

Clusteranalyse: eine Anwendung für die Berggebiete

Wüest Partner, 12. Juli 2021

Wüest Partner schlägt eine neue Einteilung der Gemeinden in den Berggebieten vor. Beauftragt durch das Staatssekretariat für Wirtschaft SECO hat sie die 811 in den Berggebieten liegenden Gemeinden in möglichst einheitliche Typologien eingeteilt. Das bedeutet, dass die in einer Typologie zusammengefassten Gebiete in sich möglichst ähnliche sozioökonomische Parameter aufweisen und sich gleichzeitig möglichst stark von den anderen Typologien unterscheiden sollen. Eine Vielzahl von sozioökonomischen Variablen bilden die Grundlage für eine neue Typologisierung. Die Einteilung wurde mittels Clusteranalyse vorgenommen.

Clusteranalyse

Eine Clusteranalyse ist ein exploratives statistisches Verfahren. Dabei werden ähnliche Strukturen in Datenbeständen gesucht und dann verschiedenen Gruppen zugeordnet. Diese Gruppen werden als Cluster bezeichnet. Die Methodik der Clusteranalyse wird im nächsten Kapitel genauer beschrieben.

Clusteranalysen werden zum Beispiel im Marketing angewendet, wenn es darum geht, die Kundschaft in Segmente einzuteilen und jedes Segment mit Produkten zu bedienen, die für die Zielgruppe massgeschneidert sind. So werden etwa geographische Gebiete zu Gruppen mit ähnlichen Absatzmerkmalen zusammengefasst. Dies hat den Vorteil, dass neu lancierte Produkte zuerst in Regionen eingeführt werden können, die vielversprechende Erfolgchancen aufweisen.

Analog zur Kundensegmentierung werden in der vorliegenden Clusteranalyse die Gemeinden der Berggebiete in Typologien eingeteilt. Die Absicht hinter dieser Einteilung ist analog zur Kundensegmentierung im Marketing: Massnahmen zur wirtschaftlichen Entwicklung der Berggebiete sollen möglichst genau zu einem Gemeindetyp passen. Ein Förderprogramm, das in der Gemeinde von Cluster 1 funktioniert, dürfte tendenziell passender sein für eine Gemeinde in demselben Cluster als für eine Gemeinde in einem anderen Cluster.

Der grosse Vorteil der Clusteranalyse liegt einerseits in der durch die Berücksichtigung einer Vielzahl von Eigenheiten angestrebten Ganzheitlichkeit und andererseits in der Objektivität. Die menschliche Eigenschaft der subjektiven Überbewertung bekannter Faktoren wird so eliminiert.

Clusteranalyse für die Berggebiete

Um sozioökonomische Ähnlichkeiten zu messen, wurden 34 Indikatoren ausgewählt und analysiert. Diese Indikatoren sind am Ende des Dokuments aufgeführt. Sie charakterisieren zum Beispiel Wachstum und Struktur sowohl der Bevölkerung als auch der Beschäftigung einer Gemeinde, und sie beschreiben deren Standortattraktivität, etwa über die Erreichbarkeit. Auch Indikatoren aus dem Bau- und Immobilienmarkt sind Teil der Datengrundlage für die folgenden Analysen. Eine Übersicht über die 34 Variablen befindet sich ganz am Ende dieses Dokuments. Die Einteilung des Berggebiets in unterschiedliche Cluster wurde als iterativer Optimierungsprozess angelegt. Das heisst, es wurden zahlreiche verschiedene

Optionen so lange getestet, bis das optimale Ergebnis gefunden wurde. Folgende Parameter standen im Zentrum unserer Überlegungen:

– **Anzahl Cluster:**

Es wurden Testläufe mit unterschiedlichen Mengen an Clustern durchgeführt. Eine Aufteilung in sechs Cluster lieferte die zufriedenstellendsten Resultate. Diese sechs Cluster werden «Gemeindetypen» genannt.

– **Raumeinheit:**

Cluster können auf Basis verschiedener Raumeinteilungen gebildet werden. Es wurden Tests mit Gemeinden, mit Arbeitsmarktregionen, aber auch mit weiteren Zellen durchgeführt. Diese weiteren Zellen ergeben sich aus Verschneidungen der Arbeitsmarktregionen mit den Raumtypen von regionsuisse oder mit Bezirken. Schliesslich stellte sich heraus, dass Gemeinden die für die geforderten Zwecke am besten geeignete Grundeinheit darstellen.

– **Variablen und deren Gewichtung:**

Es wurden Tests mit wenigen Variablen durchgeführt (etwa, indem auf die Bevölkerung oder auf die Beschäftigung fokussiert wurde) und solche mit dem vollen Set aller 34 Indikatoren. Ausserdem wurden die Auswirkungen getestet, die sich aus unterschiedlichen Gewichtungen der Indikatoren ergaben. Die Wahl fiel auf ein Konzept, das alle 34 Indikatoren mit einbezieht, diese aber unterschiedlich gewichtet. Die Gewichte von 1.0 bis 3.0 ergaben sich als Durchschnitt aus vier Expertenmeinungen.

Methodenbeschreibung: Clusteranalyse

Im Folgenden wird erläutert, wie die Clusteranalyse vorgenommen wurde. Es wird dargelegt, welcher Ansatz gewählt und wie die Cluster optimiert wurden. Schliesslich werden noch statistische Verfahren beschrieben, die in der Umsetzung zu beachten waren. Der eilige Leser kann dieses Unterkapitel überspringen.

k-means-Clustering

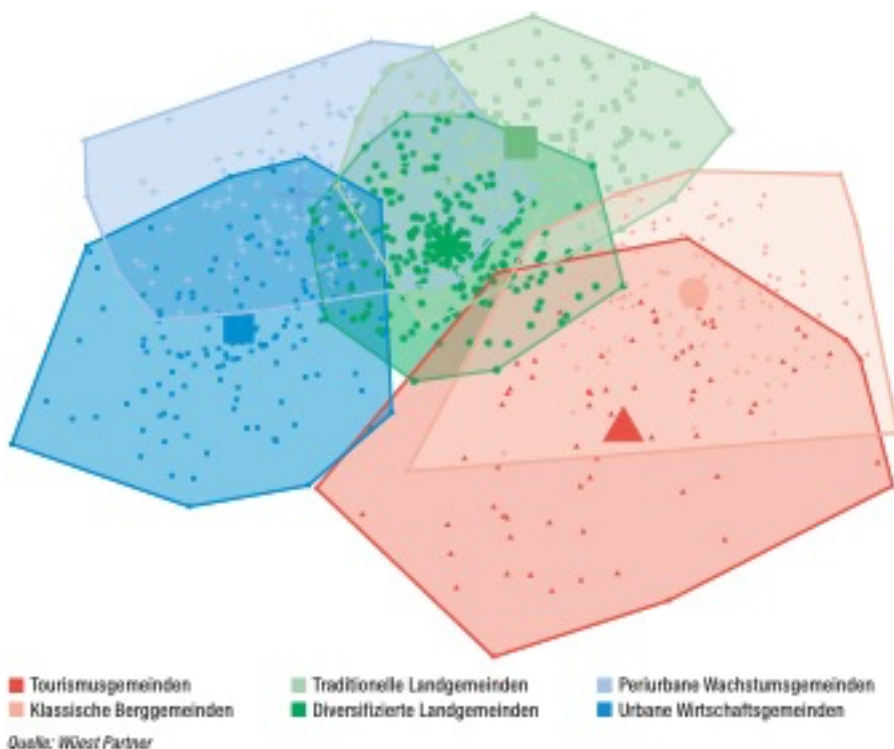
Es gibt verschiedene Methoden der Clusteranalyse. Wir haben uns für das sogenannte «k-means-Clustering» entschieden, das den populärsten Ansatz darstellt. Der Name kommt daher, dass der Nutzer die Anzahl Cluster selber wählen kann. Es muss einem jedoch bewusst sein, dass die Wahl der Anzahl Cluster sich auf das Resultat auswirkt und dieses je nachdem verschlechtern kann. Eine Stärke des «k-means-Clustering» besteht darin, dass der Algorithmus sehr effizient ist, auch bei grossen Datenmengen. Der Algorithmus ist auch deshalb effizient, weil nicht die Anzahl Gruppen optimiert wird. Vielmehr liefert das «k-means-Clustering» für die vorgegebene Anzahl Gruppen eine Einteilung, die eine optimale Homogenität innerhalb der Gruppen aufweist. Wie die Optimierung erfolgt, wird nun beschrieben.

Messung: Abstand vom Zentroiden

Gesucht werden Cluster von Gemeinden, die sozioökonomisch in sich näher beieinanderliegen als Gemeinden, die zu anderen Clustern gehören. Es gilt also, die Ähnlichkeit zwischen den Gemeinden numerisch auszudrücken. Als Masszahl verwenden wir die euklidische Distanz. Das heisst, wir messen zum Beispiel für das Bevölkerungswachstum, wie weit es für jede der 811 in den Berggebieten liegenden Gemeinden vom mittleren Bevölkerungswachstum eines Clusters entfernt ist. Es wird also für jede Ausprägung einer Gemeinde die Distanz zum

Clusterzentrum gemessen. Das Clusterzentrum wird in der Fachsprache Zentroid genannt.

Diese Abstandsmessung einer Gemeinde zu ihrem Zentroiden wird für jede Variable vorgenommen. Mit unseren 34 sozioökonomischen Variablen spannen wir so einen Raum mit 34 Dimensionen auf und bilden darin jede der 811 Gemeinden gemäss ihren Ausprägungen ab.¹ Das menschliche Auge ist sich an maximal drei Dimensionen gewöhnt; am geläufigsten sind uns Einteilungen in einen Raum mit zwei Dimensionen. Die folgende Grafik zeigt, wie sich die 811 Gemeinden der Berggebiete in einem zweidimensionalen Raum verteilen, sowie deren Einteilung in 6 Cluster.



Clustervisualisierung auf Stufe Gemeinde mit 6 Clustern und ihren Zentroiden.

Technische Anmerkung: Die beiden Dimensionen können nicht einer einzelnen Variable zugeordnet werden. Es wurde eine «Principal Component Analysis» (PCA) durchgeführt, die aus den 34 verwendeten Variablen zwei Dimensionen abbildet, sodass ein möglichst grosser Teil der Varianz erklärt wird.

Es zeigt sich, dass die 6 Cluster sich teils überlappen, teils voneinander getrennt sind. Es gilt jedoch zu berücksichtigen, dass es sich hier um eine schematische Darstellung mit nur zwei Dimensionen handelt, die der beschränkten Vorstellungskraft des menschlichen Auges entgegenkommt. In der Realität operieren wir mit 34 Variablen und haben somit eigentlich 34 Dimensionen. Je mehr Dimensionen hinzukommen, desto stärker unterscheiden sich die Cluster voneinander. Die tatsächlich errechneten Cluster mit ihren 34 Dimensionen überlappen sich also weniger, als es die hier gezeigte Abbildung suggeriert.

Je homogener ein Cluster ist, desto geringer sind die Distanzen zwischen den einzelnen Raumeinheiten und dem dazugehörigen Zentroiden. Dieser ist als

¹ Streng genommen sind nur 33 der 34 Variablen sozioökonomische Variablen. Die Variable «Meter über Meer» ist keine sozioökonomische Variable. Aber sie kann als Proxy verwendet werden für sozioökonomische Variablen wie etwa Erreichbarkeit oder Beschäftigungsmöglichkeiten. Deshalb erachtet es Wüest Partner als passend, auch die Höhe über Meer einer Gemeinde als Variable mit einzubeziehen.

Mittelpunkt eines Clusters dargestellt. In dieser Abbildung kann nun für jede der 811 Gemeinden gemessen werden, wie weit eine Gemeinde von ihrem clusterspezifischen Zentroiden entfernt ist. Im Gegensatz zur vereinfachten Darstellung mit zwei Dimensionen in der obigen Abbildung wird in der Realität der Abstand zum Zentroiden gemessen, der sich in einem Raum mit 34 Dimensionen ergibt. Dieser Abstand wird im Verlauf der Clusteranalyse minimiert.

Optimierung

Die Optimierung ist ein iterativer Prozess: Bei der Wahl von 6 Clustern wurden zuerst 6 Gemeinden je einem Cluster zugeordnet. Diese 6 Datenpunkte bildeten dann je eine Gruppe mit einem Zentroiden. Dann wurden weitere Gemeinden dem nächstliegenden Cluster zugeordnet. Durch die neu dazugekommenen Datenpunkte wurde der Zentroid neu gerechnet. Dieser Schritt wurde dann so oft wiederholt, bis keine Veränderung der Zentroiden von der vorherigen Iteration mehr auftraten respektive alle 811 Berggemeinden einem Cluster zugeordnet waren. Nach Abschluss der Clusteranalyse, also wenn die 34 Variablen der 811 Gemeinden auf die 6 Cluster verteilt sind, gibt es keine Überlappungen mehr zwischen den einzelnen Clustern, wie das in der obigen Abbildung mit zwei Dimensionen noch der Fall war.

Umsetzung

Die vorliegende Clusteranalyse wurde mit dem Statistikprogramm «R» durchgeführt. Vor der eigentlichen Clusteranalyse mussten die Daten entsprechend aufbereitet werden. Vier wichtige Vorbedingungen mussten dabei erfüllt sein:

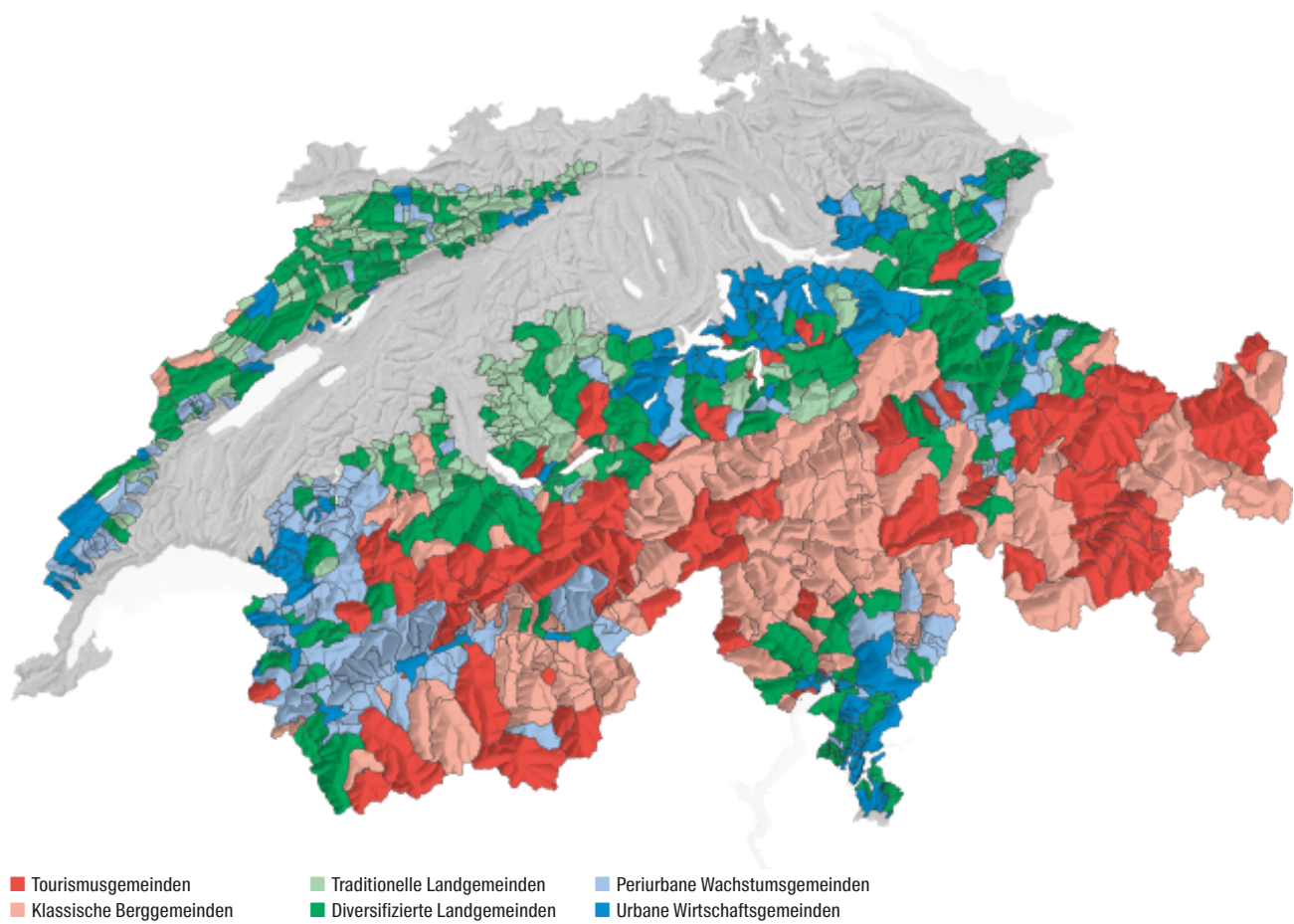
- Alle Variablen mussten normiert oder skaliert werden (z. B. zwischen -1 und $+1$). Diese Normierung ist wichtig, damit die Abstände vom Zentroiden für alle Variablen gleich gemessen werden können. Ausserdem ermöglicht diese Normierung Vergleiche zwischen unterschiedlichen Variablen, sodass z. B. eine Differenz von 0.1 bei unterschiedlichen Variablen denselben Unterschied beschreibt.
- Die zu gruppierenden Daten oder Variablen mussten vergleichbar sein. Das heisst für unser Projekt zum Beispiel, dass wir, wo nötig und möglich, mit Pro-Kopf-Werten rechnen, damit die Daten nicht von der Grösse der Gemeinde verzerrt werden.
- Fehlende Werte mussten eliminiert werden. Diese Lücken haben wir geschlossen, indem fehlende Werte zum Beispiel mit dem durchschnittlichen Variablenwert über alle Gemeinden aufgefüllt wurden.

Sehr hohe respektive sehr tiefe Ausprägungen mussten genau analysiert und allenfalls bereinigt werden. Wir entschieden uns dazu, die tiefsten bzw. höchsten 2% auf das nächste Minimum bzw. Maximum zu schreiben.

Ergebnis

Nun zeigen wir das Ergebnis der Clusteranalyse als Karte. Dort sind die 811 Gemeinden im Berggebiet eingefärbt gemäss ihrer Zugehörigkeit zu einem Cluster, die jeweils einen Namen erhalten haben.

Mehr Informationen sind in der veröffentlichten Studie «Wüest Partner (2021). Berggebiete: Sozioökonomische Analyse. Studie im Auftrag des SECO, Bern» zu finden.



Quelle: Wüest Partner

Daten

Die Tabelle auf der nächsten Seite listet detaillierte Angaben zu den verwendeten Variablen und zu den Datengrundlagen auf.

Clusteranalyse

Name	Indikator	Spezifikation	Zeit(raum)	Quelle	Gewicht	Ausprägung	Kategorie	Entw./Stand
Bevölkerungswachstum	Bevölkerungswachstum	Total Bevölkerung	2010–2019/ 2020–2040	BFS STATPOP, Wüest Partner	2.75	Wachstum	Bevölkerung/Wohnen	Entwicklung
Wachstum Erwerbstätige	Bevölkerungswachstum	im Erwerbsalter und Minderjährige	Annualisiert 2020–2040	Wüest Partner	2.50	Erwerbstätigepotenzial	Bevölkerung/Wohnen	Entwicklung
Wachstum Schweizer	Bevölkerungswachstum	Schweizer	Annualisiert 2010–2019	BFS STATPOP	1.00	Wachstum	Bevölkerung/Wohnen	Entwicklung
Wachstum Ausländer	Bevölkerungswachstum	Ausländer	Annualisiert 2010–2019	BFS STATPOP	0.75	Wachstum	Bevölkerung/Wohnen	Entwicklung
Wanderungssaldo	Bevölkerungswachstum	Verhältnis Zuzug/Wegzug innerhalb CH	Durchschnitt 2015–2019	BFS STATPOP	2.50	Lebensqualität	Bevölkerung/Wohnen	Entwicklung
Anteil Pensionäre	Bevölkerungsanteil	Pensioniert	2040	BFS STATPOP	2.00	Sozialausgaben	Bevölkerung/Wohnen	Entwicklung
Traditionelle (Werthaltung)	Bevölkerungsanteil	mit traditioneller Grundorientierung	2019	Sinus-Geo-Milieus®	1.50	Beständigkeit	Bevölkerung/Wohnen	Stand
Unterschicht	Bevölkerungsanteil	aus Unterschicht	2019	Sinus-Geo-Milieus®	1.50	Sozialausgaben	Bevölkerung/Wohnen	Stand
Kaufkraft pro Person	Steuerbares Einkommen	In Franken pro Jahr	2015	AZ-Direkt	2.50	Steuerkraft	Wirtschaft/Standort	Stand
Zweitwohnungsanteil	Anteil Zweitwohnungen	In % des Wohnungsbestandes	2019	ARE	1.75	Saisonale Belegung	Bevölkerung/Wohnen	Stand
Wohnungsbau	Wohnungszugang	In % des Wohnungsbestandes	Mittelwert 2011–2016	BFS	2.50	Dynamik	Bevölkerung/Wohnen	Entwicklung
Geschäftsbaulandpreis	Wachstum Baulandpreise in %	Geschäftsnutzung	2019 gegenüber 2010	Wüest Partner	0.50	Attraktivität	Bevölkerung/Wohnen	Entwicklung
Flächenpotenzial Wohnen	Verdichtungspotenzial in %	Bevölkerung	bis 2040	Wüest Partner	1.00	Flächenpotenzial	Bevölkerung/Wohnen	Entwicklung
Erschwingliches wohnen	Anteil preisgünstige Mieten (Inserate)	Angebot (Single, Paare, Familien)	Mittelwert 2016–2018	Wüest Partner	1.50	Resilienz	Bevölkerung/Wohnen	Stand
Baulandpreis	Baulandpreise in CHF/m ²	Einfamilienhaus	2019	Wüest Partner	1.75	Attraktivität	Bevölkerung/Wohnen	Stand
Baulandpreiswachstum	Wachstum der Baulandpreise	Einfamilienhaus	Annualisiert 2010–2019	Wüest Partner	1.50	Dynamik	Bevölkerung/Wohnen	Entwicklung
Flächenpotenz. Geschäft	Verdichtungspotenzial in %	Beschäftigte	bis 2040	Wüest Partner	0.50	Flächenpotenzial	Bevölkerung/Wohnen	Entwicklung
Baulandreserven	Anteil unbebaute Bauzonen	In % der Bauzonenfläche	2017	ARE Bauzonenstatistik	2.00	Potenzial	Bevölkerung/Wohnen	Entwicklung
Suchabos pro Inserat	Wohnungsmarktspannung	Suchabos pro Inserat		RealMatch, Wüest Partner	0.50	Wohnungsmarktspannung	Bevölkerung/Wohnen	Stand
Beschäftigungswachstum	Beschäftigungswachstum		2020–2040	Wüest Partner	2.75	Wachstum	Wirtschaft/Standort	Entwicklung
Beschäftigungsdichte	Beschäftigungsdichte	Beschäftigte VZÄ pro Einwohner	2018	BFS STATENT, STATPOP	2.75	Wirtschaftsstandort	Wirtschaft/Standort	Stand
Binnenpendler	Binnenpendler	Arbeitsort = Wohnort	2016–2018	BFS Strukturhebung	2.25	Lebensqualität	Wirtschaft/Standort	Stand
Wertschöpfung	Wertschöpfung in CHF pro Jahr	pro Beschäftigten VZÄ	2018	Wüest Partner	2.75	Steuerkraft	Wirtschaft/Standort	Stand
Tourismus	Logiernächte, Besch. Gastgewerbe		2019	BFS HESTA	2.75	Tourismusmagnet	Wirtschaft/Standort	Stand
Traditionswirtschaften	Beschäftigungsanteil	Primärsektor und Bauwirtschaft	2018	BFS STATENT	2.00	Solidität	Wirtschaft/Standort	Stand
Staatsnahe Betriebe	Beschäftigungsanteil	Verwaltung, Bildung, Heime, Sozialw., Energie, Versorgung	2018	BFS STATENT	1.75	Resilienz	Wirtschaft/Standort	Stand
Zupendler Beschäftigte	Zupendler in % der Beschäftigten			BFS	1.25	Wirtschaft	Wirtschaft/Standort	Stand
Steuern	Steuern in % des Einkommens	Verheiratet, 2 Kinder, 120'000 Bruttoeinkommen	2015	AZ-Direkt, ESTV	1.25	Sozialabgabe	Wirtschaft/Standort	Stand
Erreichbare Beschäftigte	Erreichbarkeit Beschäftigte	Indikator aus MIV, ÖV innert 30, 60, 90 Minuten	2016	TransSol, BFS, Wüest Partner	2.00	Erschlossen	Wirtschaft/Standort	Stand
Gemeindehöhe	Meter über Meer		2019	Swisstopo	1.25	Erschlossen	Wirtschaft/Standort	Stand
Standortattraktivität	Standortattraktivität	Ratingwert: 5 sehr attraktiv, 1 wenig attraktiv	2020	Wüest Partner	2.75	Attraktivität	Wirtschaft/Standort	Stand
Erreichbare Einwohner	Erreichbarkeit Einwohner	Indikator aus MIV, ÖV innert 30,60, 90 Minuten	2016	TransSol, BFS, Wüest Partner	2.25	Erschlossen	Wirtschaft/Standort	Stand
Wohneigentumspreise	Angebotspreise in CHF/m ²	Mittelwert des Medians EFH und EWG	2019	Wüest Partner	1.00	Wohnungsmarktspannung	Bevölkerung/Wohnen	Stand
Wohnungsmieten Wachstum	Wachstum Angebotsmieten	Wachstum des Medians der Gemeinde	2019 gegenüber 2010	Wüest Partner	1.50	Wohnungsmarktspannung	Bevölkerung/Wohnen	Entwicklung

