Overview:

This final project is a basic data science project that use much of the knowledge gained so far. You will also be introduced to data visualization using python modules.

Description:

You are to analyze FBI crime statistics for 2016. The data consists of crime statistics for each city (and the District of Columbia) itemized by type of crime.

The data is in the json file FBI_CrimeData_2016.json. Once the data in this file is loaded into a variable, the resultant data structure is a list where each element is a dictionary. Each dictionary consists of the following key-value pairs;

Region:	The region in which the State is located. The values are Midwest, Northeast, South, and West.
State:	The State name including the District of Columbia.
City:	The city in which the crimes occurred.
Population:	The city population.
Murder:	The number of murders that occurred in the city.
Rape:	The number of rapes that occurred in the city.
Robbery:	The number of robberies that occurred in the city.
Assault:	The number of assaults that occurred in the city.
Burglary:	The number of burglaries that occurred in the city.
Theft:	The number of thefts, exclusive of vehicle, that occurred in the city.
Vehicle_Theft:	The number of vehicle thefts that occurred in the city

For purposes of the project, the crimes murder through assault are considered violent crimes. The remaining three categories are considered non-violent crimes.

Objective:

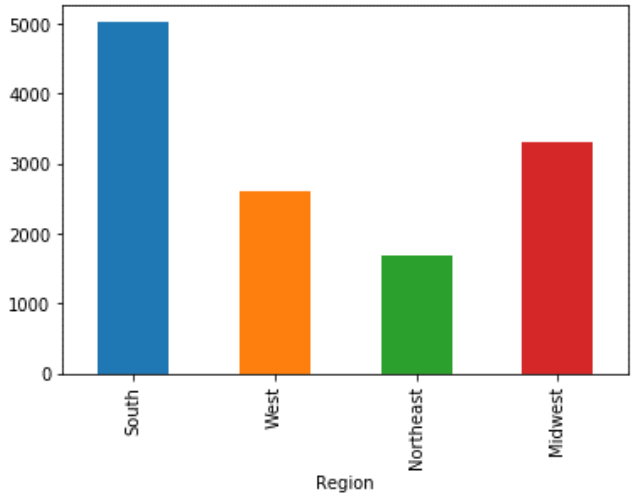
From the crime statistics data, you are to produce the following output:

1. A bar chart showing the number of “Murders by Region”.
2. A bar chart showing the number of “Violent Crimes by Region”.
3. A bar chart showing the number of “Non-violent Crimes by Region”.
4. A report which is tabular list showing the total violent crimes for each State and the distance from the national mean.

The following are images of the required output:

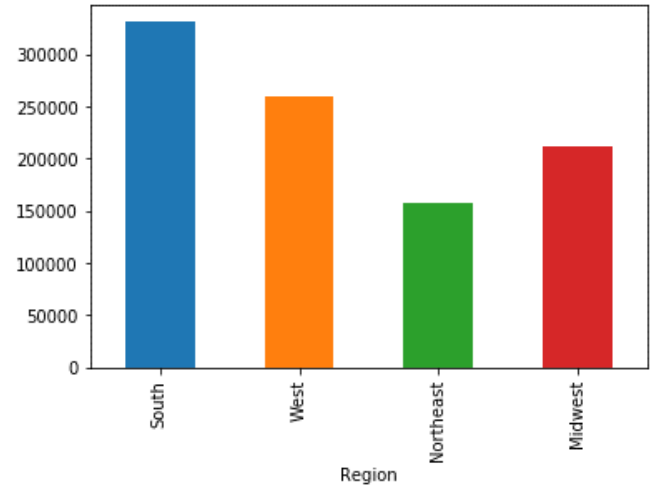
Murder by Region

	Incidents	Region
0	5018	South
1	2605	West
2	1674	Northeast
3	3304	Midwest



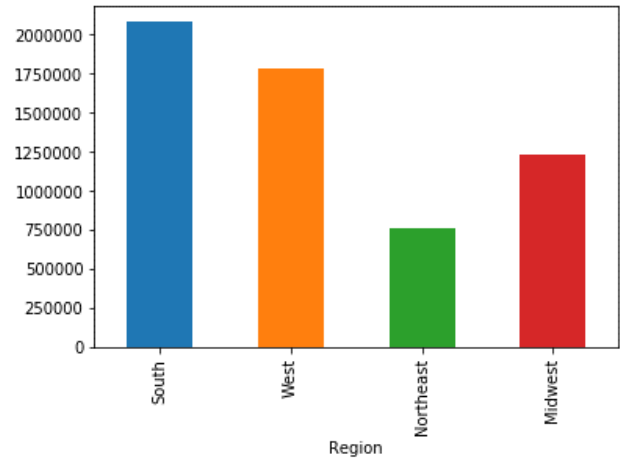
Violent Crimes by Region

	Incidents	Region
0	331011	South
1	259401	West
2	156891	Northeast
3	211854	Midwest



Non-violent Crimes by Region

	Incidents	Region
0	2080097	South
1	1781665	West
2	760200	Northeast
3	1226462	Midwest



National Average Violent Crime 18807

State	Crimes	Distance from Mean
ALABAMA	15,350	-3,457
ALASKA	4,697	-14,110
ARIZONA	23,856	5,049
ARKANSAS	12,048	-6,759
CALIFORNIA	145,205	126,398
COLORADO	15,533	-3,274
CONNECTICUT	7,754	-11,053
DELAWARE	2,359	-16,448
DISTRICT OF COLUMBIA	7,711	-11,096
FLORIDA	50,540	31,733

Truncated listed of States

Approach:

Import the modules below. Note, the last line allows charts to be displayed inline in Jupyter Notebook.

```
import json
import pandas as pd
```

Open the data file and load the json data into the variable crime_list.

Create the data structures required to produce the three bar charts. You will need to reduce crime_list into three dictionaries: murder_by_region, violent_by_region, and nonviolent_by_region. Essentially for each of the three dictionaries, iterate over crime_list and creating the appropriate dictionary using and accumulation pattern.

Print the resultant dictionaries. The contents should be identical to that shown below.

```
1 print(murder_by_region)
2 print(violent_by_region)
3 print(nonviolent_by_region)
```

```
{'South': 5018, 'West': 2605, 'Northeast': 1674, 'Midwest': 3304}
{'South': 331011, 'West': 259401, 'Northeast': 156891, 'Midwest': 211854}
{'South': 2080097, 'West': 1781665, 'Northeast': 760200, 'Midwest': 1226462}
```

Note, to produce the three above dictionaries, I chose an approach using three functions accum_crime, accum_violent_crime, and accum_nonviolent_crime. The function accum_crime has three parameters:

key: The entity upon which I'm accumulating "Region" or "State".

crime: The crime on which I'm accumulating, e.g., "Murder"

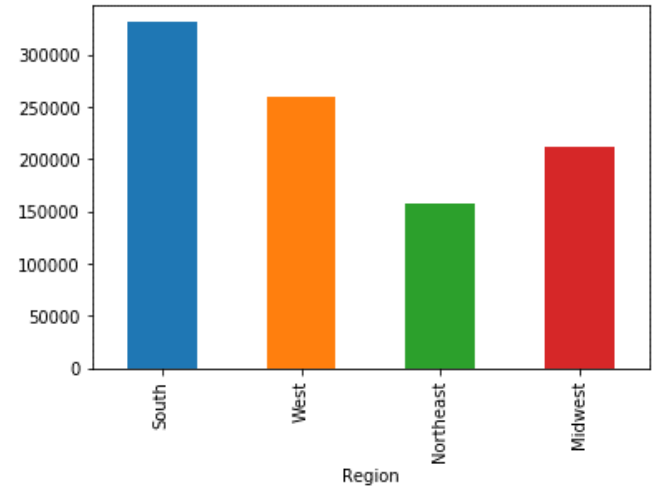
crime_list: The list containing dictionaries for each city.

Once I had a working accum_crime function which produced the murder dictionary, I copied that function as the accum_violent_crime, removed the crime parameter, and made the appropriate changes to the measurement to accumulate violent crimes. Once this function was working, I copied as accum_nonviolent_crime function and made the appropriate changes to the measurement. I recommend this approach since you will need to do a State violent crime dictionary in the next step.

The smart approach to charting the data is to create function that accepts a crime dictionary and plots the bar chart. Using this approach allows you to plot each dictionary using the function. The chart below was created by calling the function with the violent_by_region dictionary.

```
Violent Crimes by Region
   Incidents     Region
0     331011      South
1     259401       West
2     156891  Northeast
3     211854    Midwest
```

This is produced by printing the dataframe before plotting the chart.



To produce the measurements for the total violent crimes report, you must first calculate the national violent crime mean (average). To do this, you must total by the violent crimes and divide by the number of States (which includes Washington DC). There is a total of 51 States (including DC). However, you must calculate this number and not hard code it.

If you did not follow, my previous recommendation, good luck. If you did follow the recommendation, call the accum_violent_crime function and summarize by “State”. Store the returned dictionary in violent_by_state. Iterate through this variable to produce the tabular list below. Use the string format() method to produce a nice tabular alignment.

```
National Average Violent Crime 18807

State                Crimes  Distance from Mean
ALABAMA               15,350   -3,457
ALASKA                 4,697  -14,110
ARIZONA               23,856    5,049
ARKANSAS              12,048   -6,759
CALIFORNIA           145,205  126,398
COLORADO              15,533   -3,274
CONNECTICUT            7,754  -11,053
DELAWARE               2,359  -16,448
DISTRICT OF COLUMBIA   7,711  -11,096
FLORIDA               50,540   31,733
```

An additional 0 to 25 extra credit points are available for any additional charts or reports that you may choose to produce. The number of points awarded is at the discretion of the instruction. The more innovative the work, the more points are awarded.

Once the project is complete and the results verified, upload the notebook to Blackboard.