

Scraping Eventseiten

Es werden 5 Städte zu unterschiedlichen Zeitpunkten gescrapet. Ziel ist es, herauszufinden, ob die Anwesenheit eines grossen Events Auswirkungen auf die gescrapeten Daten hat im Vergleich zu Zeiträumen ohne Events.

Städte und Events

Mögliche Auswahl

- **Manchester, CL, 11 und 13 April**
- **Amsterdam, DGTL, 7-9 April und King's Day, 27 April**
- Mailand, CL, 12 und 19 April
- **Lissabon, CL, 11 und 20 April**
- Madrid, CL, 12 April und Fiesta de San Isidro, 15 Mai und Madrid Day 2 Mai
- **München, CL, 19 April und Münchner Springfestival 21 April - 7 Mai**
- Radovljica, The Chocolate Festival, 15-16 April
- **Helsinki, April Jazz Festival, 21-29 April**
- Zermatt, Zermatt Unplugged, 11-15 April
- Noordwijk, Bloemencorso Flower Parade, 22 April
- **Sevilla, Feria de Abril, 23-29 April**
- Braga, Semana Santa, 2-9 April
- Budapest, Budapest Spring Festival, 20 April - 2 Mai
- **Monaco, Monaco Grand Prix, 28 Mai**
- Emilia-Romagna, Emilia-Romagna Grand Prix, 21 Mai
- **Lyon, Nuits Sonores, 17- 21. Mai**
- Halle, Handel Festival, 26 Mai - 11 Juni
- **Brighton, Brighton Food Festival, 29 April - 1 Mai**
- **Hamburg, Hafengeburtstag, 5-7 Mai**

- Mallorca, Mallorca Live Festival, 18-20 Mai
- Barcelona, Beer Festival, 26-28 Mai
- Berlin, Karneval der Kulturen, 26-29 Mai
- Brüssel, Jazz Weekend, 26-28 Mai
- Cannes, Internationale Filmfestspiele, 16-27 Mai

→ Zusätzlich gibt es je nach Stadt noch eine Vielzahl (wahrscheinlich sogar zu viele) an Konzerte und Festivals (wie relevant?)

Websites

- Beste Übersicht und eine Website für alle Städte:
<https://www.viagogo.com/ww/>
- Für jedes Land eigene Seite: <https://www.eventim.de/campaign/international>
oder <https://www.ticketmaster.de/>
- Europaweite (abgegrenzt durch Länder aber nicht Städte):
<https://festivaly.eu/en/events/europe>

Daten - welche Daten werden gescrapet

- Name des Events
- Stadt
- Datum
- Anzahl der Events

All Events in Amsterdam

Filter By Clear all

Keywords Clear

Timeframe Clear

Select when ... v

Dates Clear

Events from... ...to

Category Clear

☒ Any category

☐ Women's Euro

☐ Upcoming

☐ Concert Tickets

☐ Sports Tickets

☐ Theatre Tickets

☐ Festival Tickets

Maximum Price/Ticket Clear

CHF

☐ Only show events with tickets available

SORT BY:	RELEVANCY	DATE ▲	PRICE	
Sat	25 Mar 2023	21:00	Rootsriders	P60, Amstelveen Netherlands View Tickets
Sun	26 Mar 2023	14:00	Grease	DeLaMar Theater - Wim Sonneveld Hall, Amsterdam Netherlands View Tickets
Sun	26 Mar 2023	14:00	Het Zwanenmeer	De Nationale Opera, Amsterdam Netherlands View Tickets
Sun	26 Mar 2023	15:00	FTBLL - Can you kick it!	Koninklijk Theater Carre, Amsterdam Netherlands View Tickets
Sun	26 Mar 2023	19:00	Mimi Webb	Melkweg Cinema, Amsterdam Netherlands View Tickets
Sun	26 Mar 2023	19:00	Mimi Webb	Melkweg Max, Amsterdam Netherlands View Tickets
Sun	26 Mar 2023	19:00	Alex G	Paradiso, Amsterdam Netherlands View Tickets

Kritische Hinterfragung

- Macht es Sinn, nur während zwei verschiedenen Datenpunkten zu scrapen?
 - Am besten wäre es, wenn man täglich scrapen würde, um die Daten im Zeitverlauf zu untersuchen
 - Problem: Wie aufwändig ist das? (auf Server auslagern?)
 - Wenn wir nur während zwei Daten scrapen besteht die Gefahr, dass wir Dynamiken in den Daten verpassen
 - (Da sich Preise häufig ändern, müsste der Scraper regelmässig aktualisiert werden, um genaue und relevante Daten zu liefern)
- Wenn wir eine Stadt mit einem grossen Event (bswp. Champions League) scrapen, werden andere Events sehr wahrscheinlich kein grosses Gewicht haben
- Problem bei Konzerten → Wahrscheinlich gehen nur Einwohner der Stadt und keiner braucht ein Hotel/AirBnB
- Eventuell zu kurzfristig → Events bereits angesagt und die "grösste" Preisentwicklung wahrscheinlich bereits passiert
- Was machen wir, wenn keine nützlichen Informationen identifiziert werden können?

- Notfallplan/-projekt? (→ Arne offen kommunizieren?)

Neuer Plan

- 2 Daten max 1-2 Wochen entfernt
- 10 Städte
 1. Datum: 1 Grosses / Mittleres Event
 2. Datum: Kein relevantes Event
- Hashtag
 - Evtl. bereits Twitter-Scraper bauen
- Wir scrapen nicht an dem Tag selbst sondern schauen die Preise für den Tag an