Public Transport Validation Dataset (pt_data.zip)

Overview

The pt_data.zip archive contains public transport validation data prepared for this course. The dataset covers the period 1 November 2019 – 30 March 2020 and provides passenger validations aggregated at a 15-minute interval for the three main modes of public transport: bus, tramway, and metro (subway). In addition, spatial reference files for bus/tram stops and metro stations are included. A Jupyter notebook is provided to demonstrate how to load and explore these data.

File Structure

```
pt_data.zip
 ├── bus_indiv_15min/
 │    └── bus_indiv_15min.csv
 ├── tramway_indiv_15min/
 │    └── tramway_indiv_15min.csv
 ├── subway_indiv_15min.csv
 ├── ref_tramway_bus.csv
 └── ref_subway.csv
```

Data Content

1. Bus Validations (bus_indiv_15min.csv)

Each row corresponds to validations observed at a bus stop within a 15-minute interval.

Columns:
- LIG_NUMERO_SAE – internal bus line number
- VAL_ARRET_CODE – stop identifier
- CRS_SENS_TRAJET – course / direction indicator (0, 1, …)
- COD_LIG_CLI – client-facing line code
- VAL_DATE – timestamp of the 15-minute interval
- Flow – number of validations observed in that interval at the stop

Example:
LIG_NUMERO_SAE,VAL_ARRET_CODE,CRS_SENS_TRAJET,COD_LIG_CLI,VAL_DATE,Flow
2,202,1,2,2019-11-01 12:15:00,1

2. Tramway Validations (tramway_indiv_15min.csv)

Same structure as the bus file, but for tramway stops.

Columns:
- LIG_NUMERO_SAE
- VAL_ARRET_CODE
- CRS_SENS_TRAJET
- COD_LIG_CLI – for example, T1, T2 …
- VAL_DATE
- Flow

Example:
LIG_NUMERO_SAE,VAL_ARRET_CODE,CRS_SENS_TRAJET,COD_LIG_CLI,VAL_DATE,Flow
520,32102,0,T1,2019-11-02 15:00:00,66

3. Metro (Subway) Validations (subway_indiv_15min.csv)

Data are aggregated differently: one row per 15-minute interval, with each column (after the timestamp) representing a station code. The value is the number of validations recorded at that station during the interval.

Columns:
- VAL_DATE – timestamp of the 15-minute interval
- Then one column per station, e.g. AMP, BEL, BRO, CHA, COR, …

Example (truncated):
VAL_DATE,AMP,BEL,BRO,CHA,COR,...
2019-11-01 00:00:00,20.0,164.0,10.0,59.0,88.0,...

4. Spatial Reference for Subway Stations (ref_subway.csv)

Reference information for metro stations.

Columns:
- IDT_SIT – station identifier
- MEAN_X, MEAN_Y – projected coordinates
- COD_TRG, COD_STA – codes used in operations
- LIB_STA_SIFO – station name

5. Spatial Reference for Bus and Tramway Stops (ref_tramway_bus.csv)

Reference information for stops.

Columns include:
- IDT_PNT – stop identifier
- NOM_PNT, NOM_HASTUS – stop names (variants)

- COD_UTE – operator code
- FLG_PNT_ARR, FLG_PNT_DPO, FLG_PNT_GAR, FLG_PNT_VIA – flags indicating stop usage (arrival, departure, garage, via)
- IDT_LIE, LIB_ADR, COD_INSEE, COD_PST, LIB_CMN – address and administrative information
- COO_X, COO_Y, COO_X_WGS84, COO_Y_WGS84 – stop coordinates (projected and WGS84)
- CAP – capacity or related field

Example row:
IDT_PNT,NOM_PNT,NOM_HASTUS,COD_UTE,FLG_PNT_ARR,...,COO_X_WGS84,COO_Y_WGS84,CAP
30093,Gare de Vaise,,UTV,O,...,4.804221040093,45.780507677991,33.0

Using the Data

- Load the CSV files with pandas (pd.read_csv) and parse VAL_DATE as datetime.
- Explore validations at different temporal scales (daily, weekly).
- Compare demand patterns across modes (bus vs. tram vs. metro).
- Use the reference files to map stops/stations spatially.

The Jupyter notebook provided illustrates how to read the files, convert dates, merge the PT usage data with the spatial reference, and make first plots.