Database Project Report

Urban Planning Analysis Team

2025-11-19

Contents

1 Cover Page

Database Project Report
DBM, HS 2025
Traffic Flow Analysis for Urban Planning in Zurich
Team: ACID

Authors: \<Author 1\>, \<Author 2\>, \<Author 3\>, \<Author 4\>
Date:

2 Introduction & Context (Proposal)

The City of Zurich faces growing challenges related to traffic congestion, mobility planning, and urban development. As the population increases and commuting patterns evolve, many intersections and arterial roads experience significant pressure, particularly during peak hours. Congestion affects not only travel times but also road safety, air quality, and the overall effectiveness of the city's transport network.

Urban planning authorities must therefore make complex decisions about where to prioritize infrastructure investments, how to optimize traffic signals, and which areas require redesign or alternative mobility solutions. These decisions rely heavily on an accurate understanding of traffic flows, temporal patterns, directional imbalances, and long-term trends. High-quality traffic data and transparent analytical methods are essential to support evidence-based planning, enable more targeted interventions, and ensure that resources are allocated efficiently.

To contribute to this goal, this project focuses on deriving meaningful insights and key performance indicators (KPIs) that can inform data-driven urban planning decisions in line with the OECD Data Value Cycle, which emphasizes the transformation of raw data into actionable value for public administration.

3 Project Idea & Use Case (Proposal)

To support Zurich's urban planners in addressing congestion and mobility challenges, this project aims to transform traffic count data into meaningful indicators that can guide evidence-based decision-making. The central use case focuses on identifying traffic pressure points at intersections and understanding how traffic patterns vary by time, location, and direction.

4 Key Performance Indicators (KPIs)

A set of KPIs is derived from the cleaned dataset to quantify traffic behavior across Zurich's intersections. These indicators are designed to highlight temporal and spatial traffic patterns, prioritize areas of concern, and offer measurable signals for potential interventions. Relevant KPIs include:

KPI	Description	Possible Calculation
Average hourly traffic volume per intersection and direction	Shows typical load and recurring high-traffic hours.	Group by intersection, direction, and hour of day; compute average hourly volume.
Daily and weekly traffic volume	Indicates overall traffic pressure at a location.	Sum hourly volumes per day or per week.
Peak-hour traffic index	Identifies the hour with the highest recurring traffic.	Calculate average volume per hour of day; select highest.

KPI	Description	Possible Calculation
Directional imbalance	Reveals asymmetry between opposing traffic flows.	Aggregate inbound and outbound volumes; compute ratio.
Year-over-year traffic trend	Detects long-term increases or decreases.	Compare annual totals or averages across years.
Weekday vs. weekend traffic pattern	Distinguishes commuter-driven from leisure-driven traffic.	Compare average weekday and weekend volumes; compute ratio.

5 Decision Rules for Supporting the Use Case

To ensure that the KPIs effectively support decision-making, threshold-based decision rules are defined. These rules indicate when a traffic location should be flagged for further analysis or planning action.

5.1 Bottleneck Identification

Locations are considered bottlenecks if peak-hour traffic exceeds around 1,000 vehicles per hour or is more than 150% above the average hourly volume. Sustained peak loads above 800 vehicles per hour signal elevated pressure and may justify signal optimization or design adjustments.

5.2 Intersection Prioritization

Intersections are prioritized if they frequently fall within the top 10% of busiest sites, or if daily volumes regularly exceed approximately 20,000 vehicles. Such locations show consistently high demand and warrant strategic planning attention.

5.3 Directional Imbalance Response

If inbound/outbound traffic ratios exceed 1.5, a notable imbalance is present; ratios above 2.0 represent strong directional dominance. These cases may require directional signal adjustments, lane reallocations, or targeted peak-hour measures.

5.4 Trend-Based Planning

Locations showing year-over-year increases greater than 5%, or cumulative growth of over 10% within three years, are treated as emerging long-term pressure points and should be considered in medium- to long-term infrastructure planning.

5.5 Weekday/Weekend Pattern Differentiation

A weekday-to-weekend ratio above 1.4 indicates commuter-dominated traffic, whereas ratios below 0.8 suggest leisure-oriented patterns. This distinction helps determine whether weekday or weekend interventions are more appropriate.

6 Data Model & Database Schema

6.1 Data Preprocessing & Preparation

- Hourly traffic count data from the City of Zurich's open data portal was collected.
- The raw dataset consisted of multiple yearly CSV files (2012–2025), all sharing an identical schema.
- All CSV files were loaded with Python using the Pandas library and combined into a single unified data frame, resulting in a consistent dataset spanning the entire time period.

6.1.1 Raw Dataset Columns

- MSID
- MSName
- ZSID
- ZSName
- Achse
- HNr
- Hoehe
- EKoord
- NKoord
- Richtung
- Knummer
- Kname
- AnzDetektoren
- D1ID–D4ID
- MessungDatZeit
- LieferDat
- AnzFahrzeuge
- AnzFahrzeugeStatus

Because the raw dataset included a mixture of analytical attributes and highly technical metadata, each field was examined to determine its relevance for traffic-flow analysis. The review was guided by the perspective of the City of Zurich's urban planning department, which is primarily interested in temporal and spatial traffic patterns at specific locations and intersections.

6.1.2 Removed Fields

- **Achse** and **HNr**, due to inconsistent content and redundancy with other location information.
- **D1ID–D4ID**, as these detector identifiers are technical metadata without analytical value.
- **LieferDat**, which is a delivery timestamp not required for traffic analysis.

All fields describing the measurement location (counting site, signal name, coordinates), the measurement configuration (measurement site, direction), and the traffic counts themselves were retained. To improve clarity and support further processing and database integration, the remaining column names were translated into English equivalents. In a further preprocessing step, several categorical values contained in German were translated to English. After all preprocessing steps were completed, the resulting cleaned dataset contained 21,721,493 rows and 14 columns.

6.1.3 Cleaned Dataset Columns

- **measurement_site_id**: Unique technical identifier of the measurement site. A measurement site represents a specific traffic-flow direction or lane at a counting location.
- **measurement_site_name**: Technical name of the measurement site. In this dataset, this field contains "Unknown" for all entries.
- **counting_site_id**: Identifier of the counting site, representing the physical location where traffic measurements are collected.
- **counting_site_name**: Human-readable name of the counting site, describing the location (e.g., street).
- **position_description**: A textual descriptor indicating where along the street segment the measurement site is located. Contains "Unknown" for many entries.
- **east_coordinate**: The east coordinate of the measurement site in the Swiss CH1903+ / LV95 reference system.
- **north_coordinate**: The north coordinate of the measurement site in the Swiss CH1903+ / LV95 reference system.
- **direction**: The direction of traffic flow being measured (e.g., "inbound", "outbound").
- **signal_id**: Identifier of the associated traffic signal or intersection controller regulating traffic at the measurement site.
- **signal_name**: Name of the associated traffic signal or intersection.
- **num_detectors**: Number of detectors installed at the measurement site.
- **timestamp**: The timestamp indicating the start of the hourly measurement interval (ISO-8601 format).
- **vehicle_count**: The number of vehicles recorded during the hourly measurement interval.
- **vehicle_count_status**: Indicates how the vehicle count was produced: "Measured", "Missing", or "Imputed".

7 Loading & Transforming the Data

8 Analyzing & Evaluating Data

9 Efficiency & Query Performance

10 Visualization & Decision Support

11 Conclusions & Lessons Learned

12 How to render this report

A single R command will render the file. In an R console run: `rmarkdown::render("report.Rmd")` Make sure your R working directory is the report folder (or use the full path). For PDF output you need a LaTeX engine (e.g. TinyTeX).