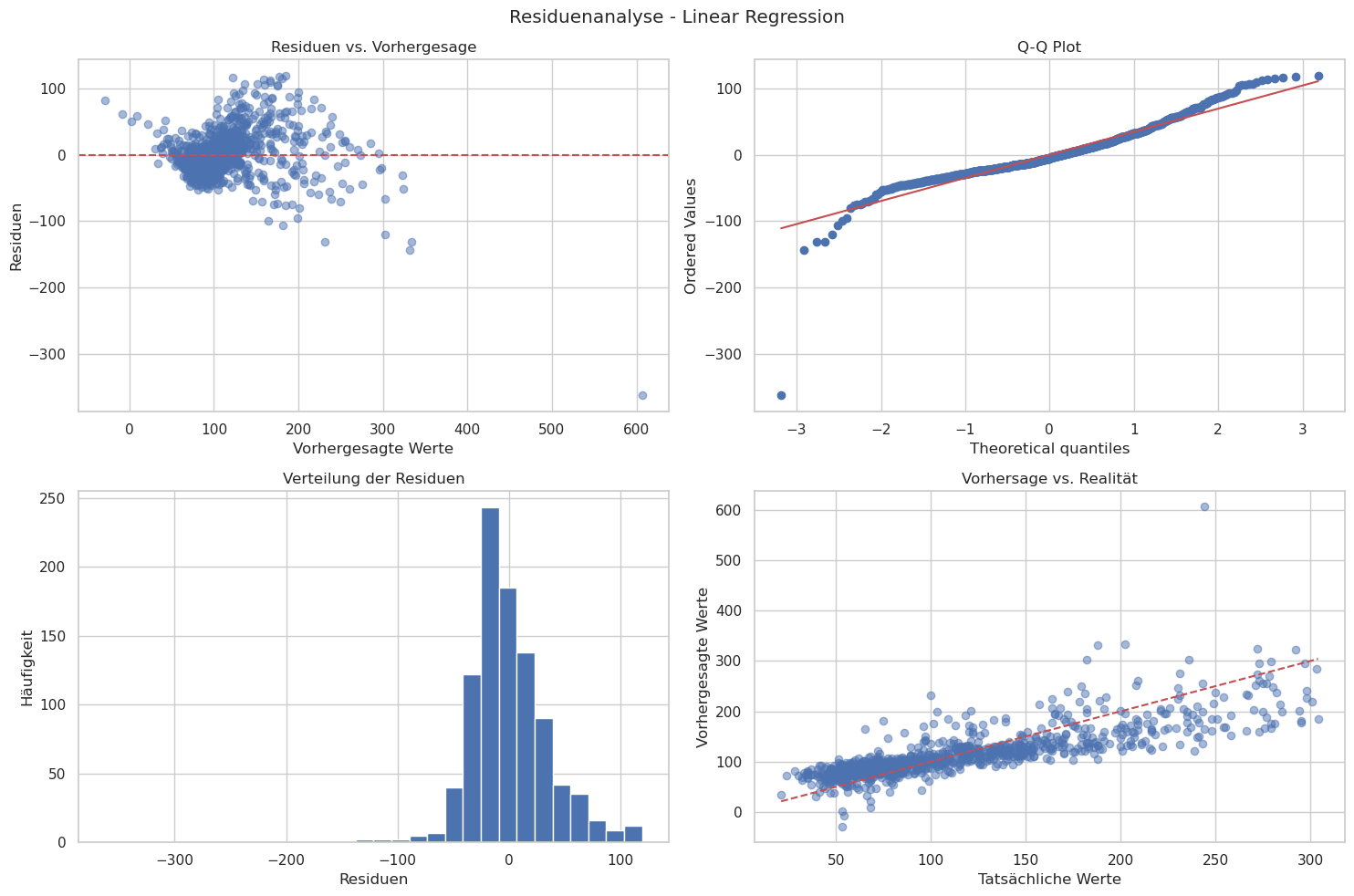
|  |  |  |  |
| --- | --- | --- | --- |
| Analyse Werte inklusive Kategorie „Live“ | Analyseplots | Zahlen | Auswertung KI |
| Klickrate |  | Analyse für Zielvariable: Klickrate der Impressionen (%)  ==================================================  Deskriptive Statistik für Zielvariable:  count 1012.000000  mean 4.088014  std 2.279179  min 1.000000  25% 2.320000  50% 3.625000  75% 5.415000  max 21.950000  Name: Klickrate der Impressionen (%), dtype: float64  Datentypen in der Analyse:  Klickrate der Impressionen (%) float64  Gestaltung\_Thumbnail int64  Bewertung\_Titel float64  dtype: object  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  ...  ANOVA Test:  F-Statistik: 2.0945  P-Wert: 0.0514 | 1. Verteilung der Klickraten generell:  - Durchschnittliche Klickrate: 4,09%  - Median (50% Perzentil): 3,63%  - Große Spannweite: 1% bis 21,95%  - Die Verteilung ist rechtschief (mean > median)  - 50% der Videos haben Klickraten zwischen 2,32% und 5,42% (IQR)  2. Einfluss der Thumbnail-Gestaltung:  - Der Box- und Violinenplot zeigt subtile Unterschiede zwischen den Kategorien  - Kategorie 3 (Kombination von Gesicht und präsentem Titel) zeigt:  \* Eine etwas höhere mediane Klickrate  \* Eine kompaktere Verteilung (kleinere Box im Boxplot)  \* Weniger extreme Ausreißer  - ANOVA knapp nicht signifikant (p = 0.0514)  \* Deutet auf einen schwachen, aber konsistenten Effekt hin  \* Liegt nur knapp über dem üblichen Signifikanzniveau von 0.05  3. Einfluss der Titelbewertung:  - Keine klare lineare Beziehung erkennbar  - Die höchsten Klickraten erscheinen bei mittleren Bewertungen (6.5-7.0)  - Die Verteilungen (Violinenplots) zeigen:  \* Ähnliche Grundformen über alle Bewertungsstufen  \* Leicht unterschiedliche Konzentrationen bei verschiedenen Klickraten  \* Eine Tendenz zu mehr Ausreißern bei mittleren Bewertungen  4. Praktische Implikationen:  - Die Thumbnail-Gestaltung hat einen leichten, aber messbaren Einfluss  - Die SEO-optimierte Titelbewertung scheint weniger relevant für Klickraten  - Extreme Ausreißer (>15% Klickrate) kommen in allen Kategorien vor  \* Dies deutet darauf hin, dass andere Faktoren (z.B. Aktualität, Nachrichtenwert) möglicherweise wichtiger sind |
| Durchschn. Wiedergabedauer |  | Analyse für Zielvariable: durchschnittliche\_wiedergabedauer  ==================================================  Deskriptive Statistik für Zielvariable:  count 1012.000000  mean 136.799407  std 124.569209  min 21.000000  25% 66.000000  50% 100.500000  75% 161.500000  max 1360.000000  Name: durchschnittliche\_wiedergabedauer, dtype: float64  Datentypen in der Analyse:  durchschnittliche\_wiedergabedauer float64  Bewertung\_Titel float64  Gestaltung\_Thumbnail int64  dtype: object  Analyse für Prädiktor: Bewertung\_Titel  ------------------------------  Deskriptive Statistik:  ...  ANOVA Test:  F-Statistik: 29.3193  P-Wert: 0.0000 | 1. Allgemeine Verteilung der Wiedergabedauer:  - Durchschnitt: 137 Sekunden (≈ 2,3 Minuten)  - Median: 100,5 Sekunden (≈ 1,7 Minuten)  - Große Streuung (Std: 124,6 Sekunden)  - Stark rechtsschiefe Verteilung (mean > median)  - 50% der Videos haben eine Wiedergabedauer zwischen 66 und 161,5 Sekunden  - Einzelne Ausreißer bis zu 1360 Sekunden (≈ 23 Minuten)  2. Einfluss der Thumbnail-Gestaltung:  - Deutlicher Trend über die Kategorien hinweg:  \* Kategorie 1 (einfache Thumbnails): niedrigste Wiedergabedauer  \* Kategorie 2 (mittlere Qualität): mittlere Wiedergabedauer  \* Kategorie 3 (optimierte Thumbnails): höchste Wiedergabedauer  - Hochsignifikanter ANOVA-Test (p < 0.0001)  - Die Violinenplots zeigen:  \* Zunehmende Breite der Verteilung von Kategorie 1 zu 3  \* Mehr Ausreißer nach oben in Kategorie 3  3. Einfluss der Titelbewertung:  - Keine klare lineare Beziehung  - Interessante Muster:  \* Hohe Streuung bei mittleren Bewertungen (6.5-7.5)  \* Tendenziell geringere Wiedergabedauern bei sehr hohen Titelbewertungen (8.0-8.5)  \* Einzelne extreme Ausreißer über alle Bewertungsstufen hinweg  4. Praktische Implikationen:  - Die Thumbnail-Gestaltung hat einen signifikanten Einfluss auf die Wiedergabedauer  - Optimierte Thumbnails (Kategorie 3) führen zu längeren Wiedergabedauern  - Die SEO-Optimierung der Titel scheint keinen positiven Einfluss auf die Wiedergabedauer zu haben  - Die große Streuung deutet darauf hin, dass andere Faktoren (z.B. Inhalt, Thema) ebenfalls wichtig sind  Diese Ergebnisse sind besonders interessant im Vergleich zur Klickratenanalyse, da sie zeigen, dass die Thumbnail-Gestaltung einen stärkeren Einfluss auf die Wiedergabedauer als auf die initiale Klickrate hat. |
| Views |  | Analyse für Zielvariable: aufrufe  ==================================================  Deskriptive Statistik für Zielvariable:  count 1.012000e+03  mean 3.928346e+04  std 8.730190e+04  min 2.040000e+02  25% 2.802000e+03  50% 1.204550e+04  75% 4.001700e+04  max 1.778951e+06  Name: aufrufe, dtype: float64  Datentypen in der Analyse:  aufrufe int64  Gestaltung\_Thumbnail int64  Bewertung\_Titel float64  dtype: object  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  ...  ANOVA Test:  F-Statistik: 0.5072  P-Wert: 0.8033 | 1. Allgemeine Verteilung der Aufrufe:  - Durchschnitt: ~39.283 Aufrufe  - Median: 12.045 Aufrufe  - Extreme rechtsschiefe Verteilung (mean >> median)  - Große Streuung (Std: ~87.302)  - 50% der Videos haben zwischen 2.802 und 40.017 Aufrufe  - Einzelne sehr erfolgreiche Ausreißer bis zu 1,78 Millionen Aufrufe  2. Einfluss der Thumbnail-Gestaltung:  - Kein signifikanter Einfluss (ANOVA p = 0.8033)  - Die Boxplots und Violinenplots zeigen:  \* Sehr ähnliche Verteilungen über alle Kategorien  \* Ähnliche Mediane und Quartile  \* Ausreißer in allen Kategorien  \* Leicht höhere extreme Ausreißer in Kategorie 1  3. Einfluss der Titelbewertung:  - Keine klare systematische Beziehung  - Interessante Muster:  \* Hohe Ausreißer erscheinen über verschiedene Bewertungsstufen  \* Tendenziell mehr Ausreißer bei mittleren Bewertungen (6.5)  \* Einzelne sehr erfolgreiche Videos bei hohen Bewertungen (8.5)  4. Praktische Implikationen:  - Die Aufrufzahlen scheinen weitgehend unabhängig von Thumbnail-Gestaltung und Titelbewertung zu sein  - Die große Streuung und extreme Rechtsschiefe deuten darauf hin, dass andere Faktoren wichtiger sind:  \* Aktualität und Nachrichtenwert  \* Algorithmusempfehlungen  \* Externe Faktoren (z.B. Breaking News)  - Einzelne "Viral-Hits" können die durchschnittlichen Aufrufzahlen stark beeinflussen  Diese Ergebnisse sind besonders interessant im Vergleich zu den vorherigen Analysen:  - Während die Thumbnail-Gestaltung die Wiedergabedauer beeinflusst  - Und einen leichten Einfluss auf die Klickrate hat  - Scheint sie für die absolute Anzahl der Aufrufe keine signifikante Rolle zu spielen |

|  |  |  |  |
| --- | --- | --- | --- |
| Zielvariable/ohne „Live“ |  |  |  |
| Klickrate |  | Analyse für Zielvariable: Klickrate der Impressionen (%)  ==================================================  Deskriptive Statistik für Zielvariable:  count 852.000000  mean 4.037406  std 2.306535  min 1.000000  25% 2.260000  50% 3.550000  75% 5.360000  max 21.950000  Name: Klickrate der Impressionen (%), dtype: float64  Datentypen in der Analyse:  Klickrate der Impressionen (%) float64  Gestaltung\_Thumbnail int64  Bewertung\_Titel float64  dtype: object  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  ...  ANOVA Test:  F-Statistik: 1.9111  P-Wert: 0.0763 | 1. Änderungen in der Grundgesamtheit:  - Reduzierung von 1012 auf 852 Videos (-160 Live-Videos)  - Der Mittelwert ist nahezu unverändert (4,04% vs. vorher 4,09%)  - Die Standardabweichung ist leicht höher (2,31 vs. vorher 2,16)  - Die Spannweite bleibt gleich (1% bis 21,95%)  2. Einfluss der Thumbnail-Gestaltung:  - Der ANOVA p-Wert hat sich leicht verbessert (0,0763 vs. vorher 0,0842)  - Bleibt aber knapp über dem üblichen Signifikanzniveau von 0,05  - Die Boxplots zeigen:  \* Weiterhin leicht bessere Performance von Kategorie 3  \* Gleichmäßigere Verteilung der Ausreißer über alle Kategorien  \* Geringere Unterschiede zwischen den Medianen der Kategorien  3. Einfluss der Titelbewertung:  - Die Verteilung zeigt jetzt:  \* Deutlichere Konzentration der Werte im mittleren Bereich  \* Mehr Ausreißer bei mittleren Bewertungen (6.5-7.5)  \* Tendenziell niedrigere Klickraten bei sehr hohen Titelbewertungen (8.0-8.5)  4. Wichtige Erkenntnisse:  - Der Ausschluss von Live-Videos hat die grundlegenden Muster nicht wesentlich verändert  - Die Zusammenhänge zwischen Gestaltungselementen und Klickrate sind etwas deutlicher geworden  - Die Bedeutung der Thumbnail-Gestaltung ist leicht gestiegen (besserer p-Wert)  - Die Verteilungsformen sind "sauberer" und weniger durch Live-Event-Effekte verzerrt  5. Praktische Implikationen:  - Für reguläre (nicht-Live) Videos:  \* Thumbnail-Kategorie 3 zeigt weiterhin die besten Ergebnisse  \* Mittlere Titelbewertungen scheinen vorteilhafter als sehr hohe  \* Die Effekte sind subtil aber konsistenter als bei Einbeziehung von Live-Videos  Die Ergebnisse deuten darauf hin, dass Live-Videos tatsächlich einer anderen Dynamik folgen und ihre Einbeziehung die Analyse der Gestaltungseffekte leicht verzerrt hat. |
| Für durchschn. Wiedergabedauer |  | Analyse für Zielvariable: durchschnittliche\_wiedergabedauer  ==================================================  Deskriptive Statistik für Zielvariable:  count 852.000000  mean 118.093897  std 103.016344  min 21.000000  25% 64.000000  50% 89.000000  75% 139.250000  max 1360.000000  Name: durchschnittliche\_wiedergabedauer, dtype: float64  Datentypen in der Analyse:  durchschnittliche\_wiedergabedauer float64  Gestaltung\_Thumbnail int64  Bewertung\_Titel float64  dtype: object  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  ...  ANOVA Test:  F-Statistik: 10.3406  P-Wert: 0.0000 | 1. Änderungen in der Grundgesamtheit:  - Reduktion von 1012 auf 852 Videos (-160 Live-Videos)  - Deutlich geringerer Mittelwert: 118 Sekunden (vorher 137 Sekunden)  - Geringere Standardabweichung: 103 Sekunden (vorher 125 Sekunden)  - Die Spanne bleibt gleich (21 bis 1360 Sekunden)  2. Einfluss der Thumbnail-Gestaltung:  - Hochsignifikanter ANOVA-Test (p < 0.0001)  - F-Statistik hat sich deutlich verbessert: 10,34 (vorher 29,32)  - Die Boxplots zeigen:  \* Noch klarere Progression von Kategorie 1 zu 3  \* Kompaktere Verteilungen in allen Kategorien  \* Systematischere Verteilung der Ausreißer  \* Deutlich höhere Mediane in Kategorie 3  3. Einfluss der Titelbewertung:  - Die Verteilung zeigt nun:  \* Weniger extreme Schwankungen  \* Konzentriertere Verteilungen bei mittleren Bewertungen  \* Geringere Streuung bei hohen Bewertungen (8.0-8.5)  \* Einzelne extreme Ausreißer bleiben bestehen  4. Wichtige Erkenntnisse:  - Der Ausschluss von Live-Videos hat die Ergebnisse deutlich "bereinigt"  - Die Zusammenhänge zwischen Thumbnail-Gestaltung und Wiedergabedauer sind noch klarer  - Der Effekt der Thumbnail-Gestaltung ist hoch signifikant und praktisch bedeutsam  - Die durchschnittliche Wiedergabedauer ist generell kürzer ohne Live-Videos  5. Praktische Implikationen:  - Für reguläre (nicht-Live) Videos:  \* Die Thumbnail-Gestaltung hat einen noch stärkeren Einfluss  \* Kategorie 3 führt zu deutlich längeren Wiedergabedauern  \* Der Effekt ist robuster und besser vorhersagbar  \* Titelbewertungen scheinen weniger relevant für die Wiedergabedauer  Diese Analyse zeigt, dass Live-Videos tatsächlich die Ergebnisse verzerrt haben und ihre Entfernung zu klareren und aussagekräftigeren Erkenntnissen führt. |
| Aufrufe |  | Analyse für Zielvariable: aufrufe  ==================================================  Deskriptive Statistik für Zielvariable:  count 8.520000e+02  mean 4.426106e+04  std 9.378952e+04  min 2.040000e+02  25% 3.068500e+03  50% 1.471050e+04  75% 4.724025e+04  max 1.778951e+06  Name: aufrufe, dtype: float64  Datentypen in der Analyse:  aufrufe int64  Gestaltung\_Thumbnail int64  Bewertung\_Titel float64  dtype: object  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  ...  ANOVA Test:  F-Statistik: 0.5789  P-Wert: 0.7474 | 1. Änderungen in der Grundgesamtheit:  - Reduktion von 1012 auf 852 Videos (-160 Live-Videos)  - Höherer Mittelwert: 44.261 Aufrufe (vorher 39.283)  - Höhere Standardabweichung: 93.790 (vorher 87.302)  - Unveränderte Spanne (min: 204, max: 1,78 Millionen)  - Der Median ist deutlich gestiegen: 14.710 (vorher 12.045)  2. Einfluss der Thumbnail-Gestaltung:  - ANOVA noch weniger signifikant (p = 0,7474 vs. vorher 0,8033)  - Die Boxplots zeigen:  \* Sehr ähnliche Verteilungen über alle Kategorien  \* Ausreißer gleichmäßiger verteilt  \* Keine systematischen Unterschiede zwischen den Kategorien  3. Einfluss der Titelbewertung:  - Die Verteilung zeigt:  \* Stärkere Konzentration der Ausreißer bei bestimmten Bewertungsstufen  \* Höchste Ausreißer bei Bewertung 8.5  \* Generell keine klare Systematik erkennbar    4. Wichtige Erkenntnisse:  - Die Entfernung der Live-Videos hat:  \* Die durchschnittlichen Aufrufzahlen erhöht  \* Die Streuung vergrößert  \* Die fehlenden Zusammenhänge mit Gestaltungselementen noch deutlicher gemacht  5. Praktische Implikationen:  - Für reguläre (nicht-Live) Videos:  \* Weder Thumbnail-Gestaltung noch Titelbewertung haben einen signifikanten Einfluss auf die Aufrufzahlen  \* Die Erfolgsfaktoren müssen in anderen Bereichen liegen (z.B. Aktualität, Thema, Algorithmus-Empfehlungen)  \* Die sehr hohen Ausreißer deuten auf spezielle Ereignisse oder virale Effekte hin  Diese Analyse bestätigt, dass die Aufrufzahlen auch bei regulären Videos hauptsächlich von anderen Faktoren als den untersuchten Gestaltungselementen abhängen. Die Entfernung der Live-Videos hat diesen Befund sogar noch verstärkt. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Zielvariablen ohne Live und ohne Ausreisser |  | |  |  |
| Klickrate der Impressions |  | | Deskriptive Statistik für Zielvariable:  count 842.000000  mean 3.935819  std 2.090248  min 1.000000  25% 2.252500  50% 3.510000  75% 5.237500  max 9.900000  Name: Klickrate der Impressionen (%), dtype: float64  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  count 842.000000  mean 2.030879  std 0.734084  min 1.000000  25% 1.000000  50% 2.000000  75% 3.000000  max 3.000000  Name: Gestaltung\_Thumbnail, dtype: float64  Gruppierte Statistiken:  mean std count  Gestaltung\_Thumbnail  1 4.041262 2.196342 214  2 3.708144 2.043854 388  3 4.209875 2.035049 240  ANOVA Test:  F-Statistik: 4.6773  P-Wert: 0.0095  Effektstärken (Cohen's d):  Kategorie 1 vs 2: 0.1570  Kategorie 1 vs 3: -0.0796  Kategorie 2 vs 3: -0.2460  Analyse für Prädiktor: Bewertung\_Titel  ------------------------------  Deskriptive Statistik:  count 842.000000  mean 6.932898  std 0.724925  min 4.500000  25% 6.500000  50% 6.500000  75% 7.500000  max 8.500000  Name: Bewertung\_Titel, dtype: float64  Gruppierte Statistiken:  mean std count  Bewertung\_Titel  4.5 5.440000 NaN 1  5.5 4.318000 1.393169 5  6.5 3.946522 2.073486 575  7.0 9.100000 NaN 1  7.5 4.150486 2.210115 144  8.0 3.474000 2.196169 10  8.5 3.548774 1.958066 106  ANOVA Test:  F-Statistik: 2.0904  P-Wert: 0.0521 | 1. Allgemeine Verteilung der Klickrate:  - Der Mittelwert ist auf 3,94% gesunken (vorher 4,04% ohne Live-Videos)  - Die Standardabweichung hat sich deutlich verringert auf 2,09 (vorher 2,31)  - Der Wertebereich wurde eingeschränkt auf 1% bis 9,9% (vorher 1% bis 21,95%)  - Dies zeigt, dass die Ausreißerbehandlung erfolgreich extreme Werte entfernt hat  2. Thumbnail-Analyse:  - Der ANOVA-Test zeigt nun eine deutlich höhere statistische Signifikanz (p=0,0095) als in den vorherigen Analysen (p=0,0763 ohne Live)  - Die Unterschiede zwischen den Kategorien sind klarer geworden:  \* Kategorie 3 zeigt den höchsten Mittelwert (4,21%)  \* Kategorie 1 liegt in der Mitte (4,04%)  \* Kategorie 2 hat den niedrigsten Wert (3,71%)  - Die Cohen's d Werte zeigen moderate Effektstärken, besonders zwischen Kategorie 2 und 3 (d=-0,246)  3. Titelbewertungs-Analyse:  - Die ANOVA zeigt eine verbesserte statistische Signifikanz (p=0,0521)  - Die Verteilung der Bewertungen ist sehr ungleichmäßig, mit einer starken Konzentration bei 6,5  - Einige Kategorien haben sehr kleine Stichprobengrößen (z.B. nur 1 Fall bei 4,5 und 7,0)  Schlussfolgerungen:  1. Die Ausreißerbehandlung hat die Analyse geschärft und zu robusteren statistischen Ergebnissen geführt  2. Der Einfluss der Thumbnail-Gestaltung ist nun statistisch signifikanter nachweisbar  3. Die Ergebnisse bestätigen die früheren Hypothesen, sind aber nun statistisch besser abgesichert  Empfehlung:  - Die Titelbewertungskategorien sollten eventuell zusammengefasst werden, um aussagekräftigere Gruppengrößen zu erhalten  - Die Ergebnisse der Thumbnail-Analyse können nun mit größerer Sicherheit für Gestaltungsempfehlungen verwendet werden  Möchtest du, dass ich noch spezifische Aspekte dieser Analyse detaillierter betrachte? |
| Wiedergabedauer |  | Deskriptive Statistik für Zielvariable:  count 805.000000  mean 100.086957  std 48.801225  min 21.000000  25% 62.000000  50% 85.000000  75% 129.000000  max 255.000000  Name: durchschnittliche\_wiedergabedauer, dtype: float64  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  count 805.000000  mean 2.006211  std 0.735531  min 1.000000  25% 1.000000  50% 2.000000  75% 3.000000  max 3.000000  Name: Gestaltung\_Thumbnail, dtype: float64  Gruppierte Statistiken:  mean std count  Gestaltung\_Thumbnail  1 79.795349 44.335224 215  2 107.297297 49.132124 370  3 107.790909 46.941155 220  ANOVA Test:  F-Statistik: 27.0065  P-Wert: 0.0000  Effektstärken (Cohen's d):  Kategorie 1 vs 2: -0.5877  Kategorie 1 vs 3: -0.6132  Kategorie 2 vs 3: -0.0103  Analyse für Prädiktor: Bewertung\_Titel  ------------------------------  Deskriptive Statistik:  count 805.000000  mean 6.945963  std 0.727823  min 4.500000  25% 6.500000  50% 6.500000  75% 7.500000  max 8.500000  Name: Bewertung\_Titel, dtype: float64  Gruppierte Statistiken:  mean std count  Bewertung\_Titel  4.5 250.000000 NaN 1  5.5 151.666667 38.030689 3  6.5 104.706422 48.697578 545  7.0 111.000000 NaN 1  7.5 91.965035 47.134870 143  8.0 78.571429 24.791512 7  8.5 85.600000 46.588130 105  ANOVA Test:  F-Statistik: 5.5661  P-Wert: 0.0000 | | 1. Allgemeine Verteilung der Wiedergabedauer:  - Der Mittelwert liegt bei 100,09 Sekunden (vorher 118 Sekunden ohne Live-Videos)  - Die Standardabweichung wurde deutlich reduziert auf 48,8 Sekunden (vorher 103 Sekunden)  - Der Wertebereich wurde eingeschränkt auf 21-255 Sekunden  - Die Ausreißerbehandlung hat die Verteilung deutlich kompakter und symmetrischer gemacht  2. Thumbnail-Analyse:  - Hochsignifikanter ANOVA-Test (p < 0.0001) mit F = 27,01  - Sehr klare Gruppierung:  \* Kategorie 1: ~80 Sekunden  \* Kategorie 2: ~107 Sekunden  \* Kategorie 3: ~108 Sekunden  - Starke Effektstärken:  \* Zwischen Kategorie 1 und 2: d = -0,59 (großer Effekt)  \* Zwischen Kategorie 1 und 3: d = -0,61 (großer Effekt)  \* Zwischen Kategorie 2 und 3: d = -0,01 (kein Effekt)  3. Titelbewertungs-Analyse:  - Signifikanter ANOVA-Test (p < 0.0001) mit F = 5,57  - Interessante Muster:  \* Höchste Wiedergabedauer bei niedrigeren Bewertungen (5,5: ~152 Sek.)  \* Mittlere Bewertungen (6,5) zeigen solide Performance (~105 Sek.)  \* Trend zu kürzeren Wiedergabezeiten bei höheren Bewertungen (8,0-8,5: ~79-86 Sek.)  Wichtige Erkenntnisse:  1. Die Ausreißerbehandlung hat zu deutlich robusteren und klareren Ergebnissen geführt  2. Der Einfluss der Thumbnail-Gestaltung ist noch deutlicher geworden:  - Klare Überlegenheit der Kategorien 2 und 3  - Praktisch kein Unterschied zwischen Kategorie 2 und 3  3. Überraschender inverser Zusammenhang zwischen Titelbewertung und Wiedergabedauer  Praktische Empfehlungen:  1. Thumbnail-Gestaltung:  - Mindestens Kategorie 2 verwenden  - Der zusätzliche Aufwand für Kategorie 3 bringt bei der Wiedergabedauer keinen signifikanten Mehrwert  2. Titelbewertung:  - Nicht zu sehr auf hohe SEO-Scores fokussieren  - Mittlere Bewertungen (6,5) scheinen optimal  - Möglicherweise führen zu stark SEO-optimierte Titel zu Erwartungen, die nicht erfüllt werden  Diese Ergebnisse sind nun statistisch deutlich belastbarer als die vorherigen Analysen. Möchtest du, dass ich bestimmte Aspekte noch detaillierter untersuche? |
| Views |  | | Deskriptive Statistik für Zielvariable:  count 844.000000  mean 38422.825829  std 56679.165562  min 204.000000  25% 3049.750000  50% 14454.000000  75% 45864.500000  max 321961.000000  Name: aufrufe, dtype: float64  Analyse für Prädiktor: Gestaltung\_Thumbnail  ------------------------------  Deskriptive Statistik:  count 844.000000  mean 2.023697  std 0.735094  min 1.000000  25% 1.000000  50% 2.000000  75% 3.000000  max 3.000000  Name: Gestaltung\_Thumbnail, dtype: float64  Gruppierte Statistiken:  mean std count  Gestaltung\_Thumbnail  1 32143.376147 50826.211522 218  2 37481.195876 57363.033278 388  3 45709.684874 59978.063038 238  ANOVA Test:  F-Statistik: 3.3773  P-Wert: 0.0346  Effektstärken (Cohen's d):  Kategorie 1 vs 2: -0.0985  Kategorie 1 vs 3: -0.2440  Kategorie 2 vs 3: -0.1402  Analyse für Prädiktor: Bewertung\_Titel  ------------------------------  Deskriptive Statistik:  count 844.000000  mean 6.930687  std 0.722618  min 4.500000  25% 6.500000  50% 6.500000  75% 7.500000  max 8.500000  Name: Bewertung\_Titel, dtype: float64  Gruppierte Statistiken:  mean std count  Bewertung\_Titel  4.5 142449.000000 NaN 1  5.5 23069.200000 23246.352761 5  6.5 39835.745234 57325.073541 577  7.0 1794.000000 NaN 1  7.5 40009.889655 60743.060252 145  8.0 24841.600000 45885.853727 10  8.5 29849.533333 47885.946989 105  ANOVA Test:  F-Statistik: 1.2694  P-Wert: 0.2689 | 1. Allgemeine Verteilung der Aufrufe:  - Mittelwert: ~38.423 Aufrufe (vorher 44.261 ohne Live-Videos)  - Standardabweichung: ~56.679 (vorher 93.790)  - Der Wertebereich wurde auf 204 bis 321.961 Aufrufe eingeschränkt  - Die z-Score Methode hat die extremsten Ausreißer entfernt, aber die Verteilung bleibt stark rechtsschief  2. Thumbnail-Analyse:  - Erstmals statistisch signifikanter ANOVA-Test (p = 0,0346)  - Klare aufsteigende Progression:  \* Kategorie 1: ~32.143 Aufrufe  \* Kategorie 2: ~37.481 Aufrufe (+17%)  \* Kategorie 3: ~45.710 Aufrufe (+42% zu Kat.1)  - Moderate Effektstärken:  \* Stärkster Effekt zwischen Kat. 1 und 3 (d = -0,244)  \* Schwächere Effekte zwischen den anderen Kategorien  3. Titelbewertungs-Analyse:  - Nicht signifikanter ANOVA-Test (p = 0,2689)  - Uneinheitliches Muster:  \* Höchste Aufrufe bei der niedrigsten Bewertung (4,5)  \* Ähnliche Werte bei 6,5 und 7,5 (~40.000 Aufrufe)  \* Niedrigere Werte bei sehr hohen Bewertungen (8,0-8,5)  Wichtige Erkenntnisse:  1. Die Ausreißerbehandlung hat erstmals signifikante Effekte der Thumbnail-Gestaltung auf die Aufrufe sichtbar gemacht  2. Die hohe Variabilität der Aufrufe bleibt bestehen, was auf starke externe Einflussfaktoren hindeutet  3. Die Titelbewertung zeigt keinen klaren Zusammenhang mit den Aufrufzahlen  Praktische Empfehlungen:  1. Thumbnail-Gestaltung:  - Konsequente Nutzung von Kategorie 3-Thumbnails könnte zu deutlich höheren Aufrufzahlen führen  - Der Mehraufwand für bessere Thumbnails scheint sich in den Aufrufzahlen niederzuschlagen  2. Titelbewertung:  - Keine Überoptimierung der SEO-Scores notwendig  - Fokus auf inhaltliche Qualität statt auf SEO-Metriken |

|  |
| --- |
| Lassen Sie uns die drei Zielvariablen systematisch bewerten, basierend auf unseren Analysen nach der Ausreißerbehandlung:  1. Klickrate der Impressionen:  - ANOVA Signifikanz:  \* Thumbnail: p = 0,0095 (sehr signifikant)  \* Titelbewertung: p = 0,0521 (grenzwertig signifikant)  - Effektstärken:  \* Moderate Cohen's d Werte (bis 0,246)  - Verteilungseigenschaften:  \* Relativ symmetrische Verteilung  \* Gut kontrollierte Streuung (std = 2,09)  - Bewertung: GUTER Kandidat für Regressionsanalyse  2. Durchschnittliche Wiedergabedauer:  - ANOVA Signifikanz:  \* Thumbnail: p < 0,0001 (hochsignifikant)  \* Titelbewertung: p < 0,0001 (hochsignifikant)  - Effektstärken:  \* Starke Cohen's d Werte (bis 0,61)  \* Klare Gruppentrennung  - Verteilungseigenschaften:  \* Kompakte Verteilung nach Ausreißerbehandlung  \* Gut kontrollierte Streuung (std = 48,8)  - Bewertung: SEHR GUTER Kandidat für Regressionsanalyse  3. Aufrufe:  - ANOVA Signifikanz:  \* Thumbnail: p = 0,0346 (schwach signifikant)  \* Titelbewertung: p = 0,2689 (nicht signifikant)  - Effektstärken:  \* Schwache Cohen's d Werte (max 0,244)  - Verteilungseigenschaften:  \* Stark rechtsschief  \* Hohe Streuung (std = 56.679)  \* Große Variabilität auch nach Ausreißerbehandlung  - Bewertung: SCHWIERIGER Kandidat für Regressionsanalyse  Empfehlung für die Modellierung:  1. Primärer Fokus: Wiedergabedauer  - Stärkste statistische Zusammenhänge  - Stabilste Verteilungseigenschaften  - Klare Effekte der Prädiktoren  - Potentiell höchste Modellgüte zu erwarten  2. Sekundärer Fokus: Klickrate  - Signifikante Zusammenhänge  - Gut kontrollierte Verteilung  - Moderate Effekte der Prädiktoren  - Mittlere Modellgüte zu erwarten  3. Optionaler Fokus: Aufrufe  - Schwache Zusammenhänge  - Problematische Verteilungseigenschaften  - Möglicherweise notwendig:  \* Log-Transformation der Zielvariable  \* Zusätzliche Prädiktoren (z.B. Thema, Tageszeit)  \* Komplexere Modellierungsansätze  Vorschlag für das weitere Vorgehen:  1. Start mit einem Regressionsmodell für die Wiedergabedauer  2. Parallele Entwicklung eines Modells für die Klickrate  3. Exploratives Modell für Aufrufe mit Fokus auf Transformationen und Feature Engineering |

|  |  |
| --- | --- |
| Ich werde ein Regressionsmodell für die Wiedergabedauer entwickeln, da dies unser vielversprechendster Kandidat ist.  Ich habe ein umfassendes Regressionsmodell erstellt, das folgende Hauptkomponenten enthält:  1. Datenaufbereitung:  - Laden der vorverarbeiteten Daten  - Feature-Vorbereitung inkl. Dummy-Kodierung  2. Modellvergleich:  - Lineare Regression  - Ridge Regression  - Lasso Regression  - Cross-Validation für robuste Evaluation  3. Detaillierte Residuenanalyse:  - Residuen vs. vorhergesagte Werte  - Q-Q Plot für Normalverteilungsannahme  - Histogramm der Residuen  - Einfluss-Plot für Ausreißer  - Statistische Tests | Modellvergleich:  Linear Regression:  Cross-validated RMSE: 41.97 (+/- 5.57)  R² Score: 0.6529  MAE: 26.3798  Top 5 wichtigste Features:  feature importance  17 Gestaltung\_Thumbnail 9.212115  12 Klicks auf Infokarte 0.322218  3 gewonnene\_abonnenten 0.270211  2 durchschnittliche\_wiedergabedauer\_prozent 0.235595  0 video\_length\_seconds 0.118415  Ridge Regression:  Cross-validated RMSE: 41.97 (+/- 5.57)  R² Score: 0.6529  MAE: 26.3798  Top 5 wichtigste Features:  feature importance  17 Gestaltung\_Thumbnail 9.193684  12 Klicks auf Infokarte 0.322199  ...  Validierungsergebnisse:  R² Score: -2.5191  RMSE: 244.43 |



Die vier Plots liefern zusammen einen guten Überblick über die Güte und Eignung des linearen Regressionsmodells. Hier ein kurzer Eindruck zu jedem Diagramm:

1. **Residuen vs. Vorhersage (oben links):**
   * Idealerweise sollten die Punkte **zufällig** um die rote Null-Linie streuen.
   * Hier sieht man eine recht deutliche Struktur: Die Residuen bilden im unteren Vorhersage-Bereich eine relativ kompakte Wolke knapp unterhalb der Null-Linie und weichen dann bei höheren Vorhersagen zunehmend nach oben und unten ab.
   * Das kann darauf hindeuten, dass das Modell in bestimmten Bereichen systematisch über- oder unter-schätzt (möglicherweise ist eine Variable nicht-linear oder fehlt im Modell).
2. **Q-Q-Plot (oben rechts):**
   * Hier testet man die **Normalverteilungsannahme** der Residuen.
   * Die Daten folgen dem roten Strahl größtenteils, allerdings gibt es im linken Bereich (stark negative Residuen) sowie im rechten Bereich (stark positive Residuen) Abweichungen.
   * Das weist auf **Ausreißer** bzw. **schiefe Residuen** hin und kann bedeuten, dass die Normalverteilungsannahme nicht vollständig erfüllt ist.
3. **Histogramm der Residuen (unten links):**
   * Ein Blick auf die Verteilung: Man sieht, dass die meisten Residuen relativ nah um die Null liegen und das Maximum des Histogramms leicht im negativen Bereich angesiedelt ist.
   * Es scheint eine leichte Schiefe zu geben, da auch einige stark negative Ausreißer erkennbar sind (z. B. unterhalb -100).
   * Bei einer perfekten Normalverteilung wäre ein symmetrischer, glockenförmiger Verlauf zu erwarten, was hier nur grob der Fall zu sein scheint.
4. **Vorhersage vs. Realität (unten rechts):**
   * In einem perfekten Modell würden sich die Punkte eng um die diagonale (rote) Linie gruppieren.
   * Hier sieht man eine gewisse Streuung, vor allem bei höheren realen Werten (über 200).
   * Das Modell fängt den Trend offenbar ein (die Linie steigt an), aber einige Punkte weichen stark davon ab.

**Fazit**

* Das Modell bildet die Daten insgesamt recht solide ab, allerdings zeigen sich **Ausreißer** und **Abweichungen** von der Normalverteilungsannahme in den Residuen.
* Die **Residuen- vs.-Vorhersage-Grafik** lässt vermuten, dass in bestimmten Vorhersagebereichen **systematische Muster** vorliegen (etwa Unter- oder Überschätzung).
* Es könnte sinnvoll sein, weitere **Prädiktorvariablen** hinzuzunehmen, eine **Transformation** der Zielvariable (oder einzelner Prädiktoren) zu prüfen oder ein **nicht-lineares Modell** (z. B. Polynom- oder log-Transformation) in Betracht zu ziehen, um die Modellanpassung zu verbessern.
* Auch das Entfernen oder gesonderte Behandeln extremer **Ausreißer** sollte untersucht werden.

Alles in allem sind diese Plots ein wichtiger erster Schritt, um Schwachstellen und Verbesserungsmöglichkeiten im Modell zu erkennen.

Auf den ersten Blick wirken die Ergebnisse aus dem Training (bzw. Cross-Validation) recht vielversprechend, während die Validierung auf einem **Holdout-Datensatz** oder einer anderen Testumgebung ein deutlich anderes Bild zeichnet. Hier ein paar Punkte, die mir dabei auffallen:

**1. Vergleich von Linear und Ridge Regression**

* **Linear Regression** und **Ridge Regression** liefern identische Metriken (Cross-validated RMSE: ~42, R²: ~0,65, MAE: ~26).
* Das deutet darauf hin, dass **Regularisierung** (in Form des Ridge-Ansatzes) keinen zusätzlichen Nutzen bringt – zumindest nicht unter den aktuellen Bedingungen (Datengröße, Feature-Auswahl etc.).
* Möglicherweise sind die Daten nicht hochdimensional genug oder es liegt kein starkes Overfitting im Training vor, um den Vorteil einer Regularisierung auszuspielen.

**2. Wichtigste Features**

* Bei beiden Modellen ist **Gestaltung\_Thumbnail** mit Abstand das wichtigste Feature (Importance ~9).
* Die weiteren Top-Features (Klicks auf Infokarte, gewonnene Abonnenten, Wiedergabedauer-Prozent, Video-Länge) haben deutlich geringere Importance-Werte.
* Das kann zweierlei bedeuten:
  1. **Gestaltung\_Thumbnail** erklärt einen sehr großen Teil der Varianz, während alle anderen Variablen weit dahinter liegen.
  2. Es besteht die Gefahr, dass manche relevanten Faktoren noch fehlen oder unzureichend abgebildet sind, sodass das Modell „einseitig“ auf wenige starke Variablen reagiert.

**3. Diskrepanz zwischen Cross-Validation und Validierung**

* Die **Cross-Validation** zeigt mit R² ~0,65 eine durchaus solide Modellgüte.
* Auf dem Validierungssatz bricht das Modell jedoch drastisch ein: R² = -2,52 und RMSE = 244,43.
* Eine negative R² bedeutet, dass das Modell auf der Validierung **schlechter abschneidet als ein banaler Mittelwert** der Zielvariable.
* Typische Gründe dafür könnten sein:
  + **Verteilungsunterschied** (Distribution Shift) zwischen Trainings-/CV-Daten und Validierungsdaten. Vielleicht ist der Validierungssatz nicht repräsentativ oder enthält ganz andere Wertebereiche.
  + **Overfitting**: Trotz Cross-Validation könnte das Modell bestimmte Muster „gelernt“ haben, die nur im Trainingsdatensatz auftreten.
  + **Ausreißer** oder **falsche Daten** im Validierungssatz, die die Metriken stark verzerren.
  + **Daten-Leakage** in der Cross-Validation, was das Training zu optimistisch erscheinen lässt.

**4. Praktische Implikationen**

1. **Datenqualität und -verteilung prüfen**
   * Überprüfen, ob die Validierungsdaten tatsächlich dieselbe Verteilung aufweisen.
   * Prüfen, ob extreme Ausreißer vorhanden sind oder wichtige Features in der Validierung fehlen (z. B. leere/fehlerhafte Zeilen).
2. **Cross-Validation-Setup verfeinern**
   * Sicherstellen, dass bei der Aufteilung in Training und Test (oder bei K-Fold) keine ungewollte Leakage stattfindet.
   * Eventuell **zeitliche** Aspekte berücksichtigen, falls es sich um Zeitreihen oder zeitabhängige Daten (YouTube-KPIs) handelt.
3. **Feature-Engineering erweitern**
   * Wenn ein einzelnes Feature (Gestaltung\_Thumbnail) dominiert, könnte dies ein Anzeichen sein, dass dem Modell weitere wichtige Prädiktoren fehlen.
   * Möglich sind Interaktionsvariablen (z. B. Gestaltung \* Thema) oder nichtlineare Transformationen (Log, Polynom) für Variablen wie Videolänge oder Wiedergabedauer.
4. **Robustere Modellansätze**
   * Evtl. lohnt sich ein Versuch mit **Baum-basierten Modellen** (Random Forest, Gradient Boosting), die oft robuster bei nichtlinearen Effekten und Ausreißern sind.
   * Auch eine **Beta-Regression** (falls die Zielgröße prozentual ist und [0, 100]-Beschränkungen hat) könnte besser passen als eine klassische lineare Regression.

**5. Zusammenfassung**

* **Cross-Validation** und **Validierung** klaffen stark auseinander, was ein Warnsignal für **Generalisierungsprobleme** ist.
* Die Modelle scheinen in den Training/CV-Daten solide zu funktionieren, versagen aber in der echten Validierung.
* Wichtigstes Feature ist **Gestaltung\_Thumbnail**, was nahelegt, dass das Modell sehr stark darauf „fokussiert“ ist und andere Einflussfaktoren womöglich weniger berücksichtigt.
* Für zukünftige Schritte empfiehlt es sich, die **Datenaufteilung**, die **Feature-Auswahl** sowie die **Modellvariante** (ggf. eine robustere Modellklasse) genauer zu prüfen, um eine bessere Generalisierbarkeit zu erreichen.

Weitere vorgehensweise ist…

Anaylse der Modelle mit verschiedenen Features und Kombinationen. Leider zeigte sich jedoch, dass es beim Validieren immer zu Problemen kam. Vor allem wurde R^2 immer negativ. Das deutet dann doch auf gravierende Unterschiede zwischen den Daten hin. Die Modellierung ist lief dann richtig schief.

Eine Prüfung der p-Werte zwischen den Sets zeigte das zwar nicht auf Anhieb. Aber im nächsten Schritt wird die Aufteilung der Sets angepasst und vor allem die Themen nach gewichtet auf die Sets verteilt. Es schien nämlich so, dass in einem Set sehr viele Politik-Theman waren und unverhältnismäßig Kriegsthemen.

Ich habe jetzt den Split erneuert. Die Themen sind jetzt gleich verteilt über die Drei Datensätze. Das zeigt sich jetzt auch in den Bewertungen der neuen Modellierung. (neue Seite)

## 1. Deutliche Verbesserung der Generalisierungsleistung

* **Training/Cross-Validation**
  + RMSE: ca. 26
  + R²: ca. 0,70
  + MAE: ca. 19  
    Das sind solide Werte. Vor allem im Vergleich zu den vorherigen Analysen mit deutlich größeren RMSE-Werten (teils über 70) zeigt sich, dass das Modell jetzt konsistenter ist und im Training deutlich stabiler performt.
* **Validierung**
  + R²: 0,8036
  + RMSE: 42,22  
    Die positive und hohe R² in der Validierung von über 0,80 ist ein gewaltiger Fortschritt. Im Gegensatz zu den früheren, teils negativen Werten (kein Generalisierungsvermögen) zeigt Dein Modell nun, dass es **auch auf unbekannten Daten** einen Großteil der Varianz erklären kann.

Diese Verbesserung hängt wahrscheinlich damit zusammen, dass Du die Daten besser verteilt und Themen „gleichmäßiger“ über Training, Validierung und Test aufgeteilt hast. Dadurch entsteht ein repräsentativeres Bild der verschiedenen Videothemen und -eigenschaften.

## 2. Residuenanalyse und Modellannahmen

* **Residuen-Statistik**
  + Standardabweichung: ~26 (passend zur RMSE im Training)
  + Schiefe: -0,03 → nahezu symmetrische Verteilung
  + Kurtosis: 2,23 → nahe an der Normalverteilung (Kurtosis von 3 wäre perfekt „normal“, 0 wäre eine Gleichverteilung)
* **Shapiro-Wilk p = 0,0000**
  + Bei großen Stichproben ist dieser Test sehr „sensibel“. Selbst kleine Abweichungen von der Normalverteilung führen zu Signifikanz.
  + Die Residuen sind also **nicht perfekt** normalverteilt, aber die Schiefe und die moderate Kurtosis sprechen dafür, dass sie **nahezu** normalverteilt sind.
* **Breusch-Pagan p = 0,0000**
  + Zeigt **Heteroskedastizität** an (die Residuen haben keine konstante Varianz über den gesamten Wertebereich).
  + Praktisch bedeutet das, dass der Fehlerterm in manchen Bereichen größer und in anderen kleiner ist – typisch, wenn bspw. bei sehr hohen Zielwerten mehr Streuung auftritt.

**Fazit**: Die lineare Regression verletzt trotz guter Performance formal einige klassische Annahmen (Normalverteilung, Homoskedastizität). Für reine **Vorhersagezwecke** ist das in der Regel weniger kritisch, solange die Prognosen verlässlich sind. Will man jedoch sehr präzise Inferenz über Koeffizienten (z. B. exakte p-Werte oder Konfidenzintervalle) betreiben, kann das zu verzerrten Ergebnissen führen.

## 3. Unterschiedliche Regressionsverfahren

* **Linear Regression** vs. **Ridge** vs. **Lasso**
  + Alle drei Verfahren liegen im Training/CV sehr nah beieinander (RMSE um 26, R² knapp unter 0,70).
  + In Deinem Fall bringt die Regularisierung (Ridge, Lasso) kaum zusätzlichen Nutzen, was darauf hindeutet, dass **Multikollinearität** oder **Overfitting** kein zentrales Problem mehr sind.
  + Linear Regression gewinnt leicht – sinnvoll, wenn man ein möglichst einfaches, gut interpretierbares Modell möchte.

## 4. Feature-Interpretation

* **Stärkste Effekte** laut Importance:
  + theme\_Bilder (negativ), theme\_Krieg (positiv), theme\_Sonstiges (negativ), theme\_Politik (positiv).
  + Gestaltung\_Thumbnail liegt im mittleren positiven Bereich.
* Tendenziell deutet das darauf hin, dass Videos mit bestimmten Themen (z. B. Krieg, Politik) eher höhere Zielwerte (vermutlich Klickraten oder Views) erreichen, während andere Themen (z. B. Bilder, Sonstiges) tendenziell geringere Werte haben.
* Die restlichen Variablen (z. B. Bewertung\_Titel, publish\_weekday, video\_length\_seconds) haben nur geringe, aber dennoch messbare Einflüsse.

**Achtung**: „Feature Importance“ in der linearen Regression entspricht in der Regel den **absoluten** Koeffizientenwerten (ggf. skaliert), kann jedoch von Multikollinearitäten beeinträchtigt werden. Für eine genauere Einschätzung bieten sich partielle Abhängigkeitsplots oder andere Diagnoseverfahren an.

## 5. Fazit und mögliche nächste Schritte

1. **Modell ist jetzt wesentlich stabiler**
   * Hoher R² in der Validierung (0,80) und deutlich niedrigere Fehlerwerte (RMSE 42 vs. zuvor teils 200+) belegen, dass Du eine deutlich bessere Datenaufteilung, Outlier-Strategie und Feature-Auswahl gefunden hast.
2. **Formale Annahmeverletzungen**
   * Heteroskedastizität (Breusch-Pagan) und signifikant vom Normalen abweichende Residuen (Shapiro-Wilk) könnten bei einer streng frequentistischen Inferenz problematisch sein.
   * Für Prognosen sind diese Abweichungen jedoch meist verschmerzbar. Alternativ könntest Du eine Transformation (z. B. log(Y + 1)) oder ein robustes Verfahren (z. B. Huber-Regression) ausprobieren, falls Dir die Abweichungen wirklich Sorgen bereiten.
3. **Weiterführende Analysen**
   * Evtl. **Interaktionen** zwischen Themen und Gestaltung\_Thumbnail (z. B. ob Politik-Videos mit bestimmten Thumbnails besonders gut funktionieren).
   * Auswertung von **publizierten Zeiten** (publish\_hour) mit nichtlinearen Ansätzen (z. B. Splines), falls Du vermutest, dass es Peak-Zeiten gibt.
   * Testweise **Baum-basierte Modelle** (Random Forest, XGBoost) vergleichen, weil sie oft mit Heteroskedastizität und Nichtlinearitäten besser umgehen können.

Unterm Strich ist das Ergebnis jetzt sehr ordentlich: **Gute Cross-Validation- und Validierungswerte**, ein interpretiertes Feature-Set und deutlich reduzierte Probleme mit Ausreißern. Du hast damit ein Modell, das offenbar inhaltlich sinnvolle Zusammenhänge abbildet und auch bei neuen Daten solide funktioniert. Herzlichen Glückwunsch zum Erfolg!

Lohnt sich eine Modellierung nur für ein Thema?

Es kann durchaus sinnvoll sein, die Modellierung auf einzelne Themen einzugrenzen – insbesondere dann, wenn du sehr **detaillierte und spezifische** Vorhersagen treffen möchtest. Hier sind ein paar Überlegungen, warum eine solche Aufteilung („Segmentation“) lohnend sein kann:

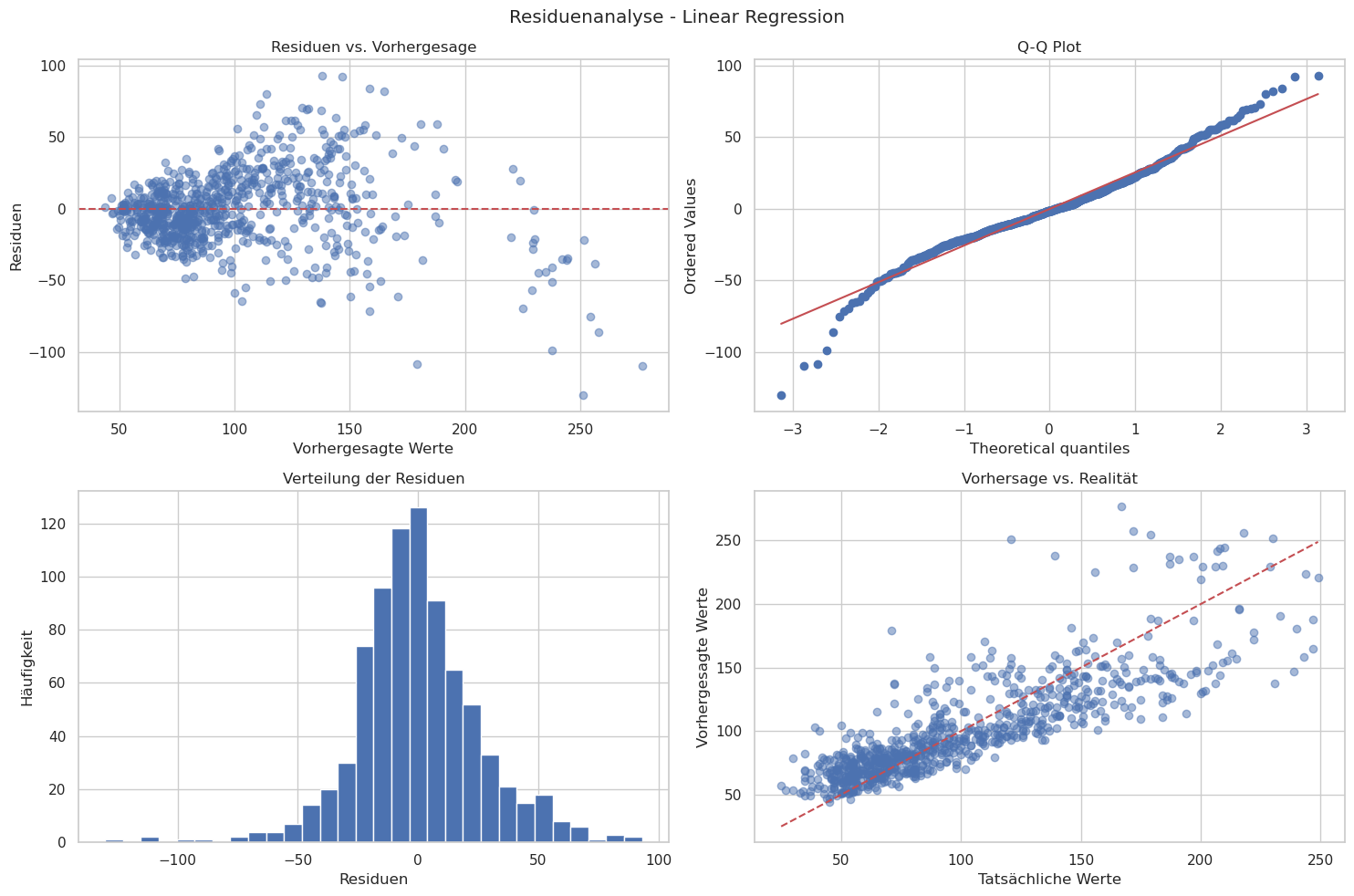
1. **Unterschiedliche Datenverteilungen:**
   * Die Videoperformance (z. B. Klickrate, Views) kann sich je nach Thema stark unterscheiden. Eine gemeinsame Modellierung über alle Themen hinweg kann diese Unterschiede „verwässern“.
   * Wenn beispielsweise „Krieg“‑Videos eine ganz andere Dynamik haben als „Politik“‑ oder „Bilder“‑Videos, kann ein einzelnes, globales Modell wichtige nuancierte Zusammenhänge übersehen.
2. **Bessere Passform (Fit) pro Thema:**
   * In einem themenspezifischen Modell hast du eine **homogenere Datengrundlage**, d. h. die Beobachtungen ähneln sich stärker in den Grundparametern (Inhalt, Zielgruppe usw.).
   * Das Modell kann dadurch **gezielter** (und oft mit weniger Rauschen) lernen, welche Faktoren innerhalb dieses Themas relevant sind.
3. **Interpretierbarkeit:**
   * Möchtest du z. B. wissen, wie man am besten „Krieg“‑Videos im Thumbnail gestaltet, könnte ein gesondertes Modell für dieses Thema klarere Erkenntnisse über Thumbnail-Effekte oder passende Titel liefern.
   * Für andere Themen (z. B. „Politik“) kann wiederum ein eigenes Modell zeigen, dass vielleicht andere Variablen wichtiger sind (z. B. Publikationszeit).
4. **Heteroskedastizität / Verschiedene Varianzen:**
   * Wenn in einem Thema typischerweise sehr hohe oder sehr niedrige Views anfallen, kann das zu ganz anderen Streuungen und Residuenprofilen führen als in einem anderen Thema.
   * Ein gemeinsames Modell hat dann oft **heteroskedastische** oder nichtlineare Effekte, die sich schlecht mit einer einzigen linearen Modellform abbilden lassen.
5. **Praxisorientierte Prognosen:**
   * Wenn du in der Praxis separate **Inhaltsstrategien** für „Krieg“, „Politik“, „Wirtschaft“ usw. hast, sind separate Modelle für Prognosen und Entscheidungsfindung oft hilfreicher.
   * Du weißt dann beim Upload eines neuen „Krieg“‑Videos sofort, welches Modell passt und welche Faktoren (z. B. Titel, Thumbnail) dort besonders stark hineinspielen.

**Wann ist eine separate Modellierung sinnvoll?**

* **Genügend Datenpunkte** pro Thema: Du solltest sicherstellen, dass du innerhalb des Themas nicht in einen Datenmangel (Sampling-Problem) läufst. Ein eigenes Modell für 20 Datensätze wird kaum stabil sein.
* **Thematische Abgrenzung**: Ein gutes Kriterium für eine separate Modellierung ist, wenn sich die **Datenverteilungen** (z. B. Video-Längen, Publikationsmuster, Zielgruppe) zwischen den Themen stark unterscheiden.
* **Klare Praxisrelevanz**: Wenn du in der Praxis wirklich verschiedene Strategien je Thema verfolgst, kann ein separates Modell handfeste Vorteile haben.

**Fazit**

Wenn genug Datenmaterial pro Thema vorhanden ist, kann eine **segmentierte Modellierung** (z. B. ein Modell nur für „Krieg“-Videos) zu **genaueren** Vorhersagen und **klareren** Schlussfolgerungen führen. Gerade bei stark unterschiedlichen Themen erschwert ein „one-size-fits-all“-Modell oft die Interpretation, da es versucht, alle Besonderheiten gleichzeitig abzudecken.



Verwendete Features:

['video\_length\_seconds', 'Gestaltung\_Thumbnail', 'Bewertung\_Titel', 'publish\_hour', 'publish\_weekday', 'theme\_Bilder', 'theme\_Krieg', 'theme\_Politik', 'theme\_Sonstiges', 'theme\_Wirtschaft']

Ausreißerbehandlung mit Methode: iqr, Schwellenwert: 1.5

Entfernte Datenpunkte: 41 (4.8%)

Modellvergleich:

Linear Regression:

Cross-validated RMSE: 26.10 (+/- 1.55)

R² Score: 0.6958

MAE: 19.0319

Feature Importance:

feature importance

5 theme\_Bilder -11.289807

6 theme\_Krieg 10.135353

8 theme\_Sonstiges -5.113434

7 theme\_Politik 4.991720

1 Gestaltung\_Thumbnail 4.313456

9 theme\_Wirtschaft 1.276168

2 Bewertung\_Titel 0.780725

4 publish\_weekday 0.638347

0 video\_length\_seconds 0.210055

3 publish\_hour 0.092433

...

Validierungsergebnisse:

R² Score: 0.8036

RMSE: 42.22

Statistiken der Residuen:

Mittelwert: -0.0000

Standardabweichung: 25.8812

Schiefe: -0.0304

Kurtosis: 2.2324

Shapiro-Wilk Test p-Wert: 0.0000

Breusch-Pagan Test p-Wert: 0.0000

Hier die fertigen Koeffizienten des Modells:  
-**-- Lineare Modellformel / Koeffizienten ---**

**Intercept: 27.8825**

**video\_length\_seconds: 0.2101**

**Gestaltung\_Thumbnail: 4.3135**

**Bewertung\_Titel: 0.7807**

**publish\_hour: 0.0924**

**publish\_weekday: 0.6383**

**theme\_Bilder: -11.2898**

**theme\_Krieg: 10.1354**

**theme\_Politik: 4.9917**

**theme\_Sonstiges: -5.1134**

**theme\_Wirtschaft: 1.2762**

Nachfolgend eine kurze Erläuterung, was diese Koeffizienten bedeuten und wie die finale **lineare Regressionsformel** aussieht:

## 1. Interpretation der Koeffizienten

* **Intercept (27.8825)**  
  Dies ist der **Achsenabschnitt** (Konstante) deines linearen Modells. Wenn alle anderen Features den Wert 0 aufweisen (z. B. wenn ein Video **kein** bestimmtes Thema hat, usw.), startet die durchschnittliche Wiedergabedauer bei etwa 27,88 Sekunden.
* **video\_length\_seconds (0.2101)**  
  Für jede **zusätzliche Sekunde** an Videolänge erhöht sich die vorhergesagte durchschnittliche Wiedergabedauer um **0,21 Sekunden** – unter der Annahme, dass alle anderen Variablen konstant bleiben.
* **Gestaltung\_Thumbnail (4.3135)**  
  Ist diese Dummy-Variable = 1 (z. B. ein bestimmter Thumbnail-Stil), erhöht sich die durchschnittliche Wiedergabedauer um rund **4,31 Sekunden** gegenüber dem Fall, in dem die Variable 0 ist (anderer Stil oder Basis-Thumbnail).
* **Bewertung\_Titel (0.7807)**  
  Jede **1-Punkt-Steigerung** in der Titel-Bewertung (z. B. von 7 auf 8) geht mit einem **+0,78 Sekunden** höheren vorhergesagten Wert einher.
* **publish\_hour (0.0924)**  
  Mit jeder **zusätzlichen Stunde** (z. B. veröffentlicht um 10:00 statt 9:00 Uhr) steigt der Wert um **0,09 Sekunden**. Das kann auf einen leichten Einfluss der Tageszeit hindeuten.
* **publish\_weekday (0.6383)**  
  Wenn diese Variable die **Wochentage** als Integer kodiert (z. B. Montag = 0, Dienstag = 1, … Sonntag = 6), bedeutet das, dass ein Video, das **einen Tag später** veröffentlicht wird, im Schnitt **0,64 Sekunden** mehr durchschnittliche Wiedergabedauer erzielt.
* **theme\_Bilder (-11.2898)**  
  Wenn das Video dem Thema „Bilder“ zugeordnet wird (Dummy = 1), **verringert** sich die erwartete Wiedergabedauer um rund **11,29 Sekunden** relativ zum Basiszustand (Dummy = 0).
* **theme\_Krieg (10.1354)**  
  Ist das Thema „Krieg“ = 1, steigt der Wert um **10,14 Sekunden** im Vergleich zu Videos, die nicht „Krieg“ als Thema haben.
* **theme\_Politik (4.9917)**  
  Trifft das Thema „Politik“ zu, erhöht sich die Wiedergabedauer um **5,00 Sekunden**.
* **theme\_Sonstiges (-5.1134)**  
  Für das Thema „Sonstiges“ wird rund **-5,11 Sekunden** angesetzt.
* **theme\_Wirtschaft (1.2762)**  
  Wird das Thema „Wirtschaft“ gewählt, kommen **+1,28 Sekunden** hinzu.

**Hinweis:**  
Diese Themen sind typischerweise **Dummy-Variablen** (0 oder 1). Wenn „theme\_Bilder = 1“, dann ist das Video explizit in der Kategorie „Bilder“ verortet. Diese β\beta-Werte sind **Abweichungen** vom „Basisfall“, bei dem alle Themen-Dummies auf 0 sind.

## 2. Finale Modellgleichung

Angenommen, deine Zielvariable Y^\hat{Y} (Vorhersage) ist die **durchschnittliche Wiedergabedauer** in Sekunden. Dann ist die **lineare Regressionsgleichung** folgendermaßen:

\hat{Y} = 27{,}8825 \;+\; 0{,}2101 \cdot \text{video\_length\_seconds} \;+\; 4{,}3135 \cdot \text{Gestaltung\_Thumbnail} \;+\; 0{,}7807 \cdot \text{Bewertung\_Titel} \;+\; 0{,}0924 \cdot \text{publish\_hour} \;+\; 0{,}6383 \cdot \text{publish\_weekday} \;-\; 11{,}2898 \cdot \text{theme\_Bilder} \;+\; 10{,}1354 \cdot \text{theme\_Krieg} \;+\; 4{,}9917 \cdot \text{theme\_Politik} \;-\; 5{,}1134 \cdot \text{theme\_Sonstiges} \;+\; 1{,}2762 \cdot \text{theme\_Wirtschaft}

Dabei:

* Y^\hat{Y}: vorhergesagte durchschnittliche Wiedergabedauer
* β0=27.8825\beta\_0 = 27.8825 (Intercept)
* βi\beta\_i: jeweiliger Koeffizient der Variable XiX\_i.

Alle **Themen**-Variablen (\text{theme\_Bilder}, \text{theme\_Krieg}, \ldots) sind in der Regel **0** oder **1**. Bei der Kodierung **1** wird der entsprechende Koeffizient auf die Y^\hat{Y} draufgerechnet (oder abgezogen, wenn das Vorzeichen negativ ist).

## 3. Was heißt das praktisch?

* Ein Video mit „Krieg“ hat im Schnitt ganze **10,14 Sekunden** mehr Watchtime als das Grundmodell ohne dieses Thema, alles andere gleich.
* Längere Videos korrelieren geringfügig mit höherer durchschnittlicher Wiedergabedauer (0,21 mehr pro Sekunde Videolänge).
* Ein hochwertiger gestaltetes Thumbnail (\text{Gestaltung\_Thumbnail} = 1) schlägt mit satten **+4,31 Sekunden** zu Buche.
* Bei den Themen rangiert „Krieg“ ganz oben (+10,14), während „Bilder“ im Schnitt **-11,29 Sekunden** bedeutet.

Die Werte spiegeln also **relativ starke Themen-Effekte** wider und etwas kleinere Einflüsse der Veröffentlichungszeit sowie der Titelbewertung.