Big Data, Spatial Analytics, Machine Learning

Workflows and Applications

Kyunam Kim

Advanced Spatial Analytics Lead

Wendy Keyes

Data Scientist

Jimmy O'Donnell

Sr. Data Scientist

esri THE SCIENCE OF WHERE

2021 ESRI USER CONFERENCE

Agenda

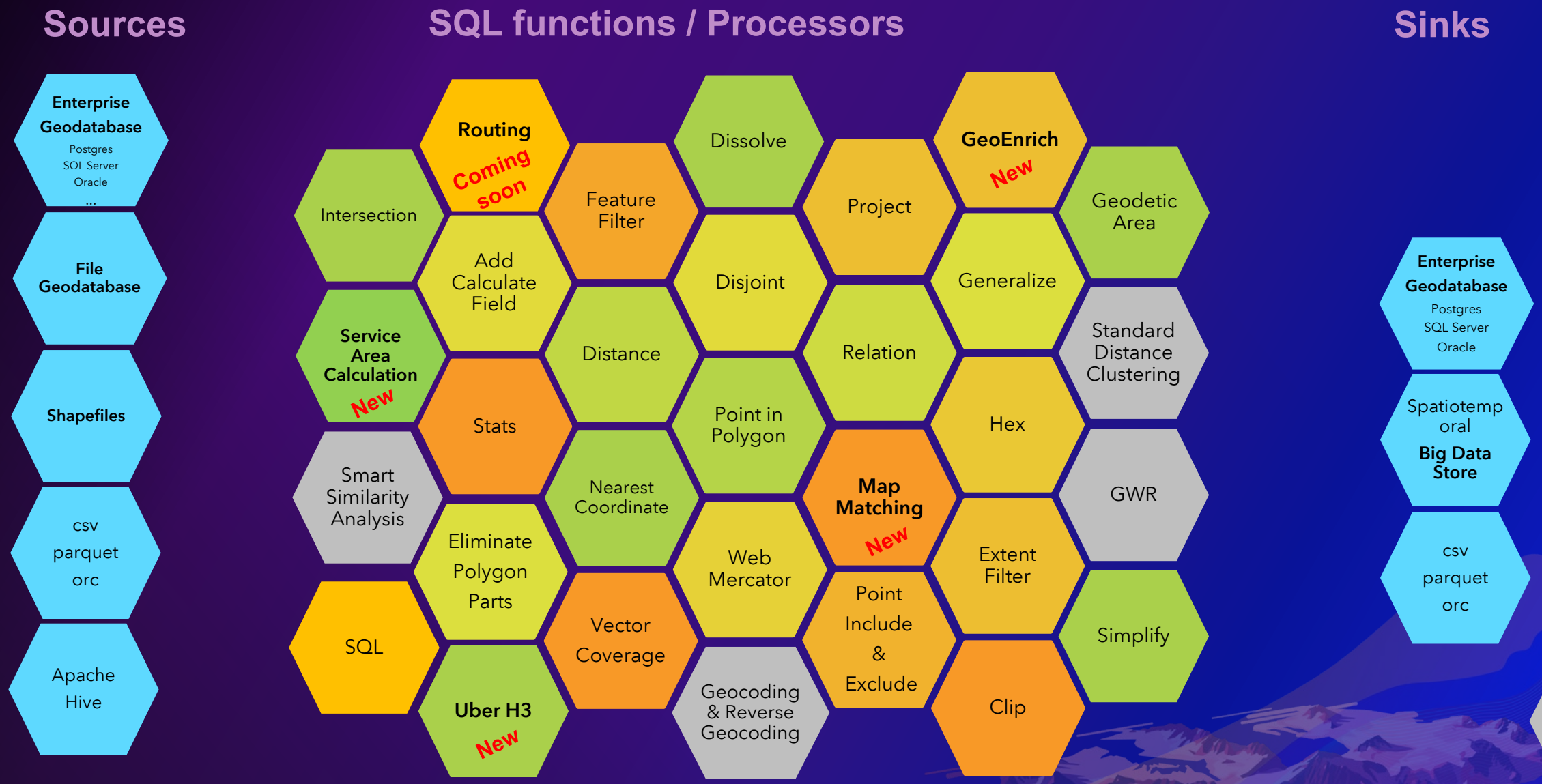
- Who We Are
- What’s New in Big Data Toolkit
- Customer Use Cases
- Live Q&A (15 mins)

Who We Are

- Advanced Spatial Analytics Team in Professional Services
  - Data Engineers
  - Data Analysts
  - Data Scientists
  - Software Developers

- Improve customer businesses by applying
  - Big Data
  - Spatial Analytics
  - Computer Vision
    - Object Detection, Object Segmentation
  - Natural Language Processing

What is Big Data Toolkit?



Big Data Toolkit

- You know SQL?
  - You can do Big Data!

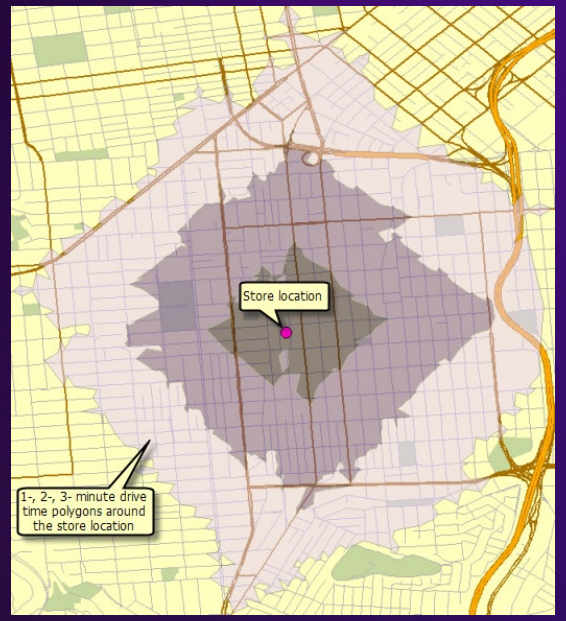
  - SELECT ST_Buffer(Shape, 2.5) FROM Locations
  - SELECT * FROM Locations A, TradeAreas B WHERE ST_Contains(B.Shape, A.Shape)

- More in-depth sessions:
  - 2020 DevSummit https://youtu.be/qgdFn6p2gpM
  - 2019 DevSummit https://youtu.be/_BJk_LKlxXc

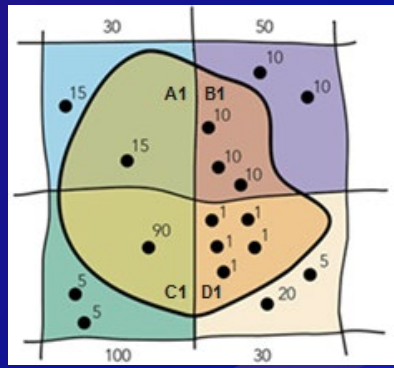
Big Data Toolkit

- Rich geo-context while feature engineering
- Understand phenomena around the world

- Massive Service Areas New



- Massive GeoEnrichment New
  - Apportions & appends 2,800 variables
    - Demographic
    - Consumer Spending
    - Market Potential

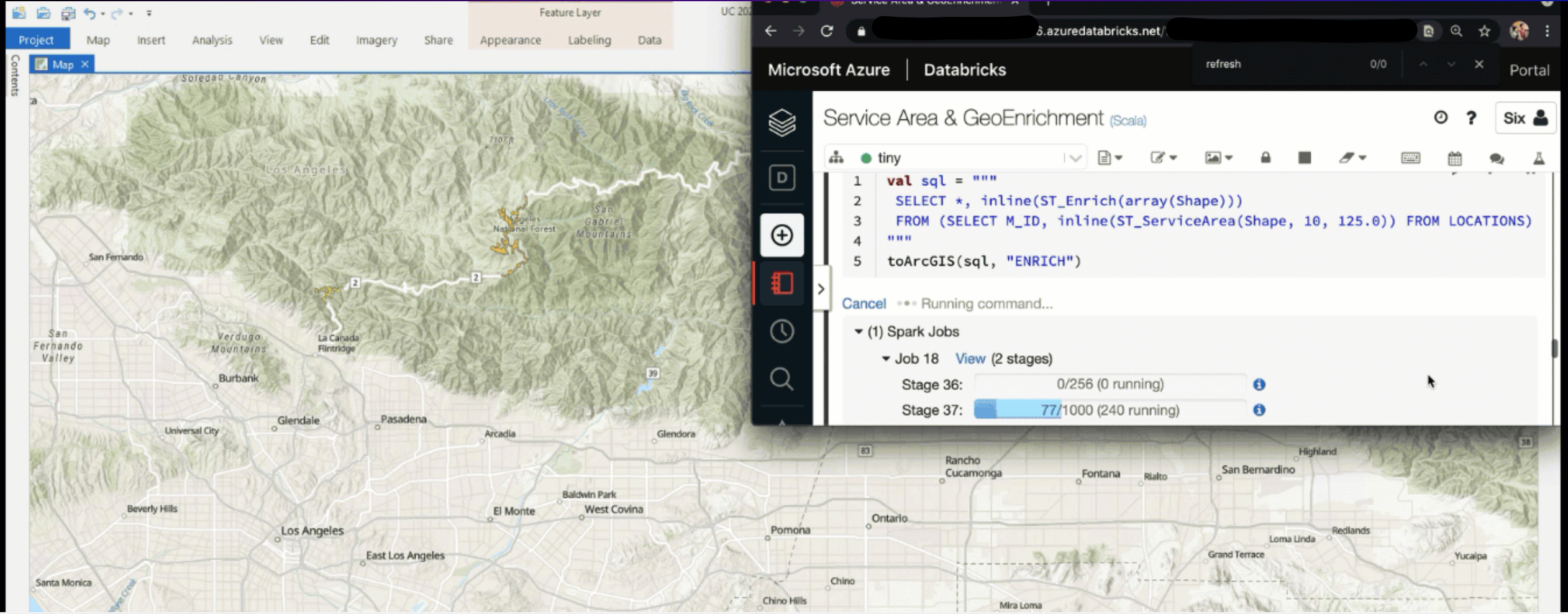


Weighted Block %

	Pop	
A1	15	(15/30 = 50% x 30)
B1	30	(30/50 = 60% x 50)
C1	90	(90/100 = 90% x 100)
D1	5	(5/30 = 16% x 30)

$$w_i = \sum_{B \in P_i \cap T} W_i(B).$$

$$W_i(B) = V^{TOTPOP}(B) \Big/ \sum_{B' \in P_i} V^{TOTPOP}(B').$$

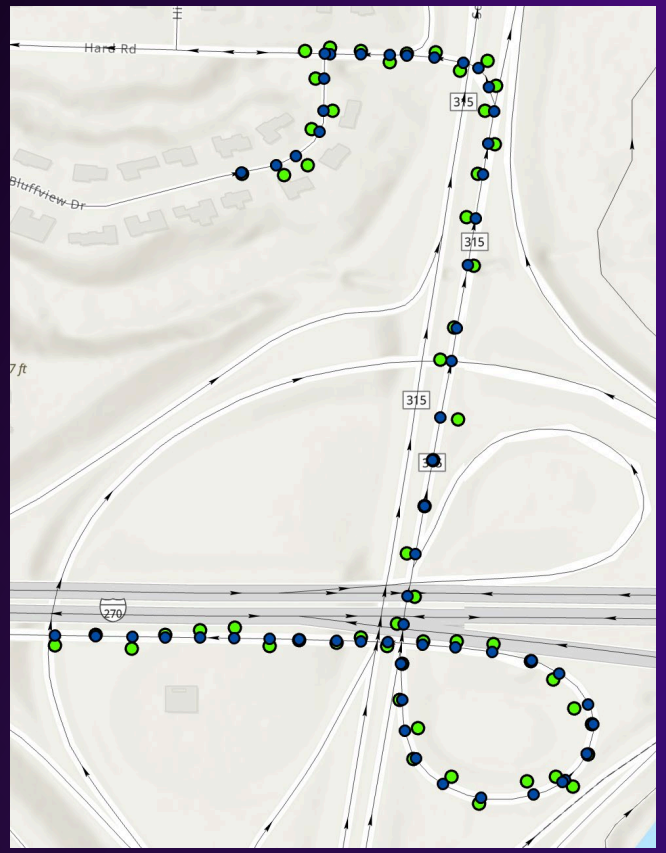


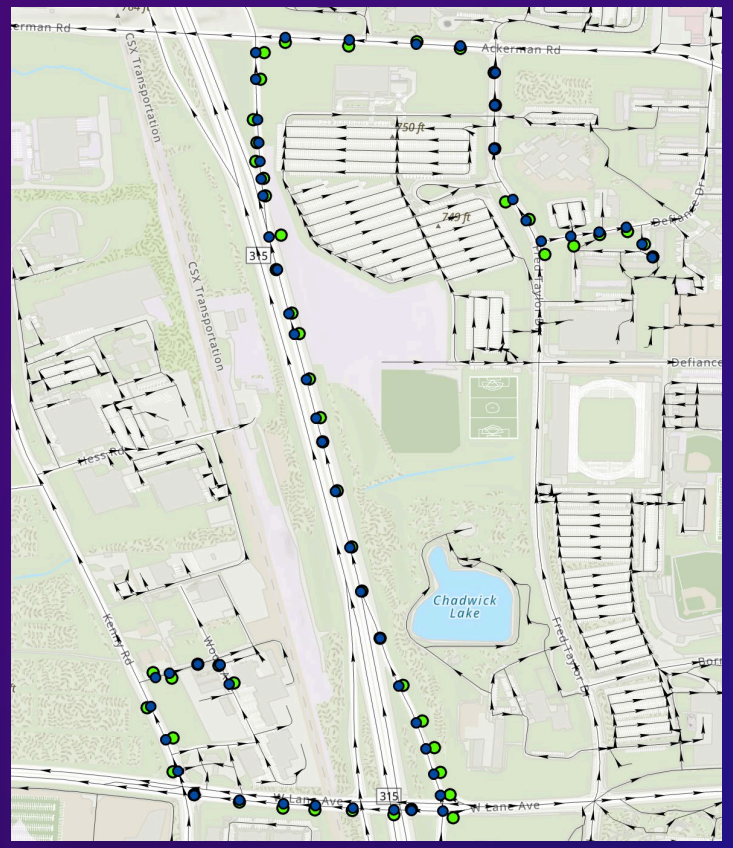
	TOTPOP_CY	HHPOP_CY	FAMPOP_CY	GQPOP_CY	POPDENS_CY	TOTHH_CY	AVGHHSZ_CY	FAMHH_CY	AVGFMSZ_CY	TOTHU_CY	OWNER_CY
1	361606	356430	308505	5175	9124.9	104694	3.4	82098	3.76	109110	54148
2	294723	290226	250936	4497	9244.8	85			3.75	89233	45327
3	225024	221428	191494	3597	9097.2	65215	3.4	51045	3.75	67875	35434

2,800+ variables

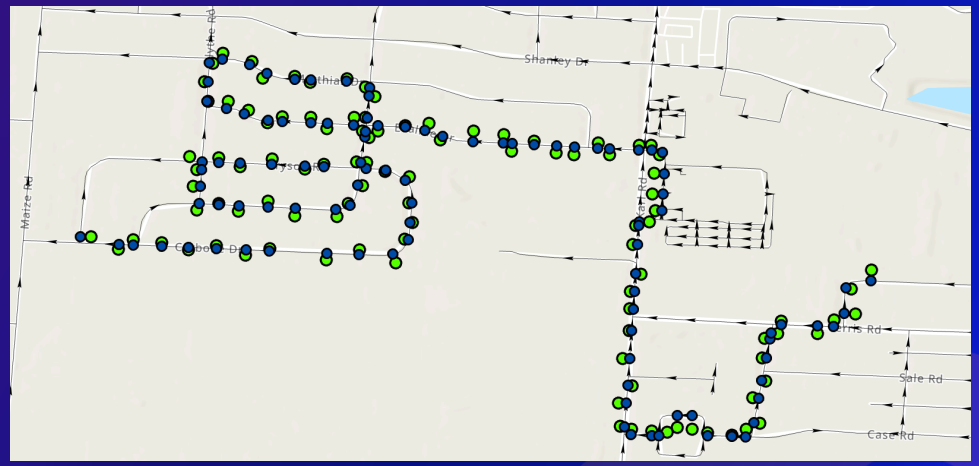
Map Matching (aka Smart Snapping)

Infer patterns & behaviours





- Snapped GPS Point
- Noisy GPS Breadcrumb



Trade Area Delineation and Location Planning
Using Human Movement Data, Business Analyst, and Big Data Toolkit

Trade Areas: Traditional Approach

- Common approaches
  - Hand-drawn using natural geographic attributes (highways, bridges, rivers, block groups, etc.)
  - Buffers, drive-times
  - Gather attributes within boundary
  - Store-centric (what customers are available to each store)
- Challenges with common approach
  - Selection bias
  - Market conversion and cannibalization
  - Boundaries and edge cases
  - Locus of competition
  - Competitive landscape
  - Temporal factors

Trade Areas and Location Planning

Data Inputs

AOIs
Stores
Competitors
Human Movement

Pre-processing

Big Data Toolkit

Customer-Centered Origin Geography

Demographics (BA)
Proximity features
Aggregate visits

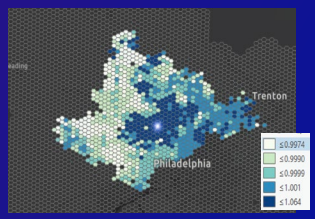
Geoenrichment

Business Analyst

Forecasting Model

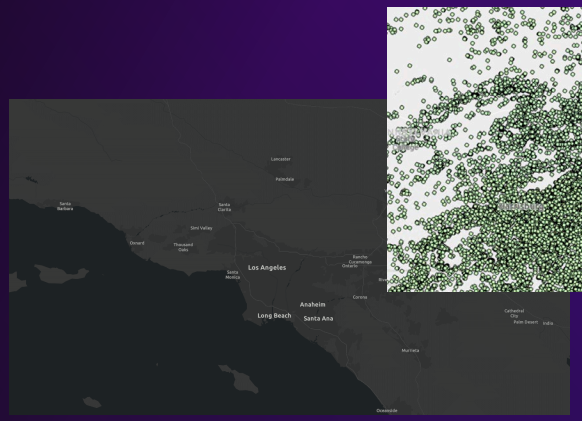
Derive Trade Areas

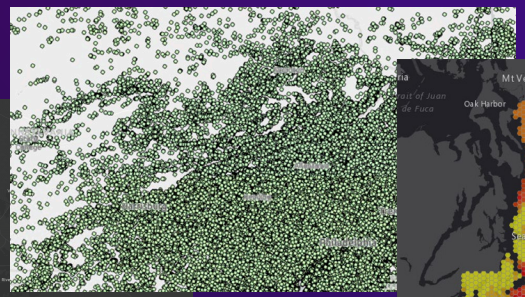
Simulation Tool

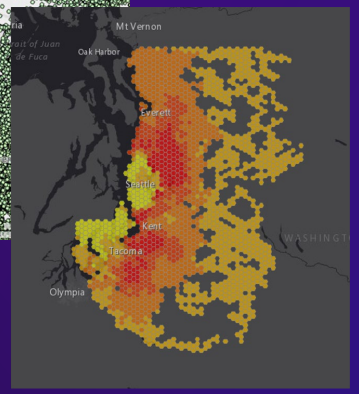


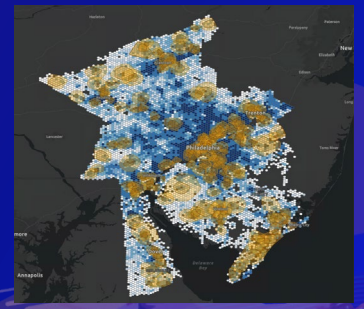
Add, Remove, and/or Relocate
Visualize areas most likely impacted
Compare to existing trade areas

Compare Trade Areas to Market Demand

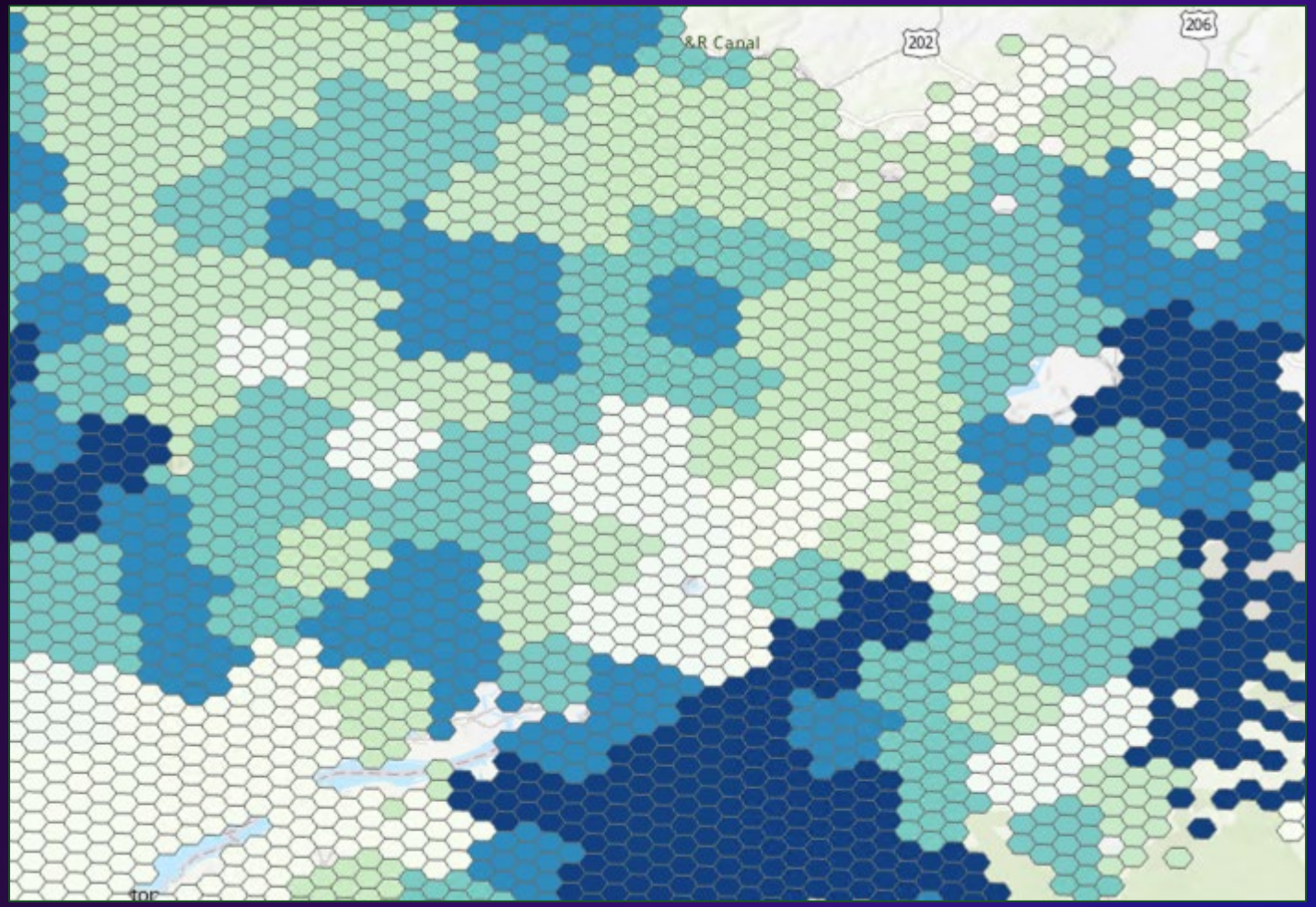






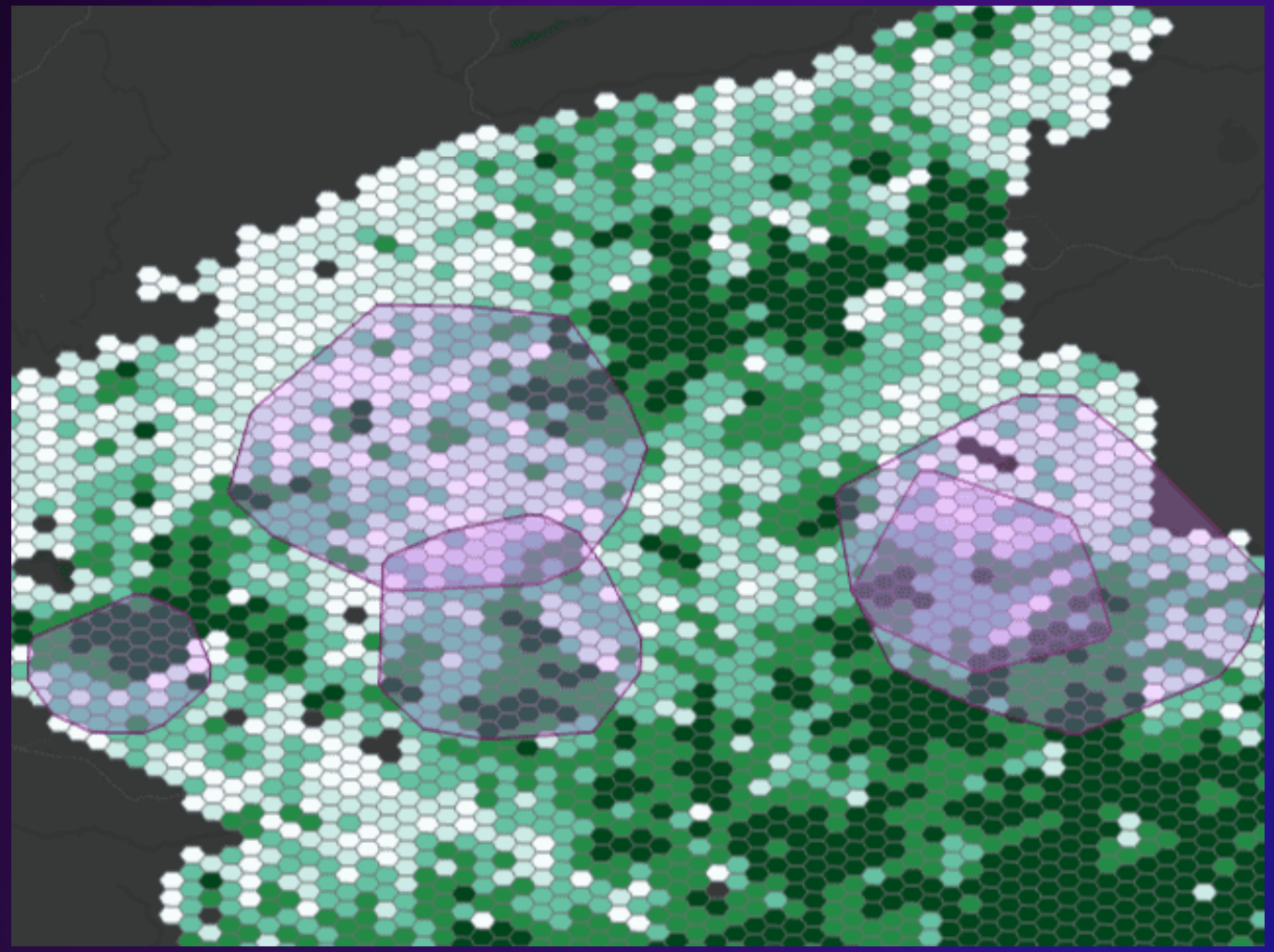


Relative Impact of Input Variable on Target



- Marginal impact of median income
- Relative impact varies across the area of interest
- Where does income matter most?

Evaluating the Market



- Trade Area (Polygons)
- Aggregated Visits (Hexbins)
- Under-served
- Over-saturated
- Optimizing locations

Asset Detection and Geo-Location Using Deep Learning and Photogrammetry (Charlotte Water)

Geo-Locating Water Meter Boxes

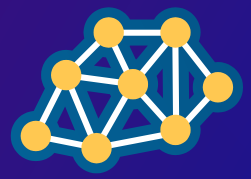
- Approximately 300,000 water meter boxes
- At least 5 different types of meters
- Up to $400,000 contracts for outside surveyors
- Goal: Find and geo-locate meter boxes
- Motivation: Significant cost savings

Deep Learning | Feature Extraction

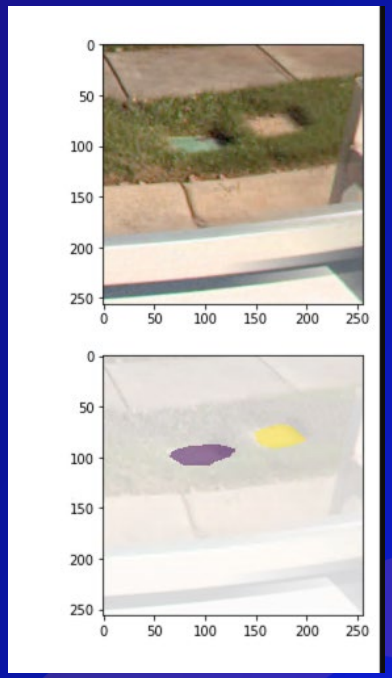
Image collection

360-degree pictures

Deep learning model

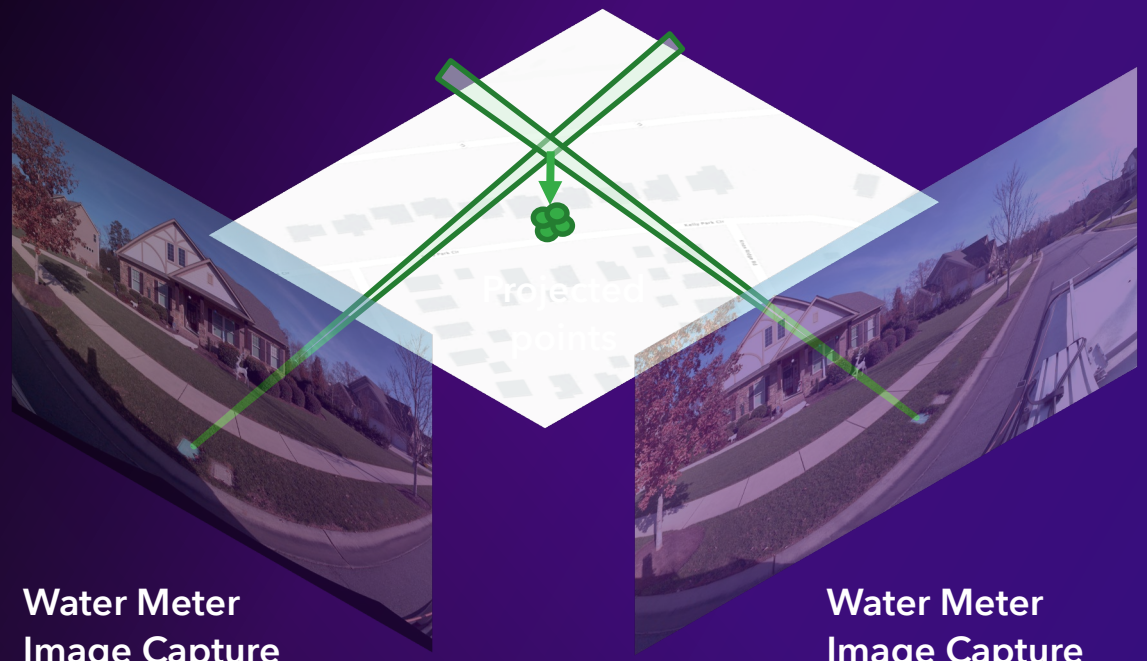


Machine learning results



Driving service routes

Photogrammetry | From Pixels to Geographic Coordinates



Water Meter Image Capture #1

Water Meter Image Capture #2

- Pictures of same meter from different angles
- Derive bearing lines
- Intersect bearing lines
- Project intersection points onto map
- Cluster points into distinct meter boxes

Crash Risk Analysis Using Hot Spots, Clustering and Space-Time Cubes (Pima County)

Use Case

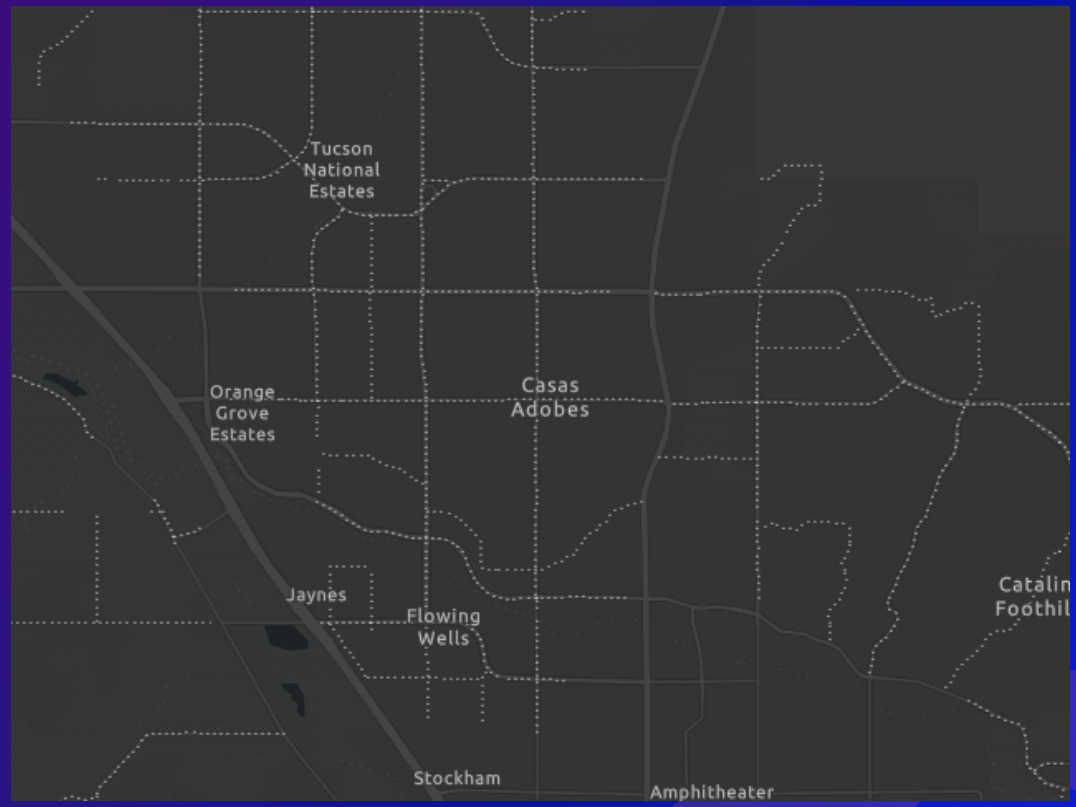
- Where are the statistically significant high incidence of crashes?
- Target high-risk areas for remediation

Preparing the Data

- Split road network into uniform segments, represented by a point
- Find nearest road point to each crash
- Aggregate crashes that occur at each road point

Hot Spot Analysis and Clustering

- Network Spatial Weights
- Hot Spot Analysis
- Clustering → Zones of Concern

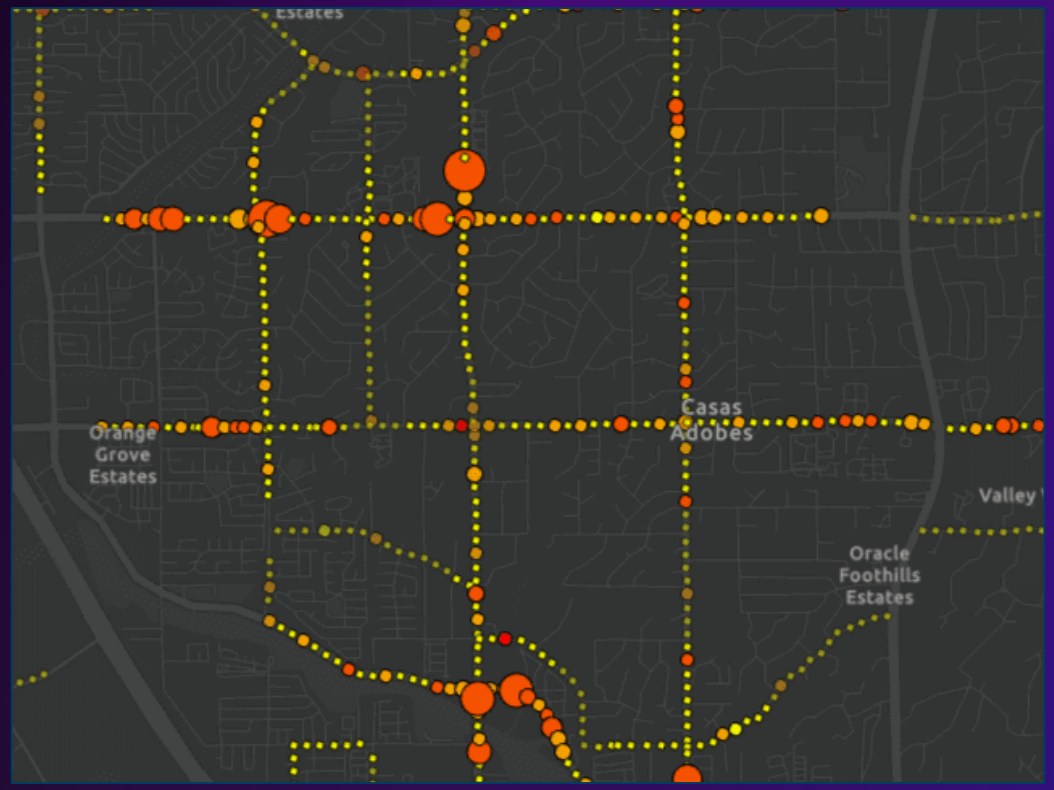


Score and Rank Zones

- Scoring criteria
  - Traffic exposure
  - Number of crashes
  - Severity of crashes

Temporal Analysis

Time-enabled map of crashes along road network



Space-Time Cube for 3D analysis

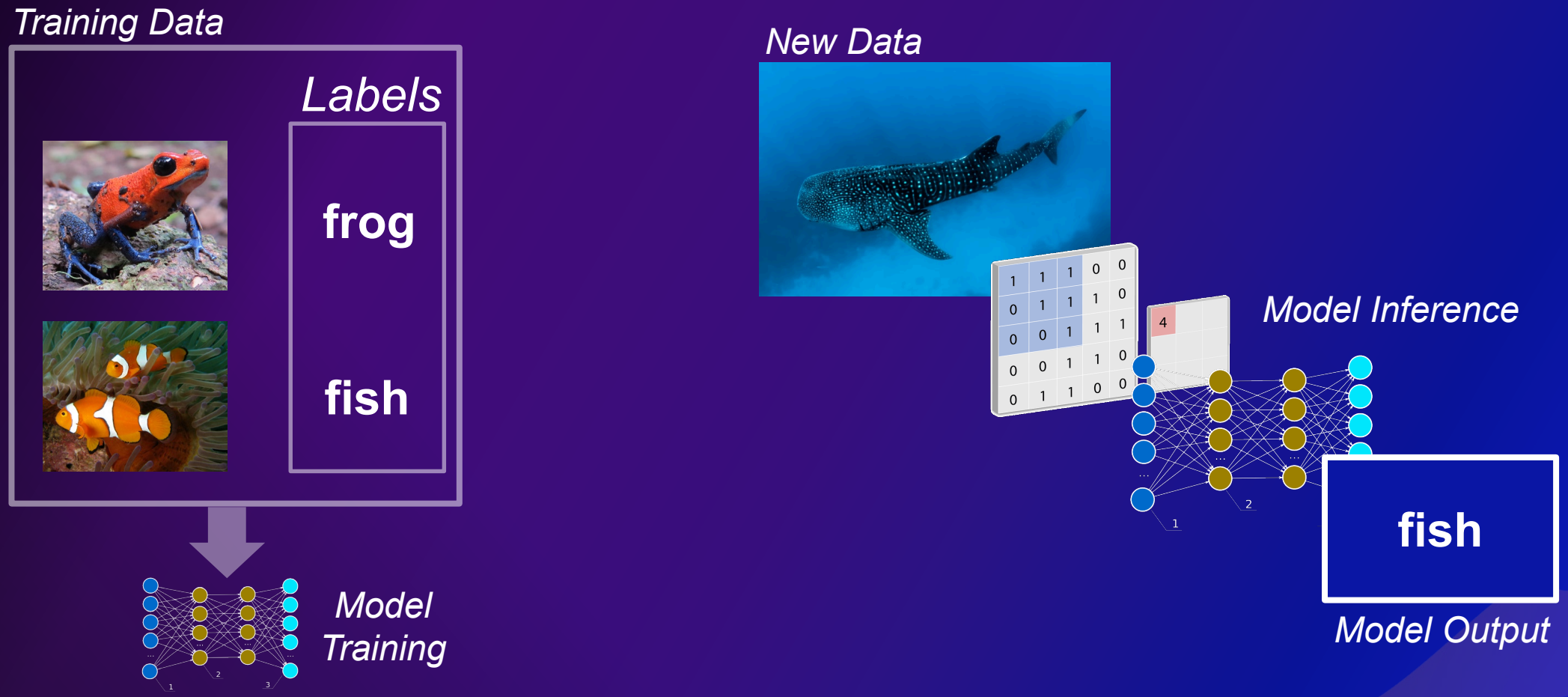
Further Iterations

- A/B Testing
- Pedestrian and Bicyclist Safety

Extracting geo-relevant data from images and video

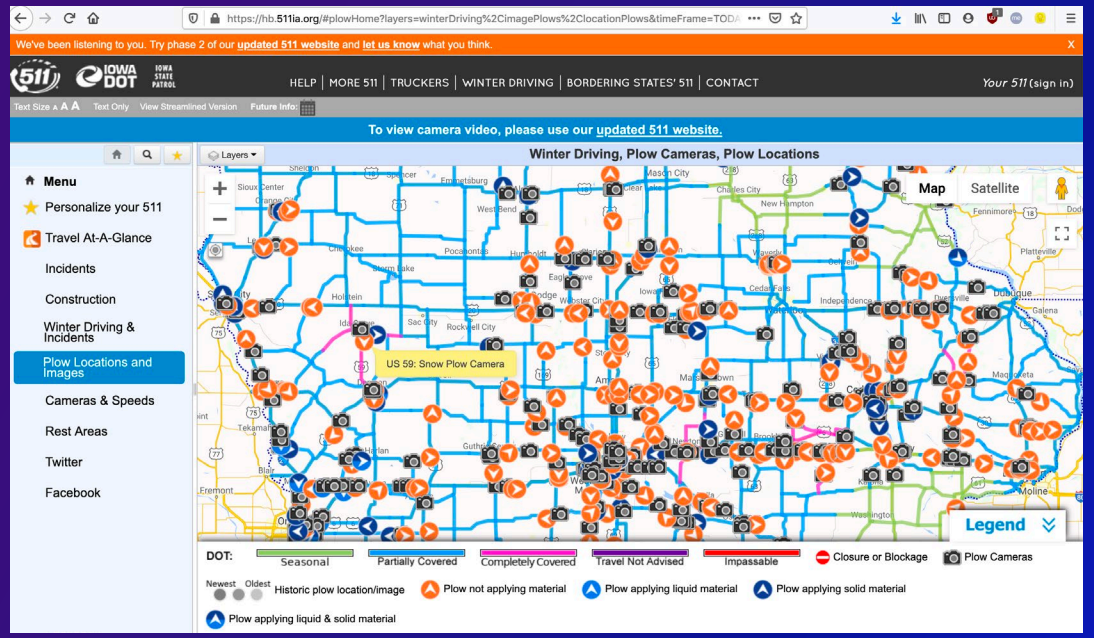
Customer use cases for machine learning

Machine learning models extract information from images and video



Winter road conditions from dashcam images

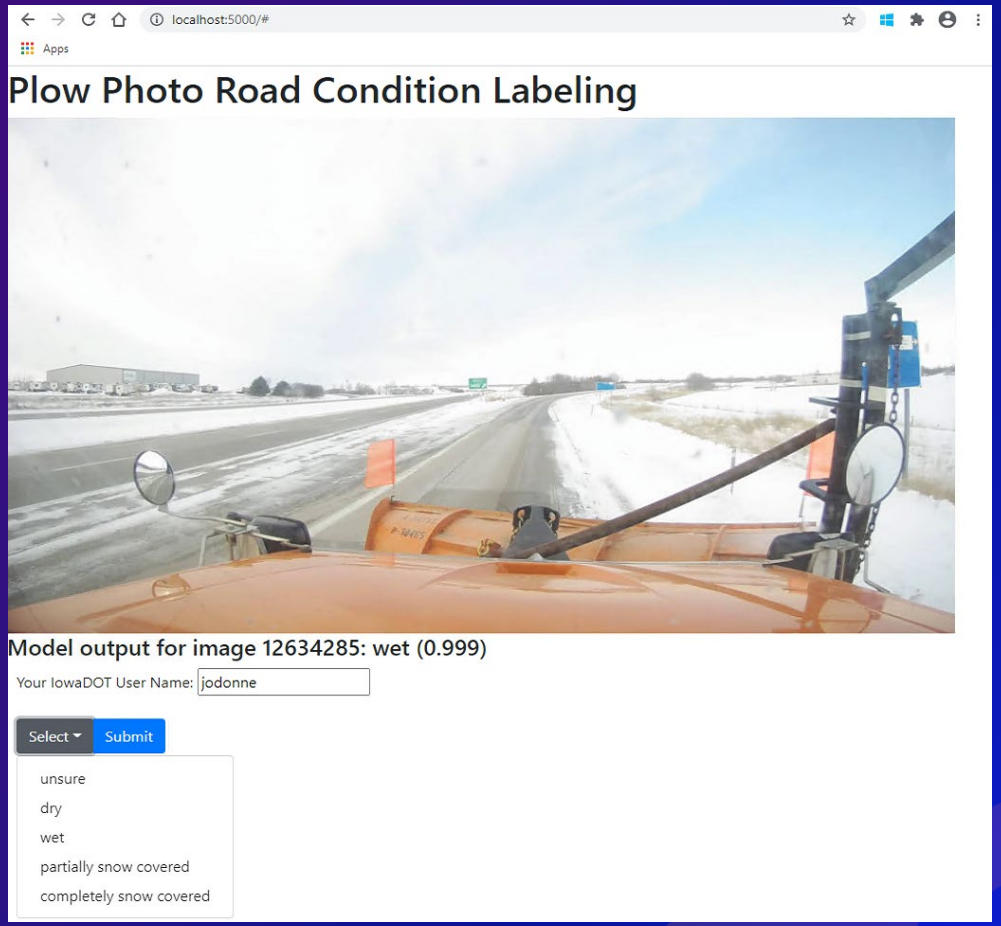
- Iowa Department of Transportation (DOT)
- Dashcams on ~700 plows
- 511 road condition reporting:
  - Dry, wet, partial or complete snow cover



- trackaplow.iowadot.gov
- data.iowadot.gov

Winter road conditions from dashcam images

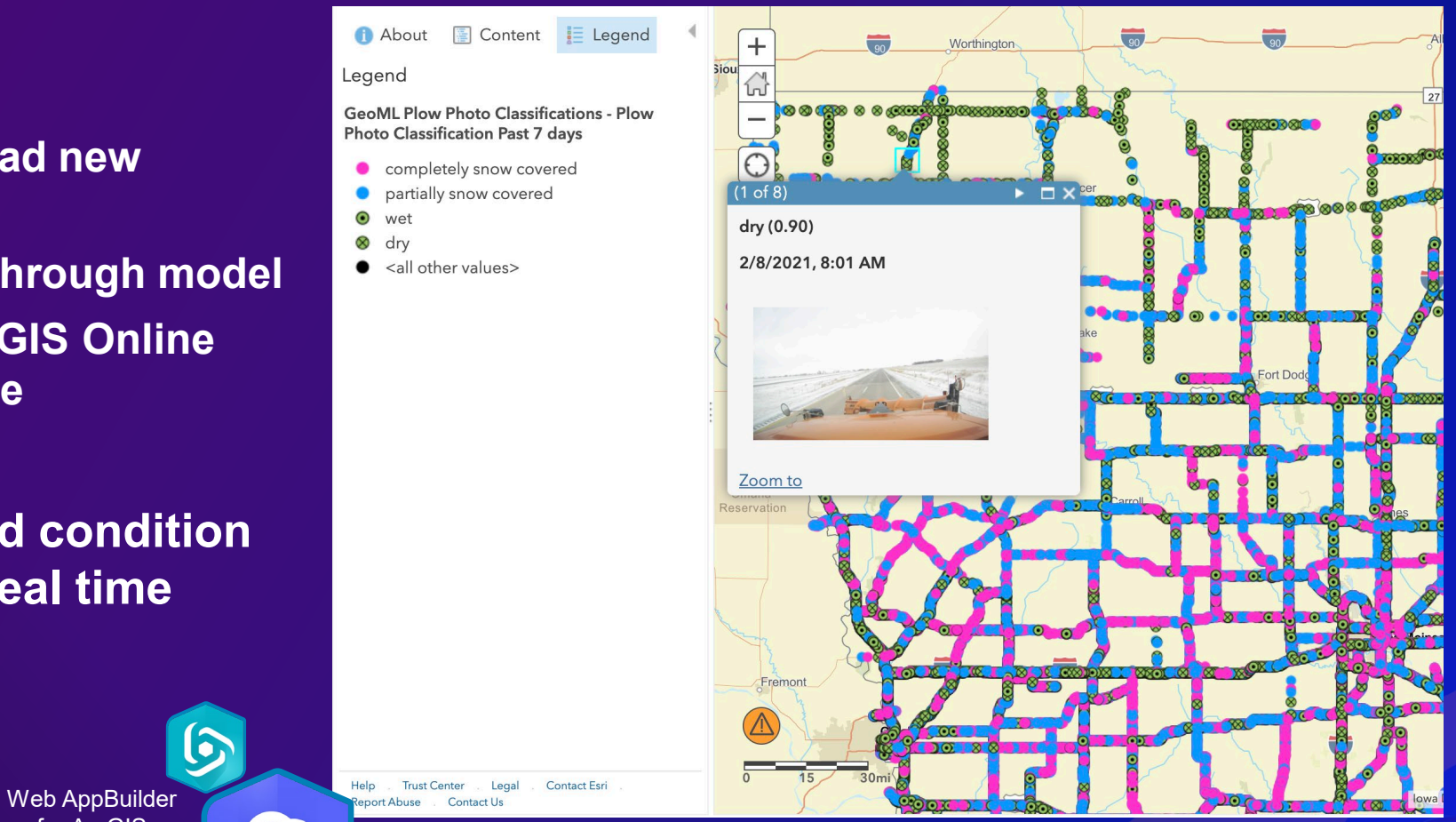
- Iowa DOT team labeled ~10,000 images to train the model
- Model trained on desktop with GPU
- Performance improved with data *quality*, not *quantity*



Winter road conditions from dashcam images

- **Every five minutes:**
  - **Check for and download new images**
  - **Run any new images through model**
  - **Write output to an ArcGIS Online Hosted Feature Service**

- **Documents winter road condition assessments in near real time**



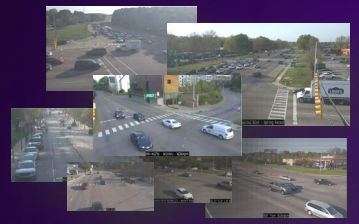
Web AppBuilder for ArcGIS

Real-time traffic analytics from streaming video

- City of Raleigh, North Carolina
- 200+ stationary networked traffic cameras
- Goal:
  - Leverage existing camera network and GIS to gain insights to traffic pattern and volume

Real-time traffic analytics from streaming video

Video Feeds (RTSP)



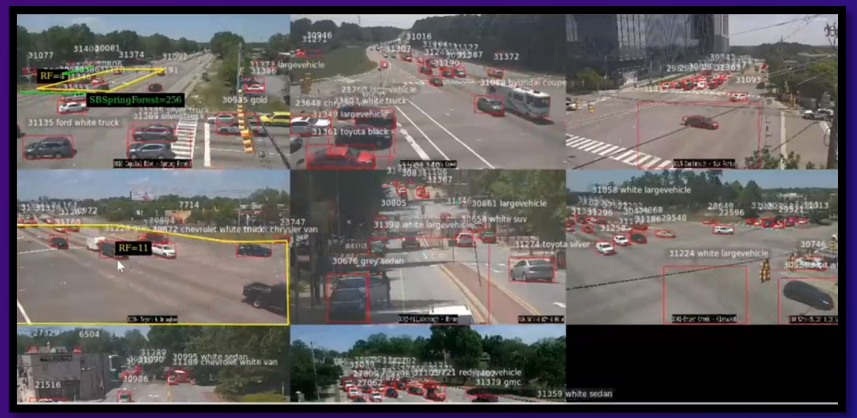
Deepstream

NVIDIA

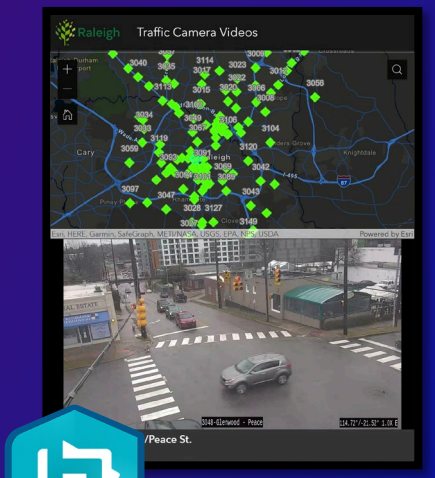
- Decode
- Pre-Process (CUDA)
- Detect, Track (TensorRT/CUDA)
- Encode, Composite

Dell R740; Nvidia Tesla T4

kafka



GeoEvent Server



ArcGIS Experience Builder

ArcGIS Online

Web AppBuilder for ArcGIS

esri

We want your feedback

Click on the Session Survey link
below this video window

Copyright © 2021 Esri. All rights reserved.