

Sergej Plovs

BI-Datenvisualisierung im Zeitalter der DSGVO

Ideen veranschaulichen und Daten visualisieren – die Vielzahl an Tools im Bereich Business Intelligence (BI) macht die Entscheidung nicht leicht. Seit dem Inkrafttreten der Datenschutz-Grundverordnung (DSGVO) am 25.05.2018 müssen die Entscheider einen weiteren Aspekt berücksichtigen. Welche Hauptmerkmale sind für mein Unternehmen relevant und welche Lösung ist damit optimal für die individuellen Anforderungen? Ein Exkurs zum Thema Datenschutz in der Datenanalyse und Datenvisualisierung bringt Antworten. Ein Toolvergleich liefert eine detaillierte Analyse der speziellen Funktionen, Schlüsselparameter und Vor- und Nachteile von Microsoft Power BI, Tableau, SAP Lumira Discovery, SAP Lumira Designer, QlikView und Qlik Sense.

Business Intelligence und Datenschutz

Die DSGVO hat nicht nur Auswirkungen auf die Art, wie die Daten gesammelt werden, sondern auch darauf, wie sie verarbeitet und visualisiert werden. An dieser Stelle benötigt das Wort „Daten“ eine klare Definition und Abgrenzung. Die DSGVO regelt den Umgang mit personenbezogenen Daten. Dabei handelt es sich um Informationen, die eine individuelle Person direkt oder auch indirekt erkennbar oder identifizierbar machen. Dazu gehören Name, Geburtsdatum, E-Mail, Anschrift, mobile Telefonnummer oder auch Kontonummer. Ein Jahr vor dem DSGVO-Stichtag hat der Bundesgerichtshof entschieden, dass in Deutschland auch die dynamische IP-Adresse des Internetnutzers zu den personenbezogenen Daten zählt (Urteil vom 16.05.2017, VI ZR 135/13).

Dies bedeutet im Umkehrschluss: Solange ein Unternehmen keine personenbezogenen Daten der Kunden sammelt und verarbeitet, ist man auf der sicheren Seite. Aber sind Analysen ohne solche Daten überhaupt noch sinnvoll, aussagekräftig genug und bieten sie den erwünschten Mehrwert für vorgesehene Anwendungsszenarien? Die Antwort lautet wie in den meisten komplizierten Fällen: „ja, aber“. Wer bestimmte grundlegende Aspekte der Anonymisierung und Pseudonymisierung beachtet, ist auch bei der Erhebung und Visualisierung solcher Daten auf der sicheren Seite.

Aufbereitung der Daten

Die DSGVO ist erst seit dem 25.05.2018 in Kraft. Die Datenbasis für die Analysen beinhaltet jedoch Millionen oder sogar Milliarden von Datensätzen, die im DSGVO-freien Datenzeitalter entstanden sind. Diese Daten sind verlockend für viele denkbaren Arten von Analysen. Dabei gilt: Bevor man die vorliegenden Daten nicht DSGVO-konform aufbereitet hat, sollte man von solchen Analysen die Finger lassen.

Die DSGVO-Konformität eines BI-Tools bedeutet nicht, dass in den entsprechenden Datenquellen keine personenbezogenen Daten mehr vorhanden sind. Vielmehr ist da-

mit der Umgang mit solchen Daten gemeint. Das Löschen der Daten ist eine einfache Option, um solche Daten aus der Verarbeitung und Visualisierung herauszuhalten. Doch die gelöschten Daten bringen keinen Mehrwert für die Analysen, sondern erzeugen einen großen Nachteil. Daher empfehlen wir, die personenbezogenen Daten in BI-Tools und deren Datenquellen nicht zu löschen, sondern zu anonymisieren.

Bei der Verfremdung eines Datensatzes geht es in erster Linie darum, dass man die verbliebenen Informationen weiterhin für gewisse Analysezwecke nutzen kann. Dabei sollten nur Daten verfremdet werden, mit denen sich die Identität der Person eindeutig feststellen lässt. So müssen beispielsweise Postleitzahl und Ort nicht verändert werden, wenn Straße und Hausnummer bereits anonymisiert wurden. Denn ohne Straße und Hausnummer lässt sich keine eindeutige und personenbezogene Adresse feststellen. Auch bei den anderen Angaben ist die Vorgehensweise vergleichbar. Löscht man den Vornamen und den Nachnamen, so muss die Anrede nicht entfernt werden. Die Anrede kann stattdessen in das Geschlecht umgewandelt werden. Auch das Geburtsdatum wird mit einem einfachen Algorithmus vor dem Löschen in das Alter der Person umgewandelt und als Zusatzmerkmal im BI-System gespeichert, denn ohne weitere Angaben ist das Alter einer Person nicht personenbezogen.

Anhand der oben genannten Beispiele wird deutlich, dass man sehr viele wertvolle Informationen in Ergänzung zu einem vollständig verfremdeten personenbezogenen Datensatz weiterhin behalten und für die Analysen verwenden kann und darf. Stehen einem Data Analyst im BI-Tool solche Informationen wie PLZ, Ort, Geschlecht und Alter zu jedem Datensatz zur Verfügung, ist er weiterhin in der Lage, komplexe Geoanalysen bis auf PLZ-Ebene durchzuführen. Dabei lässt sich nicht nur zwischen männlichen und weiblichen Kunden unterscheiden, sondern auch deren Altersstrukturen in Betracht ziehen. Der Zugriff auf solche Informationen ist für die Visualisierung der Analysen durch einen Data

Scientist Gold wert. Grundsätzlich sind alle BI-Tools technisch in der Lage, die Verfremdung der personenbezogenen Daten vorzunehmen. Die notwendigen Algorithmen dazu liefern BI-Experten oder Berater innerhalb kurzer Zeit. Die verfremdeten Daten reichen in den meisten Anwendungsfällen vollkommen aus, denn das Ziel der Auswertungen ist häufig nicht, Informationen über Einzelpersonen zu gewinnen, sondern nur, allgemeine Zusammenhänge, Strukturen und Beziehungen zu erkennen, um richtige Entscheidungen zu treffen.

Alle hier analysierten BI-Produkte bieten umfangreiche Funktionen zur Einhaltung der DSGVO bei der Verarbeitung personenbezogener Daten in den angeschlossenen Datenquellen, sowie für die Verwaltung personenbezogener Daten im BI-Tool selbst. Der Datenschutz-Fokus liegt aber eindeutig auf den Datenquellen. Zu den am häufigsten angebotenen Datenquellen zählen neben den CRM- und ERP-Systemen auch große Data Lakes oder Big Data-Datenbanken. Die Verantwortlichen der BI-Tools sollten sich nicht darauf verlassen, dass ihre Kollegen bereits in den angeschlossenen Datenquellen die Konformität zur DSGVO vollständig berücksichtigt haben.

Es empfiehlt sich, das Thema Datenschutz im gesamten ETL-Prozess sorgfältig zu betrachten und bei der Extraktion und Transformation der Daten alle notwendigen Maßnahmen zu unternehmen. Werden dennoch personenbezogene Daten in das jeweilige BI-Tool geladen, so müssen spätestens bei der Datenvisualisierung alle relevanten Informationen in Bezug auf das einzelne Individuum unkenntlich gemacht werden.

BI-Tools im Vergleich

Tableau ist Marktführer für BI-Visualisierung. Tableau-Dashboards sind besonders flexibel und bieten eine Vielzahl an Funktionen zur individuellen Gestaltung. In der umfangreichen Bibliothek aus Visualisierungsarten findet sich der Anwender innerhalb kürzester Zeit zurecht. Sie ist leicht zu bedienen und spezielle Visualisierungseffekte wie Filter und Drill-Downs lassen sich schnell finden und anwenden. Auch werden die Vorgänge gut dokumentiert und gekennzeichnet und die Daten werden klar, attraktiv und interaktiv kommuniziert. Tableau bietet einen tiefen Einblick in die Daten und ermöglicht eine effiziente Veranschaulichung für komplexe Entscheidungsprozesse. Auch für die gemeinsame Nutzung, zum Beispiel innerhalb eines Teams bietet Tableau eine optimale Lösung.

QlikView ist eines der marktführenden BI-Produkte von Qlik. Es ist die flexibelste BI-Plattform für die Umwandlung von Daten in Wissen und wurde für die gemeinsame Nutzung von BI-Berichten und Datenvisualisierungen

Abb.: BI-Datenvisualisierung-Tools



entwickelt. Es ermöglicht Benutzern, Anwendungen mit eingebetteten Analysen zu erstellen. Voraussetzung hierfür sind jedoch Qlik-spezifische Entwicklungsfähigkeiten. Qlik Sense, ein weiteres Qlik-Produkt, wurde dagegen als moderne Datenexplorationsplattform wie Tableau konzipiert.

Qlik Sense ist flexibel, wenn es darum geht, Analysen in externe Anwendungen einzubetten. Es unterstützt allgemeine Webstandards wie HTML und verfügt über ein Responsive Design. So können Anwender Berichte und Visualisierungen erstellen, die für die Anzeige auf verschiedenen Geräten optimiert sind. Zielsetzung der Produktentwicklung bei Qlik ist es, Qlik Sense weiter zu aktualisieren.

Power BI von Microsoft ist eine webbasierte Schnittstelle mit einer Vielzahl an Funktionen. So reicht der Umfang von einer anpassbaren Visualisierung bis hin zu einer Kontrolle der Datenquellen. Die primären Tools von Power-BI enthalten Power Query zum extrahieren und transformieren, Power Pivot zum modellieren und analysieren und Power View und Power Map zur Visualisierung der Daten. Der Power BI Designer konsolidiert die einzelnen Tools in einer All-in-One-Anwendung und löst die Abhängigkeiten zu Excel oder Office. So wird PowerBI.com zum Austausch von Datensätzen, Berichten und Dashboards verwendet.

Besonderer Vorteile der Power BI-Technologie ist eine In-Memory-Analyse-Engine und eine spaltenbasierte Datenbank. Darüber hinaus bietet die Technologie eine Mischung aus Performance und Anwenderfreundlichkeit aufgrund der relativ einfachen DAX-Skriptsprache. Weitere Vorteile sind die monatlichen Aktualisierungen und eine kostenlose Version, die Kaufbarrieren reduziert. Klarer Nachteil ist dagegen die unnötige Komplexität beim Verlassen der Grundfunktionen, insbesondere für Benutzer mit eingeschränkter vorheriger Exposition gegenüber Excel. Es gibt außerdem nur wenige begrenzte Datenquellen, die Echtzeitverbindungen mit Power BI ermöglichen. Es existieren auch Leistungsprobleme beim Streaming oder Importieren großer Datenmengen. Zielsetzung der Lösung ist die Erweiterung und nicht das Ersetzen bestehender Berichterstellungstools. Der gesamte Austausch einer bestehenden Enterprise Data Warehouse Lösung sollte daher im Vorfeld geprüft werden.

SAP BusinessObjects Lumira ist ein Self-Service-Tool zur Datenvisualisierung für Geschäftsanwender. Das Tool kann mit oder ohne die In-Memory-Plattform von SAP HANA verwendet werden. SAP Lumira ist sowohl Self-Service BI-Erkennungstool als auch Visualisierungstool. Dieses Tool erweckt Daten durch die Vereinfachung der Erstellung und Freigabe von interaktiven Karten, Diagrammen und Infografiken zum Leben.

Mit **Lumira Discovery** können Daten aus einer Vielzahl von On-Premise- und Cloud-basierten Quellen importiert und visuelle BI-Analysen mit intuitiven Dashboards durchgeführt werden. Lumira Discovery ist ein umfangreicher Client für Business-Key-Anwender, die eine flexible Lösung benötigen. Mit Hilfe dieses Tools lassen sich Datenquellen verbinden sowie Daten erfassen, manipulieren und zusammenführen. Darüber hinaus bietet die Lösung eine Analyse von Offline- und Onlinedaten. Durch Visualisierungen können Erkenntnisse ad hoc von allen zu erstellenden Arten gewonnen werden.

Die **SAP Lumira Designer-Edition** ermöglicht interaktive, für Mobilgeräte optimierte Visualisierungen für Business-Dashboards. So können die gewonnenen Infor-

mationen den Entscheidungsträgern präsentiert und die wichtigsten Erkenntnisse aus den Daten gewonnen werden. Lumira Designer ist für professionelle Anwender (IT-Abteilungen), um Berichte für Unternehmensanalysen zu erstellen. Es bietet die gleichen UI-Elemente wie Lumira Discovery und ermöglicht die individuelle Anpassung des Look-and-Feel und der Benutzerinteraktion durch Skripting und CSS-Stylesheets. Darüber hinaus können gemeinsam genutzte UI-Elemente erstellt werden. Die Reduktion von komplexen Anwendungen in kleine Teile ermöglicht eine die Verwaltung von umfangreichen BI-Deployments.

Fazit: Wer ist der Gewinner?

Der Teufel steckt im Detail: Eine perfekte Lösung für alle Anforderungen gibt es nicht. Die Entscheidung für das richtige Tool hängt von verschiedenen Faktoren ab. So ist es wichtig, welche Software bereits besteht, da die Kompatibilität besondere Vorteile schafft. So lässt sich Power-BI am besten in Microsoft-Produkte wie Excel oder SharePoint integrieren. SAP Lumira empfiehlt dagegen eine Verbindung zum SAP-Backend (HANA-Views oder BW). QlikView/Qlik Sense und Tableau sind dagegen herstellerunabhängige Tools. Tableau bietet darüber hinaus eine Vielzahl an Funktionen, liefert das beste UX und stellt eine Verbindung zu vielen Datenquellen her.

Aufgrund der regelmäßigen Updates und neuen Produktentwicklungen ist ein Toolvergleich aber immer nur eine Momentaufnahme. So befinden sich die Produkte in verschiedenen Phasen des Lebenszyklus, und was heute zutrifft, könnte in 12 Monaten schon nicht mehr aktuell sein.

Stichwort: BI, Business Intelligence

Autor: Sergej Plovs ist Senior Consultant und Geschäftsführer der Vision11 GmbH mit mittlerweile elf Jahren CRM-Projekterfahrung in den unterschiedlichsten Branchen. Das CRM-Beratungshaus Vision11 wurde 2011 gegründet, um Unternehmen unterschiedlicher Branchen bei der digitalen Transformation ihrer Kundenprozesse umfassend zu unterstützen. Als CRM- und Analytics Spezialist vereint Vision11 strategisches Denken, konzeptionelle Methodik sowie technologisches Wissen unter einem Dach. Zu den Hauptkunden gehören unter anderem starke Marken AUDI, MAN, TRUMPF und GROHE. Kernaufgabe ist die Schaffung von positiven digitalen Erlebnissen, um Kunden langfristig an Marken zu binden, www.visioneleven.com.

Kontakt:
redaktion@gliss-kramer.de

