

ITA Snapping Mismatch Scenarios

EDA for understanding the increased rate of snapping mismatches observed in address point (AP) ToT repo and vendor refresh of ITA territories, focusing on the Municipality/City level. Deep Analysis of Overall Changes in Territory References, FIDS, spatial makeup and any significant changes made to data with Vendor Refresh merge.

Datasets:

- **Address Point (ToT)**
: MAPS_DATA_SEMANTIC_DB.TERRITORY_APP.ITA_SOURCE_TABLE_20241020_NDM_DATA
- **Territory (MUNI/CITY)(Vendor Refresh)**
): MAPS_DATA_SEMANTIC_DB.TERRITORY_APP.ITA_VENDOR_FRESH_SOURCE_TABLE_20241014_NDM_DAT
A

Methodology:

Data Extraction:

Extract APs with feature_id = 148639752566868687 (Municipality reference).

Extract corresponding Municipality and City names from the address_component field (e.g., "Cessole").

Spatial Analysis:

Perform spatial intersection analysis between the extracted APs and vendor refresh Municipality/City boundaries.

Identify APs that fall within different Municipalities/Cities between ToT and vendor refresh data.

Mismatch Categorization:

Classify mismatches into distinct scenarios based on spatial relationships and attribute discrepancies.

Generate a comprehensive list of AP feature IDs for each mismatch scenario.

In-depth Investigation:

Utilize FusionX/QGIS to visually inspect a representative sample of APs from each mismatch scenario.

Analyze spatial context, attribute values, and boundary geometries to pinpoint the root causes of mismatches.

Root Cause Analysis:

Investigate potential reasons for the increase in mismatches with the vendor refresh data, including:

Changes in Municipality boundaries.

AP geocoding.

Errors or inconsistencies in vendor refresh data.

Differences in snapping algorithms or tolerances between ToT and vendor refresh processes.

The analysis was conducted using FME (Feature Manipulation Engine) and Snowflake. FME was used for spatial analysis, including spatial filtering, geometry extraction, and data validation. Snowflake was used for data processing and querying, including parsing nested JSON data within the address point dataset.

PRELIMINARY FINDINGS(AP):

territory_type = Municipality

Cases

	A Category	B	C Count	D Comment
1				
2	AP that are spatial mismatches with any of territories of type municipality in ITA	Case1	8,278	↳ Total AP Spatial Mismatches(Muni(Vendor Refresh)): 8,278
3	AP_FID_MATCH_SPATIAL_OVERLAP_DISTINCT	Case2	6,620	↳ (cases where Address points: address component FID's match but spatially overlap): ↳ ITA Snapping Mismatch Scenarios
4	AP_FID_MATCH_NO_OVERLAP_AP	Case3	884	↳ B. AP_FID_MATCH_NO_OVERLAP: https://apple.box.com/s/3fvncag5h3x5tssafv6i24l52fuob6pm
5	Address Point with NULL REFERENCE FID and Spatial Overlap	Case4	6,620	↳ Address Point with NULL REFERENCE FID and Spatial Overlap
6	Address Point with NULL REFERENCE FID and NO Spatial Overlap	Case5	1	↳ Address Point with NULL REFERENCE FID and NO Spatial Overlap
7	Address Points Names Mismatch	Case6	476	↳ AP Name Mismatch

Total Address Points (ToT): 9,142,095

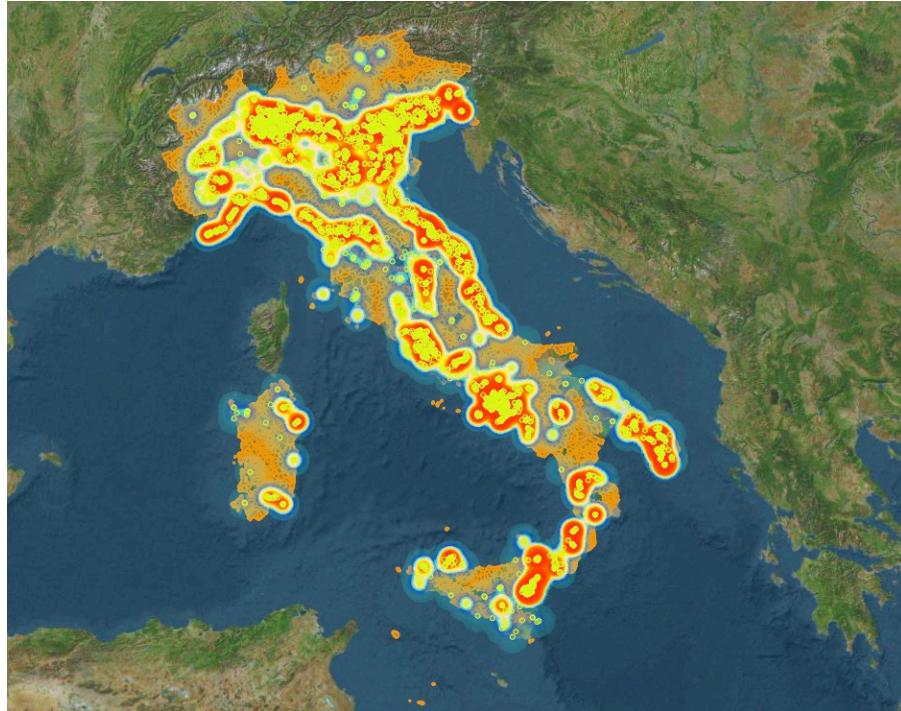
ITA Territories (Municipalities(Vendor Refresh)): 20,591

- Total AP Spatial Mismatches(Muni(Vendor Refresh)): 8,278
- AP_MUNI_MISMATCH data: <https://apple.box.com/s/co94ts7gvbyon3px9yym26g7r3yvamkh>

AP Muni Mismatch Map (spatial non-intersections muni territory):



AP Muni Mismatch (spatial non-intersections muni territory) Heat Map:



AP mismatch Territory Reference *IDs* :

*municipality REF ID

A. AP_FID_MATCH_SPATIAL_OVERLAP_DISTINCT: <https://apple.box.com/s/mx40mrs7zuw9bqqvzyy7nvggeau9w9d>

(cases where Address points: address component FID's match and spatially overlap)

Total: 9,116,771

DUPLICATE: 9,110,151 (all address points that have reference to Municipality FID)

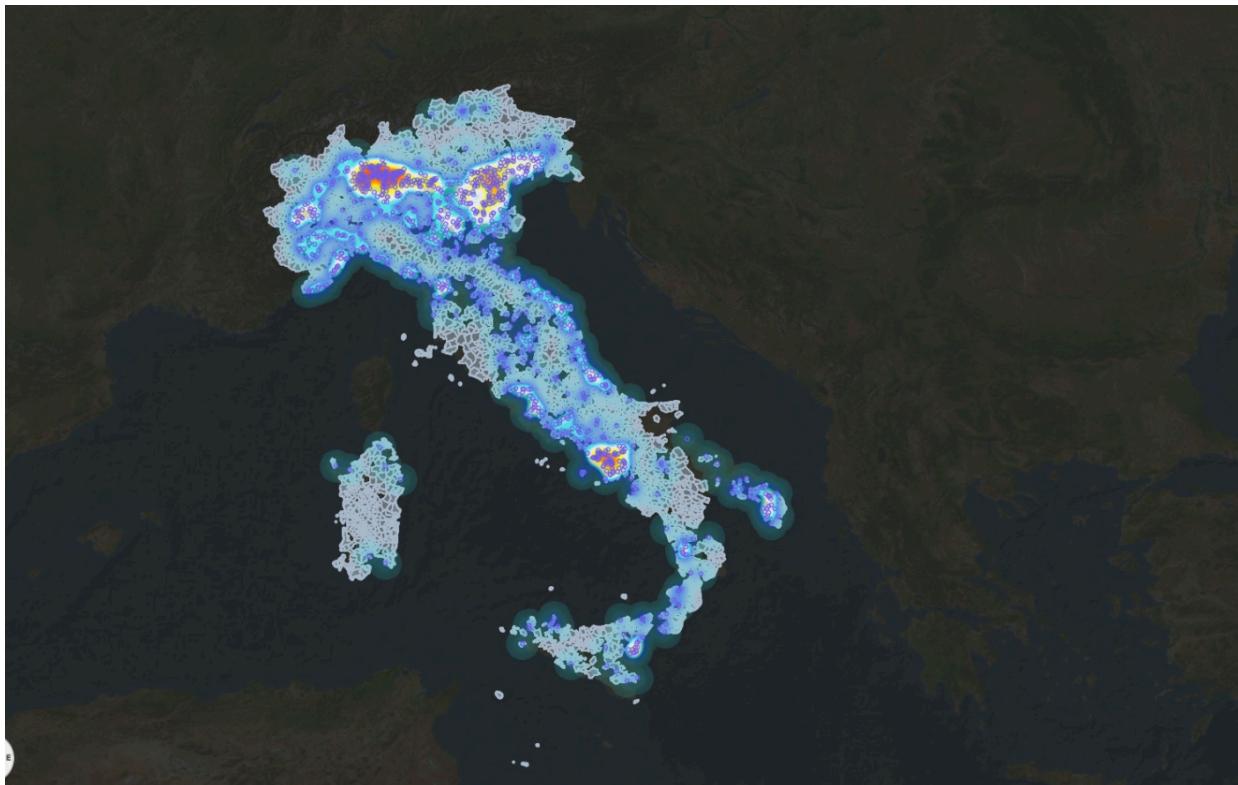
DISTINCT: 6,620 (Unique Territories referenced by Address Component in AP dataset(that match Vendor Refresh Territory Type = Municipality))

B. AP_FID_MATCH_NO_OVERLAP: <https://apple.box.com/s/3fvncag5h3x5tssafv6i24l52fuob6pm>

(cases where Address points: address component FID's match but spatially do not overlap)

Total: 6,372
DUPLICATE: 5,848
DISTINCT: 884

FID_match_no_spatial_overlap map:



Address Points with NULL reference ids: 3,180

Null reference id: NULL values for FEATURE_ID in address component for associated territory)

Address Point with NULL REFERENCE FID and Spatial Overlap

Total: 2,248
DUPLICATE: 2,247
DISTINCT: 1

*All (AP) that intersect spatially but have NULL Municipality REFERENCE FID

	FEAT ID	MUNI_NAME	TERR_REF_ID	GEOGRAPHY	TERR_NAME_AP	ISO COUNTRY CODE	predicates
1	46859145050 2726029	"Nago-Torbole"	<null>	{ "coordinates": [[[1.0891056700 00000e+01, 4.5877352300 00000e+01]], { "type": "MultiPoint" }] }	Nago Torbole	ITA	INTERSECTS

Address Point with NULL REFERENCE FID and NO Spatial Overlap

Total: 932

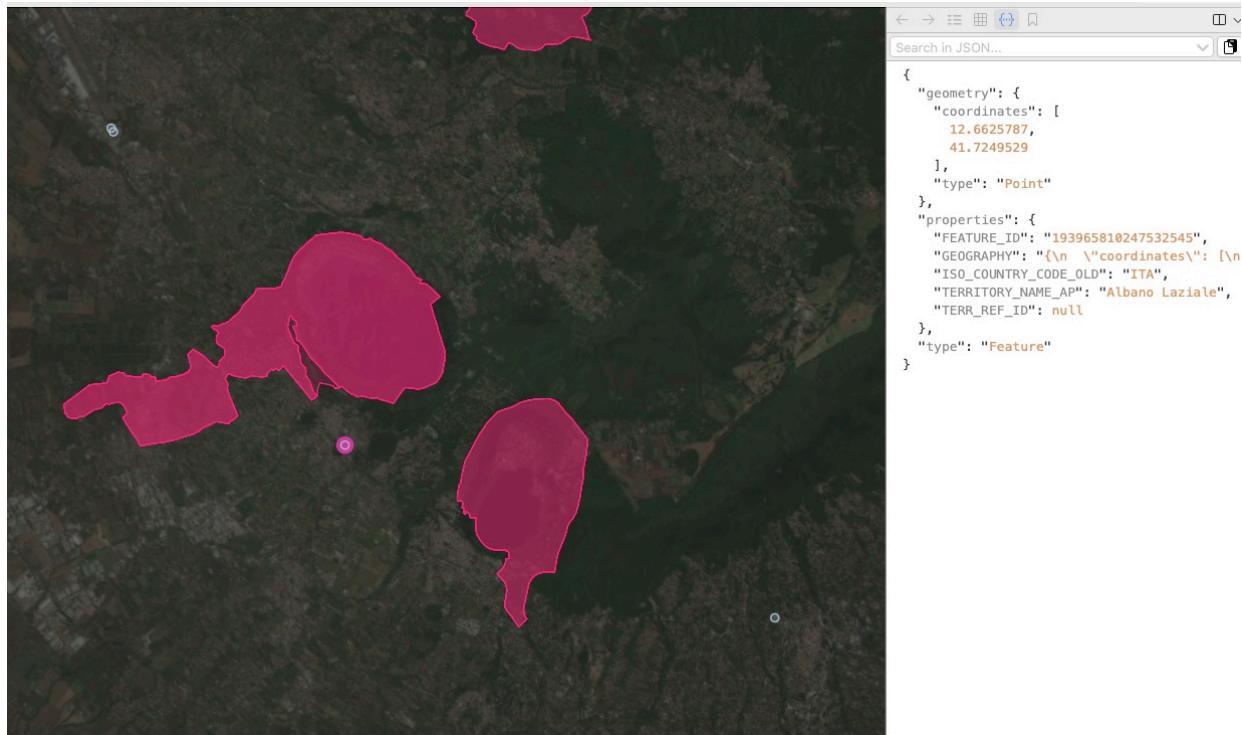
DUPLICATE: 931

DISTINCT: 1

	FEAT ID	MUNI_NAME	TERR_REF_ID	GEOGRAPHY	TERR_NAME_AP	ISO COUNTRY CODE	predicates
1	47309286270 8286987	(NOT MERGED)	<null>	{ "coordinates": [[[1.7249253300 00000e+01, 4.0471281500 00000e+01]], { "type": "MultiPoint" }] }	Taranto	ITA	N/A (No Overlap)

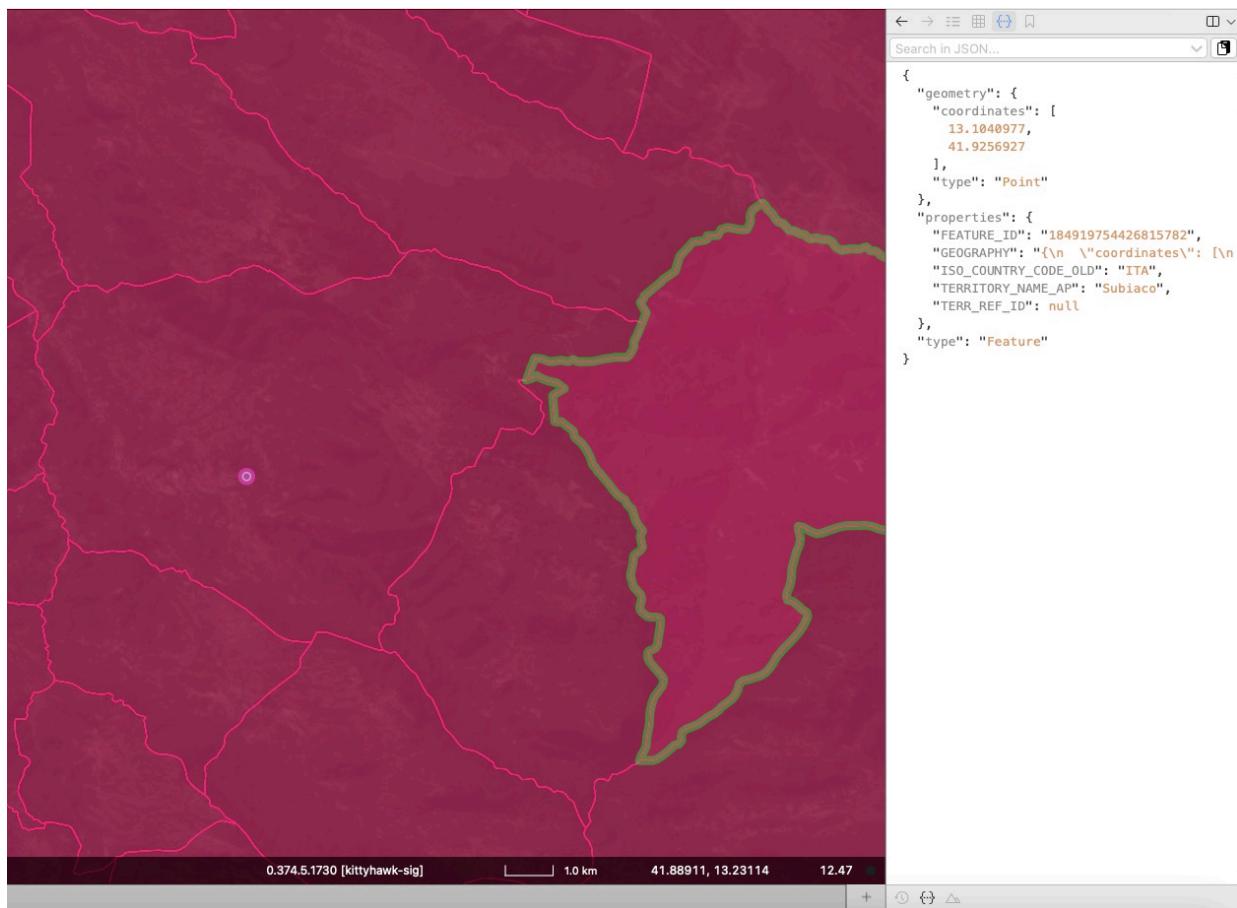
Example A:

(AP out of spatial coverage with null reference (address component) feat id)



Example B:

(AP WITHIN SPATIAL COVERAGE WITH NULL REFERENCE (ADDRESS COMPONENT) FEAT ID)



AP Name Mismatch

Overview of name changes with Vendor Refresh Merge

DISTINCT Municipality Names (ToT): 7,401

DISTINCT Municipality Names (Vendor Refresh): 6,675

Merged Names(both datasets): 6,499

Unmerged Requestor (ToT): 894

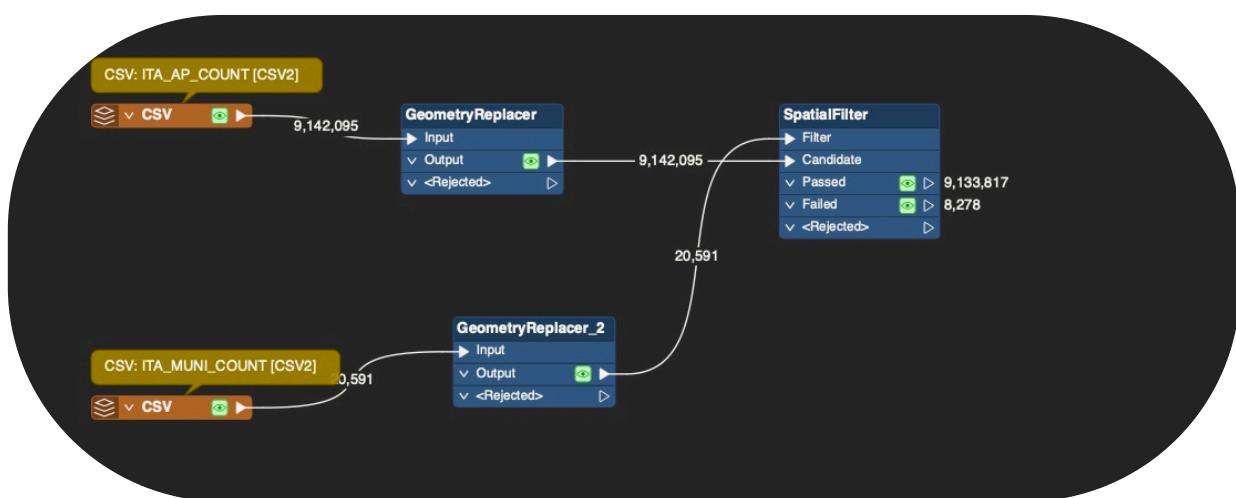
Unused Supplier (Vendor Refresh): 168

Total Mismatch: 476

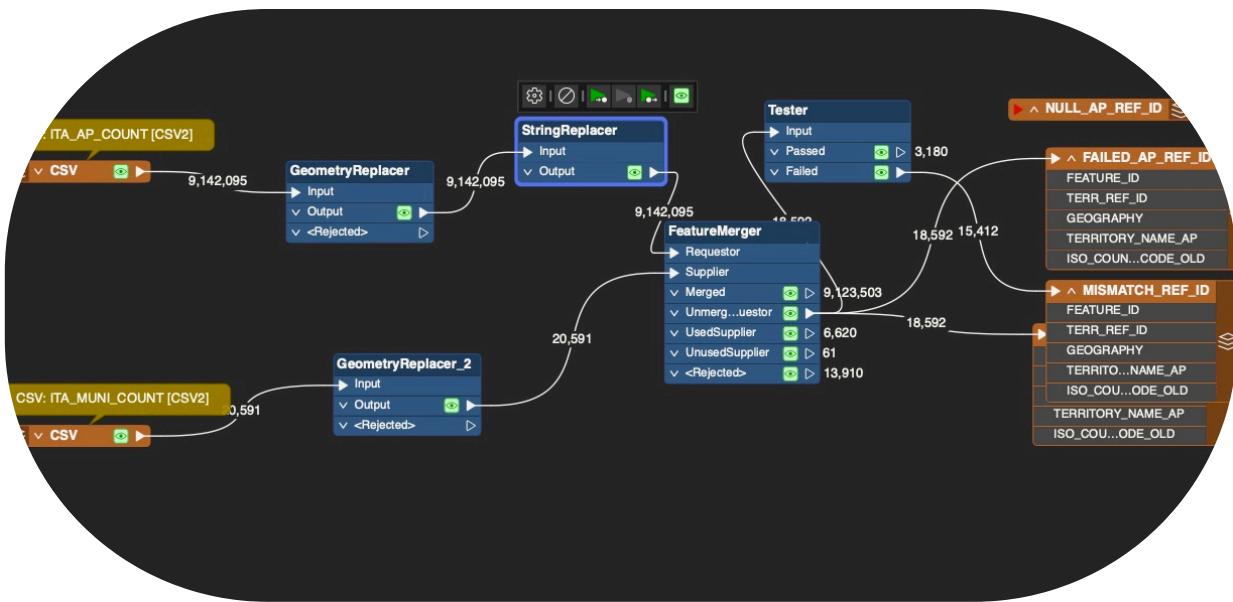
Summary: 6,499 DISTINCT municipality names carried over in Vendor Refresh, 300 Names were no longer present in Vendor Refresh and 176 new names were found in Vendor Refresh

FME Spatial Analysis Workflow

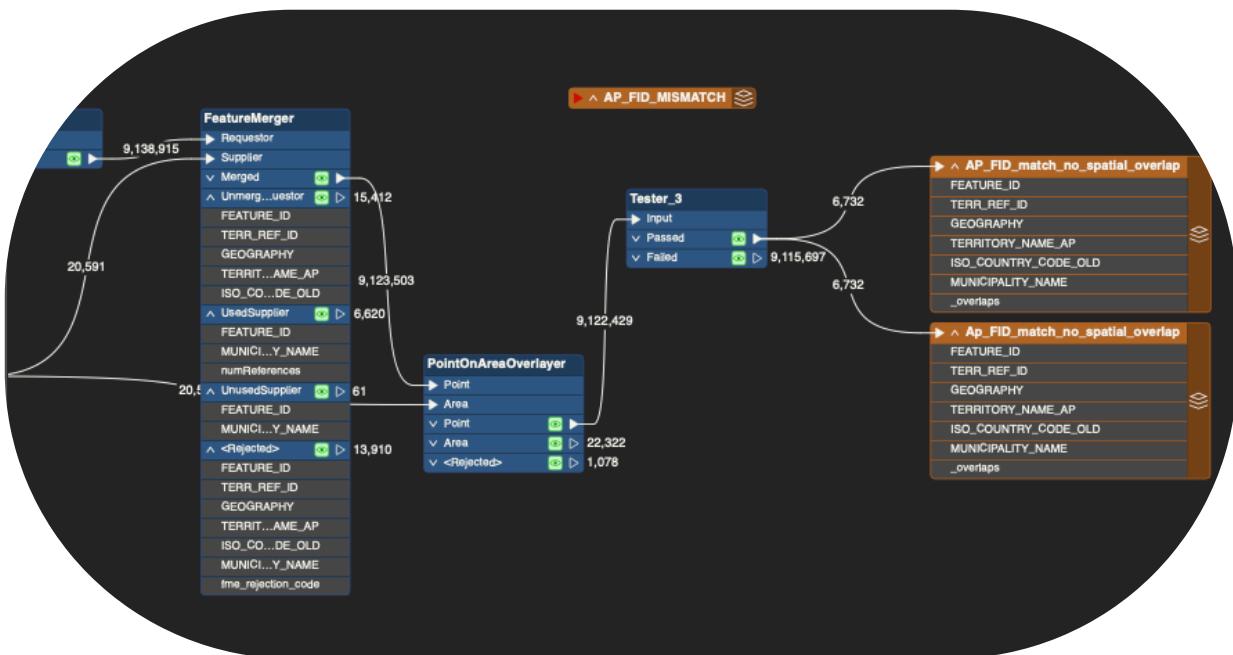
*Spatial Analysis w/ Spatial Filter (Geometry Extractor for Hex Encoded WKB conversion)



*removing “” from reference id for address points with StringReplacer



*testing if matching AP reference id's are spatially within their referenced territories



*Snowflake queries(parsing Feature Proto (nested JSON data))

```
1
2     SELECT
3         ap.feature_id,
4         ap.feature_proto:address_point.address[0].address_component[3],
5         g.wkb AS wkb,
6         g.geography,
7         ap.FEATURE_PROTO:address_point.address[0].address_component[3].name[0].unparsed_name::STRING AS territory_name_ap,
8         ap.FEATURE_PROTO:iso_country_code::STRING AS iso_country_code_old
9
10    FROM    MAPS_DATA_SEMANTIC_DB.TERRITORY_APP.ITA_SOURCE_TABLE_20241020_NOM_DATA ap
11   JOIN    MAPS_DATA_SEMANTIC_DB.TERRITORY_APP.ITA_SOURCE_TABLE_20241014_GEOMETRY_DATA g ON ap.feature_id = g.feature_id
12
13   WHERE   ap.feature_type = 'ADDRESS_POINT' AND ap.feature_proto:address_point.address[0].address_component[3].territory_type = 'MUNICIPALITY'
14
15
```