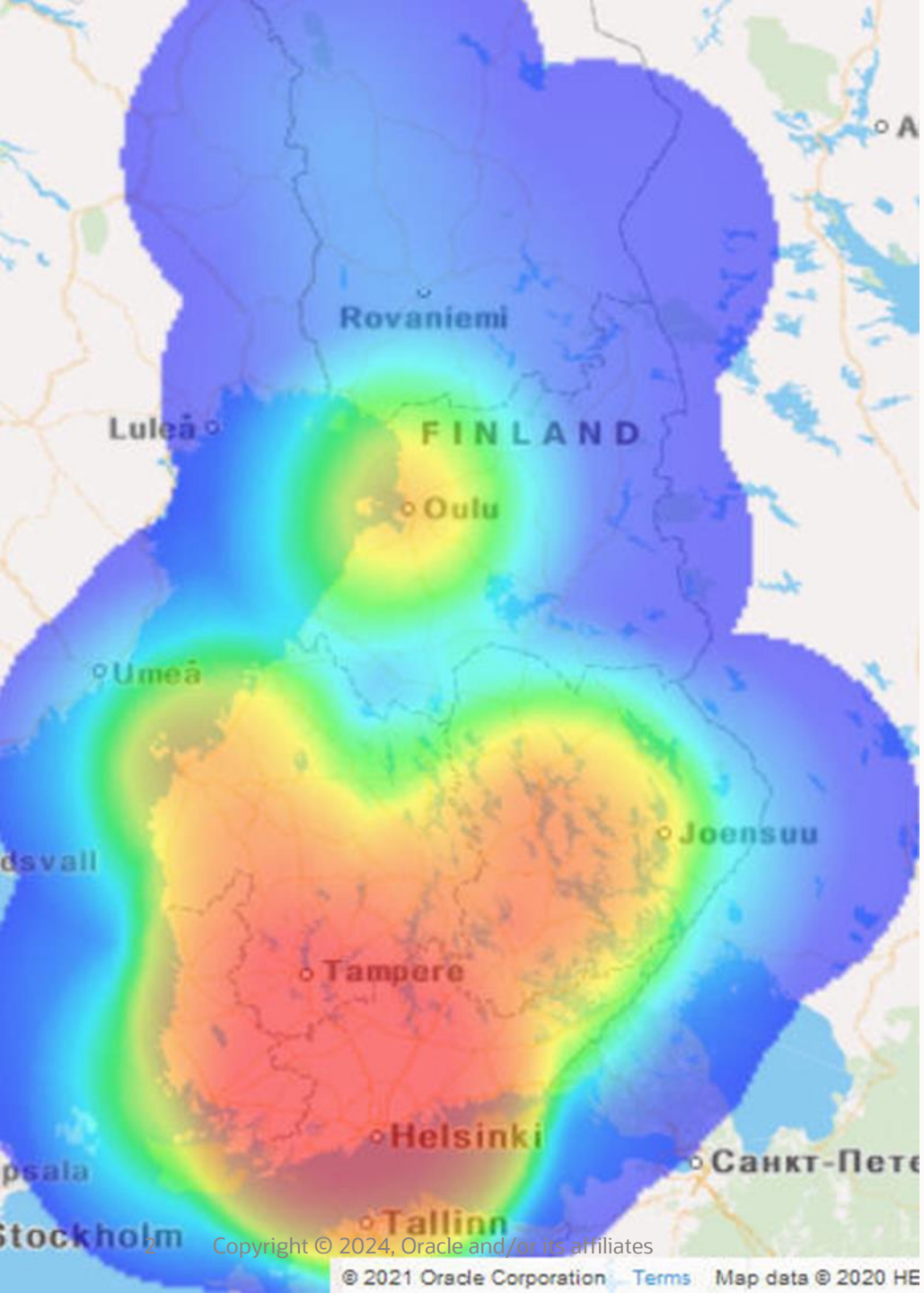
ORACLE

Von Unfallschwerpunkten bis zu Immobilienpreisen – Standortdaten und Machine Learning

SOUG Day 2024

Hans Viehmann
Director Product Management
Oracle Spatial and Graph
September 20, 2024



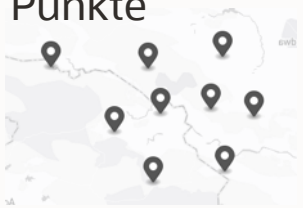
„Everything is related to everything else.
But near things are more related than distant things.”

Tobler's first law of geography

Analyse von raumbezogenen Daten

Punkte

Adressdaten
GPS Positionen
Funkmasten



Linien(züge)

Straßen
Versorgungsleitungen
Flugrouten

Flächen

Landkreise
Waldflächen
Vertriebsgebiete

Welche Gebäude befinden sich im Überflutungsgebiet?

Welche Leitungen sind weniger als 1 km vom Waldbrand entfernt?

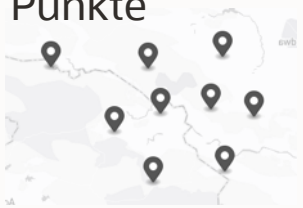
Welche Grundstücke grenzen an das neue Gewerbegebiet?

... von Netzwerkdaten, Rasterdaten, 3D Punktwolken ganz zu schweigen.

Vorhersagemodelle für raumbezogene Daten

Punkte

Adressdaten
GPS Positionen
Funkmasten

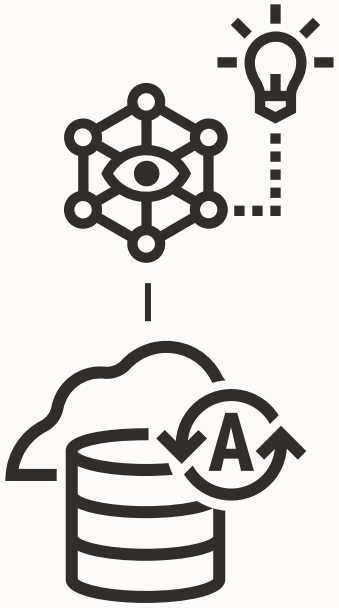


Linien(züge)

Straßen
Versorgungsleitungen
Flugrouten

Flächen

Landkreise
Waldflächen
Vertriebsgebiete



Wie werden sich die Immobilienpreise entwickeln?

Was bedeutet die Neubaustrecke für die Lärmbelastung?

Was beeinflusst das Kaufverhalten in den Marktzellen?

Typische Anwendungsfelder

Clusteranalyse

- Lokale Häufung von Unfällen, Verbrechen, etc.

Ausreißeranalyse

- Betrugserkennung in der Versicherungswirtschaft

Assoziationsanalyse

- Räumliche Korrelation von Einzelhandelsunternehmen, etc.

Regressionsanalyse

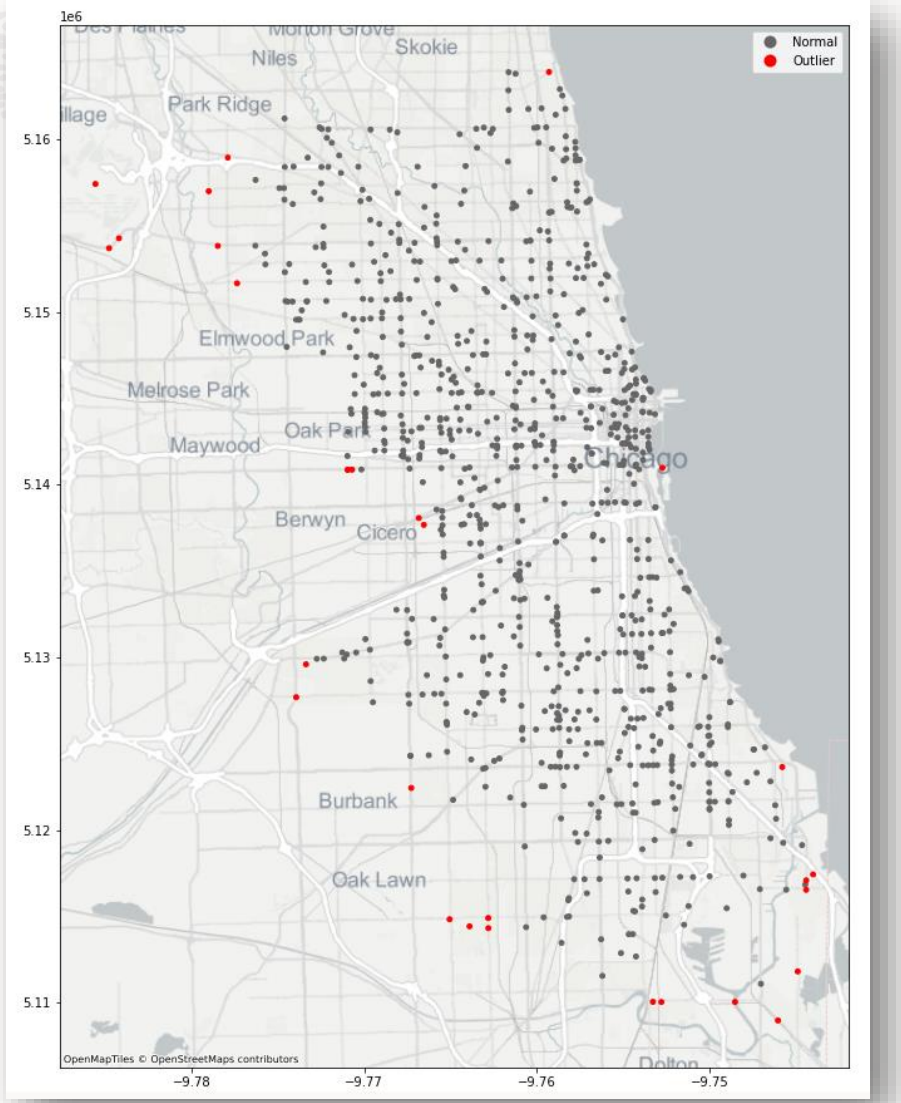
- (Geographisch gewichtete) Bewertung von Immobilien nach Lage

Klassifikation

- (Geographisch gewichtete) Einstufung von Marktzellen

Zusammenfassung

- Räumliche Aggregation in der Datenaufbereitung



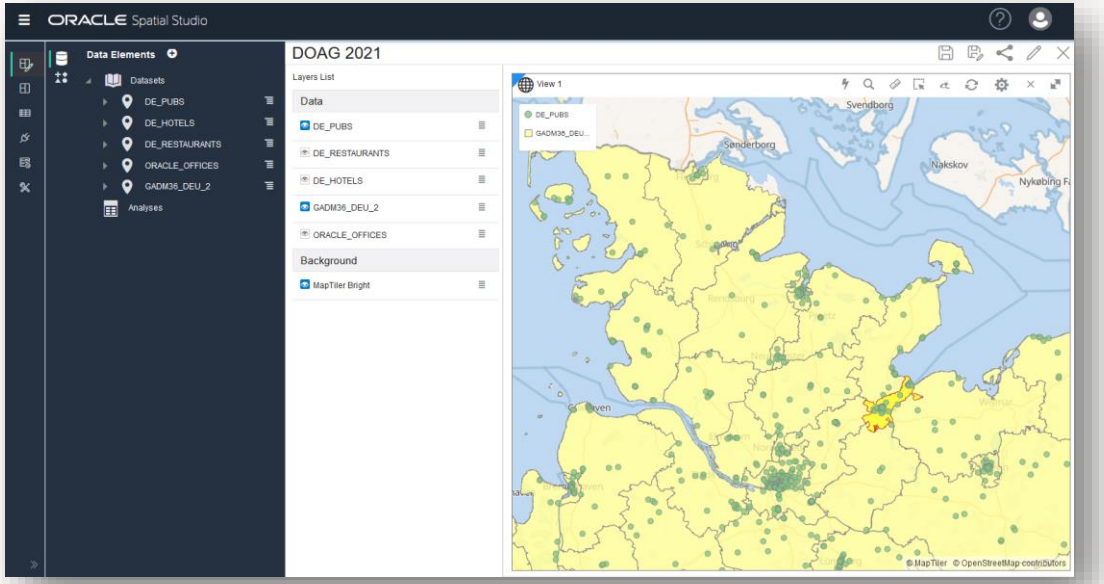
Bestandteil – Oracle Spatial

Management und Analyse von Geodaten in der Datenbank

- Integration räumlicher Daten mit anderen Unternehmensdaten
- Ausführung datenintensiver Logik dort, wo sich die Daten befinden
- Unterstützung für Vektordaten, Rasterdaten, Netzwerke, Punktwolken

Spatial Studio

- Geodaten-Analyse und Kartendarstellung für Laien
- Aufbereitung von raumbezogenen Daten (Adressen, Koordinaten, Geometrien)
- Anwendungsentwicklung und Aufbau von Verarbeitungs-Workflows



Bestandteil – OML4Py in Oracle Autonomous Database

Machine Learning in Oracle Datenbank

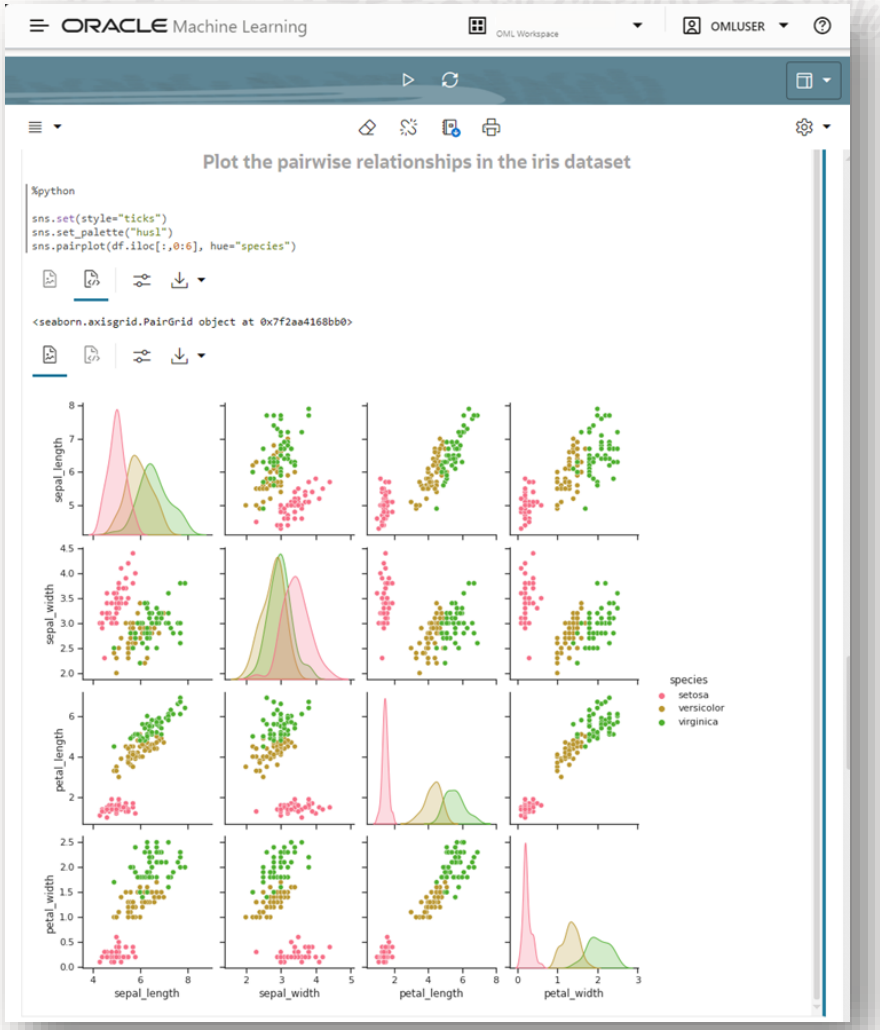
- Optimierte Algorithmen in Datenbank-Engine
- AutoML API

Python Integration

- Abstraktionsschicht für Datenobjekte
  - Daten können in Datenbank bleiben
  - Overloading von Python Funktionen zur Ausführung in SQL
- Dataframes
- Ausführung von Python aus SQL oder REST

Autonomous Database

- Fully managed environment
- Notebooks mit Python, Conda, SQL, ...



Bestandteile von OML4Py Spatial AI

Oracle Machine Learning for Python

Spatial AI

Autonomous Database Serverless

- Python API für alle gängigen räumlichen SQL Operatoren und Funktionen
- Aufbereitung und Nachverarbeitung
  - Feature engineering
  - Gap filling
  - Räumliche Gewichtungsfunktionen
- Räumliche ML Algorithmen
  - Regression
  - Classification
  - Clustering
  - Anomaly detection
  - Colocation
- Unterstützung für Pipelines (à la scikit-learn)
- Python, REST, SQL APIs zur Ausführung

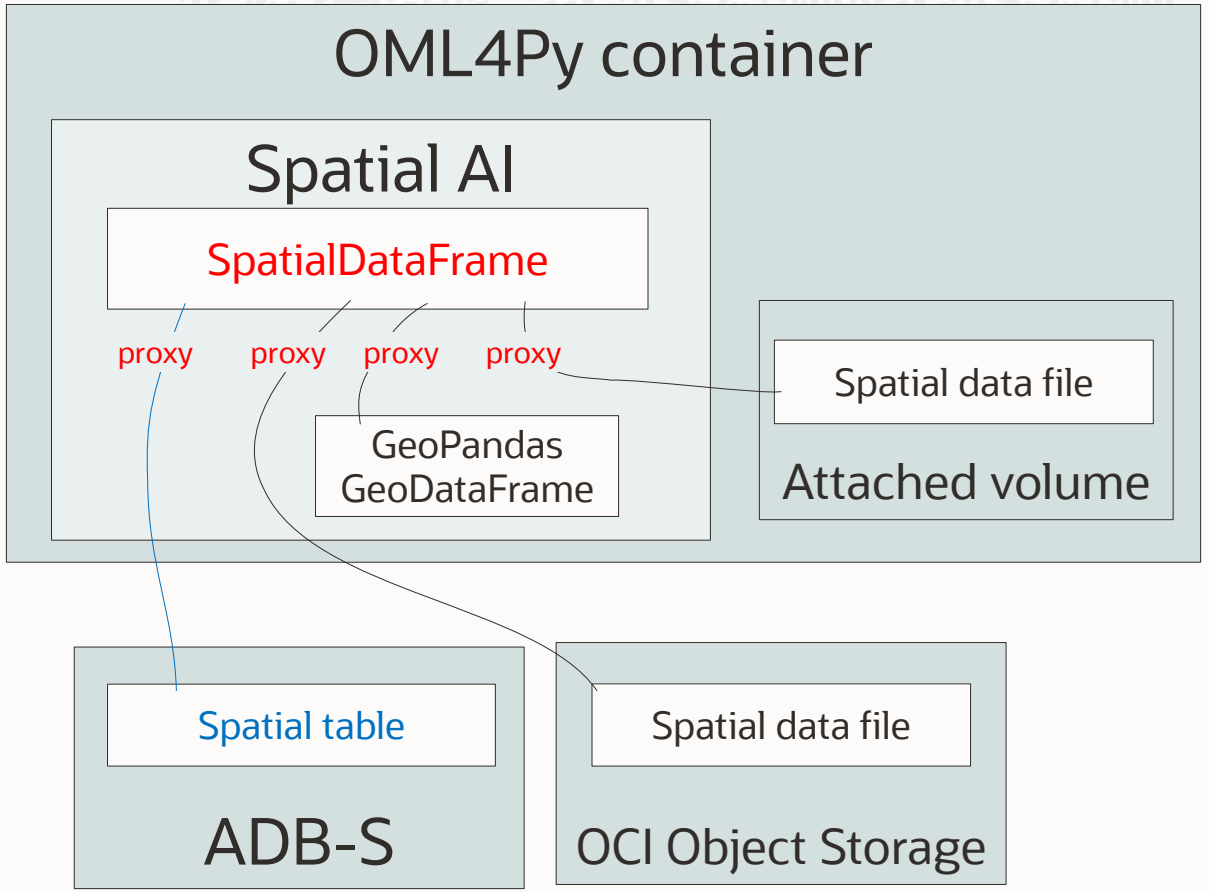
Mögliche Datenquellen

Geodaten aus Oracle Spatial, Object Storage, lokalem Dateisystem

- Gleiche Programmierschnittstelle
- Unterstützung für räumliche Operationen
- Verknüpfung mittels Joins

Räumliche Operationen

- Erfolgen ausschließlich in der Datenbank, wenn SpatialDataFrame ein Proxy-Object für Tabellen in Oracle Spatial ist



Demo

Aufgabe:

- Vorhersage von Verkauf von Elektrofahrzeugen

Daten

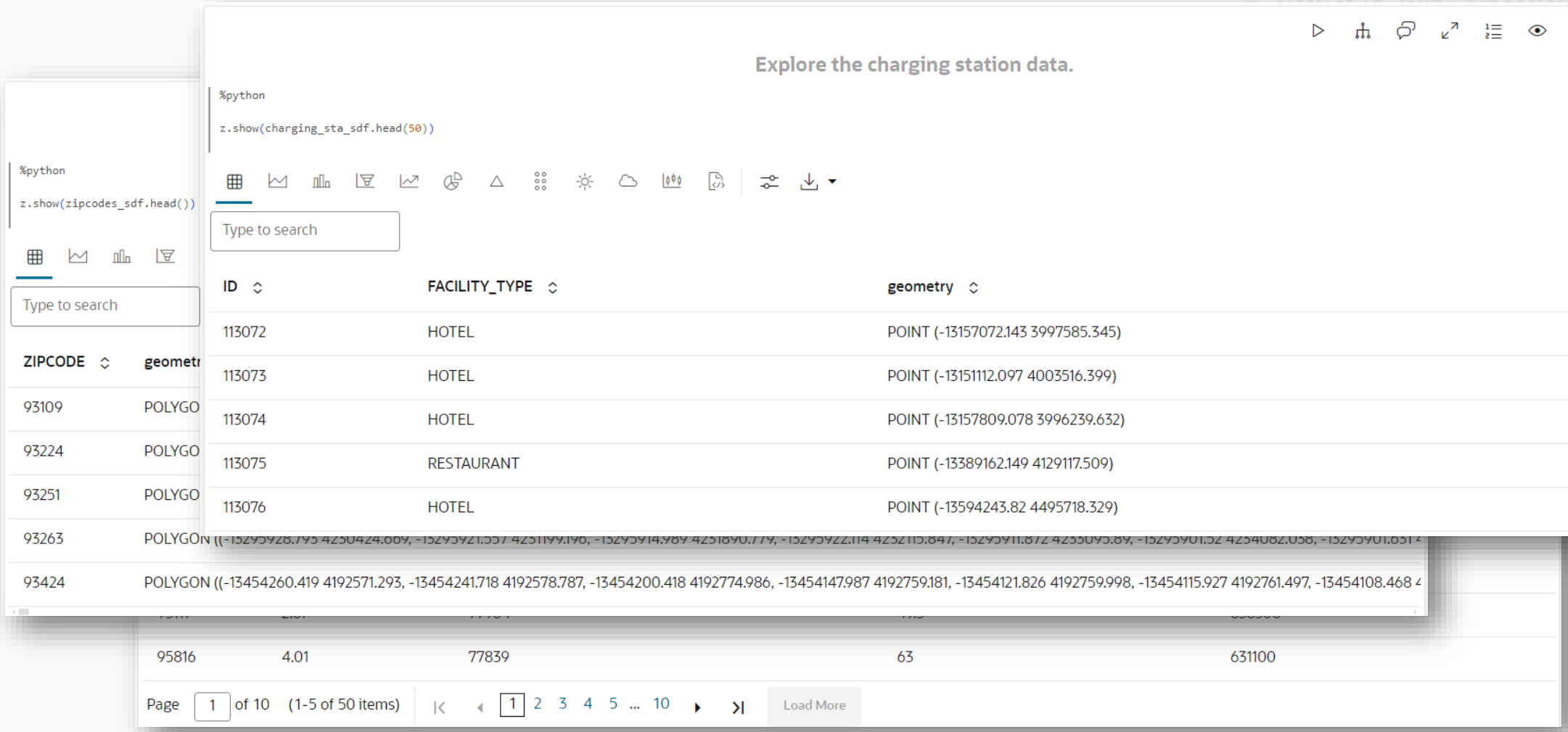
- Sozio-demographische Daten pro Postleitzahlgebiet (Einkommen, Bildung, …)
- Postitionen von Ladestationen
- Geometrien der Postleitzahlgebiete

Wichtige Konzepte

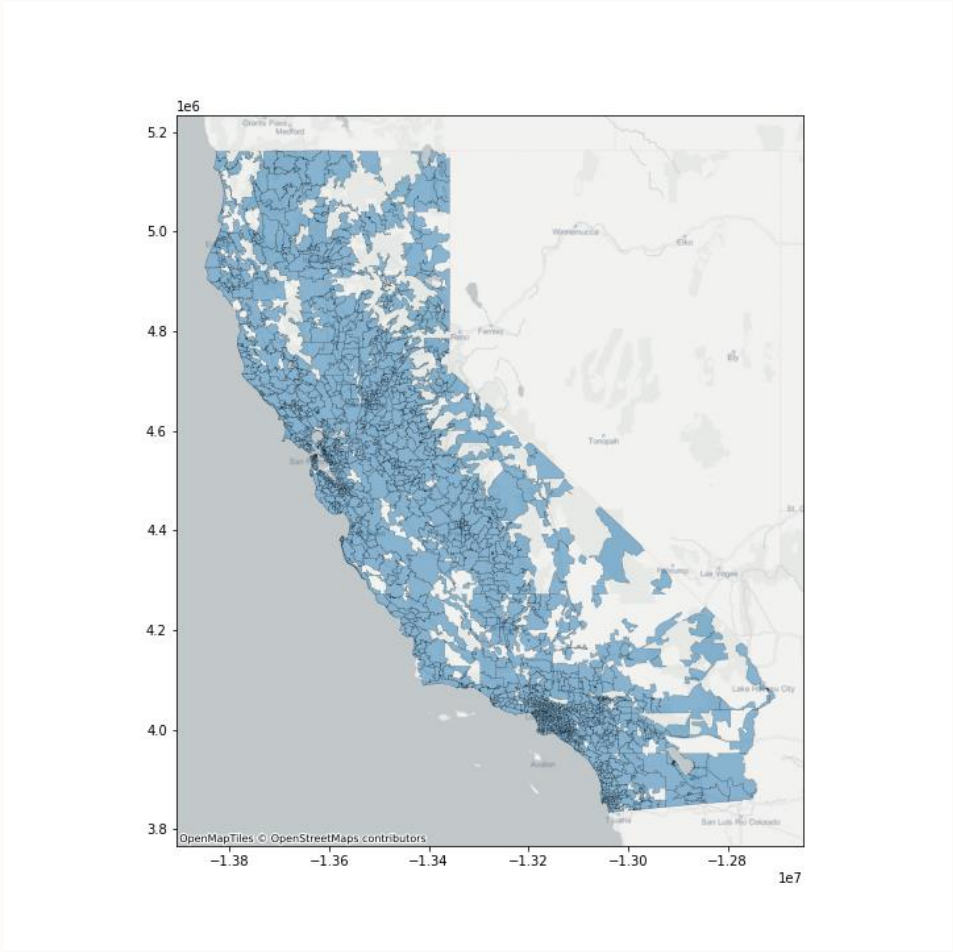
- Handhabung von Geodaten
- Visuelle Analyse
- Interpolation fehlender Werte
- Räumliche Gewichtung
- Autokorrelation (Moran's I)
- Spatial Feature Engineering
- Räumliche Vorhersagemodelle

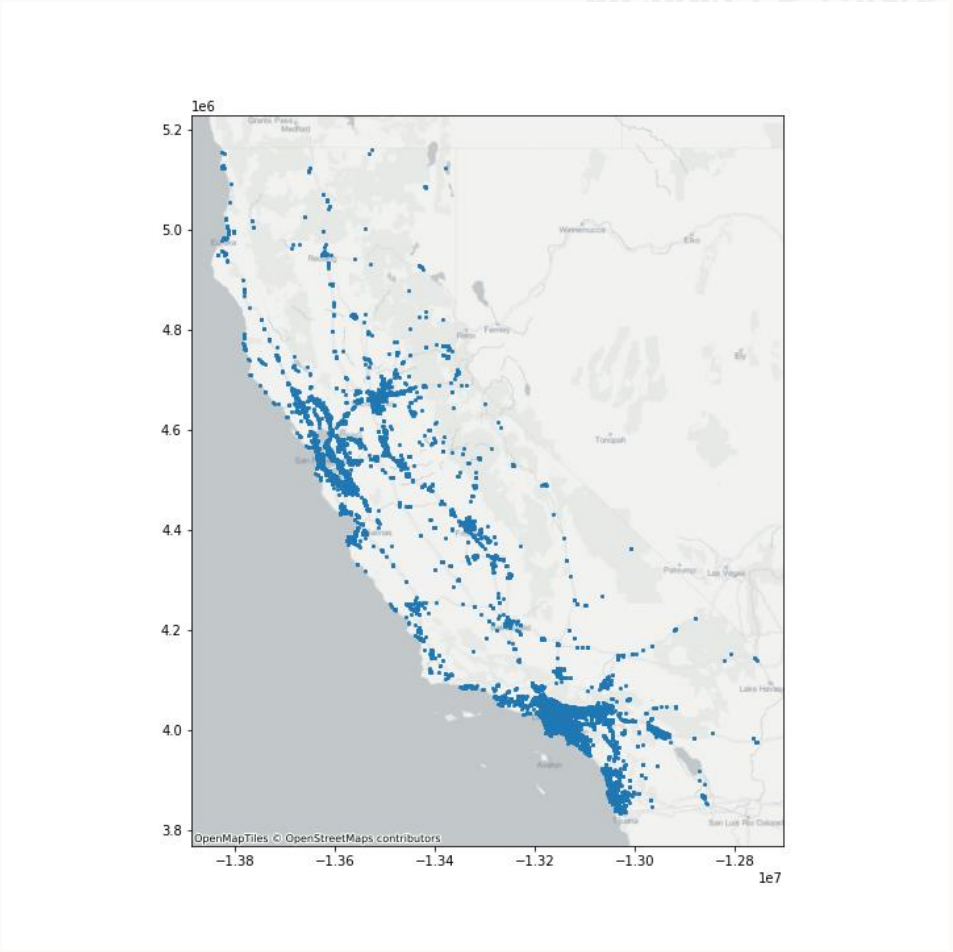
Raumbezogene Daten

Demographische Daten, Postleitzahlgebiete, Ladestationen



Visualisierung





Interpolation fehlender Werte

Räumliche Vorhersagemodelle

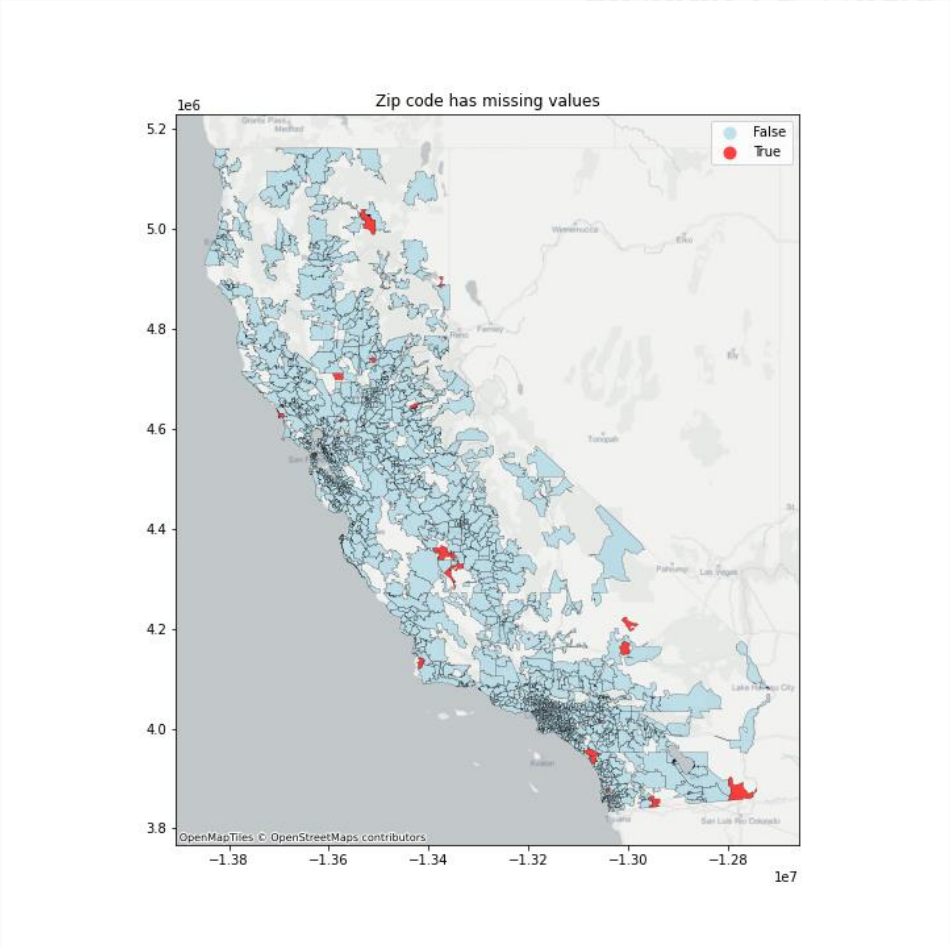
- Probleme bei fehlenden Werten
- Probleme mit fehlenden Nachbarn

Räumliche Gewichtungsmatrix

- Nächste Nachbarn (k nearest neighbors)
- Gemeinsame Grenze (oder Grenzpunkt)
- Gewichtung nach Abstand
- … Abhängig von der Fragestellung

Notwendige Voraussetzung: Gesuchte Variable zeigt räumliche Abhängigkeit

- Maß für Ungleichverteilung: Moran's I (Spatial Autocorrelation)



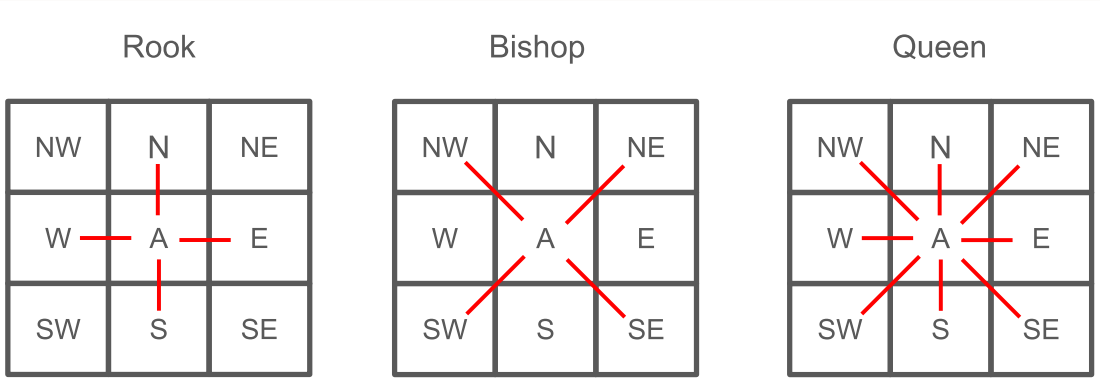
Räumliche Autokorrelation

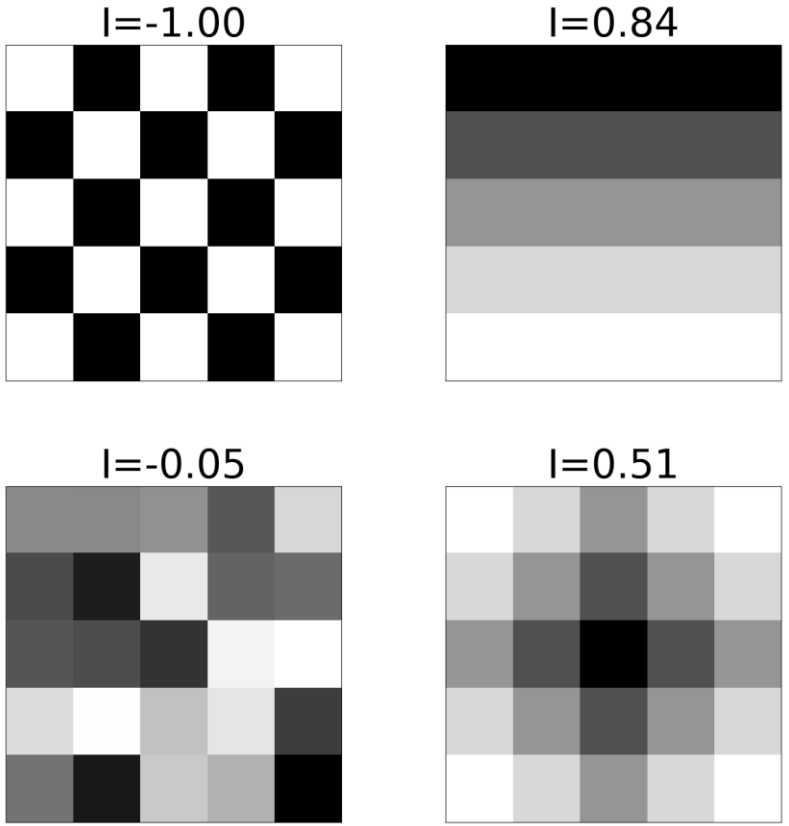
Moran's I

- Erkennung von Clusterbildung oder Streuung (s. rechts)

Nachbarschaftbeziehungen

- Für räumliche Gewichtungsmatrix z.B.





By WikiNukalito - Own work, CC BY-SA 4.0

Feature Engineering und Vorhersagemodelle

Category	Algorithms
Spatial Feature Engineering	Spatial Lag Transformer Categorical Lag Transformer Spatial Coordinates Transformer Spatial Imputer
Spatial Clustering	LISA Hotspot DBSCAN with Regionalization Agglomerative with Regionalization
Spatial Anomaly Detection	Local Outlier Factor (LOF)
Spatial Regression	Spatial Cross-Regressive Model (SLX) Spatial Lag Model (SAR) Spatial Error Model (SEM) Geographical Regressor (GR) Geographically Weighted Regression (GWR) Spatial Regimes (OLS_Regimes) Spatial Fixed Effects (SFE)
Spatial Classification	SLX Classifier GWR Classifier Geographical Classifier

Demo

SpatialDataFrame provides a proxy to Oracle Spatial tables

Versioning low Zeppelin

Create SpatialDataFrame referencing the LA_BLOCK_GROUPS table

```
%python

block_groups_sdf = SpatialDataFrame.create(DBSpatialDataset(table='la_block_groups')) [['GEOID', 'MEDIAN_INCOME', 'MEAN_EDUCATION_LEVEL
z.show(block_groups_sdf.head())
```

Type to search

GEOID	MEDIAN_INCOME	MEAN_EDUCATION_LEVEL	MEAN_AGE	geometry
060376503005	53828.0	11.473	47.585	POLYGON ((-118.359 33.866, -118.352 33.8
060376512222	60724.0	11.35	38.823	POLYGON ((-118.36 33.82, -118.351 33.816, -
060376513026	82538.0	12.801	47.808	POLYGON ((-118.39 33.805, -118.389 33.80
060376513022	143661.0	11.857	46.564	POLYGON ((-118.377 33.81, -118.376 33.811,

ORACLE Machine Learning

samples_v1
OML_USER Workspace

OML_USER

When a SpatialDataFrame references Oracle Spatial data, spatial operations are performed in ADB

Versioning

low

Zeppelin

Identify schools within 2km of a block group

```
%python
z.show(schools_sdf.within_distance(block_groups_sdf[block_groups_sdf['GEOID'] == '060376509014'], distance=2000))
```

Type to search

ID	NAME	geometry
358794039	Van Deene Elementary School	POINT (-118.29 33.838)
358783989	Fern Elementary School	POINT (-118.332 33.835)
358794865	Howard Wood Elementary School	POINT (-118.321 33.813)
358779437	John Adams Elementary School	POINT (-118.318 33.809)

Preprocessing example: feature engineering to add a spatial metric

```
%python
block_group_nn_school_sdf = schools_sdf.nearest_neighbors(\
                                        block_groups_sdf, num_neighbors=1,\
                                        distance_col='DIST_TO_NEAREST_SCHOOL', qry_win_out_cols='GEOID')
block_group_nn_school_sdf = block_group_nn_school_sdf[['GEOID','DIST_TO_NEAREST_SCHOOL']]
z.show(block_groups_sdf.merge(block_group_nn_school_sdf, how='left', on='GEOID', keep_secondary_geometry=False).head())
```

Type to search

GEOID	MEDIAN_INCOME	MEAN_EDUCATION_LEVEL	MEAN_AGE	DIST_TO_NEAREST_SCHO
060376503005	53828.0	11.473	47.585	275.45
060376512222	60724.0	11.35	38.823	195.29
060376513026	82538.0	12.801	47.808	617.2

Preprocessing example: fill gaps of data in Oracle Spatial with spatial ML algorithm

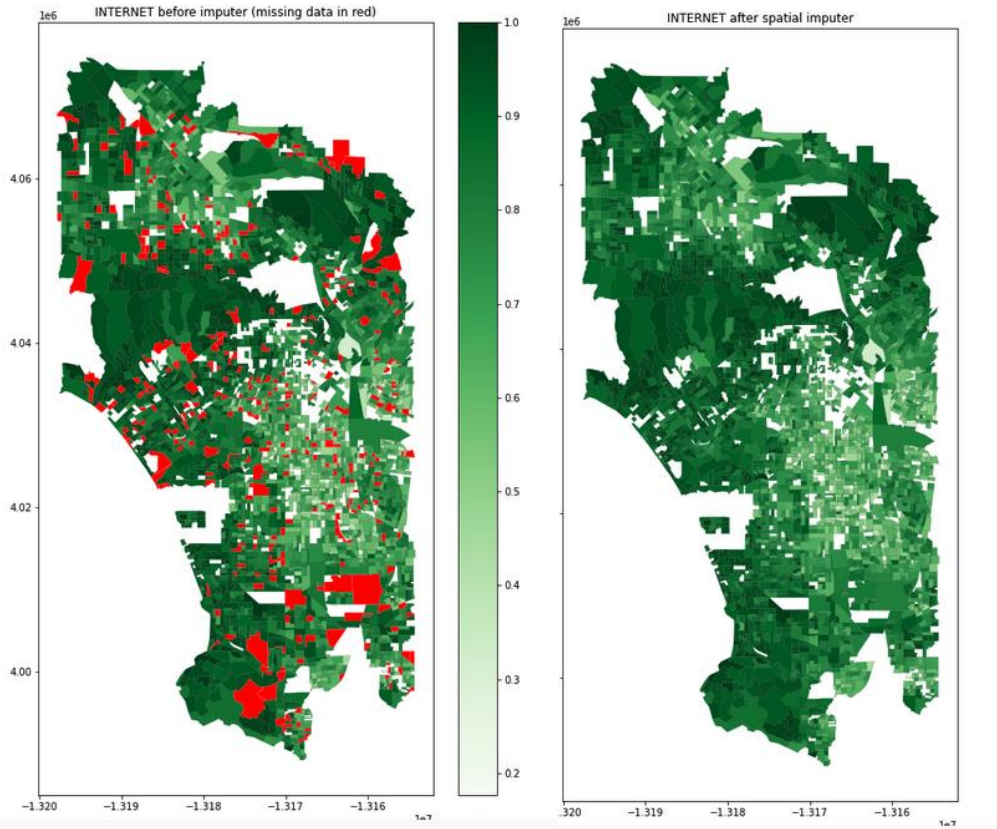
Versioning low Zeppelin

```
%python

spatial_imputer = SpatialImputer(missing_values=np.nan, spatial_weights_definition=KNNWeightsDefinition(k=10))

X_imputed = spatial_imputer.fit_transform(X, y="MEDIAN_INCOME")
X_imputed_df = pd.DataFrame(X_imputed, columns = ['MEAN_EDUCATION_LEVEL',
X_imputed_df
```

	MEAN_EDUCATION_LEVEL	MEAN_AGE	INTERNET	HOUSE_VALUE
0	12.198391	38.452568	0.894470	1566400.0
1	12.242844	40.621918	0.915416	1145300.0
2	12.353305	36.314709	0.985325	1440600.0
3	12.238597	37.198551	0.888268	1624100.0
4	13.007787	44.285275	1.000000	1417200.0
...	...	...	...	...
3432	8.288300	36.015499	0.791828	363900.0
3433	11.279384	41.882759	0.958647	782200.0
3434	11.772903	36.934315	0.837050	758900.0



Detect density clusters in data from Oracle Spatial

Versioning low Zeppelin

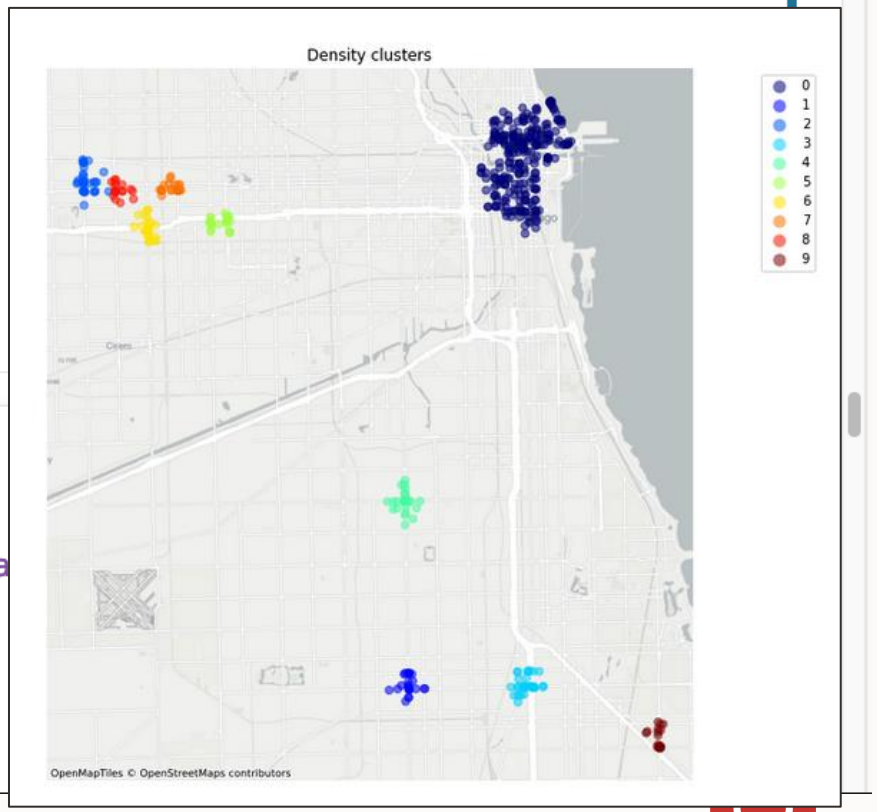
```
%python

X = accidents_injury_sdf[['geometry']].to_crs("EPSG:3857")
labels = dbscan.fit_predict(SCoordTransformer().transform(X))
np.unique(labels)
```

```
array([-1,  0,  1,  2,  3,  4,  5,  6,  7,  8,  9])
```

```
%python

_, axs = plt.subplots(figsize=(10, 10))
plot_clusters(X, labels, title='Density clusters', with_noise=False, with_bounds=Fa
```



Exploratory analysis and Global Spatial Autocorrelation of data in Oracle Spatial

Versioning

low Zeppelin

```
%python

for i, weights_spec in enumerate(weights_specs):
    mi_result = MoranITest.create(sdf, weights_spec, column_name='MEDIAN_INCOME')
    print(f"Weights definition: {mi_result.spatial_weights.definition}\nMoran I's results: I:{mi_result.i}, p value:{mi_res
```

```
Weights definition: DistanceBandWeightsDefinition(alpha=-1.0,threshold=1000.0,p=2,binary=True)
Moran I's results: I:0.5719605389665813, p value:0.001

Weights definition: KNNWeightsDefinition(k=2)
Moran I's results: I:0.6945150221963798, p value:0.001

Weights definition: KernelBasedWeightsDefinition(fixed=False,function=gaussian,bandwidth=None,k=11)
Moran I's results: I:0.663147087878812, p value:0.001

Weights definition: QueenWeightsDefinition()
Moran I's results: I:0.6765793161449112, p value:0.001
```

ORACLE Machine Learning

samples_v1
OML_USER Workspace

OML_USER

Detect hotspots and coldspots in data from Oracle Spatial

Versioning

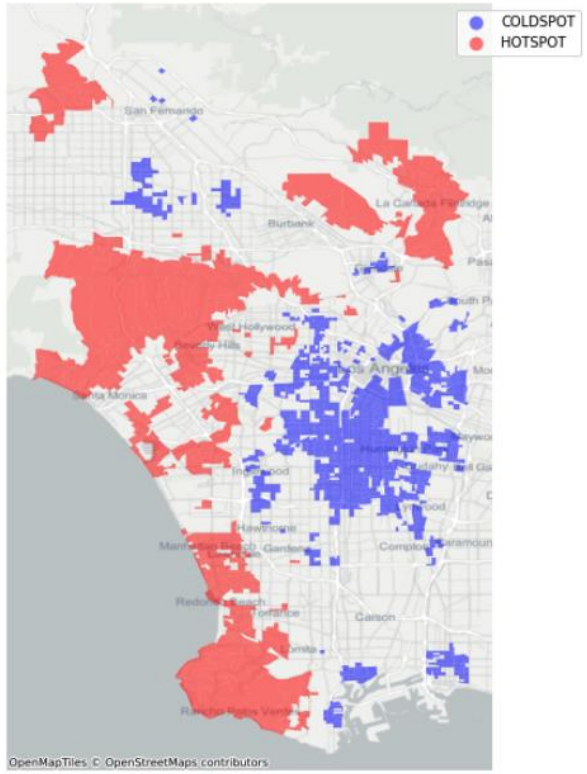
low

Zeppelin

```
%python

X = block_groups_sdf[['MEDIAN_INCOME','geometry']].to_crs('epsg:3857')
lisa = LISAHotspotClustering(max_p_value=0.05, spatial_weights_definition=KNNWeightsDefinition(k=5))
lisa.fit(X)
np.unique(lisa.labels_)
```

```
array([-1,  1,  2,  3,  4])
```



Create desciptive labels for hotspots and coldspo

```
%python

get_label = lambda x: 'HOTSPOT'  if x==1 else ( \
                      'COLDSPOT' if x==3 else ( \
                       None ))
vf = np.vectorize(get_label)
block_groups_labelled = block_groups_sdf.add_column('QUADRANT', lisa.labels_)
```

SpatialPipeline: mirrors scikit-learn Pipeline with support for spatial estimators

Versioning

low Zeppelin

```
%python

gwr_pipe = SpatialPipeline([('scale', RobustScaler()), \
                            ('gwr',   GWRRegressor(spatial_weights_definition=KNNWeightsDefinition(k=20)))])
gwr_pipe.fit(X_train, "MEDIAN_INCOME")
gwr_pipe.predict(X_test.drop(["MEDIAN_INCOME"]))[0:9]
```

```
array([[ 76037.81477983],
       [109221.70033452],
       [172357.88353423],
       [ 35776.66364485],
       [ 72946.22992853],
       [ 48016.12874966],
       [142154.11139249],
       [156999.86728209],
       [ 36118.72778437]])
```

2s 511ms @ a few seconds ago

Auto-selection of spatial regression algorithm

Versioning

low Zeppelin

```
%python
spreg_pipeline = SpatialPipeline([('scale', StandardScaler()), \
                                  ('spreg_regression', SpatialAdaptiveRegressor(spatial_weights_definition=KNNWeightsDefini
spreg_pipeline.fit(X_train, "HOUSE_VALUE")
print(f"Algorithm chosen: {spreg_pipeline.named_steps['spreg_regression'].model_type.__name__}")
```

```
Algorithm chosen: LagModel
```

```
%python
spreg_pipeline.predict(X_test.drop("HOUSE_VALUE")).flatten()
```

```
array([ 388350.98819514,  396173.58755382,  366258.12083167,
        360407.41307687,  729597.63070832,  694518.01543814,
        483702.23142709,  414668.27240398,  571455.33258045,
```

Zusammenfassung

Raumbezogene Analyse in Oracle Autonomous Database mittels Python

- Skalierbare Datenaufbereitung und Analyse
- Zugang zu vielfältigen Spatial ML Algorithmen
- Integration mit Unternehmensdaten
- Offene Schnittstellen zur Anwendungsintegration

Data Science Plattform auf Basis der Oracle Autonomous Database

- Einfaches Deployment von Data Science Lösungen
- Automatisierung erhöht die Produktivität und erleichtert den Einstieg

Weitere Informationen

- Spatial Features Homepage: oracle.com/goto/spatial
- Spatial Studio: https://www.oracle.com/.../spatial-studio/get-started.html
- YouTube Channel: youtube.com/c/OracleSpatialandGraph
- Blog: blogs.oracle.com/oraclespatial -> Database Insider (Spatial category)
- LiveLabs Workshops: http://bit.ly/golivelabs
- AskTOM sessions: https://asktom.oracle.com/pls/apex/asktom.search?oh=7761
- Oracle Spatial and Graph User Group: linkedin.com/groups/1848520/


ORACLE