

Die Vernetzung Wiens mit den Städten Europas

David Zenz

Beauftragt von



**Stadt
Wien**



Die Vernetzung Wiens mit den Städten Europas

DAVID ZENZ

David Zenz ist Statistiker am Wiener Institut für Internationale Wirtschaftsvergleiche (wiiw).

Die Durchführung der Studie wurde durch finanzielle Unterstützung der Kulturabteilung der Stadt Wien (MA 7) ermöglicht.

Zusammenfassung

Wir stellen ein Maß für die Beziehung zwischen zwei Städten/Regionen basierend auf Suchanfragen vor, ausgehend von Merkmalen der Suchanfragen-Zeitreihen nach Zerlegung der Zeitreihe mittels STL (Komponentenzerlegung mittels lokaler linearer Kernregression). Grundlage für das Maß sind einerseits die Eigenschaft 'Trendstärke', welches die Stärke des zugrundeliegenden Trends (egal ob steigend oder fallend) der Zeitreihe beschreibt, sowie das Feature 'linearity' der letzten fünf Jahre, welches uns die Richtung des Trends gibt. Die Kombination aus diesen Features der beiden Richtungen der Suchanfragen gibt uns ein Maß, welches für die Analyse der Entwicklung des vorgestellten Beziehungsmaßes über den Beobachtungszeitraum 2004-2020 in unterschiedlichen Suchkategorien zwischen zwei Städte/Regionen verwendet werden kann. Wir präsentieren Beispiele basierend auf Wien als point-of-interest im Kontext 'Wien und die Städte Europas', und schlagen ein Dashboard mit den verwendeten Indikatoren für Politik-Entscheidungen vor.

Schlüsselwörter: Zeitreihenanalyse, Big Data, Google Trends, Suchanfragen, Politik

JEL Klassifikation: C49, C80, C82, C87, C88, M30, R00, Z10, Z30

Abstract

We introduce a measure of linkage for the relationship between cities/regions, based on time series features of search engine queries. The used features are backed by time series decomposition using STL, i.e. seasonal and trend decomposition using Loess, precisely the strength of the trend and the linearity of a time series. The combination of these two features for both sides of search interest, e.g. the search interest for a certain topic in the city of Berlin based on search queries posed in Vienna, allows for the analysis of the development of this computed measure of linkage for the period 2004-2020 in various search engine categories provided by Google Trends between cities/regions in Europe. We then present examples based on the city of Vienna as a point-of-interest for selected topics and propose a dashboard for policy decisions.

Keywords: Time Series Analysis, Big Data, Google Trends, Search Engine Queries, Policy

JEL classification: C49, C80, C82, C87, C88, M30, R00, Z10, Z30

INHALT

Einleitung	1
Nutzung von Google Trends	2
Untersuchungsmethode und Wien-Bezug	3
Daten	4
Europäische Städte	4
Such-Kategorien	5
Zeitraum.....	5
Ausreißer und Brüche in den Zeitreihen	5
Deskriptive Analyse	8
Anzahl der Beobachtungen	8
Analyse	12
Zeitreihenanalyse	12
Trendstärke.....	12
Stärke der Saisonalität.....	12
Überblick.....	13
Maß für beidseitige Trendstärke	18
Beispiel Travel	20
Beispiel Business & Industrial.....	26
Vorschlag Dashboard	31
Schlussfolgerungen	34
Appendix	35
Kategorien	35
Trend & Saisonalität	35
Literatur	36

TABELLEN UND ABBILDUNGEN

Tabelle 1 / Übersicht der Variablen inkl. Ausprägungen	10
Abbildung 1 / Städte und Regionen	4
Abbildung 2 / Suchanfragen Berlin nach Wien in Kategorie "All categories"	6
Abbildung 3 / Suchanfragen in Kategorie "All categories" aus Berlin, Lissabon, Zagreb.....	6
Abbildung 4 / Fünf Städte/Regionen mit den meisten Beobachtungen als Quelle der Suchanfragen	8
Abbildung 5 / Fünf Städte/Regionen mit den wenigsten Beobachtungen als Quelle der Suchanfragen ...	9
Abbildung 6 / Fünf Städte/Regionen mit den meisten Beobachtungen als Ziel der Suchanfragen.....	9
Abbildung 7 / Fünf Städte/Regionen mit den wenigsten Beobachtungen als Ziel der Suchanfragen	9
Abbildung 8 / Wordcloud 50 häufigste Suchbegriffe nach Wien aus den Städten Europas.....	10
Abbildung 9 / Wordcloud 50 häufigste Suchanfragen aus Wien in die Städte Europas.....	11
Abbildung 10 / Trendstärke versus Stärke Saisonalität über alle Kategorien aus Deutschland, Italien und Frankreich nach Wien	13
Abbildung 11 / Trendstärke versus Stärke Saisonalität über alle Kategorien aus Deutschland, Italien und Frankreich nach Wien	14
Abbildung 12 / Städte/Regionen absteigend sortiert nach Trendstärke über alle Kategorien.....	15
Abbildung 13 / Städte/Regionen aufsteigend sortiert nach Trendstärke über alle Kategorien.....	15
Abbildung 14 / Städte/Regionen absteigend sortiert nach Stärke der Saisonalität über alle Kategorien.	16
Abbildung 15 / Städte/Regionen aufsteigend sortiert nach Stärke der Saisonalität über alle Kategorien	16
Abbildung 16 / Trendstärke und Stärke der Saisonalität über alle Kategorien.....	17
Abbildung 17 / Maß B für alle Kategorien, Ausschnitt Top 5 und Bottom 5	19
Abbildung 18 / Trendstärke versus Stärke der Saisonalität von Suchanfragen aller Kategorien von und nach Wien	20
Abbildung 19 / Trendstärke Bratislava, Paris, Rom und Galati	21
Abbildung 20 / Tabellen der Trendstärke für Bratislava, Paris, Rom und Galati.....	21
Abbildung 21 / Maß B für Paris, Bratislava, Rom und Galati	22
Abbildung 22 / Maß B für die Kategorie Travel über alle Städte/Regionen.....	22
Abbildung 23 / Ausschnitt Maß B für Top 5 und Bottom 5	23
Abbildung 24 / Wordcloud der Suchbegriffe in der Kategorie Travel	23
Abbildung 25 / Maß B für alle Städte/Regionen in der Kategorie Travel im Jahr 2007	24
Abbildung 26 / Maß B für Städte/Regionen in Italien in der Kategorie Travel über die Zeit.....	25
Abbildung 27 / Maß B für Städte/Regionen in Spanien in der Kategorie Travel über die Zeit	25
Abbildung 28 / Maß B für Städte/Regionen Tschechien, Slowakei und Ungarn in der Kategorie Travel über die Zeit	26
Abbildung 29 / Trendstärke versus Stärke der Saisonalität in der Kategorie Business & Industrial von und nach Wien	26
Abbildung 30 / Maß B in der Kategorie Business & Industrial über alle Städte/Regionen	27
Abbildung 31 / Ausschnitt Maß B in der Kategorie Business & Industrial für Top 5 und Bottom 5	27
Abbildung 32 / Wordcloud der Suchanfragen in der Kategorie Business & Industrial	28
Abbildung 33 / Maß B in der Kategorie Business & Industrial über alle Städte/Regionen im Jahr 2007 .	29
Abbildung 34 / Maß B für Städte/Regionen in Italien in der Kategorie Business & Industrial über die Zeit	29

Abbildung 35 / Maß B für Städte/Regionen in Deutschland in der Kategorie Business & Industrial über die Zeit.....	30
Abbildung 36 / Maß B für Städte/Regionen in Tschechien, Slowakei und Ungarn in der Kategorie Business & Industrial über die Zeit.....	30
Abbildung 37 / Aktuell positivste Kategorien versus aktuell positivste Städte/Regionen.....	31
Abbildung 38 / Aktuell negativste Kategorien versus aktuell negativste Städte/Regionen.....	32
Abbildung 39 / Maß B in allen Kategorien für Berlin über die Zeit.....	33
Abbildung 40 / Verwendete Kategorien	35
Abbildung 41 / Suchanfragen nach Wien mit "trend_strength"=0.....	35

Einleitung

Große Datenquellen („Big Data“) wie Google Trends werden seit dem Start im Jahr 2006 in der Forschung genutzt, mit dem Vorteil, dass die von Google Inc. bereitgestellten Daten leicht verfügbar und vorgefertigt sind, im Gegensatz zu Big Data im eigentlichen Sinn, d.h. große und unstrukturierte Daten, die nicht mit einem herkömmlichen Computer verarbeitet werden können (Varian 2014). Google Trends wird seither verwendet, um *„soziale Veränderungen zu verstehen und Vorhersagen zu treffen“* (Jun, Yoo, and Choi 2018) und ist eines der führenden Beispiele für die Anwendung großer Datenmengen in der (nicht nur ökonomischen) Forschung. Wir schlagen im Folgenden vor, diese große Datenquelle zu nutzen, um die Vernetzung von Wien mit anderen Städten, Regionen und Ländern Europas auf verschiedenen inhaltlichen Ebenen zu messen und visuell darzustellen. Unseres Wissens nach wurde dies bisher nicht gemacht. Zu den wenigen Studien, die Google Trends im Wiener Kontext verwenden zählen beispielsweise eine Analyse der Tourismus-Nachfrage für Wien und Barcelona, Österreich und Belgien (Önder 2017) oder eine Prognose Methamphetamin-bezogener Verbrechen auf regionaler und nationaler Ebene in Österreich, Deutschland und der Schweiz (Gamma et al. 2016).

Nutzung von Google Trends

Google Trends ist ein Service der Google Inc. und bietet die Möglichkeit, Suchmaschinen-Daten über einen Suchvolumenindex im Zeitverlauf zu analysieren. Dieser Zeitreihenindex basiert auf Netzsuchanfragen und zeigt, wie häufig ein bestimmter Suchbegriff im Vergleich zu allen anderen Suchbegriffen in verschiedenen Regionen und Sprachen verwendet wird, wobei die Daten bis zum 1. Januar 2004 zurückreichen. Laut Google Inc. handelt es sich bei den Daten von Google Trends um anonymisierte, unverzerrte Suchanfragen, die mittels maschineller Lernklassifizierung in vordefinierte Themen mit unterschiedlicher Granularität kategorisiert werden. Insgesamt gibt es etwa 600 Kategorien, in denen der Kontext einer Suchanfrage über das verfügbare Tool eingeschränkt werden kann (Scott and Varian 2013). Der Algorithmus schließt automatisch Suchanfragen von sehr wenigen Personen aus, so genannte "Low-Volume-Suchen" oder unbeliebte Begriffe, d.h. Google Trends stellt nur so genannte Populärbegriffe dar, wobei Suchanfragen unterhalb eines bestimmten (aber nicht öffentlich bekannten) Schwellenwerts mit 0 gekennzeichnet sind. Darüber hinaus sind doppelte Suchen und Suchanfragen mit Sonderzeichen (z.B. Anführungszeichen) ausgeschlossen.

Der resultierende Index ist proportional zu Zeit und Ort der Suchanfrage, wobei jeder Betrag einer Suchanfrage durch die Gesamtsuche nach Ort und Zeitbereich dividiert wird, um die relative Popularität darzustellen, welche in weiterer Folge auf einen Bereich von 0 bis 100 normiert werden. Dadurch wird sichergestellt, dass Regionen mit dem höchsten Suchanfragenvolumen nicht von Natur aus am höchsten eingestuft werden (Google 2019). Weiters ist zu beachten, dass die Werte für den Google Trends Index täglich neu berechnet werden, und das auf einer Basis einer 1%-igen zufälligen Stichprobe, welche für jede Berechnung neu gezogen wird. Es ist also nicht unwahrscheinlich, dass bei Abfragen von Google Trends an unterschiedlichen Zeitpunkten, auch unterschiedliche Werte in den Google Trends Zeitreihen ausgegeben werden. Laut Google ist diese Methode aber robust und garantiert, wenn auch keine direkte Vergleichbarkeit, dass der zugrundeliegende Trend und die Saisonalität der Zeitreihe annähernd ident ist (Google 2019).

Untersuchungsmethode und Wien-Bezug

Um die Vernetzung einer Stadt mit anderen Städten, Regionen oder Ländern zu messen, nehmen wir an, dass dieses Interesse, wie in unserem Fall durch Suchanfragen belegt wird und die Beziehung zwischen den Städten widerspiegelt. Hierbei werden beide Seiten der Suche berücksichtigt, d.h. Suchen in Wien nach anderen Städten, Regionen oder Ländern und umgekehrt. Wie von Horák et al. (2013) vorgeschlagen, kann mit Hilfe von Suchmaschinen-Daten ein Netzwerk aufgebaut und die zugrunde liegende Beziehung zwischen den Städten und Regionen analysiert werden. Dies schließt die traditionelle (geografische) Netzwerkanalyse ein, kann aber auch nichtlineare Methoden wie Klassifikations- und Clustering-Algorithmen zur Datenaufbereitung und -analyse erfordern (Varian 2014). Wir verwenden Google Trends Daten für verschiedene Kategorien (wie z.B. Wissenschaft, Kultur, Verkehr, Wirtschaft, für alle Städte Europas, für welche aussagekräftige Daten zur Verfügung stehen). Das Suchinteresse sowie die dazugehörigen Top-Suchbegriffe werden in einem ersten Schritt für beide Seiten der Beziehung gesammelt – z.B. die Suche im geografischen Gebiet Wien für die Stadt München in der Kategorie Wissenschaft. Dieser Schritt wird für die andere Seite wiederholt, d.h. man sucht in München nach Wien in der Kategorie Wissenschaft, um die Beziehung aus Sicht der anderen Seite abzubilden.

Um die Konnektivität Wiens zu anderen Städten in Europa in verschiedenen Themenbereichen zu messen, soll ein einziges Maß für alle Kategorien berechnet werden. Dabei wird ein Durchschnitt der abgerufenen Google Trends-Indizes (bei Verwendung von Unterkategorien oder bestimmten Suchbegriffen) verwendet, wobei 1 der höchste Wert und -1 der untere Schwellenwert ist. Anschließend erlauben uns diese Indizes das Erstellen einer Wärmekarte (heat map) für verschiedene Kategorien (und Zeiträume). Diese zeigt uns, in welchen Bereichen Wien mehr oder weniger mit den anderen Städten verbunden ist. Das erlaubt zum einen auf bestehende Verbindungen und Städtepartnerschaften besser eingehen zu können und zum anderen Verbindungen mit nur schwach ausgeprägten Intensitäten beziehungsweise Interesse auf geringem oder stagnierendem Niveau auf verschiedenen Ebenen zu stärken.

Die Betrachtung im Zeitablauf bringt die Dynamik der Interessens-Trends zwischen Wien und anderen Städten Europas zur Geltung. Neben einer kompakten schriftlichen Analyse der Wärmekarten stellen die heat maps an sich einen wesentlichen Output des Projektes dar. Diese können auch als eigenständige, selbsterklärende Bilder in sozialen Medien verbreitet werden und eignen sich daher gut um Wien im Hinblick auf andere Städte auf verschiedenen inhaltlichen Ebenen zu verorten.

Daten

EUROPÄISCHE STÄDTE

Daten zu den hier verwendeten europäischen Städten basieren unter anderem auf

- › Auflistung von Städten in der Europäischen Union auf Wikipedia,
- › Auflistung von Städten in der Europäischen Union nach Population innerhalb der Stadtgrenzen,
- › Auflistung von Ballungsräumen in Europa,
- › Auflistung von Stadtgebieten in Europa,
- › sowie auf Daten zu Städten aus dem *R* package *world*.

Zusammengefügt ergibt sich ein potentieller Datensatz von 167 Städten aus 52 Ländern. Die Abfrage der Daten von Wikipedia erfolgte automatisiert und kann dementsprechend bei einer Änderung des Artikels für einen aktuelleren Datensatz ausgeführt werden.

Abbildung 1 / Städte und Regionen

Show entries Search:

	Stadt	Region
1	Córdoba	Andalusien
2	Málaga	Andalusien
3	Seville	Andalusien
4	Bari	Apulien
5	Nur-Sultan	Aqmola
6	Zaragoza	Aragonien
7	Bilbao	Autonome Gemeinschaft Baskenland
8	Madrid	Autonome Gemeinschaft Madrid
9	Karlsruhe	Baden-Württemberg
10	Mannheim	Baden-Württemberg

Showing 1 to 10 of 130 entries Previous 2 3 4 5 ... 13 Next

Quelle: Google Trends.

Diese Städte mussten nun zu den in Google verfügbaren *geo*- und *topic*-codes gematcht werden. Hier stellte sich heraus, dass nicht alle Städte einen eigenen *geo*- beziehungsweise *topic*-code besitzen, und dementsprechend manche Städte bei den Suchanfragen als Region dargestellt werden. Dies kann sich aber unterschiedlich äußern. Ein Beispiel hierfür wäre *Paris*, welches zwar einen *topic*-code besitzt: *05qtj*, allerdings für die Anfrage nach Wien in diesem Projekt als Region mit dem *geo*-code *FR-J* für *Île-de-France* aufscheinen wird. Ein noch drastischeres Beispiel ist die Region *Nordrhein-Westfalen*, da

die Städte *Bielefeld, Bochum, Bonn, Cologne, Dortmund, Duisburg, Düsseldorf, Essen, Münster, Neuss, Wuppertal* keinen eigenen *geo-code* besitzen, sondern alle denselben: *DE-NW*. Eine Auflistung der 130 Städte mit geteiltem *geo-code*, beziehungsweise mit einem *geo-code* für eine ganze Region ist in der folgenden Tabelle zu sehen.

Wir sehen also einerseits Städte, die nur als Region verfügbar sind wie beispielsweise Paris, aber andererseits auch, dass in Regionen mehrere Städte sowie das Umland in einer Region zusammengefasst sein können. Dies erschwert die Analyse der Interessen gegenüber diesen Städten, dementsprechend sind die Analysen der betreffenden Regionen immer vor dem Hintergrund, dass es sich in diesem Fall eben um eine Region und nicht um eine Stadt handelt, zu sehen und mit entsprechender Vorsicht zu interpretieren.

SUCH-KATEGORIEN

Insgesamt wurden 75 Suchkategorien abgefragt, wobei hier die Suchkategorien von Google Trends bis zur 2. Ebene nach *Alle Kategorien* ausgewählt wurden (siehe Appendix). Google Trends selbst würde die Möglichkeit bieten Suchanfragen für 1132 Kategorien abzufragen. Alle Suchkategorien abzurufen war zeitbedingt nicht möglich, mit etwas Geduld wäre so ein Unterfangen aber prinzipiell durchführbar. Es stellt sich aber natürlich auch die Frage wie viele Suchanfragen-Ergebnisse bei einer granularen Filterung aus/nach Städten überhaupt vorhanden wären. Wir sehen teilweise in den geladenen Daten schon gewisse Einschränkungen aufgrund der geografischen Kategorisierung der Suchanfragen nach Städten/Regionen. Insofern ist die Restriktion auf die 2. Ebene von Suchanfragenkategorien auch sinnvoll mit Hinblick auf die Analyse.

ZEITRAUM

Als Zeitraum der Auswertung wurde das von Fenster 01.01.2004-01.01.2020 (Start der Zeitreihen von Google Trends) bis 01.01.2020 gewählt. Der Zeitraum der Abfragen fand im Zeitraum von Mitte März bis Ende April 2020 statt.

Die Abfrage der Daten dauerte kumulativ etwa über den Zeitraum von etwa 2 Wochen an, da nach jeder Abfrage etwa 60 Sekunden pausiert werden musste, um nicht für eine gewisse Zeit von der Datenabfrage ausgeschlossen zu werden.

AUSREIßER UND BRÜCHE IN DEN ZEITREIHEN

Die einzelnen Zeitreihen wurden automatisiert unter Verwendung des *R* packages *gtrendsR* abgefragt und für die weitere Verwendung vorbereitet. Des Weiteren wurden mittels des *R* packages *AnomalyDetection* von *twitter* Ausreißer in den einzelnen Zeitreihen unter vordefinierten Regeln automatisch behandelt, und die Originalwerte sowie Zeitpunkte abgespeichert. Wurde eine Beobachtung als Ausreißer identifiziert, so wurde diese als *missing* behandelt und mittels linearer Interpolation auf Basis der kompletten Zeitreihe approximiert. Dieser Schritt war notwendig, da einerseits immer wieder fehlende Beobachtungen, aber oft auch unplausible Datenpunkte in den Zeitreihen zu finden waren. Manchmal wurde die Ausreißer aber auch durch Brüche in der Berechnung der einzelnen Google Trends Zeitreihen, meistens aufgrund von geografischen Neuordnungen, ausgelöst. Diese Brüche

sind zwar in der Google Trends GUI eingezeichnet, können aber leider nicht automatisiert abgefragt werden. Nach Behandlung wurden die betroffenen Zeitreihen wieder zwischen 0 und 100 skaliert, um konsistent mit den Original-Reihen von *Google Trends* zu bleiben.

Die nachstehende Tabelle zeigt beispielhaft die vorbereiteten Daten nach Ausreißer-Behandlung von Suchinteresse in der Kategorie "All categories" nach Wien aus der Stadt Berlin (193 Beobachtungen).

Abbildung 2 / Suchanfragen Berlin nach Wien in Kategorie "All categories"

Show entries Search:

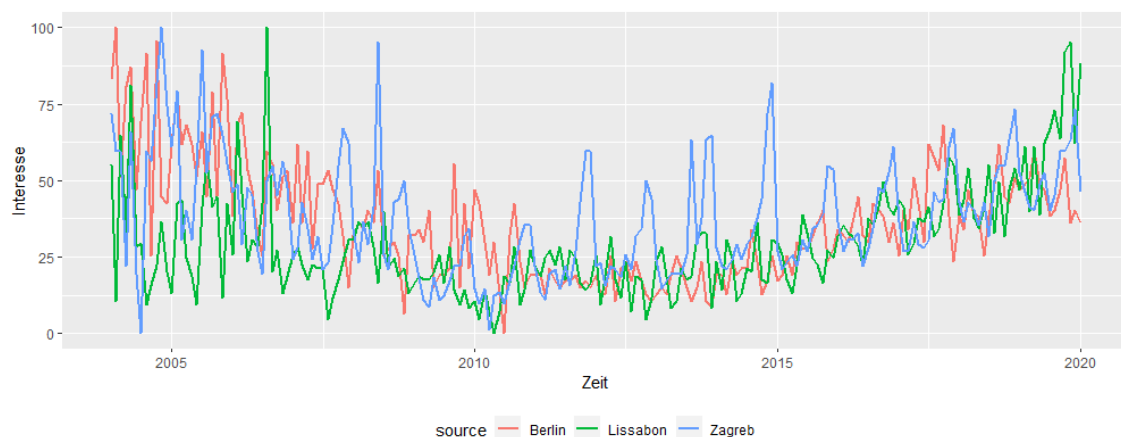
	Quelle	Kategorie	Datum	Google Trends Index
1	Berlin	All categories	2004-01-01	83
2	Berlin	All categories	2004-02-01	100
3	Berlin	All categories	2004-03-01	51
4	Berlin	All categories	2004-04-01	81
5	Berlin	All categories	2004-05-01	87
6	Berlin	All categories	2004-06-01	47
7	Berlin	All categories	2004-07-01	69
8	Berlin	All categories	2004-08-01	91
9	Berlin	All categories	2004-09-01	26
10	Berlin	All categories	2004-10-01	96

Showing 1 to 10 of 193 entries Previous 2 3 4 5 ... 20 Next

Quelle: Google Trends, eigene Berechnung.

Wenn man diese Zeitreihe nun mittels Linien-Diagramm visualisiert, erhält man für die Städte *Berlin*, *Lissabon* und *Zagreb* den folgenden Graphen, in welchem die Linie den Google Trends Index der Suchanfragen (das relative Suchinteresse) der Kategorie *all Categories* über die Zeit darstellt.

Abbildung 3 / Suchanfragen in Kategorie "All categories" aus Berlin, Lissabon, Zagreb



Quelle: Google Trends, eigene Berechnung.

Dieser Graph ist interaktiv, das heißt das Filtern von Städten, herein und heraus zoomen, sowie der Export der Zeitreihe als Bild ist hier möglich. Wir sehen ähnliche Trends, aber durchaus unterschiedliches Suchinteresse aus den drei gewählten Städten mit Bezug zu Wien. Beispielhaft ist hier der Effekt der UEFA Europameisterschaft 2008, der sich vor allem bei den Suchanfragen aus Zagreb zeigt, wohingegen das Suchinteresse aus Berlin und Lissabon hier keine große Veränderung zeigt.

Deskriptive Analyse

In einem ersten Schritt werden die Daten einer deskriptiven Analyse unterzogen. Der geladene Datensatz umfasst insgesamt 1961654 Beobachtungen über einen Zeitraum von Jänner 2004 bis Jänner 2020, wobei jede Quelle im Schnitt 12,086 Beobachtungen, und jedes Ziel 12,107 aufweist. Nachfolgend wird auf die Quellen beziehungsweise Ziele mit den meisten und den wenigsten Beobachtungen eingegangen.

ANZAHL DER BEOBACHTUNGEN

QUELLEN

Bei den fünf Quellen mit den meisten Beobachtungen zeigt sich ein homogenes Bild von gleicher Anzahl von Beobachtungen über alle aufgelisteten Städte/Regionen. Dies ist wenig überraschend, da Google Trends für viele Städte einen kompletten Datensatz liefern kann, vorausgesetzt das Sample an Suchanfragen ist groß genug. Dies kann bei kleineren Städten beziehungsweise Regionen der Fall sein, wie wir auch in der nachfolgenden Tabelle sehen können.

Abbildung 4 / Fünf Städte/Regionen mit den meisten Beobachtungen als Quelle der Suchanfragen

Show entries Search:

	Quelle	Anzahl
1	Südholland	13898
2	Griechenland	13896
3	Berlin	13896
4	Nordrhein-Westfalen	13896
5	Region Brüssel-Hauptstadt	13896

Showing 1 to 5 of 5 entries Previous Next

Quelle: Google Trends.

Bei den Städten/Regionen als Suchanfragen-Quelle mit den wenigsten Beobachtungen in unserem Datensatz sticht hervor, dass sie Städte/Regionen aus Spanien (Land Valencia, Katalonien, Autonome Gemeinschaft Baskenland, Andalusien und die Kanarischen Inseln) darstellen.

Der Grund für dieses Artefakt ist nicht ganz klar, jedoch gab es bei Abfragen mit dem *R* package *gtrendsR* öfters Probleme mit spanischen Zielen und Quellen. Einerseits sind anscheinend manche geocodes, sowie manche topic-codes mit Bezug auf Spanien fehlerhaft (implementiert). Möglicherweise könnte dies auch ein Grund für die geringe Anzahl an Beobachtungen in Hinblick auf spanische Städte sein.

Abbildung 5 / Fünf Städte/Regionen mit den wenigsten Beobachtungen als Quelle der Suchanfragen

Show entries Search:

	Quelle	Anzahl
1	Land Valencia	193
2	Katalonien	193
3	Autonome Gemeinschaft Baskenland	193
4	Andalusien	193
5	Kanarische Inseln	193

Showing 1 to 5 of 5 entries Previous Next

Quelle: Google Trends.

ZIELE

Bei den fünf Zielen mit den meisten Beobachtungen sehen wir, wie im Falle der Quellen, wieder ein ausgewogenes Bild, wohingegen bei den fünf Zielen mit den wenigsten Beobachtungen, wieder spanische Städte (Malaga, Murcia, Las Palmas) aufscheinen.

Abbildung 6 / Fünf Städte/Regionen mit den meisten Beobachtungen als Ziel der Suchanfragen

Show entries Search:

	Ziel	Anzahl
1	Amsterdam	13896
2	Barcelona	13896
3	Berlin	13896
4	Bratislava	13896
5	Budapest	13896

Showing 1 to 5 of 5 entries Previous Next

Quelle: Google Trends.

Abbildung 7 / Fünf Städte/Regionen mit den wenigsten Beobachtungen als Ziel der Suchanfragen

Show entries Search:

	Ziel	Anzahl
1	Málaga	5983
2	Murcia	8685
3	Aarhus	8878
4	Craiova	8878
5	Las Palmas	8878

Showing 1 to 5 of 5 entries Previous Next

Quelle: Google Trends.

Analyse

ZEITREIHENANALYSE

In der traditionellen Zeitreihenanalyse sind Trend und Saisonalität wichtige Faktoren mit Hilfe derer Zeitreihen in ihre Einzelkomponenten zerlegt werden können:

$$y_t = T_t + S_t + R_t$$

wobei y_t die zugrundeliegende Zeitreihe darstellt, T_t die geglättete Trend-Komponente, S_t die Saisonalität, und R_t die Rest- beziehungsweise irreguläre Komponente darstellt. Diese Zerlegung wird auch *STL* genannt, für "*Seasonal and Trend decomposition using Loess*", welche von Cleveland et al. (1990) vorgestellt wurde. Auf Basis dieser Zerlegung stellt das *R* Package *feasts* (für *Feature Extraction and Statistics for Time Series*.) die Möglichkeit bereit, diverse weitere Eigenschaften, wie die Stärke des Trends, Saisonalität, spikiness uvm., von den Grunddaten automatisiert zu berechnen. Nachfolgend konzentrieren wir uns vor allem auf die beiden Maße *Trendstärke* und *Saisonalität*, da unsere Daten keine direkten Rückschlüsse auf die tatsächlichen absoluten Zahlen von Suchanfragen darstellen, sondern ein relatives Maß, welches aber durchaus valide Trend- und Saisonalitäts-Analysen zulässt.

TRENDSTÄRKE

Die *Trendstärke* basiert auf der Annahme, dass saisonal-bereinigte Zeitreihen wesentlich mehr Variation in den Daten aufweisen als die Rest- oder irreguläre Komponente nach Zerlegung, also $\frac{Var(R_t)}{Var(T_t + R_t)}$ relativ klein ist. Für Daten mit keinem oder wenig Trend sollten die beiden Varianzen dementsprechend approximativ ident sein. Die *Trendstärke* F_T wird also wie folgt definiert:

$$F_T = \max\left(0, 1 - \frac{Var(R_t)}{Var(T_t + R_t)}\right)$$

was ein Maß für die *Trendstärke* im Intervall (0,1) garantiert.

STÄRKE DER SAISONALITÄT

Die *Stärke der Saisonalität* ist ähnlich definiert, allerdings basierend auf der Saisonalitäts-Komponente (trend-bereinigte Daten) und nicht auf der Saisonalitäts-bereinigten Komponente:

$$F_S = \max\left(0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)}\right)$$

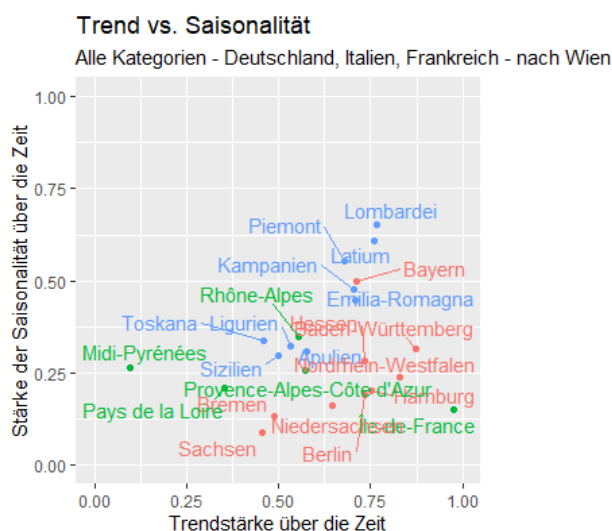
Aus der Definition wird klar, dass eine Zeitreihe mit F_S nahe oder gleich 0 keine Saisonalität beinhaltet, wohingegen eine Zeitreihe mit F_S nahe 1 eine starke Saisonalität enthält.

Maße für *Trendstärke* und *Stärke der Saisonalität* sind wichtige Maße, wenn man, wie im Falle dieses Projektes, eine große Anzahl von Zeitreihen hat und diese sinnvoll nach starken Trends und Saisonalität filtern möchte. Das Package *feasts* bietet noch weitere Auswertungsmöglichkeiten für Zeitreiheneigenschaften an wie beispielsweise *spikiness* (ein Maß für die Verbreitung von Spitzen in der Zeitreihe), *linearity* (ein Maß für die Linearität der Trend-Komponente), oder *curvature* (ein Maß für die Krümmung der Trend-Komponente). In diesem Projekt fokussieren wir uns auf die ersten zwei genannten Maße *Trendstärke* und *Stärke der Saisonalität*, um die Suchanfragen von und nach Wien aus den Städten Europas zu analysieren.

ÜBERBLICK

Wenn wir nun einen Blick auf den Zusammenhang von saisonaler Stärke und Trendstärke von Suchanfragen nach Wien aus den Ländern Deutschland, Italien und Frankreich werfen, zeigt sich, dass Suchanfragen aus Italien über den Beobachtungszeitraum einerseits eine im Vergleich relativ hohe Trendstärke, aber andererseits eine wesentlich höhere Stärke der Saisonalität ausweisen, als Suchanfragen aus Deutschland und Frankreich. Während Suchanfragen aus Deutschland eine ähnliche Trendstärke wie Italien aufweisen, scheint die Stärke der Saisonalität im direkten Vergleich zu Italien deutlich gedämpft. Frankreich hingegen zeigt sich ausgehend von den Suchanfragen aus den einzelnen Regionen homogen wenig saisonal, dafür aber relativ verstreut, was die Trendstärke der Suchanfragen nach Wien betrifft.

Abbildung 10 / Trendstärke versus Stärke Saisonalität über alle Kategorien aus Deutschland, Italien und Frankreich nach Wien



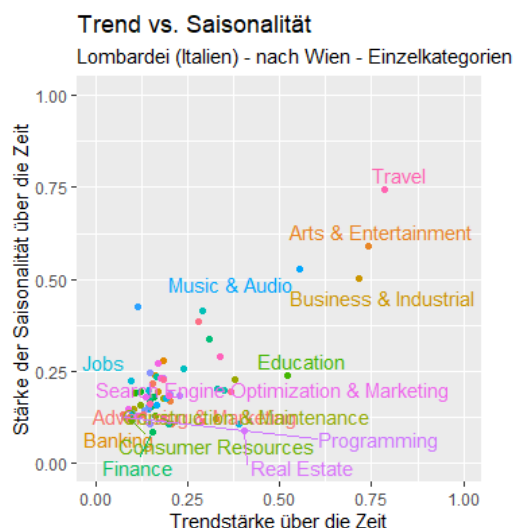
Quelle: Google Trends, eigene Berechnung.

Innerhalb der Länder zeigt sich, dass der Cluster um die Lombardei, den Piemont und das Latium in Italien, Bayern in Deutschland, und die Provinz Rhône-Alpes in Frankreich die höchste Stärke aufweisen, während Sachsen in Deutschland, die Provinz Île-de-France (Paris, Frankreich) und Sizilien (Italien) die geringste Stärke an Saisonalität zeigen. Trend-stark zeigen sich Île-de-France (Paris), Baden-Württemberg (Deutschland) und die Lombardei (Italien), während der schwächste Trend

bezüglich Suchanfragen nach Wien in allen Kategorien in den einzelnen Ländern in Midi-Pyrénées (Frankreich), Sachsen (Deutschland) und in der Toskana aufscheint.

Oben als statischer Plot inklusive Hervorhebung von Kategorien mit Trendstärke und Stärke der Saisonalität außerhalb der Schranken (0.1,0.5), sowie nachstehend nochmals als interaktiver Plot um die einzelnen Kategorien und deren Maße genau identifizieren zu können. Hier zeigt sich schön die starke Saisonalität der Suchanfragen aus der Lombardei nach Wien in der Kategorie *Travel*, sowie hohe Trendstärken in den Kategorie *Music & Audio*, *Arts & Entertainment* und *Business & Industrial*, wohingegen niedrige Stärken in beiden Maßen im Bereich *Jobs*, *Programming*, *Banking*, *Finance*, etc. zu finden sind. Die hohe Trendstärke über die Zeit seit 2004 könnte, vor allem in den Bereichen *Education*, *Business & Industrial* auf eine erhöhte Mobilität von Gütern und Studierenden hindeuten könnten, während Suchanfragen in den Kategorien wie *Jobs*, *Banking* und *Finance* einen stabilen Trend und kaum Saisonalität zeigen.

Abbildung 11 / Trendstärke versus Stärke Saisonalität über alle Kategorien aus Deutschland, Italien und Frankreich nach Wien



Quelle: Google Trends, eigene Berechnung.

TRENDSTÄRKE

Sieht man sich die Auswertung für die Trendstärke im Detail an, so zeigt sich ein interessantes Bild für die höchste Trendstärke nach Quelle und Kategorie: 9 von den 10 Zeitreihen mit der höchsten Trendstärke von Suchanfragen nach Wien stammen aus der Region Île-de-France (Paris), während die einzige Zeitreihe welche nicht aus Paris kommt, aus der Region Uusimaa (Helsinki) stammt. Bei beiden Regionen/Städten ist - nicht wirklich überraschend - die Kategorie *Travel* unter den Zeitreihen mit der höchsten Trendstärke. Aus der Region Île-de-France stammen aber auch Zeitreihen aus den Kategorien *Business & Industrial*, *Arts & Entertainment*, *Music & Audio*, welche schon bei den Zeitreihen aus der Lombardei in der Grafik oben hohe Werte für die Trendstärke auswiesen. Es ist zu erwähnen, dass alleine auf Basis der *Trendstärke* nicht ersichtlich ist, ob es sich dabei um einen zunehmenden oder abnehmenden Trend handelt. Zusätzlich muss zu der Zeitreihe von Helsinki (Uusimaa) ergänzt werden,

dass diese erst mit Juli 2013 startet und seitdem nur einen positiven Trend aufweist, was auch die hohe Trendstärke im Vergleich zu anderen Städten/Regionen erklärt.

Abbildung 12 / Städte/Regionen absteigend sortiert nach Trendstärke über alle Kategorien

Show entries Search:

	Quelle	Kategorie	Trendstärke
1	Île-de-France	Travel	0.95926
2	Île-de-France	Business & Industrial	0.93909
3	Île-de-France	Arts & Entertainment	0.93836
4	Île-de-France	Food & Drink	0.92619
5	Uusimaa	Travel	0.90975
6	Île-de-France	Shopping	0.90716
7	Île-de-France	Music & Audio	0.89959
8	Île-de-France	Apparel	0.88923
9	Île-de-France	Weather	0.87633
10	Île-de-France	People & Society	0.87398

Showing 1 to 10 of 10 entries Previous Next

Quelle: Google Trends, eigene Berechnung.

Abbildung 13 / Städte/Regionen aufsteigend sortiert nach Trendstärke über alle Kategorien

Show entries Search:

	Quelle	Kategorie	Trendstärke
1	Sachsen	Banking	0.04189
2	Sachsen	People & Society	0.04477
3	Woiwodschaft Kleinpolen	Internet & Telecom	0.05225
4	Kampanien	Computers & Electronics	0.05431
5	Ligurien	Apparel	0.05564
6	Bremen	Home & Garden	0.05658
7	Nordholland	Law & Government	0.05743
8	Woiwodschaft Kujawien-Pommern	Sports	0.05761
9	Flanders	Weather	0.05846
10	Berlin	Pets & Animals	0.05984

Showing 1 to 10 of 10 entries Previous Next

Quelle: Google Trends, eigene Berechnung.

Die Zeitreihen mit der niedrigsten Trendstärke, aber Trendstärke ungleich 0, umfassen Kategorien wie *Banking*, *People & Society*, *Internet & Telecom* etc. Diese Kategorien scheinen keine große Trend-Veränderung über die Zeit zu haben. Eine mögliche Erklärung könnte sein, dass Themen in diesem Bereich in den Städten so etwas wie Standard sind, hier also keine außergewöhnlichen Innovationen oder spezielle Themen über die vergangenen 16 Jahre aufgekommen sind - so etwas wie Themen die

ständig präsent, aber nicht auffallend breit diskutiert bzw. in veränderliche Form per Suchanfrage nachgefragt werden. Zeitreihen mit Trendstärke 0 sind hier Zwecks Übersichtlichkeit exkludiert, umfassen aber 272 Einträge (siehe Appendix).

STÄRKE DER SAISONALITÄT

Sieht man sich die Zeitreihen mit Hinblick auf die Saisonalität an, so ergibt sich ein interessantes Bild: unter den fünf Zeitreihen mit der höchsten Saisonalitäts-Stärke sind drei aus unmittelbarer Umgebung zu Wien (Budapest und Bratislava). Wenig überraschend ist, mit Hinblick auf die Politik in Ungarn, jedenfalls die Kategorie *Ethnic & Identity Groups* in *Budapest* mit einem hohen Wert für *Trendstärke* hier vertreten, wohingegen die anderen vier Kategorien leichter mit Saisonalität in Verbindung gebracht werden können.

Abbildung 14 / Städte/Regionen absteigend sortiert nach Stärke der Saisonalität über alle Kategorien

Show entries Search:

	Quelle	Kategorie	Stärke der Saisonalität
1	Lombardei	Travel	0.74428
2	Budapest	TV & Video	0.72813
3	Budapest	Ethnic & Identity Groups	0.71792
4	Bratislavský kraj	Food & Drink	0.70288
5	Zagreb	Performing Arts	0.67769

Showing 1 to 5 of 5 entries Previous Next

Quelle: Google Trends, eigene Berechnung.

Unter den Zeitreihen mit der niedrigsten Stärke an Saisonalität befinden sich jedenfalls Städte/Regionen (Paris, Krakau, Helsinki, Bydgoszcz und Wrocław), die etwas weiter entfernt von Wien sind, und das mit Kategorien die auch nicht sonderlich überraschend auf den hinteren Plätzen landen, wie *Consumer Electronics*, *Food & Drinks* etc.

Abbildung 15 / Städte/Regionen aufsteigend sortiert nach Stärke der Saisonalität über alle Kategorien

Show entries Search:

	Quelle	Kategorie	Stärke der Saisonalität
1	Île-de-France	Consumer Electronics	0.05246
2	Woiwodschaft Kleinpolen	Food & Drink	0.05396
3	Uusimaa	Jobs	0.05486
4	Woiwodschaft Kujawien-Pommern	Health	0.05562
5	Woiwodschaft Niederschlesien	Home & Garden	0.06149

Showing 1 to 5 of 5 entries Previous Next

Quelle: Google Trends, eigene Berechnung.

BLICK AUF DIE KATEGORIEN

Wenn man sich Trendstärke und Stärke der Saisonalität gemittelt über alle Kategorien von Suchanfragen nach Wien ansieht zeigt sich, dass bezüglich Trendstärke die Kategorien *Travel*, *Arts & Entertainment*, *Business & Industrial*, *Music & Audio*, *Weather* diejenigen Kategorien sind, die im Mittel den stärksten Trend haben, wohingegen *Search Engine Optimization & Marketing*, *Web Hosting & Domain Registration*, *Environmental Issues*, *Marketing Services*, *Manufacturing* im Mittel die geringste Trendstärke aufweisen. Bei der Stärke der Saisonalität gibt es hohe Werte für *Travel*, *Weather*, *Arts & Entertainment*, *Music & Audio*, *News*, und im Mittel niedrige Werte für *Finance*, *Jobs*, *Software*, *Insurance*, *Consumer Resources*. Die nachfolgende Tabelle ist dynamisch und lässt die auf- und absteigende Sortierung der Werte zu.

Abbildung 16 / Trendstärke und Stärke der Saisonalität über alle Kategorien

Show entries Search:

	Kategorie	Trendstärke	Stärke der Saisonalität
1	Home & Garden	0.20572	0.15873
2	Advertising & Marketing	0.11931	0.15991
3	Agriculture & Forestry	0.14425	0.17214
4	Antiques & Collectibles	0.13159	0.15582
5	Apparel	0.20102	0.15643
6	Arts & Entertainment	0.37464	0.22954
7	Autos & Vehicles	0.17785	0.15607
8	Banking	0.14041	0.15771
9	Beauty & Fitness	0.17478	0.15930
10	Books & Literature	0.16070	0.15769

Showing 1 to 10 of 71 entries Previous 2 3 4 5 ... 8 Next

Quelle: Google Trends, eigene Berechnung.

Niedrige Saisonalität in der Kategorie *Jobs* ist zwar etwas überraschend, lässt sich aber einfach durch eine konstante Nachfrage nach internationalen Jobs erklären, wohingegen man in den Suchanfragen innerhalb eines Landes zur Kategorie *Jobs* eine klare Saisonalität sehen kann, und dementsprechend auch ein hoher Wert bei der Stärke der Saisonalität erwartet werden kann. Auch ein niedriger Wert bei der Trendstärke zu *Environmental Issues* darf nicht überraschen, da einerseits nicht allgemeine Suchanfragen, also Suchanfragen ohne spezifischem geografischen Hintergrund, analysiert werden, sondern eben Suchanfragen von Wien in eine Stadt oder von einer Stadt nach Wien, und andererseits das Thema erst im letzten Jahr an Fahrt aufgenommen hat, sowie bei der Berechnung von der Trendstärke keine Gewichtung stattfindet: hohe Trendwerte im letzten Jahr also durch niedrige Werte der vergangenen 15 Jahre ausgemittelt werden.

MAß FÜR BEIDSEITIGE TRENDSTÄRKE

Für die Berechnung der beidseitigen Vernetzung beziehungsweise die Beziehung von Wien mit den Städten Europas, kombinieren wir die beiden Richtungen zu einem einzigen Maß, welches auf dem Intervall (0,1) skaliert sein soll. Dafür definieren wir das Beziehungsmaß B als

$$B = \sqrt{\frac{F_{TWien}^2 + F_{T-Wien}^2}{2}}$$

wobei F_{TWien} die Trendstärke der Zeitreihe aus Wien, und F_{T-Wien} die Trendstärke der Zeitreihen nach Wien bezeichnet. Dies erlaubt einerseits eine garantierte Skalierung im Intervall (0,1) und gibt im Falle des Vorkommens einer Zeitreihe mit Trendstärke $F_T = 0$ der Zeitreihe mit Trendstärke $F_T > 0$ etwas mehr Gewicht. Was allerdings bei der Berechnung und Interpretation zu beachten ist, ist dass wir bei Suchen nach Wien statt Städten zu einem Großteil Regionen in den Daten haben. Ein Aspekt, der dabei zu beachten ist, ist dass Regionen sich, abgesehen vom umliegenden Bezirk einer Stadt, auch aus mehreren Städte und deren umliegende Gebiete, zusammensetzen können. Um für diese Region ein Beziehungsmaß B berechnen zu können, muss die Abfrage aus Wien (in die Städte) in diesen Fällen zusammengefasst werden. Dafür definieren wir ein Maß $F_{TRegion}$ für die Trendstärke einer Region, als

$$F_{TRegion} = \sqrt{\frac{1}{n} \sum_{i=1}^n F_{Ti}}$$

wobei n die Anzahl der Städte in der Region ist, und i der laufende Index einer Stadt innerhalb der Region ist. F_{T-Wien} besteht also aus der Menge aller $F_{TRegion}$, sowie $F_{TStadt \neq Wien}$ i.e. $F_{T-Wien} = F_{TRegion} \cup F_{TStadt \neq Wien}$.

Um nun den Verlauf des Maßes B darzustellen, berechnen wir das Maß B für ein jährliches "expanding window", in unserem Fall wäre das erste *expanding window* B_{2004} , das zweite wäre dann $B_{2004-2005}$, das dritte $B_{2004-2006}$ etc. Formal geschrieben hätten wir B_t , mit t aus (2004, 2004 – 2005, 2004 – 2006, ..., 2004 – 2019), wobei

$$B_t = \sqrt{\frac{F_{TWient}^2 + F_{T-Wient}^2}{2}}$$

das Maß B erweitert um den Index t ist. Um einzuschätzen wie die Entwicklung der Zeitreihe in der vergangenen Zeit war, ziehen wir das *linearity*-Maß des *R* packages *feasts* hinzu. Das *linearity*-Maß gibt die Linearität der Trend-Komponente nach der Zeitreihenzerlegung an und basiert auf dem Koeffizienten einer orthogonalen quadratischen Regression auf die Trend-Komponente. Für die Richtung des Trends nehmen wir das Vorzeichen $sgn()$ des *linearity*-Maßes wie folgt

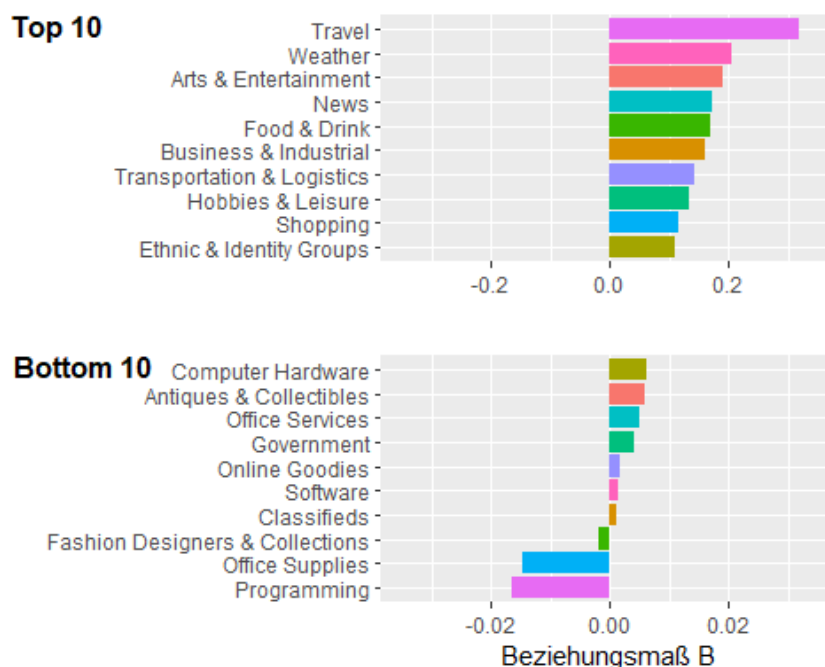
$$sgn(x) = \begin{cases} -1 & \text{falls } x < 0, \\ 1 & \text{falls } x \geq 0. \end{cases}$$

mit $x = linearity_{-wien_i}$ als Input, und kombinieren es mit unserem Beziehungsmaß B_t . Sollten wir mehrere Städte innerhalb einer Region haben, nehmen wir die Summe der Vorzeichen der Städte

$x = \frac{1}{n} \sum_{i=1}^n \text{sgn}(\text{linearity})_{T_i}$ als Input, und nehmen das Vorzeichen nochmals wie oben definiert. Wir haben 0 als positiv definiert, da wir einerseits eine Unterscheidung zwischen negativem und positivem Trend treffen müssen, und für die Analyse ein zumindest nicht negativer Trend als normativ positiv gesehen werden kann. Nachfolgend das berechnete Maß B für die Beziehung zwischen Wien und den Städten/Regionen.

Nachfolgend die *Top 10* und *Bottom 10* Kategorien dargestellt als Durchschnitt über alle Städte/Regionen. Beachten Sie bitte die unterschiedliche Skalierung auf der x-Achse zwecks Übersichtlichkeit (oberer Graph im Intervall $(-0.35, 0.35)$ wohingegen unterer Graph im Intervall $(-0.035, 0.035)$).

Abbildung 17 / Maß B für alle Kategorien, Ausschnitt Top 5 und Bottom 5



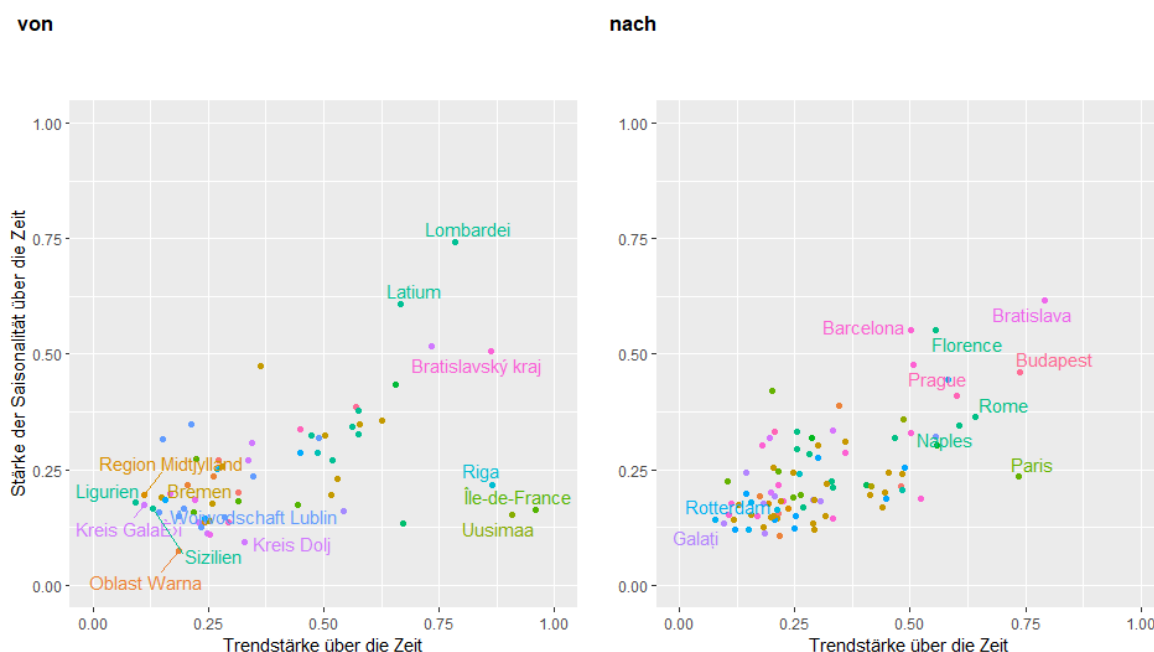
Quelle: Google Trends, eigene Berechnung.

Wenig überraschend hat die Kategorie *Travel* das stärkste Maß über alle Städte/Regionen. Auch *Arts & Entertainment*, *Food & Drink*, *Hobbies & Leisure*, *Shopping* und *Ethnic & Identity Groups* zählen zu Suchanfragen, die in den Bereich Wien-Tourismus fallen. Am unteren Ende des Rankings finden wir vor allem Suchanfragen zu Dienstleistungen oder Produkten beziehungsweise zum Thema *Government*, welche erwartungsgemäß eher lokal nachgefragt werden.

BEISPIEL TRAVEL

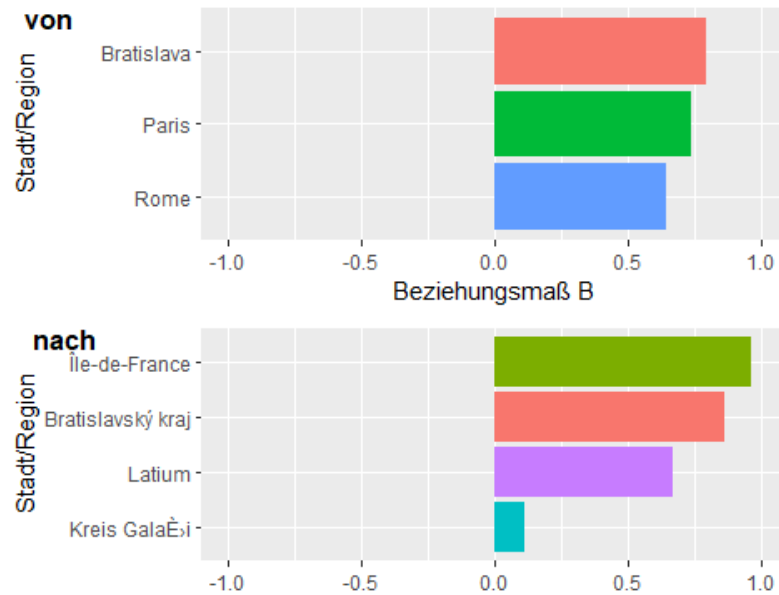
Wenn wir uns das Fallbeispiel Kategorie *Travel* ansehen, haben wir anschließend zwei Plots jeweils der Suchanfragen nach Wien und der Suchanfragen aus Wien. Gegenübergestellt wird hier wie schon oben gezeigt die *Trendstärke über die Zeit* mit der *Stärke der Saisonalität über die Zeit*, wobei das oben definierte Maß B auf der *Trendstärke* basiert. Aus beiden Graphen sehen wir, dass bei den "extremere" Werten Überschneidungen existieren, welche wir dann mittels des Maßes B (Kombination aus den beiden *Trendstärken*) genauer analysieren werden.

Abbildung 18 / Trendstärke versus Stärke der Saisonalität von Suchanfragen aller Kategorien von und nach Wien



Quelle: Google Trends, eigene Berechnung.

In beiden Grafiken gibt es Überschneidungen bei den extremere Werten bei Bratislava/Bratislavsky kraj, Rom/Latium, Paris/Ile-de-France, sowie Galati/Kreis Galati in Rumänien, wobei zweitgenanntes immer die Region ist. Schauen wir uns nun für die vier erwähnten Städte/Regionen das berechnete Trendmaß an, ergibt sich nachfolgendes Bild. Das Maß B gibt für alle vier Städte/Regionen positive Werte an, mit stark positiven Werten ($B > 0.6$) für Bratislava, Paris und Rom, sowie schwach positiven Werten ($B < 0.2$) für Galati in Rumänien.

Abbildung 19 / Trendstärke Bratislava, Paris, Rom und Galati

Quelle: Google Trends, eigene Berechnung.

Hier nochmals als Tabellen dargestellt:

Abbildung 20 / Tabellen der Trendstärke für Bratislava, Paris, Rom und Galati

Show entries Search:

	Quelle	Ziel	Kategorie	B
1	Wien	Bratislava	Travel	0.79
2	Wien	Paris	Travel	0.74
3	Wien	Rome	Travel	0.64

Showing 1 to 3 of 3 entries Previous Next

Show entries Search:

	Quelle	Ziel	Kategorie	B
1	Île-de-France	Vienna	Travel	0.96
2	Bratislavský kraj	Vienna	Travel	0.86
3	Latium	Vienna	Travel	0.67
4	Kreis Galați	Vienna	Travel	0.11

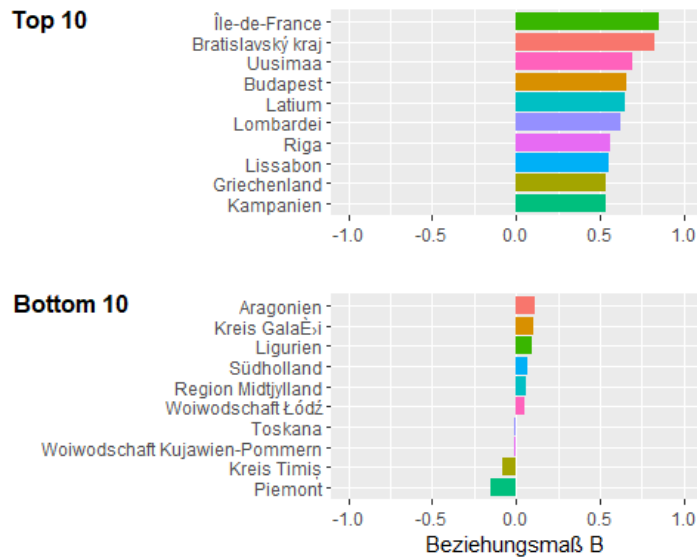
Showing 1 to 4 of 4 entries Previous Next

Quelle: Google Trends, eigene Berechnung.

Kombiniert man nun diese beiden Maße für die *Trendstärke* sowie dem *linearity* Feature wie oben beschrieben, um das definierte Maß *B* für den Trend aus beiden Suchanfragen-Seiten zu erhalten sehen wir wenig überraschend ein ähnliches Bild.

Durchgängig weist das Beziehungsmaß B für die Kategorie *Travel* sehr starke Werte aus, was einerseits für eine gute Positionierung der Stadt Wien, aber auch die Beziehung von Wien aus in die Städte/Regionen in dieser Kategorie spricht.

Abbildung 23 / Ausschnitt Maß B für Top 5 und Bottom 5



Quelle: Google Trends, eigene Berechnung.

Abbildung 24 / Wordcloud der Suchbegriffe in der Kategorie Travel



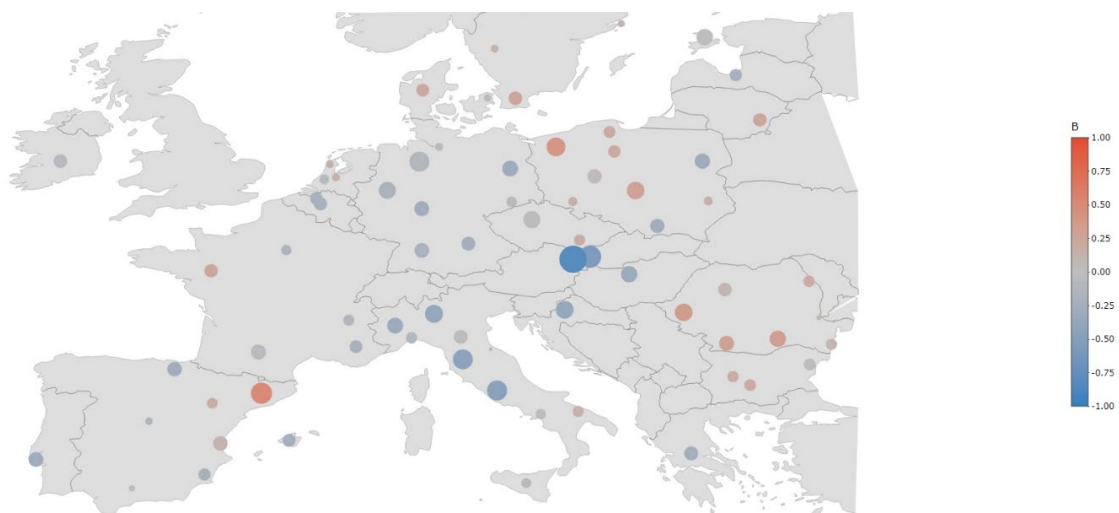
Quelle: Google Trends, eigene Berechnung.

Sieht man sich die Suchanfragen in der Kategorie *Travel* als Wordcloud an, sticht auch hier sofort "Bratislava" ins Auge. Weiters sieht man hier wenig überraschend Suchanfragen mit Inhalt von geografisch nahen Orten wie *Prag*, *Salzburg*, *Budapest*. Auch gut ersichtlich ist, dass manche

Suchbegriffe nicht thematisch beziehungsweise sprachlich zueinander gematcht wurden. So ist beispielsweise Prag mit "Prague" und "Praga", oder der Flughafen Wien mit "Vienna Airport", "Vienna International Airport", "Wien Flughafen", oder "Wien Airport" in den Top-Suchbegriffen enthalten. Solche Begriffe wären in einer optimalen Datenwelt idealerweise nochmals unter einem Thema/Bedeutung, ähnlich der Aufbereitung der amtlichen Namensstatistik, wo die einzelnen Namen nach Schreibweise, aber auch nach Aussprache dargestellt werden, zusammengefasst.

Wenn wir uns nun unser Beziehungsmaß B_t über die Zeit ansehen, kann man in folgendem interaktiven Plot sehr schön die Entwicklung zwischen den Jahren 2007 und 2019 verfolgen. Für die Jahre vor 2007 bekommt man aufgrund der Kürze der Zeitreihe und diversen fehlende beziehungsweise extremen Werten, keine sinnvolle oder gar keine *Trendstärke* und kann dementsprechend auch kein Maß für die Beziehungsstärke berechnen. Die Größe des Kreises ist die allgemeine *Trendstärke*, während die Farbe des Kreises den Wert des Maßes B_t darstellen in der Skala von -1 (blau), über 0 (grau), bis 1 (rot) im Farbverlauf. Je negativer eine Zahl, desto stärker wird der Kreis blau, und je positiver das Maß b desto stärker wird der Kreis rot eingefärbt.

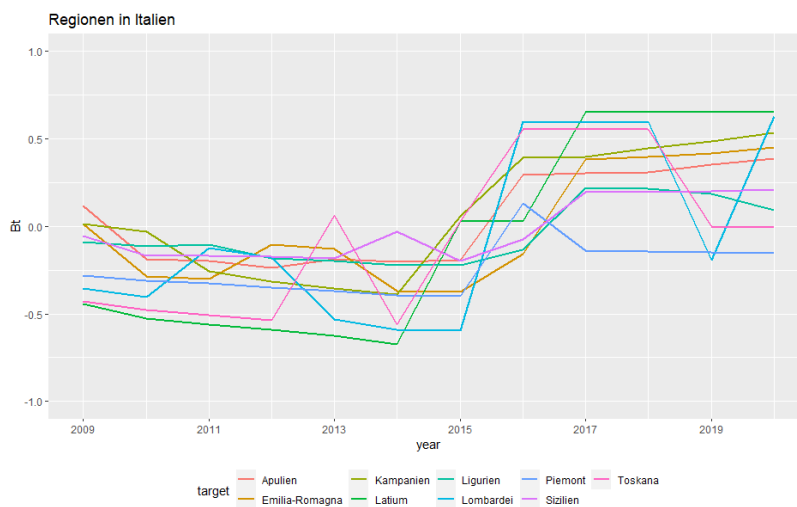
Abbildung 25 / Maß B für alle Städte/Regionen in der Kategorie Travel im Jahr 2007



Quelle: Google Trends, eigene Berechnung.

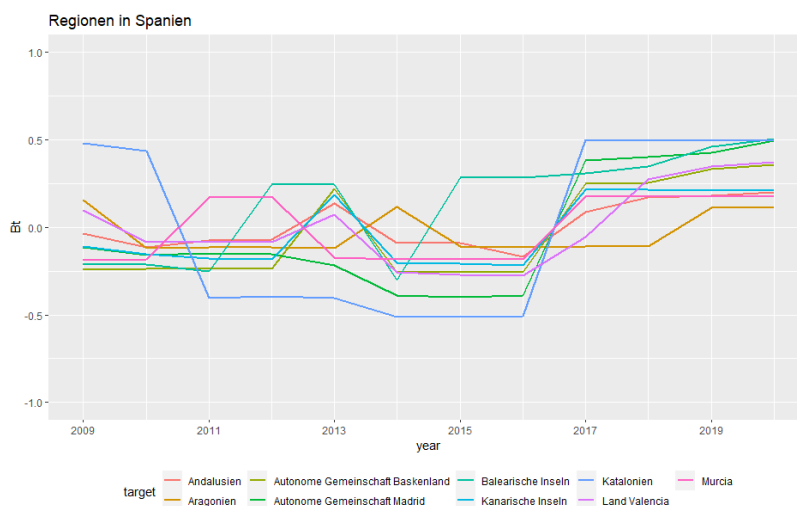
Auffallend ist hier, dass das Maß B_t , welches allgemein Anfangs leicht negativ bis leicht positiv startet, bis 2015 generell ins (leicht) negative abrutscht, ab 2015 sich aber sehr ins positive entwickelt. Das B_t um 0 am Anfang der Zeitreihe darf aber nicht verwundern, da die Anzahl der Beobachtungen beziehungsweise der Zeitbereich noch relativ kurz ist, und oft keinen starken Trend in die eine oder andere Richtung messbar sind.

Sehen wir uns zur Übersicht ausgewählte Regionen nun mittels Linien-Diagramm an, um den Trend der Zeitreihe für unser Maß B_t auf einen Blick darstellen zu können, so bestätigt sich obige Beobachtung aus dem dynamischen Graph.

Abbildung 26 / Maß B für Städte/Regionen in Italien in der Kategorie Travel über die Zeit

Quelle: Google Trends, eigene Berechnung.

Während Anfangs das Maß B_t noch weiter ins negative abrutscht, gibt es ab 2015 einen vergleichweisen starken Anstieg auf 0.2 bis 0.6. Einzig die Region *Piemont* stagniert ab mit einem leicht negativen Wert.

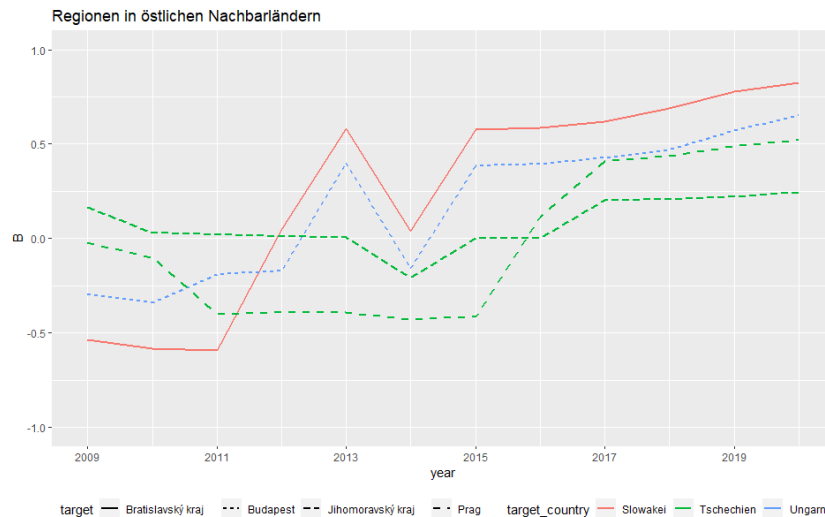
Abbildung 27 / Maß B für Städte/Regionen in Spanien in der Kategorie Travel über die Zeit

Quelle: Google Trends, eigene Berechnung.

Spanien zeigt in der Kategorie *Travel* Anfangs Werte um -0.2 bis 0.2 , welches dann ab 2016 aber in Richtung 0.1 bis 0.5 für das Maß B_t wandert.

In unseren direkten östlichen Nachbarländern *Tschechien*, *Slowakei* und *Ungarn* zeigt sich vor allem für *Bratislava* und *Budapest* ein starker positiver Trend 2016 bzw. 2017, während Regionen aus Tschechien Anfangs relativ konstant bleiben und kaum Veränderung des Maßes B_t zeigen, ab 2015 aber auch eine Entwicklung in Richtung 0.25 bis 0.5 (Prag) zeigen.

Abbildung 28 / Maß B für Städte/Regionen Tschechien, Slowakei und Ungarn in der Kategorie Travel über die Zeit

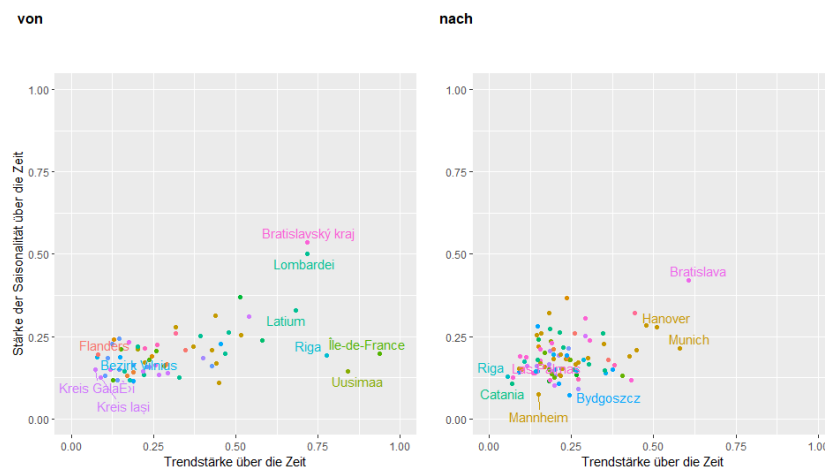


Quelle. Google Trends, eigene Berechnung.

BEISPIEL BUSINESS & INDUSTRIAL

In der Kategorie *Business & Industrial* zeigt sich ein Bild mit relativ wenig Saisonalität und Beobachtungen konzentriert in einem Bereich mit geringer Trendstärke.

Abbildung 29 / Trendstärke versus Stärke der Saisonalität in der Kategorie Business & Industrial von und nach Wien

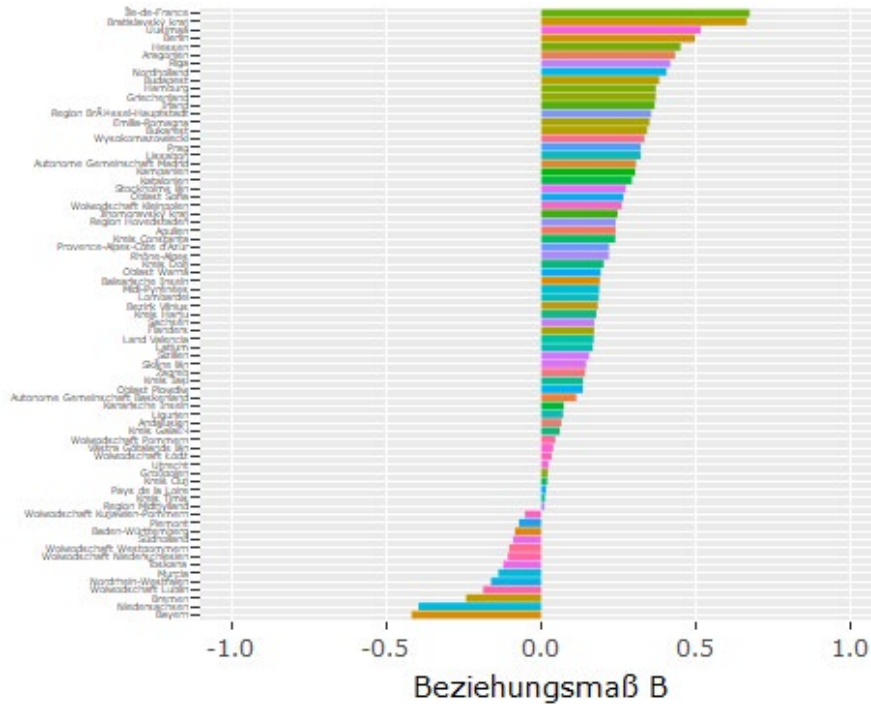


Quelle. Google Trends, eigene Berechnung.

Man sieht auch, dass die Suchanfragen aus den Regionen/Städten (im obigen Graph mit "von" deklariert) eine höhere Saisonalität sowie Trendstärke für einige ausweisen, wohingegen die Suchanfragen von Wien ausgehend in die Städte kaum Trendstärken über 0,5, sowie Werte für die

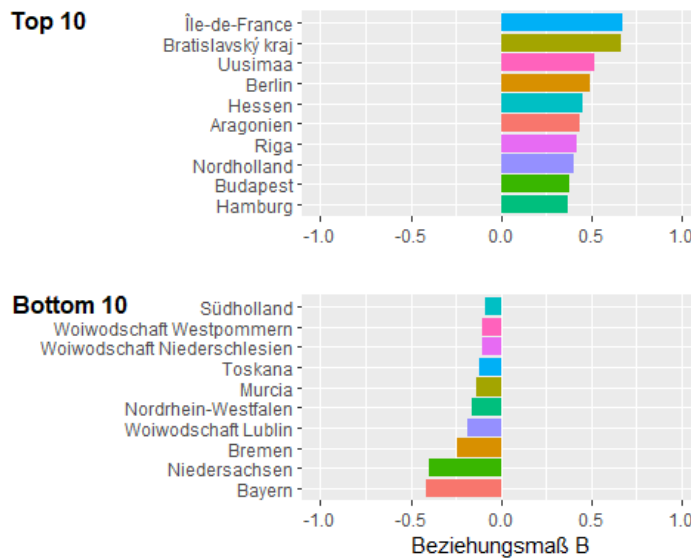
Saisonalität der Anfragen über 0.3 aufweisen. In beide Graphen zeigt sich wieder Bratislava als "stärkste" Verbindung.

Abbildung 30 / Maß B in der Kategorie Business & Industrial über alle Städte/Regionen



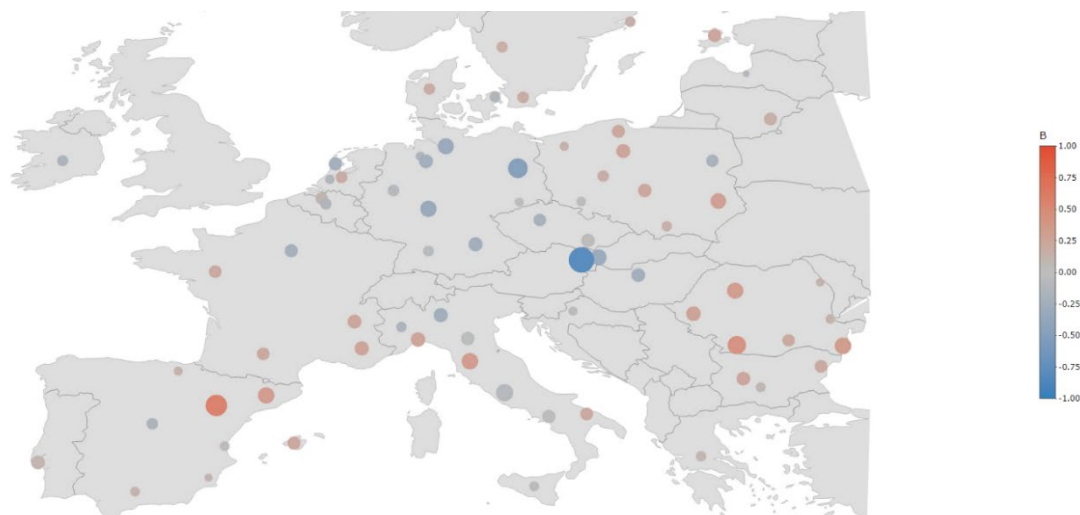
Quelle. Google Trends, eigene Berechnung.

Abbildung 31 / Ausschnitt Maß B in der Kategorie Business & Industrial für Top 5 und Bottom 5



Quelle. Google Trends, eigene Berechnung.

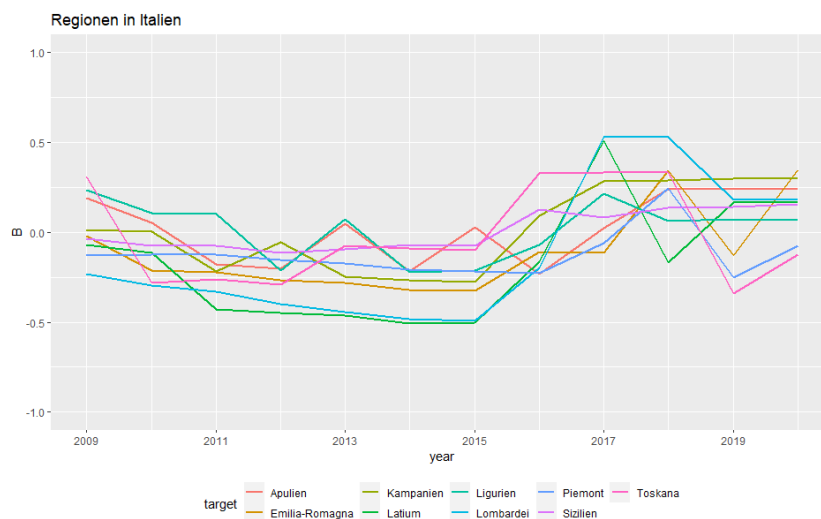
Abbildung 33 / Maß B in der Kategorie Business & Industrial über alle Städte/Regionen im Jahr 2007



Quelle: Google Trends, eigene Berechnung.

Wenn wir wieder einen genaueren Blick auf ausgewählte Regionen beziehungsweise Länder werfen, sehen wir, wie im nachstehend dargestellten Graphen der Regionen in Italien, diese Trend-Beobachtung bestätigt wird.

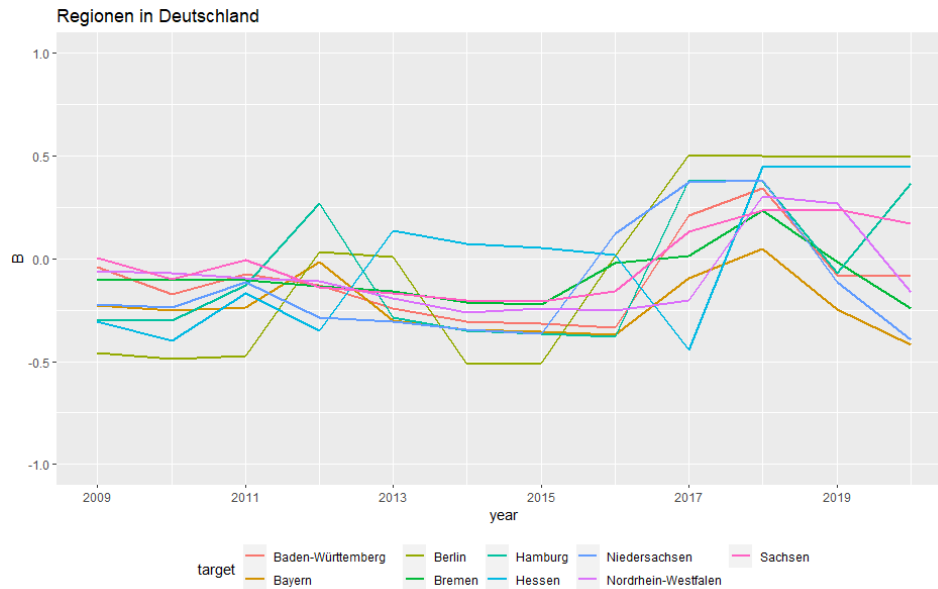
Abbildung 34 / Maß B für Städte/Regionen in Italien in der Kategorie Business & Industrial über die Zeit



Quelle: Google Trends, eigene Berechnung.

Auch in Deutschland zeigt sich ein ähnliches Bild, wobei viel Regionen ab 2018 wieder in Richtung negatives Maß wandern. Das betrifft vor allem die Regionen *Nordrhein-Westfalen*, *Bremen*, *Niedersachsen* und *Bayern*, welche ab diesem Zeitpunkt relativ stark abrutschen auf das Niveau $(-0.1, -0.4)$.

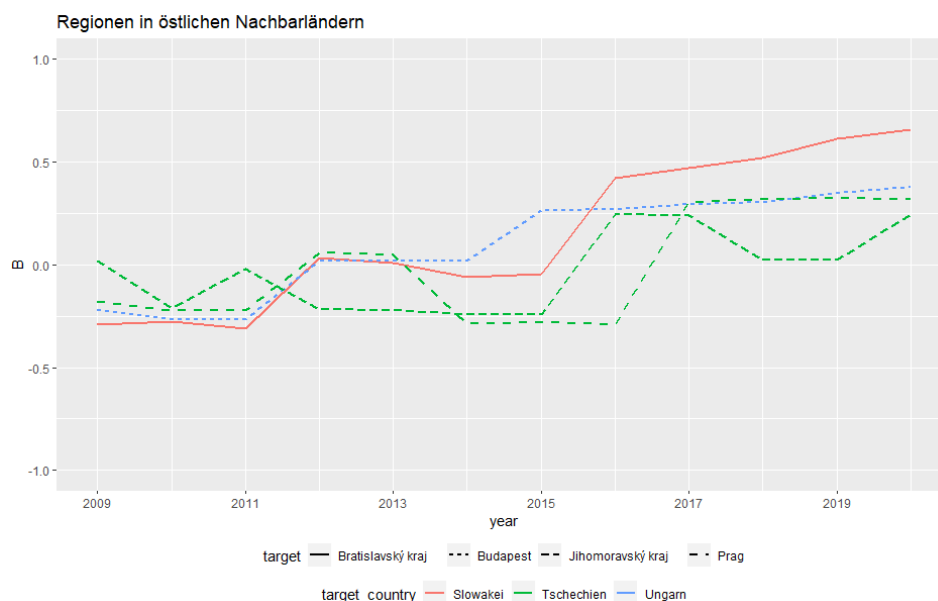
Abbildung 35 / Maß B für Städte/Regionen in Deutschland in der Kategorie Business & Industrial über die Zeit



Quelle: Google Trends, eigene Berechnung.

Ein Blick auf Österreichs östliche Nachbarländer zeigt einen Anstieg des Maßes B_t in der Kategorie *Business & Industrial* für alle beobachteten Städte/Regionen, wobei sich alle Zeitreihen von einem (leicht) negativen Niveau in Richtung eines positiven Niveaus bewegen, insbesondere die Region um *Bratislava* welche mit Ende 2019 einen relativ hohen Wert von 0.66 aufweist.

Abbildung 36 / Maß B für Städte/Regionen in Tschechien, Slowakei und Ungarn in der Kategorie Business & Industrial über die Zeit

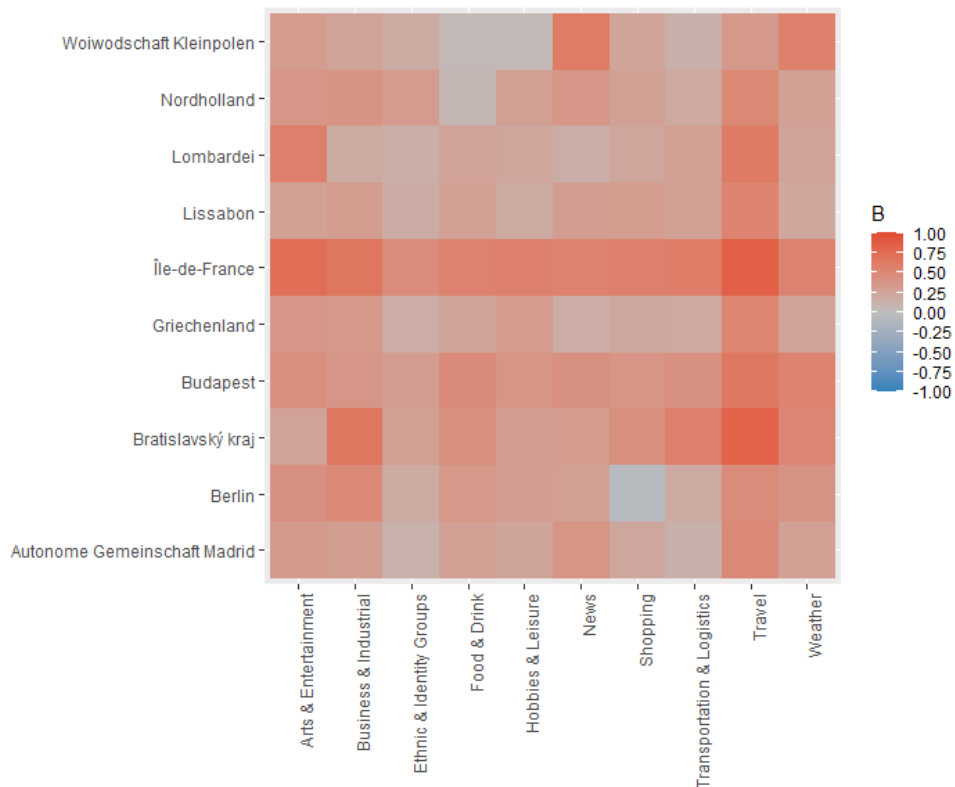


Quelle: Google Trends, eigene Berechnung.

VORSCHLAG DASHBOARD

Um einen Überblick und weitreichende Filtermöglichkeiten der hier präsentierten Kennzahlen zu ermöglichen, schlagen wir die Erstellung eines sogenannten *Dashboards* für die Präsentation vor. Ein *Dashboard* ist eine grafische Benutzeroberfläche, welche Schlüsselindikatoren grafisch und übersichtlich aufbereitet, sowie mittels Auswahlmöglichkeiten tiefere Einblicke in spezifische Indikatoren zulässt. Zu diesem Zweck kann man neben den oben bereits präsentierten Möglichkeiten beispielsweise eine heat map in Form einer Matrix für die Kategorien und Städte/Regionen mit den aktuell positivsten sowie den aktuell negativsten Werten für das Beziehungsmaß B erstellen. Sollten sich die Grunddaten ändern, dann ändert sich erwartungsgemäß auch die Zusammensetzung dieser Kombination von Beobachtungen. Die aktuell positivsten Kategorien versus die positivsten Städte/Regionen (wenig Verbesserungspotenzial):

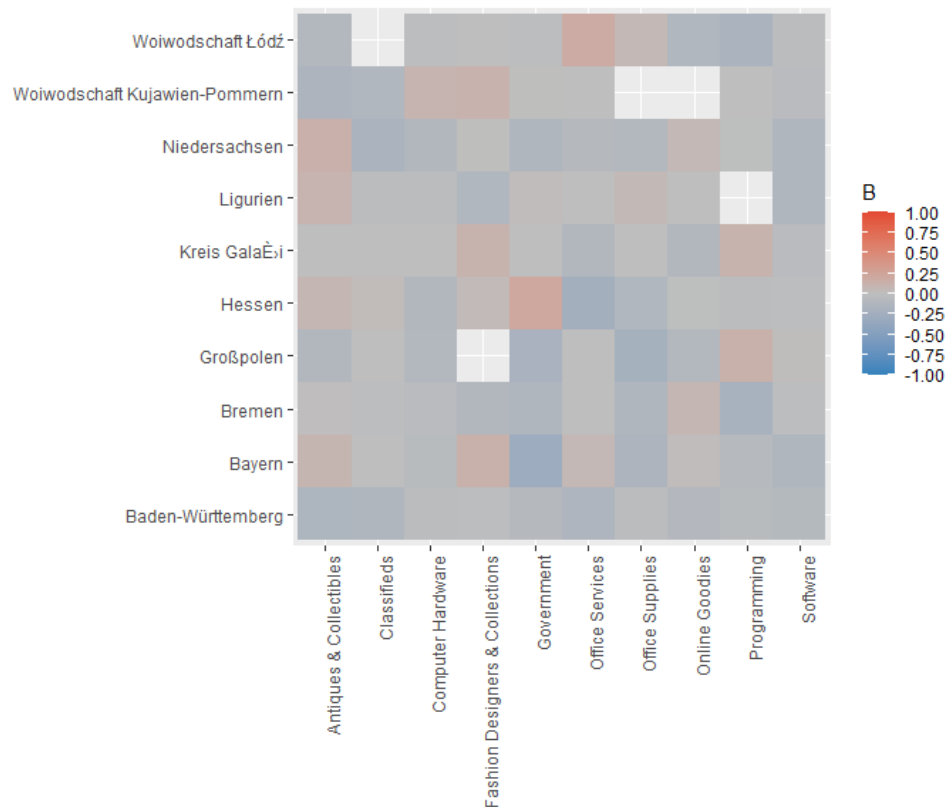
Abbildung 37 / Aktuell positivste Kategorien versus aktuell positivste Städte/Regionen



Quelle. Google Trends, eigene Berechnung.

Die aktuell negativsten Kategorien versus die aktuell negativsten Städte/Regionen (großes Verbesserungspotenzial):

Abbildung 38 / Aktuell negativste Kategorien versus aktuell negativste Städte/Regionen

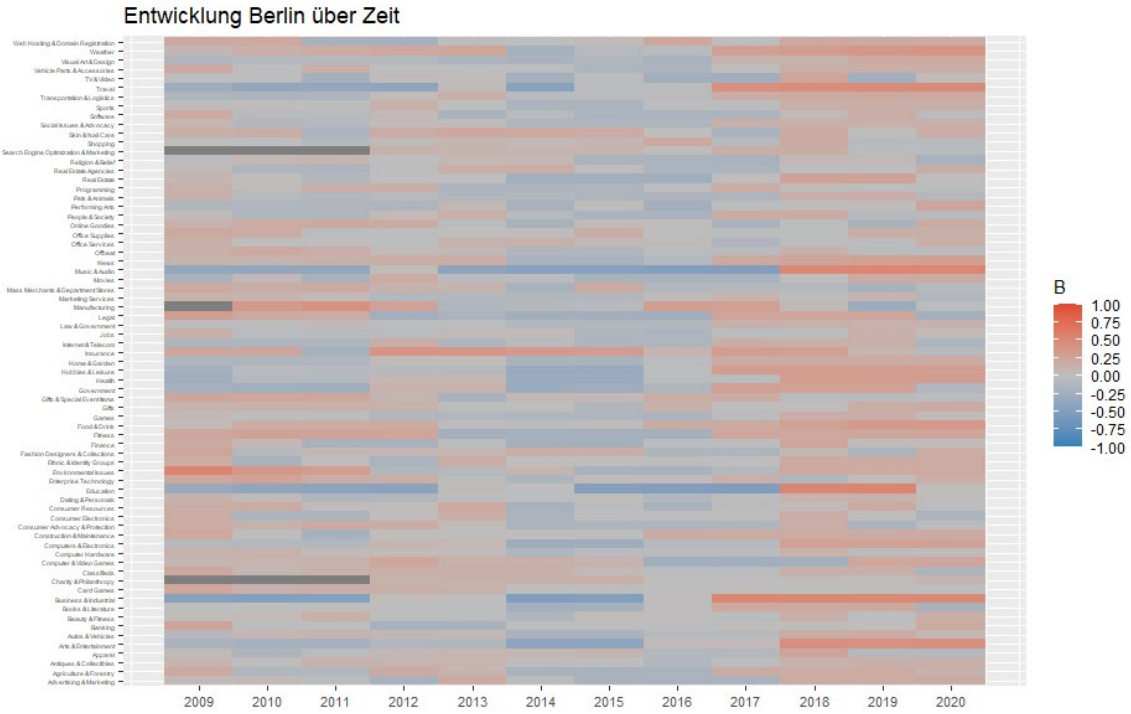


Quelle: Google Trends, eigene Berechnung.

Eine weitere Variante wäre, basierend auf der Darstellung als heat map Matrix, das Filtern nach den Regionen/Städten und die grafische Aufbereitung aller Werte für die einzelnen Kategorien über die Jahre mit dem Maß B_t , zu ermöglichen, wie nachstehend ersichtlich.

Dies erlaubt einen kompakten Überblick, sowie eine tiefere Analyse über die Entwicklung der Maßzahl B mit einer spezifischen Stadt/Region über die Zeit. Will man einen genaueren Blick auf die zugrundeliegenden Zeitreihen werfen, also Suchanfragen von Wien in die ausgewählte Stadt, sowie Suchanfragen nach Wien aus der Stadt/Region, kann man dies in einem eigenen Bereich mit Liniendiagrammen für die Zeitreihe, Wordcloud für die Suchanfragen, sowie Balkendiagrammen ähnlich den bereits verwendeten Darstellungen weiter oben pro Kategorie/Stadt Kombination dynamisch generieren. Ein Dashboard mit den in diesem Papier präsentierten Kennzahlen erlaubt somit schnell Suchanfragen-Trends sowie die Beziehung von Suchanfragen von Wien mit den Städten/Regionen Europas auf einem übersichtlichen Weg zu erkennen und, sowie spezifische Reports nach Kategorien oder Städte/Regionen zu erstellen.

Abbildung 39 / Maß B in allen Kategorien für Berlin über die Zeit



Quelle. Google Trends, eigene Berechnung.

Schlussfolgerungen

Wir haben ein neues Maß für die Suchanfragen-Beziehung zwischen zwei Städten/Regionen basierend auf beidseitigen Google Suchanfragen vorgestellt, welches es erlaubt die Entwicklung der beidseitigen Interessen über die Zeit zu verfolgen. Die vorgestellte Kennzahl ist ein Maß für die beidseitige Trendstärke und dessen Entwicklung in positive oder negative Richtung in den letzten fünf Jahren, und kann für strategische Entscheidungen im Bereich der Werbung, der gezielten Förderung einzelner Wirtschaftsbereiche in den unterschiedlichen Kategorien und für zukünftige Politikentscheidungen genutzt werden. Basierend auf den Suchanfragen aus anderen Städten/Regionen kann die Politik einerseits Rückschlüsse ziehen, welche bereits bestehenden Angebote genutzt werden, aber auch welche Interessen aktuell (noch) nicht (ausreichend) abgedeckt werden.

Wenn man sich beispielsweise das vorgeschlagene Dashboard mit den höchsten, sowie jenes mit den niedrigsten Werten vor Augen führt, zeigt sich einerseits wie stark der Fokus auf das Reisen beziehungsweise Reisen-relevante Themen gelegt wird, und zum anderen wie wichtig einzelne Städte/Regionen (wie Paris/Ile-de-France) sind. Man könnte dieses Dashboard etwa nutzen um das realisierte Touristik-Potenzial in bestimmten Städten/Regionen auszubauen (vor allem in Hinblick auf die Zeit nach Corona), oder, um ein weiteres Beispiel zu nennen, es könnte daran gedacht werden für die digitale Wirtschaftsentwicklung die richtigen Weichen in IT-relevanten Bereichen zu stellen.

Die Kombination aus der vorgestellten Kennzahl, deren Entwicklung, sowie die von Google bereitgestellten Suchanfragen bieten ein umfangreiches Tool dafür.

Appendix

KATEGORIEN

Abbildung 40 / Verwendete Kategorien

Show entries Search:

	Kategorie	ID
1	All categories	0
2	Home & Garden	11
3	Business & Industrial	12
4	Internet & Telecom	13
5	People & Society	14
6	News	16
7	Shopping	18
8	Law & Government	19
9	Sports	20
10	Books & Literature	22

Showing 1 to 10 of 75 entries Previous 2 3 4 5 ... 8 Next

Quelle: Google Trends, eigene Berechnung.

TREND & SAISONALITÄT

Abbildung 41 / Suchanfragen nach Wien mit "trend_strength"=0

Show entries Search:

	Quelle	Kategorie
1	Apulien	Classifieds
2	Apulien	Enterprise Technology
3	Apulien	Skin & Nail Care
4	Apulien	Web Hosting & Domain Registration
5	Bezirk Vilnius	Classifieds
6	Bezirk Vilnius	Office Services
7	Bratislavský kraj	Search Engine Optimization & Marketing
8	Bremen	Office Services
9	Bremen	Skin & Nail Care
10	Bukarest	Environmental Issues

Showing 1 to 10 of 272 entries Previous 2 3 4 5 ... 28 Next

Quelle: Google Trends, eigene Berechnung.

Literatur

Cleveland, Robert B, William S Cleveland, Jean E McRae, and Irma Terpenning. 1990. "STL: A Seasonal-Trend Decomposition." *Journal of Official Statistics* 6 (1): 3–73.

Google. 2019. "How Trends Data Is Adjusted." 2019.
https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052.

Gamma, Alex and Schleifer, Roman and Weinmann, Wolfgang and Buadze, Anna and Liebrez, Michael. 2016. "Could Google Trends be used to predict methamphetamine-related crime? An analysis of search volume data in Switzerland, Germany, and Austria" In *PloS one*, 11 (11). Public Library of Science.

Horák, Jiří, Igor Ivan, Pavel Kukuliač, Tomáš Inspektor, Branislav Devečka, and Markéta Návrátová. 2013. "Google Trends for Data Mining. Study of Czech Towns." In *International Conference on Computational Collective Intelligence*, 100–109. Springer.

Jun, Seung-Pyo, Hyoung Sun Yoo, and San Choi. 2018. "Ten Years of Research Change Using Google Trends: From the Perspective of Big Data Utilizations and Applications." *Technological Forecasting and Social Change* 130: 69–87.

Önder İrem. 2017. "Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities" *International Journal of Tourism Research* 19 (6): 648–660. Wiley Online Library

Scott, Steven L, and Hal R Varian. 2013. "Predicting the Present with Bayesian Structural Time Series." Available at SSRN 2304426.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–28.

IMPRESSUM

Herausgeber, Verleger, Eigentümer und Hersteller:

Verein „Wiener Institut für Internationale Wirtschaftsvergleiche“ (wiiw),
Wien 6, Rahlgasse 3

ZVR-Zahl: 329995655

Postanschrift: A 1060 Wien, Rahlgasse 3, Tel: [+431] 533 66 10, Telefax: [+431] 533 66 10 50
Internet Homepage: www.wiiw.ac.at

Nachdruck nur auszugsweise und mit genauer Quellenangabe gestattet.

Offenlegung nach § 25 Mediengesetz: Medieninhaber (Verleger): Verein "Wiener Institut für Internationale Wirtschaftsvergleiche", A 1060 Wien, Rahlgasse 3. Vereinszweck: Analyse der wirtschaftlichen Entwicklung der zentral- und osteuropäischen Länder sowie anderer Transformationswirtschaften sowohl mittels empirischer als auch theoretischer Studien und ihre Veröffentlichung; Erbringung von Beratungsleistungen für Regierungs- und Verwaltungsstellen, Firmen und Institutionen.

