

## Get NYC parking violation data:

This project uses sodapy api to download NYC parking violations data. And it also contains the docker file to build the docker container to use the python code.

The main.py call the api function to download data.

To run the main.py, just type:

```
python main.py input.txt output.json
```

in command line window.

input.txt is the input file, which contains two parameters. One is the page size, and one is the num of page. It can be set as:

```
2 3
```

Here 2 is the page size and 3 is the num of page.

src directory contains the api python code.

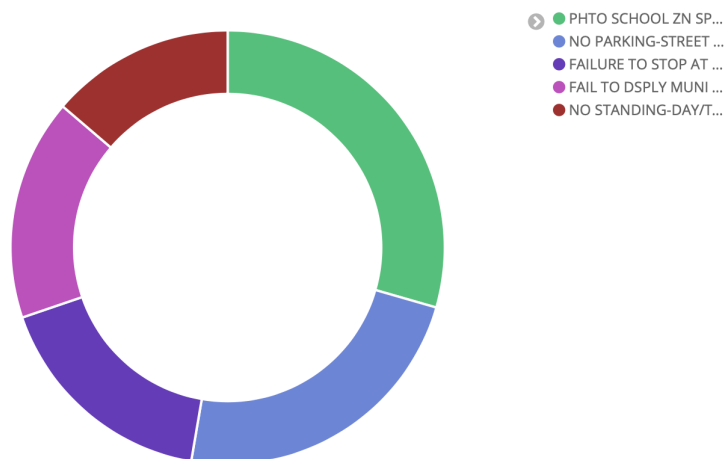
output.txt file saves the downloaded data.

token.txt file has the app token to log in the NYC data website.

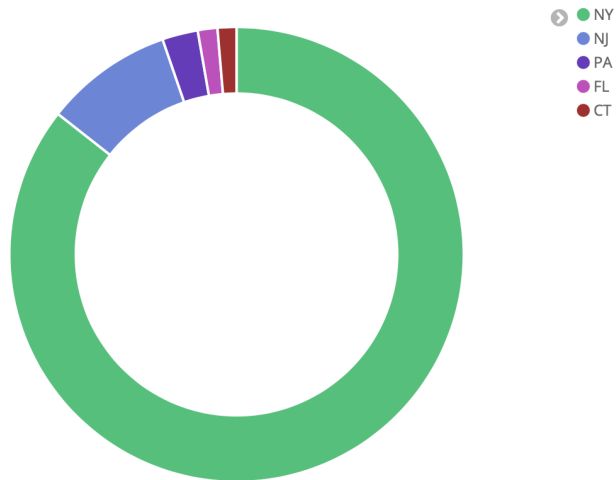
## Kibana Visualization:

The Kibana visualization is based on 100,000 data from the NYC parking violation data. The output.txt file has the results from curl command to show the data is loaded into elasticsearch.

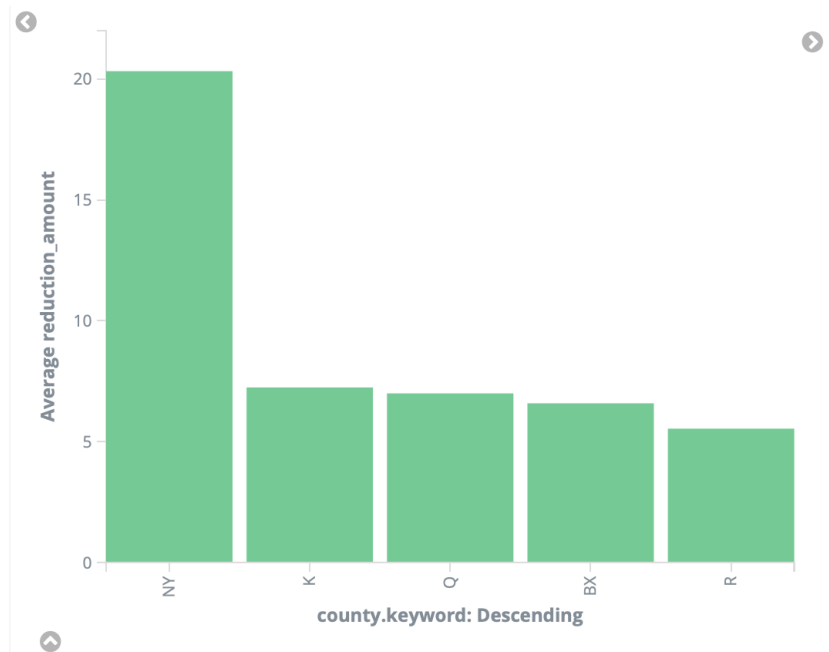
1: The first figure shows the frequency of the most several violation type. The most violation type is: NO PARKING-STREET CLEANING. And the second most violation type is: FAILURE TO STOP AT RED LIGHT.



2: The second figure shows the violation proportion in each states. NY has the majority of the violation cases in the data set, which is about 85.6% of the total violation. While NJ state has the second most violation cases, taking account 9.12% of the total violation.



3: The third figure shows 5 most average reduction amount in term of county. NY county has the most average reduction amount 20.325, and K county has the second average reduction amount 7.24.



4: The fourth figure shows the 5 most average fine amount in each county. NY has the highest average fine amount of 82.116, while county K has the second highest average fine amount of 56.716.

