

# Raum-Zeit-Analysen und räumliche Assoziationen – eine neue Funktion im IBM SPSS Modeller und IBM SPSS Statistics

von Dr. Christian Trippner

IBM, Senior Technical Sales Professional, Predictive Analytics

Das Statistikpaket IBM SPSS Statistics gibt es nunmehr schon seit mehr als 47 Jahren auf dem internationalen Markt, die Data Mining Workbench IBM SPSS Modeller gibt es immerhin seit über 19 Jahren.

Sowohl SPSS Statistics als auch der SPSS Modeller sind sogenannte „stand alone“ Produkte die unabhängig voneinander betrieben werden können.

| Case | Longitude     | Latitude      | Bevölkerung | Qualität | POP15 | POP15 | POP15 | Density | Price  | Hypothek | Prognose |
|------|---------------|---------------|-------------|----------|-------|-------|-------|---------|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1    | 96.7023707191 | 32.7182771813 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 2    | 96.6702302032 | 32.7191901939 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 3    | 96.7023048646 | 32.7171917864 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 4    | 96.7191704437 | 32.7017022478 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 5    | 96.7042302041 | 32.6919046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 6    | 96.6711448780 | 32.6904833460 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 7    | 96.7070421790 | 32.7011904647 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 8    | 96.7042302041 | 32.6702302041 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 9    | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 10   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 11   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 12   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 13   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 14   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 15   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 16   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 17   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 18   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 19   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 20   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 21   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |
| 22   | 96.6702302041 | 32.7064046697 | 964790      | 38       | 1196  | 1027  | 1.16  | 3144.73 | 100.00 | 2702.63  | 49900.00 | 1.55     | 74       | 4.54     | 9.38     | 92.00    | Voluntar |          |

Abbildung 1: SPSS Statistics Version 23

Beide Produktgruppen sind mittlerweile sehr stark im Markt etabliert und in vielen Unternehmen ein unternehmenskritisches Werkzeug um betriebliche, strategische und planerische Entscheidungen durch wissensbasierte und maschinenintelligente Modelle und automatisierte Anwendungen zu unterstützen.

Eines der charakteristischen Merkmale beider Plattformen ist die Komplexität der Applikation. Das heißt alle im Analyseprozess relevanten Schritte, wie Datenzugriff, Datenvalidierung, Datenaufbereitung, Visualisierung, Modellierung und der Modellexport, werden vom jeweiligen Werkzeug komplett und umfassend abgebildet.



Abbildung 2: SPSS Modeller Version 17

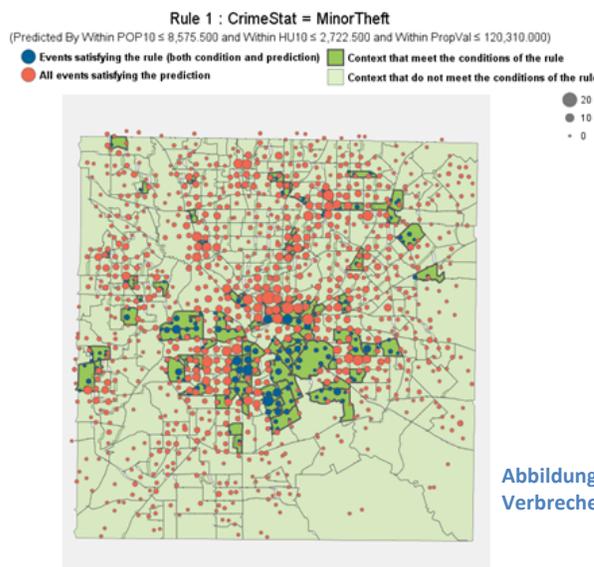
Es gibt am Markt nur wenige Programme, die diese Komplexität in der selben Bandbreite wie die IBM Analysewerkzeuge SPSS Statistics und SPSS Modeller abbilden können.

Das ist sicher auch mit einer der Gründe, warum sich beide Tools in Forschung und Lehre, aber auch in zahlreichen Unternehmen, Firmen und Organisationen als quasi Standardapplikation etablieren konnten. Aktuell ist die SPSS Statistics Version 23 und die SPSS Modeller Version 17 am Markt im Einsatz. Neben Standardanwendungen wie deskriptive Analysen (Häufigkeiten, Lage-, Streuungs- und Verteilungsmaße) und unterschiedliche hypothesengetriebene Testverfahren (CHI<sup>2</sup>-Test, t-Test, KS-Test, ...), sind auch zahlreiche weitergehende und multivariate Analysen möglich (Regression, Log. Regression, GLM-Verfahren, Cluster-, Faktor, Baumanalysen, Cox-Regression, Survivalanalyse und zahlreiche andere Verfahren). Abgerundet wird die Komplexität durch zahlreiche Analyseassistenten für eine automatische Datenvalidierung, Erkennung von Anomalien in den Daten oder auch eine automatische Datenaufbereitung.

Seit dem letzten Release bietet die IBM in den SPSS Analysetools eine neue Funktion, die es erlaubt Raum-Zeit gebundene Berechnungen und Modellierungen durchzuführen.

Diese **georäumlichen Modellierungsverfahren** ermitteln Muster in Daten, die eine räumliche Komponente (Kartenkomponente) enthalten.

Der Geomodellierungsassistent stellt Methoden für die Analyse von Geodaten mit und ohne Zeitkomponente bereit. Dabei werden im Raum vorkommende Assoziationen und zeitliche Abfolgen von Ereignissen über den Raum untersucht und danach kartographisch dargestellt. So lassen sich assoziative Raum-Zeitbezüge erstellen, die z. B. bei der Ressourcenplanung, bei der vorausschauenden Wartung (PMQ), bei der Qualitätssicherung oder im medizinischen Bereich Anwendung finden können, um nur einige wenige Anwendungsmöglichkeiten zu nennen.



Mit dieser Methode haben Landespolizeiämter in den USA (vergleichbar mit den deutschen Landeskriminalämtern) über derartige Vorhersagen, im zeitlichen Verlauf variierende Verbrechenschwerpunkte vorhergesagt, Assoziationsmuster bei Delikten erkannt, und so über eine optimierte Ressourcen- und Einsatzplanung von Streifendiensten, eine Senkung der Kriminalitätsrate erreicht.

Abbildung 3: Räumliche Verteilungsmuster von Verbrechenschwerpunkten

Zwei grundsätzliche Anwendungsszenarien sind hier verfügbar:

- **Assoziationen** auf der Basis von Ereignis- und Geodaten suchen (Geoassoziationsregeln)  
Mithilfe von Geoassoziationsregeln können Sie Muster in Daten auf der Basis von sowohl räumlichen als auch nicht räumlichen Eigenschaften suchen. Sie können beispielsweise Muster in kriminologischen Daten nach Position und demografischen Attributen ermitteln. Anhand dieser Muster können Sie Regeln erstellen, die vorhersagen, wo bestimmte Arten von Verbrechen wahrscheinlich auftreten.
- Vorhersagen mit Hilfe von **Zeitreihen** und Geodaten erstellen (räumlich-temporale Vorhersage).  
Die räumlich-temporale Vorhersage verwendet Daten, die Positionsdaten, Eingabefelder für die Vorhersage (Prädiktoren), mindestens ein Zeitfeld und ein Zielfeld enthalten. Jede Position enthält mehrere Zeilen in den Daten, die die Werte jedes Prädiktors und das Ziel in den einzelnen Zeitintervallen darstellen.

## Vorüberlegungen zum Kartenmaterial

Um eine Analyse von Geoassoziationsregeln durchführen zu können, sind bestimmte Voraussetzungen nötig:

Zuerst wird eine zu den Daten passende und georeferenzierte Karte im SHP-Format spezifiziert<sup>1</sup>. Dabei ist darauf zu achten, daß der richtige Projektionsmodus gewählt wird, um eine verzerrungsfreie Abbildung/Darstellung zu bekommen.

Bei der Projektion der Erde auf eine Karte tut sich ein grundsätzliches Problem auf, da sich die gewölbte Kugelgestalt der Erde nicht fehler- und verzerrungsfrei auf eine flache/ebene Plankarte übertragen läßt. Dadurch erhält man je nach Projektionsmodus verschieden präzise Karten.



Bildhaft kann man sich vorstellen, daß selbst eine perfekt und am Stück geschälte Orange nicht auf eine ebene Tischplatte aufgebracht werden kann, ohne daß es zu Rissen und Verzerrungen kommt.

Durch die unterschiedlichen Projektionsarten verändern sich die Längen einer Strecke, oder die Größe und Form einer Fläche bzw. der Winkel zwischen zwei Linien.

Sämtliche Kartenprojektionen enthalten mindestens eine Form dieser Verzerrungen, weshalb man sich für bestimmte Vor- und Nachteile unter diesen Abbildungseigenschaften entscheiden muß:

- längentreue (äquidistante) Abbildung – einige Strecken sind korrekt abgebildet
- flächentreue (äquivalente) Abbildung – alle Flächen sind dem Maßstab entsprechend korrekt abgebildet
- winkeltreue (konforme) Abbildung (beispielsweise zur Navigation)
- vermittelnde Verzerrungseigenschaften – Kompromisse zwischen Längentreue, Flächentreue oder Winkeltreue.

Einen weiteren weltweiten Standard stellt das World Geodetic System 1984 (WGS 84) dar. Dabei handelt es sich um ein geodätisches Referenzsystem als einheitliche Grundlage für Positionsangaben auf der Erde und im erdnahen Weltraum.

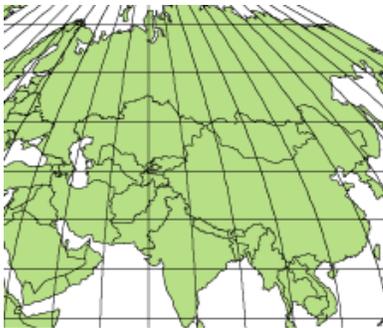
Das WGS84 schließt sich an die tatsächliche Gestalt der Erde bestmöglich an; die lokale Abweichung beträgt maximal 100 m.

Für Europa liegen die Abweichungen zwischen 50 und 60 m.

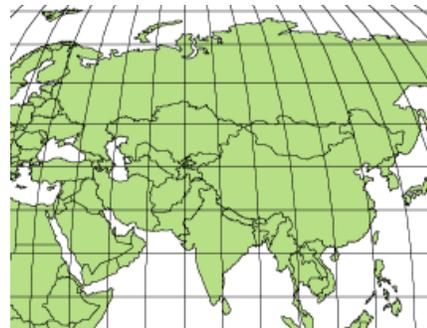
<sup>1</sup> Gutes Kartenmaterial ist bei den deutschen Landesvermessungsämtern oder dem Bundesamt für Kartographie erhältlich, sowie zahlreichen privatwirtschaftlichen Unternehmen (z. B. ESRI®, oder kostenfrei bei Open Street Map)

Die gängigen Projektionsverfahren lassen sich wie folgt gliedern (siehe auch Abbildungen unten<sup>2</sup>):

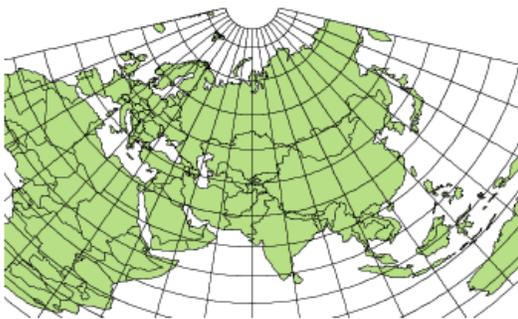
1. Längentreue Abbildungen (Längentreue Kegelprojektion, Sinusoidal-Projektion)  
Diese Abbildungen sind nicht absolut längentreu, sondern lassen nur entlang weite-  
rer Linien als den Berühr- bzw. Schnittkreisen längentreue Messungen zu.
2. Flächentreue Abbildungen (z. B. Flächentreue Kegelprojektion, Goodes-Projektion)  
Diese Abbildungen stellen die Größe einer Fläche (z. B. eines Kontinents) korrekt dar.  
Die Form kann allerdings sehr stark verzerrt werden. Insbesondere am Kartenrand  
neigen diese Abbildungen zu starken Formverzerrungen.
3. Winkeltreue Abbildungen (z. B. Mercator-Projektion, Stereographische Projektion)  
Winkeltreue Abbildungen werden insbesondere bei der Navigation in der Schifffahrt  
und im Flugverkehr zur Erstellung von Karten benötigt.
4. Vermittelnde Abbildungen (z. B. Robinson, Miller-Zylinderprojektion)  
Da keine Kartenabbildung alle Verzerrungen vollständig aufhebt, wurden einige ver-  
mittelnde Abbildungen als Kompromiss entwickelt. Bei ihnen wurde versucht die  
Verzerrungen zu minimieren.



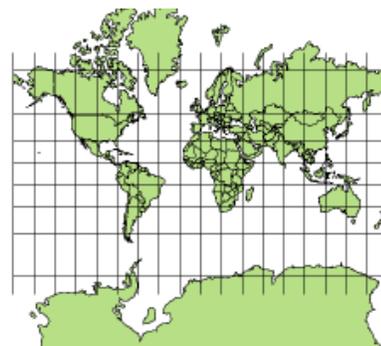
Sinusoidale Projektion



Robinson-Projektion



Flächentreue Kegelprojektion



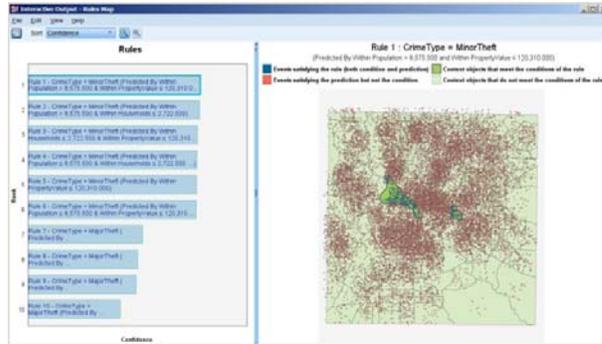
Mercator-Projektion

<sup>2</sup> Quelle: Wikipedia (<http://de.wikipedia.org/wiki/Kartennetzentwurf>), verändert

Doch zurück zur Raum-Zeit-Analyse mit den Modulen von SPSS Statistics und SPSS Modeler.

**Die georäumliche Modellierung** kann eine oder mehrere Kartendatenquellen verwenden. Kartendatenquellen enthalten Informationen, die geographische Bereiche oder andere geographische Objekte wie Straßen oder Flüsse definieren. Viele Kartenquellen enthalten darüber hinaus auch demographische oder andere beschreibende Daten und Ereignisdaten wie Kriminalberichte oder Arbeitslosenquoten.

Hier können Sie eine zuvor definierte Kartenspezifikationsdatei verwenden oder Kartenspezifikationen definieren und diese Spezifikationen für die spätere Verwendung speichern.



Die Assoziationsregeln benötigen mindestens einen Ereignisdatensatz, der Information über bestimmte Ereignisse enthält (Abwanderung, Straftat, gekauftes Produkt, Unfall, ...). Dazu können dann noch optional Kontextdaten mit Kartenmerkmalen, wie mittleres Einkommen, Anzahl der Kinder oder Bevölkerungsdichte, herangezogen werden.

Abbildung 4: Georäumliche Modellierung

Anschließend erfolgt die Definition der Assoziationsregeln. Assoziationsregeln sagen Werte der Vorhersagefelder basierend auf den Werten der Bedingungsfelder voraus. In der Regel "If x=1 and y=2, then z=3" sind die Werte von x und y beispielsweise Bedingungen und der Wert von z ist die Vorhersage.

Rules Table - Rule Support

| Rank | Rule ID | Condition   | Prediction             | Sorted By Rule Support (%) | Other Evaluation Statistics |                |      |                   |
|------|---------|---|------------------------|----------------------------|-----------------------------|----------------|------|-------------------|
|      |         |   |                        |                            | Condition Support (%)       | Confidence (%) | Lift | Deployability (%) |
| 1    | 5       | Within PropertyValue ≤ 120310   | CrimeType = MinorTheft | 10.07                      | 20.92                       | 48.13          | 1.20 | 10.85             |
| 2    | 3       | Within Households ≤ 2723<br>Within PropertyValue ≤ 120310                             | CrimeType = MinorTheft | 9.40                       | 19.45                       | 48.34          | 1.21 | 10.05             |
| 3    | 6       | Within Population ≤ 8576<br>Within PropertyValue ≤ 120310                             | CrimeType = MinorTheft | 9.02                       | 18.78                       | 48.02          | 1.20 | 9.76              |
| 4    | 4       | Within Population ≤ 8576<br>Within Households ≤ 2723<br>Within PropertyValue ≤ 120310 | CrimeType = MinorTheft | 8.55                       | 17.72                       | 48.23          | 1.21 | 9.17              |
| 5    | 16      | Within PropertyValue ≤ 120310   | CrimeType = Violent    | 4.28                       | 20.92                       | 20.46          | 2.05 | 16.64             |
| 6    | 17      | Within Households ≤ 2723<br>Within PropertyValue ≤ 120310                             | CrimeType = Violent    | 3.89                       | 19.45                       | 20.01          | 2.00 | 15.55             |
| 7    | 13      | Within Households ≥ 2723  | CrimeType = MajorTheft | 3.78                       | 15.38                       | 24.59          | 1.23 | 11.59             |
| 8    | 18      | Within Population ≤ 8576<br>Within PropertyValue ≤ 120310                             | CrimeType = Violent    | 3.69                       | 18.78                       | 19.65          | 1.97 | 15.09             |
| 9    | 10      | Within Households ≥ 2723<br>Within PropertyValue ≥ 120310                             | CrimeType = MajorTheft | 3.66                       | 13.90                       | 26.33          | 1.32 | 10.24             |
| 10   | 19      | Within Population ≤ 8576<br>Within Households ≤ 2723<br>Within PropertyValue ≤ 120310 | CrimeType = Violent    | 3.39                       | 17.72                       | 19.11          | 1.91 | 14.33             |

Abbildung 5: Regeltabelle und Maßzahlen

Nach erfolgter Analyse stehen dem Anwender je nach den im Vorfeld gewählten Optionen verschiedenste Ergebnis- und Regeltabellen, grafische Ausgaben und kartographische Ausgaben zur Verfügung.

**Die räumlich-temporale Vorhersage** analysiert im Zeitverlauf auftretende Muster und Ereignisse.

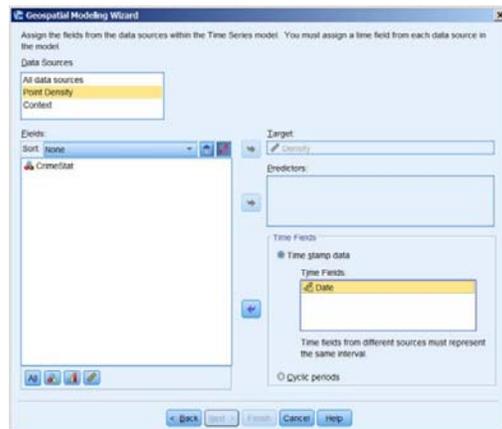


Abbildung 6: Dialogfenster zur Modellierung

Dabei werden ortsbezogene Daten, unabhängige Eingabedaten (Prädiktoren), eine oder mehrere Zeitfelder und eine Zielvariable (abhängige Variable) für die Modellierung herangezogen.

Sowohl IBM SPSS Statistics als auch das entsprechende Verfahren im IBM SPSS Modeler unterstützen den Anwender hier mit verschiedenen Assistenten und leicht verständlichen Dialogboxen, die je nach Anwendung den eigenen Ansprüchen und analytischen Erfordernissen angepaßt werden können.

Dies setzt im Vorfeld eine Definition von Karten und Datenquellen, sowie folgende Schritte im Assistenten voraus:

- Eingabe des Zielfeldes der Analyse, Zeitfelder und optionale Prädiktoren
- Definition der Zeitintervalle oder zyklischer Perioden für Zeitfelder.

Danach können die Ausgabeoptionen, die Aggregationseinstellungen und Modellerstellungsparameter angegeben werden. Die durch die Zeitreihenanalyse erhaltenen vorhergesagten Werte können später in einem SPSS Datenset in der aktuellen Sitzung oder in einer Datendatei im IBM® SPSS Statistics-Format abgelegt werden. Ebenso können die Karten- und Kontextdaten als Kartenspezifikation gespeichert werden. Dies kann auf folgende Art geschehen:

- Speichern der Kartenspezifikation in einer externen Datei (.mplan).
- Karte und Kontextdatendateien in die Spezifikation kopieren
- Daten aus Karten-Shape-Dateien, externen Datendateien und die in der Kartenspezifikation verwendeten Datasets werden in der Kartenspezifikationsdatei gespeichert.

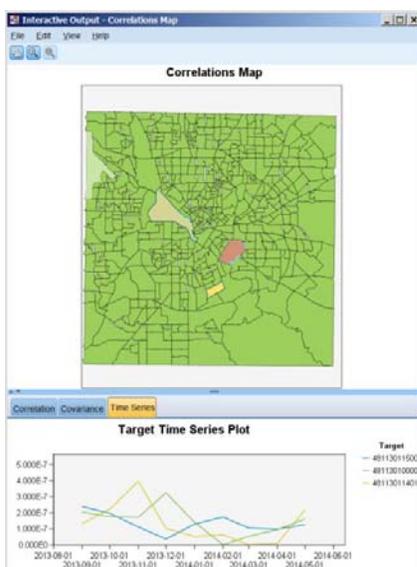


Abbildung 7: Raum-Zeit-Abbildung der Zielvariable

Die vorhergesagten Werte, inklusive Varianz und die obere und untere Konfidenzgrenze für das Zielfeld werden in einem geöffneten Dataset in der aktuellen Sitzung oder in einer Datendatei im IBM® SPSS Statistics-Format abgespeichert.

Das Analyseergebnis zeigt die untersuchten Merkmale im zeitlichen Verlauf. Zusammen mit den grafischen bzw. kartographischen Ausgaben lassen sich die räumlich-temporalen Zusammenhänge sehr gut visualisieren, so daß sich für den Anwender und auch z. B. für den Sachbearbeiter in der Fachabteilung ein klares Bild der untersuchten Verhältnisse im zeitlichen Ablauf ergibt.

Derzeit werden die Raum-Zeit-bezogenen Analysen und georäumlichen Modellierungen stark im polizeilichen Umfeld eingesetzt um Vorhersagen für Verbrechenschwerpunkte zu gewinnen.

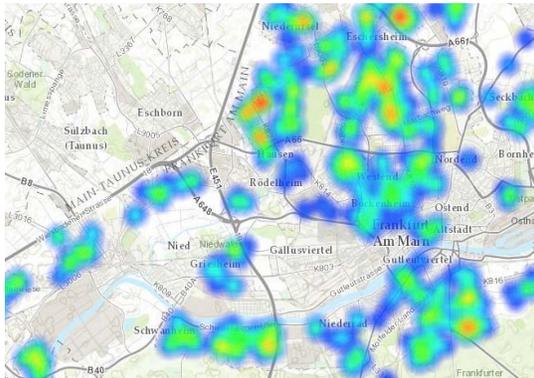
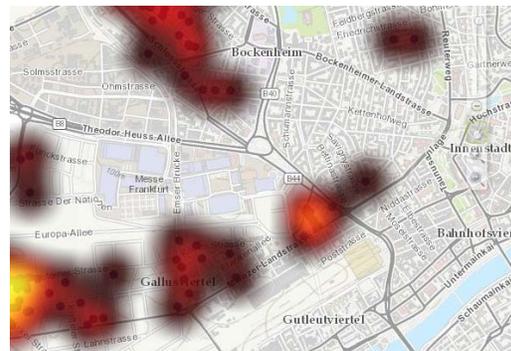


Abbildung 8: Geo-Heatmap

Damit werden dann Einsatzpläne, Streifen-dienstpläne, sowie der Personalbedarf gesteuert und im Sinne einer vorausschauenden Handlungsweise benutzt.

Damit konnten sowohl die Verbrechensraten um bis zu 30% gesenkt werden, als auch die Überstunden- und Nachtschichtzeiten beim bediensteten Personal verringert werden. Dieses wirkte sich wiederum günstig auf die Rate von Krankmeldungen und Kündigungen aus.

Gewinnbringende Einsatzszenarien finden sich u. a. im medizinischen Sektor, bei der Stau- und Verkehrsplanung, bei der Vorhersage von Migrationsverhalten, sowie im Tourismus, um nur einige weitere Möglichkeiten zu nennen.



## FAZIT

Mit der Funktion „Räumliche und temporale Modellierung“ (SPSS Statistics) bzw. dem Programmknoten „STP (Spatial-Temporal-Predictions)“ im SPSS Modeler bekommt der Fachanwender und Analysespezialist ein leistungsfähiges, einfach zu bedienendes und vielseitiges Werkzeug an die Hand mit dem raumbezogene Modellierungen durchgeführt werden können.

Erleichtert wird die Anwendung dieser Methoden noch durch die klar strukturierten Dialogfenster und Einrichtungsassistenten, sowie durch die gut verständliche Hilfefunktion im SPSS Programm. Hier kann man auch anhand eines geführten Lernprogrammes die Anwendung ausprobieren.

Alles in allem eine sehr sinnvolle Funktionalität mit der IBM erneut seine Marktposition als Branchenprimus im Bereich Predictive Analytics ausbauen kann.

Weitere Informationen zum Thema erhalten Sie unter [www.ibm.com/de/spss](http://www.ibm.com/de/spss) oder telefonisch unter 0049-89-4504 2022.

