

Projekt Data Science - Datenexploration

st178902

November 2023

1 Datenverständnis

1.1 Datensätze

1.1.1 Datensatz A ¹

Dateien:

- passagierfrequenz.csv

Beschreibung:

- Ein- und Aussteigende an Bahnhöfen

Herausgeber:

- Schweizerische Bundesbahnen SBB

Datenerfassung:

–

1.1.2 Datensatz B ²

Dateien:

- Fahrzeiten_SOLL_IST_[...].csv *52
- Haltepunkt.csv
- Haltestelle.csv

Beschreibung:

- Wöchentliche SOLL und IST Vergleiche der Fahrten und zwei Matching Tabellen

Herausgeber:

- Schweizerische Bundesbahnen SBB

Datenerfassung:

–

¹<https://data.europa.eu/data/datasets/passagierfrequenz-schweizerische-bundesbahnen-sbb>

²<https://data.europa.eu/data/datasets/878a98b8-4973-4d76-858e-eddd88652d9f-stadt-zurich>

1.2 Attribute und deren Bedeutung

1.2.1 Datensatz A

passagierfrequenz.csv ³:

Feld	Beschreibung	Datentyp
code_codice	Dienststellendokumentation-Code des Bahnhofes/der Haltestelle	string
uic	UIC-Identifikation des Bahnhofes/der Haltestelle	float
bahnhof_gare_stazione	Name des Bahnhofes/der Haltestelle	string
kt_ct_cantone	Kantonszugehörigkeit des Bahnhofs/der Haltestelle	string
isb_gi	Bahnhofseigner (Infrastrukturbetreiberin)	string
jahr_annee_anno	Bezugsjahr	float
dtv_tjm_tgm	Durchschnittlicher täglicher Verkehr (Montag bis Sonntag)	float
dww_tmjo_tfm	Durchschnittlicher werktäglicher Verkehr (Montag bis Freitag)	float
dnwv_tmjno_tmgnl	Durchschnittlicher nicht-werktäglicher Verkehr (Samstage, Sonntage, Feiertage)	float
evu_ef_itf	Einbezogene Eisenbahnverkehrsunternehmen für den Bahnhof oder die Haltestelle	string
bemerkungen	Ergänzende Informationen zum Eintrag	TODO
remarques	Bemerkungen in Französisch	TODO
note	Bemerkungen in Spanisch	TODO
remarks	Bemerkungen in Englisch	TODO
geopos	Geographische Position in Latitude, Longitude	string
Iod	Link zu der Stationenbeschreibung	string

1.2.2 Datensatz B

Haltestelle.csv ⁴:

Feld	Beschreibung	Datentyp
halt_id1	Schlüsselfeld Haltestelle.	int(19)
halt_diva	Die Haltestellennummer.	int(10)
halt_kurz	Die Haltestellenabkürzung.	varchar(8)
halt_lang	Haltestellenbezeichnung.	varchar(40)
halt_ist_aktiv	Haltestelle aktiv/inaktiv. Werte: true/false.	boolean: bit(1)

Haltepunkt.csv ⁴:

Feld	Beschreibung	Datentyp
halt_punkt_id	Schlüsselfeld Haltepunkt.	int(19)
halt_punkt_diva	Die Haltepunktnummer (redundant vorkommend in Fahrzeiten_SOLL IST). Werte 0-99.	int(2)
halt_id	Schlüsselfeld Haltestelle.	int(19)
Fortsetzung auf der nächsten Seite		

³<https://opentransportdata.swiss/dataset/02569ba8-b029-4d75-ad59-a6ffd4b6b569>

Table 3 – Fortsetzung

Feld	Beschreibung	Datentyp
GPS_Latitude	Die Y-Koordinate (geographische Breite) des Haltepunktes in WGS84.	int(10)
GPS_Longitude	Die X-Koordinate (geographische Länge) des Haltepunktes in WGS84.	int(10)
GPS_Bearing	Die Kompassrichtung des Haltepunktes.	int(10)
halt_punkt_ist_aktiv	Haltestelle aktiv/inaktiv. Werte: true/false.	boolean: bit(1)

Fahrzeiten_SOLL_IST_[...].csv *52 ⁴:

Feld	Beschreibung	Datentyp
linie	Die von der VBZ benutzte Liniennummer (z.T. nicht identisch mit publizierter Linie).	int(3)
richtung	Die Richtung einer Fahrt. Werte: 1 oder 2.	int(1)
betriebsdatum	Das Tagesdatum für das Haltestellenereignis. Betriebstag: 5:00 Uhr bis nächster Tag 01:30 Uhr.	datetime(8)
fahrzeug	Die Fahrzeugnummer.	int(10)
kurs	Die Kursnummer.	int(2)
seq_von	Die Sequenz der "von"-Haltestelle einer Fahrt.	int(10)
halt_diva_von	Die Haltestellennummer DIVA der "von"-Haltestelle.	int(10)
halt_punkt_diva_von	Die Haltepunktnummer von der "von"-Haltestelle.	int(10)
halt_kurz_von1	Die Haltestellenabkürzung der "von"-Haltestelle.	varchar(8)
datum_von	Das Tagesdatum für das Haltestellenereignis der "von"-Haltestelle.	datetime(8)
soll_an_von	Die Sollankunft der "von"-Haltestellen in Sekunden nach Mitternacht. Hinweis: ⁵	int(10)
ist_an_von	Die Istankunft der "von"-Haltestellen in Sekunden nach Mitternacht. Hinweise: ⁶ ⁷	int(10)
soll_ab_von	Die Sollabfahrt der "von"-Haltestellen in Sekunden nach Mitternacht. Hinweis: ⁸	int(10)
ist_ab_von	Die Istabfahrt der "von"-Haltestellen in Sekunden nach Mitternacht. Hinweise: ⁶ ⁷	int(10)
seq_nach	Die Sequenz der "nach"-Haltestelle einer Fahrt.	int(10)
halt_diva_nach	Die Haltestellennummer DIVA der "nach"-Haltestelle.	int(10)
halt_punkt_diva_nach	Die Haltepunktnummer von der "nach"-Haltestelle.	int(10)
halt_kurz_nach1	Die Haltestellenabkürzung der "nach"-Haltestelle.	varchar(8)

Fortsetzung auf der nächsten Seite

⁴https://data.stadt-zuerich.ch/dataset/vbz_fahrzeiten_ogd_2022

⁵Enthält Werte zwischen 0 und 86399.

⁶Kann Werte größer als 86400 enthalten, wenn "Betriebsdatum" und "Datum von"/"Datum nach" gleich sind und die Istankunft/-abfahrt nach Mitternacht erfolgte, obwohl diese vor Mitternacht geplant war.

⁷Kann negative Werte enthalten, wenn "Betriebsdatum" und "Datum von"/"Datum nach" ungleich sind und die Istankunft/-abfahrt vor Mitternacht erfolgte, obwohl diese nach Mitternacht geplant war.

⁸Kann Werte größer als 86400 enthalten, wenn die Ankunft vor und die Abfahrt nach Mitternacht geplant ist.

Table 4 – Fortsetzung

Feld	Beschreibung	Datentyp
datum_nach	Das Tagesdatum für das Haltestellenereignis der "nach"-Haltestelle.	datetime(8)
soll_an_nach	Die Sollankunft der "nach"-Haltestellen in Sekunden nach Mitternacht. Hinweis: ⁵	int(10)
ist_an_nach1	Die Istankunft der "nach"-Haltestellen in Sekunden nach Mitternacht. Hinweise: ^{6 7}	int(10)
soll_ab_nach	Die Sollabfahrt der "nach"-Haltestellen in Sekunden nach Mitternacht. Hinweis: ⁸	int(10)
ist_ab_nach	Die Istabfahrt der "nach"-Haltestellen in Sekunden nach Mitternacht. Hinweise: ^{6 7}	int(10)
fahrt_id	Die ID einer Fahrt pro Betriebstag.	int(19)
fahrweg_id	Der Fremdschlüssel zum Fahrwegstamm.	int(19)
fw_no	Die Fahrwegnummer.	int(5)
fw_typ	Der Fahrwegtyp.	int(1)
fw_kurz	Der Fahrweg-Kurzcode.	varchar(6)
fw_lang	Die Fahrwegbezeichnung.	varchar(40)
umlauf_von	Die ID eines Umlaufes von Fahrten pro Betriebstag. Entspricht einer Sammlung, Reihenfolge von Fahrten.	int(19)
halt_id_von	Fremdschlüssel zum Haltestellenstamm (halt_id); Zuordnung der "von"-Haltestelle.	int(19)
halt_id_nach	Fremdschlüssel zum Haltestellenstamm (halt_id); Zuordnung der "nach"-Haltestelle.	int(19)
halt_punkt_id_von	Fremdschlüssel zum Haltepunktstamm (halt_punkt_id); Zuordnung des "von"-Haltepunktes.	int(19)
halt_punkt_id_nach	Fremdschlüssel zum Haltepunktstamm (halt_punkt_id); Zuordnung des "nach"-Haltepunktes.	int(19)

2 Potenzielle Zusätzliche Daten

- Fahrzeiten SOLL IST für den 01.01.2022 ⁹
- Fahrzeiten SOLL IST von anderen Jahren ¹⁰
- Haltestellenliste aktuell/zukünftig ¹¹
- Datensammlung von vielen weiteren Datensätzen (z.B. Routenplanung, Haltestellen etc.) ¹²
- Weitere Daten ¹³
- Nationale Grenzen (z.B. Kantongrenzen) ¹⁴

3 Hilfreiche Links

- Beispiel Verwendung des Datensatzes ¹⁵

⁹https://data.stadt-zuerich.ch/dataset/vbz_fahrzeiten_ogd_2021/resource/063e363f-1b60-4a4d-b328-d85ecc0e11b9

¹⁰https://data.stadt-zuerich.ch/dataset?q='SOLL-IST-Vergleich'&sort=title_string+desc

¹¹https://opentransportdata.swiss/de/dataset/bav_liste

¹²https://opentransportdata.swiss/de/cookbook/#Noch_nicht_fuendig_geworden

¹³<https://data.stadt-zuerich.ch>

¹⁴<https://www.swisstopo.admin.ch/en/geodata/landscape/boundaries3d.html>

¹⁵https://github.com/VerkehrsbetriebeZuerich/ogd_examples_python

- Liniennetz Zürich ¹⁶

¹⁶https://ethz.ch/content/dam/ethz/special-interest/math/mathematical-research/fim-dam/guests/information-for-FIM-guests/Liniennetzplan_Stadt_Zuerich_2014.pdf