

In [29]: 1 `import pandas as pd`

In [30]: 1 `baby_names =pd.read_csv('https://raw.githubusercontent.com/guipsamora/pandas_`



In [31]: 1 baby_names.info

```
Out[31]: <bound method DataFrame.info of
ender State Count
0 11349 11350 Emma 2004 F AK 62
1 11350 11351 Madison 2004 F AK 48
2 11351 11352 Hannah 2004 F AK 46
3 11352 11353 Grace 2004 F AK 44
4 11353 11354 Emily 2004 F AK 41
5 11354 11355 Abigail 2004 F AK 37
6 11355 11356 Olivia 2004 F AK 33
7 11356 11357 Isabella 2004 F AK 30
8 11357 11358 Alyssa 2004 F AK 29
9 11358 11359 Sophia 2004 F AK 28
10 11359 11360 Alexis 2004 F AK 27
11 11360 11361 Elizabeth 2004 F AK 27
12 11361 11362 Hailey 2004 F AK 27
13 11362 11363 Anna 2004 F AK 26
14 11363 11364 Natalie 2004 F AK 25
15 11364 11365 Sarah 2004 F AK 25
16 11365 11366 Sydney 2004 F AK 25
17 11366 11367 Ava 2004 F AK 23
18 11367 11368 Trinity 2004 F AK 22
19 11368 11369 Haley 2004 F AK 21
20 11369 11370 Kaylee 2004 F AK 21
21 11370 11371 Taylor 2004 F AK 21
22 11371 11372 Chloe 2004 F AK 20
23 11372 11373 Ella 2004 F AK 20
24 11373 11374 Mackenzie 2004 F AK 20
25 11374 11375 Sierra 2004 F AK 19
26 11375 11376 Kayla 2004 F AK 18
27 11376 11377 Samantha 2004 F AK 18
28 11377 11378 Zoe 2004 F AK 18
29 11378 11379 Jessica 2004 F AK 17
... ..
1016365 5647396 5647397 Brooks 2014 M WY 5
1016366 5647397 5647398 Calvin 2014 M WY 5
1016367 5647398 5647399 Cameron 2014 M WY 5
1016368 5647399 5647400 Dalton 2014 M WY 5
1016369 5647400 5647401 Dawson 2014 M WY 5
1016370 5647401 5647402 Edward 2014 M WY 5
1016371 5647402 5647403 Elias 2014 M WY 5
1016372 5647403 5647404 Gage 2014 M WY 5
1016373 5647404 5647405 Hayden 2014 M WY 5
1016374 5647405 5647406 Jasper 2014 M WY 5
1016375 5647406 5647407 Jose 2014 M WY 5
1016376 5647407 5647408 Kaiden 2014 M WY 5
1016377 5647408 5647409 Kaleb 2014 M WY 5
1016378 5647409 5647410 Kasen 2014 M WY 5
1016379 5647410 5647411 Kyson 2014 M WY 5
1016380 5647411 5647412 Lukas 2014 M WY 5
1016381 5647412 5647413 Myles 2014 M WY 5
1016382 5647413 5647414 Nathaniel 2014 M WY 5
1016383 5647414 5647415 Nolan 2014 M WY 5
1016384 5647415 5647416 Oakley 2014 M WY 5
1016385 5647416 5647417 Odin 2014 M WY 5
1016386 5647417 5647418 Paxton 2014 M WY 5
```

1016387	5647418	5647419	Raymond	2014	M	WY	5
1016388	5647419	5647420	Richard	2014	M	WY	5
1016389	5647420	5647421	Rowan	2014	M	WY	5
1016390	5647421	5647422	Seth	2014	M	WY	5
1016391	5647422	5647423	Spencer	2014	M	WY	5
1016392	5647423	5647424	Tyce	2014	M	WY	5
1016393	5647424	5647425	Victor	2014	M	WY	5
1016394	5647425	5647426	Waylon	2014	M	WY	5

[1016395 rows x 7 columns]>

In [32]: 1 baby_names.head()

Out[32]:

	Unnamed: 0	Id	Name	Year	Gender	State	Count
0	11349	11350	Emma	2004	F	AK	62
1	11350	11351	Madison	2004	F	AK	48
2	11351	11352	Hannah	2004	F	AK	46
3	11352	11353	Grace	2004	F	AK	44
4	11353	11354	Emily	2004	F	AK	41

In [33]: 1 # Delete Unnamed columns

In [34]: 1 baby_names.drop(['Unnamed: 0'], axis=1)

Out[34]:

	Id	Name	Year	Gender	State	Count
0	11350	Emma	2004	F	AK	62
1	11351	Madison	2004	F	AK	48
2	11352	Hannah	2004	F	AK	46
3	11353	Grace	2004	F	AK	44
4	11354	Emily	2004	F	AK	41
5	11355	Abigail	2004	F	AK	37
6	11356	Olivia	2004	F	AK	33
7	11357	Isabella	2004	F	AK	30
8	11358	Alyssa	2004	F	AK	29
9	11359	Sophia	2004	F	AK	28
10	11360	Alexis	2004	F	AK	27

In [35]: 1 #Show the distribution of male and female

In [36]: 1 baby_names['Gender'].value_counts('F')

Out[36]: F 0.549832
M 0.450168
Name: Gender, dtype: float64

```
In [37]: 1 #Show the top 5 most preferred names
```

```
In [46]: 1 names = baby_names.groupby('Name').sum()
```

```
In [52]: 1 names.sort_values("Count", ascending = 0).head()  
2
```

Out[52]:

	Unnamed: 0	Id	Year	Count
Name				
Jacob	1665680788	1665681356	1141099	242874
Emma	1629481684	1629482250	1137085	214852
Michael	1687520717	1687521295	1161152	214405
Ethan	1660807908	1660808475	1139091	209277
Isabella	1630131220	1630131786	1137090	204798

```
In [ ]: 1 #What is the median name occurrence in the dataset
```

In [43]: 1 names[names.Count == names.Count.median()]

Out[43]:

	Unnamed: 0	Id	Year	Count
Name				
Aishani	7810519	7810526	14078	49
Alara	18841019	18841027	16079	49
Alysse	22629397	22629405	16057	49
Ameir	21780403	21780411	16086	49
Anely	4349533	4349541	16071	49
Antonina	27672241	27672250	18081	49
Aveline	7982899	7982905	12065	49
Aziah	29825399	29825407	16073	49
Baily	27406178	27406186	16064	49
Caleah	20967776	20967785	18106	49
Carlota	6971167	6971174	14077	49
Cristine	11299084	11299091	14042	49
Dahlila	8183026	8183033	14063	49
Darvin	13433465	13433473	16078	49
Deante	24229101	24229110	18064	49
Deserae	31068409	31068418	18061	49
Devean	7100053	7100057	8019	49
Elizah	13583864	13583872	16063	49
Emmaly	17488703	17488711	16075	49
Emmanuela	26771510	26771519	18074	49
Envy	13005993	13006001	16070	49
Esli	17424920	17424928	16059	49
Fay	17137611	17137619	16072	49
Gurshaan	4884923	4884930	14070	49
Hareem	16374675	16374682	14090	49
Iven	4861908	4861915	14062	49
Jaice	31447999	31448007	16098	49
Jaiyana	23805949	23805956	14068	49
Jamiracle	16247582	16247591	18091	49
Jelissa	24630708	24630716	16070	49
...
Kyndle	33887364	33887372	16082	49
Kynsley	26646466	26646473	14084	49

	Unnamed: 0	Id	Year	Count
Name				
Leylanie	4347483	4347491	16070	49
Maisha	23441757	23441764	14047	49
Malillany	17027059	17027066	14087	49
Mariann	22330704	22330712	16060	49
Marquell	18281380	18281388	16053	49
Maurilio	22999227	22999235	16058	49
Mckynzie	30075275	30075282	14068	49
Mehdi	23762219	23762227	16070	49
Nabeel	21778869	21778878	18070	49
Nalleli	17420931	17420939	16058	49
Nassir	18112860	18112868	16058	49
Nazier	31970168	31970176	16061	49
Nishant	12096440	12096448	16050	49
Rebecka	17535442	17535450	16061	49
Reghan	31154491	31154500	18073	49
Ridwan	25287418	25287426	16082	49
Riot	26477258	26477266	16104	49
Rubin	18701417	18701425	16055	49
Ryatt	25392234	25392242	16103	49
Sameera	27250232	27250241	18083	49
Sanjuanita	24554014	24554019	10035	49
Shalyn	29473992	29474001	18061	49
Skylie	27316486	27316494	16086	49
Sriram	9208196	9208203	14054	49
Trinton	32822294	32822302	16069	49
Vita	10241625	10241632	14075	49
Yoni	6439146	6439154	16060	49
Zuleima	3521366	3521373	14050	49

66 rows × 4 columns

In []: 1 *# Distribution of male and female born count by states*

```
In [45]: 1 baby_names.groupby('State')['Gender'].value_counts('F')
```

```
Out[45]: State Gender
AK      M      0.518333
        F      0.481667
AL      F      0.539870
        M      0.460130
AR      F      0.525502
        M      0.474498
AZ      F      0.572973
        M      0.427027
CA      F      0.587958
        M      0.412042
CO      F      0.554375
        M      0.445625
CT      F      0.534205
        M      0.465795
DC      F      0.504378
        M      0.495622
DE      F      0.510924
        M      0.489076
FL      F      0.562278
        M      0.437722
GA      F      0.556417
        M      0.443583
HI      M      0.521394
        F      0.478606
IA      F      0.530659
        M      0.469341
ID      F      0.504359
        M      0.495641
IL      F      0.558274
        M      0.441726
        ...
OK      F      0.539106
        M      0.460894
OR      F      0.539876
        M      0.460124
PA      F      0.552273
        M      0.447727
RI      F      0.508953
        M      0.491047
SC      F      0.535957
        M      0.464043
SD      M      0.506091
        F      0.493909
TN      F      0.552323
        M      0.447677
TX      F      0.588592
        M      0.411408
UT      F      0.536117
        M      0.463883
VA      F      0.551615
        M      0.448385
VT      M      0.536472
        F      0.463528
WA      F      0.546763
```

	M	0.453237
WI	F	0.541280
	M	0.458720
WV	F	0.535581
	M	0.464419
WY	M	0.566667
	F	0.433333

Name: Gender, Length: 102, dtype: float64

In []:

1