CSC301 – Final Project Presentation

Ahsan Fayyaz

Professor Erik Grimmelmann

AGENDA

The Project Idea

My Datasets

Computations

Visualizations

Analyisis/Conclusion

Future Prediction using ML training model

IDEA: Climate Change in NYC! Is it real?

- We all have at some point heard
  - Climate Change
  - Rising Surface Temperatures
  - Carbon Dioxide Emissions
- Let's use data to find out!

Datasets

There are two datasets that I have pulled from the web

1. Average Temperatures in NYC

https://www.weather.gov/media/okx/Climate/CentralPark/monthlyannualtemp.pdf

2. Carbon Dioxide Levels (PPM) globally

http://www.co2.earth/historical-co2-datasets

Dataset 1: Average Temperatures in NYC

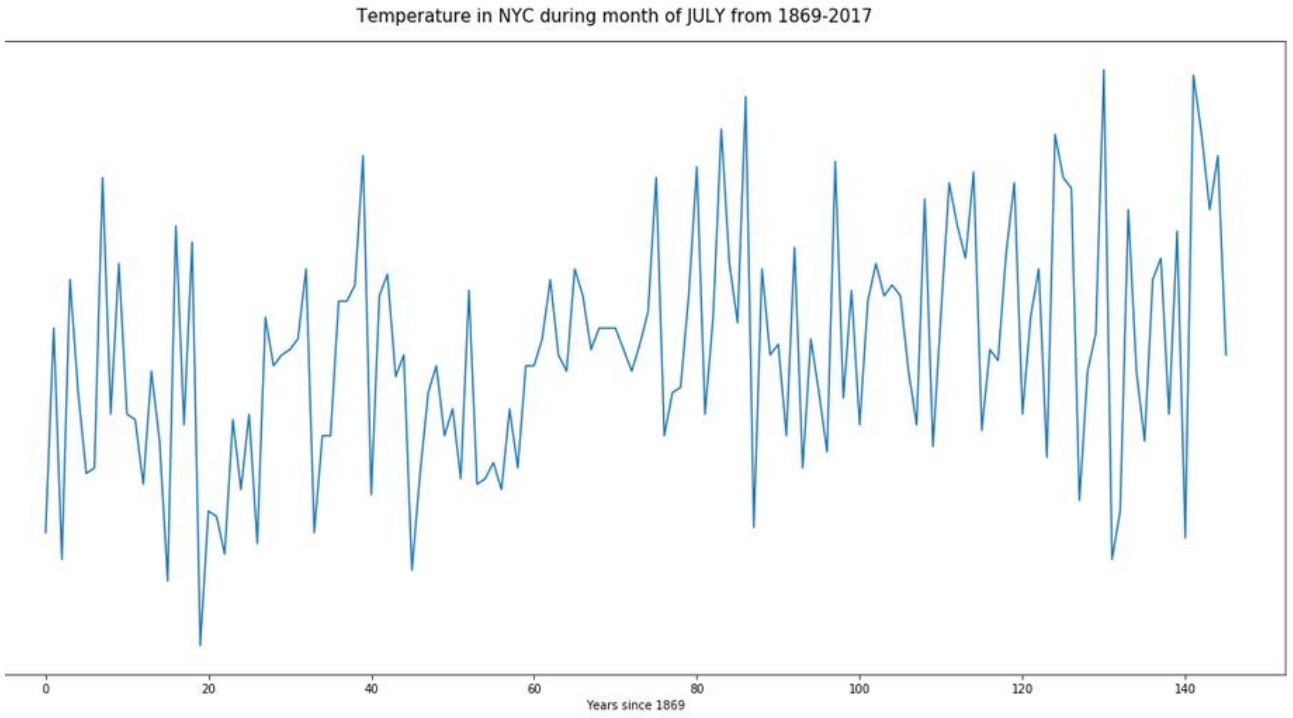
Click to add text

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
0	1869	35.1	34.5	34.8	49.2	57.7	69.3	72.8	71.8	65.6	50.9	40.3	34.7	51.4
1	1870	37.5	31.3	34.1	50.7	60.9	72.9	76.6	75.3	67.6	56.7	45.5	34.1	53.6
2	1871	28.3	30.2	44.2	52.0	60.4	68.2	72.3	73.6	60.8	55.6	38.8	29.2	51.1
3	1872	28.8	29.9	30.5	49.4	61.5	71.2	77.5	75.6	66.4	53.2	41.0	26.7	51.0
4	1873	28.6	29.5	35.7	46.7	58.8	70.3	75.4	72.0	65.4	55.8	37.0	36.5	51.0
5	1874	34.2	31.3	37.1	41.1	58.8	70.1	73.9	70.3	67.0	55.1	43.4	33.8	51.3
6	1875	23.8	25.2	34.1	43.1	60.1	69.2	74.0	72.9	64.0	53.6	39.3	33.9	49.4
7	1876	36.6	31.8	34.4	47.0	60.2	73.5	79.4	75.2	63.7	50.6	45.2	24.9	51.9
8	1877	27.7	37.0	35.8	47.7	59.6	70.2	75.0	75.4	66.9	55.8	44.5	37.4	52.8
9	1878	30.3	32.2	44.1	53.3	59.4	67.7	77.8	74.2	68.3	58.7	43.8	32.8	53.6

About this dataset:

- Average monthly & annual average temperatures
- Central Park NYC
- Starting from the year 1869
- Till 2014

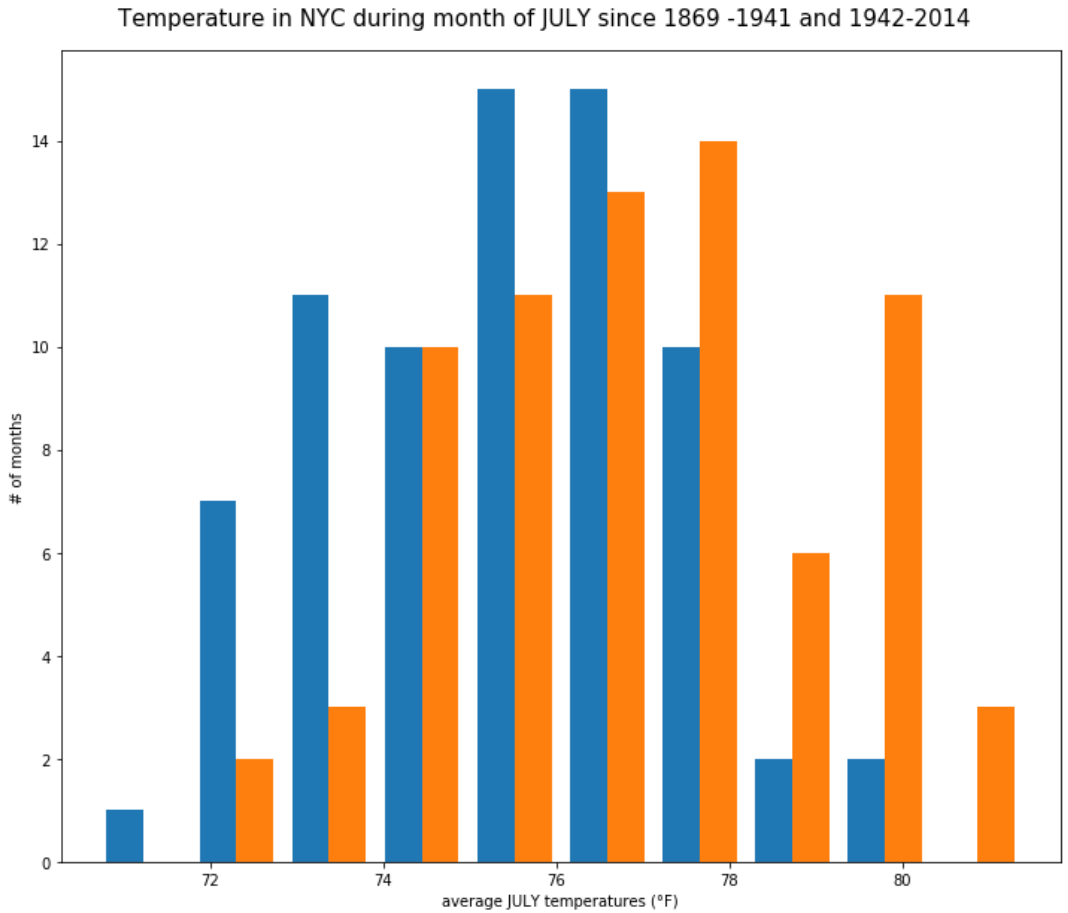
Dataset 1: Visualize



Comments:

- Created using Matplotlib (Python) Jupyter Notebook
- Upward trend observed as more years have passed since 1869
- Higher peaks

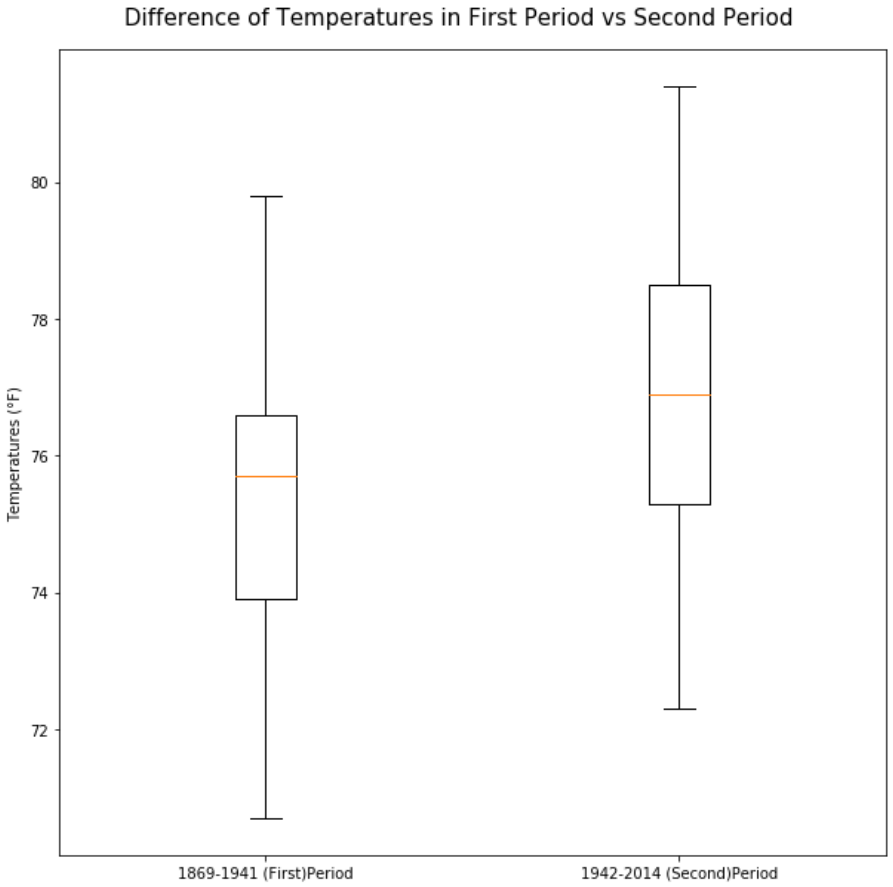
Dataset 1: Split the data into two periods (1869-1941) and (1942-2014)



Comments:

- Blue Represents (1869-1941)
- Orange Represents (1942-2014)
- Notice Higher temperatures in Period2 as compared to Period1

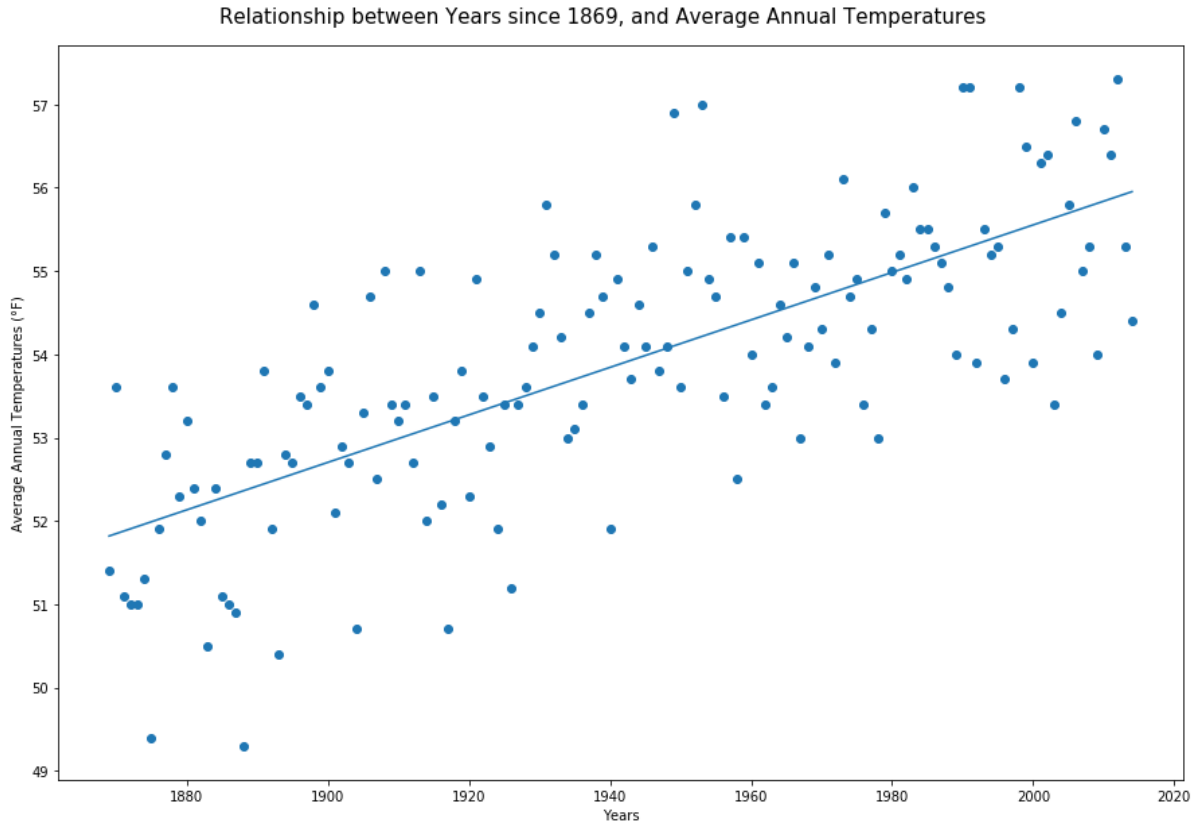
Dataset 1: Split the data into two periods (1869-1941) and (1942-2014)



Comments:

- Left (1869-1941)
- Right (1942-2014)
- Notice Higher Median temperatures in Period2 as compared to Period1

Evaluate Linear Relationship between:
1. Years Since 1869
2. Annual Average Temperatures in NYC



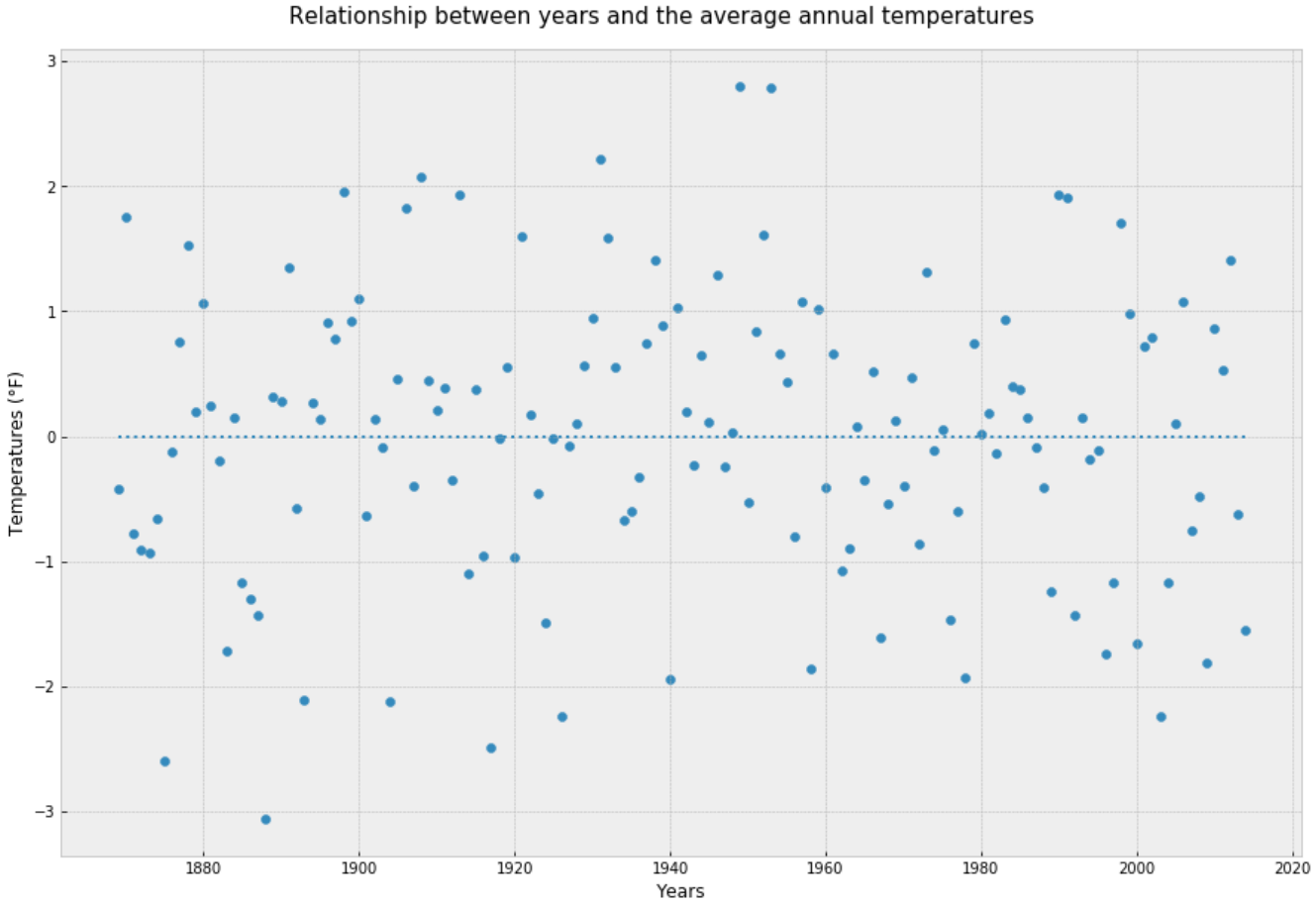
A positive linear relationship is observed between **ANNUAL average** temperatures and years since 1869

The Slope tells us that for every one year, there is an increase of 0.0285 Fahrenheit Temperature

Comments:

- Coded in Python to evaluate linear relationship
- slope = 0.0285
- This represents that temperatures have increased at a rate of 0.0285 Fahrenheit per year

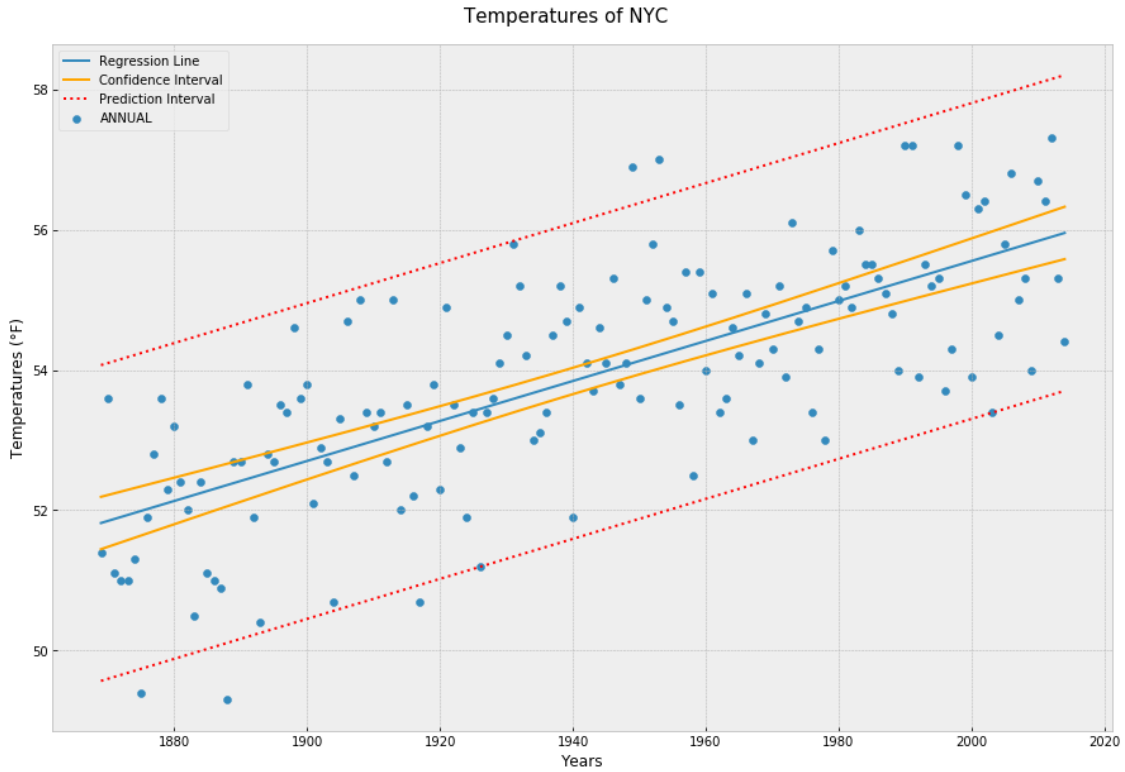
Residual Graph



Comments:

- R-squared value = 52.69% depicts 52.69% of the variation in the temperatures is accounted for by its regression on number of years passed
- Graph is relatively Homoskedastic
- Relationship isn't super strong. However, it still is positive linear

Confidence Intervals & Prediction Intervals



The orange lines are the 95% confidence interval of where the mean Y value for a given X value will lay, and the red lines are the 95% confidence interval of where a new Y value for a given X value could land. Most of the observed points are between the red lines - as 95% of them should be.

Dataset 2: Global Carbon Dioxide Levels

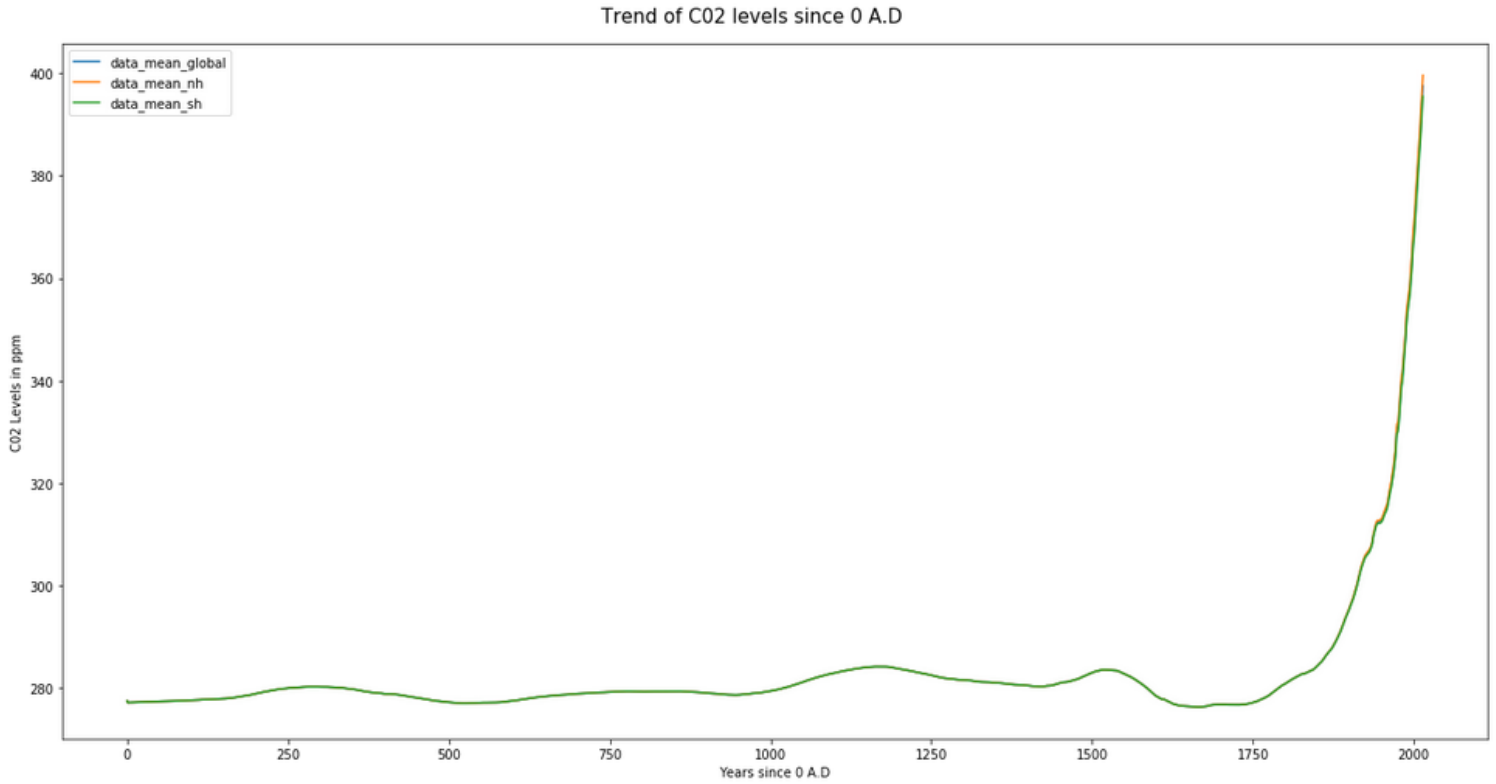
Click to add text

	year	data_mean_global	data_mean_nh	data_mean_sh
1869	1869	287.225000	287.261016	287.188983
1870	1870	287.355000	287.392248	287.317751
1871	1871	287.494000	287.533380	287.454619
1872	1872	287.663999	287.706570	287.621429
1873	1873	287.859999	287.904581	287.815418
1874	1874	288.060999	288.104895	288.017104
1875	1875	288.290999	288.337071	288.244928
1876	1876	288.519999	288.567204	288.472795
1877	1877	288.751999	288.800180	288.703818
1878	1878	288.992999	289.042084	288.943915

About this dataset:

- Carbon dioxide Levels in Air
- PPM (parts per million)
- Northern, Southern Hemisphere, Mean
- The original dataset starts from year 0 but I scraped the dataset to begin 1869 for my computations

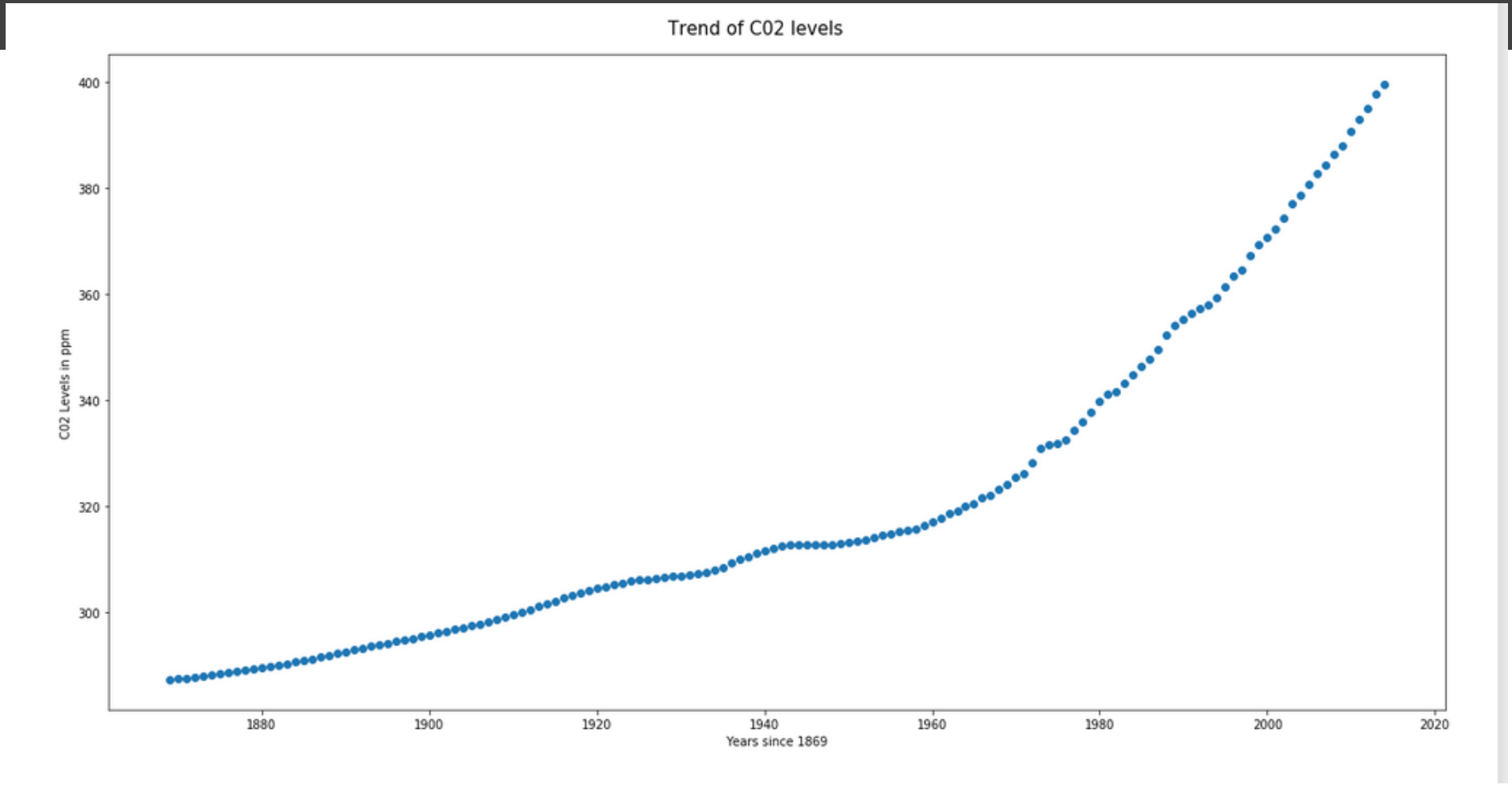
Dataset 2: Visualize (since year 0)



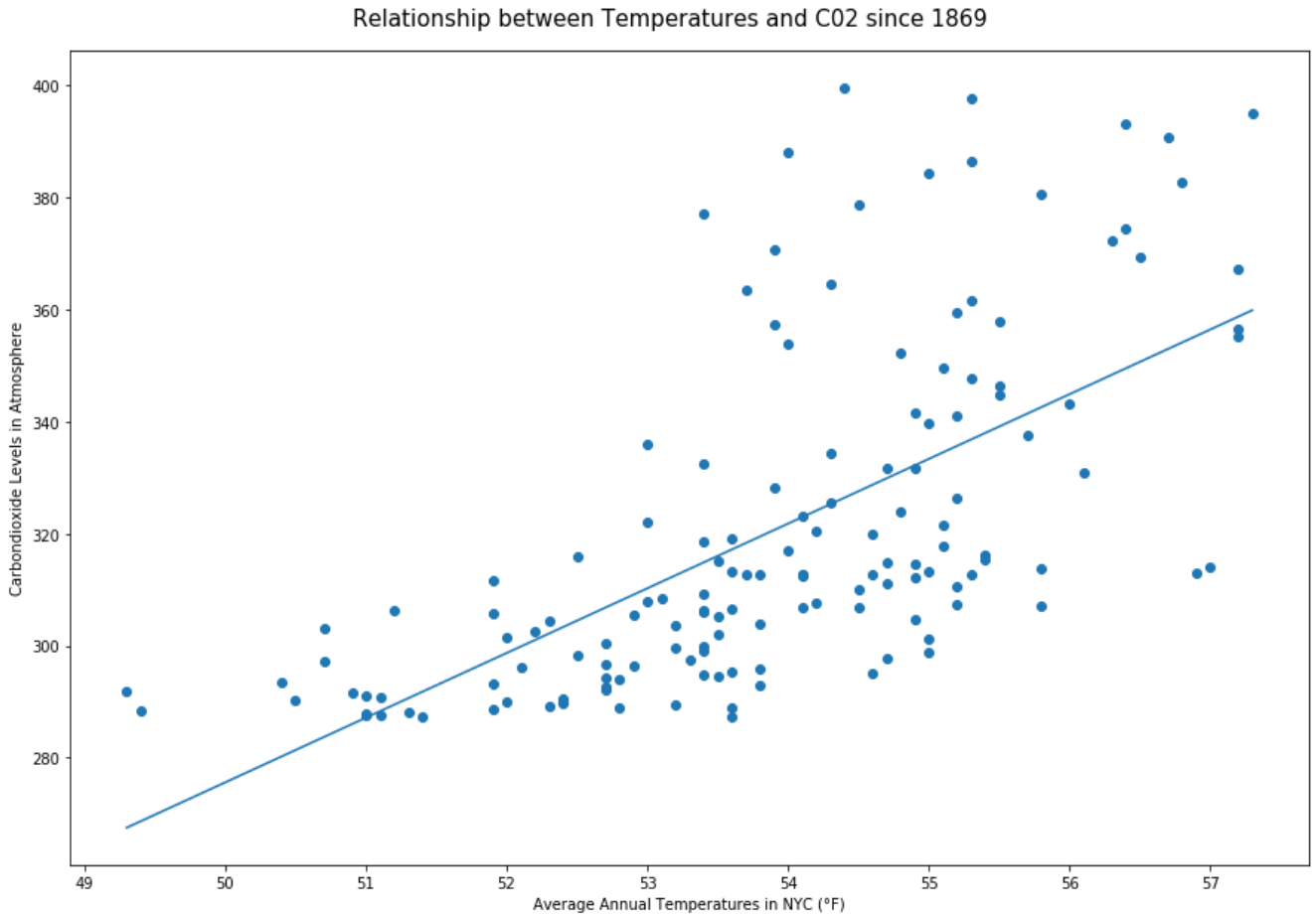
Comments:

- Created using Matplotlib (Python) Jupyter Notebook
- Upward trend observed as more years have passed since 1869
- Seems exponential after 1750s (industrialization hmmm….?)

Dataset2 : Visualize (1869-2014) in Northern Hemisphere



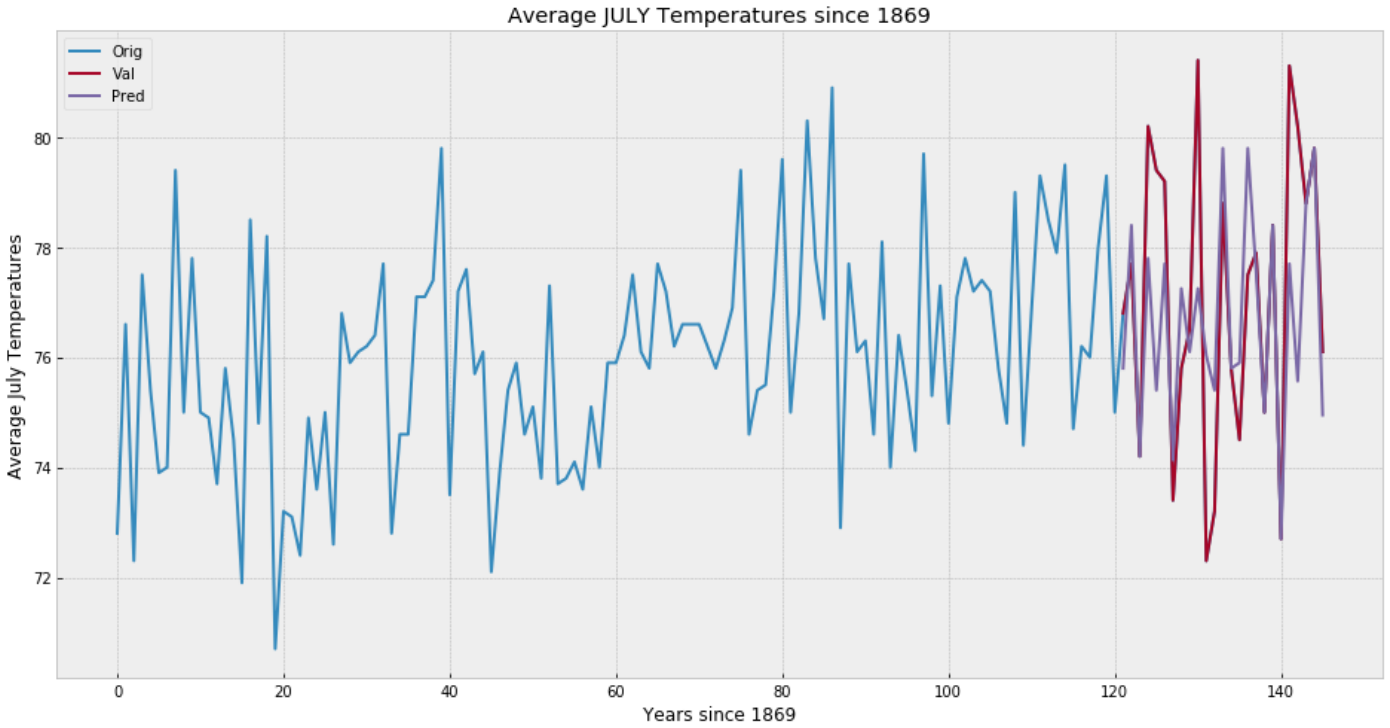
Evaluate Linear Relationship between:
1. Global Carbon Dioxide Levels
2. Annual Average Temperatures in NYC



Comments:

- Coded in Python to evaluate linear relationship
- slope = 11.54ppm
- This represents that for every 'one' degree of temperature increase, there is 11.54pm C02 that contributed
- There is an R-squared value of 41.35% here, meaning that the 41.35% of the variation in the Temperatures is accounted for by its regression on the levels of Carbondioxide in Northern Hemisphere

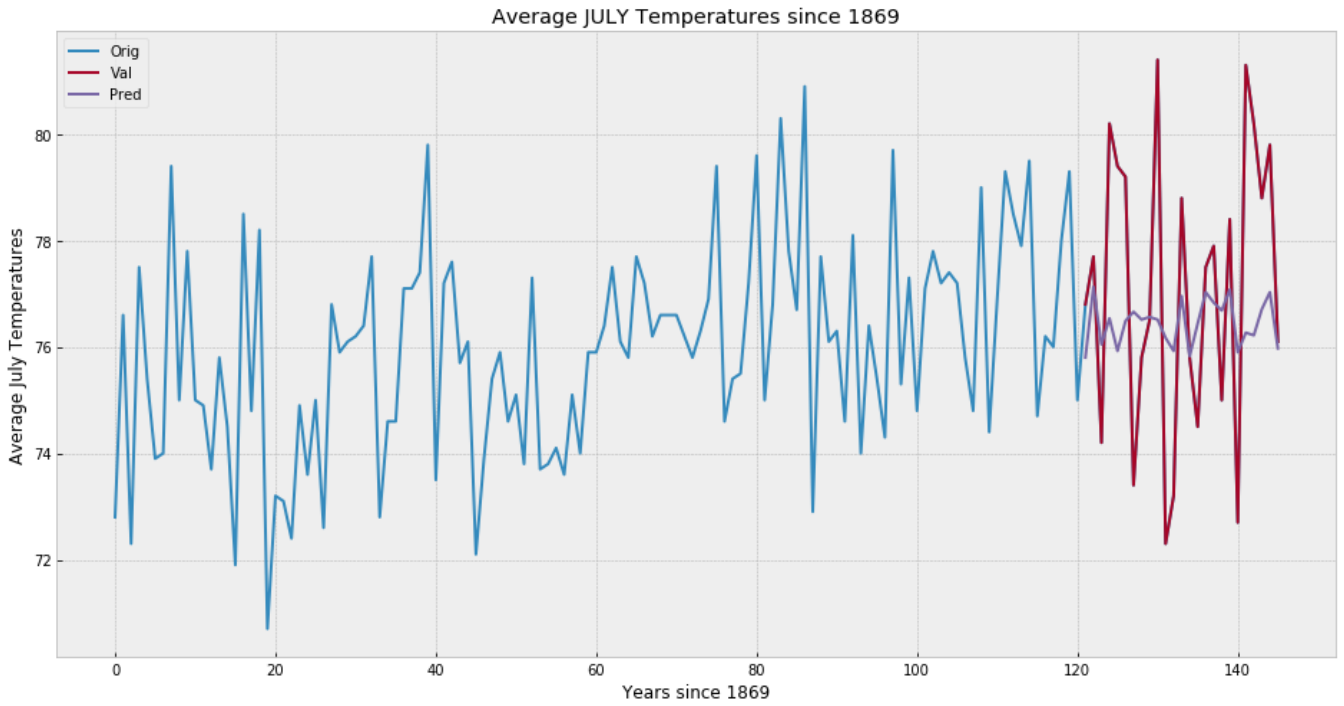
Training ML models for Temperature Dataset using sklearn



Comments:

- Coded in Python train a tree-prediction
- Split the data into 75% training and 25% testing
- Created decision tree regressor model
- Actual Value in Magenta
- Predicted Value in Purple
- Model is accurate to a good extent

Training ML models for Temperature Dataset using sklearn



Comments:

- Split the data into 75% training and 25% testing
- Created linear model regressor prediction
- Actual Value in Magenta
- Predicted Value in Purple
- Model can be improved by further training

Conclusion

- Two datasets
- The actual mean temperatures over time have increased as much as **2°Fahrenheit**
- **53%** of the variation in the Temperatures is accounted for by its regression on number of Years passed.
- My R-squared value for C02 and Annual Average Temperature was around **41%**
- **With confidence, I can state that the annual temperatures in NYC have been rising at a rate of 0.02°F / Year.**