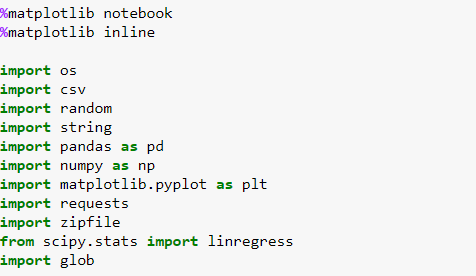
Data Exploration and Clean-up Process

* **Importing Dependencies**



* **Downloading Files**

A screenshot of a cell phone

Description automatically generated

* **Concatenation of large number of CSV Files**

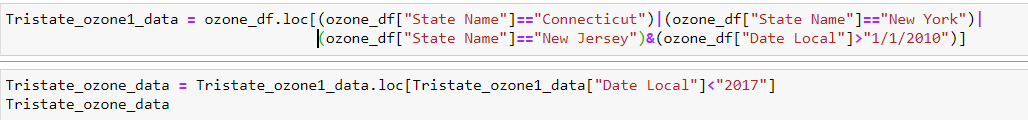
**A screenshot of a social media post

Description automatically generated**

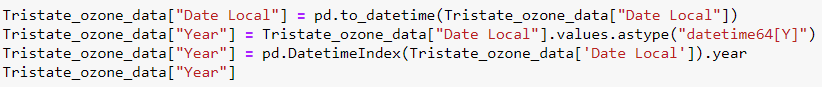
* **Reading Files**

****

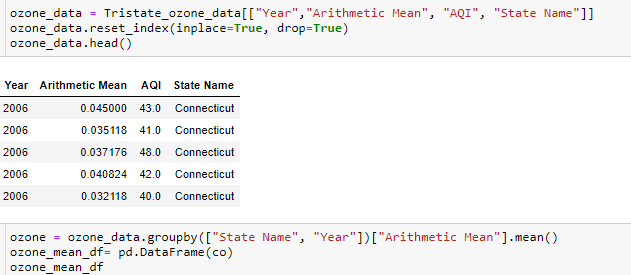
* **Data analysis for gases file**
* **Creating the data-frames:**
  + **We created a new dataframe using loc function to filter the data by tri-states and year**



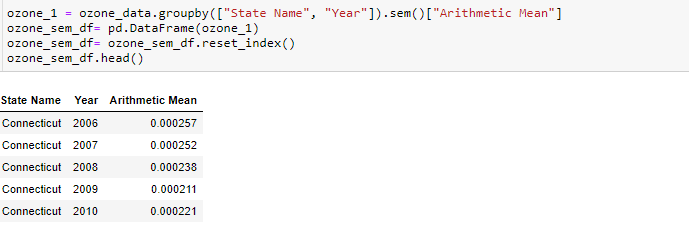
* **We required the data with a column Year and the date local column has day, so we used to datetime function for conversion**



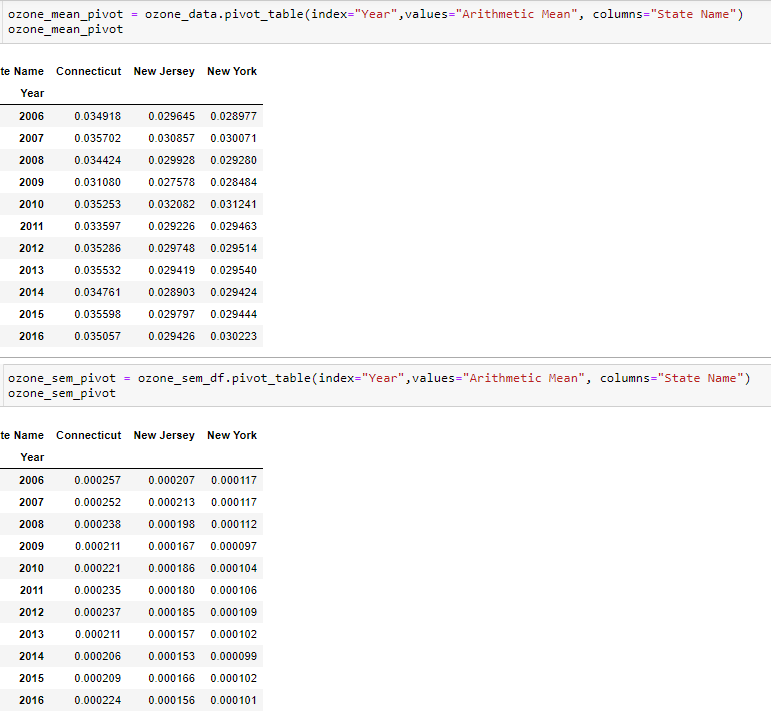
* **We pulled only the columns that were required for our analysis and calculated the mean to a dataframe**



* **We calculated the sem to a dataframe.**



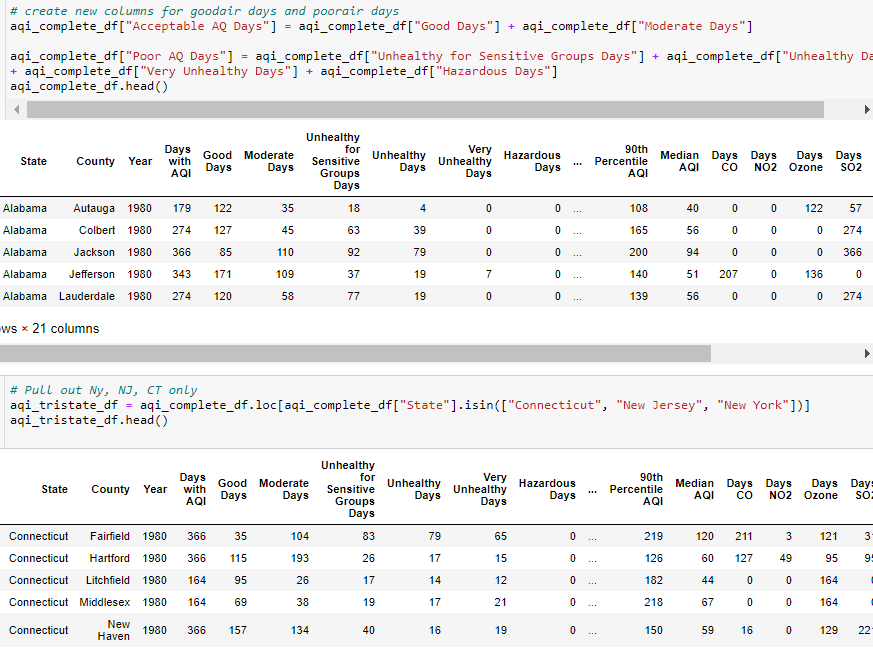
* **We created pivot tables for both mean and sem dataframes to get the states in the columns**



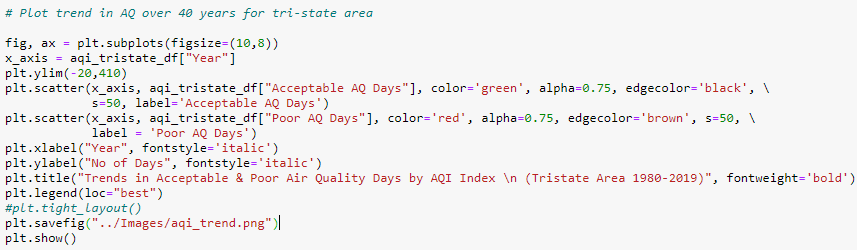
* **We created line graphs for the tristate showing the trend of average concentration for gasses (No2, So2, CO and Ozone)**



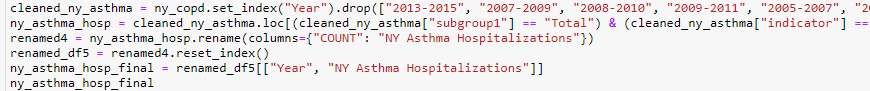
* **Data analysis for AQI file:**
  + **We created a new column as Acceptable AQ Days by adding Good Days and Moderate Days columns and pulled the tri-state data**



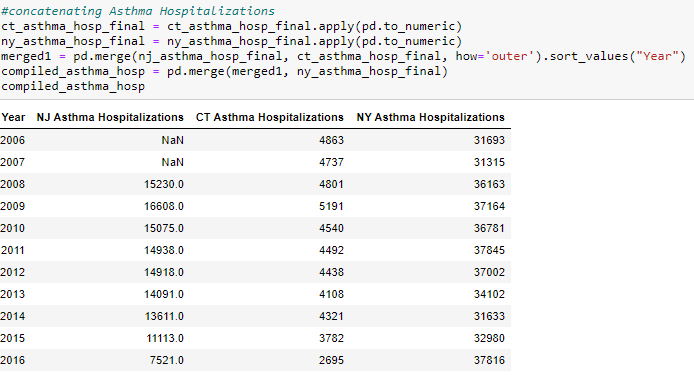
* **We created a scatter plot trend in AQ over 40 years for tri-state area**



* **Data analysis for Health Effects file:**
  + **We created a dataframe with year and number of asthma hospitalization, COPD hospitalization and ED visits for each NJ, NY and CT**



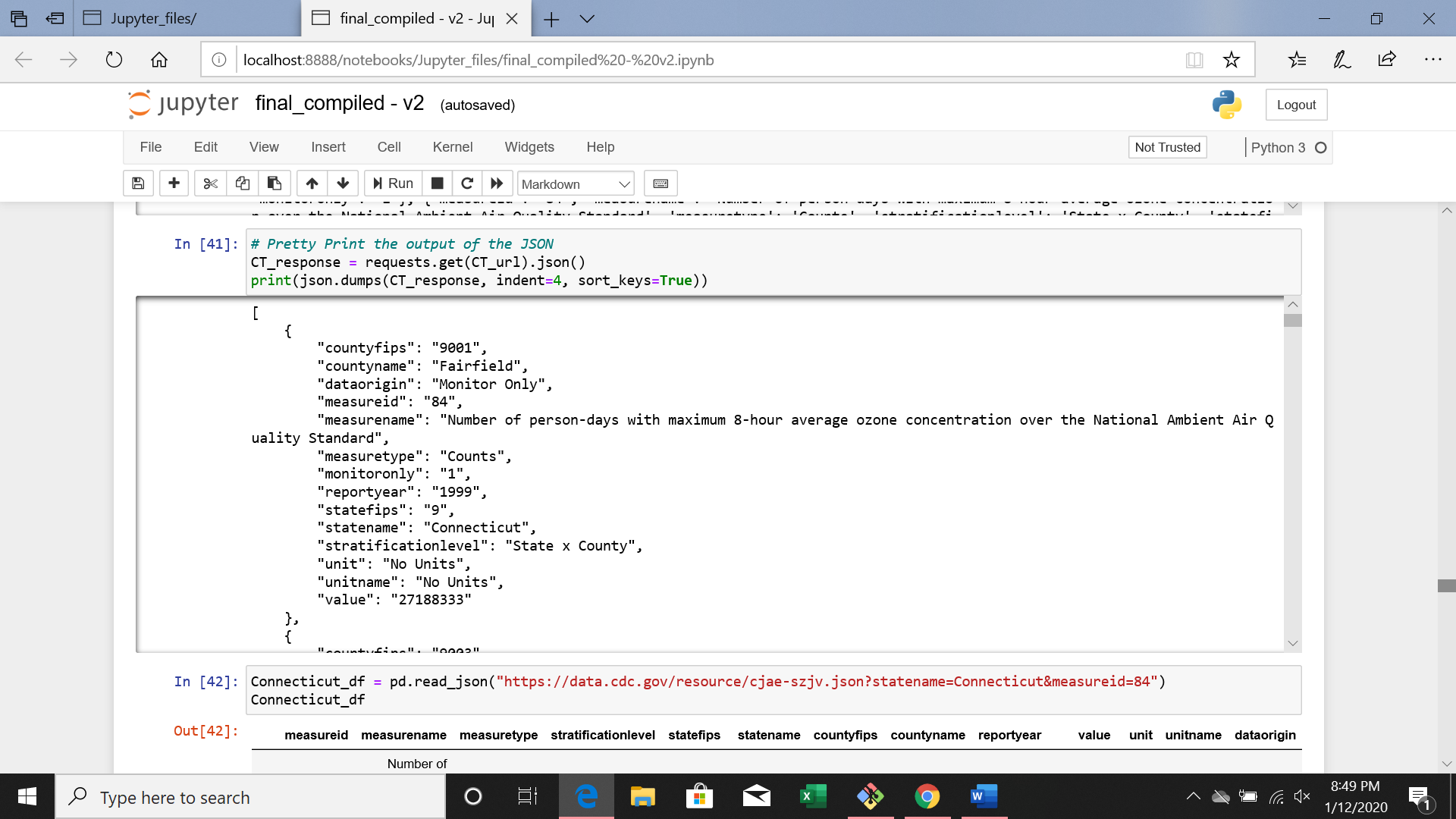
* **We concatenated the three dataframes in one table for each asthma hospitalization and ED visits and COPD hospitalization which contained all the three states**



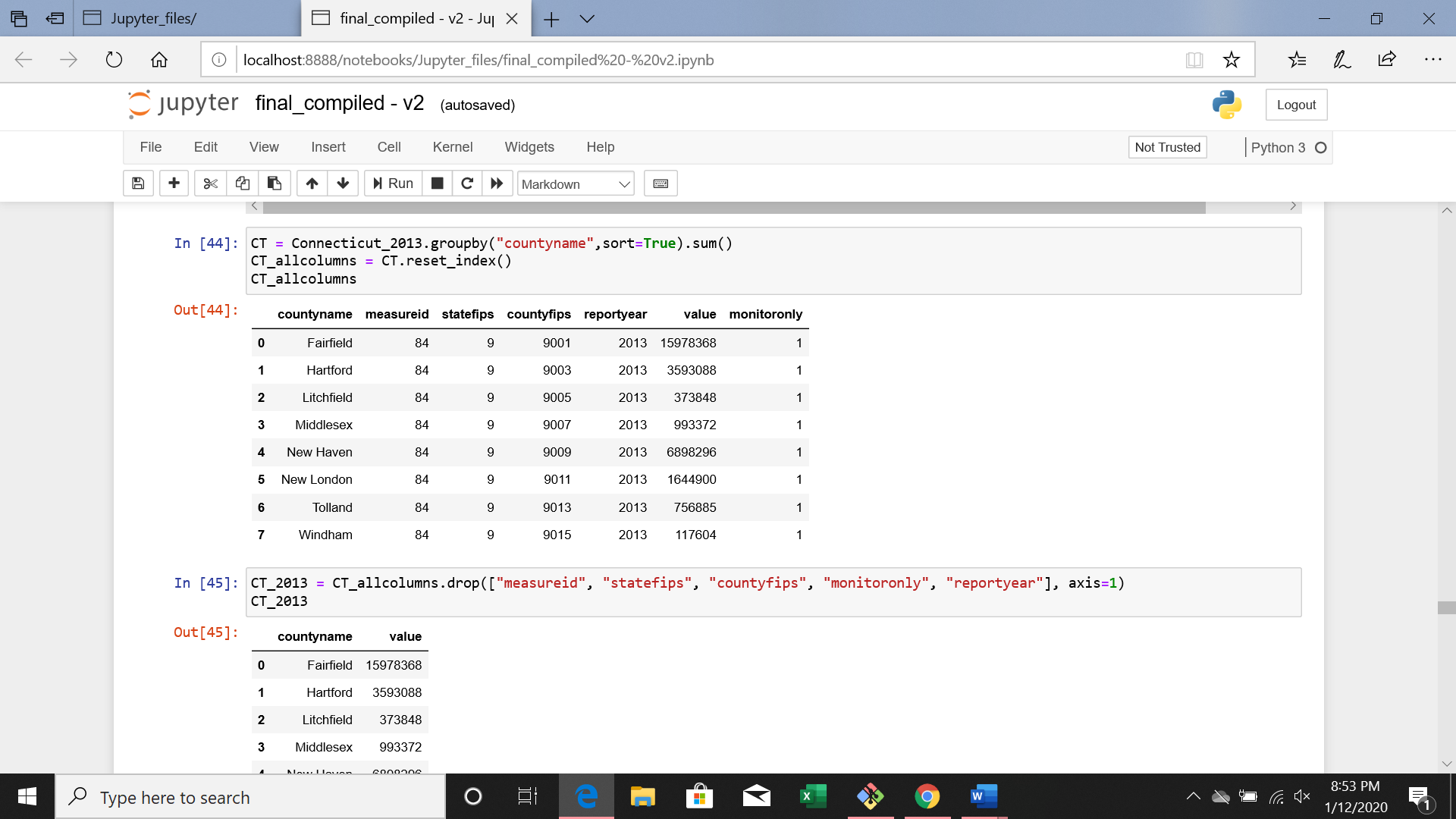
* **We created a stacked bar chart for all the three categories**



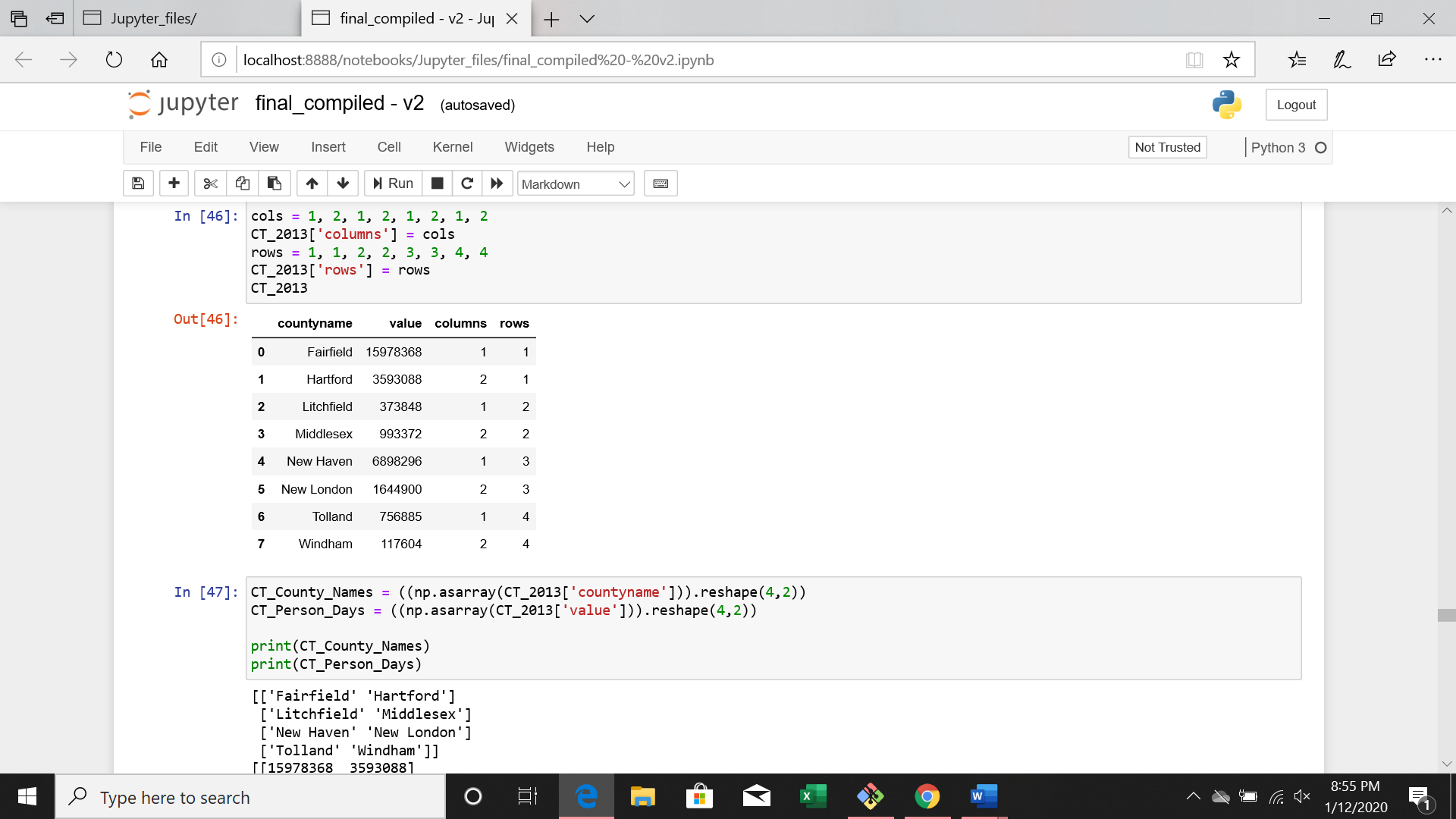
* **Creating heat map by passing JSON and read json using URL**



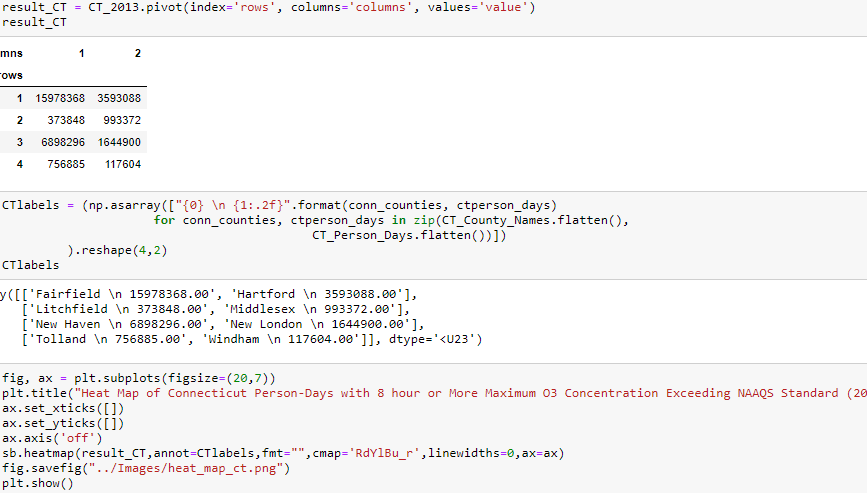
* **We pulled only the columns that were required for our analysis and calculated the mean to a dataframe**



* **We pulled specific columns and rows for county name and values.**



* **We created a heatmap with the dataframe by county in all three states (Tri-states)**



References:

<https://aqs.epa.gov/aqsweb/documents/data_api.html>

[https://www.cdc.gov/nceh/data.htm](https://slack-redir.net/link?url=https%3A%2F%2Fwww.cdc.gov%2Fnceh%2Fdata.htm)

<https://www.epa.gov/national-air-toxics-assessment/2014-nata-assessment-results#pollutant>

<https://cdc.gov/asthma/asthma_stats/attacks-current-asthma.htm>

<https://cdc.gov/asthma/nhis/default.htm>

<https://cdc.gov/asthma/nhis/default.htm#anchor_1524067853614>

<https://en.wikipedia.org/wiki/Clean_Air_Act_(United_States)#Legislation>

<https://apps.health.ny.gov/statistics/environmental/public_health_tracking/tracker/index.html#/asthmaMonth>

<https://portal.ct.gov/DPH/Health-Education-Management--Surveillance/Asthma/Asthma-Statistics>

<https://www-doh.state.nj.us/doh-shad/indicator/view/COPD.HospNJAAR.html>

<https://www.cga.ct.gov/2015/rpt/pdf/2015-R-0127.pdf>

EPA

<https://www.epa.gov/outdoor-air-quality-data/about-air-data-reports>

<https://aqs.epa.gov/aqsweb/documents/data_api.html>

CDC

<https://data.cdc.gov/resource/cjae-szjv.json?statename=Connecticut&measureid=84>

<https://data.cdc.gov/resource/cjae-szjv.json?statename=New%20Jersey&measureid=84>

<https://data.cdc.gov/resource/cjae-szjv.json?statename=New%20York&measureid=84>

NJ

<https://www-doh.state.nj.us/doh-shad/indicator/complete_profile/NJEPHTAsthmaHosp.html>

<https://www-doh.state.nj.us/doh-shad/query/selection/ub/UBSelection.html>

CT

<https://portal.ct.gov/DPH/Communications/Services--Programs/Environmental-Health>

<https://portal.ct.gov/DPH/Health-Information-Systems--Reporting/Hisrhome/Hospitalization-Statistics>

NY

<https://apps.health.ny.gov/statistics/environmental/public_health_tracking/tracker/index.html#/asthmaMonth>