

glmnetUI: Eine interaktive Benutzeroberfläche für das glmnet-Paket (Lasso und Elastic Net)

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische

Fassung ist maßgeblich.

William Bert Craytor

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

DOI: [10.67330/ja8s2k39](https://doi.org/10.67330/ja8s2k39)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Dieser Open-Access-Artikel wird unter den Bedingungen der [Creative-Commons-Namensnennung-4.0-Lizenz](https://creativecommons.org/licenses/by/4.0/) verbreitet.

Zusammenfassung. glmnetUI ist eine grafische Benutzeroberfläche für das R-Paket `glmnet`, das regularisierte generalisierte lineare Modelle mittels Lasso-, Ridge- und Elastic-Net-Straftermen anpasst. Es bietet drei Verwendungsmodi — allgemeine prädiktive Modellierung, Immobilienbewertung und Marktgebietsanalyse — und führt den Anwender durch Datenimport, Modellkonfiguration, kreuzvalidierte Modellanpassung, Koeffizientenpfade sowie herunterladbare Berichte. Dieser Artikel dokumentiert die Datenformat-Anforderungen, den Modellierungs-Workflow, die Ausgabedarstellungen und die vollständige Funktionsreferenz von glmnetUI.

Schlüsselwörter: Lasso, Ridge-Regression, Elastic Net, regularisierte Regression, glmnet, R, grafische Benutzeroberfläche, Immobilienbewertung

Anmerkung des Verlags (Juli 2026): Kurz vor Drucklegung wurde glmnetUI in earthUI (Version 0.11.0 und höher) zusammengeführt; earthUI stellt den hier beschriebenen Elastic-Net-Workflow nun neben den MARS- und GAM-Workflows bereit. glmnetUI bleibt in seiner letzten veröffentlichten Fassung (0.5.0) unter <https://github.com/wcraytor/glmnetUI> verfügbar; die Weiterentwicklung erfolgt in earthUI. Dieser Artikel bezieht sich auf glmnetUI in der veröffentlichten Form.

Inhaltsverzeichnis

1	Einleitung	3
1.1	Geschichte und Hintergrund	3
1.1.1	Das Lasso	3
1.1.2	Ridge-Regression und Elastic Net	4
1.1.3	Das Paket glmnet	4
1.1.4	Vergleich mit MARS (earth) und GAM (mgcv)	4
1.1.5	Empfohlener Arbeitsablauf	5
1.2	Was ist glmnetUI?	6
1.3	Drei Purpose-Modi	6
1.4	Immobilien-spezifische Funktionen	7
1.5	Erste Schritte	7
2	Anforderungen an die MLS-Eingabedaten	9
2.1	Dateiformat und Struktur	9
2.2	Erforderliche Spalten für den Wertermittlungsmodus	9
2.3	Konventionen für spezielle Spaltenbezeichnungen	10
2.4	Datenqualität und Vollständigkeit	10
2.5	Platzierung der Bewertungsobjektzeile	11

3	Modus General Purpose	12
3.1	Überblick	12
3.1.1	Skip First Row	12
3.1.2	Einstellungen je Purpose	12
3.2	Der Arbeitsablauf in der Seitenleiste	12
3.3	Projekte	13
3.4	Import aus earthUI (optional)	14
3.5	Registerkarten des Hauptbereichs	14
3.6	Persistenz der Einstellungen	14
3.7	Dark Mode	15
3.8	Locale- und regionale Einstellungen	15
3.8.1	Standardwerte speichern	15
4	Modus Appraisal	16
4.1	Behandlung der Bewertungsobjektzeile	16
4.2	Effective Date und Sale Age	16
4.3	Spezielle Spaltenkennzeichnungen	16
4.4	Überblick über die RCA-Anpassungen	16
5	Modus Market Area Analysis	17
5.1	Unterschiede zum Modus Appraisal	17
5.2	Wann der Market-Modus geeignet ist	17
6	Variablenauswahl	18
6.1	Zielvariable	18
6.2	Die Prädiktortabelle	18
6.3	Erkennung und Überschreiben von Datentypen	18
6.4	Erwartete Vorzeichen und deren Durchsetzung	18
6.5	Referenz der Special-Spaltentypen	19
7	Parameterauswahl	20
7.1	Alpha (Mischungsparameter)	20
7.2	Wahl von Lambda	20
7.3	Family	20
7.4	Standardize	21
7.5	Vorzeichenrestriktionen	21
7.6	Relaxed Lasso	21
7.7	Zufallsstartwert	21
7.8	Interaktionsmatrix	22
7.9	Vom Haupteffekt ausschließen (nur Interaktion)	22
7.10	Erweiterte Parameter	22
7.11	Standardwerte der Einstellungen	23
8	Anpassung des Modells	24
8.1	Die Schaltfläche zur Modellanpassung	24
8.2	Was bei der Anpassung geschieht	24
9	Ergebnis-Registerkarten	25
9.1	Data Preview	25
9.2	Equation	25
9.3	Correlation	25
9.4	Summary	25

9.5	Coefficients	26
9.6	Variable Importance	26
9.7	Contributions	26
9.8	ANOVA	26
9.9	Diagnostics	26
9.10	Glmnet Output	27
9.11	Report	27
10	Herunterladen von Daten	28
10.1	Ausgabespalten	28
10.2	CQA-Scores	28
11	RCA-Berechnungen und Herunterladen	29
11.1	Öffnen des RCA-Dialogs	29
11.2	Interpolation des CQA-Werts	29
11.3	Ausgabespalten	29
12	Sales Comparison Grid	30
12.1	Dialog zur Auswahl der Vergleichsobjekte	30
12.2	Aufbau der Arbeitsmappe	30
12.3	Blattschutz und Restmerkmalzellen	30
13	Herunterladen von Berichten	31
13.1	Der zweistufige Quarto-Ablauf	31
13.2	Berichtsformate	31
13.3	Berichtsinhalte	31
14	Vergleich mit earthUI	33
14.1	Wesentliche Unterschiede	33
14.2	Gemeinsame Funktionen	33
14.3	Wann welches Werkzeug	33
15	Demodatensatz: Appraisal_1.csv	35
15.1	Beschreibung	35
15.2	Spalten	35
15.3	Vorgeschlagener Schnelleinstieg	36
16	Systemanforderungen und Fehlerbehebung	37
16.1	Unterstützte Plattformen	37
16.2	Optionale Abhängigkeiten	37
16.3	Plattformhinweise	37
16.4	Verhalten bei fehlenden Komponenten	37
16.5	Fehlerbehebung	37

1 Einleitung

1.1 Geschichte und Hintergrund

1.1.1 Das Lasso

Das **Lasso** (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator) wurde 1996 von Robert Tibshirani eingeführt. Der Name ist ein bewusstes Wortspiel mit dem Lasso des Cowboys — einem Seil,

das einschränkt und auswählt. Tibshirani wählte ihn, weil das Verfahren die Koeffizienten buchstäblich „einfängt“: Es beschränkt ihre Absolutbeträge und zieht einige exakt auf null, wodurch ausgewählt wird, welche Variablen im Modell verbleiben.

Tibshiranis Einsicht bestand darin, der Kleinste-Quadrate-Zielfunktion einen L_1 -Strafterm (Summe der Absolutbeträge der Koeffizienten) hinzuzufügen. Anders als der L_2 -Strafterm der Ridge-Regression (Hoerl & Kennard, 1970), der die Koeffizienten in Richtung null schrumpft, sie aber nie erreicht, erzeugt der L_1 -Strafterm *dünnbesetzte* Lösungen, bei denen viele Koeffizienten exakt null sind. Damit ist das Lasso zugleich ein Regularisierungs- und ein Variablenselektionsverfahren.

1.1.2 Ridge-Regression und Elastic Net

Die **Ridge-Regression** wurde 1970 von Arthur Hoerl und Robert Kennard vorgeschlagen. Sie ergänzt einen L_2 -Strafterm (Summe der quadrierten Koeffizienten), der alle Koeffizienten proportional schrumpft, jedoch sämtliche Variablen im Modell belässt. Ridge ist wirksam, wenn die Prädiktoren korreliert sind (Multikollinearität), vereinfacht das Modell jedoch nicht.

Das **Elastic Net**, 2005 von Hui Zou und Trevor Hastie eingeführt, verbindet beide Strafterme. Der Mischungsparameter α steuert die Gewichtung: $\alpha = 1$ entspricht dem reinen Lasso, $\alpha = 0$ der reinen Ridge-Regression, Werte dazwischen ergeben einen Kompromiss. Das Elastic Net überwindet eine Schwäche des Lasso: Bei stark korrelierten Prädiktoren neigt das Lasso dazu, einen auszuwählen und die übrigen zu ignorieren, während das Elastic Net korrelierte Prädiktoren gruppiert.

1.1.3 Das Paket glmnet

Das R-Paket **glmnet** wurde von Jerome Friedman, Trevor Hastie und Robert Tibshirani an der Stanford University entwickelt. Es implementiert die Elastic-Net-Regularisierung für generalisierte lineare Modelle mithilfe eines äußerst effizienten Koordinatenabstiegs-Algorithmus (Friedman, Hastie und Tibshirani 2010); es wurde seither auf Cox-Modelle proportionaler Hazards (Simon u. a. 2011) und auf sämtliche Familien generalisierter linearer Modelle (Tay, Narasimhan und Hastie 2023) erweitert. Das Paket unterstützt die Familien gaussian (linear), binomial (logistisch), poisson und weitere GLM-Familien.

Zu den wesentlichen Merkmalen von **glmnet** zählen:

- **Regularisierungspfad:** Passt das Modell in einem einzigen Durchlauf über ein Gitter von Lambda-Werten an — weit schneller, als für jedes Lambda einzeln zu schätzen.
- **Kreuzvalidierung:** `cv.glmnet()` wählt die optimale Regularisierungsstärke automatisch über eine k -fache Kreuzvalidierung.
- **Koeffizientenschränken:** Unterstützt obere und untere Grenzen für einzelne Koeffizienten (Vorzeichenrestriktionen).
- **Relaxed Lasso:** Schätzt das Modell auf den ausgewählten Variablen mit verringertem oder ohne Strafterm neu und verringert so die Verzerrung.

1.1.4 Vergleich mit MARS (earth) und GAM (mgcv)

Die earth-Anwendung und ihre beiden Post Processing Modules (Nachverarbeitungsmodule) glmnet und mgcv verwenden unterschiedliche Modellierungsmaschinen. Die folgende Tabelle fasst die wesentlichen Unterschiede zusammen:

Merkmal	glmnet (Elastic Net)	earth (MARS)	mgcv (GAM)
Modelltyp	Linear (mit Regularisierung)	Stückweise linear (adaptive Splines)	Glatt nichtlinear (penalisierte Splines)
Nichtlinearität	Nur über Interaktionen oder Basiserweiterung	Automatisch über Hinge-Funktionen mit datengetriebenen Knoten	Automatisch über Glättungsfunktionen (Thin Plate, kubisch usw.)
Variablenselektion	Eingebaut über L_1 -Strafterm (Lasso)	Eingebaut über Vorwärts-/Rückwärtsschritte mit GCV-Pruning	Optional über Doppelstrafterm-Schrumpfung (<code>select=TRUE</code>)
Interaktionen	Paarweise Kreuzprodukte (schwer interpretierbar)	Hinge-basierte Produkte (interpretierbar, bis Grad 3)	Tensorprodukt-Glättungen (<code>te()</code> , <code>ti()</code>)
Interpretierbarkeit	Koeffizienten sind lineare Gewichte; einzeln leicht erklärbar	g-Funktionen zeigen stückweise lineare partielle Effekte je Variable	Glatte partielle Effektkurven; sehr anschaulich
Geeignet für	Hochdimensionale Daten, Variablenselektion, kleine Stichproben	Automatische Erkennung von Nichtlinearität und Interaktionen	Glatte nichtlineare Zusammenhänge, flexible Modellierung
RCA-Anpassungen	Lineare Anpassungen je Variable	Stückweise lineare Anpassungen über g-Funktionen	Glatte Anpassungen über partielle Effekte
Overfitting-Kontrolle	Kreuzvalidiertes Lambda	GCV-basiertes Pruning	REML-/GCV-Schätzung der Glättungsparameter

MARS (Multivariate Adaptive Regression Splines) wurde 1991 von Jerome Friedman eingeführt. Das Verfahren bildet stückweise lineare Modelle, indem es Hinge-Funktionen $\max(0, x - k)$ und deren Knotenpositionen adaptiv aus den Daten auswählt. Die R-Implementierung ist das Paket **earth** von Stephen Milborrow. MARS eignet sich hervorragend dafür, nichtlineare Zusammenhänge und Interaktionen automatisch aufzudecken, ohne dass der Anwender sie vorgeben muss. Seine g-Funktionen (nach Variablen gruppierte Basisterme) liefern sehr gut interpretierbare partielle Effektkurven.

GAMs (Generalized Additive Models) wurden von Hastie und Tibshirani (1990) formalisiert. Die R-Implementierung **mgcv** von Simon Wood stellt penalisierte Regressionssplines mit automatischer Glättungsauswahl über REML oder GCV bereit. GAMs erzeugen glatte partielle Effektkurven, die anschaulich darzustellen und zu erläutern sind. Wenn **earthUI** Knotenpositionen nach **mgcvUI** exportiert, dienen die aus **earth** abgeleiteten Knoten als Startpunkte für die GAM-Glättungsterme; so verbindet sich die adaptive Knotensetzung von MARS mit der glatten Schätzung des GAM.

1.1.5 Empfohlener Arbeitsablauf

Für die Immobilienbewertung und ähnliche Anwendungen, die interpretierbare und belastbare Modelle erfordern:

1. Mit **earthUI** beginnen, um die wichtigen Variablen, nichtlinearen Zusammenhänge

und Interaktionen in den Daten aufzudecken. Die automatische Basisauswahl und die g-Funktionen von earth liefern ausgezeichnete erste Erkenntnisse.

2. **Mit mgcvUI verfeinern**, wenn glattere partielle Effekte gewünscht sind. Die Knoten aus earthUI lassen sich nach mgcvUI importieren und erlauben so einen nahtlosen Übergang von stückweise linearen zu glatten Modellen.
3. **glmnetUI verwenden**, wenn eine konsequente Variablenselektion (Lasso) benötigt wird, wenn die Zahl der Prädiktoren sich der Zahl der Beobachtungen nähert oder diese übersteigt, oder wenn ein rein lineares Modell mit Regularisierung aus Gründen der Belastbarkeit gewünscht ist.

1.2 Was ist glmnetUI?

glmnetUI ist eine grafische Benutzeroberfläche für das R-Paket **glmnet**. Es läuft als lokale Shiny-Anwendung — es gibt keine Anmeldung, keinen Server und keine Benutzerkonten. Die Anwendung wird aus R heraus gestartet; anschließend wird ein Datensatz (CSV oder Excel) importiert, das Modell konfiguriert und interaktiv angepasst.

Seit der aktuellen Version ist glmnetUI kein separat installiertes Paket mehr: Es wird *innerhalb* des Pakets earthUI als dessen Elastic-Net Post Processing Module ausgeliefert und mit `earthUI::launch_glmnet()` gestartet. Die Anwendung selbst — und alles in diesem Artikel Beschriebene — bleibt unverändert; lediglich Installation und Startbefehl unterscheiden sich. Die drei Routinen (earth, glmnet, mgcv) teilen sich einen Projektbaum, eine Einstellungsdatenbank und eine Kopie der unterstützenden Infrastruktur.

Die Anwendung stellt einen vollständigen Arbeitsablauf bereit: Datenimport, Variablenkonfiguration, Modellanpassung, Diagnosegrafiken, Variablenwichtigkeit, Modellgleichungen sowie herunterladbare Berichte im Word-, PDF- oder HTML-Format.

Die Elastic-Net-Regression verbindet zwei Regularisierungstechniken:

- **Ridge-Regression** ($\alpha = 0$): Schrumpft alle Koeffizienten proportional in Richtung null. Geeignet, wenn alle Variablen als relevant angesehen werden.
- **Lasso-Regression** ($\alpha = 1$): Kann Koeffizienten exakt auf null schrumpfen und führt damit eine automatische Variablenselektion durch. Erzeugt einfachere, besser interpretierbare Modelle.
- **Elastic Net** ($0 < \alpha < 1$): Eine Mischung beider Ansätze. Nützlich bei korrelierten Prädiktoren.

1.3 Drei Purpose-Modi

Jedes glmnetUI-Projekt trägt einen **Purpose** (Verwendungszweck), der bei der Erstellung gewählt wird (Abschnitt 1, *Project*). Der Purpose des aktiven Projekts bestimmt, welche Werkzeuge und Oberflächenelemente erscheinen; um in einem anderen Modus zu arbeiten, wird ein Projekt mit dem entsprechenden Purpose angelegt oder geöffnet. Es gibt drei Modi:

- **General** (Allgemein) — Elastic-Net-Regression für beliebige Populationen oder Datensätze. Dies ist der Standardmodus. Er stellt den vollständigen regularisierten Regressionsablauf ohne domänenspezifische Ergänzungen bereit.
- **For Appraisal** (für die Wertermittlung) — Elastic-Net-Regression, zugeschnitten auf die Immobilienbewertung. Ergänzt Funktionen speziell für die Bewertung einzelner Objekte, darunter die Behandlung des Bewertungsobjekts, spezielle Spaltenkennzeichnungen sowie den Residual Constraint Approach (RCA) (dt. etwa: Residualrestriktionsansatz).
- **Market Area Analysis** (Marktgebietsanalyse) — Elastic-Net-Regression, zugeschnitten auf Marktgebietsstudien. Ergänzt Funktionen zur Analyse von Objektgruppen in einem definierten Markt, mit optionalem Ausschluss der Bewertungsobjektzeile.

In allen drei Modi ist die zentrale Modellierungsmaschine identisch — es wird stets ein Elastic-Net-Modell (glmnet) angepasst. Die Purpose-Einstellung steuert, welche zusätzlichen Werkzeuge und Oberflächenelemente verfügbar sind.

1.4 Immobilienspezifische Funktionen

Wenn entweder **For Appraisal** oder **Market Area Analysis** gewählt ist, aktiviert glmnetUI mehrere Funktionen, die für die Immobilienanalyse konzipiert sind:

- **Spezielle Spaltenkennzeichnungen** — Jeder Prädiktor kann mit einer speziellen Rolle versehen werden, etwa `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` oder `display_only`. Diese Kennzeichnungen steuern, wie die Spalte bei der Modellanpassung und in der Ausgabe behandelt wird.
- **Rundung von Breiten- und Längengrad** — Als `latitude` oder `longitude` gekennzeichnete Spalten werden automatisch auf drei Nachkommastellen gerundet, um Overfitting zu vermeiden.
- **Spalte Sale Age (Verkaufsalter)** — Ist eine Spalte als `contract_date` gekennzeichnet und ein Effective Date (Stichtag) angegeben, berechnet glmnetUI eine Spalte `sale_age` (Tage zwischen Verkaufsdatum und Stichtag) und setzt sie als Prädiktor ein.
- **RCA-Berechnungen (nur Wertermittlung)** — Im Wertermittlungsmodus kann glmnetUI nach der Modellanpassung eine Ausgabe zum Residual Constraint Approach (RCA) erzeugen, die Anpassungen je Vergleichsobjekt, Netto-/Brutto-Anpassungsübersichten sowie einen angepassten Verkaufspreis für das Bewertungsobjekt liefert.

***Hinweis:** Der Purpose-Modus General eignet sich für beliebige Datensätze — aus Finanzwesen, Wissenschaft, Ingenieurwesen oder Immobilienwirtschaft. Die Modi für Wertermittlung und Marktanalyse fügen lediglich Komfortfunktionen für Immobilienfachleute hinzu.*

1.5 Erste Schritte

Zur Verwendung von glmnetUI:

1. **earthUI in R installieren:** `install.packages("earthUI")` — die glmnet-Routine ist enthalten. (Siehe „Installing earthUI and vProlog“ im earthUI-Artikel dieser Ausgabe für die Entwicklungsversion und optionale Komponenten.)
2. **Anwendung starten:** `earthUI::launch_glmnet()` in R ausführen. Die Anwendung öffnet sich im Webbrowser auf Port 7879. Sie lässt sich auch direkt über `http://localhost:7879` im Browser aufrufen. Die Anwendung merkt sich den zuletzt verwendeten Purpose-Modus und stellt ihn automatisch wieder her.
3. **Projekt anlegen oder öffnen** in Abschnitt 1 der Seitenleiste. Ein Projekt legt den Purpose und den Arbeitsort (Land, Bundesstaat, County, Stadt) sowie seine Eingabe- und Ausgabeordner fest. Projekte werden unter einem gemeinsamen `regProj`-Wurzelverzeichnis abgelegt und mit den Schwesteranwendungen earthUI und mgcvUI geteilt. Das Öffnen eines Projekts mit anderem Purpose ist der Weg, den Modus zu wechseln.
4. **Daten importieren** in Abschnitt 2 durch Auswahl einer CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts. Dateien in diesen Ordner legen und anschließend auf **Refresh** (Aktualisieren) klicken, um sie aufzulisten.
5. **Import aus earthUI (optional)** — Abschnitt 3 ermöglicht den Import einer earthUI-Ergebnisdatei (`.rds`), um die Hinge-Basisfunktionen von earth mit der Regularisierung von glmnet zu verbinden. Dieser Schritt entfällt, wenn kein earthUI-Ergebnis vorliegt.
6. **Variablen konfigurieren** — Zielvariable und Prädiktoren wählen, Datentypen und erwartete Vorzeichen festlegen und etwaige spezielle Spaltenrollen zuweisen.

7. **glmnet-Parameter einstellen** — alpha, lambda, family, Vorzeichenrestriktionen, Interaktionen und weitere Optionen.
8. **Modell anpassen** — auf „Fit Glmnet Model“ klicken und die Ergebnisse im Hauptbereich prüfen.
9. **Exportieren** — Prognosen als Excel-Datei herunterladen, einen Quarto-Bericht erzeugen und konvertieren oder (im Wertermittlungsmodus) RCA-Anpassungen und ein Sales Comparison Grid (Vergleichsraster) berechnen.

Die Einstellungen werden automatisch im lokalen Speicher des Browsers gesichert und wiederhergestellt, sobald dieselbe Eingabedatei erneut geladen wird.

2 Anforderungen an die MLS-Eingabedaten

Für Arbeitsabläufe der Immobilienbewertung und Marktanalyse stammen die Eingabedaten üblicherweise aus einem Export eines Multiple Listing Service (MLS). Dieses Kapitel beschreibt die erwartete Dateistruktur und die Spalten, die glmnetUI verwenden kann.

2.1 Dateiformat und Struktur

glmnetUI akzeptiert **CSV**- und **Excel**-Dateien (`.xlsx`, `.xls`). Beim Import werden Spaltennamen automatisch in **snake_case** umgewandelt — aus „Living SqFt“ wird beispielsweise `living_sqft`, aus „Contract Date“ wird `contract_date` und aus „Sale Price“ wird `sale_price`. Diese Normalisierung gewährleistet einheitliche Spaltenbezüge über den gesamten Arbeitsablauf. Das beim Import verwendete CSV-Trennzeichen und Dezimaltrennzeichen ergibt sich aus den Locale-Einstellungen (siehe Kapitel 3, „Locale & Regional Settings“).

Die Datendatei sollte eine flache Tabelle mit einer Zeile je Objekt und einer Spalte je Merkmal sein. Die erste Zeile der Datei muss die Spaltenüberschriften enthalten.

2.2 Erforderliche Spalten für den Wertermittlungsmodus

glmnetUI arbeitet zwar mit beliebigen Spaltensätzen, doch der vollständige Wertermittlungsablauf profitiert von den folgenden Spalten:

Spalte	Special-Typ	Zweck
Sale Price	(Ziel)	Antwortvariable des Modells
Contract Date	<code>contract_date</code>	Dient zur Berechnung von <code>sale_age</code> (Tage ab Stichtag)
Listing Date	<code>listing_date</code>	Zusammen mit dem Vertragsdatum zur DOM-Berechnung, falls keine DOM-Spalte vorhanden ist
Days on Market	<code>dom</code>	Vermarktungsdauer in Tagen; wird in Exporten angezeigt
Concessions	<code>concessions</code>	Verkaufszugeständnisse; Netto-Verkaufspreis = Verkaufspreis – Zugeständnisse
Living Area (SF)	<code>living_area</code>	Ermöglicht Residuen je Quadratfuß (<code>residual_sf</code> , <code>cqa_sf</code>)
Lot Size	<code>lot_size</code>	Spalte für die Grundstücksgröße
Site Dimensions	<code>site_dimensions</code>	Mit der Grundstücksgröße gruppiert
Latitude	<code>latitude</code>	Auf 3 Nachkommastellen gerundet
Longitude	<code>longitude</code>	Auf 3 Nachkommastellen gerundet
Area ID	<code>area</code>	Kennung des Marktgebiets bzw. der Lage
Actual Age	<code>actual_age</code>	Objektalter
Effective Age	<code>effective_age</code>	Effektives Objektalter
Address	<code>display_only</code>	Wird in Exporten angezeigt; vom Modell ausgeschlossen

Die Spaltennamen der Tabellenkalkulation dürfen fremdsprachig sein — die „speziellen“ Bezeichnungen sind englisch, damit das R-Programm sie gesondert behandeln kann. Im Übrigen erscheinen die vergebenen Spaltennamen in den Regressionsmodellen, den Grafiken und — bei Wertermittlungen — in den Ausgabeberichten.

Nicht alle Spalten sind erforderlich. glmnetUI passt sich an — fehlt eine Spalte, entfällt schlicht die zugehörige Funktion. Für Immobilienpreismodelle sind jedoch bestimmte Spalten dringend zu empfehlen, um eine akzeptable Anpassungsgüte zu erreichen:

1. **Sale Age** (Verkaufsalter) — die Anzahl der Tage zwischen dem Vertragsdatum des Verkaufs und dem Stichtag der Wertermittlung oder Analyse. Wird eine mehrjährige Verkaufshistorie verwendet, besonders über Zeiträume von mehr als fünf Jahren, spielt `sale_age` bei der Schätzung des Verkaufspreises häufig eine zentrale Rolle.
2. **Living Area** (Wohnfläche) — auch unter Bezeichnungen wie „Living Sqft“, „GLA“ (gross living area) und ähnlich geführt. Dies ist eine weitere zentrale Bestimmungsgröße des Verkaufspreises.
3. **Total Bath Count** (Gesamtzahl der Bäder) — die Gesamtzahl der ganzen, Viertel-, halben und Dreiviertelbäder. Zwei Vollbäder und ein Gästebad ergeben beispielsweise den Wert 2,5.
4. **Garage Bays** oder **Garage Area** — die Anzahl der Garagenstellplätze oder die Garagenfläche.
5. **Lot Size** (Grundstücksgröße) — die Grundstücksfläche des Objekts, üblicherweise in Quadratfuß oder Acres.
6. **Longitude**, **Latitude** und, sofern verfügbar, **Area ID**. Diese Lagevariablen helfen dem Modell, geografische Preisunterschiede abzubilden.

2.3 Konventionen für spezielle Spaltenbezeichnungen

glmnetUI erkennt Spalten anhand ihrer **Special-Kennzeichnung**, nicht anhand ihres Spaltennamens. Die Spalten dürfen im MLS-Export beliebig benannt sein — entscheidend ist, dass in der Tabelle zur Variablenkonfiguration (Kapitel 6) der richtige Special-Typ zugewiesen wird.

So könnte das MLS die Wohnfläche etwa als „GLA“, „Living SqFt“, „liv_area“ oder „gross_living_area“ exportieren. Nach dem Import (bei dem der Name in snake_case umgewandelt wird) genügt es, sie im Special-Auswahlfeld als `living_area` zu kennzeichnen. glmnetUI verwendet sie dann unabhängig vom ursprünglichen Namen für die Berechnung der Residuen je Quadratfuß.

2.4 Datenqualität und Vollständigkeit

- **Fehlende Werte (NA)**: Zeilen mit NA-Werten in einer Prädiktor- oder Zielspalte werden vor der Modellanpassung automatisch entfernt.
- **Datumsspalten**: Zeichenspalten, die gängigen Datumsformaten entsprechen, werden beim Import automatisch geparkt. Formate mit zweistelliger Jahreszahl werden bevorzugt, wenn keine vierstellige Jahreszahl erkannt wird.
- **Numerische Spalten**: Verkaufspreis, Wohnfläche, Grundstücksgröße, Zugeständnisse und ähnliche Felder müssen numerisch sein. Exportiert das MLS Preise mit Währungssymbolen oder Tausendertrennzeichen (etwa „\$350,000“ oder „350.000,00“), sind diese gegebenenfalls vor dem Import zu bereinigen.
- **Faktorspalten**: Text-, TRUE/FALSE- und R-Faktorspalten werden automatisch als kategorial behandelt — eine Regression kann sie nur als Faktoren verarbeiten. Eine **numerische** Spalte (etwa ein numerischer Gebiets- oder Stilcode) wird nur dann zum Faktor, wenn das zugehörige Kontrollkästchen **Factor** gesetzt wird; glmnetUI leitet aus dem Wertebereich einer numerischen Spalte niemals einen kategorialen Charakter ab, sodass diskrete numerische Prädiktoren wie `bath_count` stetig bleiben. Die Special-Rolle ist von der Faktorentscheidung unabhängig.

***Hinweis:** Nach dem Import empfiehlt sich ein Blick auf die NA-Spalte in der Prädiktortabelle. Spalten mit vielen fehlenden Werten können dazu führen, dass Zeilen entfallen. Spalten mit hohem NA-Anteil sollten ausgeschlossen oder die Daten vor dem Import bereinigt werden.*

2.5 Platzierung der Bewertungsobjektzeile

Im Modus **Appraisal** muss **Zeile 1** das **Bewertungsobjekt enthalten**. Alle übrigen Zeilen sind Vergleichsverkäufe. Die Bewertungsobjektzeile wird von der Modellanpassung ausgeschlossen. Nach der Anpassung erzeugt das Modell dennoch Prognosen für diese Zeile.

Im Modus **Market Area Analysis** ist die Platzierung des Bewertungsobjekts in Zeile 1 optional.

Im Modus **General** gibt es keine besondere Zeilenbehandlung — alle Zeilen werden gleich behandelt.

3 Modus General Purpose

3.1 Überblick

Der Modus General Purpose ist beim Start von glmnetUI voreingestellt. Er stellt den vollständigen Elastic-Net-Regressionsablauf für beliebige Datensätze bereit — nicht nur für Immobilien. glmnetUI lässt sich für wissenschaftliche Daten, Finanzanalysen, ingenieurtechnische Studien oder jedes andere Regressionsproblem einsetzen.

Im Modus General entfallen in der Oberfläche die immobilienpezifischen Funktionen (Special-Spalten, Verkaufsalter, Koordinatenrundung, RCA). Die Seitenleiste ist verschlankt und konzentriert sich auf Variablenauswahl, Parameterkonfiguration, Modellanpassung und Export.

3.1.1 Skip First Row

Ein Kontrollkästchen **Skip first row** (erste Zeile überspringen) erscheint oben in der Seitenleiste, sobald ein General- oder Market-Projekt geöffnet ist. Ist es gesetzt, wird Zeile 1 von der Modellanpassung ausgeschlossen. Das ist nützlich, wenn Zeile 1 eine Ziel- oder Referenzbeobachtung enthält, die das Modell nicht beeinflussen soll. Im Modus Appraisal wird Zeile 1 (das Bewertungsobjekt) stets automatisch ausgeschlossen.

3.1.2 Einstellungen je Purpose

Einstellungen (Prädiktorauswahl, Modellparameter, Interaktionen) werden für jede Kombination aus Eingabedatei und Purpose-Modus getrennt gespeichert. Beim Wechsel zwischen General, Appraisal und Market bleiben die Einstellungen jedes Modus unabhängig voneinander erhalten.

3.2 Der Arbeitsablauf in der Seitenleiste

Die Seitenleiste ist in nummerierte, ein- und ausklappbare Abschnitte gegliedert, die von der Projektauswahl bis zum Berichtsexport führen. Die meisten Abschnitte erscheinen erst, wenn ein Projekt geöffnet ist, und mehrere sind an den Purpose des Projekts gebunden. Der Ausgabeordner ergibt sich aus dem aktiven Projekt — ein separates Feld für den Ausgabeordner gibt es nicht.

1. Project (Projekt) — Projekt anlegen oder öffnen. Ein Projekt legt den Purpose und den Arbeitsort (Land, Bundesstaat, County, Stadt) sowie seinen Eingabeordner (`<os>_in/`) und seine Ausgabeordner fest. Projekte werden unter einem gemeinsamen **regProj**-Wurzelverzeichnis abgelegt (Standard `~/regProj`, überschreibbar mit der Umgebungsvariablen `REGPROJ_ROOT`) und mit den Schwesteranwendungen earthUI und mgcvUI geteilt, sodass Modelle aller drei Anwendungen nebeneinander liegen. Ein Klick auf **Close Project** wechselt das Projekt.

2. Import Data (Daten importieren) — Eine CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts auswählen. Dateien über Finder/Explorer in den Ordner legen und anschließend auf **Refresh** klicken, um sie aufzulisten. Bei Excel-Dateien mit mehreren Blättern erscheint eine Blattauswahl. Spaltennamen werden automatisch in snake_case umgewandelt.

3. Import from earthUI (optional) (Import aus earthUI) — Eine earthUI-Ergebnisdatei (`.rds`) importieren, um die Hinge-Basisfunktionen von earth mit der Elastic-Net-Regularisierung von glmnet zu verbinden. Die Schaltfläche zum Durchsuchen ist optisch an die Dateiauswahl in Abschnitt 2 angeglichen. Dieser Schritt ist optional — er entfällt, wenn kein earthUI-Ergebnis zum Import vorliegt.

4. Variable Configuration (Variablenkonfiguration) — Auswahlfeld für die Zielvariable, Prädiktortabelle mit Kontrollkästchen für Include, Factor und Force sowie Sign-Auswahlfeldern. Die Spalte Special (zur Kennzeichnung von Vertragsdaten, Koordinaten usw.) erscheint nur in den Modi Appraisal und Market. Einzelheiten in Kapitel 6.

5. glmnet Call Parameters (glmnet-Aufrufparameter) — Die gesamte Modellkonfiguration: alpha, lambda, family, standardize, Vorzeichenrestriktionen, Relaxed Lasso, Zufallsstartwert, Interaktionsmatrix sowie erweiterte Parameter (lambda.min.ratio, nlambda, CV-Verlustmaß, Konvergenzschwelle, maximale Iterationszahl, Achsenabschnitt). Die vollständige Parameterreferenz enthält Kapitel 7.

6. Fit Glmnet Model (Modell anpassen) — Die Schaltfläche, die das Modell rechnet. Die Ergebnisse erscheinen in den Registerkarten rechts.

7. Download Output (Ausgabe herunterladen) — Exportiert Prognosen, Residuen, CQA-Scores und Beiträge je Variable als Excel-Datei in den Ausgabeordner des Projekts. Nach der Modellanpassung in allen Purpose-Modi verfügbar.

8. Calculate RCA Adjustments & Download (RCA-Anpassungen berechnen und herunterladen) — Berechnet RCA-Anpassungen je Variable für jedes Vergleichsobjekt und interpoliert das Residuum des Bewertungsobjekts über den CQA-Wert. Erscheint nur in den Modi Appraisal und Market nach der Modellanpassung. Siehe Kapitel 11.

9. Generate Sales Grid & Download (Vergleichsraster erzeugen und herunterladen) — Erzeugt das vorläufige Sales Comparison Grid und ordnet die Vergleichsobjekte nach dem prozentualen Bruttoanpassungssatz. Erscheint nur im Modus Appraisal, nachdem die RCA-Anpassungen berechnet wurden. Siehe Kapitel 12.

10. Generate Quarto Report (Quarto-Bericht erzeugen) — Schreibt ein in sich geschlossenes Quarto-Berichtspaket (.qmd) — mit allen Grafiken und Tabellen — in den Ausgabeordner des Projekts. Nach der Modellanpassung verfügbar.

11. Convert Quarto Report (Quarto-Bericht konvertieren) — Rendert das erzeugte .qmd-Paket nach HTML, Word oder PDF. Für PDF ist eine LaTeX-Installation erforderlich; wird keine erkannt, wird die PDF-Option ausgeblendet. Siehe Kapitel 13.

In den Modi Market und General entfallen die nur für die Wertermittlung vorgesehenen Schritte, und die beiden Berichtsabschnitte werden entsprechend neu nummeriert.

3.3 Projekte

glmnetUI organisiert die Arbeit in **Projekten** unterhalb eines gemeinsamen **regProj**-Wurzelverzeichnisses (desselben Baums, den auch earthUI und mgcvUI verwenden). Jedes Projekt besitzt einen Eingabeordner (<os>_in) und je Modellierungsmaschine eigene Ausgabeordner (z. B. <os>_out_glmnet).

Zum Anlegen eines Projekts auf **+ New...** in Abschnitt 1 klicken und Folgendes angeben:

- **Purpose** — General, Appraisal oder Market Area.
- **Location** — die politisch-administrative Hierarchie des Projekts. Welche Ebenen auszufüllen sind, hängt vom gewählten Land ab, sodass ausländische Benennungskonventionen automatisch unterstützt werden — ein US-Projekt verwendet State/County/City, ein französisches Region/Département/Commune, ein japanisches Präfektur/Stadt-oder-Bezirk/Distrikt und so fort.
- **Project Name** — eine kurze Bezeichnung, begrenzt auf 8 Zeichen (Buchstaben, Ziffern, _ und -).

Die Anwendung übernimmt den eingegebenen Namen nicht wörtlich: Sie bildet einen eindeutigen, mit Zeitstempel versehenen Ordnernamen, indem sie die Ortscodes und den Erstellungszeitpunkt voranstellt, in der Form <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name> — aus einem US-Projekt namens burl wird beispielsweise us_ca_081_burlin_20260604-131500_burl. Die Grenze von 8 Zeichen gilt nur für den selbst eingegebenen Teil. Nur neu angelegte Projekte erhalten das Zeitstempel-Präfix; bereits auf der Festplatte vorhandene Projekte behalten ihre ursprünglichen Namen und bleiben vollständig zugänglich.

3.4 Import aus earthUI (optional)

Abschnitt 3 der Seitenleiste — **Import from earthUI** — ermöglicht den Import einer earthUI-Ergebnisdatei (`.rds`). Dadurch kann glmnetUI die Hinge-Basisfunktionen von earth (stückweise lineare Terme) mit der Elastic-Net-Regularisierung von glmnet verbinden und so die automatische Nichtlinearitätserkennung von earth mit der Variablenselektion und den Koeffizientenrestriktionen von glmnet kombinieren.

Wird ein earthUI-Ergebnis importiert, bestimmen die Prädiktoren des earth-Modells — nicht die Include-Kontrollkästchen — die glmnet-Anpassung, denn glmnet rechnet auf der Basiserweiterung von earth. Die Tabelle zur Variablenkonfiguration macht dies deutlich: **Include** und **Type** sind gesperrt, Zeilen für Prädiktoren, die das earth-Modell nicht verwendet hat, sind deaktiviert, und ein Hinweis erläutert, dass der Prädiktorsatz aus dem earth-Modell stammt. **Force** und **Sign** bleiben für die Prädiktoren von earth bearbeitbar, da sie zum Zeitpunkt der Anpassung weiterhin auf die earth-Basis wirken.

Die Schaltfläche zum Durchsuchen ist optisch an die Dateiauswahl in Abschnitt 2 angeglichen. Dieser Schritt ist vollständig optional — liegt keine earthUI-Ergebnisdatei vor, wird Abschnitt 3 einfach übersprungen und mit Abschnitt 4 fortgefahren.

3.5 Registerkarten des Hauptbereichs

Nach der Modellanpassung stellt der Hauptbereich die folgenden Registerkarten bereit:

Registerkarte	Inhalt
Data Preview	Vorschau der importierten Daten. Im Wertermittlungsmodus aufgeteilt in Tabellen für Bewertungsobjekt und Vergleichsverkäufe.
Equation	Die angepasste Modellgleichung, dargestellt in LaTeX/MathJax, mit allen von null verschiedenen Koeffizienten und Interaktionstermen.
Correlation	Heatmap der Korrelationen zwischen numerischen Prädiktoren (bereits vor der Anpassung verfügbar).
Summary	Kennzahlen (R^2 , adj. R^2 , GR^2 , $CV\text{-}R^2$, RMSE, MAE) sowie Kennzahlen der Wertermittlung (COD, PRD, Median Ratio).
Coefficients	Tabelle der von null verschiedenen Koeffizienten mit Warnhinweisen bei Vorzeichenverletzungen.
Variable Importance	Balkendiagramm und Rangtabelle der Prädiktorwichtigkeit ($ \beta \times sd(x)$). Interaktive plotly-Tooltips, sofern verfügbar.
Contributions	Grafiken der partiellen Effekte je Variable mit Steigungsbeschriftungen. Streudiagramm plus Anpassungsgerade für numerische Prädiktoren, Box-plots für Faktoren, 3D-Fläche/Streudiagramm/Heatmap für Interaktionen.
ANOVA	Tabelle der Varianzzerlegung: SS je Variable, % der Modell-SS, Koeffizient.
Diagnostics	Vier Grafiken: Koeffizientenpfad, CV-Fehler, beobachtet gegen prognostiziert, Residuen gegen angepasste Werte.
Glmnet Output	Rohausgabe des Modells: verwendeter Zufallsstartwert, Modellausdruck, gewähltes Lambda, $\lambda_{min}/1se$, Gamma (bei Relaxed Lasso) und vollständiger Koeffizientenvektor.
Report	Berichtskonfiguration und Exportfelder.
RCA Adjustments	Ergebnisse der RCA-Analyse (Wertermittlungsmodus, nach der Berechnung).

3.6 Persistenz der Einstellungen

glmnetUI speichert die Konfiguration automatisch im lokalen Speicher des Browsers, adressiert über den Namen der Eingabedatei. Wird dieselbe Datei erneut geladen, werden alle Einstellungen

wiederhergestellt: Zielauswahl, Prädiktor-Kontrollkästchen, Datentypen, erwartete Vorzeichen, glmnet-Parameter und Interaktionsmatrix. Auch der zuletzt verwendete Purpose-Modus wird global gesichert und beim Neustart der Anwendung wiederhergestellt. Drei Optionen stehen zur Verfügung:

- **Use last settings for input file** — die für diese konkrete Datei gespeicherten Einstellungen wiederherstellen
- **Use default settings** — die gespeicherten globalen Standardwerte anwenden
- **glmnet defaults** — auf die Werkseinstellungen zurücksetzen (alpha = 1, CV mit lambda.1se, 10 Folds, gaussian, standardize aktiviert)

Eine Schaltfläche **Save current as default** speichert sämtliche aktuellen Einstellungen als globalen Standard.

3.7 Dark Mode

Ein Klick auf das Mond-/Sonnensymbol in der rechten oberen Ecke schaltet zwischen hellem und dunklem Design um. Die Designwahl wird im lokalen Speicher gesichert und bleibt über Sitzungen hinweg erhalten. Sämtliche Oberflächenelemente (Tabellen, Grafiken, Karten, Schaltflächen) passen sich über CSS-Variablen dem gewählten Design an.

3.8 Locale- und regionale Einstellungen

glmnetUI unterstützt internationale Zahlen- und CSV-Konventionen über ein länderbasiertes Locale-System. Das Auswahlmenü **Settings** in der Titelleiste bietet die Selektoren **Country** und **Paper** für 31 unterstützte Länder. Jede Voreinstellung konfiguriert:

- **CSV-Trennzeichen** — Komma (,) für USA/Vereinigtes Königreich/Japan oder Semikolon (;) für den Großteil Europas, wo das Komma als Dezimaltrennzeichen dient.
- **Dezimaltrennzeichen** — Punkt (.) oder Komma (,).
- **Tausendertrennzeichen** — Komma (USA/Vereinigtes Königreich/Japan), Punkt (Deutschland/Italien/Spanien), Leerzeichen (Finnland/Frankreich/Polen/Baltikum/Ukraine/Russland) oder Apostroph (Schweiz).
- **Papierformat** — Letter (USA/Kanada/Mexiko) oder A4 (überall sonst).

3.8.1 Standardwerte speichern

Ein Klick auf **Save as my default** speichert die Locale-Einstellungen global. Diese Standardwerte gelten für alle künftigen Sitzungen, unabhängig davon, welche Datendatei geladen wird.

***Hinweis:** Die Zahlenformatierung auf Diagrammachsen und Steigungsbeschriftungen passt sich automatisch an das eingestellte Locale an. Das deutsche Locale verwendet Punkte als Tausendertrennzeichen (200.000), das finnische Leerzeichen (200 000), das schweizerische Apostrophe (200'000). Währungssymbole werden nicht angezeigt — glmnetUI ist währungsneutral.*

4 Modus Appraisal

Wird **For Appraisal** als Purpose gewählt, richtet sich glmnetUI auf die Bewertung eines einzelnen Objekts aus. Alle in Kapitel 3 beschriebenen Funktionen bleiben verfügbar; dieses Kapitel behandelt nur die wertermittlungsspezifischen Ergänzungen.

4.1 Behandlung der Bewertungsobjektzeile

Im Wertermittlungsmodus ist **Zeile 1 des Datensatzes das Bewertungsobjekt**, und alle übrigen Zeilen sind Vergleichsverkäufe. Die Eingabedatei muss entsprechend aufgebaut sein (siehe Kapitel 2). Der Verkaufspreis des Bewertungsobjekts kann leer bleiben oder einen beliebigen Wert enthalten — glmnetUI behandelt ihn bei der Anpassung automatisch als NA.

Nach dem Import teilt sich die Registerkarte Data Preview in zwei Bereiche: **Subject Property** (Zeile 1) und **Comparable Sales** (Zeilen 2 ff.). Zeile 1 wird stets von der Modellanpassung ausgeschlossen. Nach der Anpassung erzeugt das Modell dennoch Prognosen für diese Zeile.

4.2 Effective Date und Sale Age

In den Modi Appraisal und Market erscheint im Abschnitt zur Variablenkonfiguration ein Feld **Effective Date** (Stichtag), vorgelegt mit dem heutigen Datum. Wird eine Spalte im Special-Auswahlfeld als **contract_date** gekennzeichnet, berechnet glmnetUI eine Spalte **sale_age** — die ganzzahlige Anzahl der Tage zwischen dem Vertragsdatum des jeweiligen Verkaufs und dem Stichtag. Diese Spalte wird als Prädiktor ergänzt.

Ändert sich der Stichtag, wird **sale_age** automatisch neu berechnet.

4.3 Spezielle Spaltenkennzeichnungen

In den Modi Appraisal und Market erscheint für jeden Prädiktor in der Tabelle zur Variablenkonfiguration ein Auswahlfeld **Special**. Die vollständige Referenz der Special-Typen und ihrer Wirkung enthält Kapitel 6.

4.4 Überblick über die RCA-Anpassungen

Die Schaltfläche **Calculate RCA Adjustments & Download** (Seitenleistenabschnitt 8, in den Modi Appraisal und Market nach der Modellanpassung sichtbar) berechnet marktabgeleitete Anpassungen für jedes Vergleichsobjekt relativ zum Bewertungsobjekt. Der vollständige RCA-Ablauf wird in Kapitel 11 beschrieben.

Hinweis: Der dem Bewertungsobjekt zugewiesene CQA-Wert steuert, welcher Anteil der Residualverteilung dem Bewertungsobjekt zugerechnet wird. Ein Wert von 5.00 platziert das Bewertungsobjekt am Median der Vergleichsobjekte.

5 Modus Market Area Analysis

Wird **Market Area Analysis** als Purpose gewählt, stellt glmnetUI dieselben immobilien-spezifischen Funktionen bereit wie der Wertermittlungsmodus (Special-Spalten, Verkaufsalter, Koordinatenrundung), richtet sich dabei aber auf die Analyse einer Objektgruppe statt auf die Bewertung eines einzelnen Bewertungsobjekts aus.

5.1 Unterschiede zum Modus Appraisal

- **Die Behandlung der Bewertungsobjektzeile ist flexibel** — im Market-Modus wird Zeile 1 nicht automatisch als Bewertungsobjekt behandelt.
- **Kein Sales Comparison Grid** — der Schritt „Calculate RCA Adjustments & Download“ (Abschnitt 8) steht im Market-Modus zur Verfügung, der Schritt Sales Grid (Abschnitt 9) jedoch nur im Wertermittlungsmodus.

5.2 Wann der Market-Modus geeignet ist

Der Modus Market Area Analysis empfiehlt sich, wenn Sie:

- ein Regressionsmodell für eine Lage oder ein Marktgebiet aufbauen, um die Werttreiber zu verstehen
- analysieren, wie Variablen wie Wohnfläche, Alter, Grundstücksgröße und Lage die Verkaufspreise einer Objektgruppe beeinflussen
- Unterlagen für eine Marktkonjunkturanalyse oder eine Abgrenzung der Lage vorbereiten
- mit einem Datensatz arbeiten, für den die Special-Spalten-Funktionen (Verkaufsalter, Koordinatenrundung) benötigt werden, der RCA-Anpassungsablauf jedoch nicht

Hinweis: Der Market Mode eignet sich auch für allgemeine Immobilienregressionen, bei denen die Special-Spalten-Funktionen benötigt werden, der RCA-Anpassungsablauf jedoch nicht.

6 Variablenauswahl

In Abschnitt 4 der Seitenleiste — **Variable Configuration** — wird festgelegt, welche Spalten am Modell teilnehmen und wie sie behandelt werden.

6.1 Zielvariable

Das Auswahlfeld **Target (response) variable** am Kopf von Abschnitt 4 listet sämtliche Spalten des Datensatzes auf. Eine Spalte ist als Antwortvariable zu wählen (etwa der Verkaufspreis). Die Zielspalte wird automatisch aus der Prädiktorliste entfernt.

6.2 Die Prädiktortabelle

Unterhalb der Zielauswahl listet eine Tabelle alle verbleibenden Spalten mit den folgenden Feldern auf:

Spalte	Beschreibung
Variable	Spaltenname
Type	Auswahlfeld für den Datentyp: <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>Date</code> , <code>POSIXct</code>
Include	Kontrollkästchen — diese Spalte als Prädiktor in das Modell aufnehmen
Factor	Kontrollkästchen — die Variable als kategorialen Faktor behandeln (erzeugt Dummy-Spalten in der Modellmatrix)
Force	Kontrollkästchen — in das Modell erzwingen (Straffaktor = 0, wird vom Lasso nie entfernt)
Special	Auswahlfeld (nur Appraisal/Market) — siehe die Referenz der Special-Spalentypen weiter unten
Sign	Erwartetes Koeffizientenvorzeichen: positiv, negativ oder beliebig
NAs	Anzahl fehlender Werte

6.3 Erkennung und Überschreiben von Datentypen

glmnetUI erkennt die Datentypen beim Import automatisch. Numerische, ganzzahlige und Datumsspalten werden erkannt. Zeichenspalten, die wie Datumsangaben aussehen (gängige Datumsformatmuster), werden als `Date` eingestuft.

Jede Erkennung lässt sich über das Auswahlfeld **Type** überschreiben. Eine Typänderung wirkt sich darauf aus, wie die Spalte in der Modellmatrix kodiert wird.

Text-, TRUE/FALSE- und R-Faktorspalten werden automatisch als kategorial behandelt. Eine numerische Spalte wird nur dann zum Faktor, wenn ihr Kontrollkästchen **Factor** gesetzt wird — glmnetUI leitet aus dem Wertebereich einer numerischen Spalte niemals einen kategorialen Charakter ab, sodass diskrete numerische Prädiktoren wie eine Badanzahl stetig bleiben, sofern nichts anderes angegeben wird. Die Factor-Einstellung wird über localStorage gesichert.

6.4 Erwartete Vorzeichen und deren Durchsetzung

Für jeden Prädiktor gibt das Auswahlfeld **Sign** an, ob sein Koeffizient positiv, negativ oder in beliebiger Richtung erwartet wird. Nach der Anpassung gilt:

- Koeffizienten mit unerwartetem Vorzeichen werden **in der Coefficients-Tabelle markiert**

- Vorzeichenwarnungen erscheinen in der Registerkarte Summary

Ist in den glmnet-Parametern (Kapitel 7) das Kontrollkästchen **Enforce Sign Constraints** aktiviert, wird das Modell gezwungen, Koeffizienten mit den erwarteten Vorzeichen zu erzeugen. Umgesetzt wird dies über `upper.limits` und `lower.limits` in glmnet.

6.5 Referenz der Special-Spaltentypen

In den Modi Appraisal und Market bietet das Auswahlfeld **Special** die folgenden Optionen:

Gewichtung:

- **weight** — Spalte mit Beobachtungsgewichten (in allen Modi verfügbar; nur eine zulässig; Zeilen mit Gewicht = 0 werden von der Anpassung ausgeschlossen)

Datums- und Zeittypen:

- **contract_date** — löst die automatische Berechnung von **sale_age** aus dem Stichtag aus
- **listing_date** — dient als Rückfalloption zur Berechnung der Vermarktungsdauer
- **dom** — kennzeichnet die Spalte mit der Vermarktungsdauer
- **sale_age** — kennzeichnet eine bereits vorhandene Verkaufsalter-Spalte, sodass eine beliebige Spalte als Verkaufsalter dienen kann, statt dass glmnetUI es aus **contract_date** berechnet

Monetäre Typen:

- **concessions** — kennzeichnet Verkaufszugeständnisse (Verkäufergutschriften, Käuferanreize usw.)

Transaktionstypen:

- **sale_type** — kennzeichnet die Spalte mit der Verkaufs- bzw. Transaktionsart

Größen- und Lagetypen:

- **latitude** — Werte werden automatisch auf drei Nachkommastellen gerundet
- **longitude** — gleiche Rundungsbehandlung wie beim Breitengrad
- **area** — Kennung des Marktgebiets oder der Lage (z. B. `area_id`, `neighborhood`, `market_area`)
- **living_area** — ermöglicht die Berechnung von Residuen je Quadratfuß (`residual_sf` und `cqa_sf`)
- **lot_size** — Spalte für die Grundstücksgröße
- **site_dimensions** — mit der Grundstücksgröße gruppiert (z. B. „75x120“)

Alterstypen:

- **actual_age** — Spalte mit dem Objektalter
- **effective_age** — effektives Objektalter

Anzeigetypen:

- **display_only** — die Spalte wird in Excel-Exporte aufgenommen, aber vollständig von der Modellanpassung ausgeschlossen. Geeignet für Adressfelder, MLS-Nummern, Flurstückskennungen oder sonstige Referenzdaten. Mehrere Spalten können diese Kennzeichnung tragen.

Hinweis: Als *display_only* gekennzeichnete Spalten verbleiben im Datensatz und erscheinen im Excel-Export, werden jedoch vollständig aus der Prädiktortabelle ausgeschlossen. Diese Kennzeichnung eignet sich für ID-Spalten, Adressen oder sonstige Referenzfelder.

7 Parameterauswahl

Abschnitt 5 der Seitenleiste — **glmnet Call Parameters** — eröffnet den Zugriff auf sämtliche Konfigurationsoptionen des Elastic-Net-Modells. Zu jedem Parameter gehört ein blaues Hilfesymbol (?) mit einer erläuternden Kurzinfo.

7.1 Alpha (Mischungsparameter)

Der Parameter alpha steuert die Art der Regularisierung:

Alpha-Wert	Typ	Verhalten
0	Ridge	Schrumpft alle Koeffizienten proportional; setzt keinen exakt auf null
1	Lasso	Kann Koeffizienten exakt auf null setzen (Variablenselektion)
0 bis 1	Elastic Net	Mischung aus Ridge und Lasso

Zwei Verfahren zur Wahl von alpha stehen zur Verfügung:

- **Fixed** — einen einzelnen Alpha-Wert über den Schieberegler festlegen (Standard: 1.0)
- **Grid Search** — mehrere Alpha-Werte prüfen und denjenigen mit dem geringsten CV-Fehler wählen. Bereich und Anzahl der Werte sind konfigurierbar.

7.2 Wahl von Lambda

Lambda steuert die Stärke der Regularisierung:

Verfahren	Beschreibung
Kreuzvalidierung	Empfohlen. Prüft viele Lambda-Werte mittels k -facher Kreuzvalidierung.
Manuell	Erlaubt die Eingabe eines bestimmten Lambda-Werts, sofern es dafür einen Grund gibt.

Bei Verwendung der Kreuzvalidierung stehen zwei Lambda-Varianten zur Wahl:

- **lambda.1se** (Standard) — das größte Lambda innerhalb einer Standardabweichung des Minimums. Erzeugt einfachere, stärker regularisierte Modelle. Für die Wertermittlung empfohlen.
- **lambda.min** — das Lambda, das den Kreuzvalidierungsfehler minimiert. Erzeugt das am besten angepasste, möglicherweise aber komplexere Modell.

Der Parameter **Number of CV Folds** steuert, wie die Daten für die Kreuzvalidierung aufgeteilt werden (Standard: 10 Folds).

7.3 Family

Wählen Sie die Verteilungsfamilie für Ihre Antwortvariable:

Family	Anwendungsfall
gaussian	Stetige Antworten (z. B. Verkaufspreis). Am häufigsten.

Family	Anwendungsfall
binomial	Binäre Ergebnisse (z. B. verkauft/nicht verkauft).
poisson	Zählraten (z. B. Anzahl der Verkäufe).

7.4 Standardize

Ist dieses Kontrollkästchen gesetzt (Standard), werden alle Prädiktoren vor der Anpassung auf Mittelwert 0 und Standardabweichung 1 skaliert. Damit wirkt der Strafterm unabhängig von der ursprünglichen Skala auf alle Prädiktoren gleich. Die Koeffizienten werden auf der ursprünglichen Skala zurückgegeben. Diese Option sollte in der Regel aktiviert bleiben.

7.5 Vorzeichenrestriktionen

Ist **Enforce Sign Constraints** aktiviert, werden die Koeffizienten so beschränkt, dass sie den in der Variablen-tabelle (Kapitel 6) festgelegten erwarteten Vorzeichen entsprechen:

- Ein Vorzeichen „positive“ erzwingt den Koeffizienten ≥ 0
- Ein Vorzeichen „negative“ erzwingt den Koeffizienten ≤ 0
- Auf „either“ gesetzte Variablen bleiben unbeschränkt

Umgesetzt wird dies über die Parameter `upper.limits` und `lower.limits` in `glmnet()`.

7.6 Relaxed Lasso

Das gewöhnliche Lasso verwendet denselben Strafterm für die Variablenselektion UND die Koeffizientenschätzung, was wichtige Koeffizienten zu stark in Richtung null schrumpfen kann. Das **Relaxed Lasso** trennt diese Schritte:

1. Zunächst werden die Variablen mittels Lasso ausgewählt (der Strafterm bestimmt, welche Koeffizienten überleben)
2. Anschließend werden die verbliebenen Variablen mit geringerem oder ohne Strafterm neu geschätzt, was weniger verzerrte Schätzungen liefert

Der Schieberegler **Gamma** steuert den Grad der Relaxation:

Gamma	Wirkung
0	Vollständig relaxiert (OLS-Neuschätzung, gar keine Schrumpfung)
1	Keine Relaxation (wie das gewöhnliche glmnet)
Dazwischen	Mischung aus relaxierter und penalisierter Anpassung

Bei Verwendung der Kreuzvalidierung wird Gamma automatisch so gewählt, dass sich eine optimale Leistung ergibt.

7.7 Zufallsstartwert

Ein Textfeld **Random seed**, vorgelegt mit einer zufälligen ganzen Zahl, macht die Zuordnung der Kreuzvalidierungs-Folds reproduzierbar — vor der Anpassung wird `set.seed()` aufgerufen, und der verwendete Startwert wird in der Registerkarte Glmnet Output sowie in der Statusmeldung zur Anpassung angezeigt. Nach jeder Anpassung wird ein neuer Startwert erzeugt (auf Basis der Systemzeit, nicht des Zufallszahlengenerators von R). Die letzten fünf für die aktuelle Datei verwendeten Startwerte werden als anklickbare Verweise angezeigt, sodass sie sich leicht wieder aufrufen lassen; die Startwert-Historie wird beim Laden einer neuen Datei zurückgesetzt.

7.8 Interaktionsmatrix

Eine interaktive obere Dreiecksmatrix erlaubt es zu steuern, welche Variablenpaare interagieren dürfen. Jede Zelle enthält ein Kontrollkästchen — ist es gesetzt, wird der Interaktionsterm $x_1:x_2$ in die Modellmatrix aufgenommen.

- **Allow All** — alle Interaktionspaare aktivieren
- **Clear All** — alle Paare deaktivieren (bei einer neuen Datei ist standardmäßig keines gesetzt)
- **Klick auf einen Variablennamen** — schaltet alle Interaktionen dieser Variablen um

Die Interaktionsauswahl wird je Eingabedatei im localStorage des Browsers gesichert und beim erneuten Laden derselben Datei wiederhergestellt.

Ist mindestens eine Interaktion aktiviert, enthält der exportierte Bericht eine Matrix **Enabled Interactions** (Prädiktoren auf beiden Achsen, ein Häkchen dort, wo der Produktterm $x_1:x_2$ aufgenommen wurde), im Querformat dargestellt und der Matrix Allowed Interactions von earthUI nachgebildet. Zu beachten ist, dass das Lasso den Koeffizienten eines aktivierten Terms dennoch auf null geschrumpft haben kann.

***Hinweis:** Bei vielen Prädiktoren setzt das Lasso nicht benötigte Interaktionsterme automatisch auf null. Mit „Allow All“ zu beginnen und die Auswahl dem Modell zu überlassen, ist ein vertretbares Vorgehen.*

Warnung

Werden glmnet-Interaktionen aktiviert, entstehen Anpassungen, die vor Gericht oder in Prüfungssituationen schwer zu interpretieren und zu erläutern sein können. Interaktionen sind mit Bedacht einzusetzen, und die resultierenden Anpassungen sind auf Plausibilität zu prüfen.

7.9 Vom Haupteffekt ausschließen (nur Interaktion)

Ein **Rechtsklick** auf einen Variablennamen in der Interaktionsmatrix schließt die Variable als Haupteffekt aus dem Modell aus. Nach dem Variablennamen erscheint eine fette „1“ als Anzeige. Ist eine Variable ausgeschlossen, gilt:

- Die Variable wird aus den Haupttermen der Formel **entfernt**
- Die Variable **kann weiterhin** in Interaktionstermen ($x_1:x_2$) auftreten
- Ein erneuter Rechtsklick hebt den Ausschluss auf
- Der Zustand wird je Eingabedatei über localStorage gesichert

Ein typischer Anwendungsfall: Sollen Zeitanpassungen mit der Objektgröße skalieren, wird **sale_age** vom Haupteffekt ausgeschlossen und nur in einer Interaktion mit **living_area** zugelassen. So entsteht ein Interaktionsterm **sale_age:living_area** anstelle eines konstanten **sale_age**-Koeffizienten — die zeitliche Wertsteigerungsrate variiert dann mit der Hausgröße.

7.10 Erweiterte Parameter

Ein ausklappbarer Abschnitt „Advanced“ am Ende der glmnet Call Parameters macht weitere Einstellungen zugänglich. Sie verwenden sinnvolle Standardwerte, sind aber sichtbar, damit sämtliche Modelleinstellungen für Gericht oder Prüfung dokumentiert werden können.

- **Lambda Min Ratio** — Verhältnis des kleinsten zum größten Lambda im Regularisierungspfad. Standard: 0.00001. Bei kleinen Datensätzen (<50 Beobachtungen) kann ein größerer Wert (etwa 0.01) die Leistung verbessern.

- **Number of Lambda Values** — Anzahl der Lambda-Werte im Pfad. Standard: 100.
- **CV Loss Metric** — Verlustfunktion für die Kreuzvalidierung: MSE (Standard für gaussian), MAE (robust gegenüber Ausreißern) oder Devianz.
- **Convergence Threshold** — Konvergenzschwelle des Koordinatenabstiegs. Standard: $1e-07$.
- **Max Iterations** — maximale Anzahl der Koordinatenabstiegs-Iterationen. Standard: 100.000. Bei Konvergenzwarnungen erhöhen.
- **Fit Intercept** — einen Achsenabschnitt (konstanten Term) aufnehmen. Standard: aktiviert. Der Achsenabschnitt ist der „Basiswert“ im RCA der Wertermittlung.

7.11 Standardwerte der Einstellungen

Drei Optionen steuern, wie die Parameter beim Laden einer Datei initialisiert werden:

- **Use last settings for input file** — genau die zuletzt mit dieser Datei verwendeten Einstellungen wiederherstellen
- **Use default settings** — die gespeicherten globalen Standardwerte anwenden
- **glmnet defaults** — auf die Werkseinstellungen zurücksetzen ($\alpha = 1$, CV mit λ_{1se} , 10 Folds, gaussian, standardize aktiviert)

Ein Klick auf **Save current as default** speichert sämtliche aktuellen Einstellungen als globalen Standard für künftige Dateien.

8 Anpassung des Modells

8.1 Die Schaltfläche zur Modellanpassung

Abschnitt 6 der Seitenleiste enthält die Schaltfläche **Fit Glmnet Model**. Ein Klick darauf rechnet das Modell mit der aktuellen Konfiguration.

8.2 Was bei der Anpassung geschieht

Beim Klick auf Fit führt glmnetUI Folgendes aus:

1. **Daten vorbereiten** — Zeilen mit NA-Werten entfernen, Gewichte anwenden (sofern gekennzeichnet), die Bewertungsobjektzeile ausschließen (im Wertermittlungsmodus) und Faktorvariablen als Dummy-Spalten kodieren.
2. **Modellmatrix aufbauen** — die Designmatrix einschließlich aller in der Interaktionsmatrix zugelassenen Interaktionsterme erzeugen.
3. **Restriktionen anwenden** — ist die Vorzeichendurchsetzung aktiviert, werden `upper.limits` und `lower.limits` für die Koeffizienten gesetzt.
4. **Kreuzvalidierung durchführen** — ist CV gewählt, wird `cv.glmnet()` aufgerufen, um das optimale Lambda zu bestimmen. Ist das Relaxed Lasso aktiviert, erfolgt der Aufruf als `cv.glmnet(..., relax = TRUE)`.
5. **Endgültiges Modell anpassen** — das angepasste Modell, die Koeffizienten, die Prognosen und die Residuen werden gespeichert.
6. **Alle Ergebnis-Registerkarten aktualisieren** — Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA und Diagnostics werden befüllt.

Eine Statusmeldung unterhalb der Fit-Schaltfläche zeigt die Anzahl der Beobachtungen, die ausgeschlossenen Zeilen und das gewählte Lambda an.

9 Ergebnis-Registerkarten

9.1 Data Preview

Zeigt die importierten Daten als interaktive DataTable. In den Modi Appraisal und Market ist die Vorschau in zwei Tabellen aufgeteilt: **Subject Property** (Zeile 1) und **Comparable Sales** (Zeilen 2 ff.).

Jede Zelle wird auf eine Zeile begrenzt und auf die Spaltenbreite gekürzt, sodass breite Freitextfelder (etwa Objektbemerkungen) eine Tabellenzeile nicht über die Seite ziehen. Ein **Doppelklick auf eine beliebige Zelle** öffnet ein Fenster mit ihrem vollständigen, ungekürzten Inhalt.

9.2 Equation

Zeigt die angepasste Modellgleichung, über MathJax in LaTeX gerendert. Dargestellt werden alle von null verschiedenen Koeffizienten in korrekter mathematischer Formatierung:

- Positive Koeffizienten mit $+$ -Zeichen
- Negative Koeffizienten mit $-$ -Zeichen
- Interaktionsterme als $x_1 \times x_2$
- Zahlen mit Tausendertrennzeichen und angemessener Anzahl von Nachkommastellen

9.3 Correlation

Eine Heatmap-Matrix der Pearson-Korrelationen zwischen allen numerischen Prädiktoren und der Antwortvariablen. Unmittelbar nach dem Datenimport verfügbar — eine Modellanpassung ist nicht erforderlich.

- Farbverlauf: blau (-1) $\rightarrow$ weiß (0) $\rightarrow$ rot ($+1$)
- Korrelationswerte in jeder Zelle ausgewiesen
- Schrift- und Achsengröße passen sich der Zahl der Variablen an
- Nützlich, um Multikollinearität vor der Anpassung zu erkennen

9.4 Summary

Kennzahlen zur Anpassungsgüte, als Karten dargestellt:

- **R^2** — Anteil der erklärten Varianz
- **Adj R^2** — um die Zahl der Prädiktoren bereinigt
- **GR^2** — generalisiertes R^2 aus dem Devianzverhältnis von glmnet beim gewählten Lambda
- **CV R^2** — R^2 aus der Kreuzvalidierung (wenn CV verwendet wird)
- **RMSE** — Wurzel des mittleren quadratischen Fehlers
- **MAE** — mittlerer absoluter Fehler

In den Modi Appraisal und Market zeigen zusätzliche Karten:

- **Median Ratio** — Median von (prognostiziert / beobachtet)
- **COD** — Coefficient of Dispersion (Streuungskoeffizient; niedriger ist besser)
- **PRD** — Price-Related Differential

Eine **Overfitting-Warnung** erscheint, wenn das Trainings- R^2 das CV- R^2 um mehr als 0,1 übersteigt.

Unterhalb der Karten dupliziert eine Koeffiziententabelle mit Vorzeichenwarnungen der Bequemlichkeit halber die Registerkarte Coefficients.

9.5 Coefficients

Eine Tabelle aller von null verschiedenen Koeffizienten mit:

- Variablenname
- Koeffizientenwert
- Erwartetes Vorzeichen (aus der Variablenkonfiguration)
- Vorzeichenwarnung, wenn die Richtung des Koeffizienten nicht der Erwartung entspricht

9.6 Variable Importance

Standardisierte Koeffizientengröße: $|\beta| \times \text{sd}(x)$ je Prädiktor, bei Faktorvariablen über die Dummy-Spalten aggregiert. Zwei Darstellungsformen:

- **Interaktives plotly** (wenn das Paket `plotly` installiert ist): waagerechtes Balkendiagramm mit Tooltips, die die genaue Wichtigkeit, den Koeffizienten und den relativen Anteil zeigen.
- **Statisches ggplot2** (Rückfalloption): waagerechtes Balkendiagramm mit Achsenbeschriftungen samt Tausendertrennzeichen.

Unterhalb der Grafik zeigt eine DataTable Variable, Importance, Coefficient und Relative %.

9.7 Contributions

Grafiken der partiellen Effekte je Variable, die den Beitrag jedes Prädiktors zu den Modellprognosen zeigen. Die Variable wird über ein Auswahlfeld gewählt.

Für numerische Prädiktoren: ein Streudiagramm der x-Werte der Variablen gegen ihren Beitrag ($\beta \times x$), mit einem roten Liniensegment und einer Steigungsbeschriftung. Die Steigungsbeschriftung gibt den marginalen Effekt je Einheit an (etwa +1,234.56/unit), mit adaptiven Einheiten für Variablen mit geringer Spannweite wie geografische Breite und Länge (etwa +0.12/0.001). Der Untertitel zeigt den Wert des Achsenabschnitts (Basiswert).

Für Faktorvariablen: ein Boxplot der Verteilung der Beitragswerte je Faktorstufe.

Für Interaktionen zweier Variablen: drei übereinander angeordnete Darstellungen — eine 3D-Flächengrafik des gemeinsamen Beitrags (interaktiv mit `plotly`, sofern verfügbar, sonst statisch über `persp()`), ein Streudiagramm mit nach Beitrag eingefärbten Punkten sowie eine Heatmap des mittleren Beitrags über beide Wertebereiche.

9.8 ANOVA

Tabelle der Varianzzerlegung, die für jeden Prädiktor Folgendes zeigt:

- **SS** — Quadratsumme: $\sum(\text{contribution}_j - \overline{\text{contribution}}_j)^2$
- **% of Model SS** — Anteil an der gesamten Quadratsumme des Modells
- **Coefficient** — der dominierende Koeffizient dieses Prädiktors

Enthält eine Zeile für den Achsenabschnitt sowie eine Zeile TOTAL MODEL. Interaktionsterme (erkennbar an „:“) werden als eigene Gruppen ausgewiesen.

9.9 Diagnostics

Vier Diagnosegrafiken mit großen Schriften (Basisgröße 16 pt), 15 Achsenteilstrichen und Beschriftungen mit Tausendertrennzeichen (keine wissenschaftliche Notation):

1. **Koeffizientenpfad** — zeigt, wie sich alle Koeffizienten mit wachsendem $\log(\lambda)$ verändern. Die Linien sind nach Variablen eingefärbt. Hilft dabei, den Regularisierungspfad und die Variablenselektion zu veranschaulichen.

2. **CV-Fehler** — Kreuzvalidierungsfehler als Funktion von $\log(\lambda)$. Die Punkte zeigen den mittleren CV-Fehler, die Fehlerbalken ± 1 Standardfehler. Gestrichelte senkrechte Linien markieren $\lambda_{\min}$ (blau) und λ_{1se} (grün).
3. **Beobachtet gegen prognostiziert** — Streudiagramm der beobachteten gegen die prognostizierten Werte. Die Punkte sollten sich um die gestrichelte 45-Grad-Referenzlinie gruppieren. Eine breitere Streuung deutet auf einen größeren Prognosefehler hin.
4. **Residuen gegen angepasste Werte** — Streudiagramm der Residuen gegen die angepassten Werte. Es sollte eine zufällige Streuung um die (gestrichelte) Nulllinie zeigen. Muster deuten auf eine Fehlspezifikation des Modells hin (etwa Nichtlinearität oder Heteroskedastizität).

Jede Grafik verfügt über eine Schaltfläche **Download PNG** zum Speichern mit 150 DPI.

9.10 Glmnet Output

Rohausgabe des Modells: der verwendete Zufallsstartwert, der Modellausdruck, das gewählte Lambda, $\lambda_{\min}/\lambda_{1se}$, Gamma (wenn das Relaxed Lasso aktiviert ist) sowie der vollständige Koeffizientenvektor.

9.11 Report

Berichtsfelder konfigurieren (Name des Sachverständigen, Objektadresse, Berichtsdatum, Aktennummer) und unmittelbar aus der Registerkarte heraus nach Word (.docx) oder PDF exportieren. Der vollständige, auf Quarto gestützte Berichtsablauf (HTML, Word oder PDF) ist in Kapitel 13 beschrieben.

10 Herunterladen von Daten

Nach der Modellanpassung exportiert die Schaltfläche **Download Output (Excel)** (Seitenleistenabschnitt 7, in allen Purpose-Modi verfügbar) eine Excel-Datei mit Prognosen und Diagnosegrößen in den Ausgabeordner des Projekts. Die Abschnittsüberschrift lautet in den Modi Appraisal und Market „Download Estimated Sale Prices & Residuals“ und im Modus General „Download Estimated Target Variable(s) & Residuals“.

10.1 Ausgabespalten

Spalte	Beschreibung
<code>est_<target></code>	Modellprognose (z. B. <code>est_sale_price</code>)
<code>residual</code>	Beobachtet – prognostiziert
<code>cqa</code>	Condition-Quality-Appeal-Score (Skala 0–10)
<code>residual_sf</code>	Residuum / Wohnfläche (wenn <code>living_area</code> gekennzeichnet ist)
<code>cqa_sf</code>	CQA, berechnet aus der Rangfolge über <code>residual_sf</code>
<code><var>_contribution</code>	Beitrag der jeweiligen Variablen zur Prognose
<code>basis</code>	Wert des Achsenabschnitts (für alle Zeilen gleich)

Nach erfolgreichem Abschluss erscheint auf der Download-Schaltfläche ein weißes Häkchen.

10.2 CQA-Scores

Der CQA-Wert ordnet das Residuum jeder Zeile im Vergleich zu allen übrigen auf einer Skala von 0 bis 10 ein:

- **Hoher CQA (~9–10):** deutlich über der Prognose verkauft — vermutlich überdurchschnittlicher Zustand bzw. überdurchschnittliche Qualität oder Attraktivität
- **Niedriger CQA (~0–1):** deutlich unter der Prognose verkauft — vermutlich unterdurchschnittlicher Zustand oder ein Notverkauf
- **CQA ~5:** nahe dem Median

In den Modi Appraisal und Market werden die Zeilen absteigend nach `residual_sf` sortiert, sofern eine Spalte `living_area` gekennzeichnet ist.

11 RCA-Berechnungen und Herunterladen

Der RCA-Ablauf (Residual Constraint Approach) steht in den Modi **Appraisal** und **Market Area Analysis** nach der Modellanpassung zur Verfügung. Das darauf aufbauende Sales Comparison Grid (Kapitel 12) ist dem Wertermittlungsmodus vorbehalten.

11.1 Öffnen des RCA-Dialogs

Klicken Sie auf die Schaltfläche **Calculate RCA Adjustments & Download** in Seitenleistenabschnitt 8. Zu wählen sind:

- Score-Typ **CQA** oder **CQA per SF**. Bei „CQA per SF“ wird das Residuum des Bewertungsobjekts mit der Wohnfläche skaliert.
- Eingabe des CQA-Werts für das Bewertungsobjekt (0.00–9.99, Standard 5.00)

11.2 Interpolation des CQA-Werts

1. Die CQA-Werte und Residuen der Vergleichsobjekte werden sortiert
2. Eine lineare Interpolation bildet den gewählten CQA-Wert auf ein Residuum ab
3. Wert des Bewertungsobjekts = Modellprognose + interpoliertes Residuum

11.3 Ausgabespalten

Spalte	Beschreibung
subject_value	Modellprognose + interpoliertes Residuum
<var>_adjustment	Beitrag des Bewertungsobjekts – Beitrag des Vergleichsobjekts
residual_adjustment	Residuum des Bewertungsobjekts – Residuum des Vergleichsobjekts
net_adjustments	Summe aller Anpassungen
gross_adjustments	Summe der absoluten Anpassungen
adjusted_sale_price	Verkaufspreis des Vergleichsobjekts + Nettoanpassungen

Nach erfolgreicher Berechnung erscheint auf der Schaltfläche ein weißes Häkchen.

12 Sales Comparison Grid

Das Sales Comparison Grid (Vergleichsraster) steht ausschließlich im Modus **Appraisal** zur Verfügung, nachdem die RCA-Anpassungen (Schritt 8) berechnet wurden. Es erzeugt eine formatierte Excel-Arbeitsmappe, die sich zur Aufnahme in Wertermittlungsberichte eignet.

12.1 Dialog zur Auswahl der Vergleichsobjekte

Ein Klick auf die Schaltfläche **Generate Sales Grid & Download** in Seitenleistenabschnitt 9 öffnet einen modalen Dialog, der alle Vergleichsverkäufe aus der RCA-Ausgabe auflistet. Die Vergleichsobjekte sind in zwei Gruppen aufgeteilt:

- **Recommended comps** (empfohlene Vergleichsobjekte, vorausgewählt): Bruttoanpassung < 25 % des Verkaufspreises, sortiert nach Verkaufsalter (jüngste zuerst).
- **Other comps** (übrige Vergleichsobjekte, standardmäßig nicht ausgewählt): die verbleibenden Vergleichsobjekte, sortiert nach Verkaufsalter.

Bis zu 30 Vergleichsobjekte auswählen und anschließend auf **Generate Sales Grid** klicken, um die Arbeitsmappe zu erzeugen.

12.2 Aufbau der Arbeitsmappe

Die erzeugte Arbeitsmappe enthält bis zu 10 Blätter mit jeweils drei nebeneinander angeordneten Vergleichsobjekten:

- **Kopfzeilen:** Adressen von Bewertungsobjekt und Vergleichsobjekten, Verkaufspreise, Verkaufsdaten und Vermarktungsdauer
- **Gruppierte Zeilen:** Lage (Breitengrad, Längengrad, Gebiet), Grundstücksgröße (Grundstücksfläche, Grundstücksmaße) und Alter (tatsächliches Alter, effektives Alter) mit Gruppenüberschriften und Unterpunkten
- **Zeilen der Modellvariablen:** eine Zeile je von null verschiedenem Modellkoeffizienten, mit dem Beitrag des Bewertungsobjekts, den Beiträgen der Vergleichsobjekte und den prozentualen Anpassungen
- **Zeilen für Restmerkmale:** leere, entsperrte Zellen für Eingaben des Sachverständigen zu Merkmalen, die das Modell nicht erfasst (etwa Zustand, Qualität, Aussicht)
- **Zusammenfassungszeilen:** Nettoanpassung in %, Bruttoanpassung in % und angepasster Verkaufspreis mit aktiven Excel-Formeln

12.3 Blattschutz und Restmerkmalzellen

Jedes Blatt ist mit einem Kennwort gegen versehentliche Änderungen geschützt, die Wertzellen für die Restmerkmale sind jedoch ausdrücklich entsperrt. So können Sachverständige Werte für Merkmale eintragen, die nicht im Modell enthalten sind, während die formelgestützten Anpassungsberechnungen erhalten bleiben.

13 Herunterladen von Berichten

13.1 Der zweistufige Quarto-Ablauf

Der Berichtsexport aus der Seitenleiste ist ein zweistufiger Quarto-Ablauf:

1. **Generate Quarto Report** (Seitenleistenabschnitt 10) — schreibt ein in sich geschlossenes Quarto-Paket (die `.qmd`-Quelle nebst Grafiken, `report_data.rds` und `reference.docx`) in den glmnet-Ausgabeordner des Projekts.
2. **Convert Quarto Report** (Seitenleistenabschnitt 11) — rendert eine `.qmd`-Datei in die gewählten Ausgabeformate. Der Pfad ist mit dem zuletzt erzeugten Paket vorbelegt, es lässt sich jedoch zu jeder beliebigen `.qmd`-Datei navigieren.

13.2 Berichtsformate

Im Konvertierungsschritt stehen drei Formate zur Verfügung:

- **HTML** — in sich geschlossenes HTML-Dokument. Funktioniert auf allen Plattformen ohne zusätzliche Software außer pandoc. Empfohlener Standard.
- **Word (.docx)** — formatiertes Dokument. Geeignet zur Bearbeitung und Weitergabe. Funktioniert auf allen Plattformen.
- **PDF** — über Quarto mit LaTeX erzeugt. Das Papierformat richtet sich nach der Locale-Einstellung (Letter oder A4). **Erfordert eine LaTeX-Installation** (siehe den Anhang zu den Systemanforderungen). Wird kein LaTeX erkannt, wird die PDF-Option automatisch aus den Format-Kontrollkästchen ausgeblendet.

Berichte werden nach Möglichkeit mit Quarto erzeugt, ersatzweise mit rmarkdown. Die Registerkarte Report bietet darüber hinaus Schaltflächen für den direkten Export nach Word und PDF.

13.3 Berichtsinhalte

Die Berichte enthalten sämtliche Inhalte der Registerkarten:

- Beschreibung des Datensatzes (Dateiname, Beobachtungen, Antwortvariable, Prädiktoren, Purpose)
- Modellspezifikation (alpha, lambda, family, standardize, relaxed, von null verschiedene Koeffizienten)
- Ergebnisübersicht (R^2 , adjustiertes R^2 , generalisiertes R^2 , CV- R^2 , RMSE, MAE)
- Modellgleichung (in LaTeX gerendert)
- Koeffiziententabelle
- Variablenwichtigkeit (Balkendiagramm und Tabelle)
- Grafiken der Beiträge je Variable, entsprechend der Registerkarte Contributions:
 - **Numerische Prädiktoren:** Streudiagramm mit linearer Anpassungsgeraden
 - **Faktorvariablen:** Boxplot je Stufe
 - **Interaktionen:** nach Beitrag eingefärbtes Streudiagramm, Heatmap des mittleren Beitrags und statische Momentaufnahme der 3D-Fläche (persp)
- Matrix Enabled Interactions (wenn mindestens eine Interaktion aktiviert ist) — Prädiktoren auf beiden Achsen, mit einem Häkchen dort, wo der Produktterm $x_1:x_2$ aufgenommen wurde, im Querformat dargestellt
- Heatmap der Korrelationsmatrix
- Tabelle der ANOVA-Zerlegung
- Diagnosegrafiken:

- Koeffizientenpfad
 - Kreuzvalidierungsfehler
 - Beobachtet gegen prognostiziert
 - Residuen gegen angepasste Werte
 - Normal-Q-Q-Diagramm
- Modellausgabe

Sämtliche Achsenbeschriftungen und Farblegenden verwenden Zahlen mit Tausendertrennzeichen (keine wissenschaftliche Notation).

Nach erfolgreicher Erzeugung erscheint auf der Berichtsschaltfläche ein weißes Häkchen; während der Bericht gerendert wird, läuft eine Zeitanzeige mit.

14 Vergleich mit earthUI

Die earth-Anwendung und ihr glmnet Post Processing Module bilden ein gemeinsames Werkzeugpaket für die Regressionsmodellierung. Sie teilen Datenformat, Special-Spaltentypen, RCA-Ablauf und Demodatensätze, verwenden jedoch unterschiedliche Modellierungsmaschinen.

14.1 Wesentliche Unterschiede

Merkmal	glmnetUI	earthUI
Verfahren	Elastic Net (glmnet)	MARS/Earth (earth)
Variablenselektion	Automatisch über den Lasso-Strafterm	Automatisch über Vorwärts-/Rückwärts-Pruning
Koeffizientenbeziehungen	Nur lineare Effekte	Stückweise linear (Hinge-Funktionen)
Interaktionen	Paarweise über die Interaktionsmatrix	Bis Grad 3
Koeffizientenvorzeichen	Über Restriktionen durchsetzbar	Nicht unmittelbar durchsetzbar
Relaxierte Anpassung	Relaxed Lasso verfügbar	Nicht anwendbar
Alpha-Optimierung	Gittersuche verfügbar	Nicht anwendbar (kein Mischungsparameter)
Sales Comparison Grid	Verfügbar (Schritt 9)	Verfügbar (Schritt 8)
g-Funktionsgrafiken	Nicht anwendbar	Darstellung gruppierteter Terme
Berichtsformate	HTML, Word, PDF	HTML, Word, PDF

14.2 Gemeinsame Funktionen

Beide Werkzeuge bieten:

- CSV- und Excel-Import mit Umwandlung der Spaltennamen in snake_case
- Locale-Unterstützung für 31 Länder mit CSV-Trennzeichen, Dezimaltrennzeichen und Papierformat
- Spezielle Spaltenkennzeichnungen (contract_date, living_area, area, latitude, longitude usw.)
- Automatische Berechnung von sale_age aus dem Stichtag
- Umschalter zwischen dunklem und hellem Design
- Persistenz der Einstellungen über localStorage
- Korrelations-Heatmap, Variablenwichtigkeit, Beiträge, ANOVA, Diagnostik
- CQA-Bewertung und RCA-Anpassungsablauf
- Sales Comparison Grid mit Auswahl der Vergleichsobjekte und formatierter Excel-Ausgabe
- Kompatibilität mit denselben Demodatensätzen (Appraisal_1.csv)

14.3 Wann welches Werkzeug

- **glmnetUI verwenden**, wenn ein einfaches, interpretierbares lineares Modell mit automatischer Variablenselektion gewünscht ist. Das Elastic Net eignet sich besonders, wenn viele Prädiktoren vorliegen und das Modell die wichtigsten auswählen soll.
- **earthUI verwenden**, wenn nichtlineare Zusammenhänge zu erwarten sind (etwa wenn sich die Wirkung der Wohnfläche auf den Preis je nach Größe ändert) oder Hinge-Funktionen benötigt werden, um Schwelleneffekte abzubilden. Earth-Modelle können

zudem Interaktionen höherer Ordnung (Grad 2 und 3) erfassen. earthUI bietet darüber hinaus eine experimentelle, auf Prolog/DCG gestützte Funktion zur Extraktion von Bemerkungen, für die es in glmnetUI keine Entsprechung gibt.

- **Beide verwenden**, um lineare und nichtlineare Modelle auf denselben Daten zu vergleichen. Ergeben beide ähnliche R^2 -Werte, ist das einfachere glmnet-Modell aus Gründen der Interpretierbarkeit möglicherweise vorzuziehen.

15 Demodatensatz: Appraisal_1.csv

15.1 Beschreibung

Der Demodatensatz wird mit glmnetUI ausgeliefert (und mit earthUI geteilt). Er lässt sich wie folgt auffinden:

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "glmnetUI")
```

Die Datei enthält 1.502 Wohnimmobilienverkäufe (zuzüglich eines Bewertungsobjekts in Zeile 1) aus einem simulierten MLS-Export. Die Daten bilden Verkäufe von Einfamilienhäusern in einem mehrgebietigen Markt mit einer Bandbreite von Objektgrößen, Baualtern und Lagen ab.

Es handelt sich nicht um reale Daten, sie beruhen jedoch auf einer realistischen Wohnlage in Nordkalifornien. Sämtliche identifizierenden Angaben wurden verändert oder entfernt.

15.2 Spalten

Spalte	Typ	Special-Typ	Beschreibung
weight	numeric	weight	Beobachtungsgewicht (0 = von der Anpassung ausschließen)
id	numeric	display_only	Interne Datensatzkennung
property_id	numeric	display_only	MLS-Objektkennung
listing_id	character	display_only	MLS-Angebotsnummer
parcel_number	character	display_only	Flurstücksnummer der Katasterbehörde (APN)
street_address	character	display_only	Objektadresse
city_name	character	display_only	Stadt
postal_code	character	display_only	Postleitzahl
county_name	character	display_only	County
contract_date	Date	contract_date	Vertragsdatum des Verkaufs (berechnet sale_age)
sale_age	numeric	—	Tage vom Vertragsdatum bis zum Stichtag
sale_price	numeric	(Ziel)	Verkaufspreis — Antwortvariable
living_sqft	numeric	living_area	Bruttowohnfläche in Quadratfuß
beds_total	integer	—	Anzahl der Schlafzimmer
baths_total	numeric	—	Gesamtzahl der Bäder (z. B. 2,5 = 2 Vollbäder + 1 Gästebad)
lot_size	numeric	lot_size	Grundstücksfläche in Quadratfuß
area_id	integer	area	MLS-Gebietskennung
age	numeric	actual_age	Objektalter in Jahren
latitude	numeric	latitude	Breitengrad (für das Modell auf 3 Nachkommastellen gerundet)
longitude	numeric	longitude	Längengrad (für das Modell auf 3 Nachkommastellen gerundet)
garage_spaces	integer	—	Anzahl der Garagenstellplätze
days_on_market	integer	dom	Vermarktungsdauer in Tagen
listing_date	Date	listing_date	Datum des Angebots
sale_concessions	numeric	concessions	Verkäuferzugeständnisse

15.3 Vorgeschlagener Schnelleinstieg

1. Die glmnet-Routine starten: `earthUI::launch_glmnet()`
2. Ein neues Projekt (Abschnitt 1) mit dem Purpose **For Appraisal** anlegen
3. `Appraisal_1.csv` in den Eingabeordner des Projekts legen und in Abschnitt 2 auswählen (bei Bedarf auf **Refresh** klicken)
4. `sale_price` als Zielvariable wählen
5. Die Special-Typen gemäß der obigen Tabelle zuweisen
6. Folgende Prädiktoren aufnehmen: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (als Faktor), `age`, `latitude`, `longitude`, `garage_spaces`
7. $\alpha = 1$ (Lasso) beibehalten, CV mit `lambda.1se`
8. Auf **Fit Glmnet Model** klicken
9. Die Registerkarten Summary, Coefficients und Diagnostics prüfen
10. Die Ausgabe herunterladen (Schritt 7) und die CQA-Rangfolge prüfen
11. RCA-Anpassungen (Schritt 8) mit einem CQA-Wert von ~5.00 berechnen
12. Ein Sales Comparison Grid (Schritt 9) mit den empfohlenen Vergleichsobjekten erzeugen

16 Systemanforderungen und Fehlerbehebung

16.1 Unterstützte Plattformen

glmnetUI läuft unter **macOS**, **Windows** und **Linux**. RStudio Desktop (2023.06+) wird dringend empfohlen — es bringt **pandoc** mit, das für die Berichte benötigt wird.

16.2 Optionale Abhängigkeiten

- **LaTeX** (TinyTeX, MiKTeX oder MacTeX): nur für die Erzeugung von PDF-Berichten. Wird keine Installation erkannt, wird die PDF-Option automatisch ausgeblendet. Installation mit: `tinytex::install_tinytex()`
- **sysfonts + showtext**: Schriftart Roboto Condensed für Grafiken. Ist sie nicht verfügbar oder besteht keine Internetverbindung, wird automatisch auf die serifenlose Systemschrift zurückgegriffen.
- **earth** ($\geq 5.3.0$): nur für die Importstrecke aus earthUI erforderlich.

16.3 Plattformhinweise

macOS: die wenigsten Probleme. Für PDF TinyTeX installieren: `tinytex::install_tinytex()`

Windows: funktioniert gut mit RStudio. Auf Firmenrechnern mit eingeschränkten Rechten kann es Beschränkungen für das temporäre Verzeichnis geben — die Anwendung gibt dann eine Warnung in der Konsole aus, läuft jedoch ohne Persistenz der Einstellungen weiter.

Linux: möglicherweise sind Systembibliotheken erforderlich: `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

16.4 Verhalten bei fehlenden Komponenten

Fehlende Komponente	Verhalten
Kein LaTeX	PDF-Option ausgeblendet. HTML und Word bleiben verfügbar.
Kein Internet / Schriften nicht ladbar	Serifenlose Systemschrift wird verwendet. Konsolenmeldung wird protokolliert.
Kein RSQLite / schreibgeschütztes Dateisystem	Einstellungen bleiben nicht erhalten. Die Anwendung läuft normal.
Temporäres Verzeichnis nicht beschreibbar	Die Berichtserzeugung schlägt mit einer klaren Fehlermeldung fehl. <code>TMPDIR</code> auf einen beschreibbaren Ort setzen.

16.5 Fehlerbehebung

„**PDF option not available**“: In R `tinytex::install_tinytex()` ausführen und die Anwendung neu starten.

„**Settings will not persist**“: Die Berechtigungen für das Benutzerdatenverzeichnis prüfen (macOS: `~/Library/Application Support/R/glmnetUI/`, Linux: `~/.local/share/R/glmnetUI/`, Windows: `%APPDATA%/R/data/glmnetUI/`).

Port 7879 bereits belegt: `lsof -ti:7879 | xargs kill` (macOS/Linux) oder `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell) ausführen.

Literatur

- Friedman, Jerome, Trevor Hastie und Robert Tibshirani (2010). „Regularization Paths for Generalized Linear Models via Coordinate Descent“. In: *Journal of Statistical Software* 33.1, S. 1–22. DOI: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01).
- Simon, Noah u. a. (2011). „Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent“. In: *Journal of Statistical Software* 39.5, S. 1–13. DOI: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).
- Tay, J. Kenneth, Balasubramanian Narasimhan und Trevor Hastie (2023). „Elastic Net Regularization Paths for All Generalized Linear Models“. In: *Journal of Statistical Software* 106.1, S. 1–31. DOI: [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).