AUTOMATED POI VALIDATION & SCENARIO CLASSIFICATION

Analysis through Rule-Based and ML Approaches

OUR TEAM

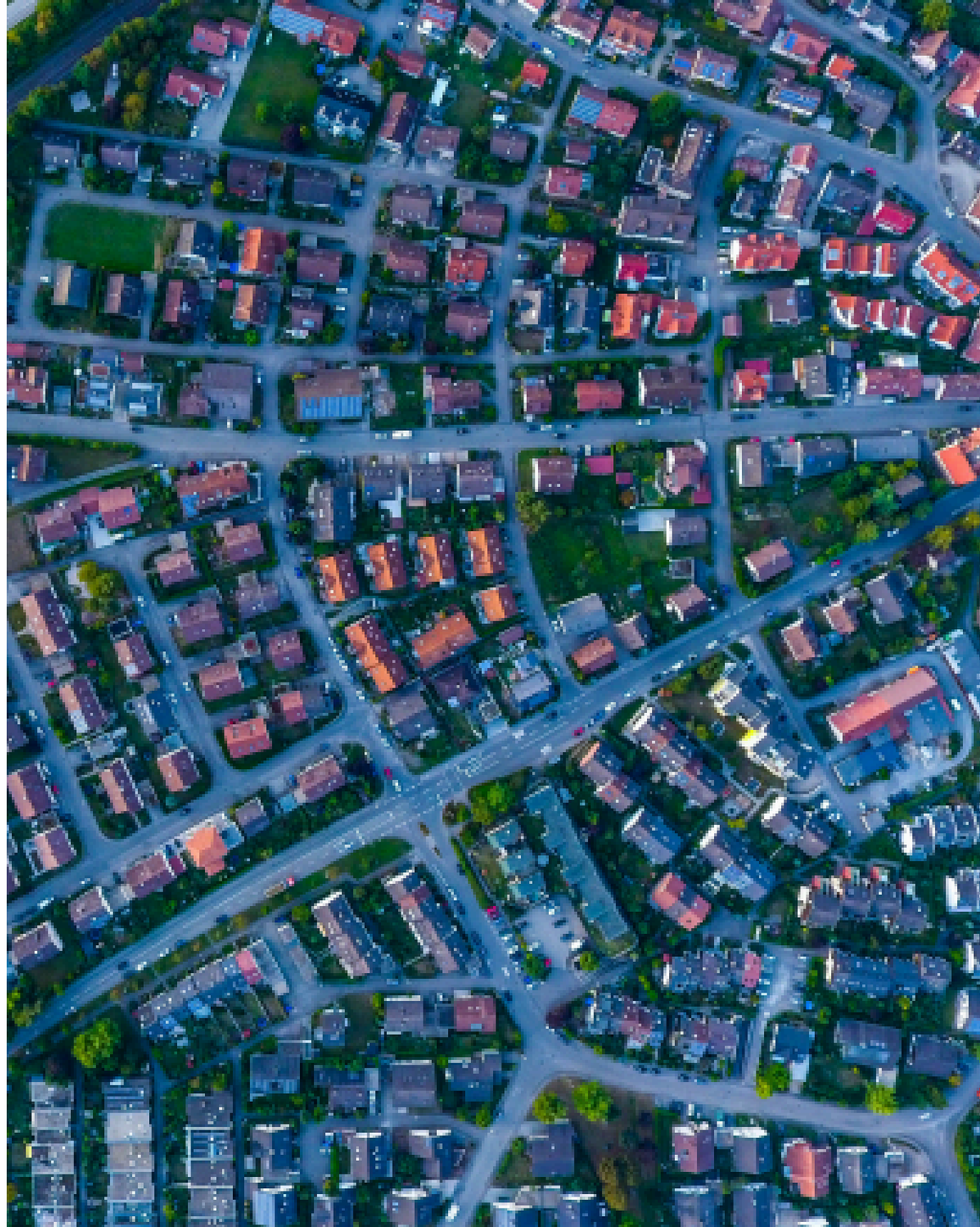
Jimena Díaz Franco

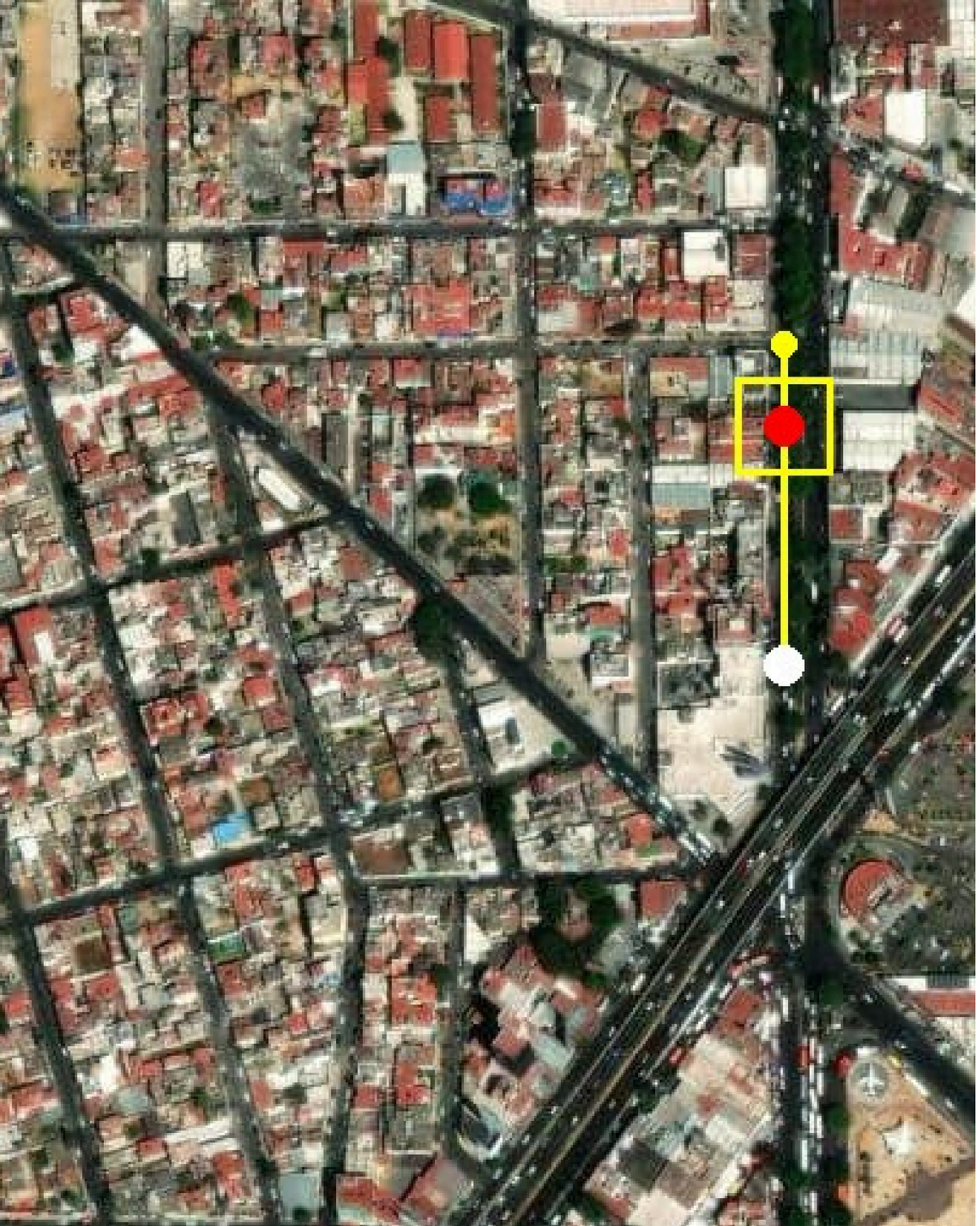
Ilan Gómez Guerrero

María Guadalupe Soto Acosta

Problem statement

Manually correcting large geospatial datasets, like verifying POI placements, is impractical at scale due to the need for precise spatial alignment, street-side accuracy, and consistency. Traditional QA relies on slow, error-prone visual checks, creating a bottleneck as datasets grow. Automation is essential to ensure quality efficiently.





WHAT WE BUILT

- Satellite Viewer for spatial inspection
- Rule-based engine for scenario detection
- ML model for scenario classification
- Duplicate identifier
- Real-time feedback overlay

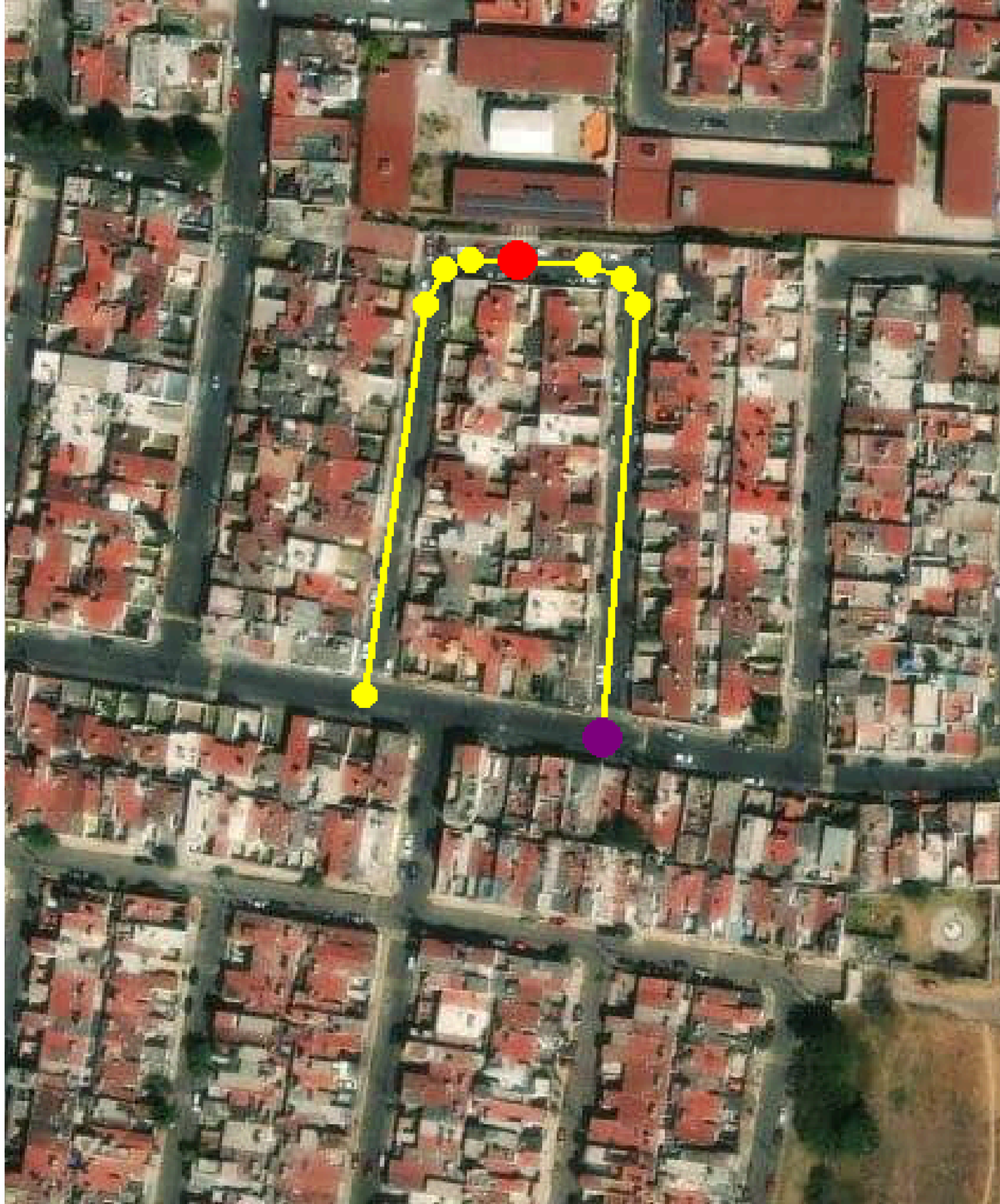
Rule-Based Validation Logic

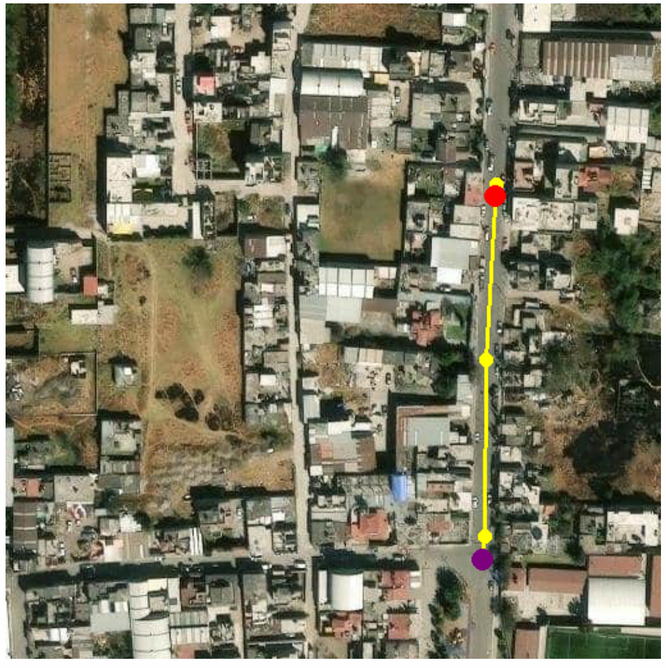
- Check for link presence
- Identify expected coordinate
- Compare actual vs expected distance
- Compare attributes
- Assign scenario + reason

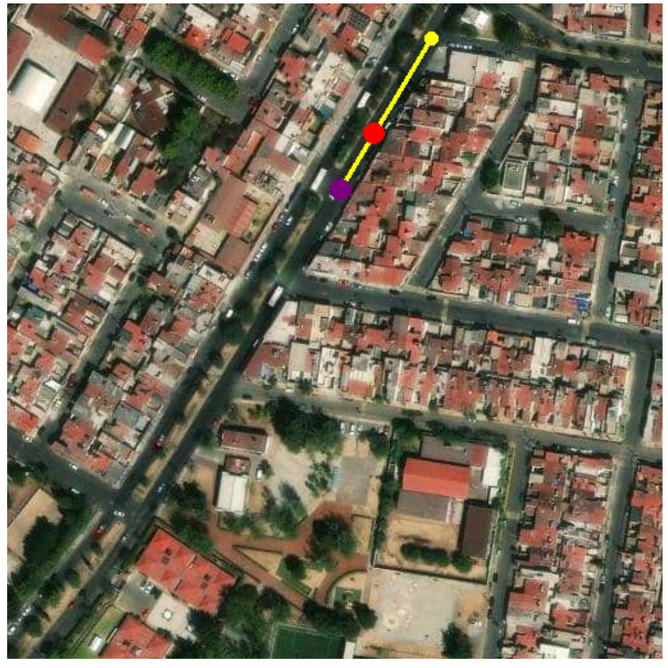
LINK_ID	POI_ID	POI_NAME	ST_NAME	link_id	_scenario	_reason
1346737015	1154356661	TACO SALSA	AVENIDA JUA	1346737015	3	Name mismatch (link: "AVENIDA BENITO JUÁREZ", POI: "AVENIDA JUAN PABLO II")
1346737015	1154356662	OXXO	AVENIDA JUA	1346737015	3	Name mismatch (link: "AVENIDA BENITO JUÁREZ", POI: "AVENIDA JUAN PABLO II")
702650117	1178940778	EMMANUEL	BOULEVARD S	702650117	3	Name mismatch (link: "BOULEVARD LÁZARO CÁRDENAS", POI: "BOULEVARD SOLIDARIDAD LAS TORRES")
721496780	1179042774	MISCELANEA	BOULEVARD S	721496780	3	Name mismatch (link: "BOULEVARD LÁZARO CÁRDENAS", POI: "BOULEVARD SOLIDARIDAD LAS TORRES")
848010952	1202700681	MISCELÁNEA	CALLE RÍO VIE	848010952	3	Name mismatch (link: "CALLE MARIO ORTEGA MONROY", POI: "CALLE RÍO VIEJO")
848010952	1202700682	TLAPALERÍA F	CALLE RÍO VIE	848010952	3	Name mismatch (link: "CALLE MARIO ORTEGA MONROY", POI: "CALLE RÍO VIEJO")
702719788	1202700660	TIENDITA DE	CALLE RÍO VIE	702719788	3	Name mismatch (link: "CALLE MARIO ORTEGA MONROY", POI: "CALLE RÍO VIEJO")
1119528348	1192852594	RÓTULOS EN	AVENIDA PRO	1119528348	3	Name mismatch (link: "CALLE ANTONIO J. TORRES", POI: "AVENIDA PROFESOR R. HERIBERTO ENRÍQUEZ")
1347025288	1048914837	FARMACIAS G	AVENIDA YUF	1347025288	3	Name mismatch (link: "AVENIDA EMILIANO ZAPATA", POI: "AVENIDA YURIRIA")

Visual Inspector (Satellite View)

- Satellite Raster Tile API
- Graphic Markers:
  - Street LineString
  - Reference Node
  - Point of Interest (POI)





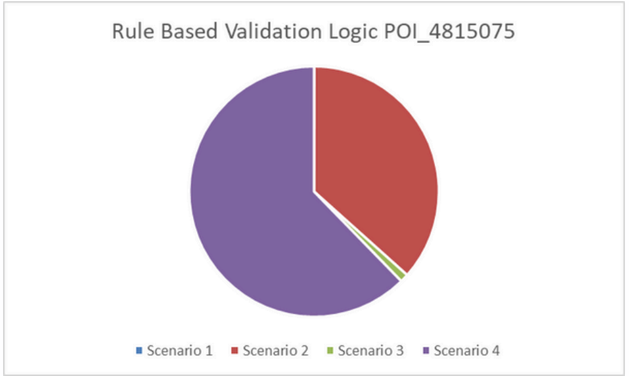


Machine Learning Integration

- Detection through color of the important points and lines
- Calculation of the right and left side of the road given the reference node
- ML integration for structures and buildings identification
- Comparation to data base location

Results & Insights

- Through the integration of a rule-based engine, a satellite viewer, and a machine learning model, we were able to automatically classify POIs into different scenarios.
- A large portion of POIs were flagged due to invalid atributes.
- Offset distance was consistently detected by both the rule-based engine and the ML model, showing strong alignment between logic-based and learned behavior.
- The model revealed inconsistencies in the data, often harder to catch manually.
- A significant number of POIs were validated as correct, helping isolate errors from clean data.



Future work: Our next steps

Improve ML accuracy with more data

- Collect a larger labeled dataset of POI–link associations (valid/invalid) across regions.
- Engineer richer features:
  - Distance to link
  - Percent along link (PERCFRREF)
  - Road type & POI category
  - Local POI density
- Use scalable models: Random Forest, LightGBM, or spatial transformers.
- Apply Active Learning: flag uncertain cases for human review and retrain iteratively.
- Optimize for big data with batch/cloud processing (e.g., Dask, Spark, AWS/GCP).

Building smarter geospatial systems starts with cleaner data, and scalable intelligence makes it sustainable.

Thank you!

See you again soon!