

## Datensatzbeschreibung

# Felduntersuchung zu Behavioral Energy Efficiency Potentialen von privaten Haushalten

## Inhalt

1	Beschreibung des Datensatzes .....	1
2	Gesamtverbrauchsdaten .....	1
3	Eventbasierte Geräteverbräuche .....	2
4	Variablen .....	3
5	Kontaktinformationen .....	4

## 1 Beschreibung des Datensatzes

Im Rahmen der Veröffentlichung des Working Papers “Felduntersuchung zu Behavioral Energy Efficiency Potentialen von privaten Haushalten” wird ein Beispieldatensatz veröffentlicht. Ziel dieses Working Papers war es, zu untersuchen inwiefern sich die Zusendung eines monatlichen Stromreports auf das Stromverbrauchsverhalten privater Haushalte in Deutschland auswirkt und ob es vor dem Hintergrund Behavioral Energy Efficiency zu Einsparungen kommt.

Dieser begleitende Datensatz umfasst Daten zum Stromverbrauch von 200 Haushalten im Zeitraum vom 01.01.2019 bis zum 31.01.2020. Er unterscheidet sich prinzipiell von den im Working Paper verwendeten Daten. Die Abweichung ist zum einen durch explizites Einverständnis von den Teilnehmern zur Veröffentlichung ihrer Daten begründet. Zum anderen wurden Haushalte gewählt, deren Verbrauchsdaten über den gesamten Zeitraum vollständig sind.

Der Datensatz ist in drei einzelne Dateien unterteilt, deren Aufbau und Variablen in den folgenden Kapiteln beschrieben werden. Die Haushalte sind durch eine durchgehende ID gekennzeichnet, was eine eindeutige Zuordnung eines individuellen Haushalts innerhalb der drei Dateien ermöglicht.

## 2 Gesamtverbrauchsdaten

Diese Datei beinhaltet den jeweiligen Stromverbrauch der 200 Haushalte. Das Zeitintervall der Erfassung beträgt 15 Minuten. Die ersten zwei Spalten beinhalten den Zeitstempel und die Einheit. Jede weitere Spalte enthält den Stromverbrauch als viertelstündige Zeitreihe mit der jeweiligen Haushalts ID als Spaltenkopf.

Die Messdaten wurden durch Smart Meter erhoben, welche ein Datenintervall von 0,5 Hz aufweisen (ein Datenpunkt alle zwei Sekunden). Diese wurden auf 15 minütige Daten aggregiert.

Variablen im Überblick:

- [rec\_time]
  - o Zeitstempel der Erfassung des Verbrauchs
  - o Format 1: YYYY-MM-DD hh:mm:ss+01:00 UTC (Winterzeit)
  - o Format 2: YYYY-MM-DD hh:mm:ss+02:00 UTC (Sommerzeit)
- [Einheit]
  - o Einheit des gemessenen Verbrauchs
  - o Standardmäßig in kWh
- [1, 2, 3,...,200]
  - o IDs der Haushalte (ID 1 entspricht Haushalt 1)
  - o Jede Spalte enthält die Verbrauchsdaten entsprechend dem Zeitstempel
  - o Die Verbrauchsdaten wurden in der auf den Zeitstempel folgenden 15 Minuten realisiert

### 3 Eventbasierte Geräteverbräuche

Diese Datei enthält disaggregierte Verbrauchsdaten je Haushalt. Für die Disaggregation wurde die Gesamlastkurve auf einzelne Stromverbraucher mithilfe von Algorithmen heruntergebrochen. Die disaggregierten Verbrauchsdaten weisen eine eventbasierte zeitliche Auflösung auf. Eine eventbasierte Darstellung zeigt, wann ein Event (beispielsweise das Anschalten einer Spülmaschine) registriert wurde und wie viel Stromverbrauch diesem Event zugeschrieben wird.

Im Datensatz bezeichnet jeder Zeileneintrag ein durch die Algorithmen erkanntes Event. Die Spaltenköpfe enthalten die Variablen ID, Gerät, Startzeit, Stoppzeit, Verbrauch und Einheit (Erläuterung s.u.). Bei den Geräteklassen StandBy, Kühlschrank und Beleuchtung ist zu beachten, dass diese Verbräuche über den Tag summiert und als ein Event pro Tag ausgegeben werden.

Es ist zu beachten, dass die Algorithmen mittels maschinellem Lernen eine bestmögliche Abschätzung ("Best-Guess") treffen. Die Genauigkeiten variieren dabei zwischen den einzelnen Geräteklassen und hängen insbesondere auch von der Komplexität der Geräteausstattung der Haushalte ab. Die verwendeten Algorithmen weisen im Durchschnitt dabei eine Genauigkeit von etwa 80% auf und sind auf bestimmte Geräteklassen beschränkt.

Die Differenz zwischen den Gesamtverbrauchsdaten und den eventbasierten Geräteverbräuchen ist der Verbrauch, der durch die Algorithmen nicht zugeordnet werden konnte. Der Anteil des erkannten Verbrauchs am Gesamtverbrauch ist allerdings kein Indikator für die Qualität der Erkennung durch die Algorithmen. Im Einzelfall kann der erkannte Verbrauch prozentual einen geringen Anteil am Gesamtverbrauch haben, aber eine gute Erkennungsqualität aufweisen (beispielsweise bei Haushalten mit einem sehr hohen Verbrauch).

Variablen im Überblick:

- [ID]
  - o zugeordnete ID des Haushaltes
  - o Mögliche Werte: [1, 2, 3..., 200]
- [Gerät]
  - o zugeordnete Gerätekategorie des Events
  - o Mögliche Werte:
    - [AlwaysOn] – der erkannte Standby-Verbrauch
    - [Refrigeration] – der erkannte Verbrauch von Kühlgeräten; einzelne Geräte werden summiert

- [Lighting] – der erkannte Verbrauch von Beleuchtung
- [Dryer] – der erkannte Verbrauch von Wäschetrocknern
- [FlowWaterHeater] – der erkannte Verbrauch von Durchlauferhitzern
- [WashingMachine] – der erkannte Verbrauch von Waschmaschinen
- [Dishwasher] – der erkannte Verbrauch von Spülmaschinen
- [Oven] – der erkannte Verbrauch von Elektro-Herden (beinhaltet die Herdplatte und den Backofen)
- [Startzeit]
  - Registrierte Startzeit des Events
  - Format 1: [YYYY-MM-DD hh:mm:ss+01:00 UTC] - Winterzeit
  - Format 2: [YYYY-MM-DD hh:mm:ss+02:00 UTC] – Sommerzeit
- [Stoppzeit]
  - Registrierte Stoppzeit des Events
  - Format 1: [YYYY-MM-DD hh:mm:ss+01:00 UTC] - Winterzeit
  - Format 2: [YYYY-MM-DD hh:mm:ss+02:00 UTC] - Sommerzeit
- [Verbrauch]
  - Erkannter Stromverbrauch des Events
- [Einheit]
  - Einheit der Messdaten des Stromverbrauchs
  - Standardmäßig in [kWh]

## 4 Variablen

Die Datei “Variablen” liefert weitere soziodemografische Informationen zu den einzelnen Haushalten.

Variablen im Überblick:

- [ID]
  - zugeordnete ID des Haushaltes
  - Mögliche Werte: [1, 2, 3..., 200]
- [Gruppe]
  - Zuordnung des Datensatzes zur Test- oder Kontrollgruppe in Bezug auf das Experiment des Working Papers
  - Mögliche Werte:
    - [0] – Haushalt hat nicht am Experiment teilgenommen
    - [1] – Haushalt hat als Testkunde im Experiment teilgenommen
- [Engagement]
  - Einordnung des Engagements des Testkunden innerhalb des Experiments
  - Bei einer Öffnungsrate der E-Mails mit den Reports von 100% wurde ein hohes Engagement angenommen
  - Mögliche Werte:
    - [0] – Haushalt ist Teil der Kontrollgruppe, keine Indikation des Engagements
    - [1] – Hohes Engagement, Teilnehmer hat alle E-Mails mit den Reports geöffnet
    - [2] – Niedriges Engagement, Teilnehmer hat nicht alle E-Mails mit den Reports geöffnet

- [Bundesland]
  - o Regionale Einordnung des Haushaltes
  - o Mögliche Werte: [Baden-Württemberg, Bayern, Berlin, Brandenburg, Bremen, Hamburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Sachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen]
- [Altersgruppen]
  - o Zuordnung des Alters in Altersgruppen
  - o Indikation der Zugehörigkeit durch binäre Variablen
    - [0] – Alter des Kunden liegt nicht in der Altersgruppe
    - [1] – Alter des Kunden liegt in der Altersgruppe
  - o Verwendete Altersgruppen (in Anlehnung an Destatis)
    - [Alter bis unter 20]
    - [Alter von 20 bis unter 40]
    - [Alter von 40 bis unter 60]
    - [Alter von 60 bis unter 80]
    - [Alter 80 und mehr]
- [Geschlecht]
  - o Angegebenes Geschlecht des Kunden (Anmerkung: keine Aussage über weitere Bewohner des Haushalts möglich)
  - o Mögliche Werte<sup>1</sup>:
    - [w] – weiblich
    - [m] – männlich

## 5 Kontaktinformationen

Bei Rückfragen zum Datensatz stehen Ihnen folgende Kontaktpersonen zur Verfügung:

Adrian Beyertt, Business Development Manager, Fresh Energy GmbH  
[adrian.beyertt@getfresh.energy](mailto:adrian.beyertt@getfresh.energy)

Stephan Seim, Wissenschaftlicher Mitarbeiter, Technische Universität Berlin  
[stephan.seim@tu-berlin.de](mailto:stephan.seim@tu-berlin.de)

---

<sup>1</sup> Die Berücksichtigung von diversen Geschlechtern ist berücksichtigt worden. Es gab dazu keine Angaben, daher wird in den Werten keine separate Gruppe ausgewiesen