**Session 9: Assignment 1**

Problem Statement

Read the dataset from the below link

https://raw.githubusercontent.com/guipsamora/pandas\_exercises/master/06\_Stats/US

\_Baby\_Names/US\_Baby\_Names\_right.csv

Question 1) Delete unnamed columns

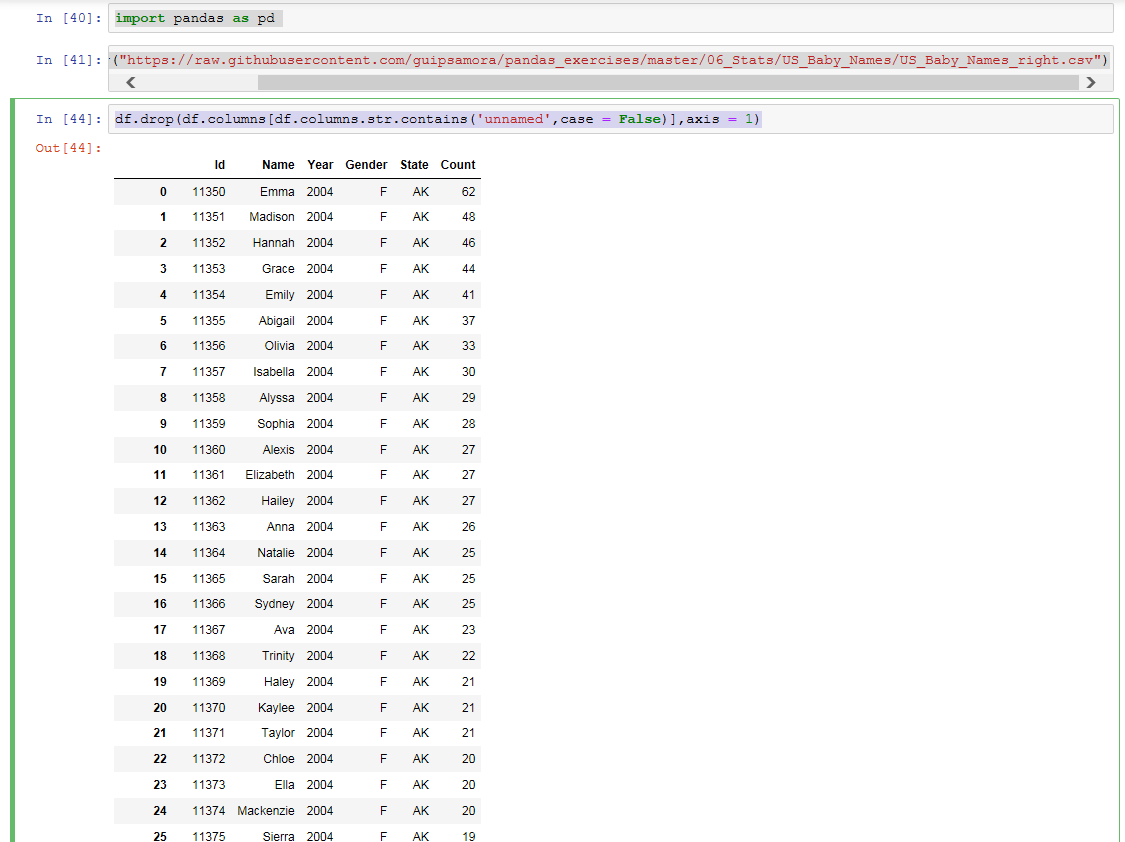
Code :

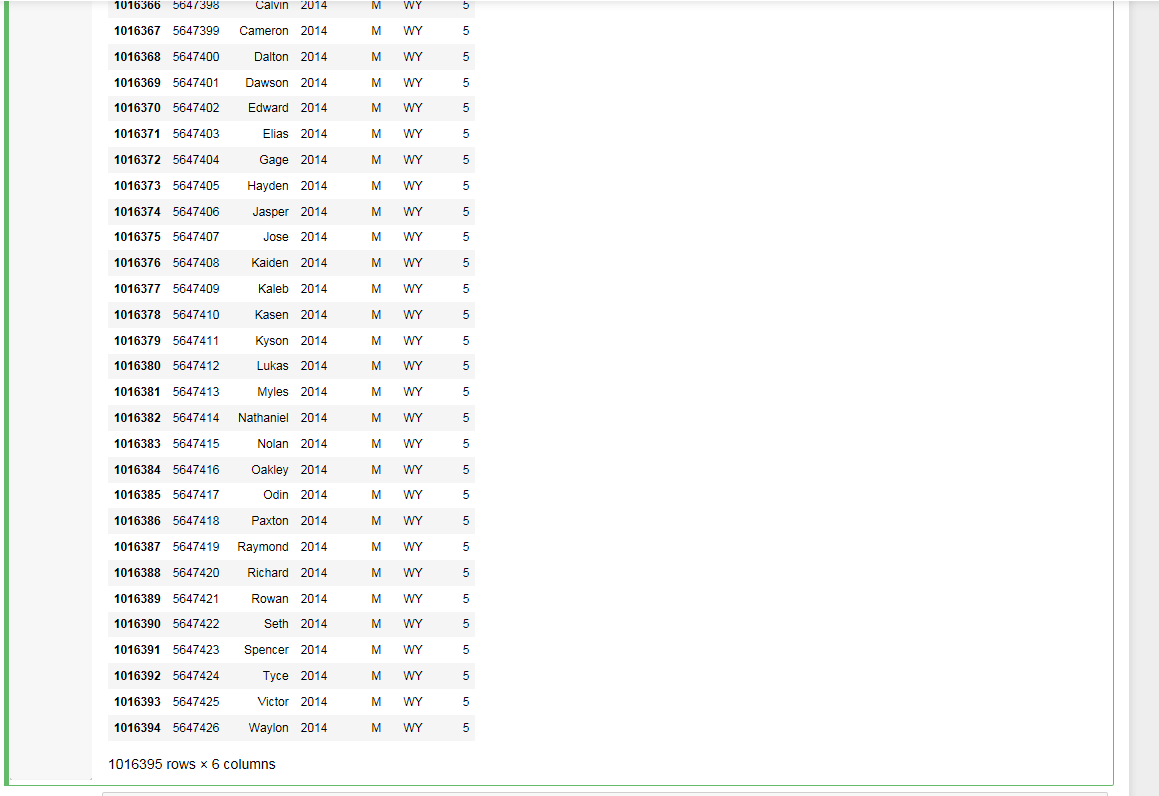
import pandas as pd

df = pd.read\_csv("https://raw.githubusercontent.com/guipsamora/pandas\_exercises/master/06\_Stats/US\_Baby\_Names/US\_Baby\_Names\_right.csv")

df.drop(df.columns[df.columns.str.contains('unnamed',case = False)],axis = 1)

Output:





Question 2) Show the distribution of male and female

Code:

import pandas as pd

import matplotlib.pyplot as plt

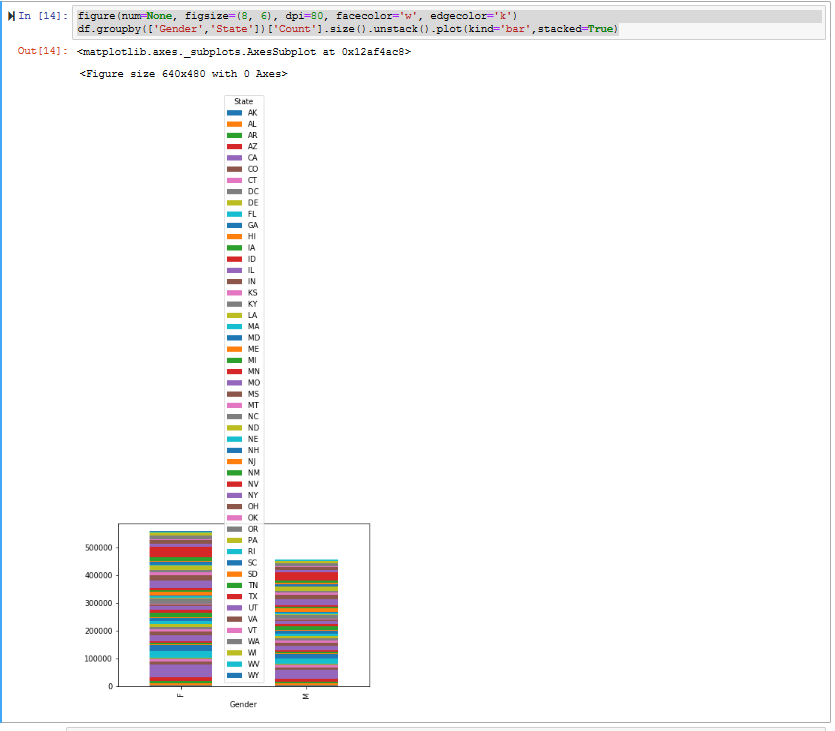
from matplotlib.pyplot import figure

df = pd.read\_csv("https://raw.githubusercontent.com/guipsamora/pandas\_exercises/master/06\_Stats/US\_Baby\_Names/US\_Baby\_Names\_right.csv")

figure(num=None, figsize=(8, 6), dpi=80, facecolor='w', edgecolor='k')

df.groupby(['Gender','State'])['Count'].size().unstack().plot(kind='bar',stacked=True)

Output:



Question 3) Show the top 5 most preferred names

Code:

import pandas as pd

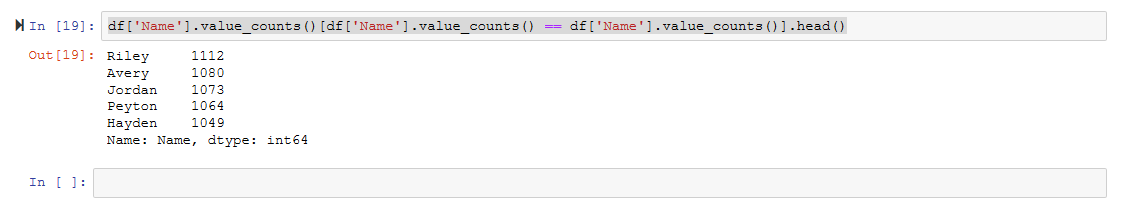
import matplotlib.pyplot as plt

from matplotlib.pyplot import figure

df = pd.read\_csv("https://raw.githubusercontent.com/guipsamora/pandas\_exercises/master/06\_Stats/US\_Baby\_Names/US\_Baby\_Names\_right.csv")

df['Name'].value\_counts()[df['Name'].value\_counts() == df['Name'].value\_counts()].head()

Output:



Question 4) What is the median name occurence in the dataset

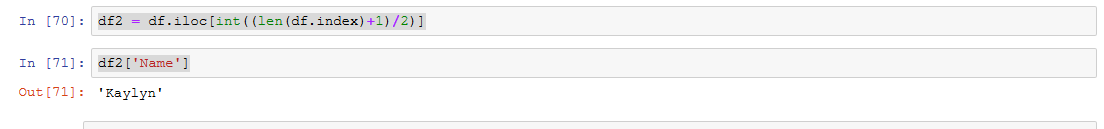
Code:

import pandas as pd

df = pd.read\_csv("https://raw.githubusercontent.com/guipsamora/pandas\_exercises/master/06\_Stats/US\_Baby\_Names/US\_Baby\_Names\_right.csv")

df2 = df.iloc[int((len(df.index)+1)/2)]

df2['Name']



Question 5) Distribution of male and female born count by states

Code:

counts = df.groupby(['State','Gender']).count()

counts

