Critical Thinking - Module 1: Option #1

Taylor Weese

MIS500-1 - Foundations of Data Analytics

Colorado State University - Global Campus

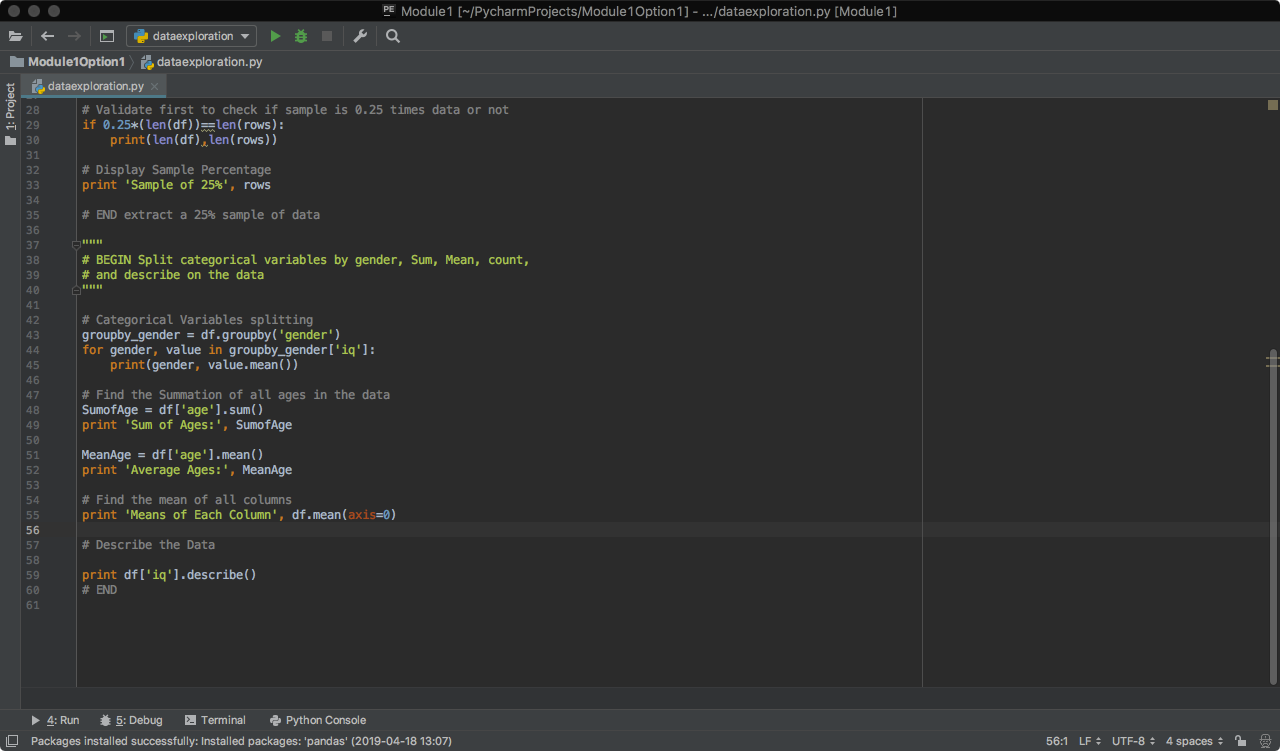
Dwight Davis

April 21, 2019

1. The code (copied from PyCharm Edu)

import pandas as pd  
  
# 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  
"""  
  
# Create data\_frame of array values  
df = pd.DataFrame({'name': ['matt', 'lisa', 'richard', 'john', 'Julia', 'jane', 'marlon'],  
 'gender': ['M', 'F', 'M', 'M', 'M', 'F', 'M'], 'age': [23, 78, 22, 19, 45, 33, 20],  
 '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  
"""  
  
# Set 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

1. Screenshot(s) of PyCharm Edu executing the program



1. Screenshot of the output of the program

