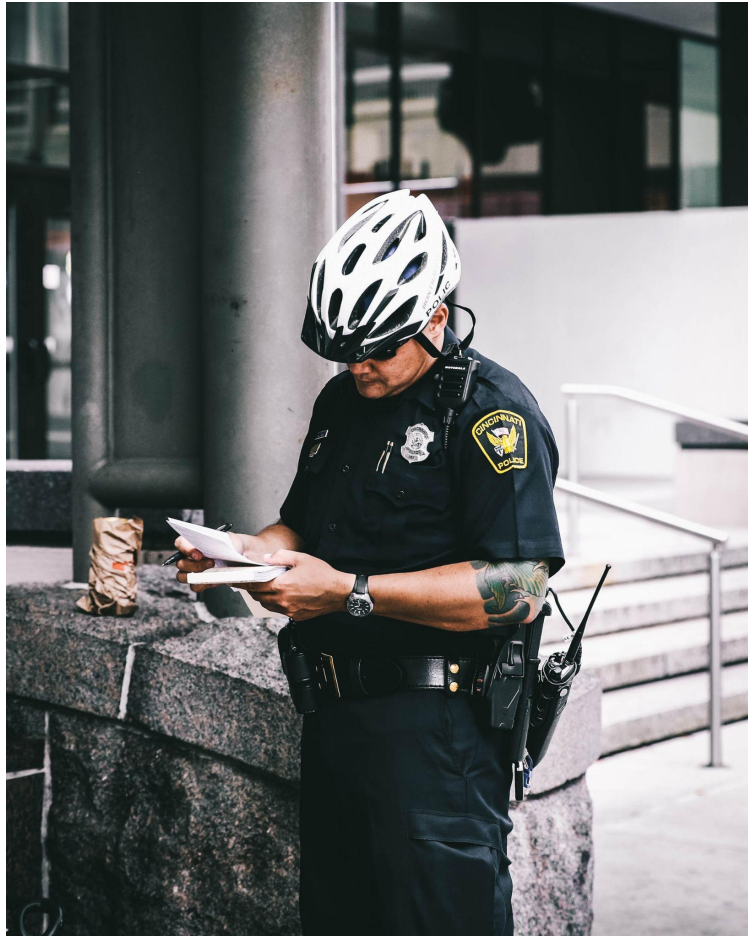


NYC Open Parking and Camera Violations



PROBLEM STATEMENT

NYC issues a large number of parking and camera violations. Currently there is not an easy way to identify detailed information on these violations.

CONTEXT

This project will provide a processing approach to retrieve historical data since 2016. Up to date records will also be processed on a daily basis. This will allow issuing agencies and the general public more insight into violations. This could be used to obtain metrics such as outstanding violations and breakdown by all boroughs.

SCOPE OF SOLUTION SPACE

The Scope of the solution will be to retrieve all the data from 2016 to present and ongoing basis. Once obtained it will be cleaned for any missing or incorrect data processed and loaded into the database for easy retrieval. Some sample metrics will be provided.

DATA SOURCE

The Data will be obtained from NYC Open DATA issued by the Department of Finance found at:

<https://data.cityofnewyork.us/City-Government/Open-Parking-and-Camera-Violations/nc67-uf89>

A python script will be used to leverage the initial data from excel. The ongoing daily updates will be obtained from the SODA API

[Open Parking and Camera Violations | Socrata API Foundry](#)

CHOICE OF TECHNOLOGY

Extraction - Python Scripts will be used to obtain the data. It will use cloud storage to house it.

Transformation - The data processing will leverage Airflow and Spark

Load - Data will be loaded both into MySQL database and ElasticSearch for additional Searching capabilities

Overall infrastructure will leverage Kubernetes

CRITERIA FOR SUCCESS

The completed project will enable parking and traffic violations metrics to be easily retrieved.

All work will be captured in a GitHub repo which will include slides and readme file describing project.