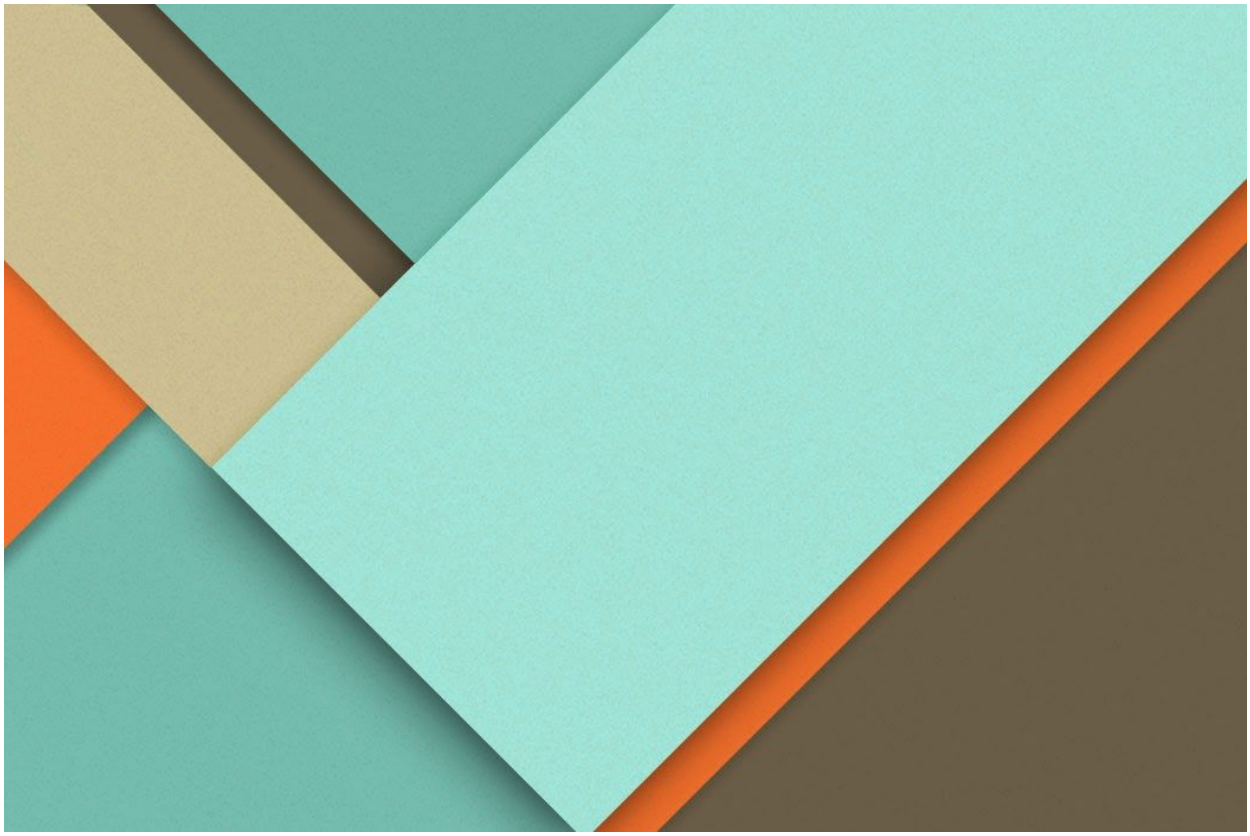


Explore Weather Trends

02/03/2020

Ankit Chaudhary

Username:- ankitchaudhary2303@gmail.com

Overview

I have been provided with temperature database from the portal from where I have extracted data related to Global temperature and City Temperature. I have analyzed the temperature of city (NEW YORK) with global temperature by extracting data from database.

Goals

1. Extraction of data from database and export to csv
2. Making a visualization chart based on extracted data.
3. Observations based on chart.

Tools Used:

1. SQL: Extract data from database
2. Python: for analyzing the data
3. Jupyter Notebook: for writing python codes and making observations.
4. Google Sheets

I. Step 1:- Extraction of dataset

=>

select * from city_list where country = 'United States'

=> extract global data

ALTER TABLE city_data RENAME COLUMN avg_temp to cavt

ALTER TABLE global_data RENAME COLUMN avg_temp to gavt

select g.year, g.gavt, c.cavt

from global_data as g, city_data as c

where g.year = c.year AND city = 'New York'

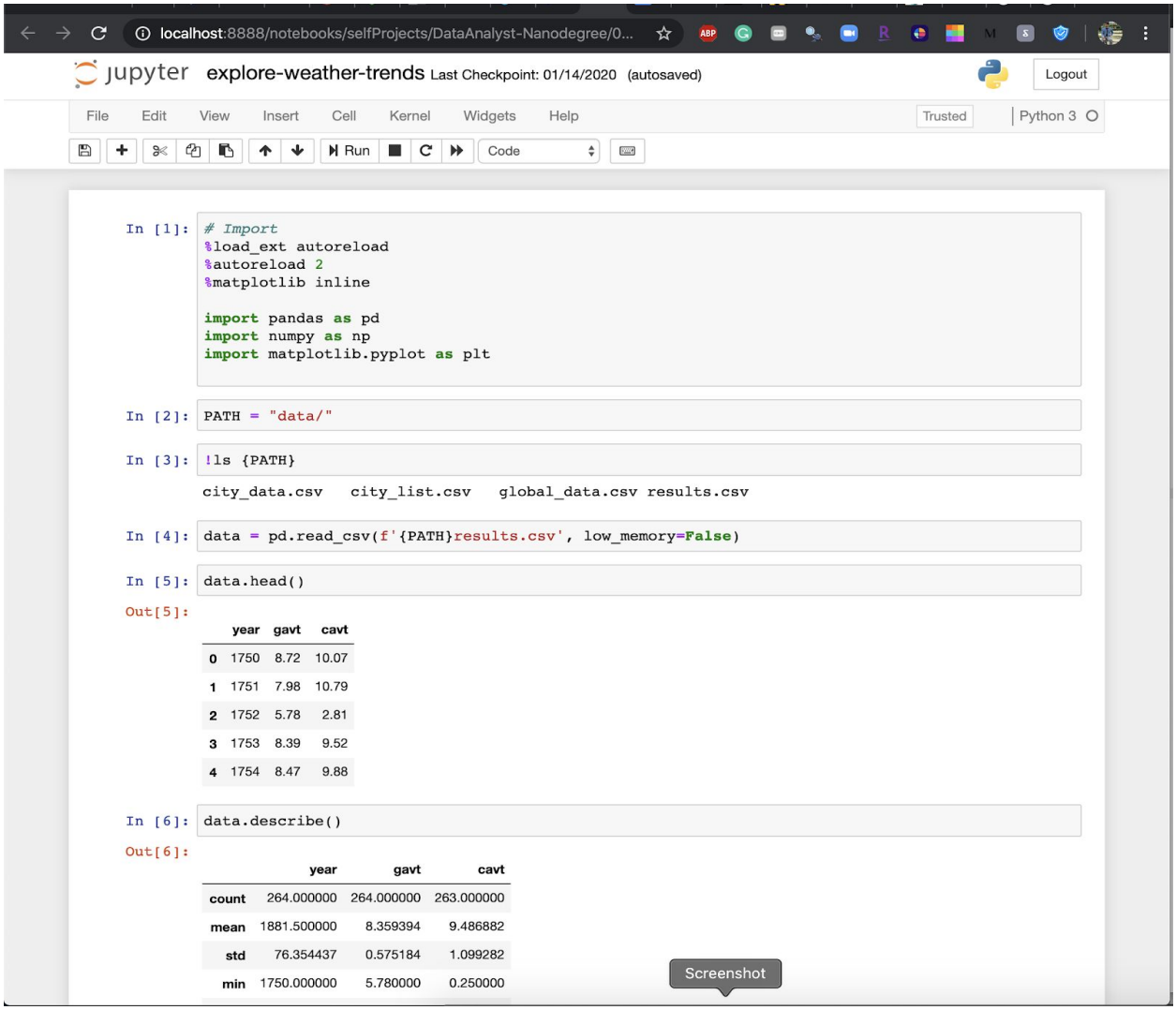
select g.year, g.gavt, c.cavt

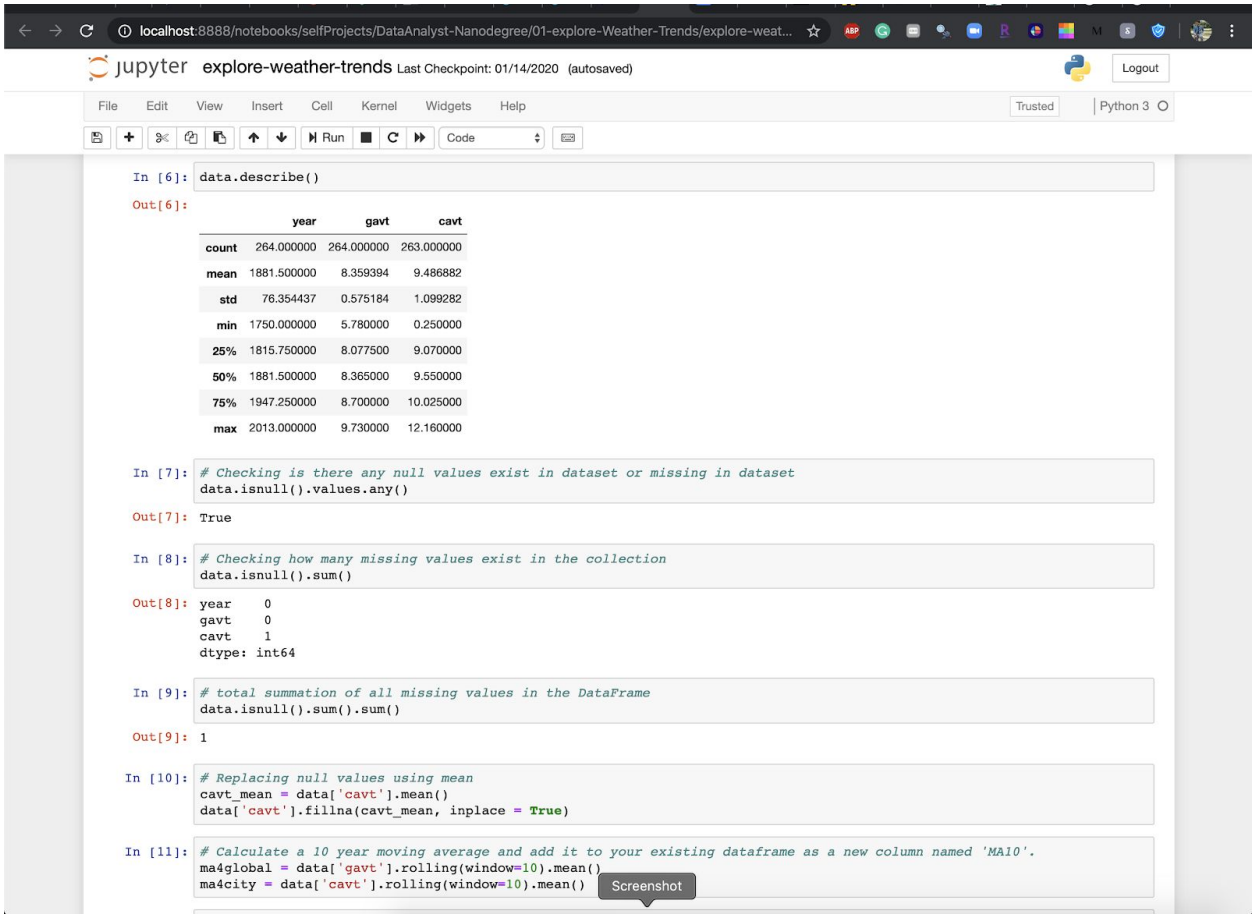
from global_data as g JOIN city_data as c

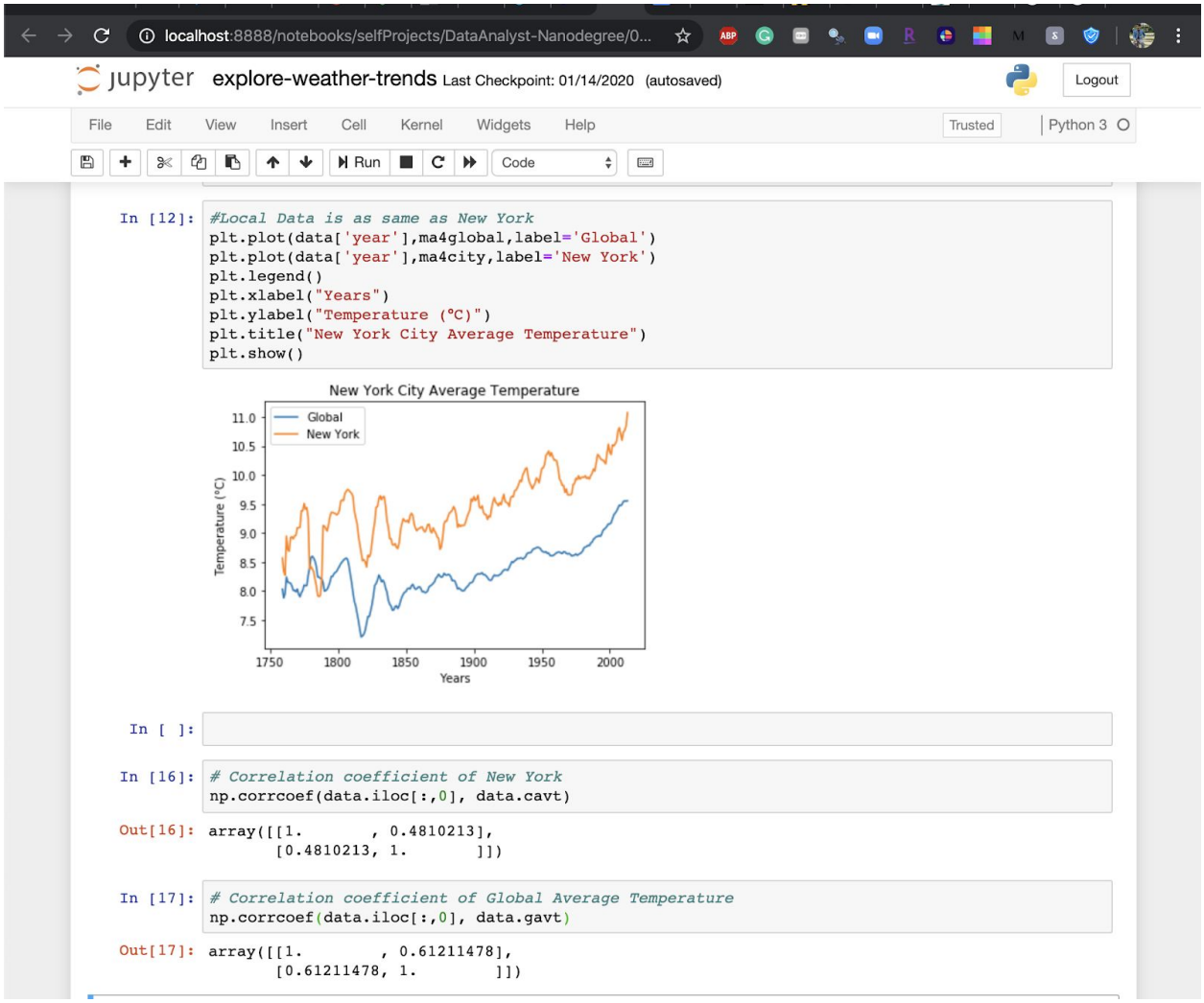
ON g.year = c.year

where city = 'New York'

II. Step 2:- Python Code for Making Line Chart




localhost:8888/notebooks/selfProjects/DataAnalyst-Nanodegree/01-explore-Weather-Trends/explore-weat...
jupyter explore-weather-trends Last Checkpoint: 01/14/2020 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [6]: data.describe()
Out[6]:
year gavt cavt
count 264.000000 264.000000 263.000000
mean 1881.500000 8.359394 9.486882
std 76.354437 0.575184 1.099282
min 1750.000000 5.780000 0.250000
25% 1815.750000 8.077500 9.070000
50% 1881.500000 8.365000 9.550000
75% 1947.250000 8.700000 10.025000
max 2013.000000 9.730000 12.160000
In [7]: # Checking is there any null values exist in dataset or missing in dataset
data.isnull().values.any()
Out[7]: True
In [8]: # Checking how many missing values exist in the collection
data.isnull().sum()
Out[8]: year 0
gavt 0
cavt 1
dtype: int64
In [9]: # total summation of all missing values in the DataFrame
data.isnull().sum().sum()
Out[9]: 1
In [10]: # Replacing null values using mean
cavt_mean = data['cavt'].mean()
data['cavt'].fillna(cavt_mean, inplace = True)
In [11]: # Calculate a 10 year moving average and add it to your existing dataframe as a new column named 'MA10'.
ma4global = data['gavt'].rolling(window=10).mean()
ma4city = data['cavt'].rolling(window=10).mean()
Screenshot



Results:

The Line chart is shown above tells us:

1. Overall trends looks like that the temperature is increasing in all over the world.
2. Even the temperature is increasing in New York also.
3. As you can see the positive correlation. It means that it is increasing but not linearly increasing.
4. Correlation coefficient is the deciding factor.
5. There is more rapid increase in global temperature comparative to the New York temperature.
6. The below shown figure is of correlation coefficient of New York.