MAPPING NYC HISTORICAL LIQUIDATIONS

Department of City Planning, Data Engineering Team
City of New York, NY
Amanda Doyle — Director of Enterprise Data Management

coding it forward >

NYC

ALI THURSLAND
Duke University ‘24
Computer Science B.S.

DEA BARDHOSHI
UC Berkeley ‘23
Data Science B.A.

PROBLEM:

- NYC planners want to consider historical trends in capital spending by location when deciding where to invest in future infrastructure projects.
- No such dataset exists to inform these decisions.

cif>

SOLUTION: HISTORICAL CAPITAL SPENDING

- This data product aggregates historical liquidations of the NYC capital budget with geospatial information where possible
- **What is the potential impact?**
  - Identify areas of underinvestment as areas of focus for future investment
  - Improve access to green spaces and recreational areas by helping Parks prioritize where to close the walking-to-park (WTP) gap
  - Provide a means of analyzing what resiliency spending looks like in the wake of natural disasters (i.e., post-Sandy)
  - … and more!

cif>

KEY TERMS

- **capital project:** a project that costs $35,000 or more and has a lifespan of 5+ years (i.e. fire trucks, bridges, and sewers)
- **historical liquidations:** checks issued by NYC for capital projects over time
- **geospatial information:** information that allows data to be mapped to a location (i.e. coordinates or building footprints)

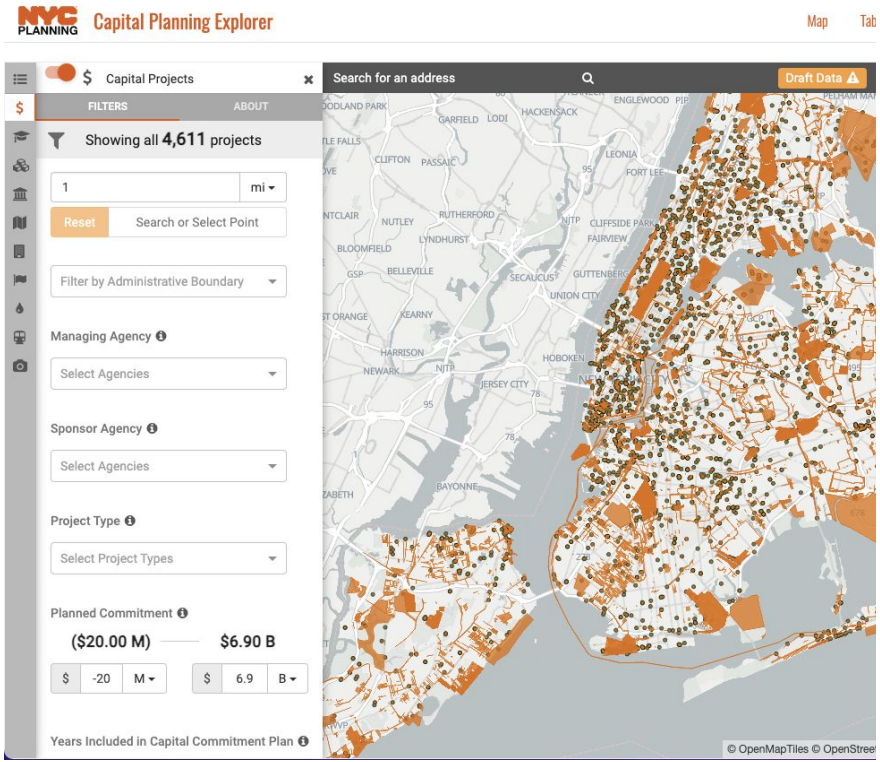
cif>

CHECKBOOK NYC

- Open-source dataset and interactive tool
- Tracks every check issued by NYC between 2010 and present
- **1,990,934** checks
- **$133.6B** in capital spending
- Rows contain **no geospatial data**

CAPITAL PROJECTS DATABASE (CPDB)

- Current + future NYC capital projects → subset containing geospatial information
- CPDB versions 2017-present
- Projects are categorized as Fixed Asset, ITT, Vehicles and Equipment, Lump Sum or None



GOALS

- Categorize Checkbook NYC records as we do in CPDB
    - Capital Projects that are categorized as Fixed Assets are of particular interest
- Join Checkbook NYC to CPDB to get geospatial data
- Create a data pipeline for historical capital spending

cif>

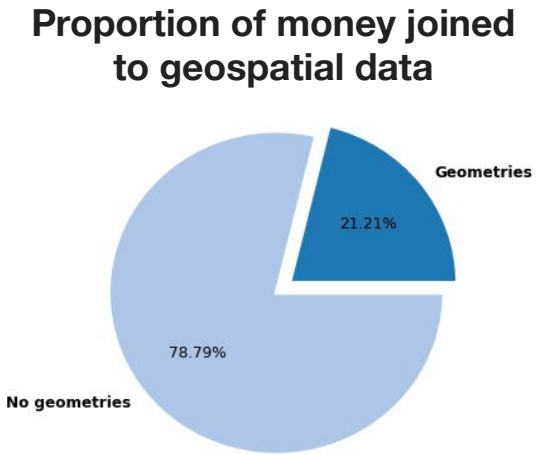
DATA PROCESSING

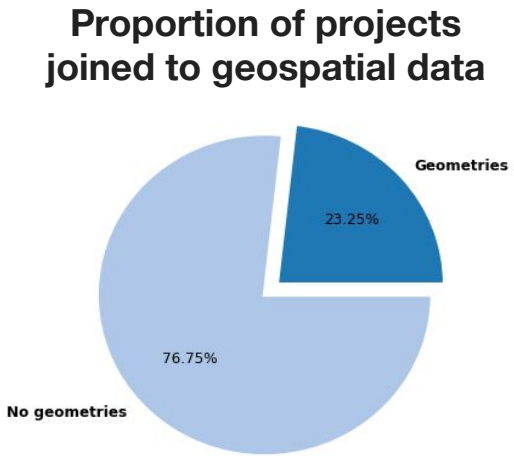
- **16,687** rows
  - Aggregated individual checks to the project level
  - one row = one capital project
- Added geospatial data from CPDB to capital projects where possible
- Assigned categories to projects using the same CPDB logic
  - Fixed Asset, Lump Sum, ITT Vehicles and Equipment, or None
  - Maximize number of projects categorized as Fixed Asset

cif>

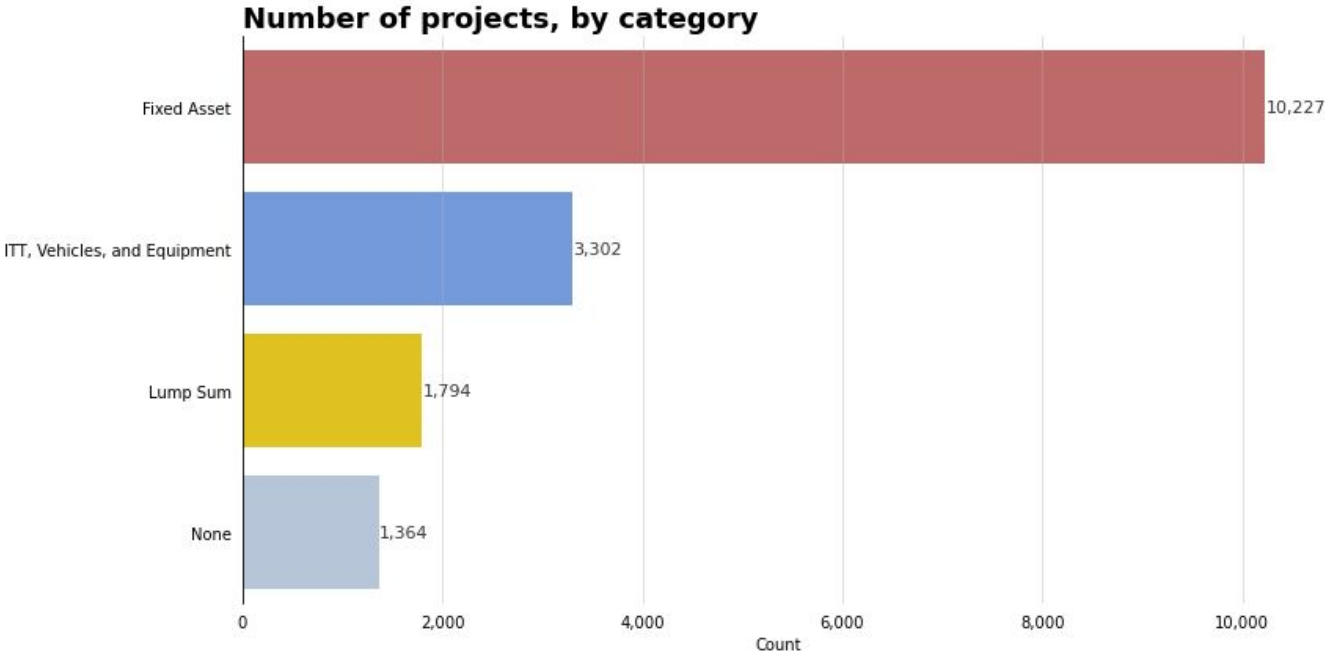
RESULTS OF JOINING CPDB TO CHECKBOOK

- Number of projects mapped to geospatial data
  - **3,880**
- Amount of money mapped to geospatial info
  - **$28.9B**



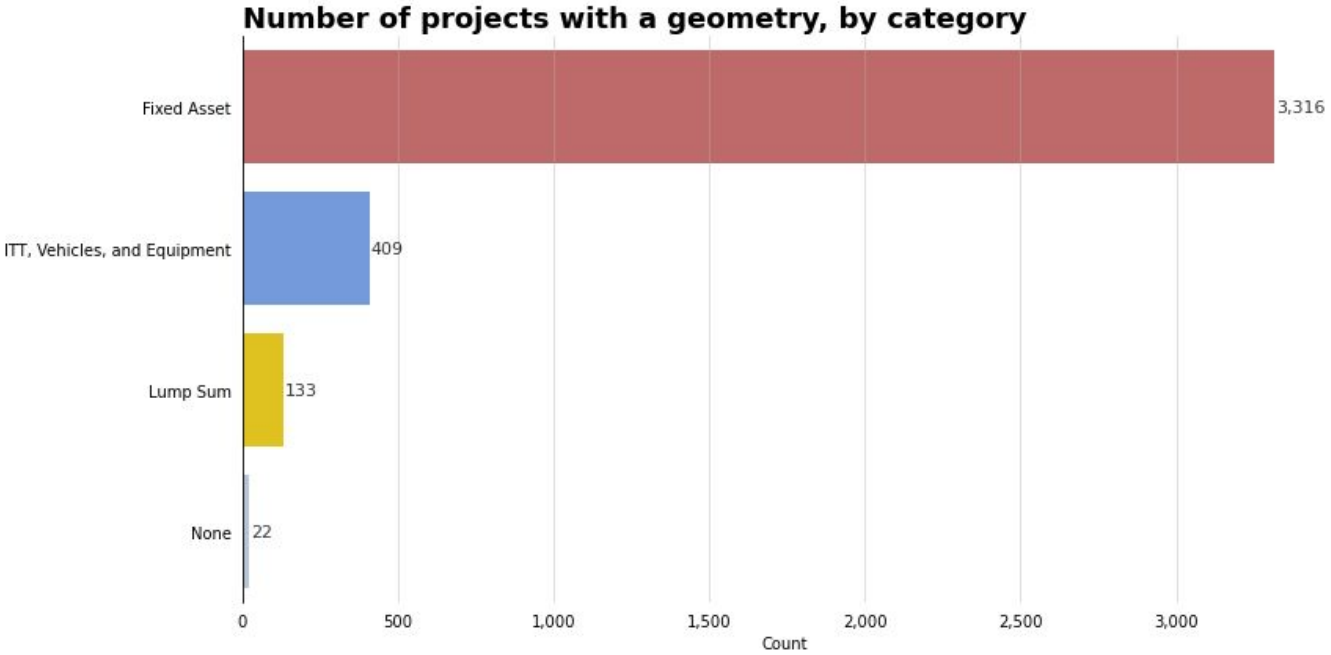


CHECKBOOK NYC CAPITAL PROJECTS: CATEGORIZATION



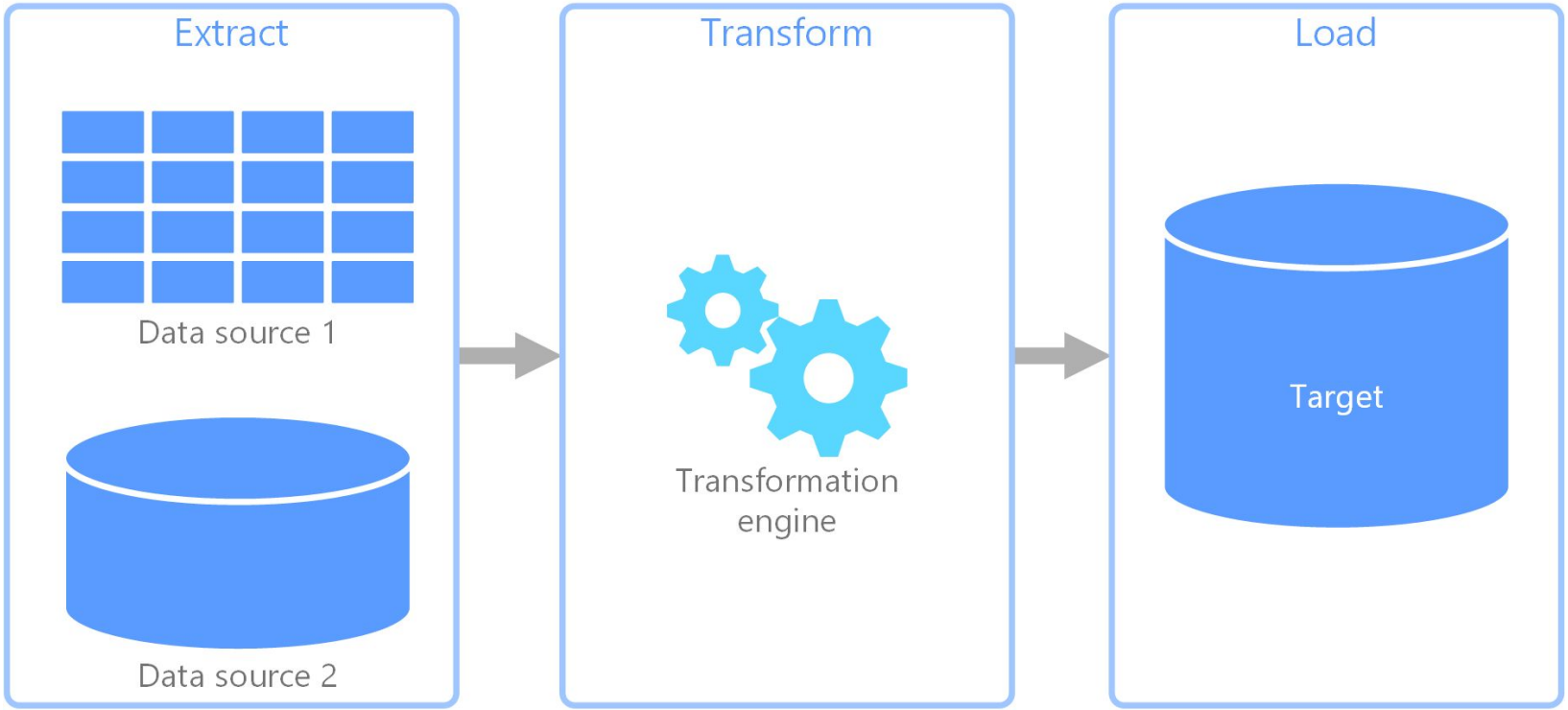
92% of projects were assigned a category
61% of projects were assigned to “Fixed Asset”

RESULTS OF ADDING GEOSPATIAL DATA, BY CATEGORY



85.4% of projects mapped to geospatial information categorized as ‘Fixed Asset’

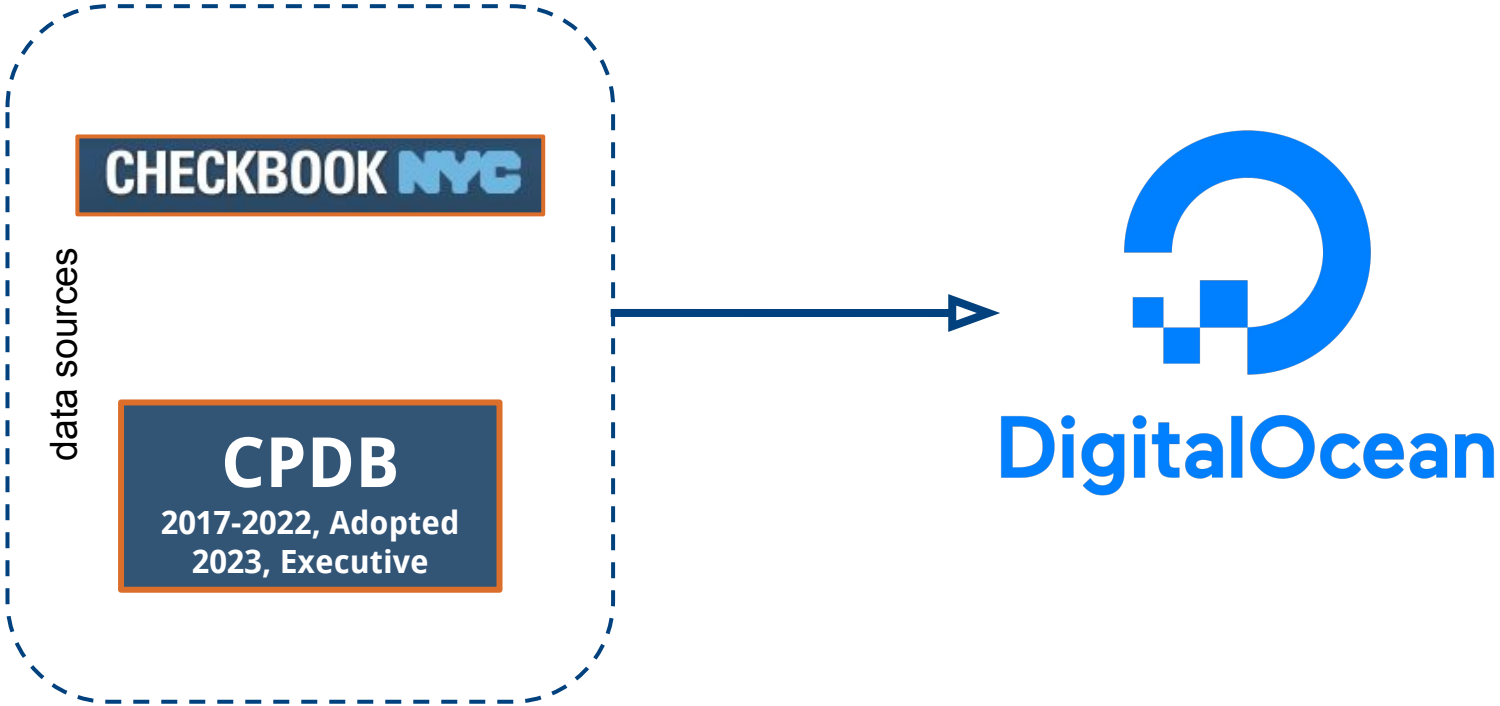
WHAT MAKES A DATA PIPELINE?



cif>

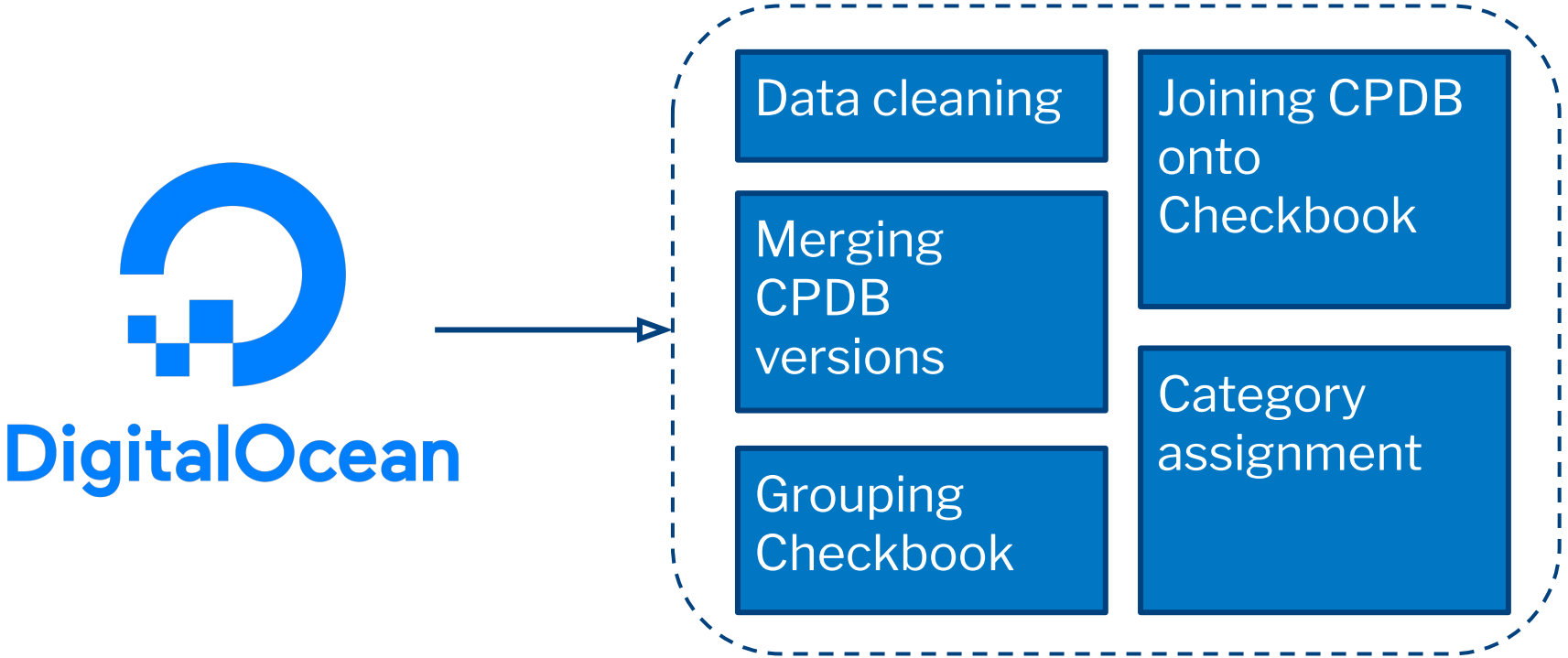
[source]

EXTRACTION



cif>

TRANSFORMATION



LOADING

Historical capital spending dataset

- each row is a capital project
- capital projects have geospatial data where possible

cif>

WHY DOES THIS MATTER?

- The Historical Capital Spending data product will:
  - Empower planners to draw insights from historical capital expenditures at the community level
  - Steer agencies to allocate funds to historically underinvested neighborhoods
  - Increase transparency into capital spending trends by neighborhood

cif>

POSSIBLE EXTENSIONS

- **Improving source data:**
  - Using APIs to streamline data ingestion
  - Layering in other databases as sources for locations
- **Enriching location information:**
  - Use name matching to extract location mentions from descriptive text fields
- **Making data publicly accessible:**
  - Publishing data, creating interactive tool

cif>

KEY TAKEAWAYS

- **What skills does one need to be a data engineer?**
  - Technical tools, ability to communicate effectively with stakeholders, data storytelling, resourcefulness
- **What skills does one need to work for city government?**
  - Working collaboratively with people across disciplines who share a common passion

cif>

THANK YOU!

- **Special thanks to:**
  - Amanda Doyle, Damon McCullough, Finn van Krieken and Alex Richey
  - Rachel, Yuyang and Ariana at CIF
  - Our CIF mentors, Elisa Lee and Maria Filippelli
  - The entire CIF 2023 cohort
  - Everyone tuning in!

cif>

APPENDIX

HOW DID WE ASSIGN CATEGORIES?

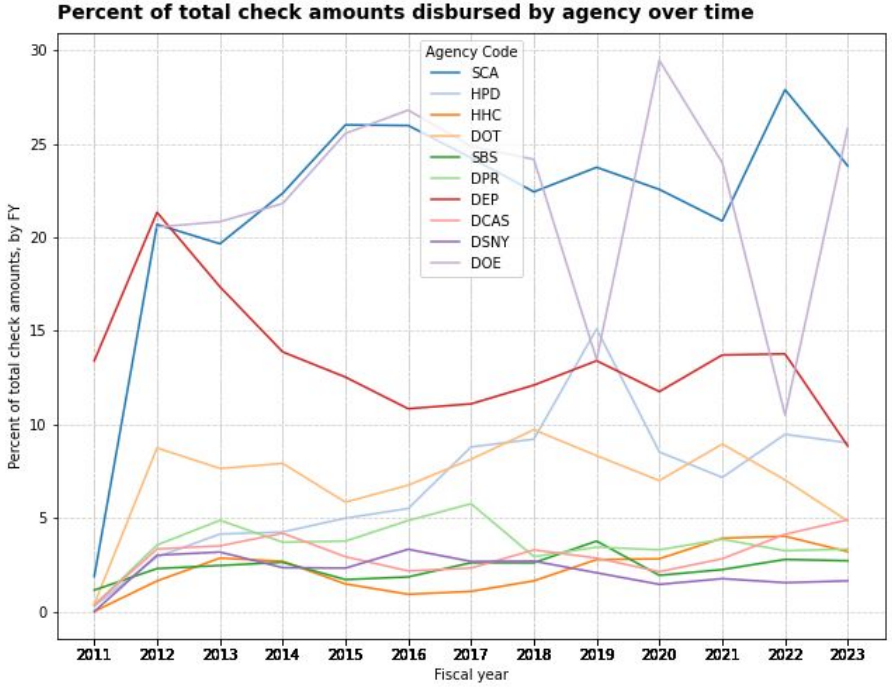
- *Goal:* maximize the number of Fixed Assets because those are the projects most likely to have physical locations
- *How:* use keyword mentions in Budget Code, Contract Purpose to assign categories to projects, then assign final category, prioritizing Fixed Asset keywords

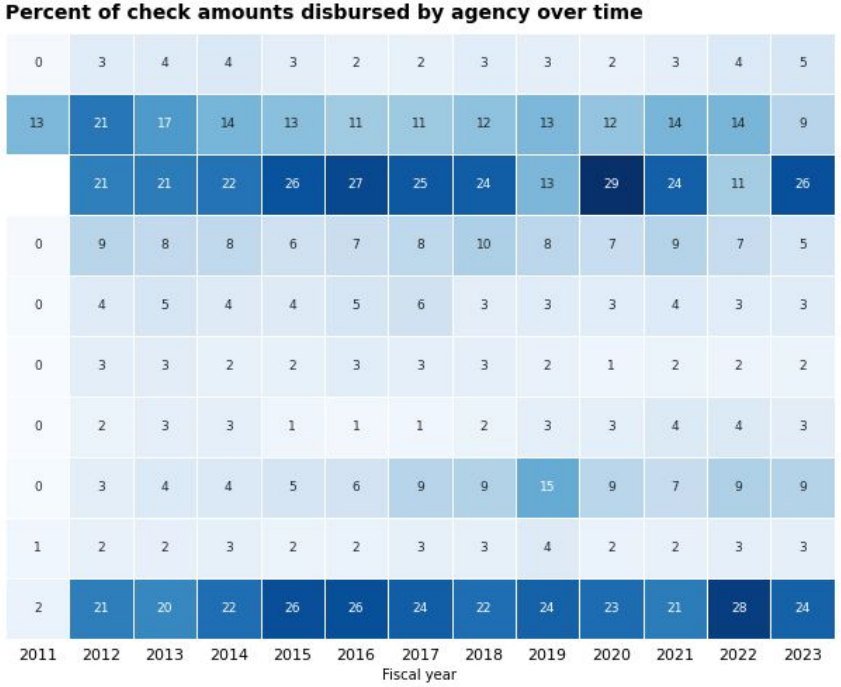
Agency	FMS ID	Budget Code	Contract Purpose	bc_category	cp_category	cpdb_category
Department of Information Technology and Telec...	858PSAC1FAC	S911 (DOITT: SYSTEM INTEGRATION SVCS FOR THE 9)	Emergency Communications Transformation Program	ITT, Vehicles, and Equipment	ITT, Vehicles, and Equipment	Lump Sum
Department of Citywide Administrative Services	858PW77BAT	BAT1 (BROOKLYN ARMY TERMINAL (140 58TH ST, BK)	Renewal No. 1 for the Citywide Mobile Wireless...	Fixed Asset	ITT, Vehicles, and Equipment	ITT, Vehicles, and Equipment
Department of Citywide Administrative Services	858PW77BAT	BAT1 (BROOKLYN ARMY TERMINAL (140 58TH ST, BK)	Renewal No. 1 for the Citywide Mobile Wireless...	Fixed Asset	ITT, Vehicles, and Equipment	ITT, Vehicles, and Equipment
Department of Citywide Administrative Services	858PW77QBP	QBP2 (QUEENS BOROUGH PRESIDENT OFFICE (120-55)	Maint, Repair, and Mod for Intellipath, Key Sy...	None	ITT, Vehicles, and Equipment	ITT, Vehicles, and Equipment
Department of Citywide Administrative Services	858PW77QBP	QBP2 (QUEENS BOROUGH PRESIDENT OFFICE (120-55)	Maint, Repair, and Mod for Intellipath, Key Sy...	None	ITT, Vehicles, and Equipment	ITT, Vehicles, and Equipment

final_category
ITT, Vehicles and Equipment
Fixed Asset
Fixed Asset
ITT, Vehicles, and Equipment
ITT, Vehicles, and Equipment

cif>

EXPLORATORY DATA ANALYSIS: CAPITAL SPENDING BY AGENCIES OVER TIME





PERCENTS PROJECTS AND MONEY MAPPED TO GEOSPATIAL INFORMATION

	Percent Mapped By Category	Percent Money Mapped By Category	Category
0	0.324240	0.497146	Fixed Asset
1	0.123864	0.053494	ITT, Vehicles and Equipment
2	0.074136	0.070974	Lump Sum
3	0.016129	0.011344	Null

cif>