MIS500 Module 1, Critical Thinking Assignment, Option 1

Janice Gordon

Colorado State University – Global Campus

August 16, 2019

**Code Copied from PyCharm**

*#BEGIN*

*""" Program Begin HERE*

*Some data Exploration using Python. Assuming that all the needed packages*

*are already install for your IDE to find them.*

*"""*

*#############################################################*

*#Program name - Data Exploration*

*#input - NONE*

*#output - Some Exploration statistics*

*###############################################################*

**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 a 25% sample of 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*

*############################################################*

*# BEGIN Split categorical variables by gender, Sum, Mean, count,*

*# and describe on the data*

*############################################################*

*#Categorical Variables splitting*

groupby\_gender = df.groupby(**'gender'**)

**for** gender, value **in** groupby\_gender[**'iq'**]:

**print**((gender, value.mean()))

*# 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*

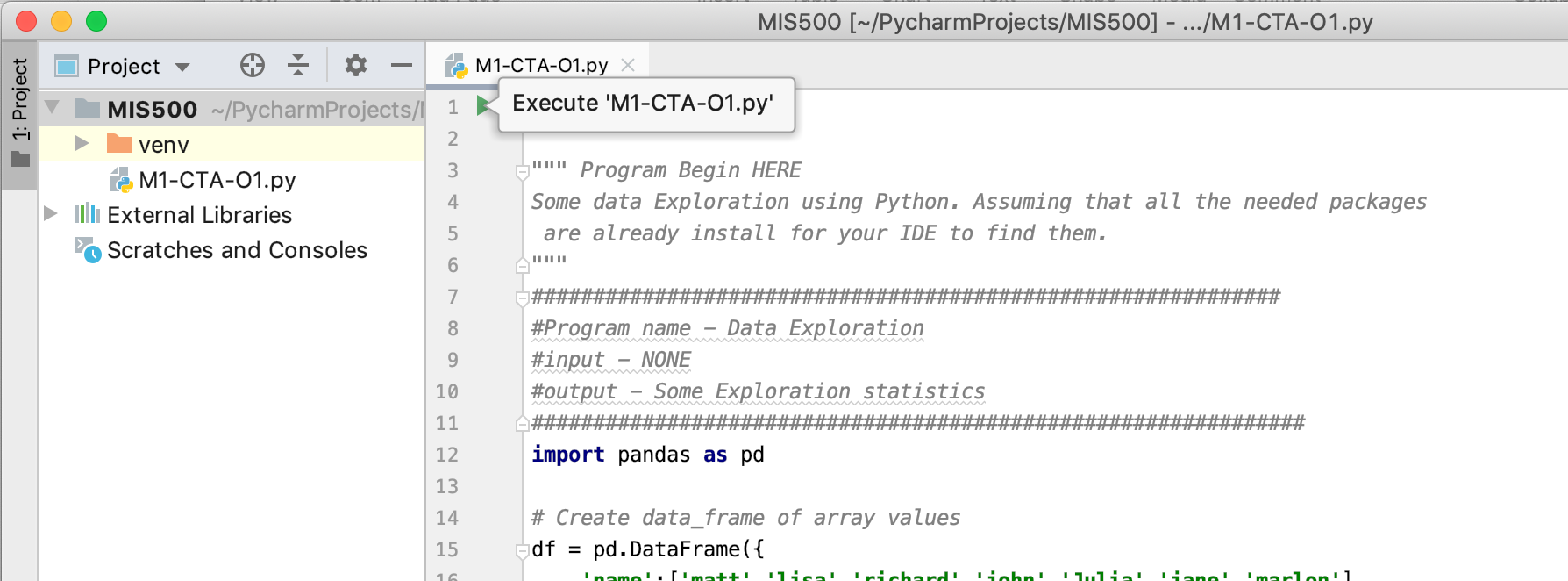
**print** (**'Means of each column'**, df.mean(axis=0))

*# Describe the Data*

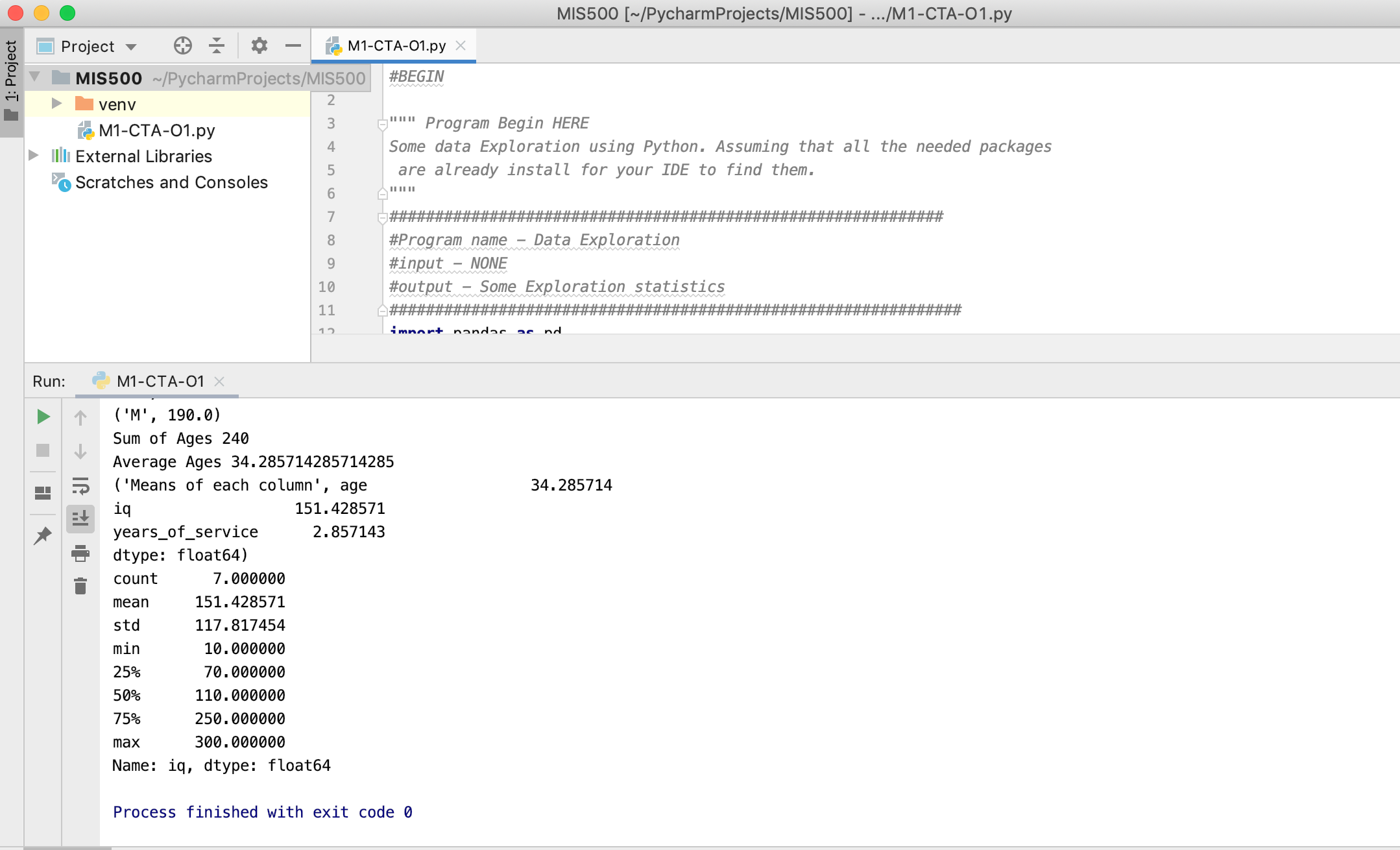
**print** df[**'iq'**].describe()

*#END*

**Screenshots of Running the Code in PyCharm**

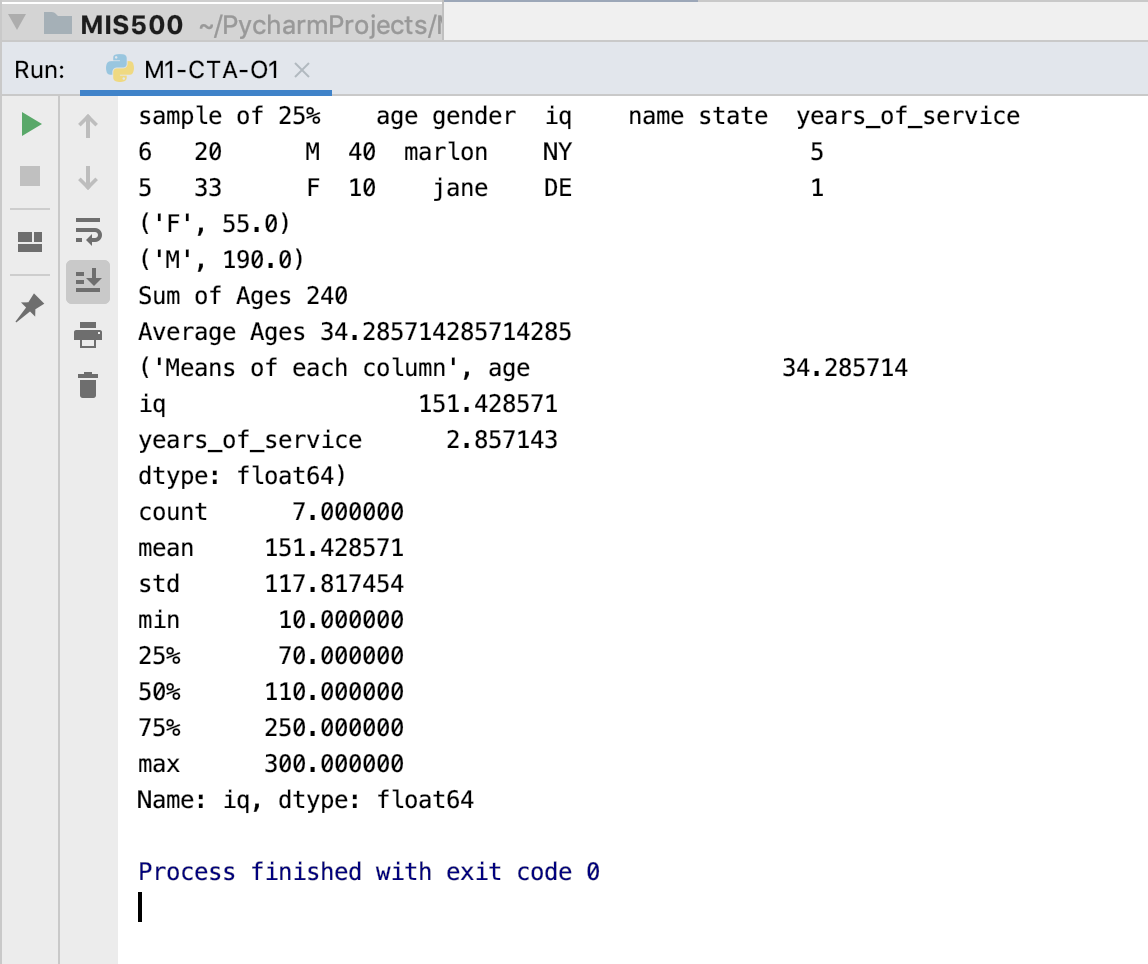
Figure 1 shows the start of code execution and Figure 2 shows the PyCharm interface just after pressing the execute button.****

*Figure 1.* Screenshot of pressing execute button.

****

*Figure 2.* Screenshot of code run completion.

**Screenshot of Code Execution Results in PyCharm**

The results of the code execution in PyCharm are shown in Figure 3.

*Figure 3.* Code results output.