In [1]:

```
"""
Read the dataset from the below link
https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/
Questions:
1. Delete unnamed columns
2. Show the distribution of male and female
3. Show the top 5 most preferred names
4. What is the median name occurence in the dataset
5. Distribution of male and female born count by states
"""


import pandas as pd
import numpy as np
dataset=pd.read_csv("https://raw.githubusercontent.com/guipsamora/pandas_exerc

dataset.info()
dataset.drop(labels="Unnamed: 0",axis=1,inplace=True) #Deleting unnamed columr
print("\n ******************** \n Deleting unnamed columns")
dataset.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1016395 entries, 0 to 1016394
Data columns (total 7 columns):
Unnamed: 0    1016395 non-null int64
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(4), object(3)
memory usage: 54.3+ MB

 ********************
 Deleting unnamed columns
```

Out[1]:

	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

In [2]:
```
print("distribution of male and female")
dataset.Gender.value_counts()
```

```
distribution of male and female
```

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

In [3]:
```
print("top 5 most preferred names")
dataset.Name.value_counts()[0:5]
```

```
top 5 most preferred names
```

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

In [4]:
```
medName=int(dataset["Id"].median())
MedianName=dataset[dataset["Id"]==medName]["Name"]
print("Median name occurence in the dataset with index and name is {}".format(
```

```
Median name occurence in the dataset with index and name is 508197    Kasey
Name: Name, dtype: object
```

In [5]:
```
#Distribution of male and female born count by states
bystate=dataset.groupby("State")
print("Distribution of male and female born count by states")
bystate.Gender.value_counts()
```

```
Distribution of male and female born count by states
```

Out[5]:
```
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
```

In [ ]:
```
1
```