PyDev console: starting.

Python 3.6.4 (v3.6.4:d48ecebad5, Dec 18 2017, 21:07:28)

[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin

#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

sample of 25% name age gender state years\_of\_service iq

4 Julia 45 M MD 2 300

6 marlon 20 M NY 5 40

('F', 55.0)

('M', 190.0)

Sum of Ages 240

Average Ages 34.285714285714285

Means of each column age 34.285714

years\_of\_service 2.857143

iq 151.428571

dtype: float64

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