

Otto-Friedrich-Universität Bamberg
Lehrstuhl für Statistik und Ökonometrie
Standort: Bamberg
Sommersemester 2022
Vorlesung: Blockseminar Survey-Methodik
Prüfer: Dr. Sara Bleniger

Statistische Geheimhaltung: Cell Key Methode

Joshua Simon
joshua-guenter.simon@stud.uni-bamberg.de
Master Survey Statistik, 4. Fachsemester
Matrikelnummer: 2032411
15. Juni 2022

Inhaltsverzeichnis

1. Einführung	4
2. Etablierte Geheimhaltungsverfahren	4
2.1. Überblick der Statistiken	4
2.2. Datenfluss	5
3. Cell Key Methode	5
3.1. Methodik	5
3.2. Besonderheiten der Cell Key Methode	6
4. Ergebnisse und Auswertungen	6
5. Zusammenfassung und Fazit	6
Literatur	7

Abbildungsverzeichnis

1. Ablaufdiagramm der Cell Key Methode 5

Tabellenverzeichnis

Listings

1. Einführung

Die amtliche Statistik sorgt mit einer Vielzahl an Veröffentlichungen für die Bereitstellung von aufbereiteten statistischen Informationen. Damit geht sie dem Ziel nach, Bürgern, Institutionen und anderen gesellschaftlichen Einrichtungen eine Datengrundlage für die Entscheidungsfindung zu bieten. Weiter dient die amtliche Statistik auch der Politik und der Wissenschaft als Datenquelle. Das Sammeln und Erheben dieser Daten stellt in vielen Fällen einen Eingriff auf das Recht der informationellen Selbstbestimmung für Personen und Entitäten dar. Dieses Recht ist das Fundament des modernen Datenschutzes und wird über Artikel 2 des Grundgesetzes abgedeckt. Es steht außer Frage, dass dieses Recht besonders schützenswert ist. Demnach steht auch die amtliche Statistik in der Pflicht dieser Verantwortung nachzukommen. Konkret manifestiert sich das Einhalten dieser Pflicht in dem sog. Statistikgeheimnis. Aus dem Bundesstatistikgesetz lässt sich hierzu der folgende Absatz aufgreifen:

Einzelangaben über persönliche und sachliche Verhältnisse, die für eine Bundesstatistik gemacht werden, sind von den Amtsträgern und für den öffentlichen Dienst besonders Verpflichteten, die mit der Durchführung von Bundesstatistiken betraut sind, geheim zu halten, soweit durch besondere Rechtsvorschrift nichts anderes bestimmt ist. (§ 16 Abs. 1 Satz 1 BStatG).

Konkret möchte man mit der statistischen Geheimhaltung die folgenden Punkte bedienen (nach Begründung zum BStatG; BT-Drucks. Nr. 10/5345 vom 17. April 1986).

Im weiteren Verlauf dieser Arbeit sollen Geheimhaltungsverfahren und Geheimhaltungsregeln präsentiert werden, die im Einzelnen die statistische Geheimhaltung gewährleisten und damit dem Statistikgeheimnis der amtlichen Statistik nachkommen. Besondere Beachtung wird dabei der Cell Key Methode (CKM) geschenkt. Dieses Verfahren bietet ein Ansatz, welcher gegenüber anderen Verfahren gut zu Implementieren und zu Automatisieren ist. Gerade dieser Punkt ist in einem immer weiter werdenden technologischen Umfeld nicht außer Acht zu lassen.

2. Etablierte Geheimhaltungsverfahren

2.1. Überblick der Statistiken

Für die oben genannten Statistiken werden jeweils in unterschiedlichen Abständen Daten von den Meldern erhoben. Dabei besitzen die meisten dieser Statistiken einen eigenen Datensatz, der mit seinen individuellen Merkmalen auf den jeweiligen Zweck

der Statistik abgestimmt ist.

2.2. Datenfluss

In diesem Abschnitt soll der Datenfluss innerhalb des Bayerischen Landesamts für Statistik skizziert werden. Ein vollständiger und reibungsloser Datenfluss ist das höchste Ziel im Data Engineering. Einige übergeordnete Konzepte und Definitionen zu dem Bereich Data Engineering finden sich auch in [Cri20].

3. Cell Key Methode

3.1. Methodik

Dieses Kapitel des Berichts beschäftigt sich mit der technischen Umsetzung des in Kapitel 2 beschriebenen Datenflusses. Für die technische Unterstützung kommt hierbei eine eigens im LfStat entwickelte Softwarelösung zum Einsatz. Die Software soll dabei die folgenden Ziele erfüllen:

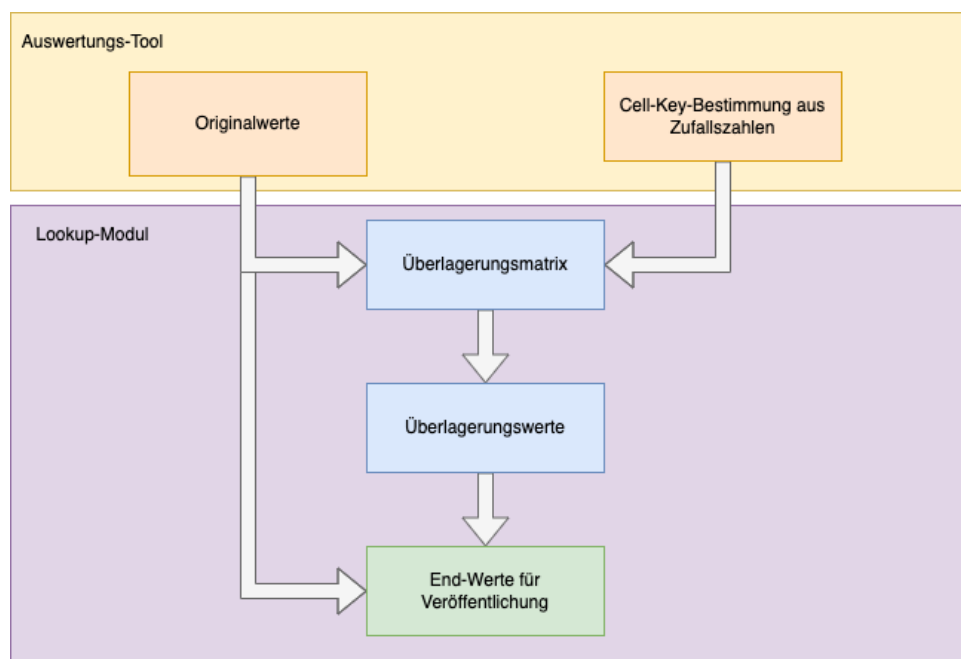


Abbildung 1: Ablaufdiagramm der Cell Key Methode

Die Kernaufgaben der Softwarelösung umfassen also verschiedenen File-Prozesse, Ladeprozesse in Datenbanken und Auswertungssysteme, Datenanalysen und Datenauswertungen sowie Verwaltungsaufgaben, beispielsweise durch das Logging von Arbeitsschritten in einer Datenbank. Allgemeine und weiterführende Konzepte werden

auch in Hinblick auf eine technische Umsetzung in [Cri20] behandelt.

3.2. Besonderheiten der Cell Key Methode

4. Ergebnisse und Auswertungen

In diesem Kapitel sollen Ergebnisse über die Datenqualität zusammengetragen werden. Die Analyse solcher (Meta-)Daten ist zentral für das Data Engineering, da hiermit das implementierte System validiert und eine mögliche Verbesserung der Datenqualität durch verschiedene Aufbereitungsprozesse gemessen werden kann. Im Anschluss folgt ein kurzer Überblick über die erhobenen Daten selbst.

5. Zusammenfassung und Fazit

Damit das Bayerische Landesamt für Statistik seiner gesetzlich vorgeschriebenen Pflichten nachkommen kann, sind die unterschiedlichsten Technologien und Verfahren notwendig. Neben dem Erheben der Daten bei den Meldern, gehören auch die Plausibilisierung und die Aufbereitung der Daten zu seinen Aufgaben. Um eine so große Datenmenge effizient zu verarbeiten, wird auf moderne Lösungen aus der Informationstechnologie zurückgegriffen. Hierzu zählen verschiedene Skriptsprachen, Datenbanken und Datawarehouses. Um die Qualität der Daten zu wahren, sind neben rein technischer Kontrollen auch sehr fachliche Zusammenhänge zu prüfen und ggf. zu korrigieren. Dies erfordert ein tiefes inhaltliches Verständnis der Daten und ihrer Merkmale. Eine gewisse Interpretation und Analyse sind also auch im Data Engineering notwendig, um die Datenqualität und den Datenfluss zu erhalten, sowie die Daten zu publizieren und damit die Pflicht der amtlichen Statistik zu erfüllen.

Literatur

- Bayerisches Landesamt für Statistik (o.J.): *Hochschulen*, online verfügbar unter: https://www.statistik.bayern.de/statistik/bildung_soziales/hochschulen/index.html#link_1 (Zugriff am 29.12.2020).
- Crickard, Paul (2020): *Data Engineering with Python*. Birmingham: Packt Publishing.
- Kimball, Ralph and Ross, Margy (2013): *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Indianapolis: Wiley.
- Statistisches Bundesamt (o.J.): *Gesetz über die Statistik für das Hochschulwesen (Hochschulstatistikgesetz HStatG)*, online verfügbar unter: https://www.destatis.de/DE/Methoden/Rechtsgrundlagen/Statistikbereiche/Inhalte/505_HStatG.pdf?__blob=publicationFile (Zugriff am 29.12.2020).
- Wikipedia (2019): *T-SQL - Wikipedia, Die freie Enzyklopädie*, online verfügbar unter: <https://de.wikipedia.org/wiki/Transact-SQL> (Zugriff am 10.01.2021).
- Wikipedia (2021): *Datawarehouse - Wikipedia, Die freie Enzyklopädie*, online verfügbar unter: https://de.wikipedia.org/wiki/Data_Warehouse (Zugriff am 10.01.2021).
- Wikipedia (2021): *ETL-Prozess - Wikipedia, Die freie Enzyklopädie*, online verfügbar unter: <https://de.wikipedia.org/wiki/ETL-Prozess> (Zugriff am 10.01.2021).

Ich erkläre hiermit, dass ich die Seminararbeit mit dem Titel *Statistische Geheimhaltung: Cell Key Methode* im *Sommersemester 2022* selbständig angefertigt, keine anderen Hilfsmittel als die im Literaturverzeichnis genannten benutzt und alle aus den Quellen und der Literatur wörtlich oder sinngemäSS übernommenen Stellen als solche gekennzeichnet habe.

Bamberg, den 15. Juni 2022

Unterschrift