Traffic Detector Data

Start Date	2022-01-01 00:00:00
End Date	2023-06-30 23:59:59
N. of Files	18 (1 per month)

Origin

This data comes from traffic detectors spread throughout the Porto municipality. These detectors are induction loops (see Wikipedia entry to get an idea) which can detect large metallic objects above them - such as cars, buses, trucks, etc.
Porto's detectors measure two things over a **time period of ten minutes**:

- Count: the number of vehicles per unit of time.
- Occupancy: the percentage of time that the a metallic object was detected over the loop.

Note that during the period January 2022 - June 2023 there was rapid expansion of the number of detectors in place. Therefore, for some detectors you do not have data from January 2022, but only starting later on.

Transformations

Data has been transformed to follow the TrafficFlowObserved data model from Fiware. This means that:

- **Count** is converted to **intensity** (also called **flow** in traffic engineering), in vehicles per hour (VPH). This is just the count over 10 minutes multiplied by 6 (hence all values will be multiples of 6 - that's expected).
    - For example, if 4 vehicles were detected in 10 minutes, then the intensity is 24 vehicles per hour.
- Occupancy is a number between 0 and 100, representing a percentage of time occupied.
    - For example, a value of 5 represents 5% of time where a vehicle was detected over the loop.

Data Dictionary

Each .ZIP file contains data for a whole month, for all detectors. Inside each .ZIP file there will be one file per detector, with a name like CMP_001.json. CMP stands for "Câmara Municipal do Porto", and "001" is the Controller ID. Each controller collects data from 1-3 intersections, and it has multiple detectors feeding data into it.

Inside each JSON file you'll have a long list of objects, each of which with the following keys and values in the table below.

Name	Description	Type	Default	Example
dateObservedFrom	Start of the observation period	Datetime	null	"2023-01-01T00:00:00+00:00"
dateObservedTo	End of the observation period	Datetime	null	"2023-01-01T00:10:00+00:00"
Id	ID of the traffic detector. The last part of this ID should be of the form CMP_xxx-yyyy, where xxx is the ID of the controller and yyyy the ID of the detector within that controller. **Remember that the pair (xxx, yyyy) is what uniquely identifies an induction loop!**	String	null	"urn:ngsi-ld:TrafficFlowO…orto:traffic:CMP_001-e1" (Corresponding to detector CMP_001-e1)
location:value:coordinates	Geolocation of the controller. This is NOT the geolocation of the loop, but of the controller, which roughly maps to the intersection where the detector is placed (but not always).	Geolocation	null	0: 41.159844 1: -8.624687
Intensity	Vehicles per hour observed in that time period. Corresponds to 6 times the actual count over the 10 minute period.	Int	0	12 (Corresponding to 2 vehicles observed in those 10 minutes)

Occupancy	Percentage of time that a metallic object was detected over the loop.	Int	100	2 (Corresponding to 2% of the 10 minute interval having a metallic object over the loop)

Important Notes

- Remember that you need both the controller ID and the detector ID to uniquely identify a detector. For example, just “e3” as a detector ID might mean CMP_001-e3, or CMP_002-e3, etc.
- **Sadly, a count of zero during a 10-minute period is not stored for that time period, and therefore it’s difficult to distinguish that from a genuine missing value.**
- Not all loops have the same size:
    - Some loops cover a larger part of the street lengthwise, compared to others. You can assume a length (along the street) of the order of 2 meters, as in the Wikipedia link above, but you cannot assume they all have the same length.
    - Likewise, some loops cover a single lane, while others cover 2 or even 3 lanes.
    - It’s safe to assume that the same detector (in the same controller) has the same size throughout the dataset’s time period.
- As mentioned earlier, some detectors (most, in fact) only started having data sometime during the period Jan 2022 - Jun 2023 that you are studying. **For this reason, inside each .ZIP file there are multiple JSON files that are just an empty list “[]” (opening square brackets, then closing square brackets).** Those represent controllers for which there was no data during that month.
- Some detectors start with the letter “b”, as in CMP_001-b1. In these cases, intensity is counting the number of button presses by pedestrians in crosswalks with traffic lights. We recommend you start with the “e” detectors and ignore the “b” ones at first, but if the rest of your solution is stable, feel free to explore these as well.
- Some detectors have names like “rdp” or “rdm”, as in CMP_001-rdm1. These are also counting traffic, but they do so using radars instead of induction loops. From our experience, the calibration of these radars is quite different from induction loops, so you should be cautious when comparing absolute values from one type to the other type
- There is a separate file under a folder called “OpenStreetMap” which contains a mapping from detector names to OpenStreetMap ways. You can see more information in the README in that folder.