1. Open Government & Municipal Portals

Cities around the world publish traffic sensor, camera, and vehicle-GPS data. Examples:

- **NYC Open Data**: Taxi & Limousine Commission trip records (pick-up/drop-off lat-longs, timestamps) and DOT traffic speed cameras.
- **Chicago Data Portal**: Traffic tracker data with sensor counts and speeds.
- **London Datastore**: Live traffic camera snapshots and historical counts.

Most portals let you query via REST API or download CSV/JSON in bulk.

---

2. Public Transportation & Mobility APIs

Transit agencies often expose real-time feeds:

- **GTFS-RT (General Transit Feed Specification – Real Time)**: Vehicle positions, trip updates for buses/trains.
- **OpenTraffic (HERE)** or **TomTom Traffic API**: Global traffic flow and incidents (free tiers available).

You'll need to register for API keys, and you can pull JSON or Protocol Buffer feeds at regular intervals.

---

3. Crowdsourced & Probe Data

Services that aggregate anonymized GPS traces from apps:

- **Strava Metro**: Cycling and running activity heatmaps over time.
- **Mozilla Location Service**: Contributes Wi-Fi and cell-tower–based positions.
- **OpenStreetMap (OSM) Tile Usage**: Not directly traffic, but you can infer popular routes from tile-request heatmaps.

Some require partnerships or academic access.

---

4. Satellite & Aerial Imagery

For batch analysis of congestion (e.g., parking lot fill rates or roadway density):

- **Sentinel-2** or **Landsat**: Free, multispectral imagery every 5–16 days.
- **Planet Labs**: High-cadence commercial imagery (academic/free trial for research).

You'll classify roads/vehicles via computer vision pipelines.

5. Simulated & Synthetic Data

If real data is constrained, generate synthetic traffic:

- **SUMO (Simulation of Urban Mobility)**: Open-source traffic micro-simulator.
- **MATSim**: Large-scale agent-based transport simulator.

Great for testing algorithms when you can't get live feeds.

---

6. Web-Scraping or Mobile App Data

- **Google Maps / Waze**: While there's no official public API for third-party traffic, you can capture periodic speed data via headless browser scraping (be mindful of ToS).
- **City Bike / Scooter Schemes**: Many share station-level availability and trip logs (e.g., Citi Bike in NYC).

Tips for Data Collection & Use

- **APIs + ETL Pipeline**: Automate pulls (e.g., cron jobs) and store raw feeds before cleaning.
- **Geo-Indexing**: Use spatial databases like PostGIS to efficiently query by region/road segments.
- **Time Windowing**: Collect data over multiple weeks to capture peak/off-peak variance.
- **Privacy & Compliance**: Always anonymize any raw GPS/vehicle IDs and follow data-use policies.