#Data Exploration  
import pandas as pd

#create data\_frame of array values  
df = pd.DataFrame({  
 'name':['matt','lisa','richard','john','julia','jane','marlon'],  
 'age':[23,78,22,19,45,33,20],  
 'gender':['M','F','M','M','M','F','M'],  
 'state':['DC','CO','DE','VA','MD','DE','NY'],  
 'years\_of\_service':[10,0,2,0,2,1,5],  
 'iq':[300,100,110,200,300, 10, 40]  
})

# BEGIN extract of 25% sample data

rows = df.sample(frac = .25) #validate first to check if sample is 0.25 times data or not if(0.25\*(len(df))==len(rows)): print(len(df), len(rows))

# Display Sample Percentage

print (‘sample of 25%’),rows #END extract a 25% sample of data

#Categorical Variable Splitting  
groupby\_gender = df.groupby('gender')  
for gender, value in groupby\_gender['iq']:  
 print((gender,value.mean()))

('F', 55.0)  
('M', 190.0)

#Find the Summation of all ages in the data  
SumofAge=df['age'].sum()  
print ('Sum of Ages'), SumofAge  
MeanAge = df['age'].mean()  
print ('Average Ages'), MeanAge  
# Find the mean of all columns  
print ('Means of each column'), df.mean(axis=0)  
#Describe the Data  
print (df['iq'].describe())

Sum of Ages  
Average Ages  
Means of each column  
count 7.000000  
mean 151.428571  
std 117.817454  
min 10.000000  
25% 70.000000  
50% 110.000000  
75% 250.000000  
max 300.000000  
Name: iq, dtype: float64