

Project 1

By Christy Wong

Date: April 10th, 2022

Overview

The Open Parking and Camera Violations was initially loaded with all violations contained in the mainframe database as of May 2016 and made by the NYC city government. It was last updated April 9 2022. There is 78.6M rows and 19 columns in this data.

Frame Work

- Docker
- The service that was used in the project was ElasticSearch, Kibana
- While using main.py - These were used:
 - From sodapy import Socrata
 - import requests
 - from requests.auth import HTTPBasicAuth
 - import argparse
 - import sys
 - import os
 - import json

BULK API

it was used to decrease the time to download the data information into Elasticsearch. This was used to process large amounts of data in batches. It was more efficient to use because it would have taken 3-4 hours to get 100k hits.

To load the Elastic Search:

you need to build the docker image by using this command:

```
docker build -t nycparking:1.0 .
```

This command was use to load the amount of data into Elastic Search Dashboard:

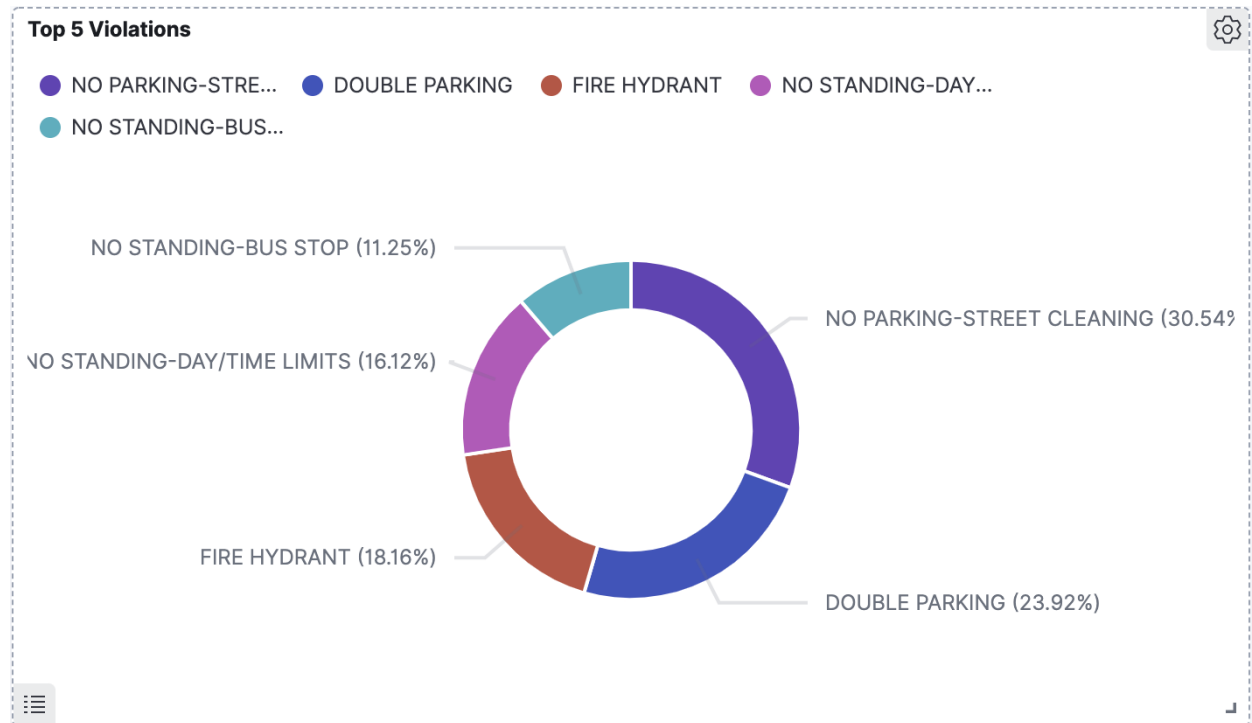
```
In [2]: # docker run \  
# -e DATASET_ID="nc67-uf89" \  
# -v /path/to/data:/data
```

```
#-e APP_TOKEN="your_token" \  
#-e ES_HOST="your_domain" \  
#-e ES_USERNAME="your_username" \  
#-e ES_PASSWORD="your_password" \  
#-e INDEX_NAME="christyproject5" \  
#nycparking:1.0 --page_size=10000 --num_page=20
```

With Kibana, A dashboard was made with 4 visualizations:

The First Visualization is the Top 5 Violations:

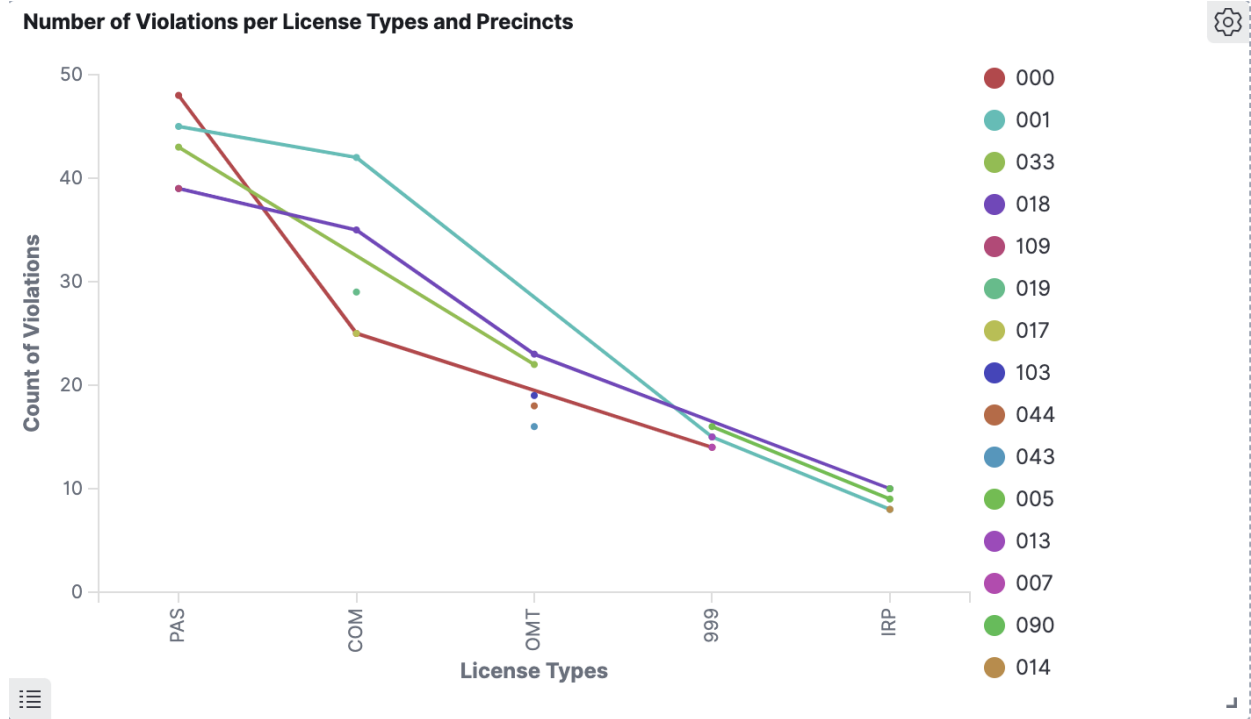
- The top violation is "No Parking- Street Cleaning" which is 30.54%



The Second Visualizations is count of violations per license type and which precincts have the highest:

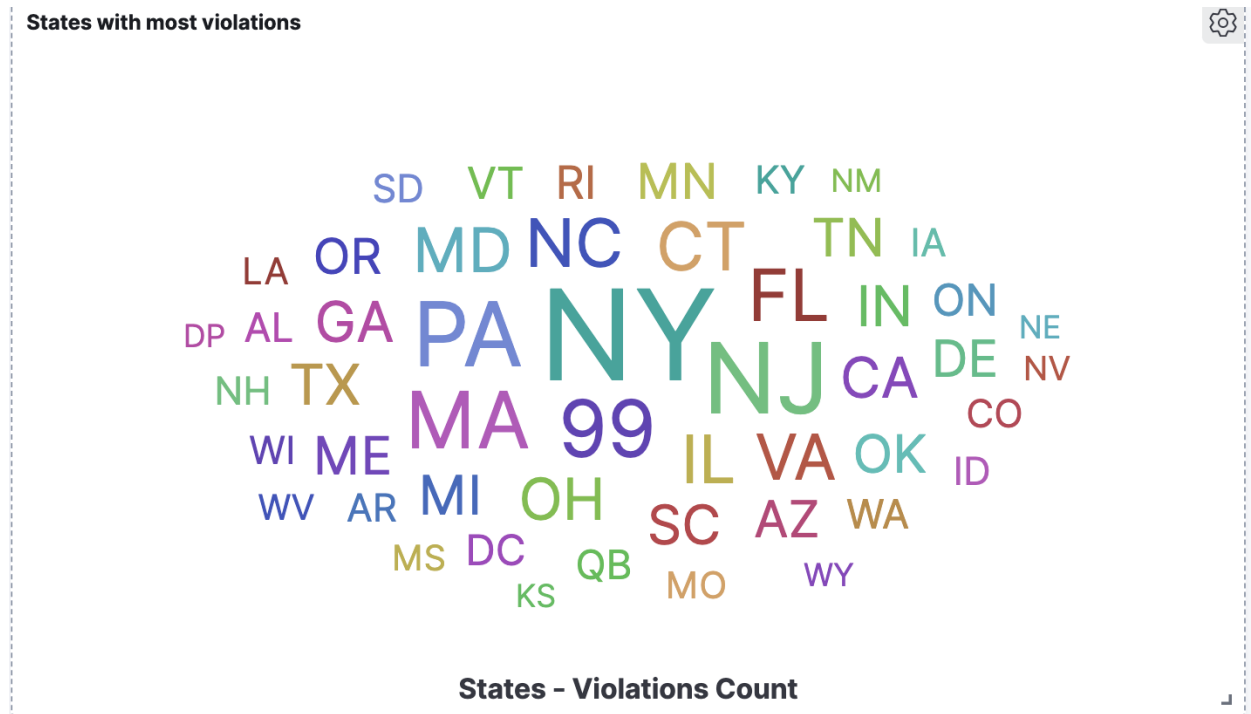
- PAS is the top license type that have the highest count of violations - PAS is Passenger Vehicles
- Midtown Precinct North is 000
- New York Police Department - 1st Precinct is 001
- New York City Police Department - 33rd Precinct is 033

These precincts are located in NYC



The Third Visualization is which 50 states have a high Violations count:

- NY have the highest violation count because its letters are bigger than the rest
- NJ and PA are the next states that have a high violation count compare to the rest



The Fourth Visualization is the top 10 states that have a high average of penalty/interest/fine amount:

- SD, BC, SC are the top three states that have a high average of penalty/interest/fine amount.

- South Dakota (SD) have an penalty average is 52 dollars, interest average is 42 dollars, and fine amount average is 78 dollars.

Average Penalty, Interest, Fine in Top 10 States

