

Methodisches Vorgehen

Methode des regelbasierten Ansatzes zur Selektion der Personas

Der regelbasierte Analyseansatz umfasst drei methodische Schritte. Ausgangspunkt ist ein Datensatz von infas 360 mit rund 19 Millionen adressgenauen Haushaltsdatensätzen, die insgesamt ca. 42 Millionen Haushalte in Deutschland abbilden. In einem zweiten Schritt wurden auf diesen Adressen fünf zentrale Merkmale betrachtet, die sich in Analyse als ausschlaggebend für klimabezogene Handlungsfähigkeit erwiesen haben: der Energieträger der Heizung (z. B. Wärmepumpe, Fernwärme, Gas, Öl), die Eigentümerstruktur (Eigentümer, Mieter, institutionelle oder private Vermietung), das Nettoäquivalenzeinkommen, das Baujahr des Gebäudes sowie die Gebäudeform (z. B. Einfamilienhaus, Reihenhaushaus, Mehrfamilienhaus).

Anhand dieser Merkmale erfolgt im dritten Schritt die regelbasierte, schrittweise Zuordnung der Haushalte zu einem von 16 definierten Haushaltstypen (Personas). Diese Zuordnung folgt einem Regelwerk, das auf den fünf erwähnten inhaltlich begründeten Kriterien basiert.

Das erste Kriterium ist der Energieträger, um zu differenzieren, ob die Haushalt bereits klimaneutral mit Wärme versorgt werden bzw. versorgt werden können. Das zweite ist die Handlungsebene. Sie beschreibt, in welchem Maß ein Haushalt eigenständig über klimarelevante Investitionen entscheiden kann. In der Modelllogik spiegelt sich dies in der Differenzierung nach Eigentumsverhältnissen (Mieter/Eigentümer) und Gebäudeformen (Einfamilienhaus, Reihenhaushaus, Mehrfamilienhaus) wider. Das dritte Kriterium betrifft die verfügbaren Ressourcen und institutionellen Strukturen. Dieses Kriterium wird im Regelwerk über die Eigentümerstruktur operationalisiert. Das vierte Kriterium ist der Investitionsbedarf. Das Baujahr der Gebäude wird als zentrales Kriterium in der Zuordnung berücksichtigt. Als fünftes Kriterium fließt das Nettoäquivalenzeinkommen in die Analyse ein. Es erlaubt Rückschlüsse auf die wirtschaftliche Handlungsfähigkeit eines Haushalts. Das regelbasierte System nutzt hierzu abgestufte Einkommensklassen, die in die Entscheidungskette zur Persona-Zuordnung integriert sind.

Die konkrete Anwendung dieser Regeln erfolgt schrittweise anhand der Merkmalskombinationen. Zunächst wird nach Energieträgern unterschieden. So werden beispielsweise Haushalte mit Wärmepumpe als eine Gruppe definiert, da sie bereits weitgehend klimaneutral sind. Daraufhin werden die weiteren Kriterien nicht mehr angewandt. Dasselbe gilt für Pellet-Heizungen oder Haushalte mit Nachtspeicher. Weiter ausdifferenziert werden die Haushalte mit Gas- und Ölheizungen. Bei ihnen kommen die weiteren Merkmale zum Tragen, die dann jeweils eine oder mehrere neue Untergruppen ergeben.

Durch diese systematische Kombination adressgenauer Merkmale, gestützt durch eine empirisch valide Datengrundlage, entsteht ein konsistentes und differenziertes Bild der Ausgangslage deutscher Haushalte im Hinblick auf die Wärmewende und klimafreundliche Mobilität. Der regelbasierte Ansatz ermöglicht eine datenbasierte Einordnung von Belastungs- und Handlungspotenzialen – und bietet eine fundierte Grundlage für sozial gerechte und zielgerichtete klimapolitische Maßnahmen.

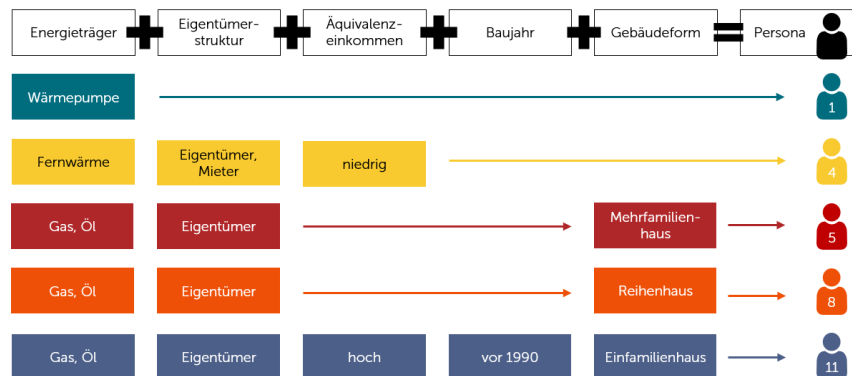
So lässt sich das Verfahren zusammenfassen:

Regelbasierter Ansatz

1. Selektion der Grundgesamtheit:
Alle Adressen mit Haushalten
(N = 19 Mio. Adressen mit
42 Mio. Haushalten)

2. Anreicherung mit den relevanten, adressgenau vorliegenden Merkmalen:
Energieträger, Eigentümerstruktur, Netto-Äquivalenzeinkommen, Baujahr des Gebäudes, Gebäudeform

3. Schrittweise Zuordnung der Adressen anhand der Ausprägungen der Merkmale (Selektionskriterien) zu den vordefinierten Persona



Erläuterungen zum Datensatz von infas 360

Die infas360 GmbH ist auf die systematische Nutzung öffentlicher und privater Datenquellen zur Beantwortung wirtschaftlicher und wissenschaftlicher Fragestellungen spezialisiert. Ein besonderer Schwerpunkt liegt auf der Ermittlung kleinräumiger Kennzahlen mithilfe moderner statistischer Verfahren.

Die Datengrundlage besteht aus großen, miteinander verknüpften Datenbanken. Zentral ist eine Gebäudedatenbank, die alle Adressen in Deutschland abdeckt und für jedes Gebäude Informationen zu Typ, Höhe, Volumen, Baujahr, Nutzung sowie Energieversorgung, Internetanbindung und Garten- oder Garagenausstattung bereitstellt. Jeder Datenpunkt ist somit einem konkreten Gebäude zugeordnet. Auf diese Weise wird die gesamte Grundgesamtheit der Wohn- und Nichtwohngebäude in Deutschland abgebildet, wobei Analysen auf Personenebene – etwa Persona-Analysen – sich ausschließlich auf Wohngebäude beziehen.

Ergänzend zu den Gebäudedaten stehen mikrogeografische Informationen zu Haushalten und Einwohnern zur Verfügung, darunter Angaben zu Haushaltstypen, Einkommensverteilungen und Altersstrukturen auf adressgenauer Ebene. Zusätzlich werden georeferenzierte Objekte wie Straßen, E-Ladesäulen oder Haltestellen des öffentlichen Nahverkehrs in einer Points-of-Interest-Datenbank erfasst.

Die Daten stammen aus unterschiedlichen Quellen. Sie umfassen eine Mischung aus amtlichen Registerdaten, direkt erhobenen Echtfalldaten sowie Schätzungen. Wo amtliche Registerdaten fehlen, werden verfügbare Datenquellen durch eigene Befragungen von infas 360 ergänzt und mithilfe statistischer Verfahren auf die Grundgesamtheit hochgerechnet.

Am Beispiel der Einkommensdaten lässt sich das Vorgehen wie folgt illustrieren: Mehrere tausend Personen werden zu ihrem Einkommen und ihrer Adresse befragt. Auf Basis dieser Stichprobe werden Merkmale identifiziert, die das Einkommen statistisch signifikant erklären. Mit diesen Modellen werden die bekannten Einkommensdaten der Befragten auf die Gesamtheit der rund 40 Millionen Haushalte in Deutschland übertragen. Parallel erfolgt eine Validierung anhand weiterer externer Datensätze, um die Schätzgenauigkeit zu überprüfen.

Auf dieser Datengrundlage können verschiedene Analysen durchgeführt werden, die stets auf die Grundgesamtheit aller Haushalte bezogen sind. Dazu zählen Marktpotenzialanalysen, Kundensegmentierungen, Standortbewertungen und Vertriebsgebietsanalysen. Dabei kommen sowohl klassische statistische Methoden als auch spezialisierte Verfahren der Small Area

Estimation zum Einsatz, um auch für kleinste geographische Einheiten verlässliche Aussagen zu ermöglichen.

Die Daten für die Selektion der Personas

Für die Selektion der Personas dienen die folgenden Merkmale. Sie bestimmen maßgeblich den Zuschnitt der 16 Gruppen. Auf welchen Datensätzen und Methodiken die Merkmale beruhen wird im Folgenden ausgeführt.

Energieträger der Heizung

Um die Verteilung der Energieträger in Deutschland zu erfassen, greift infas360 auf Millionen Datensätze von Echtfällen zu Heizungstypen vom Immobilienportal Immoscout zurück. Dazu werden Daten des Zensus 2022, einer großangelegten alle 10 Jahre stattfindenden kleinräumigen statistischen Erhebung zu Einwohnerzahl, Haushaltsgröße und weiteren Daten, verwendet. Der Zensus arbeitet dabei mit Haushaltsbefragungen, Registerdaten und Hochrechnungen. Auf Basis dieser Datensätze wird ein Schätzmodell entwickelt, das die Wahrscheinlichkeit angibt, mit der Öl, Gas Fernwärme oder weitere Energieträger verwendet werden.

Eigentumsform pro Gebäude

Um die vorherrschenden Eigentumsform pro Gebäude festzustellen, werden amtliche Gebäude- und Adressdaten erfasst. Mittels Haushaltsbefragungen und Registerdaten des Zensus 2022 kann den Gebäuden für lokale Raster (bis zu 100mx100m) eine Eigentumsform, wie z.B. Wohnungseigentümergeinschaft, Privatperson oder Wohnungsunternehmen zugewiesen werden. Unklare Fälle werden geschätzt, um die Eigentumsform mit der höchsten Wahrscheinlichkeit zu erhalten.

Eigentümerstruktur pro Gebäude

Die Eigentümerstruktur ergibt sich aus der Eigentumsform in Verbindung mit dem Gebäudetyp. Dafür wird zuerst zwischen Gebäudetypen Ein- und Zweifamilienhäusern und Mehrfamilienhäusern unterschieden. Die Haushaltsbefragungen und Erhebungen aus dem Zensus liefern Daten zur Eigentumsform in einem Siedlungsgebiet. Der jährlich stattfindende Mikrozensus gibt Informationen, in welchen Gebäudetypen welche Haushalte Mieter oder Eigentümer sind. So ergibt sich z.B. die Menge an Wohnungen in Mehrfamilienhäusern in einem Gebiet, die von Eigentümern bewohnt werden.

Äquivalenzeinkommen

Zur Vergleichbarkeit verschiedener Haushalte dient das Nettoäquivalenzeinkommen als zentrale Einkommensgröße in der Persona-Analyse. Infas360 verfügt über einen durch Befragungen entwickelten und durch Schätzungen auf die Grundgesamt ausgedehnten adressscharfen Datensatz der monatliche Kaufkraft pro Haushalt. Dieser Wert muss mit Hilfe der Haushaltszusammensetzung in das Nettoäquivalenzeinkommen umgerechnet werden.

Die Zusammensetzung der Haushalte hat Infas 360 auf Basis von eigenen stichprobenartigen Haushaltsbefragungen und Datenerhebungen (Mikrozensus 2021) ermittelt. Damit können dann Erwachsene (Faktor 1), Jugendliche (Faktor 0,5) und Kinder (Faktor 0,3) gewichtet werden, um ein Haushaltseinkommen einer Familie mit dem eines Einpersonenhaushalts vergleichen zu können. Das sich ergebende Äquivalenzeinkommen wurde anschließend angepasst an den regionalen Einkommensmedian (NUTS-II) und an das bundesweite durchschnittliche Jahres-

Äquivalenzeinkommen. Es ergeben sich in 500er bis 1000er Schritten Einkommensklassen von unter 1000 Euro bis über 4000 Euro Nettoäquivalenzeinkommen.

Baujahresklassen

Zur Ermittlung des Gebäude-Baujahresklassen greift infas360 auf 3,4 Mio. echter Baualterswerte bestehender Immobilien von Immoscout24 zurück. Diese Werte werden als Grundlage verwendet, um Schätzungen der Baualtersklasse vorzunehmen. Handelt es sich bei einem Echtfall beispielsweise um ein Reihenhaus, wird für die Gebäude in der gleichen Baureihe das gleiche Baujahr angenommen. Abgeglichen werden die Ergebnisse der Schätzungen mit den Altersklassen, die statistische Zensus-Erhebung 2022 gebietsweise ermittelt hat. So ergeben sich in Klassen von 10 Jahren Werte vom Baujahr 1919 bis 2010 und später.

Die Daten für die Profilierung der Personas

Neben den Daten zur Selektion der 16 Persona-Gruppen kommen weitere Merkmale zum Einsatz, die die Profile der Haushalte konkretisieren sollen. Die Methodik ihre Zustandekommens wird im Folgenden erläutert.

Wohnlage/Gemeindetypologie

Zur Einordnung der Wohnlage der jeweiligen Personas werden die Wohnorte der Haushalte dem zugehörigen Gemeindetyp zugewiesen. Amtliche Daten des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR) ordnen alle Gemeinden in Deutschland den Kategorien, Groß-, Klein-Mittelstadt und Landgemeinden zu. Diese Kategorisierung übernimmt infas 360 für die Persona-Analyse. Verknüpft werden diese Daten mit den jeweiligen Adressen der Gebäude

Die Einordnung der Wohnlage der Personas in Großstadt, Mittelstadt, Kleinstadt und Dorf (Landgemeinde) wird zunächst über die Einteilung zwischen Stadt und Landgemeinde vorgenommen. Hat eine Gemeinde mind. 5.000 Einwohner oder eine zentrale Funktion, wird diese als "Stadt" bezeichnet. Wenn nicht, handelt es sich um eine Landgemeinde.

Bei den Städten werden folgende Typen unterschieden:

- Großstadt: Gemeinde mit mindestens 100.000 Einwohnern
- Mittelstadt: Gemeinde mit 20.000 bis unter 100.000 Einwohnern
- Kleinstadt: Gemeinde mit 5.000 Einwohnern bis unter 20.000 Einwohnern

Personen pro Haushalt

Wie viele Personen in einem Haushalt leben, ermittelt infas 360 über Zahlen der stichprobenartigen Mikrozensus-Erhebung von 2021. Hier wird für ein 100x100m-Raster ermittelt, wie groß die Haushalte in diesem Bereich sind. Verbunden wird das mit vorhandenen adressscharfen Daten zur Wohnfläche und dem Gebäudetyp, um die Größe der Haushalte zu ermitteln. Die Ergebnisse werden anschließen abgeglichen mit der Einwohnerzahl in diesen Gebieten. Bei unschlüssigen Daten wird die Haushaltsgröße geschätzt.

Energieverbrauch

Der Energieverbrauch wird auf Basis des Energiebedarf ermittelt. Dieser spiegelt den jährlichen Bedarf für Heizung und Warmwasser in KWh pro qm Wohnfläche wider. Er gibt also den theoretischen Energiebedarf eines Gebäudes aufgrund der Bauweise an und nicht den tatsächlichen Energieverbrauch. Das Immobilienportal Immoscout hat echte Daten zum Energiebedarf von 1,2 Mio. Gebäuden. Die Werte werden auf den Quadratmeter heruntergerechnet und auf alle Wohngebäude übertragen. Zusätzlich fließen Daten zu

Haushalten im Siedlungsblock, Baujahresklasse, Sanierungsquoten und weitere ein, um den Energiebedarf möglichst genau zu schätzen.

Der Energieverbrauch wird mit einer Formel aus dem Energiebedarf errechnet. Dabei wird davon ausgegangen, dass Haushalte in besonders ineffizienten Gebäuden sparsam heizen und somit weniger Energie verbrauchen als der Zustand des Gebäudes theoretisch nahelegen würde.

Dominierende Altersstruktur

Die Verteilung der Altersklassen je Persona ergeben sich aus Zustelldaten der Deutschen Post. Diese werden verglichen mit Daten des Zensus von 2022, die auf 100x100m genau den Anteil der über 65-Jährigen ermitteln. Zusätzlich gibt es Gemeindedaten von 2022 zur Anzahl der über 60-Jährigen pro Gemeinde.

Bevölkerungsprognose

Die Bevölkerungsprognose schätzt die Bevölkerungsentwicklung auf Gemeindeebene bis 2035 ab. Sie wird insbesondere von natürlichen Vorgängen, also Sterbefälle und Lebendgeburten, und Wanderungen, also Zu- und Fortzügen, beeinflusst. Diese Daten lassen sich über die Regionaldatenbank Deutschland abrufen. In der Regionaldatenbank der statistischen Ämter in Deutschland finden sich zudem Zahlen aus unterschiedlichen Befragungen, Verwaltungsdaten und statistischen Modellen. Für die Personas werden die Bevölkerungsentwicklungen seit 2008 analysiert und darauf basierend bis 2035 geschätzt, ob die Bevölkerung schrumpfen wird, stabil bleibt oder wachsen wird.

ÖPNV-Verfügbarkeit

Die ÖPNV-Verfügbarkeit wird über einen ÖPNV-Index ermittelt. Er gibt die Verfügbarkeit öffentlicher Verkehrsmittel je Adresse an. Dafür wird ermittelt, welche Haltestellenart (Bus-, Straßenbahn-, U-Bahn, DB-Haltestellen) von der Adresse aus gut erreichbar ist. Dabei werden schienengebundene Verkehrsmittel höher gewichtet. Es ergeben sich verschiedene Klassen auf Adressebene (z.B. nur ein Typ, drei Typen von Haltestellen). Es ergibt sich ein Index, der zeigt, wie gut die entsprechende Anbindung an den ÖPNV vor Ort ist. Je näher und vielfältiger die Haltestellen, desto besser Index-Wert.

E-Auto-Anteil

Auf Basis von Befragungen von infas 360 werden E-Auto-Besitzer und E-Auto-Interessierte ermittelt. Anschließend wird geschaut, wie die Antworten mit Merkmalen wie Solaranlage vorhanden, Distanz zur nächsten Ladestation und weiteren zusammenhängen. So kann geschätzt werden, wie wahrscheinlich es ist, dass sich ein Haushalt an einer bestimmten Adresse ein E-Auto zulegt, oder schon zugelegt hat.