**Deutsch: Beschreibung der Datensätze**

Bereitgestellt werden Daten aus dem Handelsregister ([https://www.handelsregister.de](https://www.handelsregister.de/)) seit 2007. Speziell für das DataFest sind nur Unternehmen mit Sitz in Bayern oder Baden-Württemberg inklusive abgeleiteter Informationen enthalten.Das gemeinsame Registerportal der Länder wurde zum 01.08.2022 umgestellt. Seitdem wird der öffentlich verfügbare chronologische Abdruck verwendet. Die Umstellung führt dazu, dass es in der Datenbank von infas aber auch noch Datensätze gibt, die auf Volltexten basieren, welche vor dem 01.08.2022 verwendet wurden.

Der Datensatz besteht aus 9 Dateien:

1. Firmen\_hist.csv: Enthält alle Merkmale aus den Handelsregistereinträgen von Firmen (aktiv, inaktiv, aufgelöst). Enthält 2.689.784 Einträge und 39 Spalten. Die Einträge reichen vom 1. Januar 2007 bis zum 27. März 2023.
2. Firmen\_inaktiv.csv: Beinhaltet alle aktuellen Merkmale gelöschten Unternehmen zum Zeitpunkt der Lieferung. Dies bedeutet, dass diese Firmen aus dem Handelsregister gelöscht wurden. Enthält 289.976 Einträge und 36 Spalten.
3. Firmen.csv: Enthält alle aktuellen Merkmale aktiver/ aufgelösten Unternehmen zum Zeitpunkt der Lieferung. Aufgelöst ist eine Firma, wenn sie sich in Liquidation befindet. Dies kann bspw. ein Insolvenzverfahren sein. Die Liquidation, also die Auflösung einer Firma, ist in der Regel die Grundlage für die darauffolgende Löschung der entsprechenden Firma. Besteht aus 568.811 Einträgen und 36 Spalten.
4. Personen.csv: Liefert alle aktuellen Merkmale von Personen. Beinhaltet 979.103 Einträge und 11 Spalten.
5. Personen\_hist.csv: Enthält alle Merkmale aus den Handelsregistereinträgen von Personen. Enthält 2.813.767 Einträge und 15 Spalten.
6. Relationen\_person\_firma.csv: Beinhaltet alle aktiven Verbindungen zwischen Personen und Firmen. Enthält 717.146 Einträge und 9 Spalten.
7. Relationen\_person\_firma\_hist.csv: Enthält alle früheren Verbindungen zwischen Personen und Unternehmen. Besteht aus 1.137.360 Einträgen und 9 Spalten.
8. Relationen\_firma\_firma.csv: Enthält alle aktiven Verbindungen zwischen zwei Unternehmen. Beinhaltet 76.051 Einträge und 10 Spalten.
9. Relationen\_firma\_firma\_hist.csv: Enthält alle früheren Verbindungen zwischen zwei Unternehmen. Enthält 105.464 Einträge und 10 Spalten.

**English. Description of the datasets**

The provided dataset contains comprehensive information from the commercial register ([https://www.handelsregister.de](https://www.handelsregister.de/)) dating back to 2007. For DataFest, the dataset is limited to companies with their registered offices in Bavaria or Baden-Württemberg and includes derived information. The joint register portal of the federal states underwent a conversion on August 1, 2022. Since then, the publicly available chronological imprint has been used. However, due to the changeover, some data records in the infas database still rely on full texts utilized before August 1, 2022.

The dataset is comprised of 9 files:

1. Firmen\_hist.csv: Contains all characteristics from the commercial register entries of companies (active, inactive, liquidated). Includes 2,689,784 entries and 39 columns. Entries range from January 1, 2007, to March 27, 2023.
2. Firmen\_inaktiv.csv: includes only companies with the status 'inactive'. This indicates that these companies have been removed from the commercial register. Contains 289,976 entries and 36 columns.
3. Firmen.csv: Features all up-to-date characteristics of active/dissolved companies at the time of delivery. A company is considered dissolved if it is undergoing liquidation, which could be due to reasons such as insolvency proceedings. Typically, the liquidation process, or the dissolution of a company, serves as the foundation for the eventual removal of the associated company from the records. Consists of 568,811 entries and 36 columns.
4. Personen.csv: Provides all up-to-date characteristics of individuals. Includes 979,103 entries and 11 columns.
5. Personen\_hist.csv: Contains all characteristics from the commercial register entries of individuals. Features 2,813,767 entries and 15 columns.
6. Relationen\_person\_firma.csv: Includes all active connections between individuals and companies. Contains 717,146 entries and 9 columns.
7. Relationen\_person\_firma\_hist.csv: Features all former connections between individuals and companies. Consists of 1,137,360 entries and 9 columns.
8. Relationen\_firma\_firma.csv: Contains all active connections between two companies. Includes 76,051 entries and 10 columns.
9. Relationen\_firma\_firma\_hist.csv: Provides all former connections between two companies. Features 105,464 entries and 10 columns.