How to run Final_Project_Code_IS590PR.py file

These instructions are with respect to PyCharm as the IDE but we believe that they should be relevant to other IDE's as well.

1. Download the file from the GitHub repository - https://raw.githubusercontent.com/rahulrohri/final_project_2020Sp/master/Final_Project_Code_IS590PR.py
2. Download the necessary datasets from the google drive link https://drive.google.com/open?id=1g_StaWiaWQyNjNOu3wlFKG2dsIJZjyjF
3. Open PyCharm and load this file in your environment.
4. Update the corresponding file paths in the program code (lines 47 – 50)
5. Click on the run command that will initiate the process of running the Final_Project_Code_IS590PR.py file.
6. Next you will be shown the following message in the run terminal assuming you have loaded the correct files and not made any changes to the base data.
   The columns necessary for analysis are present in the EMS data file
   The columns necessary for analysis are present in the Complaints data file
   The columns necessary for analysis are present in the Arrests data file
7. Since our program is designed in a way to input only the necessary columns for analysis and save memory and RAM , you will be prompted 5 times to input the relevant columns. Click in the run terminal section when prompter and perform the following:

   a. In the First prompt , enter the following column names
      enter your column name 1 and press enter: CMPLNT_NUM
      enter your column name 2 and press enter: CMPLNT_FR_DT
      enter your column name 3 and press enter: BORO_NM

   b. In the Second prompt , enter the following column names
      enter your column name 1 and press enter: ARREST_BORO
      enter your column name 2 and press enter: ARREST_DATE
      enter your column name 3 and press enter: ARREST_KEY

   c. In the Third prompt , enter the following column names
      enter your column name 1 and press enter: VIC_RACE
      enter your column name 2 and press enter: OFNS_DESC
      enter your column name 3 and press enter: CMPLNT_NUM
      After a few seconds you will see a message on the screen that says "The correlation between Population Density and the crime per capita is : 0.7275769206466038"

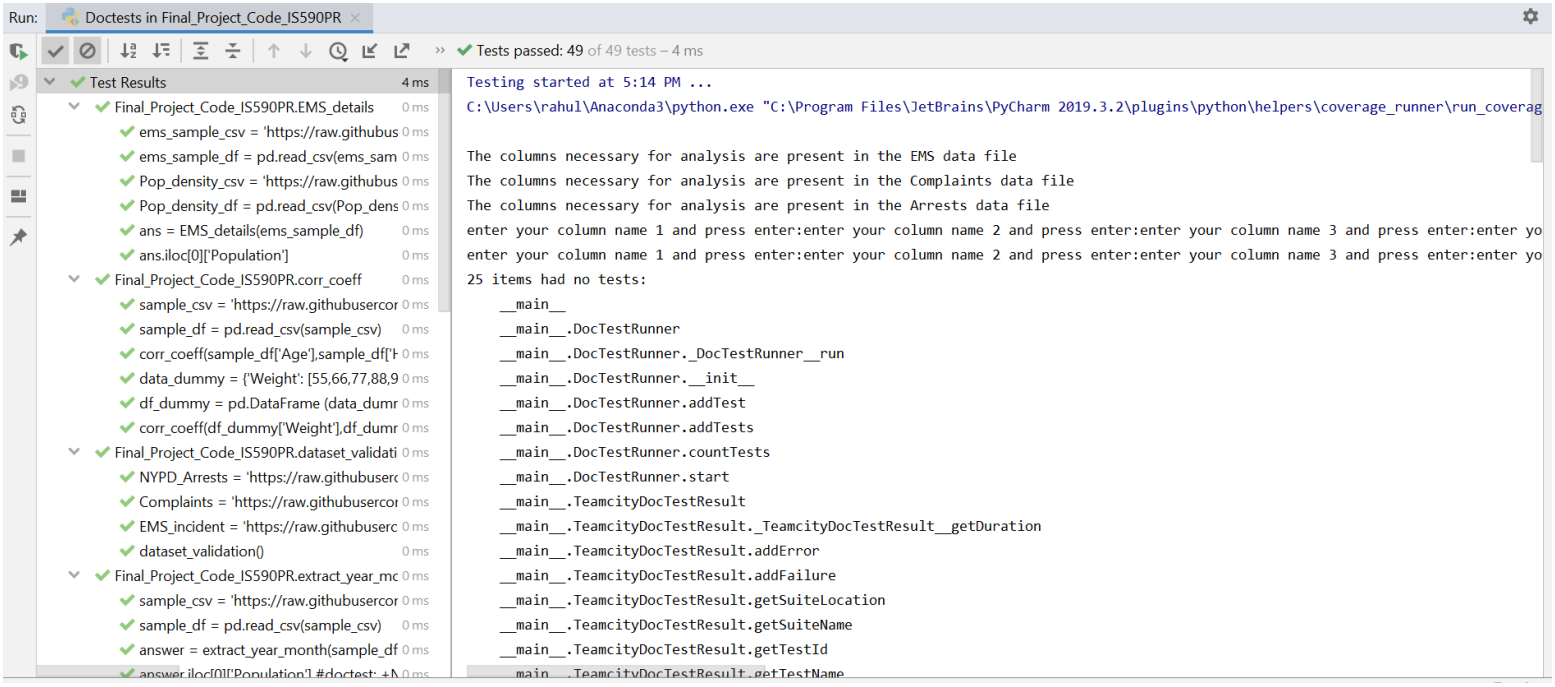
   d. In the Fourth prompt , enter the following column names
      enter your column name 1 and press enter: OFNS_DESC
      enter your column name 2 and press enter: BORO_NM
      enter your column name 3 and press enter: CMPLNT_NUM

e. In the Fifth prompt , enter the following column names
enter your column name 1 and press enter: INCIDENT_RESPONSE_SECONDS_QY
enter your column name 2 and press enter: INCIDENT_DATETIME
enter your column name 3 and press enter: BOROUGH

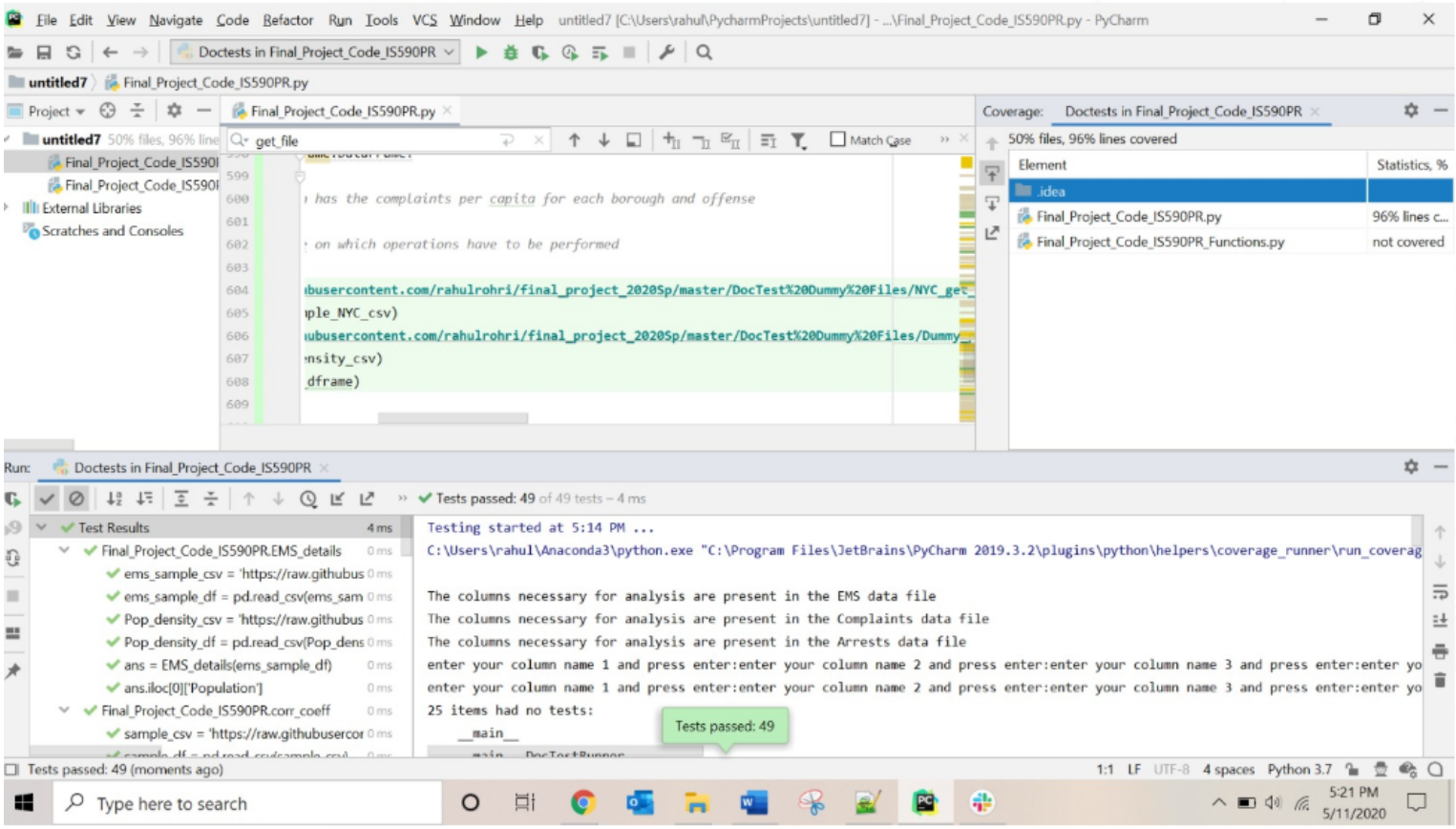
8. After this wait for a few minutes and the output will be shown on screen. The message will be "The correlation coefficient between Population Density and the Incident response time is : 0.4639141825587108"

You should also get some plots on your screen assuming the packages are installed in your environment.

Additional Note – if you choose to run the doctests as well, then after all the above 8 steps are completed and run successfully, the doctests will run and at a point for the **get_file function** it will be waiting for the input from you. Since this is not a jupyter notebook , the input may not be visible . Just click on the run terminal Screen and type "AircraftHex" and press "enter key". The type "SessionID" and press "enter key". The reason is that we have fed a dummy csv file to check if the function is running and thus it is waiting for the user to input the 2 necessary columns for analysis. After that all doctests should pass and you should get a screen similar to the image below



If running with coverage feature of pycharm, then you should get something similar to the below image



End of program