

```
In [1]: # Import the necessary Libraries  
import pandas as pd
```

```
In [2]: # Import the dataset from this address.  
url = 'https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_Names_right.csv'  
  
# Assign it to a variable called baby_names  
baby_names = pd.read_csv(url, index_col = 0 , parse_dates = [0])  
baby_names.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 1016395 entries, 11349 to 5647425  
Data columns (total 6 columns):  
Id          1016395 non-null int64  
Name        1016395 non-null object  
Year        1016395 non-null int64  
Gender      1016395 non-null object  
State       1016395 non-null object  
Count       1016395 non-null int64  
dtypes: int64(3), object(3)  
memory usage: 54.3+ MB
```

In [3]: *# See the first 10 entries*
 baby_names.head(10)

Out[3]:

| | Id | Name | Year | Gender | State | Count |
|--------------|-----------|-------------|-------------|---------------|--------------|--------------|
| 11349 | 11350 | Emma | 2004 | F | AK | 62 |
| 11350 | 11351 | Madison | 2004 | F | AK | 48 |
| 11351 | 11352 | Hannah | 2004 | F | AK | 46 |
| 11352 | 11353 | Grace | 2004 | F | AK | 44 |
| 11353 | 11354 | Emily | 2004 | F | AK | 41 |
| 11354 | 11355 | Abigail | 2004 | F | AK | 37 |
| 11355 | 11356 | Olivia | 2004 | F | AK | 33 |
| 11356 | 11357 | Isabella | 2004 | F | AK | 30 |
| 11357 | 11358 | Alyssa | 2004 | F | AK | 29 |
| 11358 | 11359 | Sophia | 2004 | F | AK | 28 |

In [4]: *#Delete the column 'Id' as there are no unnamed columns.*

```
# deletes Unnamed: 0
del baby_names['Id']
baby_names.head()
```

Out[4]:

| | Name | Year | Gender | State | Count |
|--------------|-------------|-------------|---------------|--------------|--------------|
| 11349 | Emma | 2004 | F | AK | 62 |
| 11350 | Madison | 2004 | F | AK | 48 |
| 11351 | Hannah | 2004 | F | AK | 46 |
| 11352 | Grace | 2004 | F | AK | 44 |
| 11353 | Emily | 2004 | F | AK | 41 |

```
In [5]: #Show the distribution of male and female  
print(baby_names.groupby("Gender")['Gender'].size())
```

```
Gender  
F    558846  
M    457549  
Name: Gender, dtype: int64
```

```
In [6]: #Group the dataset by name and assign to names  
# group the data  
names = baby_names.groupby("Name").sum()  
  
# print the first 5 observations  
names.head()  
  
# print the size of the dataset  
print(names.shape)  
  
# sort it from the biggest value to the smallest one  
names.sort_values("Count", ascending = 0).head()
```

```
(17632, 2)
```

Out[6]:

| | Year | Count |
|----------|---------|--------|
| Name | | |
| Jacob | 1141099 | 242874 |
| Emma | 1137085 | 214852 |
| Michael | 1161152 | 214405 |
| Ethan | 1139091 | 209277 |
| Isabella | 1137090 | 204798 |

In [7]: *#4. What is the median name occurrence*
names[names.Count == names.Count.median()]

Out[7]:

| | Year | Count |
|-----------|-------|-------|
| Name | | |
| Aishani | 14078 | 49 |
| Alara | 16079 | 49 |
| Alysse | 16057 | 49 |
| Ameir | 16086 | 49 |
| Anely | 16071 | 49 |
| Antonina | 18081 | 49 |
| Aveline | 12065 | 49 |
| Aziah | 16073 | 49 |
| Baily | 16064 | 49 |
| Caleah | 18106 | 49 |
| Carlota | 14077 | 49 |
| Cristine | 14042 | 49 |
| Dahlila | 14063 | 49 |
| Darvin | 16078 | 49 |
| Deante | 18064 | 49 |
| Deserae | 18061 | 49 |
| Devean | 8019 | 49 |
| Elizah | 16063 | 49 |
| Emmaly | 16075 | 49 |
| Emmanuela | 18074 | 49 |
| Envy | 16070 | 49 |
| Esli | 16059 | 49 |

| | Year | Count |
|-----------|-------|-------|
| Name | | |
| Fay | 16072 | 49 |
| Gurshaan | 14070 | 49 |
| Hareem | 14090 | 49 |
| Iven | 14062 | 49 |
| Jaice | 16098 | 49 |
| Jaiyana | 14068 | 49 |
| Jamiracle | 18091 | 49 |
| Jelissa | 16070 | 49 |
| ... | ... | ... |
| Kyndle | 16082 | 49 |
| Kynsley | 14084 | 49 |
| Leylanie | 16070 | 49 |
| Maisha | 14047 | 49 |
| Malillany | 14087 | 49 |
| Mariann | 16060 | 49 |
| Marquell | 16053 | 49 |
| Maurilio | 16058 | 49 |
| Mckynzie | 14068 | 49 |
| Mehdi | 16070 | 49 |
| Nabeel | 18070 | 49 |
| Nalleli | 16058 | 49 |
| Nassir | 16058 | 49 |
| Nazier | 16061 | 49 |

| | Year | Count |
|------------|-------|-------|
| Name | | |
| Nishant | 16050 | 49 |
| Rebecka | 16061 | 49 |
| Reghan | 18073 | 49 |
| Ridwan | 16082 | 49 |
| Riot | 16104 | 49 |
| Rubin | 16055 | 49 |
| Ryatt | 16103 | 49 |
| Sameera | 18083 | 49 |
| Sanjuanita | 10035 | 49 |
| Shalyn | 18061 | 49 |
| Skylie | 16086 | 49 |
| Sriram | 14054 | 49 |
| Trinton | 16069 | 49 |
| Vita | 14075 | 49 |
| Yoni | 16060 | 49 |
| Zuleima | 14050 | 49 |

66 rows × 2 columns

```
In [8]: #5. Distribution of male and female born count by states  
state_babies = baby_names.groupby(["State", "Gender"]).size().sort_values(ascending=False).reset_index(name=  
'Count')  
state_babies
```


Out[8]:

| | State | Gender | Count |
|----|-------|--------|-------|
| 0 | CA | F | 45144 |
| 1 | TX | F | 39760 |
| 2 | CA | M | 31637 |
| 3 | NY | F | 28158 |
| 4 | TX | M | 27791 |
| 5 | FL | F | 25781 |
| 6 | NY | M | 22585 |
| 7 | IL | F | 21268 |
| 8 | FL | M | 20070 |
| 9 | GA | F | 19385 |
| 10 | OH | F | 18143 |
| 11 | PA | F | 17480 |
| 12 | NC | F | 17357 |
| 13 | IL | M | 16828 |
| 14 | MI | F | 16038 |
| 15 | GA | M | 15454 |
| 16 | NJ | F | 15041 |
| 17 | VA | F | 14759 |
| 18 | AZ | F | 14518 |
| 19 | OH | M | 14318 |
| 20 | PA | M | 14171 |
| 21 | NC | M | 13530 |
| 22 | WA | F | 13329 |

| | State | Gender | Count |
|-----|-------|--------|-------|
| 23 | MI | M | 13243 |
| 24 | TN | F | 13063 |
| 25 | IN | F | 13056 |
| 26 | NJ | M | 12274 |
| 27 | VA | M | 11997 |
| 28 | MO | F | 11948 |
| 29 | CO | F | 11424 |
| ... | ... | ... | ... |
| 72 | NE | M | 5029 |
| 73 | NM | M | 4966 |
| 74 | ID | F | 4918 |
| 75 | ID | M | 4833 |
| 76 | WV | F | 4305 |
| 77 | WV | M | 3733 |
| 78 | HI | M | 3546 |
| 79 | HI | F | 3255 |
| 80 | DC | F | 3053 |
| 81 | DC | M | 3000 |
| 82 | MT | M | 2986 |
| 83 | ME | F | 2976 |
| 84 | NH | F | 2957 |
| 85 | SD | M | 2908 |
| 86 | SD | F | 2838 |
| 87 | ME | M | 2777 |

| | State | Gender | Count |
|-----|-------|--------|-------|
| 88 | MT | F | 2690 |
| 89 | NH | M | 2659 |
| 90 | AK | M | 2587 |
| 91 | ND | M | 2581 |
| 92 | RI | F | 2558 |
| 93 | DE | F | 2549 |
| 94 | RI | M | 2468 |
| 95 | DE | M | 2440 |
| 96 | AK | F | 2404 |
| 97 | ND | F | 2399 |
| 98 | WY | M | 1904 |
| 99 | VT | M | 1618 |
| 100 | WY | F | 1456 |
| 101 | VT | F | 1398 |

102 rows × 3 columns