Behind the Gates

Deciphering Milan’s Area C Traffic DNA

Ca’ Foscari University of Venice
CM90 – Computer Science and Information Technology
CM0471 – Statistical Inference Learning

Gabriele Pilotto – 902388@stud.unive.it
Academic Year 2025/2026

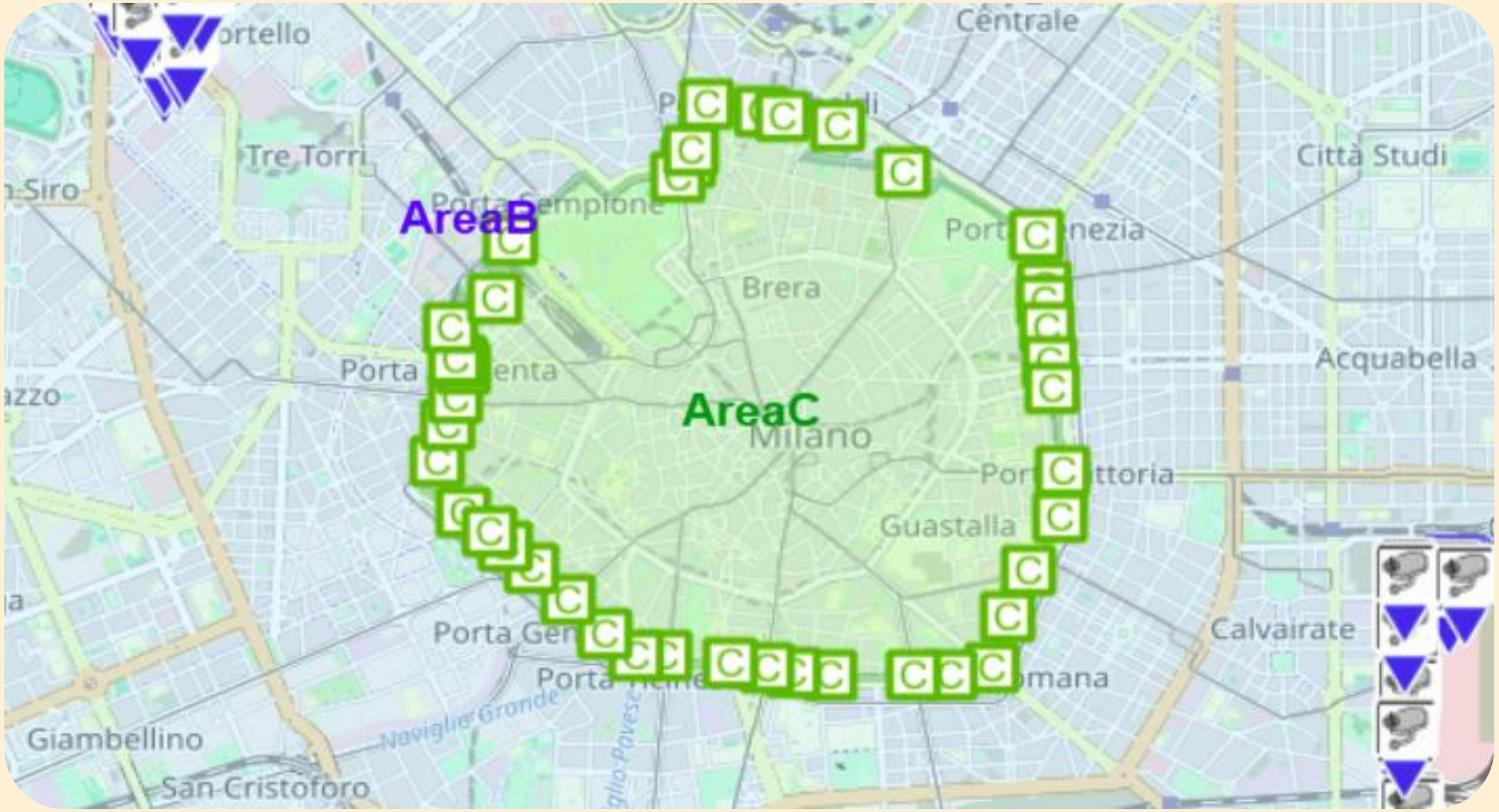
What is Area C?

Area C is the **Limited Traffic Zone (LTZ) of Milan**'s city center, delimited by the *Cerchia dei Bastioni*.

40.000.000[1]

Registered transits
between Jan-Nov 2024

[1] total transits: 39.693.644



Logged information

Area C has **43 gates** logging aggregated information about the number transits and the access details.

Vehicle Info

Including: Motorcycle (y/n), Euro Class, Fuel Type, Vehicle Class, Service Vehicle, FAP

Access Detail

Including: Time, Location, Policy Status, Excluded Users, Resident (y/n), Policy Class

Who used Area C ?

Research questions

Using the aggregated data stored, we can answer the question from **three perspectives**.

Habits

Understand the traffic trends and patterns

Environment

Verify the effectiveness of the enforced policies

Profiling

Identify the residents among all the uses

01

Habits

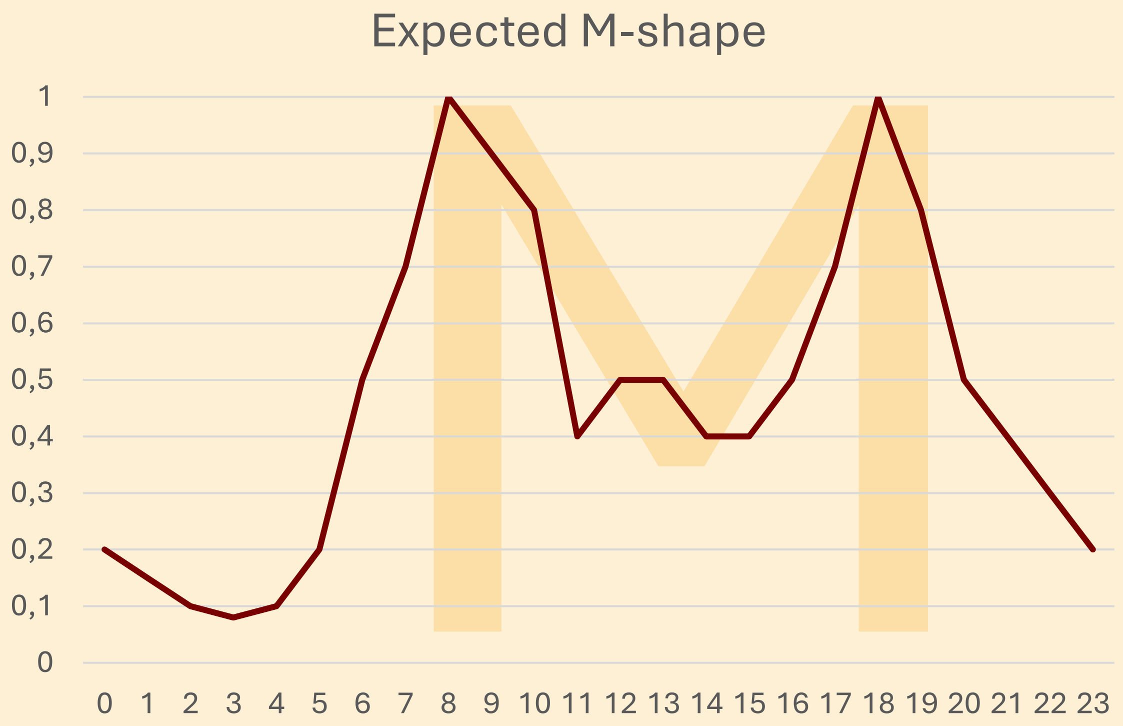
Goal: Understand the influence of rush hour, identify traffic patterns and traffic-increasing contexts.

The rush-hour influence (1/4)

Milan’s city center hosts a lot of offices. **Does the rush hour correspond to office hour?**

Assuming **office hours are 9-18** we would expect a trend characterized by an **“M” shape:**

- **Peak 1:** around 8:00, people go to the offices
- Small increase at launch break
- **Peak 2:** around 18:00, people come back home
- Lowest peak around 3:00

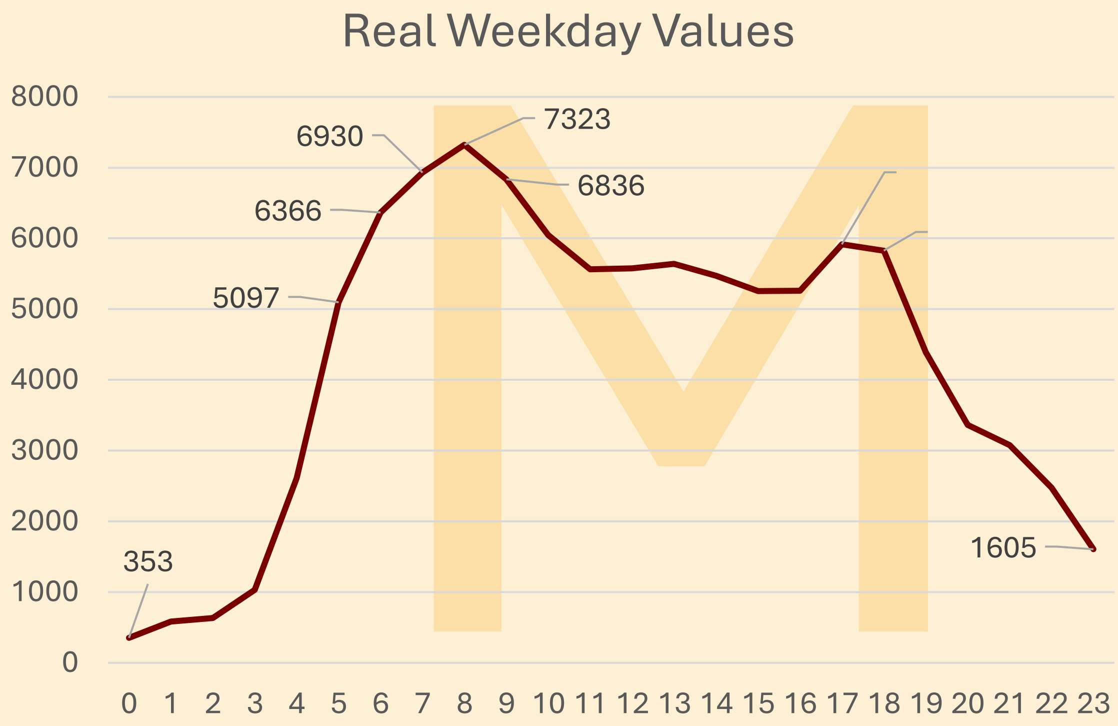


The rush-hour influence (2/4)

Milan’s city center hosts a lot of offices. **Does the rush hour correspond to office hour?**

The actual values for the weekdays show instead:

- **Severe slope** starting 4:00
- **Peak** at 8:00
- **No peak** at 17:00
- **Gentle slope** from 19:00

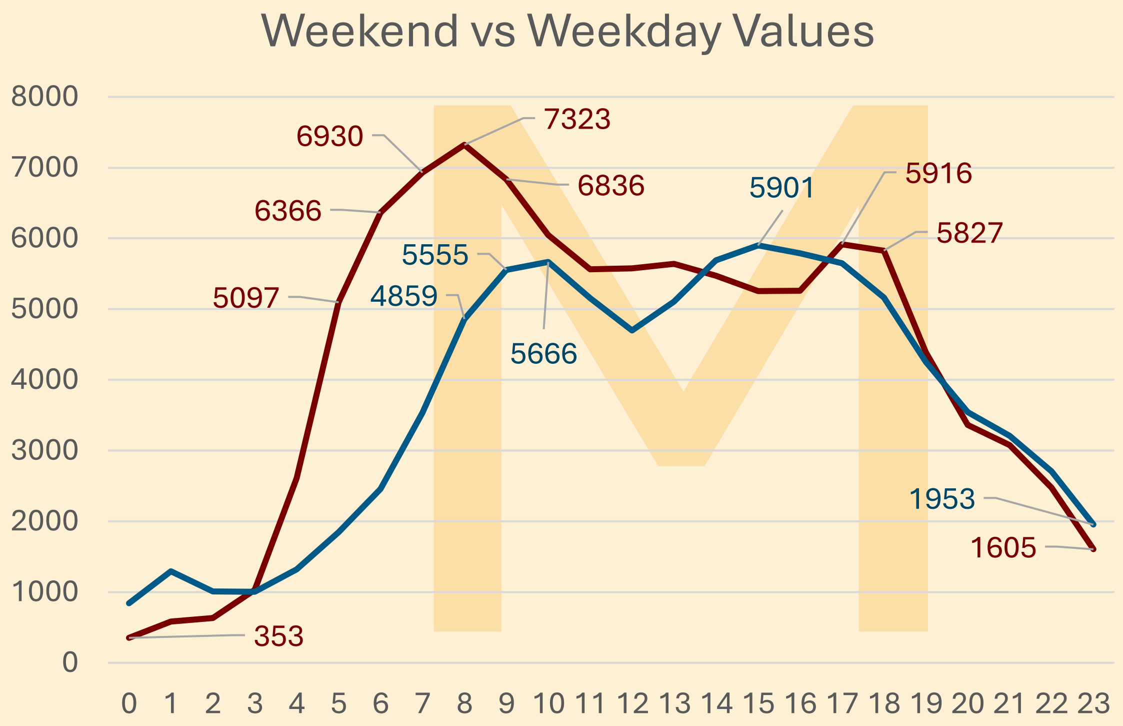


The rush-hour influence (3/4)

Milan’s city center hosts a lot of offices. **Does the rush hour correspond to office hour?**

Are the working activities the real cause? What happen **in the weekends?**

- **Lazy Milan** starts its morning later
- **Less traffic** overall
- **More night** life and traffic
- **Almost no “M”** shape



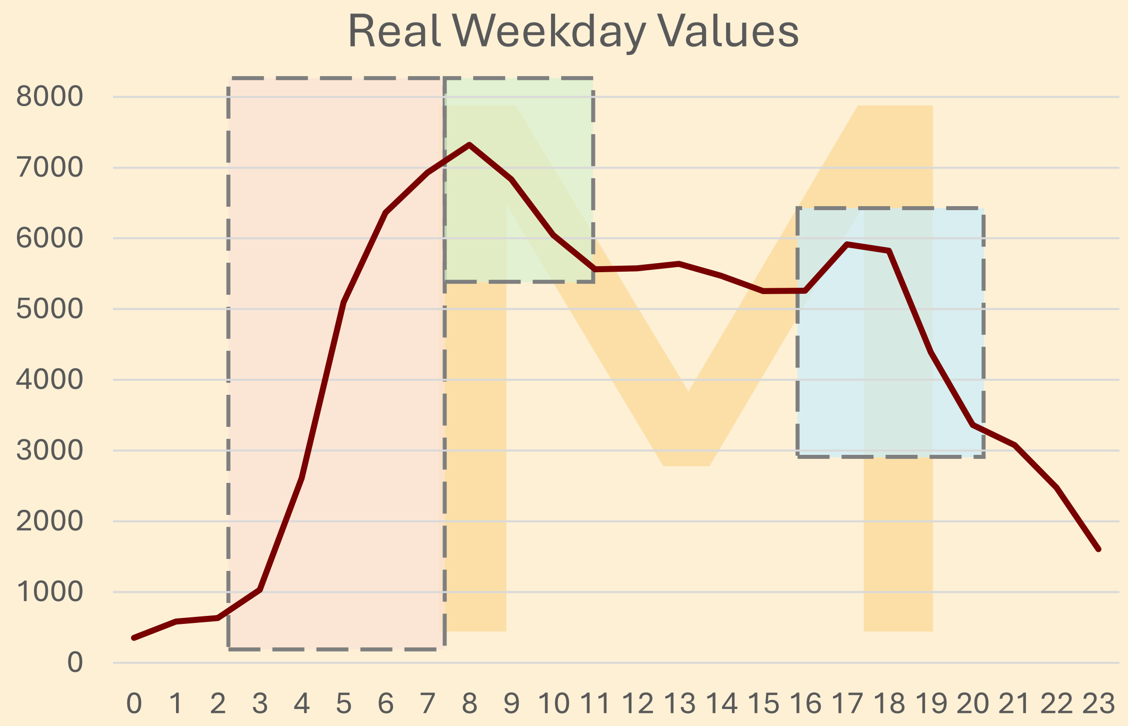
The rush-hour influence (4/4)

Milan’s city center hosts a lot of offices. **Does the rush hour correspond to office hour?**

Yes, but there are two rush hours:

- **Service rush** in pink for logistic
- **Office rush** in green for white collars

A **lazy return** at home can be spotted as show in the box highlighted in blue

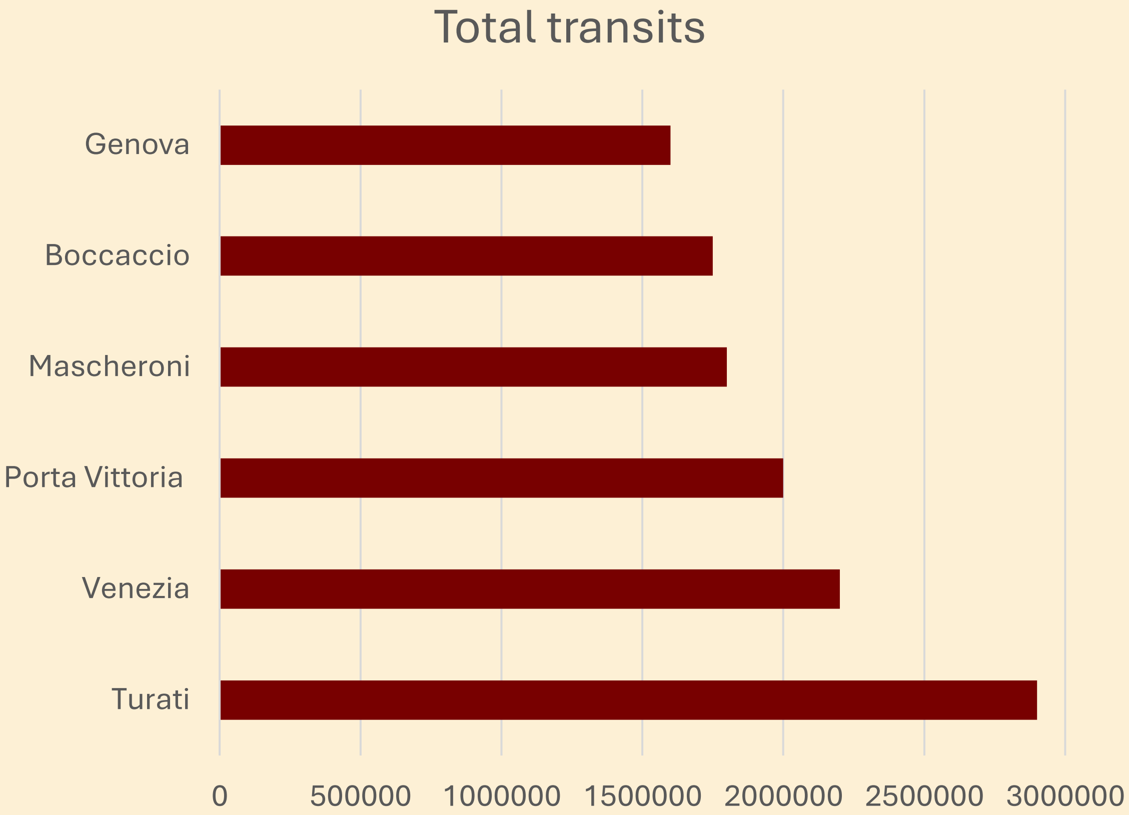


Location influence (1/2)

Among the 43 gates, **are some gates more utilized?**

Plotting the total transits per gate we can spot:

- The **most used** gates:
  - *Turati* with 2,9M transits
  - *Venezia* with 2,15M transits
- The **less used** gates:
  - *Milazzo*
  - *Baretti*



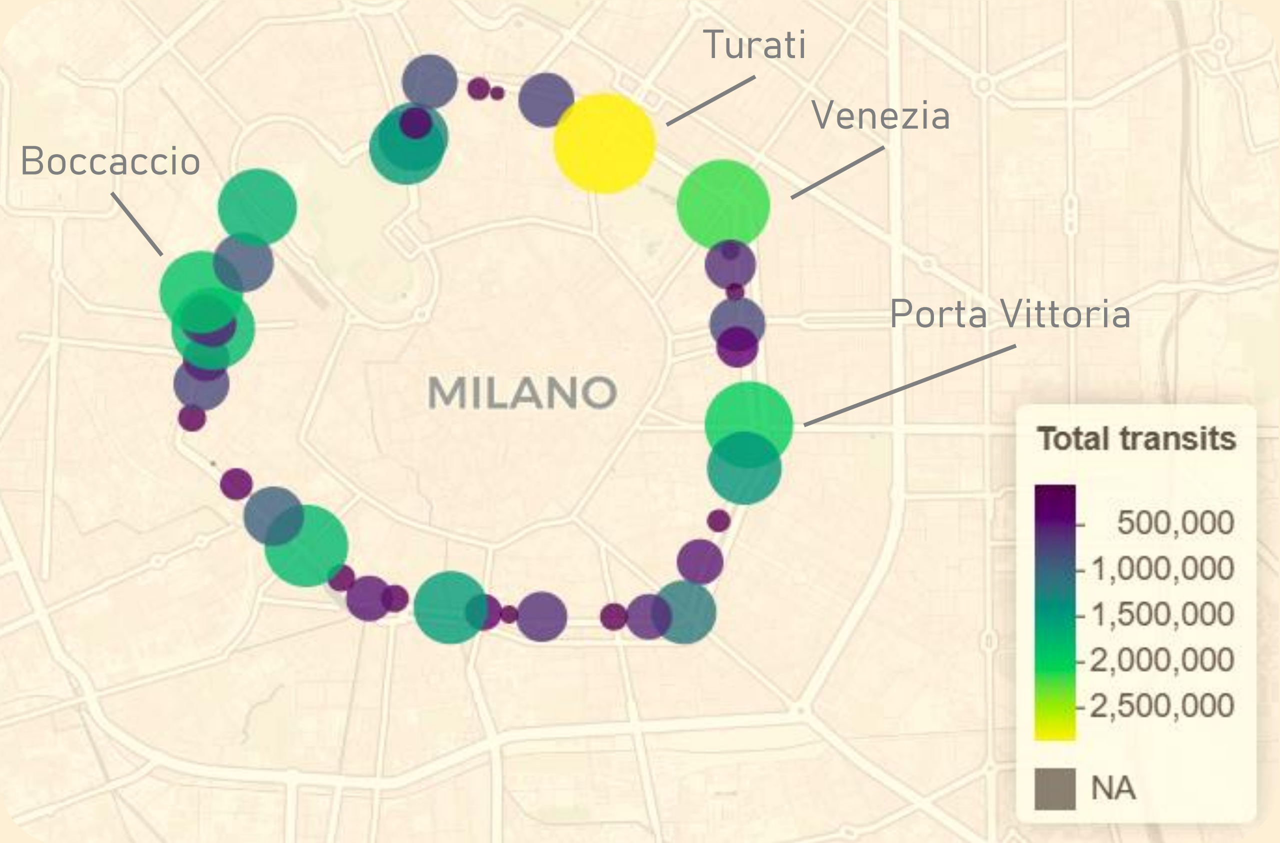
Note how *Turati* stands out among the biggest.

Location influence (2/2)

Among the 43 gates, **are some gates more utilized?**

Why those gates?

- **North-East Axis**
  - *Turati* → Central Station
  - *Venezia* → Commercial District
- **East-West Axis**
  - *Porta Vittoria* → Linate and suburban
  - *Boccaccio* → Residents and Fiera



Predicting the traffic (1/3)

Combining other predictor, such as **location, day of week and month,** we can define a predictive model.

4

Multiple Linear Regression without location

Verdict: treating the traffic as a linear phenomenon isn't a good idea: R^2_{adj} is just at 4%. Good p-values.

3

Polynomial Regression without location

Verdict: more flexibility has led to R^2_{adj} at 18%. Traffic is represented as a reversed "U", not like a "M".

2

Polynomial Iterative Model

Verdict: R^2_{adj} at 72% is a great improvement. Location is very significative. Still high residuals (838 max)

1

GAM with Negative Binomial

Verdict: this is the best model. With an R^2_{adj} at 84% we capture the "M" shape. Deviance explained: 82%

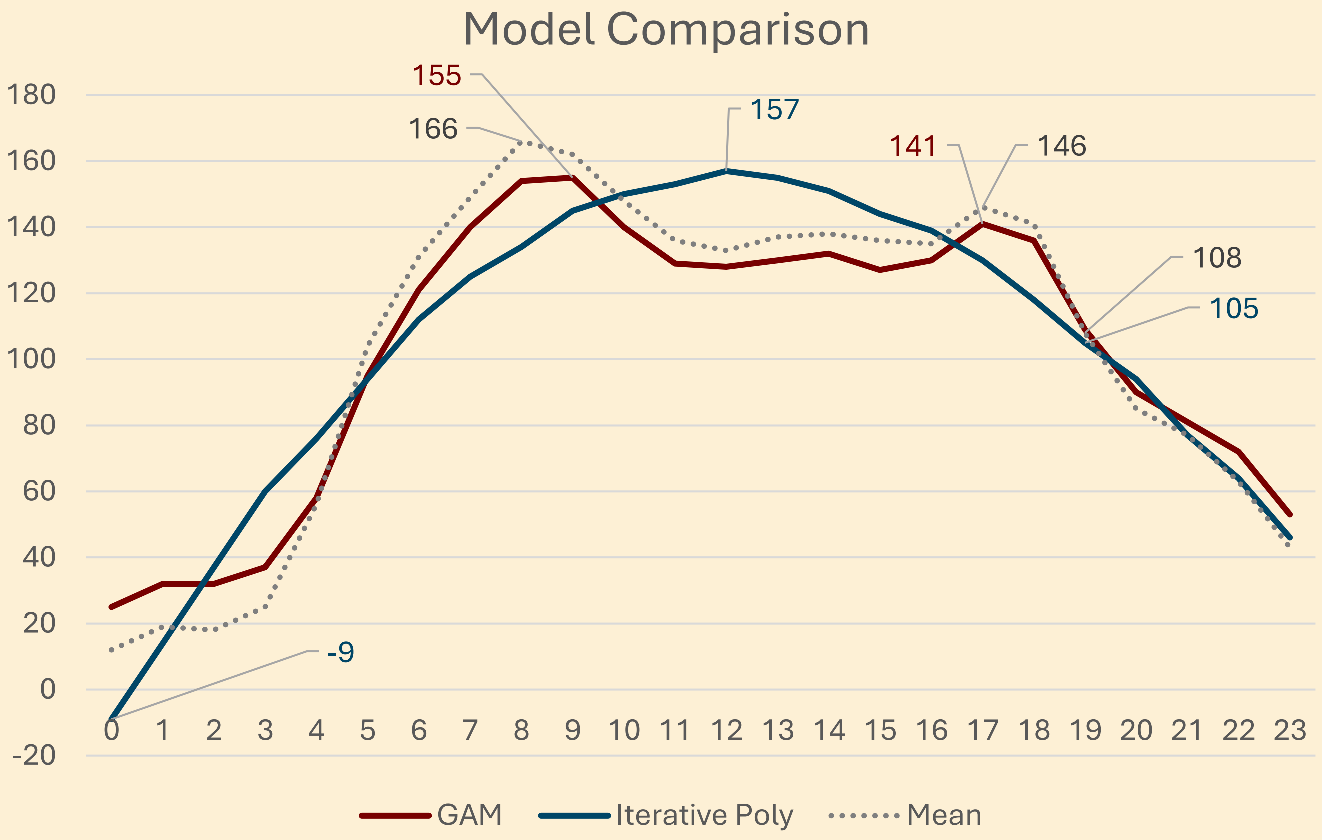
Predicting the traffic (2/3)

Does the **GAM** overfit? Does the **Iterative Poly** R^2_{adj} is a good indicator? What about the **RMSE**?

- **Iterative Poly:** under/over estimate:
  - Not so precise
  - Predicts negative value
- **GAM:** sticks to the mean:
  - With k=20 is really accurate
  - NB family perform the best

About the **RMSE**:

Iterative Poly	61.68
GAM	46.17



Predicting the traffic (3/3)

Is possible **predict the traffic**? **Yes,** traffic is given by several factors:

Working routine

Primary cause
(both offices + logistic)

Day of week

Less traffic in the weekend

Season and holidays

Peaks in August and in the coldest seasons

02

Environment

Goal: Understand what lead to more pollution and the effectiveness of Area C policies over time

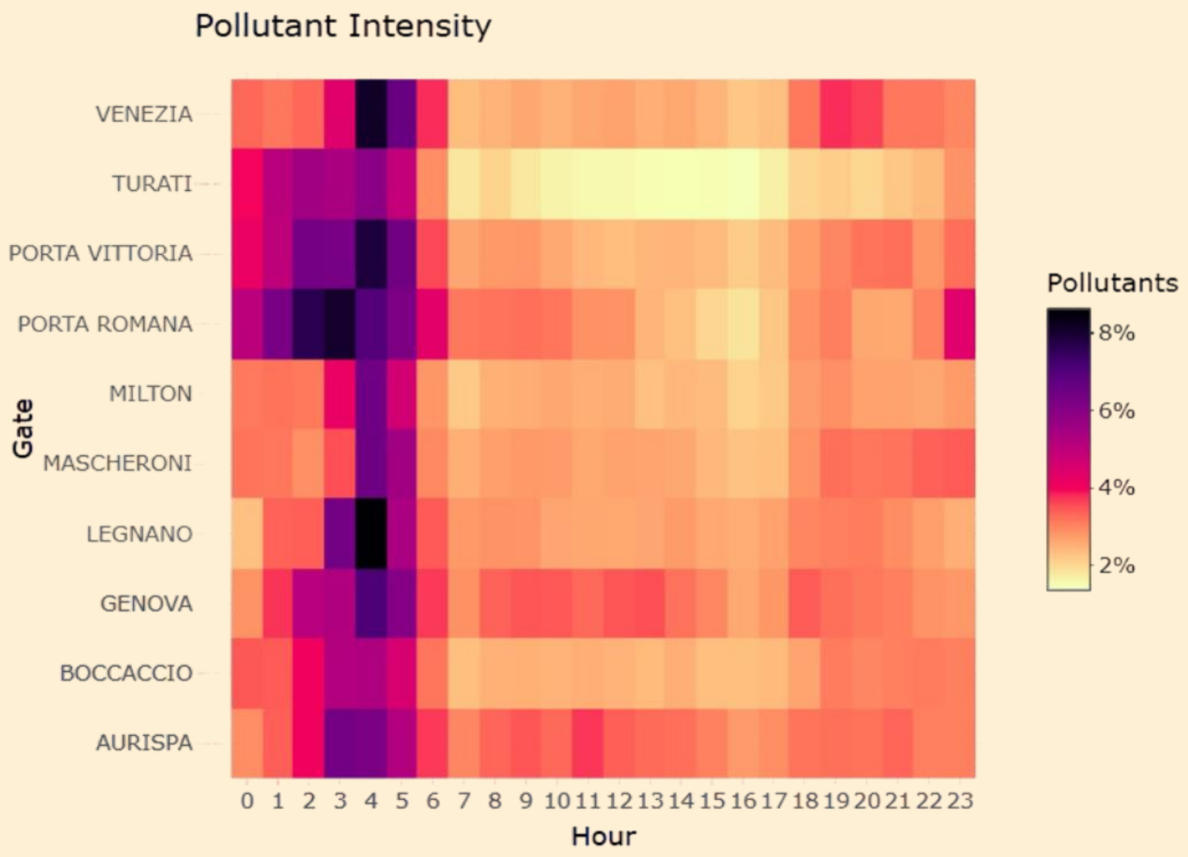
Most polluted context (1/2)

Where and **when** the most pollutant vehicle access Area C? **Who** they are?

Those are the **top 10 polluted** gates.

The **morning concentration** suggests the pollution is given by **logistic activities**:

- Mostly **east side**
- No concentration in the afternoon
- **Trucks** avoid fees
- Trucks are **more pollutant**



Most polluted context (2/2)

Let’s check the hypothesis using **LDA** considering day of week, hour, resident (y/n), fuel type and vehicle.

Looking at the most influent predictors, being identified as pollutant is given by:

- Using **Diesel** as fuel
- Accesses in the **morning**
- **Goods** as vehicle category

This **confirms** the EDA hypothesis with an accuracy of **80%** but a precision of just **25%.**

Predictor	**True**	**False**
Hour	11.74 (11:45)	12.34 (12:20)
Goods	0.29	0.15
Diesel	0.88	0.21
Resident	0.33	0.16
Weekend	0.30	0.25

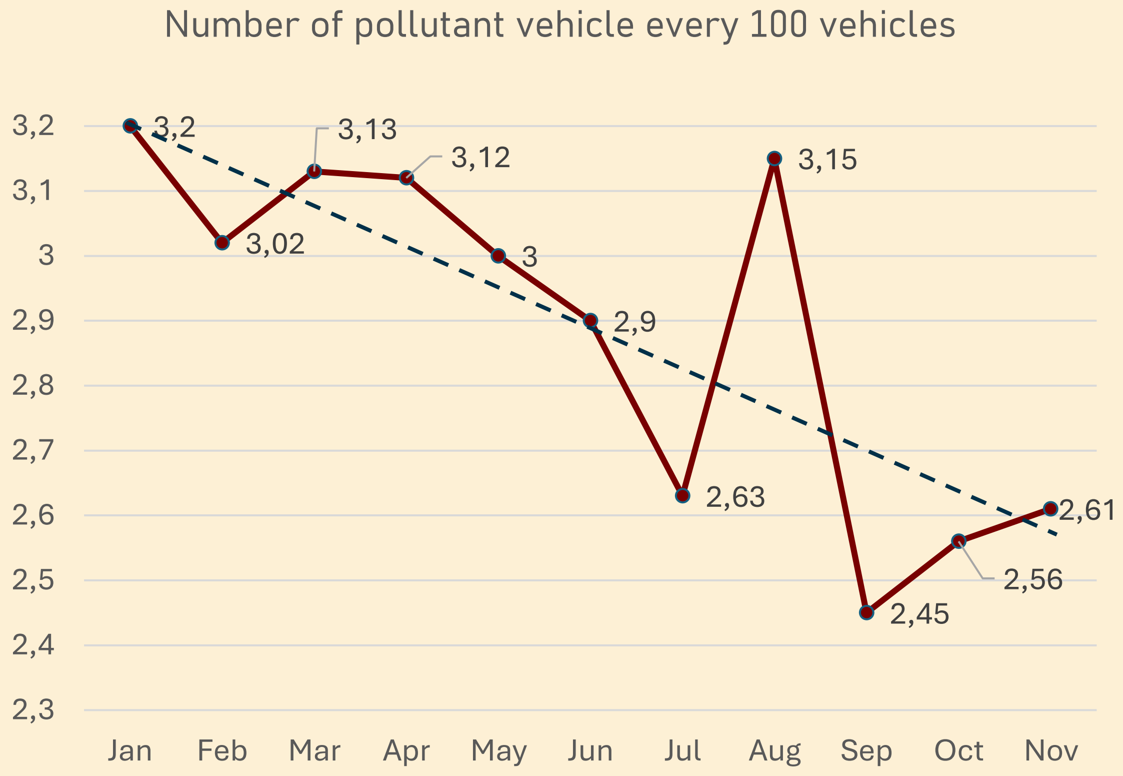
Policies effectiveness

Are the **stricter policies working**? There is a **change** after the new October policy?

The policies **seem to work** but:

- Less total transit during **hotter seasons**
- **Anomalous peak** in August (holidays?)
- **Limited reliability**:
    - R^2_{adj} 58%
    - No comparison with another year

New policy (Oct 24) appears effective.


Number of pollutant vehicle every 100 vehicles

03

Profiling

Goal: Identify the residents, and their preferences, among all the users.

Resident characteristics

What does **distinguish a resident**? There are **multiple predictors** to consider.

Hour
They differs from non-residents?

Weekend
Workers shouldn’t access Area C

Fuel type
They prefer a specific fuel type?

Is pollutant
They benefit from less strict policies

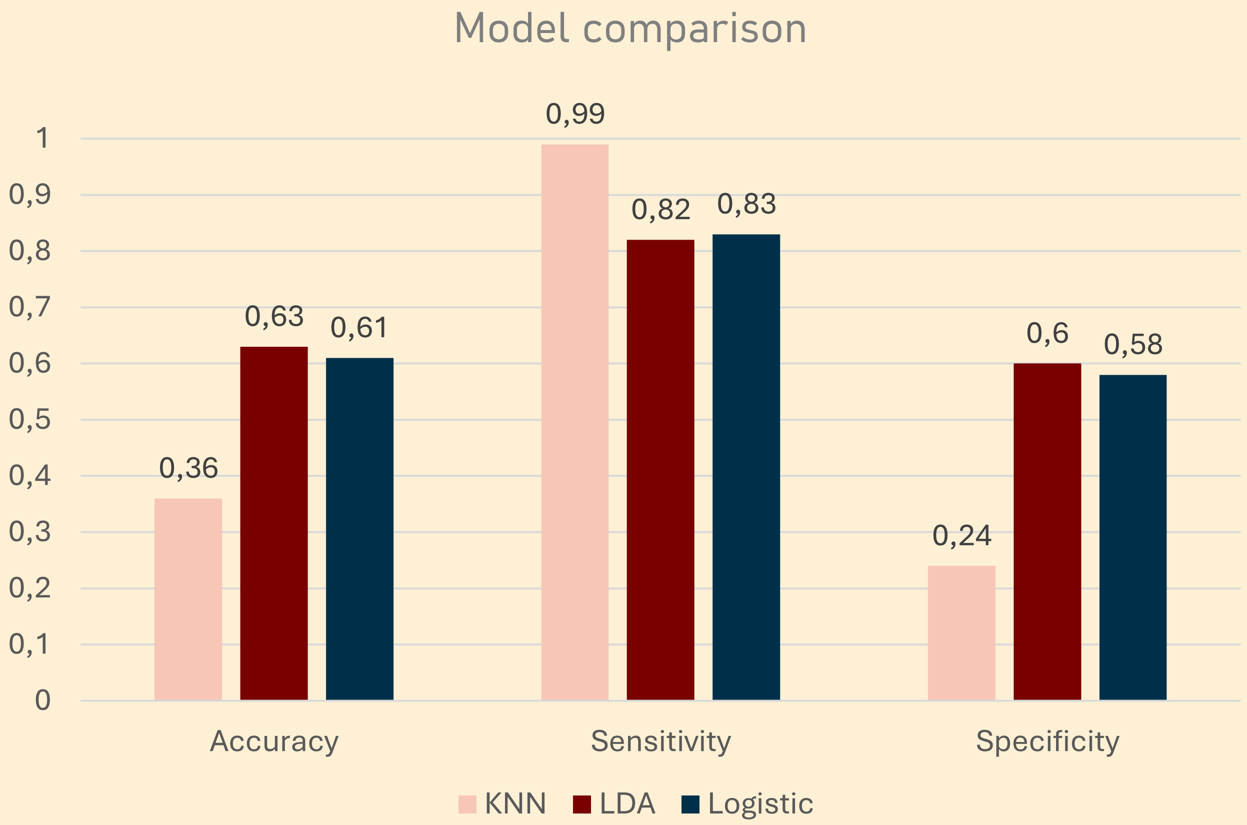
Vehicle type
They shouldn’t drive a truck… right?

How to spot a Resident

This time, let's consider **only the best model** among the three tested.

Looking at the **performance** metrics:

- **LDA and Logistic:**
  - Performed almost the same
  - Give the same results
  - Good metrics
- **KNN:**
  - To few samples
  - Really low specificity



We’ll use Logistic Regression and LDA

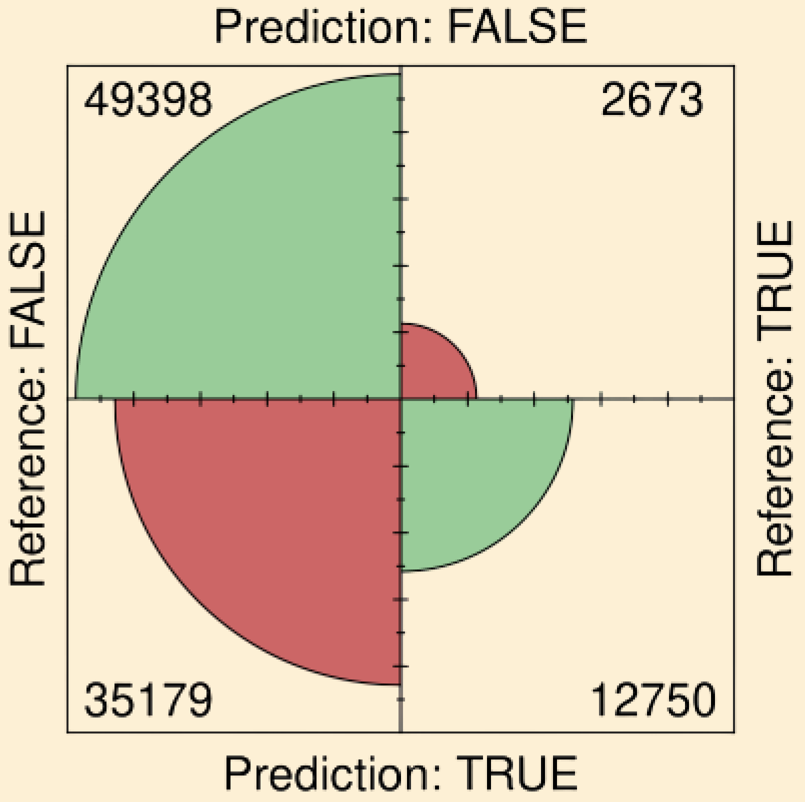
Overall, the models have **performed the same**. However, they **are not perfect**.

With an **accuracy of 62.5%** we can say that a resident

Uses **Petrol** as fuel
Estimate: 1.0336620 | LD1: 1.09501077

Drives a **Car**
Estimate: 1.1296062 | LD1: 1.18140372

Accesses in the **afternoon**
Group mean: 13.93109 (14:00)

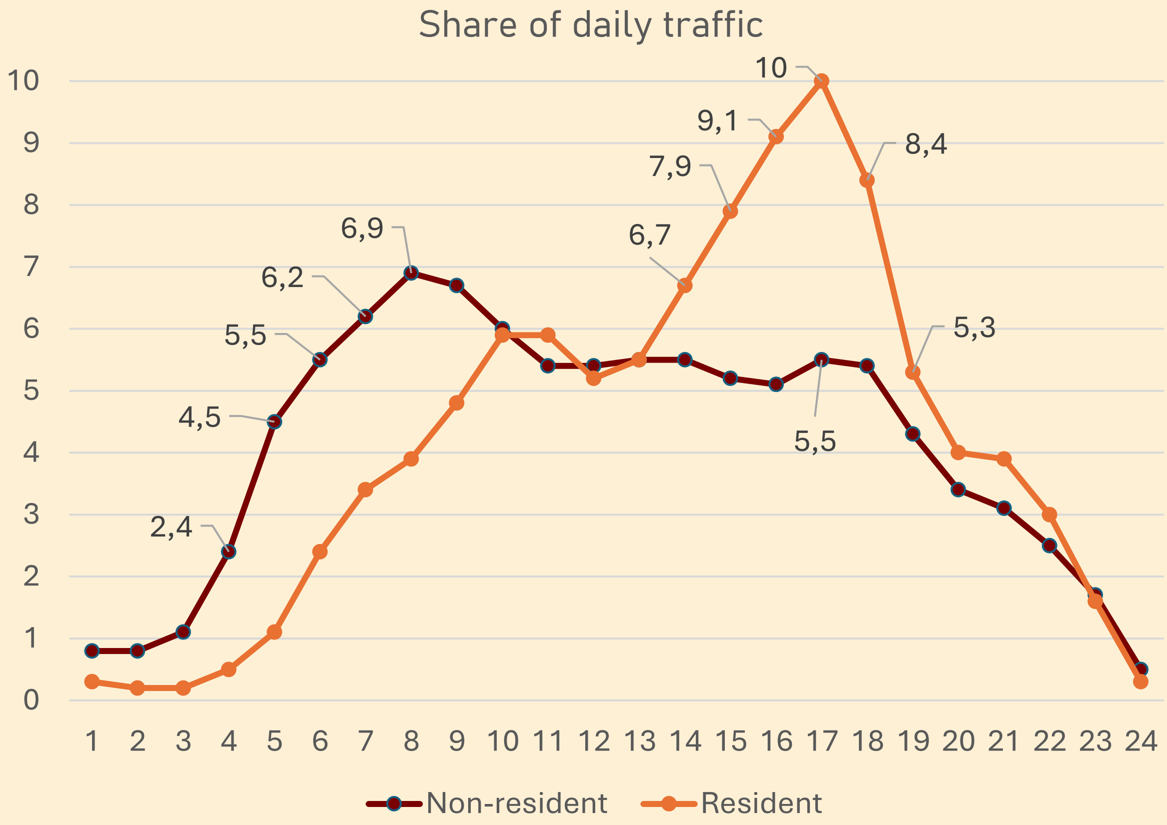


About the time

Comparing the **share of daily traffic** for residents and non-residents we can see the **afternoon peak.**

The chart **confirms** some analysis:

- **Non-resident**:
  - Heavy morning logistic traffic
  - Drop after 07:30 (policy: on)
- **Resident**:
  - Evening return at home
  - Quiet nights

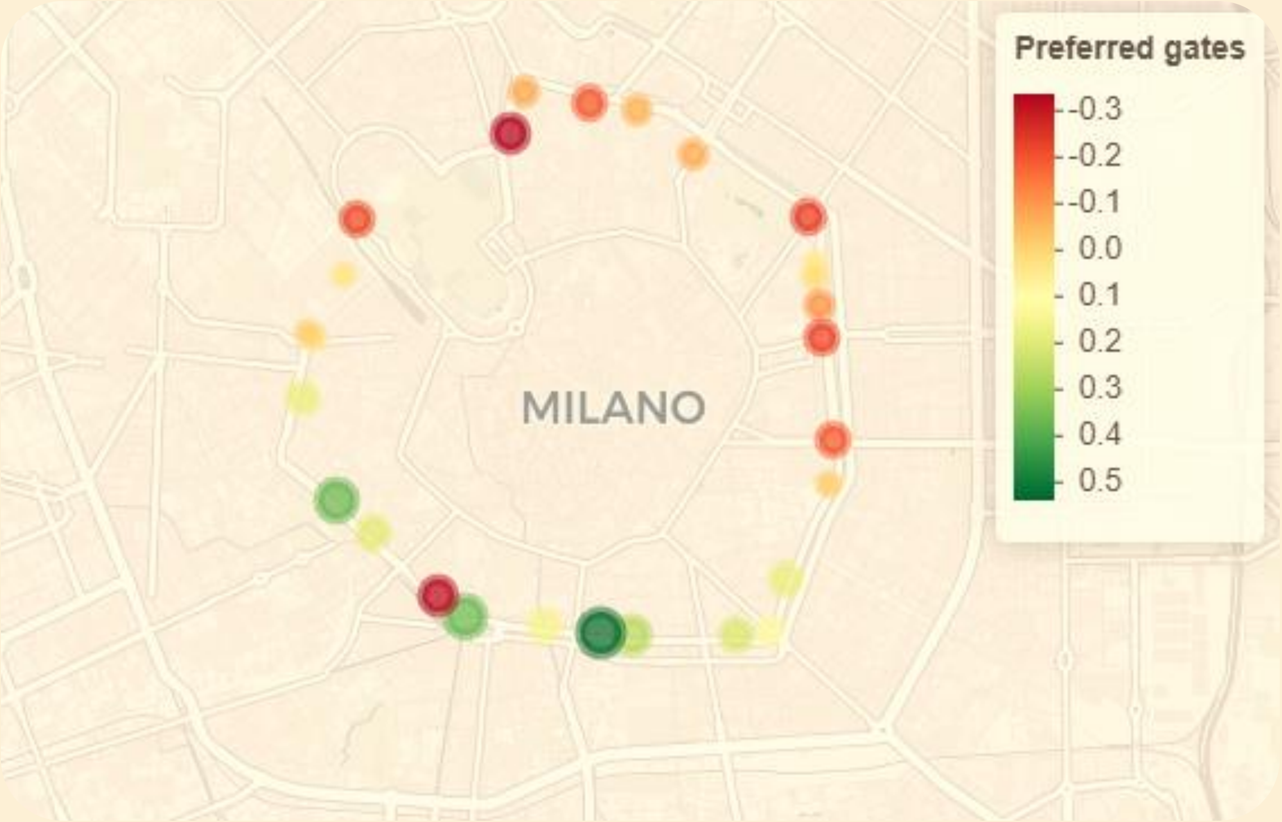


Where they live?

Can we understand **where the residents' houses are**? (without stalking them)

Due to the **low global number of residents**, **it's difficult** because Lasso can obtain only **41% of accuracy**. However:

- **South-west:**
  - Max: *Melegnano* and *Servio Tullio*
  - Historical residential area
- **East:**
  - Min: *Venezia* and *Monforte*
  - Commercial roads



04

Conclusions

Is possible to decipher Milan’s Area C Traffic DNA?

YES

BUT...

Issues and limitations

The dataset does **not consider unpredictable variables** useful for fine tuning.

Weather

Rain and cold may increase traffic

Events

May cause street deviations

Working sites

Construction or renovation sites

Thanks for you attention :)

Any question?