

Data Science Project

Ideas

Henri Heyden Mika Materzok Nike Pulow Senanur Şimşek
Group “Für Fortnite”

Christian-Albrechts-Universität zu Kiel

March 7, 2025

Die Quellen sind noch nicht up to date,
morgen finden wir noch paar mehr, die
man gut verwenden kann!

Overview

- 1 Music
- 2 Kiel
- 3 Gaming
- 4 Gigachad-Gpt
- 5 Supermarket
- 6 Illnesses & University
- 7 Rich people
- 8 AI embeddings on images

Music

Spotify API, Songsterr/Ultimate-Guitar webscraping

- What chord-progressions repeat most frequent in songs by artist (e.g. Taylor Swift), timespan or genre?

Music

Spotify API, Songsterr/Ultimate-Guitar webscraping

- What chord-progressions repeat most frequent in songs by artist (e.g. Taylor Swift), timespan or genre?
- How did song length change through the years?

Music

Spotify API, Songsterr/Ultimate-Guitar webscraping

- What chord-progressions repeat most frequent in songs by artist (e.g. Taylor Swift), timespan or genre?
- How did song length change through the years?
- What genres are more popular in which countries?

Music

Spotify API, Songsterr/Ultimate-Guitar webscraping

- What chord-progressions repeat most frequent in songs by artist (e.g. Taylor Swift), timespan or genre?
- How did song length change through the years?
- What genres are more popular in which countries?
- Do song tempo and popularity correlate?

Music

Spotify API, Songsterr/Ultimate-Guitar webscraping

- What chord-progressions repeat most frequent in songs by artist (e.g. Taylor Swift), timespan or genre?
- How did song length change through the years?
- What genres are more popular in which countries?
- Do song tempo and popularity correlate?
- Does the season influence the music people listen to?

Overview

- 1 Music
- 2 **Kiel**
- 3 Gaming
- 4 Gigachad-Gpt
- 5 Supermarket
- 6 Illnesses & University
- 7 Rich people
- 8 AI embeddings on images

- What buslines/trains are most frequently late, and how late?

Kiel

Stadt Kiel Open Data, Immowelt API, Tourismusverband Schleswig-Holstein, Public GPS-traces

- What buslines/trains are most frequently late, and how late?
- Which districts have the highest density of coffee shops and/or bars?

Kiel

Stadt Kiel Open Data, Immowelt API, Tourismusverband Schleswig-Holstein, Public GPS-traces

- What buslines/trains are most frequently late, and how late?
- Which districts have the highest density of coffee shops and/or bars?
- What influence did **THE EVENT**¹ have on local retail and gastronomic sales?

¹Kieler Woche / Corona?

Kiel

Stadt Kiel Open Data, Immowelt API, Tourismusverband Schleswig-Holstein, Public GPS-traces

- What buslines/trains are most frequently late, and how late?
- Which districts have the highest density of coffee shops and/or bars?
- What influence did **THE EVENT**¹ have on local retail and gastronomic sales?
- How do search requests about Kiel change during the “Kieler Woche” over the year?

¹Kieler Woche / Corona?

- What buslines/trains are most frequently late, and how late?
- Which districts have the highest density of coffee shops and/or bars?
- What influence did **THE EVENT**¹ have on local retail and gastronomic sales?
- How do search requests about Kiel change during the “Kieler Woche” over the year?
- How high is the economical influence of the “Kieler Woche” on the city?

¹Kieler Woche / Corona?

- What buslines/trains are most frequently late, and how late?
- Which districts have the highest density of coffee shops and/or bars?
- What influence did **THE EVENT**¹ have on local retail and gastronomic sales?
- How do search requests about Kiel change during the “Kieler Woche” over the year?
- How high is the economical influence of the “Kieler Woche” on the city?
- How did rental costs change in different parts of Kiel?

¹Kieler Woche / Corona?

- What buslines/trains are most frequently late, and how late?
- Which districts have the highest density of coffee shops and/or bars?
- What influence did **THE EVENT**¹ have on local retail and gastronomic sales?
- How do search requests about Kiel change during the “Kieler Woche” over the year?
- How high is the economical influence of the “Kieler Woche” on the city?
- How did rental costs change in different parts of Kiel?
- Are there patterns in the apartment listings over the seasons?

¹Kieler Woche / Corona?

- What buslines/trains are most frequently late, and how late?
- Which districts have the highest density of coffee shops and/or bars?
- What influence did **THE EVENT**¹ have on local retail and gastronomic sales?
- How do search requests about Kiel change during the “Kieler Woche” over the year?
- How high is the economical influence of the “Kieler Woche” on the city?
- How did rental costs change in different parts of Kiel?
- Are there patterns in the apartment listings over the seasons?
- ~~Where are hotspots of GPS-activity in Kiel?~~

¹Kieler Woche / Corona?

Overview

- 1 Music
- 2 Kiel
- 3 Gaming**
- 4 Gigachad-Gpt
- 5 Supermarket
- 6 Illnesses & University
- 7 Rich people
- 8 AI embeddings on images

Gaming

Steam API, Twitch API, Kaggle Gaming set

- Do playtime and ratings correlate?

²Oder auch: What makes a game popular? Also konkret bezogen zum Beispiel auf Tags, die ein Spiel ausgezeichnet bekommt/hat.

Gaming

Steam API, Twitch API, Kaggle Gaming set

- Do playtime and ratings correlate?
- Which games have the longest active communities?

²Oder auch: What makes a game popular? Also konkret bezogen zum Beispiel auf Tags, die ein Spiel ausgezeichnet bekommt/hat.

Gaming

Steam API, Twitch API, Kaggle Gaming set

- Do playtime and ratings correlate?
- Which games have the longest active communities?
- How do Twitch-Streams influence the sales of games?

²Oder auch: What makes a game popular? Also konkret bezogen zum Beispiel auf Tags, die ein Spiel ausgezeichnet bekommt/hat.

Gaming

Steam API, Twitch API, Kaggle Gaming set

- Do playtime and ratings correlate?
- Which games have the longest active communities?
- How do Twitch-Streams influence the sales of games?
- Which factors influence the popularity of a game?²

²Oder auch: What makes a game popular? Also konkret bezogen zum Beispiel auf Tags, die ein Spiel ausgezeichnet bekommt/hat.

Overview

- 1 Music
- 2 Kiel
- 3 Gaming
- 4 Gigachad-Gpt**
- 5 Supermarket
- 6 Illnesses & University
- 7 Rich people
- 8 AI embeddings on images

- What are frequent prompts?

³Mehr interessante Fragen

⁴Die API ist nicht frei verfügbar, aber wir könnten theoretisch prompts automatisch eingeben indem wir heimlich nen “virtuellen” browser steuern (aka. interactive session based webscraping) und prompts eingeben lassen. Verstößt aber eventuell gegen deren AGB.

- What are frequent prompts?
- How do answers change with the same prompt remade?

³Mehr interessante Fragen

⁴Die API ist nicht frei verfügbar, aber wir könnten theoretisch prompts automatisch eingeben indem wir heimlich nen “virtuellen” browser steuern (aka. interactive session based webscraping) und prompts eingeben lassen. Verstößt aber eventuell gegen deren AGB.

- What are frequent prompts?
- How do answers change with the same prompt remade?
- ...³

³Mehr interessante Fragen

⁴Die API ist nicht frei verfügbar, aber wir könnten theoretisch prompts automatisch eingeben indem wir heimlich nen “virtuellen” browser steuern (aka. interactive session based webscraping) und prompts eingeben lassen. Verstößt aber eventuell gegen deren AGB.

Overview

- 1 Music
- 2 Kiel
- 3 Gaming
- 4 Gigachad-Gpt
- 5 Supermarket**
- 6 Illnesses & University
- 7 Rich people
- 8 AI embeddings on images

Supermarket

LIDL/Rewe, etc.

- What are frequent itemsets?

Supermarket

LIDL/Rewe, etc.

- What are frequent itemsets?
- How do sales change during the year?

Supermarket

LIDL/Rewe, etc.

- What are frequent itemsets?
- How do sales change during the year?
- What do people eat during different seasons?

Supermarket

LIDL/Rewe, etc.

- What are frequent itemsets?
- How do sales change during the year?
- What do people eat during different seasons?
- What do people buy during different holidays?

Overview

- 1 Music
- 2 Kiel
- 3 Gaming
- 4 Gigachad-Gpt
- 5 Supermarket
- 6 Illnesses & University**
- 7 Rich people
- 8 AI embeddings on images

- When do students/lecturers get ill?

- When do students/lecturers get ill?
- Do students/lecturers get ill at different times in the semester?

- When do students/lecturers get ill?
- Do students/lecturers get ill at different times in the semester?
- Do students get ill during the examination period?

⁵Gute Quellen finden!

- When do students/lecturers get ill?
- Do students/lecturers get ill at different times in the semester?
- Do students get ill during the examination period?
- Are students/lecturers suffering from higher stress compared to other people?

⁵Gute Quellen finden!

- When do students/lecturers get ill?
- Do students/lecturers get ill at different times in the semester?
- Do students get ill during the examination period?
- Are students/lecturers suffering from higher stress compared to other people?
- What are the influences of university to the mental health of students/lecturers?

⁵Gute Quellen finden!

Overview

- 1 Music
- 2 Kiel
- 3 Gaming
- 4 Gigachad-Gpt
- 5 Supermarket
- 6 Illnesses & University
- 7 Rich people**
- 8 AI embeddings on images

Rich people

Webscraping, public social media APIS⁷,

- At which time or day in the week do some famous billionaires post the most on social media?

⁶Fun fact: Taylor Swift is a billionaire according to Forbes.

⁷konkretisieren

Rich people

Webscraping, public social media APIS⁷,

- At which time or day in the week do some famous billionaires post the most on social media?
- Which billionaires post most frequently on social media?

⁶Fun fact: Taylor Swift is a billionaire according to Forbes.

⁷konkretisieren

Rich people

Webscraping, public social media APIS⁷,

- At which time or day in the week do some famous billionaires post the most on social media?
- Which billionaires post most frequently on social media?
- Which billionaires fly the most?⁶

⁶Fun fact: Taylor Swift is a billionaire according to Forbes.

⁷konkretisieren

Rich people

Webscraping, public social media APIS⁷,

- At which time or day in the week do some famous billionaires post the most on social media?
- Which billionaires post most frequently on social media?
- Which billionaires fly the most?⁶
- ...

⁶Fun fact: Taylor Swift is a billionaire according to Forbes.

⁷konkretisieren

Overview

- 1 Music
- 2 Kiel
- 3 Gaming
- 4 Gigachad-Gpt
- 5 Supermarket
- 6 Illnesses & University
- 7 Rich people
- 8 AI embeddings on images

AI embeddings on images

Publicly available vector embedders of Images, own autoencoder⁹, MNIST, 80 Million Tiny Images, etc.

- Do different encoders have similar variance on the same datasets?

⁸Hab das zufälligerweise gestern Abend gemacht mit den Bildern aus dem Medienpädagogik-Projekt vom dem ich erzählt hatte, war sehr interessant. Der Encoder hatte dabei 512 Dimensionen, also kann sowas auch sehr lange dauern zu berechnen.

⁹Falls wir genug Zeit haben solch ein Modell zu trainieren

AI embeddings on images

Publicly available vector embedders of Images, own autoencoder⁹, MNIST, 80 Million Tiny Images, etc.

- Do different encoders have similar variance on the same datasets?
- How do embeddings of different datasets behave?

⁸Hab das zufälligerweise gestern Abend gemacht mit den Bildern aus dem Medienpädagogik-Projekt vom dem ich erzählt hatte, war sehr interessant. Der Encoder hatte dabei 512 Dimensionen, also kann sowas auch sehr lange dauern zu berechnen.

⁹Falls wir genug Zeit haben solch ein Modell zu trainieren

AI embeddings on images

Publicly available vector embedders of Images, own autoencoder⁹, MNIST, 80 Million Tiny Images, etc.

- Do different encoders have similar variance on the same datasets?
- How do embeddings of different datasets behave?
- How performant are trained encoders on different datasets?

⁸Hab das zufälligerweise gestern Abend gemacht mit den Bildern aus dem Medienpädagogik-Projekt vom dem ich erzählt hatte, war sehr interessant. Der Encoder hatte dabei 512 Dimensionen, also kann sowas auch sehr lange dauern zu berechnen.

⁹Falls wir genug Zeit haben solch ein Modell zu trainieren

AI embeddings on images

Publicly available vector embedders of Images, own autoencoder⁹, MNIST, 80 Million Tiny Images, etc.

- Do different encoders have similar variance on the same datasets?
- How do embeddings of different datasets behave?
- How performant are trained encoders on different datasets?
- What embeddings of one encoder form clusters across one dataset?

⁸Hab das zufälligerweise gestern Abend gemacht mit den Bildern aus dem Medienpädagogik-Projekt vom dem ich erzählt hatte, war sehr interessant. Der Encoder hatte dabei 512 Dimensionen, also kann sowas auch sehr lange dauern zu berechnen.

⁹Falls wir genug Zeit haben solch ein Modell zu trainieren

AI embeddings on images

Publicly available vector embedders of Images, own autoencoder⁹, MNIST, 80 Million Tiny Images, etc.

- Do different encoders have similar variance on the same datasets?
- How do embeddings of different datasets behave?
- How performant are trained encoders on different datasets?
- What embeddings of one encoder form clusters across one dataset?
- How do clusters change during dimensionality reduction via PCA?

⁸Hab das zufälligerweise gestern Abend gemacht mit den Bildern aus dem Medienpädagogik-Projekt vom dem ich erzählt hatte, war sehr interessant. Der Encoder hatte dabei 512 Dimensionen, also kann sowas auch sehr lange dauern zu berechnen.

⁹Falls wir genug Zeit haben solch ein Modell zu trainieren

AI embeddings on images

Publicly available vector embedders of Images, own autoencoder⁹, MNIST, 80 Million Tiny Images, etc.

- Do different encoders have similar variance on the same datasets?
- How do embeddings of different datasets behave?
- How performant are trained encoders on different datasets?
- What embeddings of one encoder form clusters across one dataset?
- How do clusters change during dimensionality reduction via PCA?
- Can we identify different clusters by eye with a 2D-PCA?⁸

⁸Hab das zufälligerweise gestern Abend gemacht mit den Bildern aus dem Medienpädagogik-Projekt vom dem ich erzählt hatte, war sehr interessant. Der Encoder hatte dabei 512 Dimensionen, also kann sowas auch sehr lange dauern zu berechnen.

⁹Falls wir genug Zeit haben solch ein Modell zu trainieren