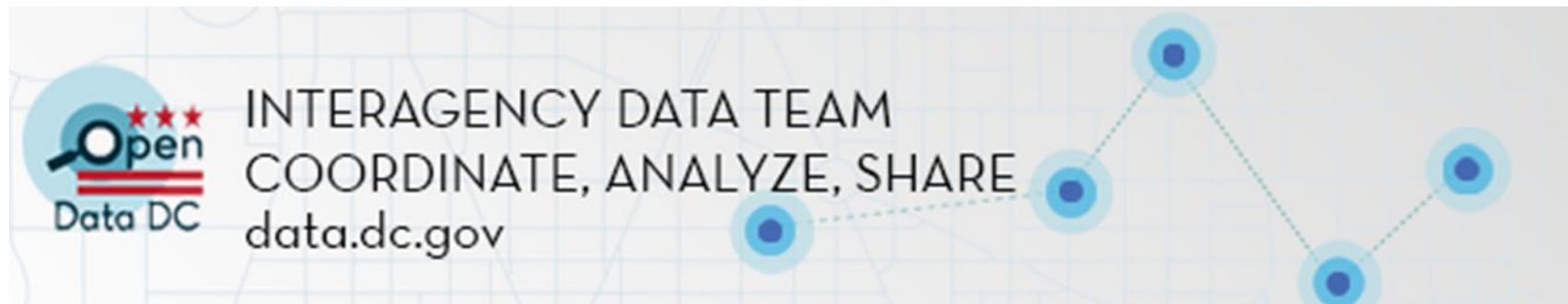


Interagency Data Team

October 16, 2019



Welcome, agenda Data team news

Michael Bentivegna

**Data Visualization and Analysis Program Manager
Office of the Chief Technology Officer**

Agenda

Welcome, News & Updates

Michael Bentivegna, Data Visualization and Analysis Program Manager
Office of the Chief Technology Officer

Beginning the 2020 Enterprise Dataset Inventory

Eva Reid, Data & Visualizations Senior Analyst
Office of the Chief Technology Officer

Mobility Data Specifications (MDS) & Dockless Data within the DC Data Lake

James Graham, GIS and Applications Manager
District Department of Transportation

Data Report

Mario Field, Data Curation & Governance Program Manager
Office of the Chief Technology Officer

Business Intelligence - Tableau Updates



**Tableau
 Internal
 Production @
 version
 2019.1.2**

**Next update:
 2019.2.4 in late
 Oct. Will
 include New
 License Model**



**Tableau Public
 Production @
 version
 2019.1.3**

**9 Agency
 websites in
 production**

**3 additional
 projects in
 progress**



**Tableau
 Test/Dev
 version
 2019.2.4**

**Login based license
 management(LBLM)**

**Tableau server for R
 and Rserve testing**

**MIT kerberos ticket
 authentication for
 Hadoop**

**Whitelist DDOT's
 Wiki page & MapBox**



**Tableau Site
 Administrator
 quarterly
 meeting**

**Next meeting:
 Oct. 22, 2019**

Business Intelligence - MicroStrategy Updates



MicroStrategy Production @ version 10.11

Version 2019.3 will be upgraded in the last week of October with new physical server implementation.



Lower MicroStrategy Environments 2019.3 (11.3)

Enterprise/Self
Service

Development
Test
QA



MicroStrategy Public/SharePoint

DGS SharePoint
live

DBH SharePoint in
progress

Guidelines are
available now



MicroStrategy Stake Holders Meeting

Next meeting will be tentatively schedule for November 13th

@ OCTO, 200 I
Street SE.

Geographic Information Systems - DC GIS

Office of the Chief Technology Officer

**ArcGIS Server
Production @
version 10.6.1**

**ESRI Desktop
License Manager to
10.7.1**

**ArcGIS Portal @
version 10.7.1**

AGOL new [Data Stories](#)
gallery on
opendata.dc.gov

ArcGIS Hub Initiatives -

DMPED has OZ
Marketplace

DDOT has ParkDC

HSEMA will have
Special Events

DDOT underway...

**Streaming
services**

ESRI GeoEvent
@10.6.1

Apache Kafka DSP
Streaming data

**Data Cyclomedia /
Globespotter
Imagery (360
degree, spatially
referenced
imagery) [Online](#)
now has data
available from
2013, 2015, 2016,
and now end of
2018.**

DC Government



Tableau Training

Introductory training

November 13 and 14 (Full)

December 4 and 5 (Full)

New dates coming soon...

Intermediate training

November 19 and 20 (One seat left!)



<https://octo.in.dc.gov/service/tableau-training-courses>



MicroStrategy Training

Introductory training

- November 26 and 27
- December 17 and 18
- January 28 and 29
- February 25 and 26
- March 17 and 18

Architect/Developer training

- First training next week
- By invitation only



<https://octo.in.dc.gov/service/microstrategy-introductory-training>



GIS Training

230 Overview DC GIS Services: November 6th

233 ArcGIS Online: November 13th

232 ArcGIS Desktop (Pro): November 20th

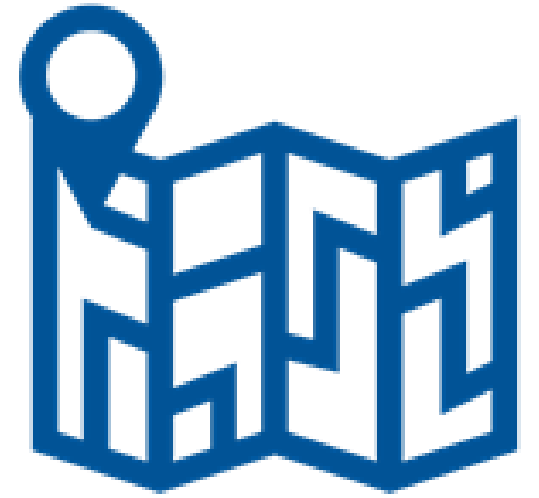


- Requires access to PeopleSoft
- More information: <https://octo.in.dc.gov/service/enroll-gis-training>

GIS / Data Science User Meeting

•GIS – November 21st (Via WebEx)
contact Eva.Reid@dc.gov
Topics: Visualization; Cartography

•Data Science @ DC – Today 3:45 pm
Location: OCTO, Room 5009
contact
Tomash.Bukowiecki@dc.gov
Topic: Rent Stabilization and Housing
Insecurity



Enterprise Data Inventory (EDI)

- About the EDI
- New Requirement: Level 3 and 4 datasets
- Dates to remember
- Expectations and next steps for EDI

Enterprise Data Inventory 2019

- DC Data Policy (Mayor's Order 2017-115, amended 2018-050) requires agencies under mayoral purview to record metadata for any enterprise dataset
- Enterprise dataset:
 - Directly supports the mission of one or more public bodies.
 - Is stored in a named information technology system that is accessible to multiple workforce members

Enterprise Data Inventory 2019

- Second Enterprise Data Inventory (EDI) published on March 10, 2019
- 1779 enterprise datasets (increased 7%)
- 73% of Level 0 (Open) datasets recorded in 2019 are in publicly-available Open Data Portal

Enterprise Data Inventory 2020

- Updated application
- New requirement to include Sensitivity Flag to all Level 3 and Level 4 records (existing and new)
- New OCTO Liaisons

EDI: Dates to Remember

Date	Activity
September 4, 2019	EDI open for editing
November 11	Recommended date for agency general counsel review to begin
December 10	All changes to agency inventory must be complete
January - March 2020	OCTO Quality Control, CA office review, EDI Report development
March 10, 2020:	Public distribution of the full Inventory

EDI: Sensitivity Flag

- Existing and new datasets classified at Levels 3 and 4 must now contain a sensitivity flag, indicating the specific reason for the classification
 - Meets the requirement laid out in the DC Data Policy
- Sensitivity flag is in **addition** to the Description (but they should match)
- Sensitivity related to Personally Identifiable Information (PII), or other reasons data might be classified as confidential

EDI: Sensitivity Flag

Please select all Sensitivity Flags that apply.

- PII
 - HIPPA
 - RFPA
 - FERPA
 - COPPA
- Other
 - Attorney-Client Privileged Information
 - Criminal Justice Information
 - Critical Infrastructure Information
 - Federal Tax Information (FTI)
 - For Official Use Only (FOUO)
 - Law Enforcement Sensitive
 - Legally privileged
 - Sensitive but Unclassified

Personally Identifiable Information (PII)

- National Institute of Standards and Technology (NIST) Guide to Protecting the Confidentiality of Personally identifiable Information (PII) (Special Publication 800-122)
- Definition of PII: Any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information. [Read More....](#)

Health/Medical Privacy

HIPAA (Health Insurance Portability and Accountability Act of 1996) establishes standards for the electronic exchange, privacy and security of health information. The Privacy Rule applies to any health care provider who transmits health information in electronic form.

Definition of Protected/Personal Health Information (PHI): Individually identifiable health information that is transmitted or maintained in any form or media by a covered entity. Individually identifiable health information related to:

- An individual's past, present or future physical or mental health or condition
- The provision of healthcare to an individual
- The past, present or future payment for the provision of healthcare to an individual.

Financial Privacy

Right to Financial Privacy Act (RFPA) prohibits financial institutions from granting unrestricted access to customers' financial records except under specific exceptions. The Payment Card Industry (PCI) Security Standards Council establishes rules and procedures for payment card data to ensure confidentiality of PII contained in its systems.

Student Privacy

The Family Educational Rights and Privacy act (FERPA) Protection of education records, defined as those records that contain information directly related to a student and which are maintained by an educational agency or institution or by a party acting for the agency or institution. FERPA prohibits a school from disclosing PII from an eligible student's education records to a third party without written consent unless one of the limited FERPA exceptions applies. FERPA also defines the term PII, (34 C.F.R. sec 99.3.) FERPA defines education records: records that are (1) directly related to a student and (2) maintained by an educational agency or institution or by a party acting for the agency or institution. /See



Expectations and next steps for EDI

- Enter data early!
- Identify new datasets: New programs? New staff?
- OCTO Liaisons are available to help!
 - See contact list: <https://octo.in.dc.gov/page/enterprise-data-inventory-data-policy>

Mobility Data Specifications (MDS) & Dockless Data within the DC Data Lake

**James Graham, GIS and Applications Manager
District Department of Transportation**

The data seemed so simple...
Hey, maybe we could reuse the code we already had?
Yeah, right...



DOCKLESS IN DC

Outline

Two horizontal dotted lines in the top right corner of the slide.

A brief history

Policy team's (evolving) data needs

New challenges

Dip into the 'Lake'

A decorative graphic at the bottom of the slide consisting of a thick red curved line and a thinner blue curved line below it.

History of Dockless



September 2017: Dockless Vehicle Demonstration Period Begins

Initial Need: GBFS (No Problem) Somewhat simple, daily summaries, count by Ward





General Bikeshare Feed Specification (GBFS)

This document explains the types of files and data that comprise the General Bikeshare Feed Specification (GBFS) and defines the fields used in all of those files.

Table of Contents

- [Revision History](#)
- [Introduction](#)
- [Files](#)
- [File Requirements](#)
- [Field Definitions](#)
 - [gbfs.json](#)
 - [system_information.json](#)
 - [station_information.json](#)
 - [station_status.json](#)
 - [free_bike_status.json](#)
 - [system_hours.json](#)
 - [system_calendar.json](#)
 - [system_regions.json](#)
 - [system_pricing_plans.json](#)
 - [system_alerts.json](#)

History of Dockless

September 2017: Dockless Vehicle Demonstration Period Begins

Initial Need: GBFS (No Problem) Somewhat simple, daily summaries, count by Ward

Summer 2018: Scooters!

New/More complex data: Uh-oh





Mobility Data Specification

The Mobility Data Specification (**MDS**) is a set of data specifications and data sharing requirements focused on dockless e-scooters and bicycles. Inspired by [GTFS](#) and [GBFS](#) the goals of the MDS are to provide API and data standards for municipalities to help ingest, compare and analyze *mobility as a service* provider data.

The **MDS** helps the City ingest and analyze information from for-profit companies who operate dockless scooters and bicycles in the public right-of-way. MDS is a key piece of digital infrastructure that helps cities and regulators such as Los Angeles Department of Transportation (LADOT) understand how dockless mobility operate.

Mobility providers are required to share data with LADOT as part of the City of Los Angeles' Dockless Mobility Permit. Standardizing data collection between different providers improves cooperation, innovation, and efficiency of the City's transportation network.

MDS is currently comprised of two distinct components:

- The [provider](#) Application Program Interface (API) was first released May 2018 to be implemented by mobility providers. When a municipality queries information from a mobility provider, the Provider API has a historical view of operations in a standard format.
- The [agency](#) API was first released in April 2019 to be implemented by regulatory agencies. The first implementation went live in February 2019. Providers query the Agency API when an event occurs, like a trip starting or ending.

History of Dockless

September 2017: Dockless Vehicle Demonstration Period Begins

Initial Need: GBFS (No Problem) Somewhat simple, daily summaries, count by Ward

Summer 2018: Scooters!

New/More complex data: Uh-oh

New/More complex questions: Double Uh-oh



Crossroads



Data are streaming/realtime

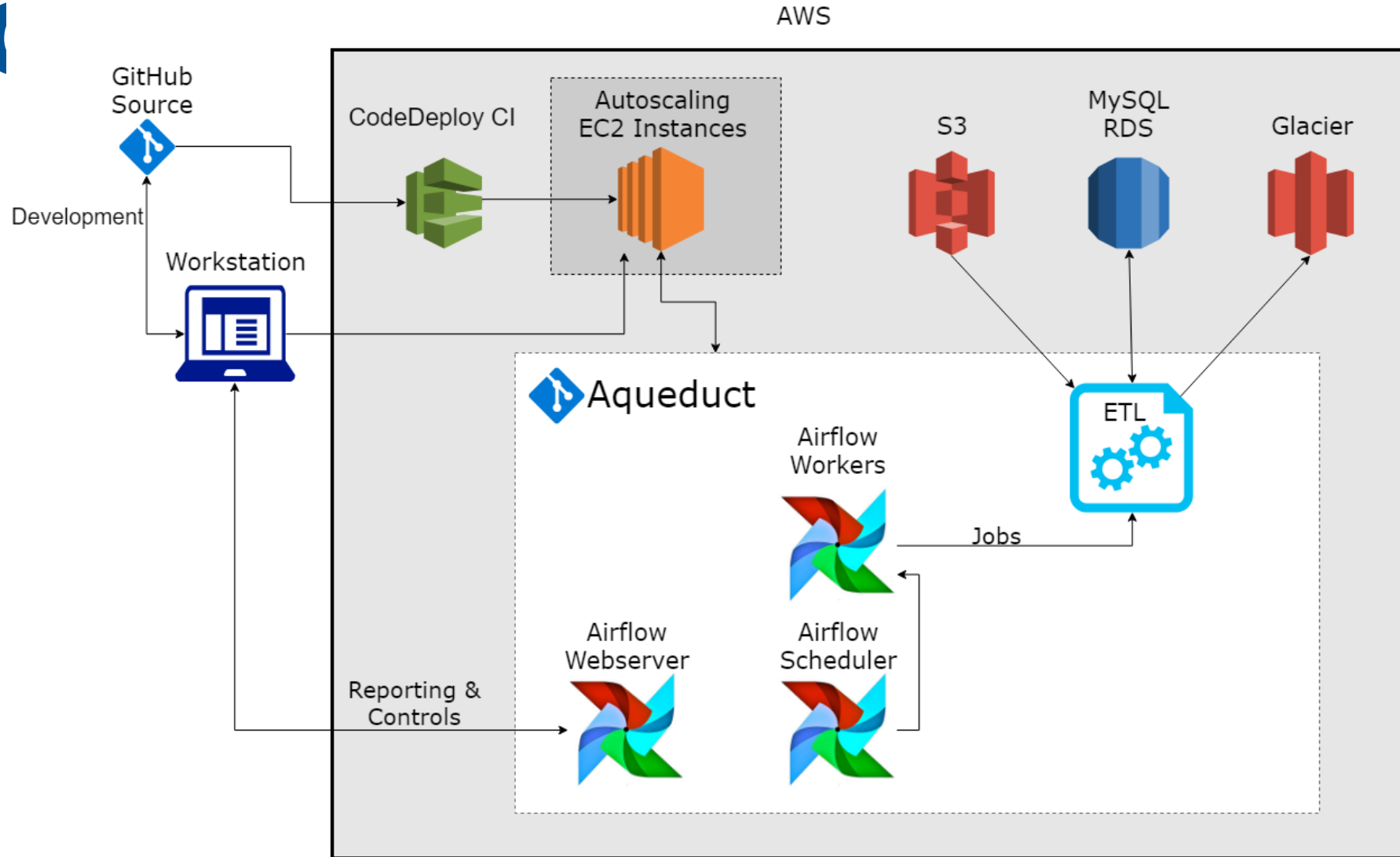
Query?

Storage?



LADOT's

Proc



New Challenges

Data are streaming / “Real-time”

Query?

Storage?

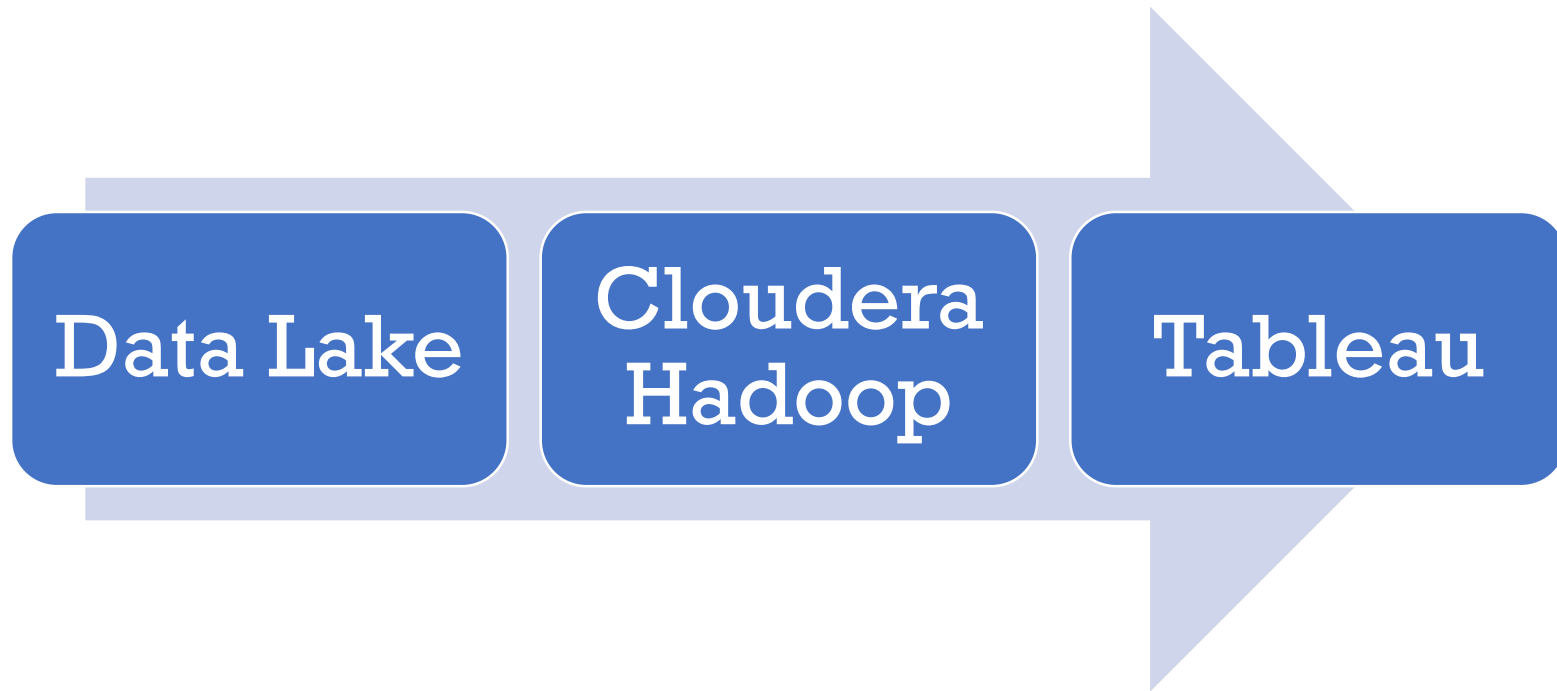
Funding?

Staff?

The Data Lake

- ✓ Storage challenges
 - ✓ Automated Ingestion
 - ✓ Security/Access (via AD)
 - ✓ Querying capabilities
 - ✓ OCTO-supported
- 

Data Lake to Tableau



Cloudera Hadoop

Server: Port:

Enter information to sign in to the server:

Type:

Authentication:

Transport:

Realm:

Host FQDN:

Service Name:

Require SSL

[Initial SQL...](#)

Tableau - DataLakeConnection

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Analytics dockless_lime_status_ch...

Search

Dimensions

- Recordid
- Provider Id
- Provider Name
- Device Id
- Vehicle Id
- Vehicle Type
- Propulsion Type1
- Propulsion Type2
- Event Type
- Event Type Reason
- Geometry Type
- Associated Trip
- DateTime
- First Event at 6 AM
- Measure Names

Measures

- Event Time
- Publication Time
- Longitude
- Latitude
- Battery Pct
- Number of Records
- Measure Values

Pages: MONTH(DateTime)

Columns: Longitude

Rows: Latitude

Filters: Event Type Reason: user_drop_off

Marks: Automatic

Color: Vehicle Type

Size: Label

Detail: Tooltip

Drop Off Locations

Event Type Reason

- (All)
- low_battery
- maintenance
- maintenance_drop_off
- maintenance_pick_up
- rebalance_drop_off
- rebalance_pick_up
- user_drop_off
- user_pick_up

MONTH(DateTime)

October 2019

Vehicle Type

- bicycle
- scooter

Data Source | DateTime | Battery Pct by Event Type | Count of Vehicle ID by Day | Device ID vs. Vehicle ID | Monthly Animation | O-D | Sheet 7 | Dashboard 1 | Dashboard 2 | Sheet 8 | Sheet 9

77967 marks 1 row by 1 column SUM of AVG(Longitude): -6.005,841.8358

Ting Ma



Tableau - DataLakeConnection

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Analytics

dockless_lime_status_ch...

Search

Dimensions

- Abc Recorduid
- Abc Provider Id
- Abc Provider Name
- Abc Device Id
- Abc Vehicle Id
- Abc Vehicle Type
- Abc Propulsion Type1
- Abc Propulsion Type2
- Abc Event Type
- Abc Event Type Reason
- Abc Geometry Type
- Abc Associated Trip
- Abc DateTime
- Abc First Event at 6 AM
- Abc Measure Names

Measures

- # Event Time
- # Publication Time
- # Longitude
- # Latitude
- # Battery Pct
- # Number of Records
- # Measure Values

Pages: YEAR(DateTime)

Columns: Longitude

Rows: Latitude

Filters: Associated Trip, Vehicle Type: scooter

Marks: Line, Color, Size, Label, Detail, Tooltip, Path, Associated Trip

Origin-Destination of Selected Dockless Vehicle Trips (An Example)

YEAR(DateTime): 2019

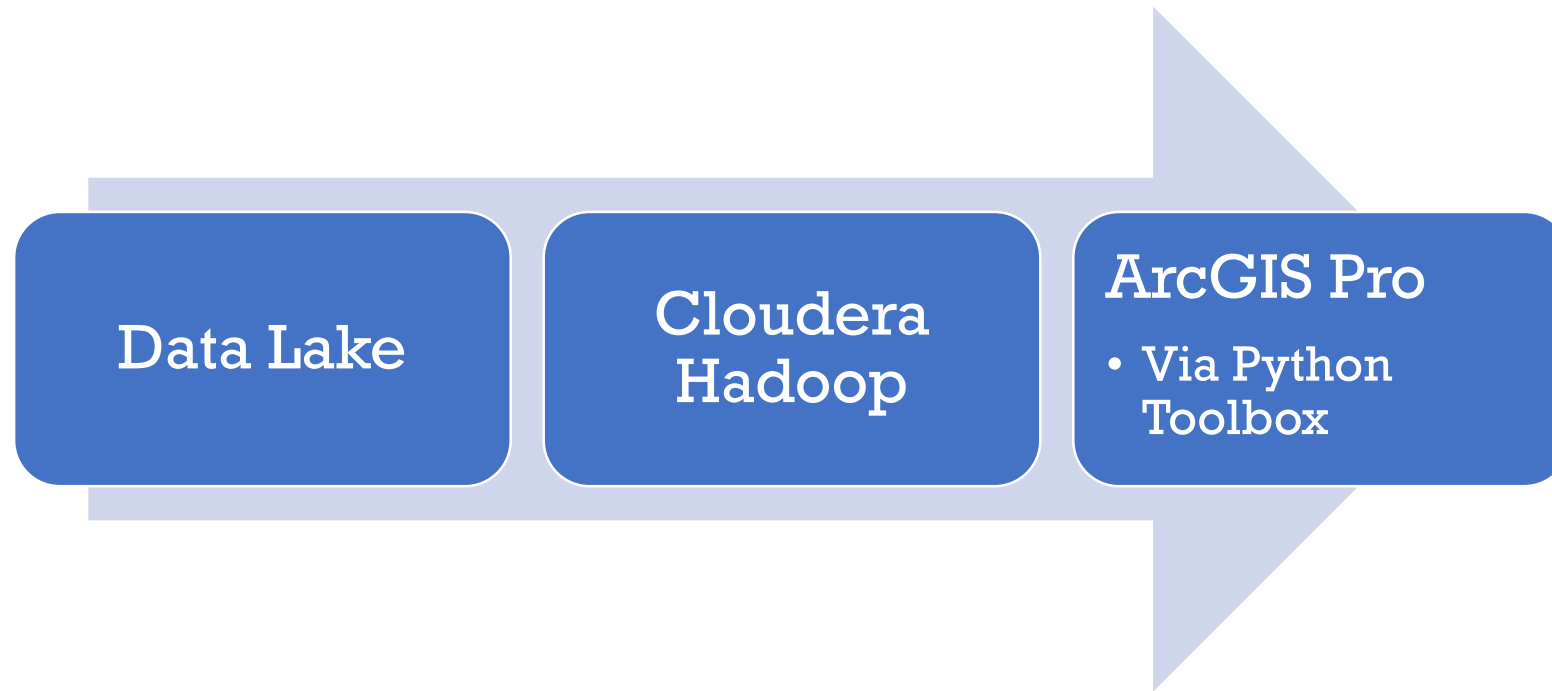
Show history

Data Source | DateTime | Battery Pct by Event Type | Count of Vehicle ID by Day | Device ID vs. Vehicle ID | Monthly Animation | O-D | Sheet 7 | Dashboard 1 | Dashboard 2 | Sheet 8 | Sheet 9

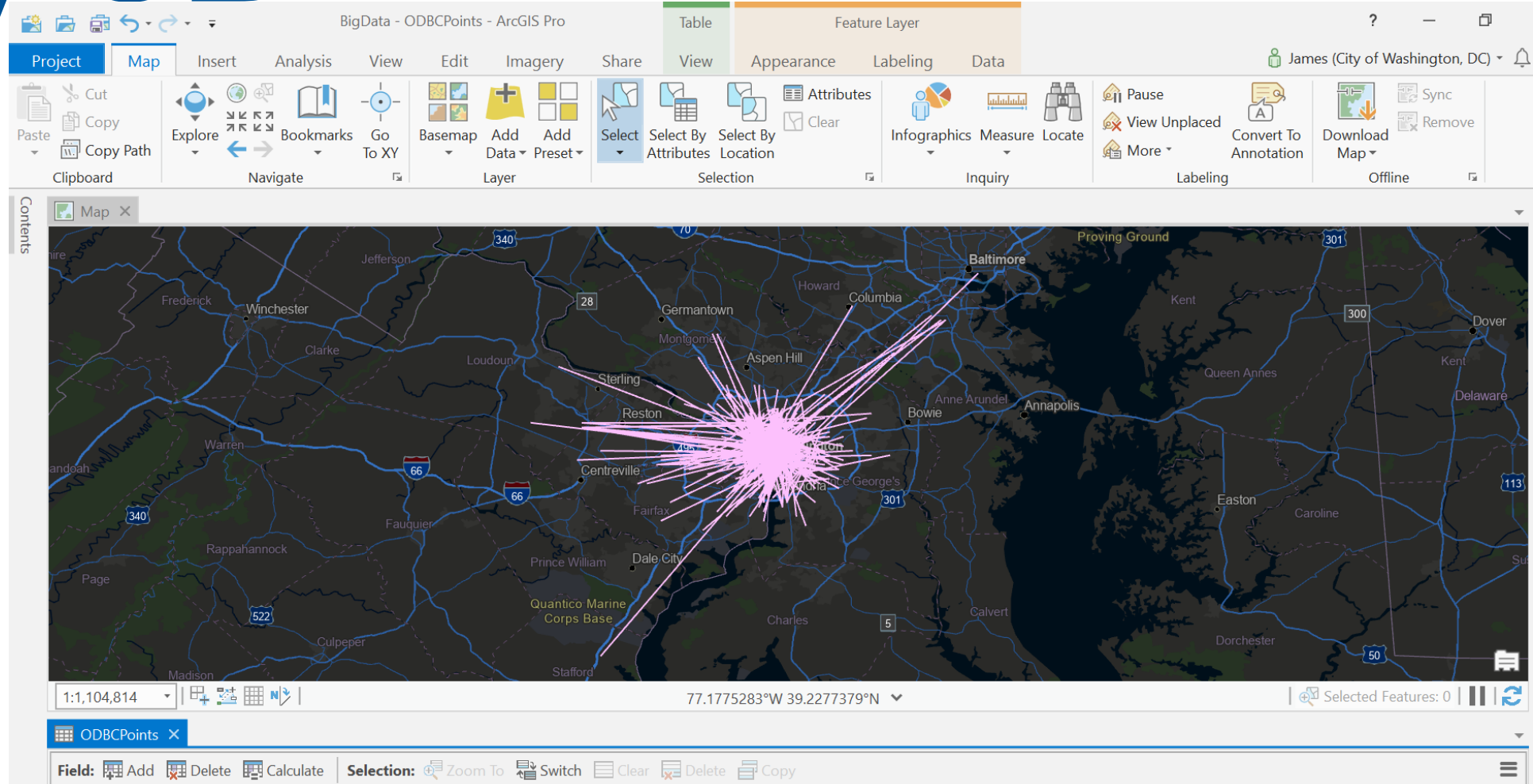
33 marks 1 row by 1 column SUM of AVG(Longitude): -2,542.14436

Ting Ma

Data Lake to ArcGIS Pro



Uber, Lyft Trips by OD



Next Steps:



- Work with OCTO to load remaining providers
 - Improve links to GIS
- 

Thanks!

Questions?

Dockless Program Manager: Sharada.Strasmore@dc.gov

Data/Automations: James.Graham2@dc.gov

Data Vizualizations: Ting.Ma@dc.gov



Data Report

Mario Field, Data Curation Program Manager

Office of the Chief Technology Officer

Updated Data

- Architect of the Capitol
- Business Improvement District
- Central Employment Area
- Cooling Center
- DC Property
- Entertainment and Sports
Gambling Restriction Zone
- Federal Location
- Great Street
- Liquor License
- Main Street
- Metro Line
- Notary Public
- Polling Place
- Public School
- Recreation Center
- School Attendance Zone
- Shotspotter
- Wireless Hot Spot

New Datasets

- **MPD**
 - Adult Arrest (MPD)
 - Marijuana Arrest (MPD)
- **DMV**
 - Driver's License (DMV)
 - Identification Card (DMV)
 - Customer Disability Placard (DMV)
 - Personalized Plate (DMV)
 - Reciprocity Parking Permit (DMV)

Upcoming Datasets

- AIDS Drug Assistance Program
- Boot and Tow
- Building Benchmark
- High School Lottery Results
- Piers

Open Discussion

Questions?

Idea's for future
topics/presentations?

Next Inter-Agency Data Team
meeting likely next year

Thank you 😊

