Read the dataset from the below link

https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_Na
(https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_N

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

url="https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_Names_right.csv"
c=pd.read_csv(url)
c
```

Out[1]:

	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
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

	Unnamed: 0	Id	Name	Year	Gender	State	Count
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 × 7 columns

Show the distribution of male and female

In [2]:
```
c.groupby(by='Gender')['Gender'].count()
```

Out[2]:
```
Gender
F    558846
M    457549
Name: Gender, dtype: int64
```

Show the top 5 most preferred names

In [3]:
```
c.groupby(by='Name')['Name'].count().sort_values(ascending=False).head()
```

Out[3]:
```
Name
Riley     1112
Avery     1080
Jordan    1073
Peyton    1064
Hayden    1049
Name: Name, dtype: int64
```

Distribution of male and female born count by states

In [4]:
```
c.groupby(['State'])['Gender'].value_counts()
```

Out[4]:

State	Gender	
AK	M	2587
	F	2404
AL	F	9878
	M	8419
AR	F	7171
	M	6475
AZ	F	14518
	M	10820
CA	F	45144
	M	31637
CO	F	11424
	M	9183
CT	F	6575
	M	5733
DC	F	3053
	M	3000
DE	F	2549
	M	2440
FL	F	25781
	M	20070
GA	F	19385
	M	15454
HI	M	3546
	F	3255
IA	F	7131
	M	6307
ID	F	4918
	M	4833
IL	F	21268
	M	16828
		...
OK	F	9519
	M	8138
OR	F	8604
	M	7333
PA	F	17480
	M	14171
RI	F	2558
	M	2468
SC	F	9465
	M	8195
SD	M	2908
	F	2838
TN	F	13063
	M	10588
TX	F	39760
	M	27791
UT	F	9515
	M	8233
VA	F	14759
	M	11997
VT	M	1618
	F	1398
WA	F	13329
	M	11049
WI	F	10549

```
        M          8940
WV      F          4305
        M          3733
WY      M          1904
        F          1456
Name: Gender, Length: 102, dtype: int64
```

Delete unnamed columns

In [5]:
```
c.columns
```

Out[5]:
```
Index(['Unnamed: 0', 'Id', 'Name', 'Year', 'Gender', 'State', 'Count'], dtype
='object')
```

In [6]:
```
c.drop(columns=['Unnamed: 0'],axis=0)
```

Out[6]:

	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
11	11361	Elizabeth	2004	F	AK	27
12	11362	Hailey	2004	F	AK	27
13	11363	Anna	2004	F	AK	26
14	11364	Natalie	2004	F	AK	25
15	11365	Sarah	2004	F	AK	25
16	11366	Sydney	2004	F	AK	25
17	11367	Ava	2004	F	AK	23
18	11368	Trinity	2004	F	AK	22
19	11369	Haley	2004	F	AK	21
20	11370	Kaylee	2004	F	AK	21
21	11371	Taylor	2004	F	AK	21
22	11372	Chloe	2004	F	AK	20
23	11373	Ella	2004	F	AK	20
24	11374	Mackenzie	2004	F	AK	20
25	11375	Sierra	2004	F	AK	19
26	11376	Kayla	2004	F	AK	18
27	11377	Samantha	2004	F	AK	18
28	11378	Zoe	2004	F	AK	18
29	11379	Jessica	2004	F	AK	17
...	...	...	...	...	...	...
1016365	5647397	Brooks	2014	M	WY	5
1016366	5647398	Calvin	2014	M	WY	5
1016367	5647399	Cameron	2014	M	WY	5
1016368	5647400	Dalton	2014	M	WY	5

	Id	Name	Year	Gender	State	Count
1016369	5647401	Dawson	2014	M	WY	5
1016370	5647402	Edward	2014	M	WY	5
1016371	5647403	Elias	2014	M	WY	5
1016372	5647404	Gage	2014	M	WY	5
1016373	5647405	Hayden	2014	M	WY	5
1016374	5647406	Jasper	2014	M	WY	5
1016375	5647407	Jose	2014	M	WY	5
1016376	5647408	Kaiden	2014	M	WY	5
1016377	5647409	Kaleb	2014	M	WY	5
1016378	5647410	Kasen	2014	M	WY	5
1016379	5647411	Kyson	2014	M	WY	5
1016380	5647412	Lukas	2014	M	WY	5
1016381	5647413	Myles	2014	M	WY	5
1016382	5647414	Nathaniel	2014	M	WY	5
1016383	5647415	Nolan	2014	M	WY	5
1016384	5647416	Oakley	2014	M	WY	5
1016385	5647417	Odin	2014	M	WY	5
1016386	5647418	Paxton	2014	M	WY	5
1016387	5647419	Raymond	2014	M	WY	5
1016388	5647420	Richard	2014	M	WY	5
1016389	5647421	Rowan	2014	M	WY	5
1016390	5647422	Seth	2014	M	WY	5
1016391	5647423	Spencer	2014	M	WY	5
1016392	5647424	Tyce	2014	M	WY	5
1016393	5647425	Victor	2014	M	WY	5
1016394	5647426	Waylon	2014	M	WY	5

1016395 rows × 6 columns

What is the median name occurence in the dataset

In [9]:
```
c.groupby(by='Name')['Name'].count().median()
```

Out[9]: 8.0

In [ ]:
```

```