Big Data Project 2025

NYC Parking Tickets

Lukas Zandomeneghi | Miriam Palmetshofer

DATASET

- **NYC Parking tickets**
- Collected every year since 2014 (~10M per year)
- **43 columns** like Plate ID, Violation Code, Issue Date, etc
- fiscal year **2014**, which spans from July 1, 2013, to June 30, 2014

TECHNOLOGY STACK

Data processing:
- Preprocessing
- Filtering
- Aggregating

Search and Indexing:
- Stores data from Spark
- Fast and scalable search

Data Visualization:
- Fetches data from Elasticsearch
- Lots of visualization possibilities

RESEARCH QUESTIONS

1. Do **repeat offenders** cluster in certain areas or vehicle types?
2. Are parking violations more common during **certain times of day or days of the week**?
3. Define **hotspots** of parking violations.
4. Analyse the **enforcement strength/severity** of the **agency.**

R1

Do repeat offenders cluster in certain areas or vehicle types?

REPEATED OFFENDERS – more than 5 tickets

Total offenders (identified by plate)

2,719,581

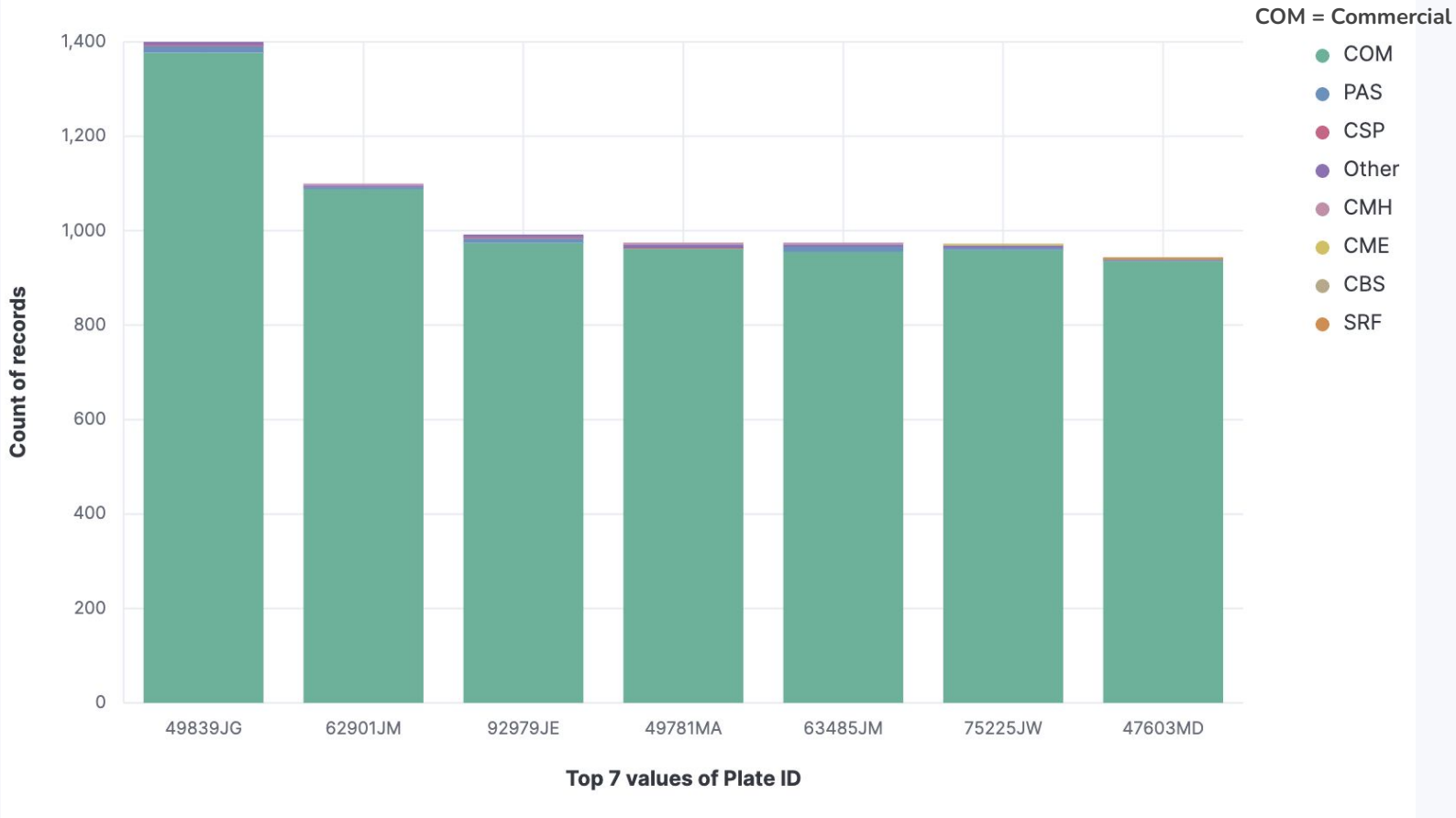
Unique count of plate_id for repeated offenders

12,2%

334,103

REPEATED OFFENDERS

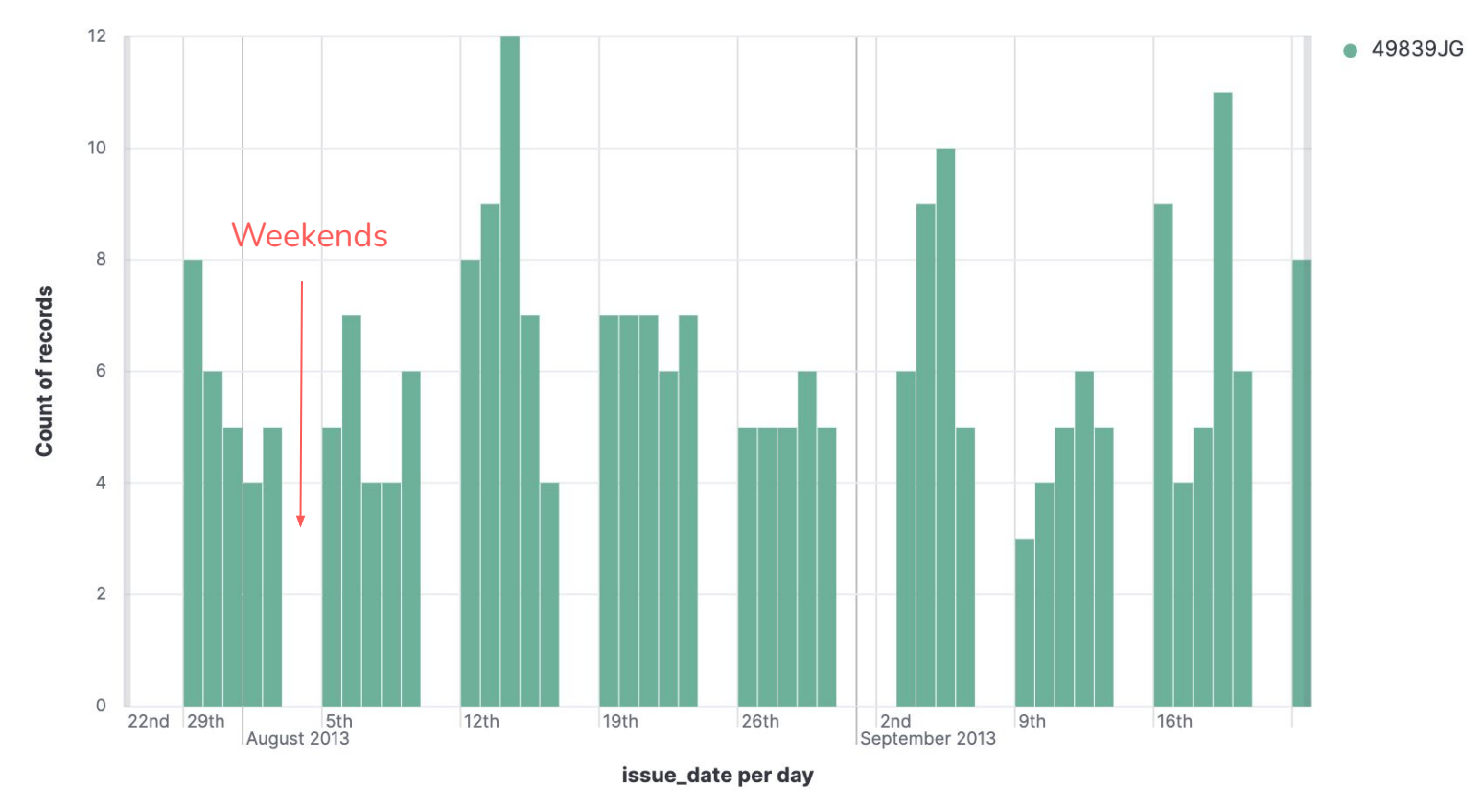
REPEATED OFFENDERS – Plate type



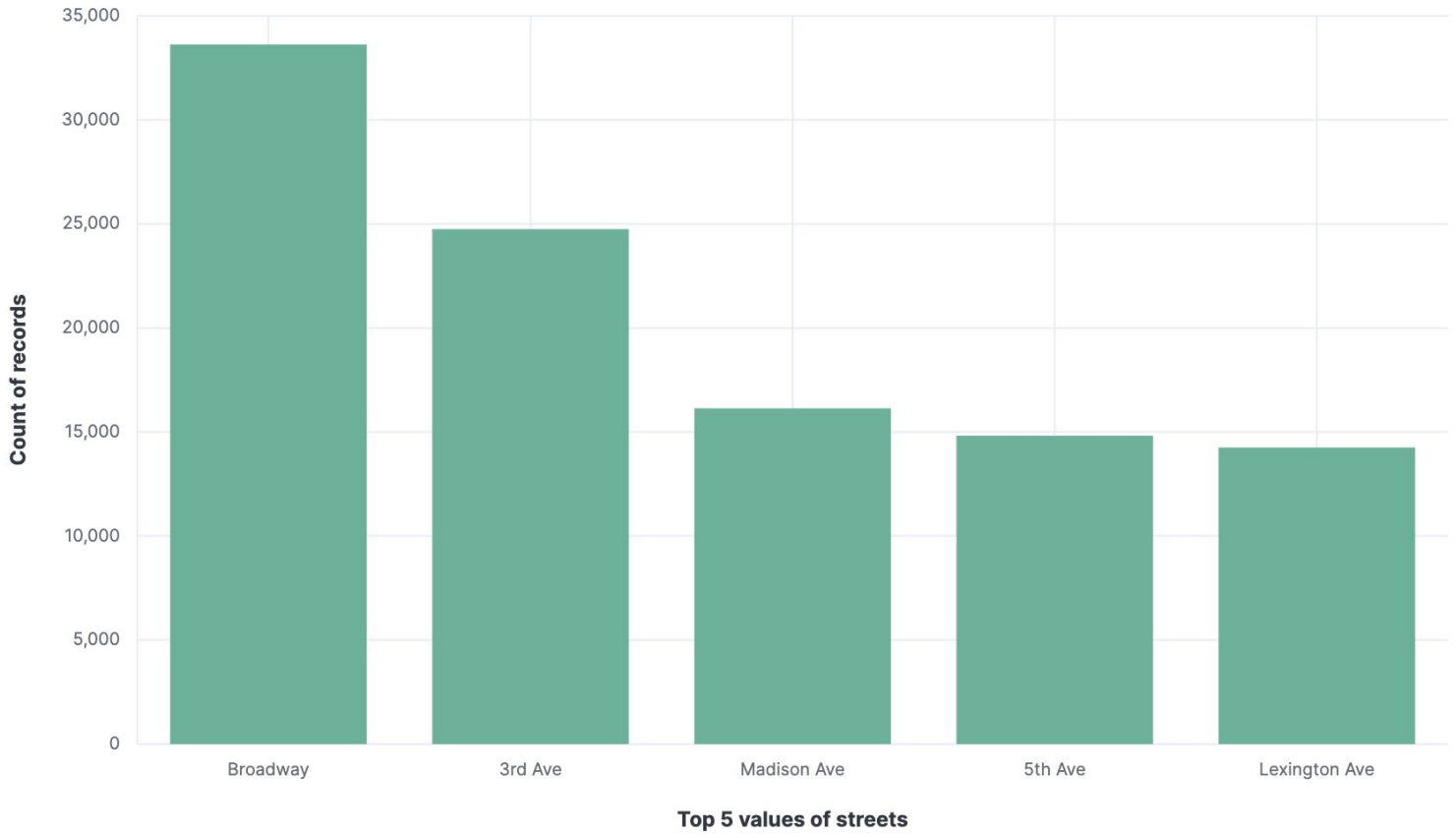
REPEATED OFFENDERS – Plate lookup

- **Plate lookup** with https://www.lookupaplate.com/new-york
- Top 3 are UPS trucks, very likely that others are delivery trucks as well (e.g. 47603MD is also a truck)

REPEATED OFFENDERS – violation times



REPEATED OFFENDERS – most common streets



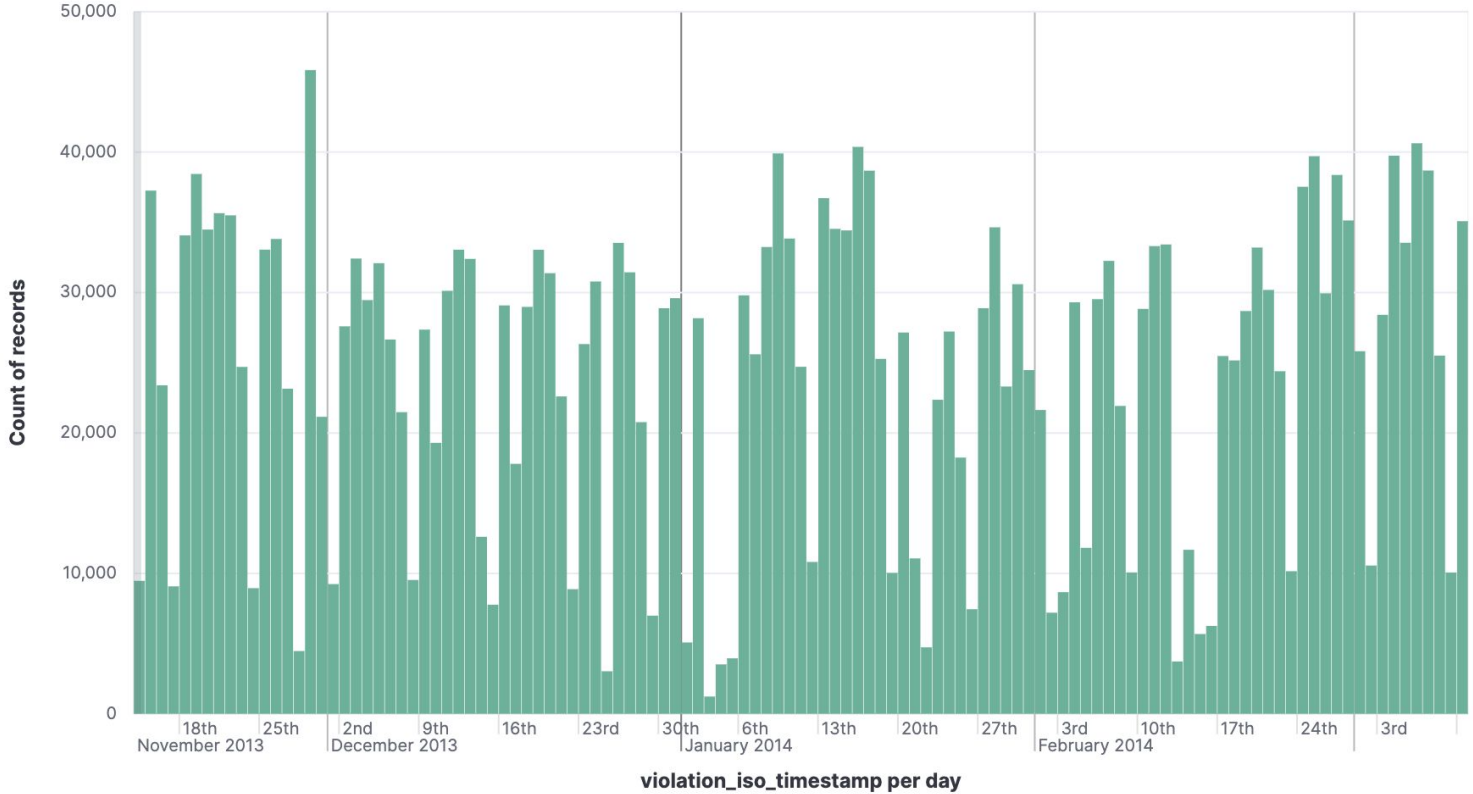
REPEATED OFFENDERS – Streets

- High Traffic Volume
- Commercial and Delivery Zones
- Strict Parking Regulations

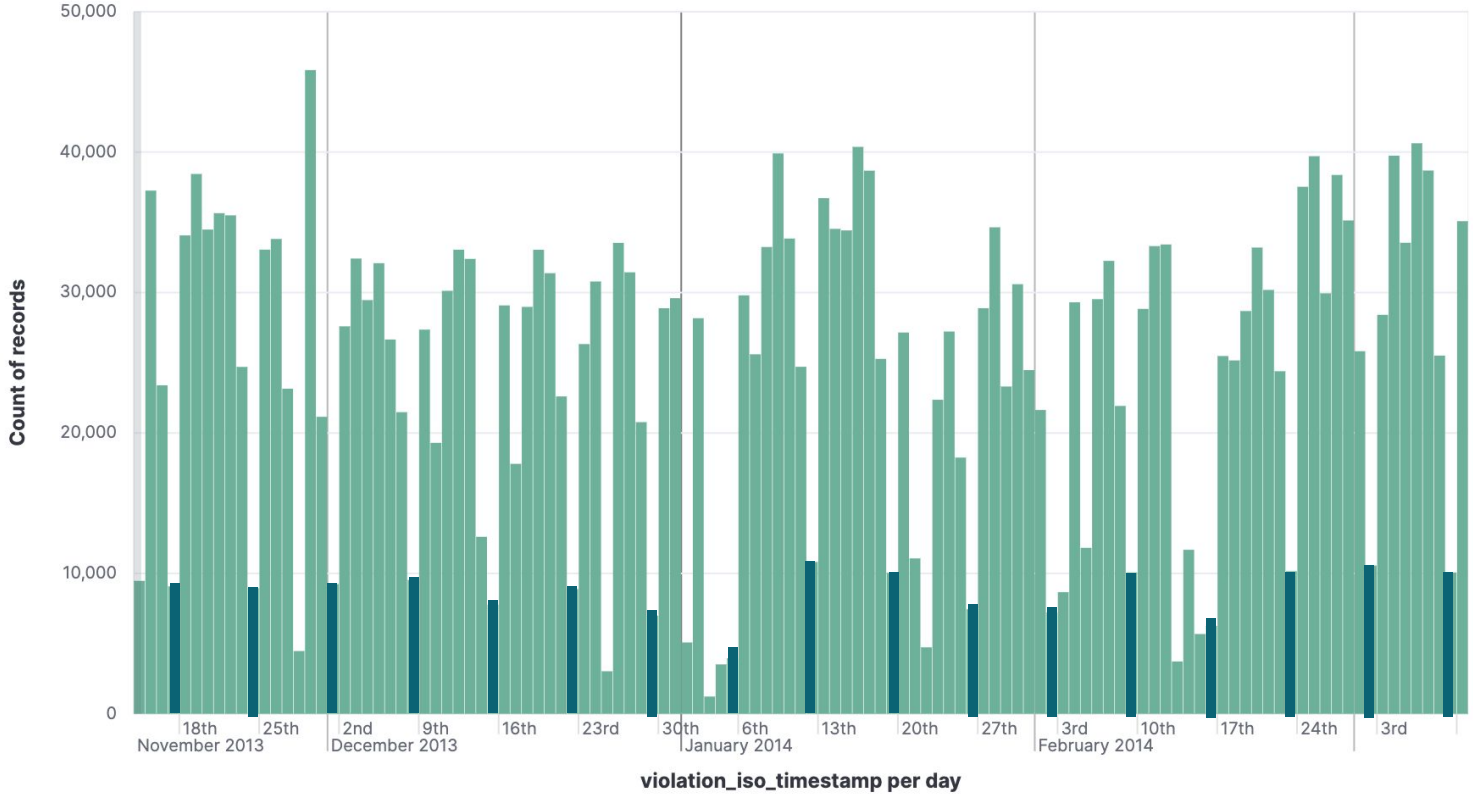
R2

Are parking violations more common during certain times of day or days of the week?

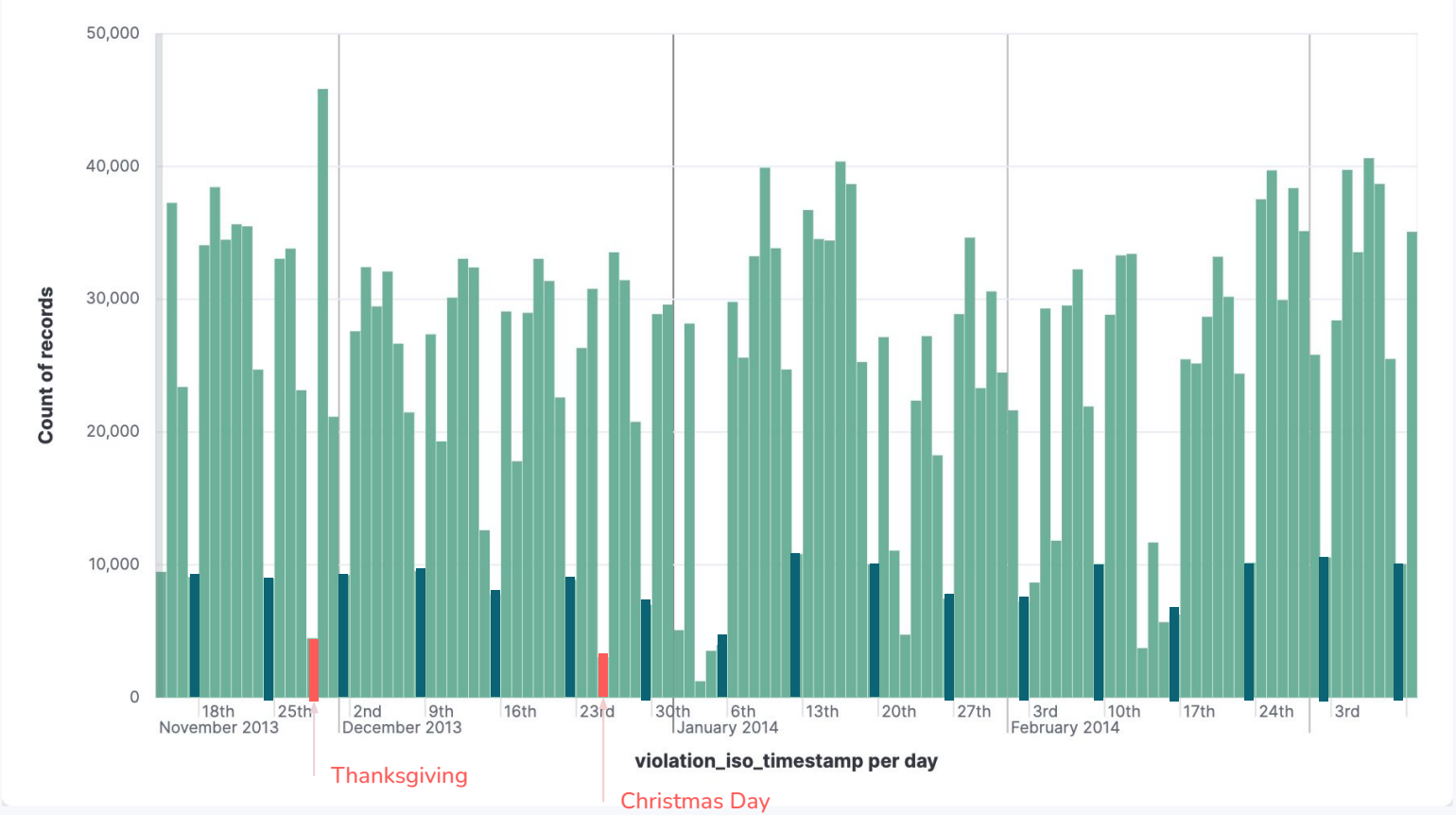
TIME ANALYSIS – 14th Nov - 14 March 2013



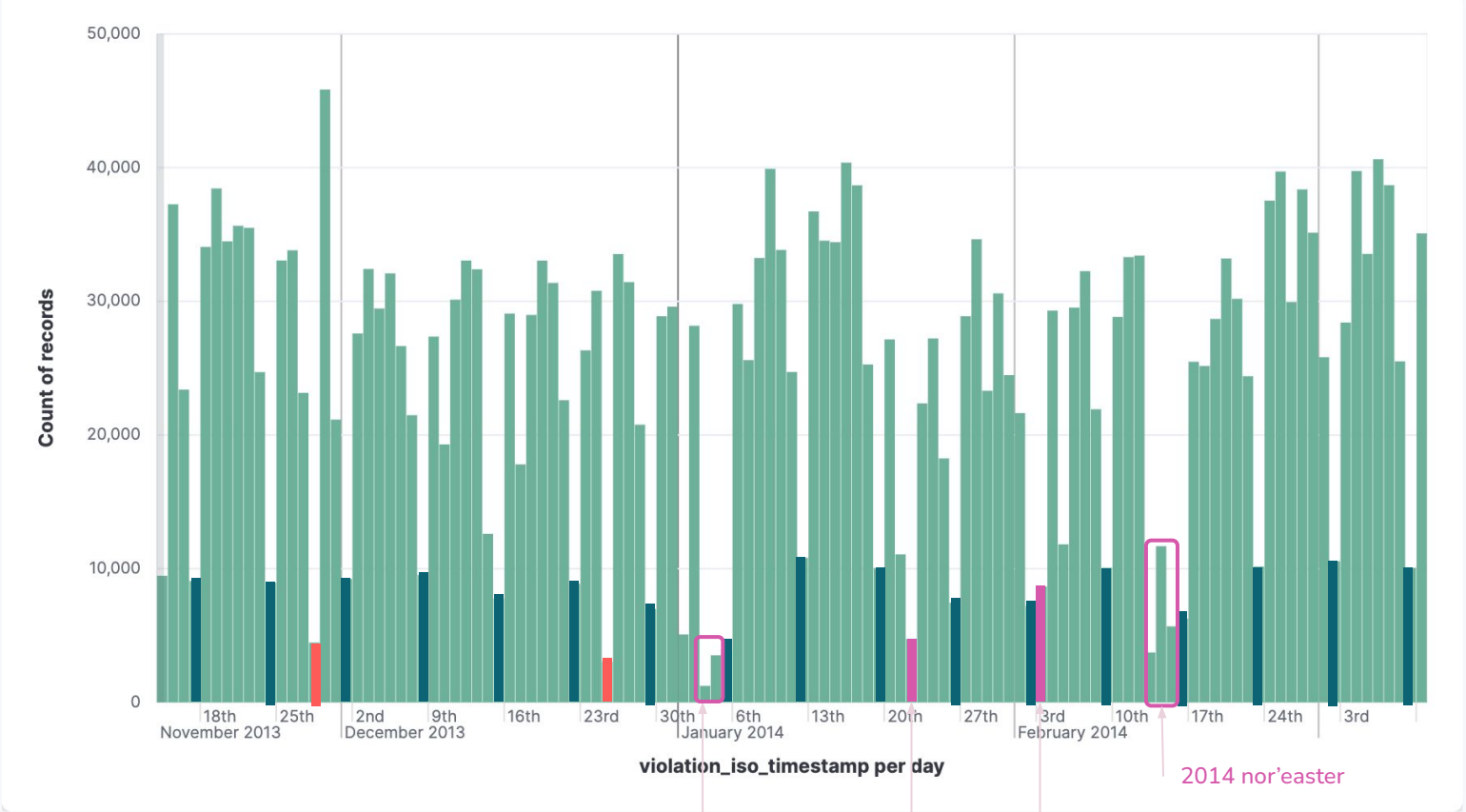
TIME ANALYSIS – Sundays



TIME ANALYSIS – Holidays



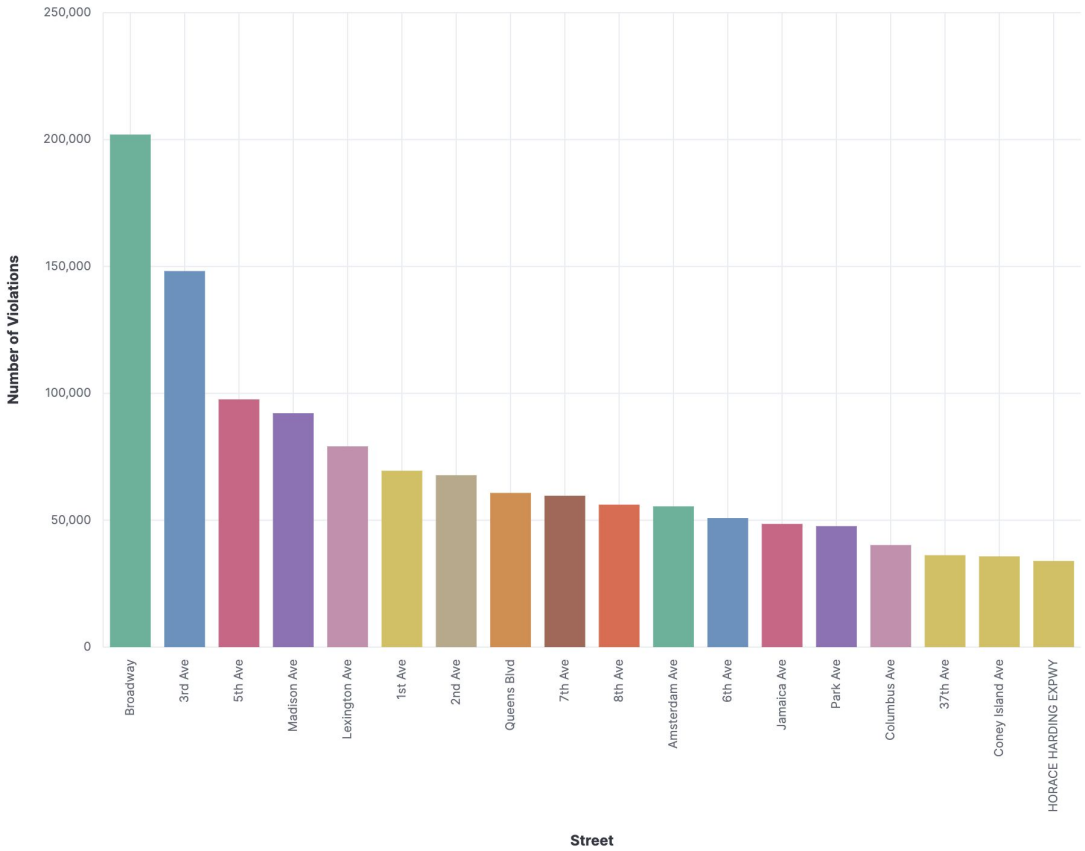
TIME ANALYSIS – Snow storms



R3

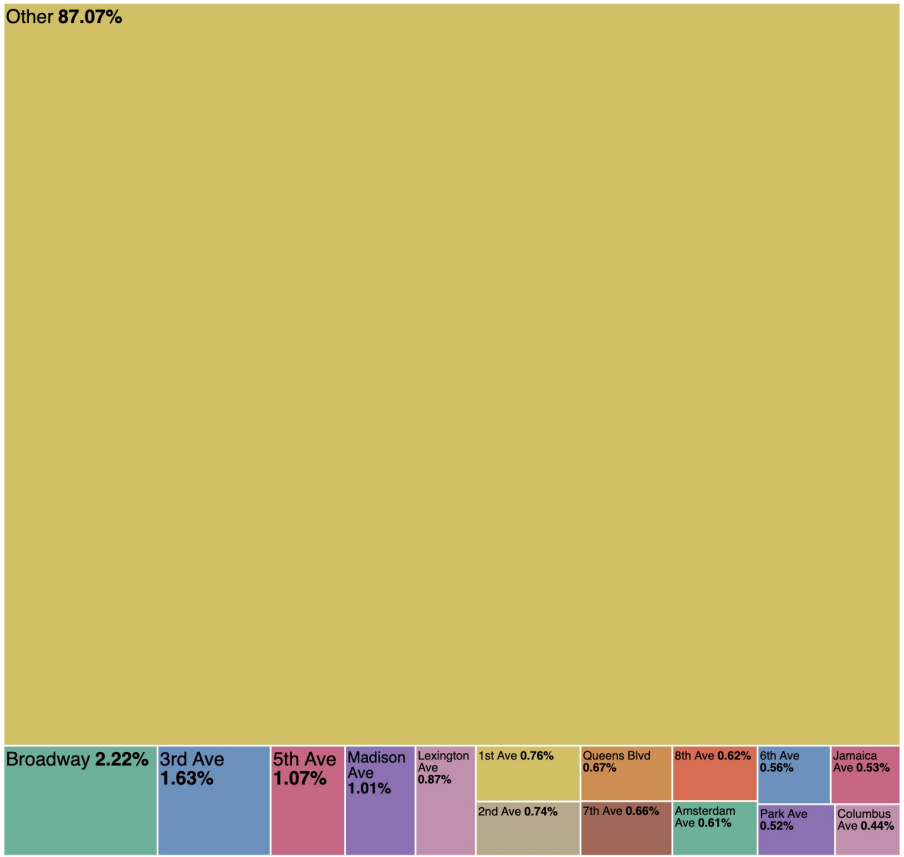
Define hotspots of parking violations.

HOTSPOTS - Streets

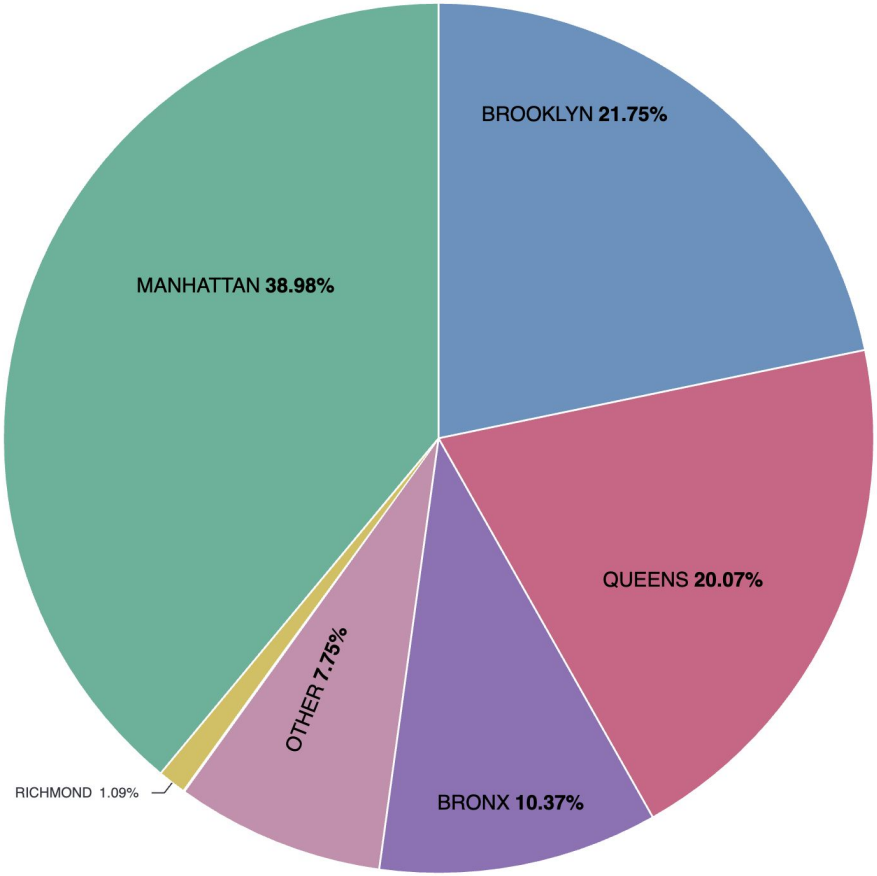


- Correlate with findings of RQ2
- More popular and important streets ➜ more violations

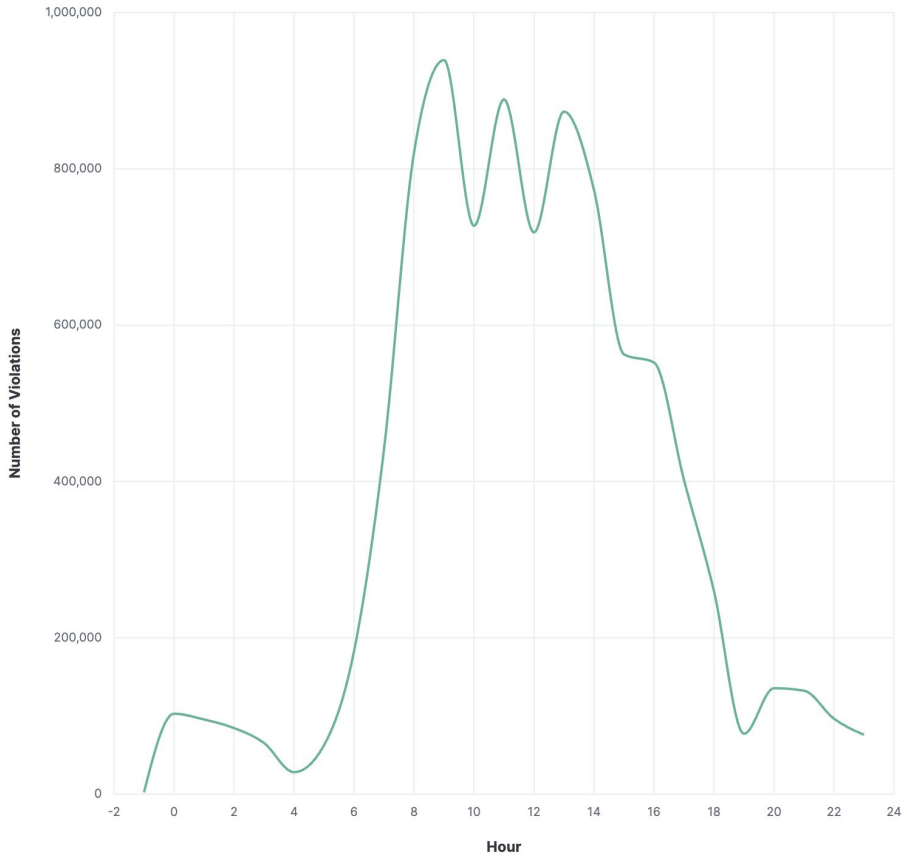
HOTSPOTS - Streets



HOTSPOTS - Counties



HOTSPOTS - Throughout Day

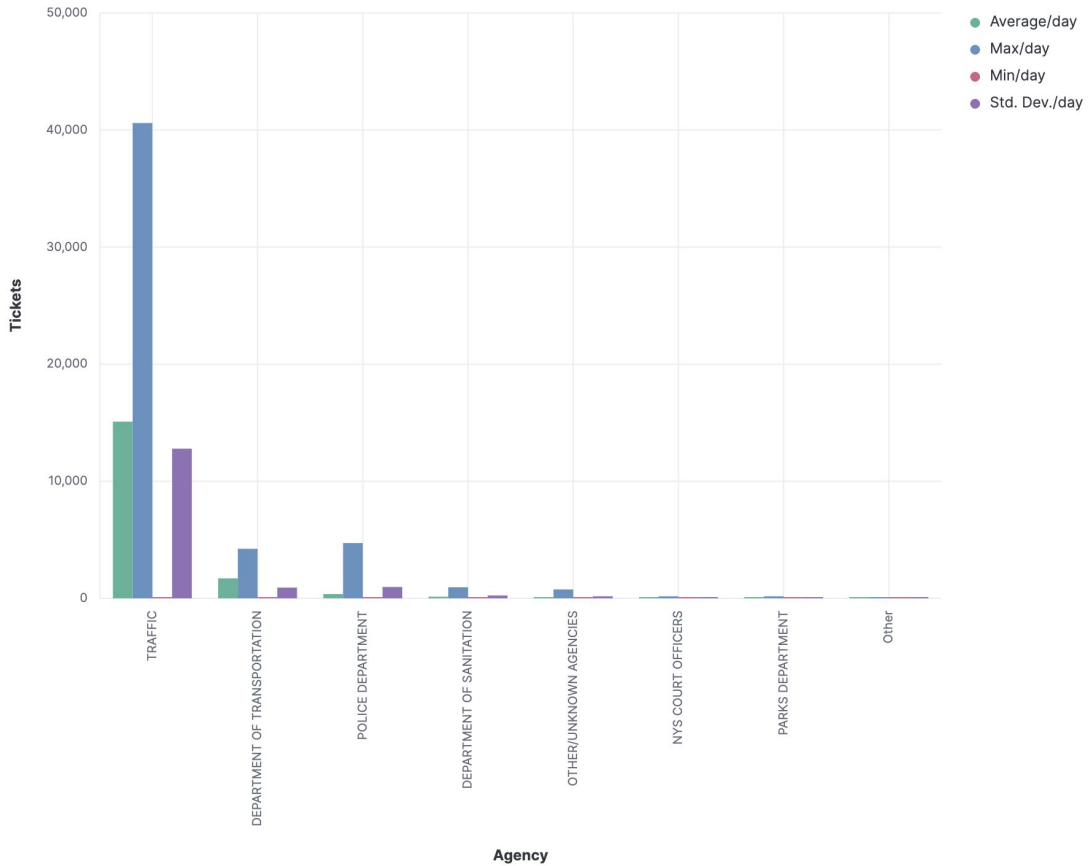


- More tickets during day ➜ Business hours
- Lunch break is visible

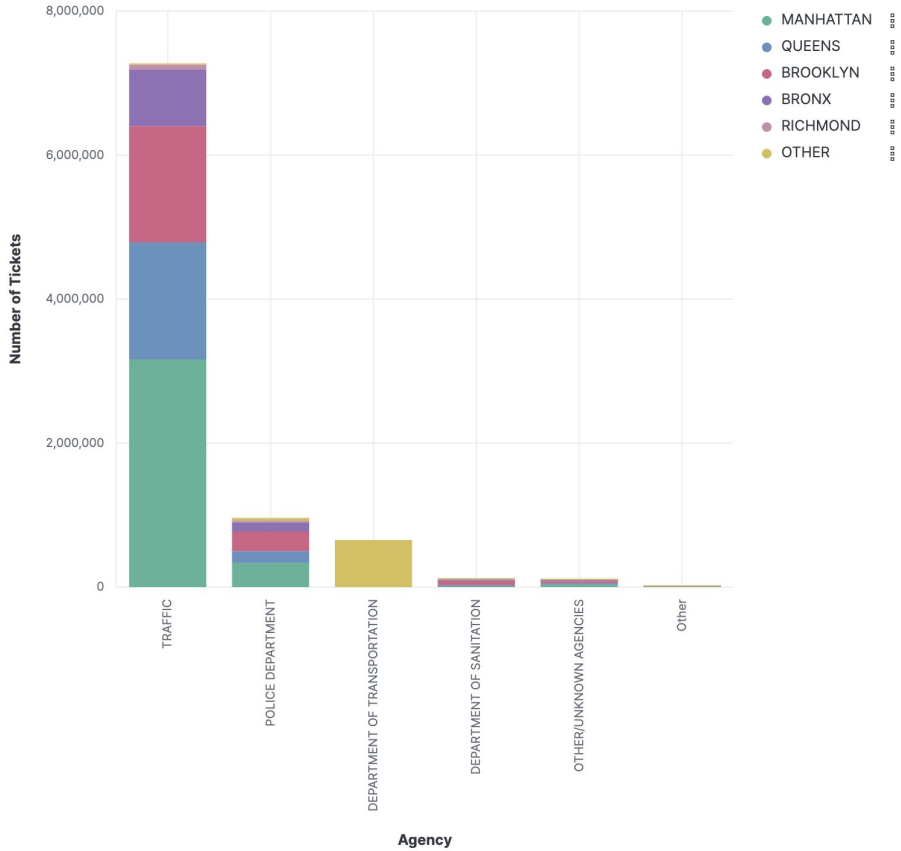
R4

Analyse the enforcement strength/severity of the agency.

AGENCY ENFORCEMENTS - Tickets per Agency



AGENCY ENFORCEMENTS - Tickets per Agency and County



Pitfalls and Problems

PITFALLS AND PROBLEMS

- Setup of the Tech Stack (Problems with the docker container)
  - Missing user in spark image
  - Missing jars in spark image
- Timestamps in Elasticsearch/Kibana
- Writing to Elasticsearch
  - Appending ➜ Needed to change mode to overwrite