## 1. Old data (2006 only)

A list of names and numbers

AI-generated content may be incorrect.

Only 4335 have pedestrian counts (people walking across the street)

Most of the scat is INT (intersection) except 4335 have both of them. (These 4335 INT and POS have different *HF VicRoads Internal, 15772 and 5485)* by observing, the 15772 have less flow at 00:00 => might be pedestrian count since people may be less than cars at that time (common sense ?)

4/5/2025: current solution, flow (car count) += ½ pedestrian count (15772).

2. Data Source:

- public holiday: [https://data.gov.au/dataset/ds-dga-b1bc6077-dadd-4f61-9f8c-002ab2cdff10/details?q=Australian%20public%20holidays%20combined%C2%A0(2021%E2%80%91{})](https://data.gov.au/dataset/ds-dga-b1bc6077-dadd-4f61-9f8c-002ab2cdff10/details?q=Australian%20public%20holidays%20combined%C2%A0(2021%E2%80%91%7b%7d))

- 2014-2024: <https://opendata.transport.vic.gov.au/dataset/traffic-signal-volume-data>

- traffic light (geo information for each scat): <https://discover.data.vic.gov.au/dataset/traffic-lights>

- school zone: <https://discover.data.vic.gov.au/dataset/school-locations-2024?utm_source=chatgpt.com>

5/5/2025

Since scat\_type.csv lack of some scat, these missing scats then fill by using Traffic\_lights.csv.

5/5/2025:

Data flow: Raw data from 2014-2024 => transform\_traffic\_data.py => 2014-2024\_transformed\_scats\_data.csv => feature\_engineering.py => 11 featured\_scats\_data.csv

(but when I upload to huggingface, these datasets already aggregated)