

NYPD PUBLIC SAFETY DASHBOARD

Data Processing

ISLG

2023-03-06

✨ Today's Game Plan ✨

1. Workflow overview

2. Data collection set up

- OpenData
- Scraping
- Sent from agency

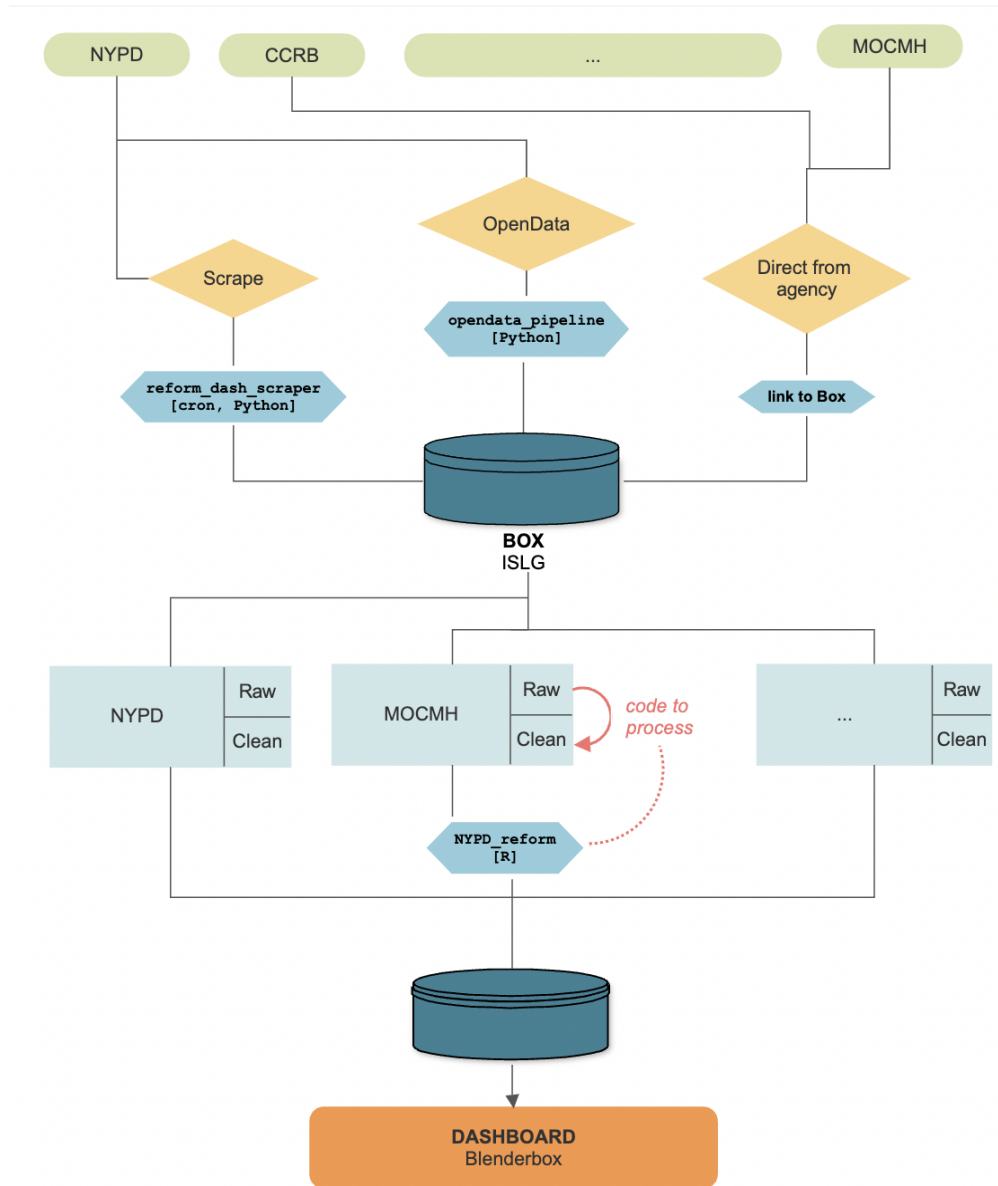
3. Reading/writing data on Box

- Boxr
- Box Sync
- Box.com

4. Data storage folder structure

- Agency
- Bucket?
- Metric?

5. Next time: Code repository and Code reviews



Tools & Software

- Box / [Box Sync](#)
 - We will use Box to store raw and clean data. [\[here\]](#) No data will be stored in Git repository.
- Git / GitHub
 - NYPD_Reform will be our team's primary codebase, which we will use to track and review scripts [\[here\]](#)
 - reform_dash_scraper is separate repository used for automated scraping. [\[here\]](#)
- R / RStudio

Overview of process

- Data collection streams:
 1. OpenData (Annie)
 2. Scraped via cron job (Annie)
 3. Agencies drop us data through Box folder (Jamel & Melanie)
- Wrangling (Jamel, Melanie & Annie)
- QA and Code review (Jamel, Melanie & Annie)
- Pushing to Blenderbox's database can be done either:
 1. programmatically or (Annie)
 2. with drag/drop UI (Jamel & Melanie)

An example: vehicle stops

[code]

What is Git and GitHub? Why Git?



	COMMENT	DATE
○	CREATED MAIN LOOP & TIMING CONTROL	14 HOURS AGO
○	ENABLED CONFIG FILE PARSING	9 HOURS AGO
○	MISC BUGFIXES	5 HOURS AGO
○	CODE ADDITIONS/EDITS	4 HOURS AGO
○	MORE CODE	4 HOURS AGO
○	HERE HAVE CODE	4 HOURS AGO
○	AAAAAAA	3 HOURS AGO
○	ADKFJSLKDFJSDKLFJ	3 HOURS AGO
○	MY HANDS ARE TYPING WORDS	2 HOURS AGO
○	HAAAAAAAANDS	2 HOURS AGO

AS A PROJECT DRAGS ON, MY GIT COMMIT MESSAGES GET LESS AND LESS INFORMATIVE.

- a popular Version Control System (VCS)
 - git creates a "snapshot" of your code and allows you manage your code history
- **GitHub** is the cloud-based service that helps us keep track of Git repositories
- many people can work on chunks of code simultaneously within a project

Structure of NYPD_Reform Repository