```
array(['Alexis Sanchez', 6.0], dtype=object)
```

In [106]:

```
1 players.lookup([0,132], ('market_value', 'name'))
```

Out[106]:

```
array([65.0, 'Connor Wickham'], dtype=object)
```

In [107]:

```
1 players.head()
```

Out[107]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [108]:

```
1 names = ['Petr Cech', 'Mesut Ozil', 'Alexis Sanchez']
2 attributes = ['age', 'market_value', 'page_views']
```

In [109]:

```
1 players.set_index('name').lookup(names, attributes)
```

Out[109]:

```
array([ 35.,  50., 4329.])
```

## Sorting by index or column

In [110]:

```
1 players.head()
```

Out[110]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [111]:

```
1 players.sort_values(by='market_value', ascending=False)
```

Out[111]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
96	Eden Hazard	Chelsea	26	LW	1	75.00	4220	10.5	2.30%
267	Paul Pogba	Manchester+United	24	CM	2	75.00	7435	8.0	19.50%
0	Alexis Sanchez	Arsenal	28	LW	1	65.00	4329	12.0	17.10%
244	Kevin De Bruyne	Manchester+City	26	AM	1	65.00	2252	10.0	17.50%
245	Sergio Aguero	Manchester+City	29	CF	1	65.00	4046	11.5	9.70%
...	...	...	...	...	...	...	...	...	...
287	Joel Castro Pereira	Manchester+United	21	GK	4	0.10	395	4.0	1.00%

In [112]:

```
1 players.index
```

Out[112]:

```
RangeIndex(start=0, stop=465, step=1)
```

In [113]:

```
1 type(players.index)
```

Out[113]:

```
pandas.core.indexes.range.RangeIndex
```

In [114]:

```
1 players.info(verbose= False)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 465 entries, 0 to 464
Columns: 17 entries, name to new_signing
dtypes: float64(2), int64(10), object(5)
memory usage: 61.9+ KB
```

In [115]:

```
1 players.set_index('name', inplace=True)
```

In [116]:

```
1 players.head(4)
```

Out[116]:

	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	fpl
name									

Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%	
Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%	
Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%	
Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%	

◀ ▶

In [117]:

```
1 players.sort_index().head(10)
```

Out[117]:

	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_index
name								
Aaron Cresswell	West+Ham	27	LB	3	12.0	380	5.0	1.30
Aaron Lennon	Everton	30	RW	1	5.0	504	5.5	0.20
Aaron Mooy	Huddersfield	26	CM	2	5.0	588	5.5	2.50
Aaron Ramsey	Arsenal	26	CM	2	35.0	1040	7.0	5.10
Abdoulaye Doucoure	Watford	24	CM	2	6.0	124	5.0	0.00
Adam Federici	Bournemouth	32	GK	4	1.0	126	4.0	1.50
Adam Lallana	Liverpool	29	AM	1	25.0	1808	7.5	6.40
Adam Smith	Bournemouth	26	RB	3	5.0	200	5.0	0.90
Ademola Lookman	Everton	19	LW	1	5.0	1387	5.5	0.30
Adrian	West+Ham	30	GK	4	8.0	266	4.5	0.80

In [118]:

```
1 players.sort_index(inplace=True)
```

In [119]:

```
1 players.head()
```

Out[119]:

	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_s
name								
Aaron Cresswell	West+Ham	27	LB	3	12.0	380	5.0	1.30
Aaron Lennon	Everton	30	RW	1	5.0	504	5.5	0.20
Aaron Mooy	Huddersfield	26	CM	2	5.0	588	5.5	2.50
Aaron Ramsey	Arsenal	26	CM	2	35.0	1040	7.0	5.10
Abdoulaye Doucoure	Watford	24	CM	2	6.0	124	5.0	0.00

In [120]:

```
1 players.sort_index(axis=1)
```

Out[120]:

	age	age_cat	big_club	club	club_id	fpl_points	fpl_sel	fpl_value	market_value	nat
name										
Aaron Cresswell	27	3	0	West+Ham	20	60	1.30%	5.0	12.0	E
Aaron Lennon	30	4	0	Everton	7	22	0.20%	5.5	5.0	E
Aaron Mooy	26	3	0	Huddersfield	8	0	2.50%	5.5	5.0	AI
Aaron Ramsey	26	3	1	Arsenal	1	56	5.10%	7.0	35.0	
Abdoulaye Doucoure	24	2	0	Watford	18	38	0.00%	5.0	6.0	
...	...	...	...	...	...	...	...	...	...	...

In [121]:

```
1 players.reset_index(inplace=True)
```

In [122]:

```
1 players.head()
```

Out[122]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
0	Aaron Cresswell	West+Ham	27	LB	3	12.0	380	5.0
1	Aaron Lennon	Everton	30	RW	1	5.0	504	5.5
2	Aaron Mooy	Huddersfield	26	CM	2	5.0	588	5.5
3	Aaron Ramsey	Arsenal	26	CM	2	35.0	1040	7.0
4	Abdoulaye Doucoure	Watford	24	CM	2	6.0	124	5.0

### Sorting Vs Reordering

In [123]:

```
1 #reindex
```

In [124]:

```
1 players.iloc[:4, :4]
```

Out[124]:

	name	club	age	position
0	Aaron Cresswell	West+Ham	27	LB
1	Aaron Lennon	Everton	30	RW
2	Aaron Mooy	Huddersfield	26	CM
3	Aaron Ramsey	Arsenal	26	CM

In [125]:

```
1 players_lite = players.iloc[:4, :4]
```

In [126]:

```
1 #row_order: 2,1 ,3, 0
2 #column_order: age, name, position, club
```

In [127]:

```
1 players_lite.reindex(index=[2,1,3,0], columns=['age', 'name', 'position', 'club'])
```

Out[127]:

	age	name	position	club
2	26	Aaron Mooy	CM	Huddersfield
1	30	Aaron Lennon	RW	Everton
3	26	Aaron Ramsey	CM	Arsenal
0	27	Aaron Cresswell	LB	West+Ham

In [128]:

```
1 players.reindex(index=[2,1,3,0], columns=['age', 'name', 'position', 'club'])
```

Out[128]:

	age	name	position	club
2	26	Aaron Mooy	CM	Huddersfield
1	30	Aaron Lennon	RW	Everton
3	26	Aaron Ramsey	CM	Arsenal
0	27	Aaron Cresswell	LB	West+Ham

In [129]:

```
1 #get all the columns in alphabetical order
```

In [130]:

```
1 players.reindex(index=[2,1,3,0]).sort_index(axis=1)
```

Out[130]:

	age	age_cat	big_club	club	club_id	fpl_points	fpl_sel	fpl_value	market_value
2	26	3	0	Huddersfield	8	0	2.50%	5.5	5.0
1	30	4	0	Everton	7	22	0.20%	5.5	5.0
3	26	3	1	Arsenal	1	56	5.10%	7.0	35.0
0	27	3	0	West+Ham	20	60	1.30%	5.0	12.0

In [131]:

```
1 #OR
2 #players.reindex(index=[2,1,3,0], columns=[])
```

In [132]:

```
1 # how do we get a sorted list of column labels?
```

In [133]:

```
1 players.columns
```

Out[133]:

```
Index(['name', 'club', 'age', 'position', 'position_cat', 'market_value',
       'page_views', 'fpl_value', 'fpl_sel', 'fpl_points', 'region',
       'nationality', 'new_foreign', 'age_cat', 'club_id', 'big_club',
       'new_signing'],
      dtype='object')
```

In [134]:

```
1 iter(players.columns)
```

Out[134]:

```
<map at 0x1ea1217bd48>
```

In [135]:

```
1 sorted(players.columns)
```

Out[135]:

```
['age',
 'age_cat',
 'big_club',
 'club',
 'club_id',
 'fpl_points',
 'fpl_sel',
 'fpl_value',
 'market_value',
 'name',
 'nationality',
 'new_foreign',
 'new_signing',
 'page_views',
 'position',
 'position_cat',
 'region']
```

In [136]:

```
1 players.reindex(index=[2,1,3,0], columns=sorted(players.columns))
```

Out[136]:

	age	age_cat	big_club	club	club_id	fpl_points	fpl_sel	fpl_value	market_value
2	26	3	0	Huddersfield	8	0	2.50%	5.5	5.0
1	30	4	0	Everton	7	22	0.20%	5.5	5.0
3	26	3	1	Arsenal	1	56	5.10%	7.0	35.0
0	27	3	0	West+Ham	20	60	1.30%	5.0	12.0 C



In [137]:

```
1 players.reindex(index=[2,1,3,0], columns=sorted(players.columns)[:6])
```

Out[137]:

	age	age_cat	big_club	club	club_id	fpl_points
2	26	3	0	Huddersfield	8	0
1	30	4	0	Everton	7	22
3	26	3	1	Arsenal	1	56
0	27	3	0	West+Ham	20	60

In [138]:

```
1 #Another way to sort columns
```

In [139]:

```
1 players.reindex(index=[2,1,3,0], columns=players.columns.sort_values())
```

Out[139]:

	age	age_cat	big_club	club	club_id	fpl_points	fpl_sel	fpl_value	market_value
2	26	3	0	Huddersfield	8	0	2.50%	5.5	5.0
1	30	4	0	Everton	7	22	0.20%	5.5	5.0
3	26	3	1	Arsenal	1	56	5.10%	7.0	35.0
0	27	3	0	West+Ham	20	60	1.30%	5.0	12.0 C



**Avoid Sorting like this**

In [140]:

```
1 #anti-pattern
```

In [141]:

```
1 df = players.iloc[:6, :6]
```

In [142]:

```
1 df
```

Out[142]:

	name	club	age	position	position_cat	market_value
0	Aaron Cresswell	West+Ham	27	LB	3	12.0
1	Aaron Lennon	Everton	30	RW	1	5.0
2	Aaron Mooy	Huddersfield	26	CM	2	5.0
3	Aaron Ramsey	Arsenal	26	CM	2	35.0
4	Abdoulaye Doucoure	Watford	24	CM	2	6.0
5	Adam Federici	Bournemouth	32	GK	4	1.0

In [143]:

```
1 df.swapaxes(1,0)
```

Out[143]:

	0	1	2	3	4	5
name	Aaron Cresswell	Aaron Lennon	Aaron Mooy	Aaron Ramsey	Abdoulaye Doucoure	Adam Federici
club	West+Ham	Everton	Huddersfield	Arsenal	Watford	Bournemouth
age	27	30	26	26	24	32
position	LB	RW	CM	CM	CM	GK
position_cat	3	1	2	2	2	4
market_value	12	5	5	35	6	1

In [144]:

1 df.T

Out[144]:

	0	1	2	3	4	5
name	Aaron Cresswell	Aaron Lennon	Aaron Mooy	Aaron Ramsey	Abdoulaye Doucoure	Adam Federici
club	West+Ham	Everton	Huddersfield	Arsenal	Watford	Bournemouth
age	27	30	26	26	24	32
position	LB	RW	CM	CM	CM	GK
position_cat	3	1	2	2	2	4
market_value	12	5	5	35	6	1

In [145]:

1 df.T.sort\_index()

Out[145]:

	0	1	2	3	4	5
age	27	30	26	26	24	32
club	West+Ham	Everton	Huddersfield	Arsenal	Watford	Bournemouth
market_value	12	5	5	35	6	1
name	Aaron Cresswell	Aaron Lennon	Aaron Mooy	Aaron Ramsey	Abdoulaye Doucoure	Adam Federici
position	LB	RW	CM	CM	CM	GK
position_cat	3	1	2	2	2	4

In [146]:

1 df.T.sort\_index().T

Out[146]:

	age	club	market_value	name	position	position_cat
0	27	West+Ham	12	Aaron Cresswell	LB	3
1	30	Everton	5	Aaron Lennon	RW	1
2	26	Huddersfield	5	Aaron Mooy	CM	2
3	26	Arsenal	35	Aaron Ramsey	CM	2
4	24	Watford	6	Abdoulaye Doucoure	CM	2
5	32	Bournemouth	1	Adam Federici	GK	4

In [147]:

```
1 df.sort_index(axis=1)
```

Out[147]:

	age	club	market_value	name	position	position_cat
0	27	West+Ham	12.0	Aaron Cresswell	LB	3
1	30	Everton	5.0	Aaron Lennon	RW	1
2	26	Huddersfield	5.0	Aaron Mooy	CM	2
3	26	Arsenal	35.0	Aaron Ramsey	CM	2
4	24	Watford	6.0	Abdoulaye Doucoure	CM	2
5	32	Bournemouth	1.0	Adam Federici	GK	4

## Skill Challenge

In [148]:

```
1 #1. Sort the players in the dataframe by age in ascending order. Who is the youngest fo
```

In [149]:

```
1 players.head()
```

Out[149]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl
0	Aaron Cresswell	West+Ham	27	LB	3	12.0	380	5.0	1.
1	Aaron Lennon	Everton	30	RW	1	5.0	504	5.5	0.
2	Aaron Mooy	Huddersfield	26	CM	2	5.0	588	5.5	2.
3	Aaron Ramsey	Arsenal	26	CM	2	35.0	1040	7.0	5.
4	Abdoulaye Doucoure	Watford	24	CM	2	6.0	124	5.0	0.

◀ ▶

In [150]:

```
1 players.sort_values(by='age', ascending=True)
```

Out[150]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
53		Ben Woodburn	Liverpool	17	LW	1	1.50	1241	4.5	0.10%
217		Jonathan Leko	West+Brom	18	RW	1	1.50	169	4.5	0.20%
434		Trent Alexander-Arnold	Liverpool	18	RB	3	1.50	327	4.5	0.30%
229		Josh Tymon	Stoke+City	18	LB	3	1.00	120	4.5	0.10%
45		Axel Tuanzebe	Manchester+United	19	CB	3	1.00	279	4.0	1.70%
...		...	...	...	...	...	...	...	...	...

In [151]:

```
1 players.sort_values(by='age', ascending=True).nsmallest(1, columns='age')
```

Out[151]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_s
53		Ben Woodburn	Liverpool	17	LW	1	1.5	1241	4.5	0.10

Ben Woodburn is the youngest player, age 17

In [152]:

```
1 #2. Set the club column as the index of the dataframe. Then sort the dataframe index in
2 # changes are applied to the underlying dataframe and carry over to the next question
```

In [153]:

```
1 players.set_index('club', inplace=True)
```

In [154]:

```
1 players.sort_index(axis=0, inplace=True)
```

In [155]:

1 players

Out[155]:

		name	age	position	position_cat	market_value	page_views	fpl_value	fpl_se
	club								
<b>Arsenal</b>	David Ospina	28	GK	4	7.0	544	5.0	0.20%	
<b>Arsenal</b>	Alexandre Lacazette	26	CF	1	40.0	1183	10.5	26.50%	
<b>Arsenal</b>	Alexis Sanchez	28	LW	1	65.0	4329	12.0	17.10%	
<b>Arsenal</b>	Laurent Koscielny	31	CB	3	22.0	912	6.0	0.70%	
<b>Arsenal</b>	Mesut Ozil	28	AM	1	50.0	4395	9.5	5.60%	
...	...	...	...	...	...	...	...	...	..
<b>West+Ham</b>	Mark Noble	30	CM	2	7.0	425	5.5	0.10%	
<b>West+Ham</b>	Michail Antonio	27	RW	1	18.0	1142	7.5	0.50%	
<b>West+Ham</b>	Robert Snodgrass	29	RW	1	8.0	1210	6.0	6.50%	
<b>West+Ham</b>	Ashley Fletcher	21	CF	1	1.0	412	4.5	5.90%	
<b>West+Ham</b>	Aaron Cresswell	27	LB	3	12.0	380	5.0	1.30%	

465 rows × 16 columns

In [156]:

```
1 #3. Sort the dataframe values by club and market_value, where the club is alphabetical(A-Z)
2 # descending order(within each team, the most valuable player first)
```

In [157]:

```
1 players.sort_values(by=['club', 'market_value'], ascending=(True, False))
```

Out[157]:

		name	age	position	position_cat	market_value	page_views	fpl_value	fpl_se
club									
<b>Arsenal</b>	Alexis Sanchez	28	LW	1	65.0	4329	12.0	17.10%	
<b>Arsenal</b>	Mesut Ozil	28	AM	1	50.0	4395	9.5	5.60%	
<b>Arsenal</b>	Alexandre Lacazette	26	CF	1	40.0	1183	10.5	26.50%	
<b>Arsenal</b>	Granit Xhaka	24	DM	2	35.0	1815	5.5	2.00%	
<b>Arsenal</b>	Granit Xhaka	24	DM	2	35.0	1815	5.5	2.00%	
...	...	...	...	...	...	...	...	...	..
<b>West+Ham</b>	Edimilson Fernandes	21	CM	2	5.0	288	4.5	0.40%	
<b>West+Ham</b>	Sam Byram	23	RB	3	4.5	198	4.5	0.30%	
<b>West+Ham</b>	Darren Randolph	30	GK	4	2.5	459	4.5	0.40%	
<b>West+Ham</b>	James Collins	33	CB	3	2.0	187	4.5	0.90%	
<b>West+Ham</b>	Ashley Fletcher	21	CF	1	1.0	412	4.5	5.90%	

465 rows × 16 columns

## Identifying Dups

In [158]:

```
1 players.head()
```

Out[158]:

		name	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	f
club										
<b>Arsenal</b>	David Ospina	28	GK	4	7.0	544	5.0	0.20%		
<b>Arsenal</b>	Alexandre Lacazette	26	CF	1	40.0	1183	10.5	26.50%		
<b>Arsenal</b>	Alexis Sanchez	28	LW	1	65.0	4329	12.0	17.10%		
<b>Arsenal</b>	Laurent Koscielny	31	CB	3	22.0	912	6.0	0.70%		
<b>Arsenal</b>	Mesut Ozil	28	AM	1	50.0	4395	9.5	5.60%		

In [159]:

```
1 players.duplicated()
```

Out[159]:

club	
Arsenal	False
	...
West+Ham	False
Length:	465, dtype: bool

In [160]:

```
1 players[players.duplicated()]
```

Out[160]:

		name	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
	club								
<b>Arsenal</b>		Granit Xhaka	24	DM	2	35.0	1815	5.5	2.00%
<b>Arsenal</b>		Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.80%
<b>Arsenal</b>		Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.80%

In [161]:

```
1 players.reset_index(inplace=True)
```

In [162]:

```
1 #the subset parameter
```

In [163]:

```
1 #unique -> club, position, market value
```

In [164]:

```
1 players.duplicated(subset=['club', 'position', 'market_value'])
```

Out[164]:

```
0      False
1      False
2      False
3      False
4      False
...
460     False
461     False
462     False
463     False
464     False
Length: 465, dtype: bool
```

In [165]:

```
1 players.loc[players.duplicated(subset=['club', 'position', 'market_value'])]
```

Out[165]:

	club	name	age	position	position_cat	market_value	page_views
14	Arsenal	Granit Xhaka	24	DM	2	35.00	1815
17	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.00	1519
22	Arsenal	Danny Welbeck	26	CF	1	15.00	1521
24	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.00	1519
30	Arsenal	Petr Cech	35	GK	4	7.00	1529
46	Bournemouth	Tyrone Mings	24	CB	3	5.00	422
47	Bournemouth	Artur Boruc	37	GK	4	1.00	436
55	Bournemouth	Callum Wilson	25	CF	1	10.00	337
65	Brighton+and+Hove	Shane Duffy	25	CB	3	5.00	243
89	Burnley	Steven Defour	29	CM	2	6.00	336
178	Huddersfield	Dean Whitehead	35	CM	2	0.25	161
182	Huddersfield	Martin Cranie	30	CB	3	0.50	67
183	Huddersfield	Mark Hudson	35	CB	3	0.50	70
242	Liverpool	Jordan Henderson	27	CM	2	28.00	1236
254	Manchester+City	Fernandinho	32	DM	2	18.00	595
256	Manchester+City	Yaya Toure	34	CM	2	8.00	2318
266	Manchester+United	Marcos Rojo	27	CB	3	18.00	1063
299	Newcastle+United	Aleksandar Mitrovic	22	CF	1	10.00	781
301	Newcastle+United	Lascelles	27	CB	3	5.00	400
309	Newcastle+United	Florian Lejeune	26	CB	3	5.00	65
315	Southampton	Jordy Clasie	26	DM	2	10.00	201
353	Stoke+City	Marc Muniesa	25	CB	3	3.00	172
392	Tottenham	Victor Wanyama	26	DM	2	25.00	1002
407	Watford	Tom Cleverley	27	CM	2	8.00	703

	club	name	age	position	position_cat	market_value	page_views
418	Watford	Sebastian Prodl	30	CB	3	4.00	169
434	West+Brom	Ahmed Hegazy	26	CB	3	1.00	82

In [166]:

```
1 players.loc[players.duplicated(subset=['club', 'position', 'market_value'])].shape
```

Out[166]:

(27, 17)

In [167]:

```
1 """When dealing with duplicates, the first is the original by default unless the keep parameter is used"""
    
```

Out[167]:

'When dealing with duplicates, the first is the original by default unless the keep parameter is used'

In [168]:

```
1 players.loc[players.duplicated(subset=['club', 'age', 'position', 'market_value'], keep='first')]
```

Out[168]:

(7, 17)

In [169]:

```
1 players.loc[players.duplicated(subset=['club', 'age', 'position', 'market_value'], keep='last')]
```

Out[169]:

(7, 17)

In [170]:

```
1 players.loc[players.duplicated(subset=['club', 'age', 'position', 'market_value'], keep=False)]
```

Out[170]:

(13, 17)

## Removing Duplicates

In [171]:

```
1 players[players.duplicated()]
```

Out[171]:

	club	name	age	position	position_cat	market_value	page_views	fpl_value	fpl_
14	Arsenal	Granit Xhaka	24	DM	2	35.0	1815	5.5	2.0
17	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.8
24	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.8

In [172]:

```
1 players[players.duplicated(subset='None', keep='first')]
```

Out[172]:

	club	name	age	position	position_cat	market_value	page_views	fpl_value	fpl_
14	Arsenal	Granit Xhaka	24	DM	2	35.0	1815	5.5	2.0
17	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.8
24	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.8

In [173]:

```
1 players.market_value.mean()
```

Out[173]:

11.125649350649349

In [174]:

```
1 players.drop_duplicates(keep='first')
```

Out[174]:

	club	name	age	position	position_cat	market_value	page_views	fpl_value	f
0	Arsenal	David Ospina	28	GK	4	7.0	544	5.0	
1	Arsenal	Alexandre Lacazette	26	CF	1	40.0	1183	10.5	2
2	Arsenal	Alexis Sanchez	28	LW	1	65.0	4329	12.0	1
3	Arsenal	Laurent Koscielny	31	CB	3	22.0	912	6.0	
4	Arsenal	Mesut Ozil	28	AM	1	50.0	4395	9.5	
...	...	...	...	...	...	...	...	...	...
460	West+Ham	Mark Noble	30	CM	2	7.0	425	5.5	
461	West+Ham	Michail Antonio	27	RW	1	18.0	1142	7.5	
462	West+Ham	Robert Snodgrass	29	RW	1	8.0	1210	6.0	
463	West+Ham	Ashley Fletcher	21	CF	1	1.0	412	4.5	
464	West+Ham	Aaron Cresswell	27	LB	3	12.0	380	5.0	

462 rows × 17 columns

In [175]:

```
1 players.drop_duplicates(keep='first').market_value.mean()
```

Out[175]:

11.026252723311545

In [176]:

```
1 #NOTE: Duplicates are not necessarily bad
```

## Removing DataFrame Rows

In [177]:

```
1 players[players.duplicated()]
```

Out[177]:

	club	name	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
14	Arsenal	Granit Xhaka	24	DM	2	35.0	1815	5.5	2.0%
17	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.8%
24	Arsenal	Alex Oxlade-Chamberlain	23	RM	2	22.0	1519	6.0	1.8%

In [178]:

```
1 players = pd.read_csv('soccer.csv')
```

In [179]:

```
1 players.head()
```

Out[179]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [180]:

```
1 players[players.duplicated()]
```

Out[180]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
11	Granit Xhaka	Arsenal	24	DM	2	35.0	1815	5.5	2.0%
13	Alex Oxlade-Chamberlain	Arsenal	23	RM	2	22.0	1519	6.0	1.8%
23	Alex Oxlade-Chamberlain	Arsenal	23	RM	2	22.0	1519	6.0	1.8%

In [181]:

```
1 players.drop(labels=11, axis=0)
```

Out[181]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	1
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	
...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB	3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5	
464	Diafra Sakho	West+Ham	27	CF	1	10.0	214	5.5	

464 rows × 17 columns



In [182]:

```
1 players.drop(index=[23, 11])
```

Out[182]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	1
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	
...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB	3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5	
464	Diafra Sakho	West+Ham	27	CF	1	10.0	214	5.5	

463 rows × 17 columns

## Removing Columns

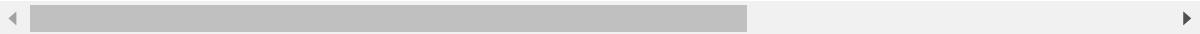
In [183]:

```
1 players.drop(columns=['club', 'position', 'market_value'])
```

Out[183]:

		name	age	position_cat	page_views	fpl_value	fpl_sel	fpl_points	region	nationalit
0		Alexis Sanchez	28	1	4329	12.0	17.10%	264	3	Chil
1		Mesut Ozil	28	1	4395	9.5	5.60%	167	2	German
2		Petr Cech	35	4	1529	5.5	5.90%	134	2	Czec Republi
3		Theo Walcott	28	1	2393	7.5	1.50%	122	1	Englan
4		Laurent Koscielny	31	3	912	6.0	0.70%	121	2	Franc
...		...	...	...	...	...	...	...	...	...
460		Edimilson Fernandes	21	2	288	4.5	0.40%	38	2	Switzerlan
461		Arthur Masuaku	23	3	199	4.5	0.20%	34	4	Congo Dl
462		Sam Byram	23	3	198	4.5	0.30%	29	1	Englan
463		Ashley Fletcher	21	1	412	4.5	5.90%	16	1	Englan
464		Diafra Sakho	27	1	214	5.5	0.10%	12	4	Seneg

465 rows × 14 columns



## Alternative

In [184]:

```
1 # using reindex to remove rows and columns
```

In [185]:

```
1 players.reindex(index=[0, 1, 2])
```

Out[185]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%

In [186]:

```
1 unwanted_rows = [1,2,3,4]
```

In [187]:

```
1 players.reindex(index=set(players.index)).difference(unwanted_rows))
```

Out[187]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	1
5	Hector Bellerin	Arsenal	22	RB	3	30.0	1675	6.0	1
6	Olivier Giroud	Arsenal	30	CF	1	22.0	2230	8.5	1
7	Nacho Monreal	Arsenal	31	LB	3	13.0	555	5.5	1
8	Shkodran Mustafi	Arsenal	25	CB	3	30.0	1877	5.5	1
...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB	3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5	
464	Diafra Sakho	West+Ham	27	CF	1	10.0	214	5.5	

461 rows × 17 columns

In [188]:

```
1 unwanted_columns = ['name', 'position']
```

In [189]:

```
1 players.reindex(columns=set(players.columns).difference(unwanted_columns))
```

Out[189]:

	big_club	nationality	fpl_points	club_id	new_foreign	fpl_value	age_cat	market_value	age	position_
0	1	Chile	264	1	0	12.0	4	65.0	28	
1	1	Germany	167	1	0	9.5	4	50.0	28	
2	1	Czech Republic	134	1	0	5.5	6	7.0	35	
3	1	England	122	1	0	7.5	4	20.0	28	
4	1	France	121	1	0	6.0	4	22.0	31	
...	...	...	...	...	...	...	...	...	...	...
460	0	Switzerland	38	20	0	4.5	1	5.0	21	
461	0	Congo DR	34	20	0	4.5	2	7.0	23	
462	0	England	29	20	0	4.5	2	4.5	23	

## Null Values in DFs

In [190]:

```
1 players.head()
```

Out[190]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [191]:

```
1 #series refresher
```

In [192]:

```
1 players.age
```

Out[192]:

```
0    28  
1    28  
2    35  
3    28  
4    31  
..  
460   21  
461   23  
462   23  
463   21  
464   27  
Name: age, Length: 465, dtype: int64
```

In [193]:

```
1 ages = players.age
```

In [194]:

```
1 ages
```

Out[194]:

```
0    28  
1    28  
2    35  
3    28  
4    31  
..  
460   21  
461   23  
462   23  
463   21  
464   27  
Name: age, Length: 465, dtype: int64
```

In [195]:

```
1 ages.isna()
```

Out[195]:

```
0    False  
1    False  
2    False  
3    False  
4    False  
..  
460   False  
461   False  
462   False  
463   False  
464   False  
Name: age, Length: 465, dtype: bool
```

In [196]:

```
1 ages[ages.isna()]
```

Out[196]:

Series([], Name: age, dtype: int64)

In [197]:

```
1 players.age.isna().sum()
```

Out[197]:

0

In [198]:

```
1 players.age.isna().isna().sum()
```

Out[198]:

0

In [199]:

```
1 #DataFrame
```

In [200]:

```
1 players.isna()
```

Out[200]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	f
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
460	False	False	False	False	False	False	False	False	False	False
461	False	False	False	False	False	False	False	False	False	False
462	False	False	False	False	False	False	False	False	False	False
463	False	False	False	False	False	False	False	False	False	False
464	False	False	False	False	False	False	False	False	False	False

465 rows × 17 columns

In [201]:

```
1 #How many NAs are there in this dataframe?
```

In [202]:

```
1 #For DataFrame
2 np.count_nonzero(players.isna())
```

Out[202]:

4

In [203]:

```
1 # For Series
2 players.position.isna().sum()
```

Out[203]:

1

In [204]:

```
1 #which records have NA in this dataframe?
```

In [205]:

```
1 players.isna().values
```

Out[205]:

```
array([[False, False, False, ..., False, False, False],
       [False, False, False, ..., False, False, False],
       [False, False, False, ..., False, False, False],
       ...,
       [False, False, False, ..., False, False, False],
       [False, False, False, ..., False, False, False],
       [False, False, False, ..., False, False, False]])
```

In [206]:

```
1 players[players.isna().values]
```

Out[206]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka		Arsenal	24	NaN	2	NaN	1815	5.5
30	Granit Xhaka		Arsenal	24	NaN	2	NaN	1815	5.5
192	Steve Mounie		Huddersfield	22	CF	1	NaN	56	6.0
195	Kasper Schmeichel		Leicester+City	30	GK	4	NaN	1601	5.0

In [207]:

```
1 players[players.isna().values].drop_duplicates()
```

Out[207]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	NaN	2	NaN	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	NaN	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	NaN	1601	5.0

## Dropping And Filling DataFrame NAs

In [208]:

```
1 players.fillna('something meaningful').loc[[30,192,195]]
```

Out[208]:

	name	club	age	position	position_cat	market_value	page_views	fpl_val
30	Granit Xhaka	Arsenal	24	something meaningful	2	something meaningful	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	something meaningful	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	something meaningful	1601	5.0

In [209]:

```
1 players.iloc[20:30]
```

Out[209]:

		name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_
20	Danny Welbeck	Arsenal	26	CF		1	15.0	1521	7.5	0.5
21	Rob Holding	Arsenal	21	CB		3	7.0	666	5.0	1.7
22	Mohamed Elneny	Arsenal	25	DM		2	10.0	683	4.5	0.8
23	Alex Oxlade-Chamberlain	Arsenal	23	RM		2	22.0	1519	6.0	1.8
24	Lucas Perez	Arsenal	28	CF		1	15.0	2055	7.5	0.1
25	Emiliano Martinez	Arsenal	24	GK		4	1.0	372	4.0	4.2
26	David Ospina	Arsenal	28	GK		4	7.0	544	5.0	0.2
27	Carl Jenkinson	Arsenal	25	RB		3	5.0	561	4.5	0.4
28	Per Mertesacker	Arsenal	32	CB		3	6.0	702	5.0	1.0
29	Sead Kolasinac	Arsenal	24	LB		3	15.0	618	6.0	6.9

In [210]:

```
1 players.index
```

Out[210]:

```
RangeIndex(start=0, stop=465, step=1)
```

In [211]:

```
1 type(players.index)
```

Out[211]:

```
pandas.core.indexes.range.RangeIndex
```

In [212]:

```
1 players.position.unique()
```

Out[212]:

```
array(['LW', 'AM', 'GK', 'RW', 'CB', 'RB', 'CF', 'LB', 'DM', 'RM', 'CM',
       nan, 'SS', 'LM'], dtype=object)
```

In [213]:

```
1 players.club.unique()[0]
```

Out[213]:

'Arsenal'

In [214]:

```
1 players.club.unique()[10]
```

Out[214]:

'Manchester+City'

In [215]:

```
1 players.fillna({'market_value':100, 'position':'RM'}).loc[[30, 192, 195]]
```

Out[215]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	RM	2	100.0	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	100.0	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	100.0	1601	5.0

In [216]:

```
1 players.fillna({'position':'wetin u dey play', 'market_value': 'Elo lowo ori e'}).loc[[30, 192, 195]]
```

Out[216]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	wetin u dey play	2	Elo lowo ori e	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	Elo lowo ori e	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	Elo lowo ori e	1601	5.0

In [217]:

```
1 #How about dropping nulls?
```

In [218]:

```
1 players.dropna()
```

Out[218]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	1
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	
...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB	3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5	
464	Diafra Sakho	West+Ham	27	CF	1	10.0	214	5.5	

462 rows × 17 columns

In [219]:

```
1 #OR
2 players.dropna(axis=1, how='any').loc[[30, 192, 195]] #this drops the columns with na v
```

Out[219]:

	name	club	age	position_cat	page_views	fpl_value	fpl_sel	fpl_points	re
30	Granit Xhaka	Arsenal	24	2	1815	5.5	2.00%	85	
192	Steve Mounie	Huddersfield	22	1	56	6.0	0.60%	0	
195	Kasper Schmeichel	Leicester+City	30	4	1601	5.0	2.40%	109	

In [220]:

```
1 players.dropna(axis=1, how='all').loc[[30, 192, 195]]# drops none of the columns because
```

Out[220]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	NaN	2	NaN	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	NaN	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	NaN	1601	5.0

In [221]:

```
1 #players.dropna(axis=0, how='any').loc[[30, 192, 195]]-this would give an error because
```

In [222]:

```
1 players.dropna(axis=0, how='all').loc[[30, 192, 195]]# this won't drop any row it does
```

Out[222]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	NaN	2	NaN	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	NaN	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	NaN	1601	5.0

## Methods And Axes with fillna()

In [223]:

```
1 players[players.isna().values]
```

Out[223]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	NaN	2	NaN	1815	5.5
30	Granit Xhaka	Arsenal	24	NaN	2	NaN	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	NaN	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	NaN	1601	5.0

In [224]:

```
1 players[players.isna().values].drop_duplicates()
```

Out[224]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	NaN	2	NaN	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	NaN	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	NaN	1601	5.0

In [225]:

```
1 players.fillna(method='ffill').loc[[30, 192, 195]]
```

Out[225]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	LB	2	15.0	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	3.0	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	30.0	1601	5.0

In [226]:

```
1 players.market_value.loc[29]
```

Out[226]:

15.0

In [227]:

```
1 players.fillna(method='bfill').loc[[30, 192, 195]] # fills across rows
```

Out[227]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	CF	2	40.0	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	15.0	56	6.0
195	Kasper Schmeichel	Leicester+City	30	GK	4	6.0	1601	5.0

In [228]:

```
1 players.market_value.loc[[31, 193, 196]]
```

Out[228]:

```
31    40.0
193   15.0
196   6.0
Name: market_value, dtype: float64
```

In [229]:

```
1 players.fillna(method='bfill', axis=1).loc[[30, 192, 195]] # fills across columns
```

Out[229]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value
30	Granit Xhaka	Arsenal	24	2	2	1815	1815	5.5
192	Steve Mounie	Huddersfield	22	CF	1	56	56	€
195	Kasper Schmeichel	Leicester+City	30	GK	4	1601	1601	£

## Skill Challenge

In [230]:

```
1 #From players dataframe, remove the label 2,10,21, & market_value column. Assign the result to df2
```

In [231]:

```
1 df2 = players.drop(index=[2,10,21], columns='market_value')
```

In [232]:

```
1 df2.shape
```

Out[232]:

(462, 16)

In [233]:

1 df2

Out[233]:

	name	club	age	position	position_cat	page_views	fpl_value	fpl_sel	fpl_poir
0	Alexis Sanchez	Arsenal	28	LW	1	4329	12.0	17.10%	2
1	Mesut Ozil	Arsenal	28	AM	1	4395	9.5	5.60%	1
3	Theo Walcott	Arsenal	28	RW	1	2393	7.5	1.50%	1
4	Laurent Koscielny	Arsenal	31	CB	3	912	6.0	0.70%	1
5	Hector Bellerin	Arsenal	22	RB	3	1675	6.0	13.70%	1
...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	288	4.5	0.40%	
461	Arthur Masuaku	West+Ham	23	LB	3	199	4.5	0.20%	
462	Sam Byram	West+Ham	23	RB	3	198	4.5	0.30%	
463	Ashley Fletcher	West+Ham	21	CF	1	412	4.5	5.90%	
464	Diafra Sakho	West+Ham	27	CF	1	214	5.5	0.10%	

462 rows × 16 columns

In [234]:

1 #Does the nationality column in df2 have any NA value? How many unique nationalities are there?

In [235]:

1 df2.nationality[df2.nationality.isna()]

Out[235]:

Series([], Name: nationality, dtype: object)

In [236]:

1 df2.nationality.isna().sum()

Out[236]:

0

In [237]:

```
1 df2.nationality.isna().isna().sum()
```

Out[237]:

0

In [238]:

```
1 np.count_nonzero(df2.nationality.isna())
```

Out[238]:

0

In [239]:

```
1 df2.nationality.unique()
```

Out[239]:

```
array(['Chile', 'Germany', 'England', 'France', 'Spain', 'Nigeria',
       'Switzerland', 'Wales', 'Brazil', 'Egypt', 'Argentina', 'Colombia',
       'Bosnia', 'Norway', 'Poland', 'Scotland', 'Congo DR', 'Ireland',
       'Netherlands', 'Australia', "Cote d'Ivoire", 'Finland', 'Cameroon',
       'Austria', 'Israel', 'Czech Republic', 'Northern Ireland',
       'Canada', 'Belgium', 'Iceland', 'Serbia', 'Portugal', 'Ghana',
       'South Korea', 'Mali', 'Senegal', 'Curacao', 'Denmark', 'Slovenia',
       'Trinidad and Tobago', 'Bermuda', 'Benin', 'Algeria', 'Jamaica',
       'Japan', 'Tunisia', 'Croatia', 'Estonia', 'Ecuador', 'Armenia',
       'Italy', 'Sweden', 'United States', 'Morocco', 'The Gambia',
       'Kenya', 'Greece', 'Uruguay', 'Romania', 'Venezuela',
       'New Zealand'], dtype=object)
```

In [240]:

```
1 df2.nationality.describe().loc['unique']
```

Out[240]:

61

In [241]:

```
1 df2.nationality.unique().size
```

Out[241]:

61

In [242]:

```
1 help(np.array)
```

Help on built-in function array in module numpy:

```
array(...)  
    array(object, dtype=None, copy=True, order='K', subok=False, ndmin=0)  
  
Create an array.  
  
Parameters  
-----  
object : array_like  
    An array, any object exposing the array interface, an object whose  
    __array__ method returns an array, or any (nested) sequence.  
dtype : data-type, optional  
    The desired data-type for the array. If not given, then the type  
will  
    be determined as the minimum type required to hold the objects in  
the  
    sequence.  
copy : bool, optional  
    To copy or not to copy the data contained in these arrays into a new  
    array. If a copy is not desired, a view is returned instead.
```

In [243]:

```
1 df2[['age', 'position']].nunique()
```

Out[243]:

```
age      22  
position  13  
dtype: int64
```

In [244]:

```
1 df2[['age', 'position']].describe()
```

Out[244]:

age
count 462.000000
mean 26.777056
std 3.939906
min 17.000000
25% 24.000000
50% 27.000000
75% 29.000000
max 38.000000

In [245]:

```
1 df2.set_index('club')
```

Out[245]:

	name	age	position	position_cat	page_views	fpl_value	fpl_sel	fpl_points	region	nation
--	------	-----	----------	--------------	------------	-----------	---------	------------	--------	--------

club

<b>Arsenal</b>	Alexis Sanchez	28	LW	1	4329	12.0	17.10%	264	3	C
<b>Arsenal</b>	Mesut Ozil	28	AM	1	4395	9.5	5.60%	167	2	Germ
<b>Arsenal</b>	Theo Walcott	28	RW	1	2393	7.5	1.50%	122	1	Eng
<b>Arsenal</b>	Laurent Koscielny	31	CB	3	912	6.0	0.70%	121	2	Fra
<b>Arsenal</b>	Hector Bellerin	22	RB	3	1675	6.0	13.70%	119	2	S
...	...	...	...	...	...	...	...	...	...	...

In [246]:

```
1 players
```

Out[246]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
--	------	------	-----	----------	--------------	--------------	------------	-----------	---

0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	1
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	
...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB	3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5	
464	Diafra Sakho	West+Ham	27	CF	1	10.0	214	5.5	

465 rows × 17 columns

In [247]:

```
1 #3.
2 df2.drop_duplicates(subset=['age', 'club', 'position'], keep='first').loc[:, ['age', 'p
```

Out[247]:

	age	position
0	28	LW
1	28	AM
3	28	RW
4	31	CB
5	22	RB
...	...	...
460	21	CM
461	23	LB
462	23	RB
463	21	CF

**31st of july 2020*****Calculating aggregates with agg()***

In [248]:

```
1 players.head()
```

Out[248]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [249]:

```
1 players.agg('mean')
```

Out[249]:

```
age           26.776344
position_cat   2.178495
market_value    11.125649
page_views     771.546237
fpl_value       5.450538
fpl_points     57.544086
region          1.989247
new_foreign     0.034409
age_cat         3.195699
club_id         10.253763
big_club        0.309677
new_signing     0.144086
dtype: float64
```

In [250]:

```
1 players.fpl_value.mean()
```

Out[250]:

```
5.450537634408602
```

In [251]:

```
1 players.agg('min')
```

Out[251]:

```
name            Aaron Cresswell
club           Arsenal
age              17
position_cat      1
market_value     0.05
page_views        3
fpl_value         4
fpl_sel          0.00%
fpl_points       0
region            1
nationality      Algeria
new_foreign       0
age_cat           1
club_id           1
big_club          0
new_signing       0
dtype: object
```

In [252]:

```
1 players.agg(min)
```

Out[252]:

```
name           Aaron Cresswell
club            Arsenal
age              17
position_cat      1
market_value     0.05
page_views        3
fpl_value         4
fpl_sel          0.00%
fpl_points       0
region            1
nationality      Algeria
new_foreign       0
age_cat           1
club_id            1
big_club          0
new_signing       0
dtype: object
```

In [253]:

```
1 players.agg(np.min)
```

Out[253]:

```
name           Aaron Cresswell
club            Arsenal
age              17
position_cat      1
market_value     0.05
page_views        3
fpl_value         4
fpl_sel          0.00%
fpl_points       0
region            1
nationality      Algeria
new_foreign       0
age_cat           1
club_id            1
big_club          0
new_signing       0
dtype: object
```

In [254]:

```
1 'a' < 'b'
```

Out[254]:

True

In [255]:

```
1 ord('a')
```

Out[255]:

97

In [256]:

```
1 ord('d')
```

Out[256]:

100

In [257]:

```
1 ls = ['a', 'b', 'c', 'day']
```

In [258]:

```
1 min(ls)
```

Out[258]:

'a'

In [259]:

```
1 max(ls)
```

Out[259]:

'day'

In [260]:

```
1 players.agg('min')
```

Out[260]:

```
name           Aaron Cresswell
club          Arsenal
age            17
position_cat      1
market_value    0.05
page_views        3
fpl_value         4
fpl_sel        0.00%
fpl_points        0
region           1
nationality      Algeria
new_foreign        0
age_cat            1
club_id            1
big_club           0
new_signing        0
dtype: object
```

In [261]:

```
1 #OR
2 players.select_dtypes(np.number).agg('min')
```

Out[261]:

age	17.00
position_cat	1.00
market_value	0.05
page_views	3.00
fpl_value	4.00
fpl_points	0.00
region	1.00
new_foreign	0.00
age_cat	1.00
club_id	1.00
big_club	0.00
new_signing	0.00
	dtype: float64

In [262]:

```
1 import sys
2 sys.version
```

Out[262]:

```
'3.7.6 (default, Jan 8 2020, 20:23:39) [MSC v.1916 64 bit (AMD64)]'
```

In [263]:

```
1 players.select_dtypes(np.number).agg(['min', 'max', 'mean'])
```

Out[263]:

	age	position_cat	market_value	page_views	fpl_value	fpl_points	region	new
min	17.000000	1.000000	0.050000	3.000000	4.000000	0.000000	1.000000	0.000000
max	38.000000	4.000000	75.000000	7664.000000	12.500000	264.000000	4.000000	4.000000
mean	26.776344	2.178495	11.125649	771.546237	5.450538	57.544086	1.989247	1.989247

## Some-Shape Transform

In [264]:

```
1 players.head(3)
```

Out[264]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%

In [265]:

```
1 #fx rate = 1.02% per year for banks
```

In [266]:

```
1 #usdeur = 0.91
```

In [267]:

```
1 players.loc[:, ['market_value', 'fpl_value']]
```

Out[267]:

	market_value	fpl_value
0	65.0	12.0
1	50.0	9.5
2	7.0	5.5
3	20.0	7.5
4	22.0	6.0
...	...	...
460	5.0	4.5
461	7.0	4.5
462	4.5	4.5
463	1.0	4.5

In [268]:

```
1 players.loc[:, ['market_value', 'fpl_value']].transform(lambda x: x * 0.91)
```

Out[268]:

	market_value	fpl_value
0	59.150	10.920
1	45.500	8.645
2	6.370	5.005
3	18.200	6.825
4	20.020	5.460
...	...	...
460	4.550	4.095
461	6.370	4.095
462	4.095	4.095
463	0.910	4.095

In [269]:

```
1 #ASIDE
```

In [270]:

```
1 #choice
```

In [271]:

```
1 from random import choice
```

In [272]:

```
1 names = ['bread', 'john', 'sam', 'mike']
```

In [273]:

```
1 choice(names)
```

Out[273]:

'bread'

In [274]:

```
1 #string methods -> .str.
```

In [275]:

```
1 'ad'.upper()
```

Out[275]:

'AD'

In [276]:

```
1 ser = pd.Series(names)
```

In [277]:

```
1 ser
```

Out[277]:

```
0    bread
1    john
2    sam
3    mike
dtype: object
```

In [278]:

```
1 #ser.upper() - this will give an error
```

In [279]:

```
1 ser.str.upper()
```

Out[279]:

```
0    BREAD
1    JOHN
2    SAM
3    MIKE
dtype: object
```

In [280]:

```
1 ser.str.lower()
```

Out[280]:

```
0    bread
1    john
2    sam
3    mike
dtype: object
```

In [281]:

```
1 ser.str.title()
```

Out[281]:

```
0    Bread
1    John
2    Sam
3    Mike
dtype: object
```

In [282]:

```
1 ser.str.swapcase()
```

Out[282]:

```
0    BREAD
1    JOHN
2    SAM
3    MIKE
dtype: object
```

In [283]:

```
1 #Creating a function that should:
2 #-apply a random string capitalization
3 #-to a sequence of values
4 #-return the transformed sequence
5 from random import choice
```

In [284]:

```
1 def random_case(x):
2     funcs=[x.str.swapcase(), x.str.lower(), x.str.upper(), x.str.title()]
3     return choice(funcs)
4 players.select_dtypes(include=object).transform(random_case)
```

Out[284]:

	name	club	position	fpl_sel	nationality
0	alexis sanchez	arsenal	Lw	17.10%	CHILE
1	mesut ozil	arsenal	Am	5.60%	GERMANY
2	petr cech	arsenal	Gk	5.90%	CZECH REPUBLIC
3	theo walcott	arsenal	Rw	1.50%	ENGLAND
4	laurent koscielny	arsenal	Cb	0.70%	FRANCE
...	...	...	...	...	...
460	edimilson fernandes	west+ham	Cm	0.40%	SWITZERLAND
461	arthur masuaku	west+ham	Lb	0.20%	CONGO DR
462	sam byram	west+ham	Rb	0.30%	ENGLAND
463	ashley fletcher	west+ham	Cf	5.90%	ENGLAND

In [285]:

```
1 for i in range(4):
2     print(players.select_dtypes(include=object).transform(random_case))
```

	name	club	position	fpl_sel	nationality
0	alexis sanchez	Arsenal	LW	17.10%	Chile
1	mesut ozil	Arsenal	AM	5.60%	Germany
2	petr cech	Arsenal	GK	5.90%	Czech Republic
3	theo walcott	Arsenal	RW	1.50%	England
4	laurent koscielny	Arsenal	CB	0.70%	France
..	...	...	...	...	...
460	edimilson fernandes	West+Ham	CM	0.40%	Switzerland
461	arthur masuaku	West+Ham	LB	0.20%	Congo Dr
462	sam byram	West+Ham	RB	0.30%	England
463	ashley fletcher	West+Ham	CF	5.90%	England
464	diafra sakho	West+Ham	CF	0.10%	Senegal

[465 rows x 5 columns]

	name	club	position	fpl_sel	nationality
0	ALEXIS SANCHEZ	Arsenal	lw	17.10%	Chile
1	MESUT OZIL	Arsenal	am	5.60%	Germany
2	PETR CECH	Arsenal	gk	5.90%	Czech Republic
3	THEO WALCOTT	Arsenal	rw	1.50%	England
..	LAURENT KOSCIELNY	...	...	...	...

**More Flexibility with apply()**

In [286]:

```
1 players.head()
```

Out[286]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [287]:

```

1 def round_floats(x):
2     if x.dtype == np.float64:
3         return round(x)
4     return x
5 players.apply(round_floats)

```

Out[287]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	fpl_poi
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%	1
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	10.0	5.60%	1
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	6.0	5.90%	1
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	8.0	1.50%	1
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%	1
...	...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.0	0.40%	1

In [288]:

```
1 players.market_value.dtype, players.fpl_value.dtype
```

Out[288]:

(dtype('float64'), dtype('float64'))

In [289]:

```

1 def round_floats(x):
2     if x.dtype == np.float64:
3         return round(x)
4     return x
5 players.transform(round_floats)

```

Out[289]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	fpl_poi
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%	1
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	10.0	5.60%	1
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	6.0	5.90%	1
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	8.0	1.50%	1
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%	1
...	...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.0	0.40%	1

In [290]:

```
1 players.select_dtypes(np.float64).head()
```

Out[290]:

	market_value	fpl_value
0	65.0	12.0
1	50.0	9.5
2	7.0	5.5
3	20.0	7.5
4	22.0	6.0

In [291]:

```
1 players.select_dtypes(np.int64)
```

Out[291]:

	age	position_cat	page_views	fpl_points	region	new_foreign	age_cat	club_id	big_club
0	28	1	4329	264	3	0	4	1	1
1	28	1	4395	167	2	0	4	1	1
2	35	4	1529	134	2	0	6	1	1
3	28	1	2393	122	1	0	4	1	1
4	31	3	912	121	2	0	4	1	1
...	...	...	...	...	...	...	...	...	...
460	21	2	288	38	2	0	1	20	0
461	23	3	199	34	4	0	2	20	0
462	23	3	198	29	1	0	2	20	0
463	21	1	412	16	1	0	1	20	0
464	27	1	214	12	4	0	3	20	0

465 rows × 10 columns

In [292]:

```
1 players.select_dtypes(np.float64).apply(round_floats).head()
```

Out[292]:

	market_value	fpl_value
0	65.0	12.0
1	50.0	10.0
2	7.0	6.0
3	20.0	8.0
4	22.0	6.0

In [293]:

```
1 #apply = agg() + transform()
```

In [294]:

```
1 # apply as aggregates
```

In [295]:

```
1 players.agg('mean')
```

Out[295]:

```
age           26.776344
position_cat   2.178495
market_value    11.125649
page_views     771.546237
fpl_value       5.450538
fpl_points      57.544086
region          1.989247
new_foreign     0.034409
age_cat          3.195699
club_id          10.253763
big_club         0.309677
new_signing      0.144086
dtype: float64
```

In [296]:

```
1 players.apply('mean')
```

Out[296]:

```
age           26.776344
position_cat   2.178495
market_value    11.125649
page_views     771.546237
fpl_value       5.450538
fpl_points      57.544086
region          1.989247
new_foreign     0.034409
age_cat          3.195699
club_id          10.253763
big_club         0.309677
new_signing      0.144086
dtype: float64
```

In [297]:

```
1 #flipping the axis parameter
```

In [298]:

```
1 players.apply('mean', axis=1)
```

Out[298]:

```
0      392.333333
1      388.208333
2      143.708333
3      214.875000
4      91.916667
       ...
460     31.875000
461     24.791667
462     23.750000
463     39.875000
464     24.708333
Length: 465, dtype: float64
```

In [299]:

```
1 players.apply('mean', axis=0)
```

Out[299]:

```
age           26.776344
position_cat   2.178495
market_value    11.125649
page_views     771.546237
fpl_value       5.450538
fpl_points     57.544086
region          1.989247
new_foreign     0.034409
age_cat         3.195699
club_id         10.253763
big_club        0.309677
new_signing     0.144086
dtype: float64
```

In [300]:

```
1 players.loc[460,:]
```

Out[300]:

```
name          Edimilson Fernandes
club          West+Ham
age            21
position        CM
position_cat    2
market_value     5
page_views      288
fpl_value       4.5
fpl_sel         0.40%
fpl_points      38
region          2
nationality     Switzerland
new_foreign      0
age_cat          1
club_id          20
big_club         0
new_signing      1
Name: 460, dtype: object
```

In [301]:

```
1 players.loc[460, [dtype != object for dtype in players.dtypes]]
```

Out[301]:

```
age            21
position_cat    2
market_value     5
page_views      288
fpl_value       4.5
fpl_points      38
region          2
new_foreign      0
age_cat          1
club_id          20
big_club         0
new_signing      1
Name: 460, dtype: object
```

In [302]:

```
1 players.dtypes
```

Out[302]:

```
name          object
club          object
age           int64
position      object
position_cat  int64
market_value  float64
page_views   int64
fpl_value    float64
fpl_sel      object
fpl_points   int64
region        int64
nationality  object
new_foreign  int64
age_cat      int64
club_id      int64
big_club     int64
new_signing  int64
dtype: object
```

In [303]:

```
1 players.loc[460, [dtypes != object for dtypes in players.dtypes]].mean()
```

Out[303]:

31.875

## Element-wise Operations with applymap()

In [304]:

```
1 #vectorized ops: apply(), transform(), agg()
```

In [305]:

```
1 #non-vectorized ops: applymap()
```

In [306]:

```
1 players.head()
```

Out[306]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [307]:

```
1 inflation = 1.02
```

In [308]:

```
1 mini_df = players.loc[:, ['market_value', 'fpl_value']]
```

In [309]:

```
1 mini_df * inflation
```

Out[309]:

	market_value	fpl_value
0	66.30	12.24
1	51.00	9.69
2	7.14	5.61
3	20.40	7.65
4	22.44	6.12
...	...	...
460	5.10	4.59
461	7.14	4.59
462	4.59	4.59
463	1.02	4.59
464	10.20	5.61

465 rows × 2 columns

In [310]:

```
1 #OR
2 players.loc[:, ['market_value', 'fpl_value']] * inflation
```

Out[310]:

	market_value	fpl_value
0	66.30	12.24
1	51.00	9.69
2	7.14	5.61
3	20.40	7.65
4	22.44	6.12
...	...	...
460	5.10	4.59
461	7.14	4.59
462	4.59	4.59
463	1.02	4.59
464	10.20	5.61

465 rows × 2 columns

In [311]:

```
1 from datetime import datetime
```

In [312]:

```
1 counter = 0
2 def log_and_transform(x):
3     global counter
4     counter += 1
5     if counter % 100 == 0:
6         print(f"It's {datetime.now()} i just adjusted the {counter}th value")
7     return x*inflation
```

In [313]:

```
1 mini_df.applymap(log_and_transform)
```

It's2021-01-30 01:30:21.522098 i just adjusted the 100th value  
 It's2021-01-30 01:30:21.522098 i just adjusted the 200th value  
 It's2021-01-30 01:30:21.530093 i just adjusted the 300th value  
 It's2021-01-30 01:30:21.530093 i just adjusted the 400th value  
 It's2021-01-30 01:30:21.530093 i just adjusted the 500th value  
 It's2021-01-30 01:30:21.530093 i just adjusted the 600th value  
 It's2021-01-30 01:30:21.538096 i just adjusted the 700th value  
 It's2021-01-30 01:30:21.538096 i just adjusted the 800th value  
 It's2021-01-30 01:30:21.538096 i just adjusted the 900th value  
 It's2021-01-30 01:30:21.546098 i just adjusted the 1000th value  
 It's2021-01-30 01:30:21.546098 i just adjusted the 1100th value  
 It's2021-01-30 01:30:21.546098 i just adjusted the 1200th value  
 It's2021-01-30 01:30:21.546098 i just adjusted the 1300th value

Out[313]:

	market_value	fpl_value
0	66.30	12.24
1	51.00	9.69
2	7.14	5.61
3	20.40	7.65
4	22.44	6.12
...	...	...
460	5.10	4.59
461	7.14	4.59
462	4.59	4.59
463	1.02	4.59
464	10.20	5.61

465 rows × 2 columns

**31st july continued....****Skill Challenge**

In [314]:

```
1 #Create a standalone function
```

In [315]:

```
1 df = players.loc[3:10, ['page_views', 'market_value']]
```

In [316]:

```
1 type(players.loc[3:10, ['page_views', 'market_value']])
```

Out[316]:

pandas.core.frame.DataFrame

In [317]:

```
1 def popularity(x):
2
3     if x < 220:
4         return "'relatively unknown'"
5     elif x < 600:
6         return "'kind of popular'"
7     elif x < 2000:
8         return "'popular'"
9     else:
10        return "'super popular'"
```

In [318]:

```
1 players['popularity']=players.page_views.apply(popularity)
```

In [319]:

```
1 players.popularity
```

Out[319]:

```
0      'super popular'
1      'super popular'
2      'popular'
3      'super popular'
4      'popular'
...
460    'kind of popular'
461    'relatively unknown'
462    'relatively unknown'
463    'kind of popular'
464    'relatively unknown'
Name: popularity, Length: 465, dtype: object
```

In [320]:

```
1 players.popularity.value_counts().loc["'super popular'"]
```

Out[320]:

37

In [321]:

```
1 players[players.popularity=='super popular']
```

Out[321]:

	name	club	age	position	position_cat	market_value	page_views	fpl_v
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	
6	Olivier Giroud	Arsenal	30	CF	1	22.0	2230	
24	Lucas Perez	Arsenal	28	CF	1	15.0	2055	
33	Jermain Defoe	Bournemouth	34	CF	1	5.0	3213	
96	Eden Hazard	Chelsea	26	LW	1	75.0	4220	
97	Diego Costa	Chelsea	28	CF	1	50.0	4454	
99	Marcos Alonso Mendoza	Chelsea	26	LB	3	25.0	3069	
103	David Luiz	Chelsea	30	CB	3	30.0	2745	
104	Cesc Fabregas	Chelsea	30	CM	2	35.0	2378	
106	Victor Moses	Chelsea	26	RM	2	18.0	2537	
108	N'Golo Kante	Chelsea	26	DM	2	50.0	4042	
147	Wayne Rooney	Everton	31	SS	1	15.0	7664	
193	Jamie Vardy	Leicester+City	30	CF	1	15.0	2988	
217	Roberto Firmino	Liverpool	25	SS	1	38.0	2196	
218	Philippe Coutinho	Liverpool	25	AM	1	45.0	2958	
219	Sadio Mane	Liverpool	25	LW	1	40.0	3219	
230	Daniel Sturridge	Liverpool	27	CF	1	20.0	2042	
244	Kevin De Bruyne	Manchester+City	26	AM	1	65.0	2252	
245	Sergio Aguero	Manchester+City	29	CF	1	65.0	4046	
246	Raheem Sterling	Manchester+City	22	LW	1	45.0	2074	

	name	club	age	position	position_cat	market_value	page_views	fpl_v
249	Leroy Sane	Manchester+City	21	LW	1	35.0	2302	
252	Yaya Toure	Manchester+City	34	CM	2	8.0	2318	
255	Gabriel Jesus	Manchester+City	20	CF	1	30.0	4254	
264	Romelu Lukaku	Manchester+United	24	CF	1	50.0	3727	
265	David de Gea	Manchester+United	26	GK	4	40.0	2126	
267	Paul Pogba	Manchester+United	24	CM	2	75.0	7435	
269	Juan Mata	Manchester+United	29	AM	1	30.0	2280	
272	Anthony Martial	Manchester+United	21	CF	1	30.0	2430	
274	Marcus Rashford	Manchester+United	19	CF	1	18.0	3668	
275	Henrikh Mkhitaryan	Manchester+United	28	AM	1	35.0	3312	
326	Manolo Gabbiadini	Southampton	25	CF	1	15.0	2012	
340	Peter Crouch	Stoke+City	36	CF	1	1.5	2656	
380	Dele Alli	Tottenham	21	CM	2	45.0	4626	
381	Harry Kane	Tottenham	23	CF	1	60.0	4161	
383	Son Heung-min	Tottenham	25	LW	1	30.0	2185	

◀ ▶

In [322]:

```
1 players[players.popularity=='super popular'].shape
```

Out[322]:

(37, 18)

## Setting DataFrame Values

In [323]:

```
1 players.head(10)
```

Out[323]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel	fpl_points	re
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%	264	
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%	167	
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%	134	
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%	122	
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%	121	
5	Hector Bellerin	Arsenal	22	RB	3	30.0	1675	6.0	13.70%	119	

In [324]:

```
1 players.head()
```

Out[324]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [325]:

```
1 players.loc[3, 'position'] = 'CM'
```

In [326]:

```
1 players.iloc[3,3] = 'RW'
```

In [327]:

```
1 players.head()
```

Out[327]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [328]:

```
1 #at and iat[] should be preferred for single value assignment
```

In [329]:

```
1 players.at[3,'position'] = 'CM'
```

In [330]:

```
1 players.head()
```

Out[330]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	CM	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [331]:

```
1 players.iat[3,3] = 'RW'
```

In [332]:

```
1 players.head()
```

Out[332]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	GK	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	RW	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [333]:

```
1 %timeit players.at[2, 'position'] = 'CM'
```

26.7 µs ± 2.98 µs per loop (mean ± std. dev. of 7 runs, 10000 loops each)

In [334]:

```
1 %timeit players.iat[3, 3] = 'CM'
```

26.3 µs ± 1.48 µs per loop (mean ± std. dev. of 7 runs, 10000 loops each)

In [335]:

```
1 %timeit players.loc[2, 'position'] = 'CM'
```

1.18 ms ± 140 µs per loop (mean ± std. dev. of 7 runs, 1000 loops each)

**View Vs Copy**

In [336]:

```
1 players.head()
```

Out[336]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	LW	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	AM	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	CM	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	CM	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [337]:

```
1 players.loc[0:3, 'position'] = ['CM', 'RW', 'CB', 'GK']
```

In [338]:

```
1 players.head()
```

Out[338]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	CM	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	RW	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	CB	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	GK	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [339]:

```
1 #Aaron -> nickname
```

In [340]:

```
1 players.loc[players.name.str.startswith('Aaron')]
```

Out[340]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
15	Aaron Ramsey	Arsenal	26	CM	2	35.0	1040	7.0	
157	Aaron Lennon	Everton	30	RW	1	5.0	504	5.5	
176	Aaron Mooy	Huddersfield	26	CM	2	5.0	588	5.5	
455	Aaron Cresswell	West+Ham	27	LB	3	12.0	380	5.0	

In [341]:

```
1 players.set_index('name').filter(regex='(?i)Aaron', axis=0)
```

Out[341]:

name	club	age	position	position_cat	market_value	page_views	fpl_value	f
Aaron Ramsey	Arsenal	26	CM	2	35.00	1040	7.0	
Aaron Lennon	Everton	30	RW	1	5.00	504	5.5	
Aaron Mooy	Huddersfield	26	CM	2	5.00	588	5.5	
Rolandos Arons	Newcastle+United	21	LW	1	0.75	170	4.5	
Aaron Cresswell	West+Ham	27	LB	3	12.00	380	5.0	

In [342]:

```
1 players.loc[players.name.str.startswith('Aaron'), 'name'] = 'Ronny'
```

In [343]:

```
1 players.loc[[15,157,176,455]]
```

Out[343]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
15	Ronny	Arsenal	26	CM	2	35.0	1040	7.0	5.1%
157	Ronny	Everton	30	RW	1	5.0	504	5.5	0.2%
176	Ronny	Huddersfield	26	CM	2	5.0	588	5.5	2.5%
455	Ronny	West+Ham	27	LB	3	12.0	380	5.0	1.3%

## Adding DataFrame Cols

In [344]:

```
1 players.head()
```

Out[344]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	CM	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	RW	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	CB	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	GK	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [345]:

```
1 players.popularity
```

Out[345]:

```
0      'super popular'
1      'super popular'
2      'popular'
3      'super popular'
4      'popular'
...
460     'kind of popular'
461     'relatively unknown'
462     'relatively unknown'
463     'kind of popular'
464     'relatively unknown'
Name: popularity, Length: 465, dtype: object
```

In [346]:

```
1 'MVtoFPL' in players
```

Out[346]:

False

In [347]:

```
1 'name' in players
```

Out[347]:

True

In [348]:

```
1 players['MVtoFPL'] = 1.0
```

In [349]:

```
1 players.head()
```

Out[349]:

	<b>name</b>	<b>club</b>	<b>age</b>	<b>position</b>	<b>position_cat</b>	<b>market_value</b>	<b>page_views</b>	<b>fpl_value</b>	<b>fpl_sel</b>
<b>0</b>	Alexis Sanchez	Arsenal	28	CM	1	65.0	4329	12.0	17.10%
<b>1</b>	Mesut Ozil	Arsenal	28	RW	1	50.0	4395	9.5	5.60%
<b>2</b>	Petr Cech	Arsenal	35	CB	4	7.0	1529	5.5	5.90%
<b>3</b>	Theo Walcott	Arsenal	28	GK	1	20.0	2393	7.5	1.50%
<b>4</b>	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [350]:

```
1 players['MVtoFPL'] = players['market_value']/players['fpl_value']
```

In [351]:

```
1 players
```

Out[351]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	f
0	Alexis Sanchez	Arsenal	28	CM	1	65.0	4329	12.0	1
1	Mesut Ozil	Arsenal	28	RW	1	50.0	4395	9.5	
2	Petr Cech	Arsenal	35	CB	4	7.0	1529	5.5	
3	Theo Walcott	Arsenal	28	GK	1	20.0	2393	7.5	
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	
...	...	...	...	...	...	...	...	...	...
460	Edimilson Fernandes	West+Ham	21	CM	2	5.0	288	4.5	
461	Arthur Masuaku	West+Ham	23	LB	3	7.0	199	4.5	
462	Sam Byram	West+Ham	23	RB	3	4.5	198	4.5	
463	Ashley Fletcher	West+Ham	21	CF	1	1.0	412	4.5	
464	Diafra Sakho	West+Ham	27	CF	1	10.0	214	5.5	

465 rows × 19 columns



In [352]:

```
1 players.head()
```

Out[352]:

	name	club	age	position	position_cat	market_value	page_views	fpl_value	fpl_sel
0	Alexis Sanchez	Arsenal	28	CM	1	65.0	4329	12.0	17.10%
1	Mesut Ozil	Arsenal	28	RW	1	50.0	4395	9.5	5.60%
2	Petr Cech	Arsenal	35	CB	4	7.0	1529	5.5	5.90%
3	Theo Walcott	Arsenal	28	GK	1	20.0	2393	7.5	1.50%
4	Laurent Koscielny	Arsenal	31	CB	3	22.0	912	6.0	0.70%

In [353]:

```
1 df_mini = players.iloc[:4, 1:5]
```

In [354]:

```
1 df_mini
```

Out[354]:

	club	age	position	position_cat
0	Arsenal	28	CM	1
1	Arsenal	28	RW	1
2	Arsenal	35	CB	4
3	Arsenal	28	GK	1

In [355]:

```
1 player_names = pd.Series(['Brandon', 'Bradley', 'Ronald', 'Ronny'])
```

In [356]:

```
1 player_names
```

Out[356]:

```
0    Brandon
1    Bradley
2    Ronald
3    Ronny
dtype: object
```

In [357]:

1 df\_mini

Out[357]:

	club	age	position	position_cat
0	Arsenal	28	CM	1
1	Arsenal	28	RW	1
2	Arsenal	35	CB	4
3	Arsenal	28	GK	1

In [358]:

1 df\_mini.insert(0, 'nickname', player\_names)

In [359]:

1 df\_mini

Out[359]:

	nickname	club	age	position	position_cat
0	Brandon	Arsenal	28	CM	1
1	Bradley	Arsenal	28	RW	1
2	Ronald	Arsenal	35	CB	4
3	Ronny	Arsenal	28	GK	1

In [360]:

1 #the assign approach -&gt; this allows to add

In [361]:

1 df\_mini.assign(career\_goals=[12, 67, 179, 49])

Out[361]:

	nickname	club	age	position	position_cat	career_goals
0	Brandon	Arsenal	28	CM	1	12
1	Bradley	Arsenal	28	RW	1	67
2	Ronald	Arsenal	35	CB	4	179
3	Ronny	Arsenal	28	GK	1	49

## Adding Rows to DataFrame

In [362]:

```
1 df_mini
```

Out[362]:

	nickname	club	age	position	position_cat
0	Brandon	Arsenal	28	CM	1
1	Bradley	Arsenal	28	RW	1
2	Ronald	Arsenal	35	CB	4
3	Ronny	Arsenal	28	GK	1

In [363]:

```
1 # the append approach() method -> series, dfs, or a collection of them
```

In [364]:

```
1 cristiano = pd.Series({
2     'nickname' : 'cristiano',
3     'age' : 32,
4     'position' : 'RW',
5     'club' : 'Juventus',
6     'position_cat' : 1
7 }, name=4)
```

In [365]:

```
1 cristiano
```

Out[365]:

```
nickname      cristiano
age            32
position       RW
club          Juventus
position_cat    1
Name: 4, dtype: object
```

In [366]:

```
1 df_mini.append(cristiano)
```

Out[366]:

	nickname	club	age	position	position_cat
0	Brandon	Arsenal	28	CM	1
1	Bradley	Arsenal	28	RW	1
2	Ronald	Arsenal	35	CB	4
3	Ronny	Arsenal	28	GK	1
4	cristiano	Juventus	32	RW	1

In [367]:

```

1 other_players = pd.DataFrame({
2     'nicknames' : ['Gianluigi','lionel'],
3     'age' : [37,32],
4     'club' : ['Juventsu', 'Barcelona'],
5     'position' : ['GK', 'CF'],
6     'position_cat' : [4,2]
7 }, index=[5,6])

```

In [368]:

```
1 df_mini.append(other_players)
```

Out[368]:

	nickname	club	age	position	position_cat	nicknames
0	Brandon	Arsenal	28	CM	1	NaN
1	Bradley	Arsenal	28	RW	1	NaN
2	Ronald	Arsenal	35	CB	4	NaN
3	Ronny	Arsenal	28	GK	1	NaN
5	NaN	Juventsu	37	GK	4	Gianluigi
6	NaN	Barcelona	32	CF	2	lionel

In [369]:

```
1 df_mini = df_mini.append(other_players)
```

In [370]:

```
1 df_mini.loc[9] = 'some row value'
```

In [371]:

```
1 #adding rows to DFs is inefficient(very expensive)
```

## Skill Challenge

**1. From the players dataframe, select 4 columns and 4 rows of no particular order. Assign the resulting dataframe to df\_random.**

In [372]:

```
1 df_random = players.sample(4, axis=0).sample(4, axis=1)
```

In [373]:

```
1 df_random
```

Out[373]:

	nationality	position_cat	position	fpl_sel
47	England	2	CM	1.10%
293	England	3	CB	3.60%
28	Germany	3	CB	1.00%
121	Wales	4	GK	2.10%

In [374]:

```
1 df_random.shape
```

Out[374]:

(4, 4)

2. Extend df\_random1)vertically by adding a new row , and 2)horizontally by adding a new column.Do this as two seperate operations.

In [375]:

```
1 un = pd.Series({
2     'nationality' : 'Nigerian',
3     'age' : 32,
4     'market_value' : 20,
5     'popularity' : 'popular'
6 }, name=412)
```

In [376]:

```
1 df_random = df_random.append(un, ) #1. adding the row to the dataframe
```

In [377]:

```
1 df_random.shape
```

Out[377]:

(5, 7)

In [378]:

1 df\_random

Out[378]:

	nationality	position_cat	position	fpl_sel	age	market_value	popularity
47	England	2.0	CM	1.10%	NaN	NaN	NaN
293	England	3.0	CB	3.60%	NaN	NaN	NaN
28	Germany	3.0	CB	1.00%	NaN	NaN	NaN
121	Wales	4.0	GK	2.10%	NaN	NaN	NaN
412	Nigerian	NaN	NaN	NaN	32.0	20.0	popular

In [379]:

1 fav\_food = ['rice', 'beans', 'yam', 'bread', 'cassava']

In [380]:

1 df\_random = df\_random.assign(fav\_food = fav\_food)*# 2. adding a new column to the dataframe*

In [381]:

1 %timeit df\_random.assign(fav\_food = fav\_food)  
2 %timeit df\_random.append(un)

860 µs ± 14 µs per loop (mean ± std. dev. of 7 runs, 1000 loops each)  
 16.3 ms ± 91.2 µs per loop (mean ± std. dev. of 7 runs, 100 loops each)

Adding a column to a dataframe is faster compared to adding a row : approximately 16.5 times faster

In [ ]:

1