

# 6\_Conclusion

September 3, 2025

## 1 General Conclusion

The objective of this project was to predict passenger ridership on the Geneva Public Transport (TPG) network using open data, complemented by contextual variables such as weather and calendar information.

### 1.1 Two-Level Methodology

The approach was structured into two complementary parts:

- **Global level:** modeling daily ridership aggregated across the entire network, with a comparison of several Machine Learning models.
- **Local level:** application on a specific stop and line, in order to illustrate the operational applicability of the results.

### 1.2 Main Results

The data exploration highlighted clear temporal patterns:

- a strong weekly effect (significant difference between weekdays and weekends),
- the influence of holiday periods and public holidays,
- a strong correlation with past values (lags and rolling means).

Regarding the models:

- **Random Forest** proved to be the most accurate, but at the cost of very high computational requirements.
- **Ridge Regression** delivered nearly equivalent results, with the advantage of being simple, interpretable, and resource-efficient. This is why it was chosen for the local case study.
- **Decision Trees** and **k-NN** provided correct but slightly lower performance.
- **Classic Neural Networks (MLP)** were not well suited to this problem and showed poor results.

The feature importance analysis confirmed that temporal factors (day of the week, weekends, holidays) and historical ridership values (lags, rolling means) are key drivers. Weather, by contrast, only had a marginal secondary effect.

### 1.3 Contributions and Perspectives

This project demonstrates that predicting daily TPG ridership with satisfactory accuracy is feasible using relatively classical Machine Learning models. The proposed methodology is reproducible, robust, and adaptable to other transport networks.

Several avenues for improvement are possible:

- integrating **real-time data** (disruptions, exceptional events),

- refining the local analysis, for example by working with **hourly segments** or focusing on strategic stops such as train stations.

In conclusion, this study confirms the importance of temporal factors in urban mobility and demonstrates the relevance of Machine Learning approaches to anticipate public transport demand.

## **1.4 Personal Contribution**

This project also gave me the opportunity to experiment with different models, face concrete computational constraints, and confirm the importance of a pragmatic approach to modeling. I also realized that there are always further opportunities for improvement or optimization, but at some point it is necessary to conclude the work.