IT 166 Assignment 4 (30 points)

Due: October 18 at 11:55 PM

Data file: crime\_incident\_data2017.csv

1. Read data file using Pandas library
2. Display first 5 and last 5 row of your data
3. Display total number of rows and columns using Pandas

Before dropping Null values: (-, -)

1. Copy df to df1 and drop all missing values from df1
2. Display total number of rows and columns using

After dropping Null values: (-, -)

1. Refer back to df DataFrame and treat missing values and provide your justification about your strategy.
2. We want to know the earliest and latest report date, based on the data you have.

Hints: Example below

The earliest report date is 2017-01-01 00:00:00

The latest report date is 2017-08-31 00:00:00

1. We are also interested in knowing the best and worst day in terms of the number of reports, tell the minimum and maximum number of cases reported.
2. We are interested to know the number of crimes reported on July 4, 2017 from the data that you have.

Hints:

1. First filter the crimes based on the date and Offense Type and Offense Count

2. Groupby the filtered crimes based on Offense Type and take the sum()

1. Display All offense categories
2. Now we are interested to see the counts of offenses under each category
3. Drop or delete ‘Offense Type’ column from df DataFrame.

Requirements:

1. Have all the code saved in assignment4.py and upload to Reggienet