

Datenpipeline zur Verarbeitung von Daten der Datenquelle

HyStreet.com

Mathieu, Daniel, Kai

01-01-2023

hystreet.com

Inhaltsverzeichnis

| | | |
|----------|------------------------------|----------|
| 1 | Einführung | 1 |
| 1.1 | Motivation | 1 |
| 2 | Datenpipeline | 2 |
| 2.1 | Einlesen der Daten | 2 |

1 Einführung

Ziel dieser Arbeit ist die Erstellung einer Datenpipeline, die die Daten der Datenquelle HyStreet verarbeitet, speichert und anschließend analysiert sowie visualisiert. Die Datenquelle stellt Informationen zur Anzahl an Passanten pro Stunden zu verschiedenen Standorten zur Verfügung, Genauerer dazu in den nächsten Kapiteln. Zunächst soll geklärt werden, aus welcher Motivation und mit welchem Ziel die Daten analysiert wurden.

1.1 Motivation

— Motivation —

2 Datenpipeline

Im nachfolgenden Teil der Arbeit werden die einzelnen Elemente der erstellte Datenpipeline nun genauer besprochen und erläutert. Dazu wird in Abbildung 1 zunächst die Datenpipeline als Ganzen betrachtet.

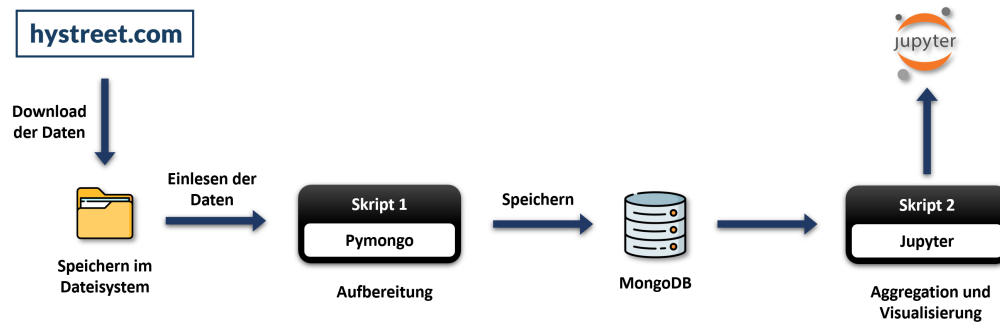


Figure 1: Datenpipeline zur Verarbeitung der HyStreet-Daten

Wie bereits erwähnt, stammen die Daten aus der Internetseite HyStreet.com. Die Webseite stellt somit auch den Startpunkt der Datenpipeline dar. Da die Daten nur kostenpflichtig automatisiert über eine API Schnittstelle heruntergeladen werden konnten, mussten wir uns für die Alternative entscheiden, die Daten manuell über deren Webseite herunterzuladen. Über den manuellen Download ist der Erhalt der Daten kostenlos. Zum genauen Vorgang des Downloads im folgenden Abschnitt Genauerer. Das Ergebnis dieses Schrittes ist eine Dateien im CSV-Format für den jeweiligen Standort. Der Aufbau dieser Dateien wurde bereits im einführenden Kapitel genauer beschrieben. Nach dem Herunterladen, werden die Dateien nun zunächst im lokalen Dateisystem gespeichert, um anschließend im Python Skript *Skript 1* verarbeitet werden zu können.

In diesem Skript werden die Daten aus allen heruntergeladenen Dateien mit Hilfe des Packages *Pandas* eingelesen und verarbeitet. Dabei werden diverse Spalten in ihre Einzelteile zerlegt und Spalten, die für unsere Zwecke nicht weiter erforderlich sind entfernt. Nach der Aufbereitung der Daten, werden diese schließlich in das JSON-Format umgewandelt, um im letzten Schritt in eine MongoDB eingelesen zu werden.

2.1 Einlesen der Daten

List of Figures

| | | |
|---|---|---|
| 1 | Datenpipeline zur Verarbeitung der HyStreet-Daten | 2 |
|---|---|---|

List of Tables