Mélisande Rappo
melisande.rappo@gmail.com

Data Science Project

Projections of the School-Age Population in Lausanne

Conceptual Design Report

5th of October 2025

Abstract

Understanding how the number of students will change in the city of Lausanne is important for planning schools. The School Department needs information about future school populations to organize classrooms, teachers, and school facilities. In this project, we try to predict the number of children in primary and secondary schools using data science methods. We use historical communal data, including births, population changes, migrations, and past school enrollments. First, the data are cleaned, checked, and combined to make one complete dataset. Then, we test different models, including linear regression and some simple machine learning methods, to estimate the number of students in each school level. We also study which variables have the most influence on school population trends. The models are evaluated using metrics like errors and explained variance, and results are shown in tables and charts. This report describes the data, methods, and analysis process. It also gives ideas for future work, for example, testing different scenarios if migration or birth rates change.

Table of Contents

1 Project Objectives

The main goal of this project is to estimate how many students will attend primary and secondary schools in Lausanne in the next years. This information is very important for the School Department to plan classrooms, assign teachers, and organize school buildings properly. We want to predict the number of students for each school cycle and for each year, using the past trends and current demographic data. We will use data about births, population movements, school enrollments, and urban development projects. First, we will clean and combine all this data to make one complete dataset. Then, we will test different models, like linear regression and simple machine learning methods, to forecast the student numbers. The project will produce numbers for each school level and also visualizations, such as plots showing the trends over time. We will also check which factors have the strongest effect on school population, for example migration, birth rates, or new housing projects. At the end, the goal is to give the School Department clear and usable forecasts by school district, with charts and tables, to help them plan for teachers, classrooms, and school facilities in the future.

2 Methods

This project aims to forecast the number of students in Lausanne schools using data science methods. Python will be the main programming language, with pandas and NumPy for data handling, and Matplotlib and Seaborn for visualizations. The workflow will run on Google Colab and local Python installations with Anaconda, allowing flexibility and easy testing of models.The workflow begins with cleaning and combining datasets, including school enrollments, childcare availability, population statistics, births, migration, housing stock, and urban development projects. Additional features, such as family structure, trends in public versus private school attendance could be included along the way. Derived features like students per housing unit and cohort-based projections will also be calculated. For predictions, we will use linear regression and machine learning models such as Random Forest and Gradient Boosting. Models will be evaluated using MAE, RMSE, and R^2. Scenario testing will adjust key variables (housing delivery dates, net migration, and childcare availability) to explore different outcomes and assess how sensitive forecasts are to changes. We'll also look at the previous study realized by Statistique Vaud 10 years ago on the same topic.[1] Finally, results will be presented in tables and plots to illustrate trends and support decision-making. Methods may be adjusted as the project progresses, but the goal is to produce clear and actionable forecasts for the School Services Department.

[1] Statistique Vaud, « Projections scolaires à Lausanne », internal report, 2015.

3 Data

For this project, we will use several datasets. The main data sources are:

- **Historical school enrollments** from the city of Lausanne, including the number of students by school cycle and year.
- **Public childcare places** from the City of Lausanne including number of places funded for each age group (preschool care and after-school care) by school district and number of occupied places.
- **Population statistics and demographic data** from the city of Lausanne, including birth records, migration, and population by age and school grade, household structure.
- **Housing stock and urban development projects** data from the city of Lausanne, including information about new housing.

3.1 Historical school enrollments

The enrolment data will be delivered by the school department of the city of Lausanne[2]. They will include number of students by grades and schools. We only consider compulsory education, as post-compulsory education is managed by the canton rather than at the communal level. There are 8 primary school buildings (grades 1-6P) and 7 mixed buildings (grades 7-11PS) in Lausanne. In 2024, almost 750 students attended private schools. To improve our school enrollment projections, we will need to examine the number of students in private schools and how this trend evolves over time.

3.2 Public childcare availability

This dataset will be delivered by the Lausanne Family Office[3] and contains the number of funded places for each age group and school district, which can influence families's decisions on whether to stay in Lausanne or move elsewhere. Including these variables will help the model account for how the availability of childcare and after-school care affects the distribution of children across schools in the city.

3.3 Population statistics and demographic data

The variable AgeScol is calculated using the child's date of birth and the school starting rules in Lausanne. Children must be 4 years old by July 31 to enter school in August, so we subtract the birth date from the reference school year date to determine their age at the beginning of the school year. We then assign the child to the correct school cycle, from 1P to 11S, creating a

[2] City of Lausanne, Service des Ecoles, School enrolment data, internal dataset, 2024.
[3] City of Lausanne, Lausanne Family Office, Public childcare/after-school care data, internal dataset, 2024.

consistent variable that reflects each child's school age and allows us to project future enrollment numbers accurately.

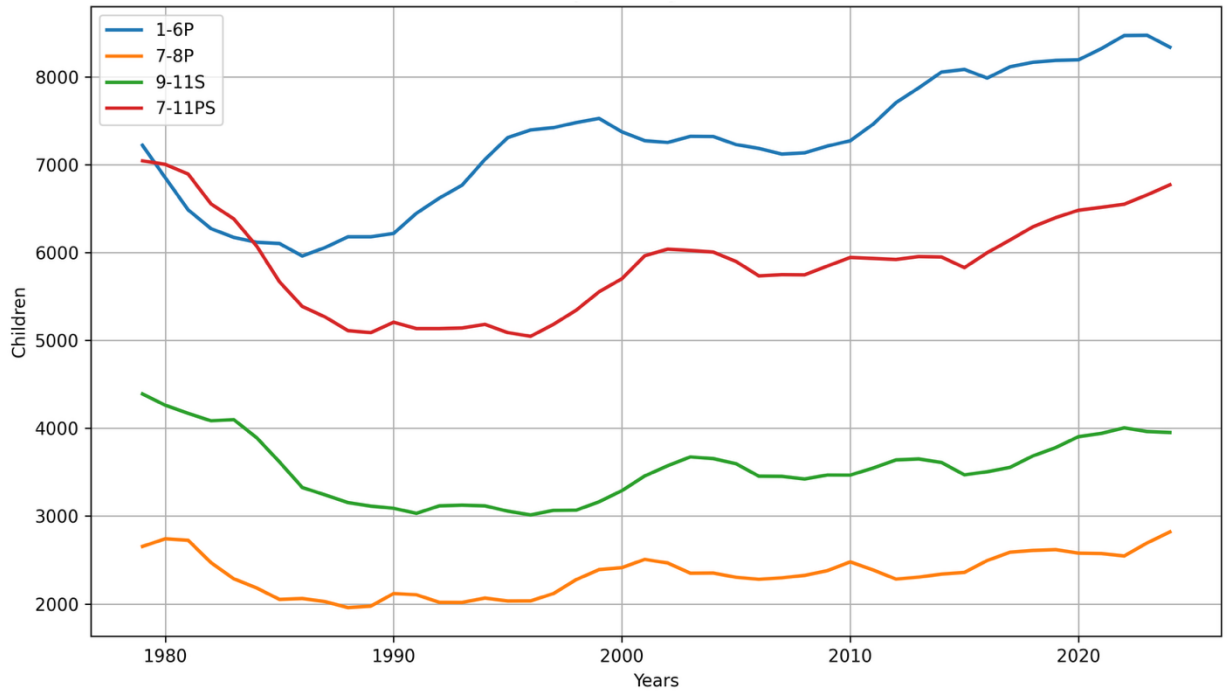
Table 1 : School Age Groups and Corresponding Grades

AgeScol	Grades
4-9	1-6P
10-11	7-8P
12-14	9-11S
10-14	7-11PS

The degree segmentation in Table 1 is related to the spatial organization of school buildings in Lausanne. Primary school students in grades 1–6P attend the same building for six years. Primary school students in grades 7–8P share the same building with secondary school students (9-11S). The 7–11PS group is created to include all students located in those mixed school buildings.

After assigning a school grade to each child aged 4 to 14 (AgeScol), we determine the school building they will attend based on their home address. Normally, student placement is determined by the family's residence, but exceptions can be made to allow attendance in a different school zone.

Figure 1 : Number of children in Lausanne by school grade (based on AgeScol) since 1979

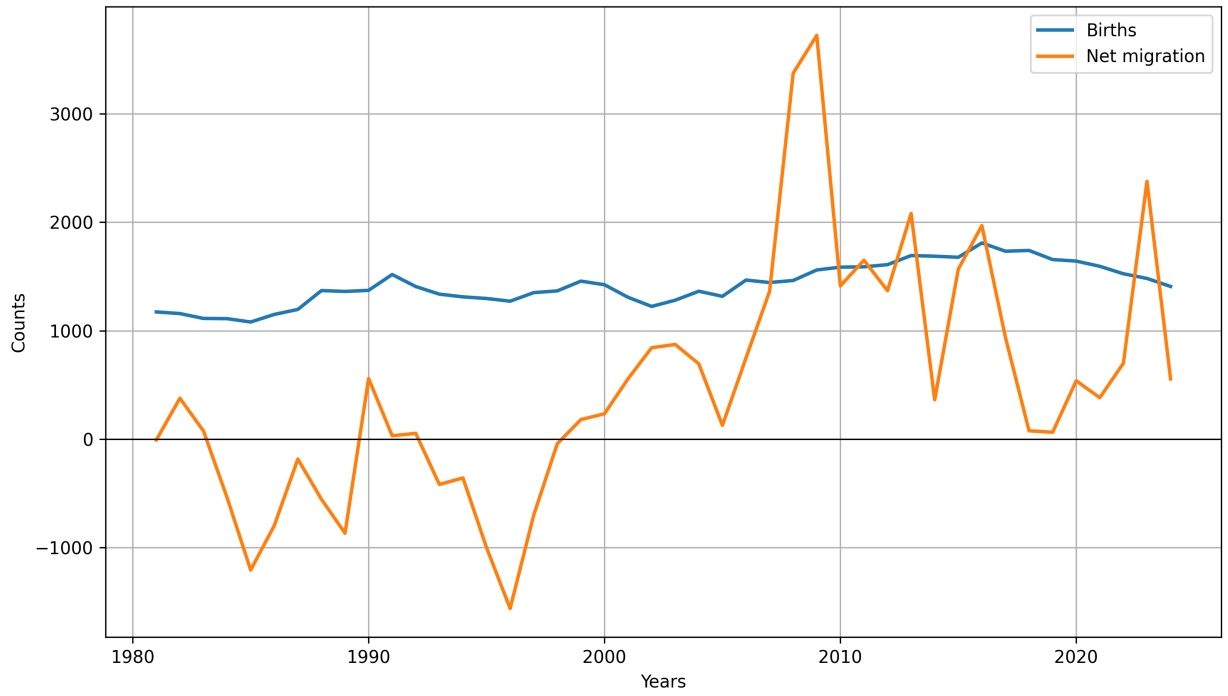


Source: Residents Registration Office, City of Lausanne

In Figure 1, we can see the evolution of children living in Lausanne and their corresponding school grades. Another important factor to consider is the migration of this population. For this purpose, our data are still being consolidated. We could then look at net migration, calculated as (Arrivals + Births) – (Departures + Deaths), which will help us understand trends in residential mobility among families in Lausanne. For now, we can look at the overall migration and births in Lausanne (see Figure 2). But previous intern analyses have shown that Lausanne is an attractive city for young professionals, workers, and students. People over 30 tend to leave the city, [4] possibly to raise a family in the suburbs. Further investigation of net migration by age is still needed.

We will also use the population projections produced by Statistique Vaud[5], including the usual three demographic scenarios (low, medium, and high), to better inform and guide our model.

Figure 2 : Net migration and number of births since 1981 in Lausanne



Source: Residents Registration Office, City of Lausanne

3.4 Housing stock and Urban development projects

The data on the evolution of the housing stock come from the Federal Statistical Office (Buildings and Housing Statistics, StatBL)[6]. Figure 3 shows how the number of housing units and the

[4] City of Lausanne, Office d'appui économique et statistique, Portrait statistique de Lausanne, 2025.
[5] Statistique Vaud, Perspectives 2021–2040: résultats régionaux.
[6] Office fédéral de la statistique, Statistique des bâtiments et des logements (StatBL).

population in Lausanne have grown over time, along with the average number of people per housing unit. We can see that even though the city is getting more populated and adding more housing, households are getting slightly smaller. This matters for our project because smaller households may mean fewer children per housing unit, which can influence future school enrollments. Figure 4 breaks down housing units by the number of rooms. The availability of family-sized housing affects where children live and therefore which schools they will attend. Overall, these trends in housing and population help us understand one of the main drivers behind future student numbers in Lausanne and should be included in our forecasting models. For this study, the analysis should be conducted at the school district level, allowing us to capture differences in student populations and housing within each district.

Figure 3 : Housing units and population in Lausanne (left axis) and people per housing unit (right axis) since 1981

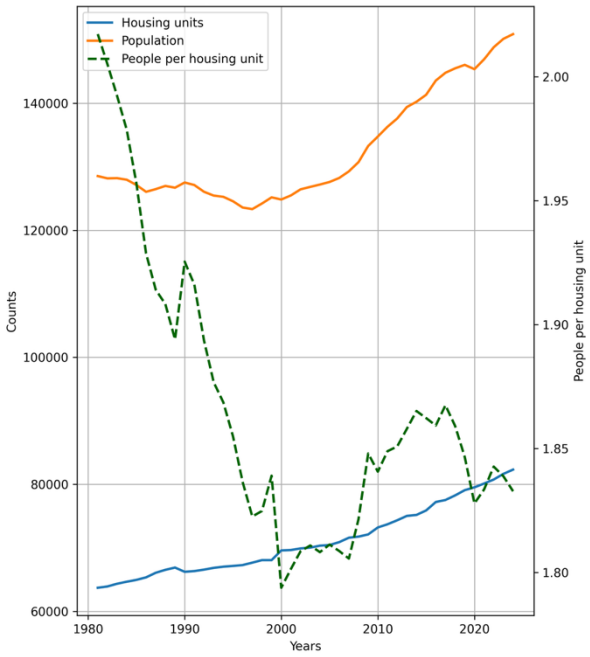
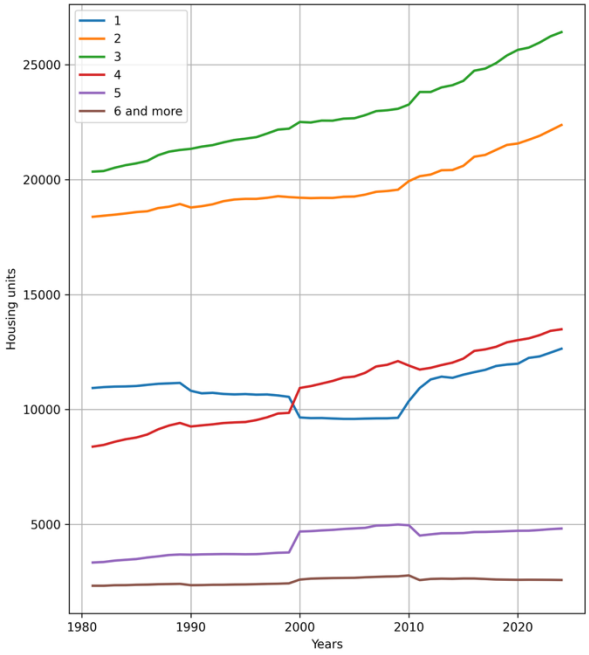


Figure 4 : Housing units in Lausanne by number of rooms since 1971



Source : Federal statistical office, Statistique des bâtiments et des logements StatBL

Regarding urban development projects, the Urban Planning Department prepares projections of the number of housing units in collaboration with the various property management companies and the Building Permit Office. They provide forecasts for each development project, including the estimated completion date of the construction. These projections tend to fluctuate more than other housing data due to uncertainties in construction schedules and approvals.

4 Metadata

We are using the datasets described in Chapter 3, which include historical school enrollments, births, migrations, and information on urban development projects in Lausanne. These data cover multiple years, typically from 1979 to the most recent available year, and are stored on a secure server at the City of Lausanne because they contain sensitive information.

For each dataset, we maintain a catalogue of characteristics that labels every column, describes its type (e.g., integer, string, date), and lists possible values. This helps us understand exactly what each variable represents and ensures consistent use across the project.

The data are initially accessed through RStudio, where we explore the datasets, clean them, and extract the subsets that we need for our analysis. Once the relevant data are prepared, they are exported and used locally in Python, mainly within Conda environments, to perform projections and modeling. This workflow allows us to combine the flexibility of R for data exploration with the power of Python for modeling.

5 Data Quality

The quality of the data largely depends on how accurately it was entered. For data concerning residents and schools, the datasets are raw and come directly from the relevant services, without any cleaning. It is our responsibility to process and standardize them. Changes in software or administrative frameworks can sometimes cause inconsistencies in the data, so we need to acquire sufficient domain knowledge to smooth these irregularities. We are continuously improving our understanding of the data and maintain close contact with the IT department to enhance the quality of the extractions we receive. I do not expect these factors to have a significant impact on the project.

For data on existing housing, the information is consolidated by the Statistical Federal Office. Regarding urban development projects, the delivery dates of new buildings can vary considerably, especially in a city affected by multiple public transport projects (metro, tram, train station) and by the number of legal objections to ongoing projects. I expect these construction delays to have a more noticeable impact. Therefore, it will be necessary to include delay scenarios when developing the model.

6 Data Flow

Data Acquisition

- School enrollments
- Childcare
- Population statistics
- Population projections
- Housing stock
- Urban development projects

Data Exploration & Cleaning

- Access the raw data using RStudio.
- Check for missing values, inconsistencies, and anomalies.
- Extract relevant subsets (school-age children, school districts segmentation).
- Standardize formats (dates, numeric types, school cycles).

Feature engineering

- Create AgeScol variable based on date of birth and school entry rules.
- Aggregate housing data by district and compute derived variables (e.g., persons per housing unit).
- Integrate school, population, and housing datasets into a single analysis-ready dataset.

Modeling & Forecasting (Python / Conda)

- Load cleaned datasets into Python (using Pandas, NumPy).
- Apply statistical models (linear regression, time-series models).
- Apply machine learning models (Random Forest, Gradient Boosting) to capture complex interactions.
- Evaluate models using MAE, RMSE, R^2.
- Test different scenarios: migration changes, housing delays, birth rate, child care fluctuations.

Outputs

- Generate plots of historical and predicted student numbers by cycle and district.
- Create tables summarizing predictions and key metrics.
- Prepare figures linking housing, demographics, and enrollments to highlight drivers of school population changes.

Interpretation & Recommendations

- Show impact of different assumptions
- Validate model outputs with School Services Department.
- Adjust models or data preprocessing based on feedback or new data.

7 Data Model

At the conceptual level, the data model is designed to forecast the number of students in Lausanne schools. The main goal is to provide reliable predictions for each school cycle and district, enabling the School Services Department to plan classrooms, allocate teachers, and prepare school infrastructure effectively.

At the logical level, we plan to use a combination of linear regression and machine learning models such as Random Forest or Gradient Boosting. The features used in the models include historical school enrollments, the AgeScol variable, births, net migration, housing units (including number of rooms), urban development project data (number of units and estimated delivery dates), population projections under low, medium, and high scenarios, and public childcare availability (number of funded preschool and after-school care places, occupancy per district). Derived features, such as students per housing and cohort-based projections by school cycle, are also included to better capture the effect of housing and childcare availability on student numbers.

Table 2 : Logical Data Model: Features and Descriptions

Entity	Columns / Features	Description
Students	Age, AgeScol, school cycle, sssigned school building	Used to assign students to grades and predict future enrollments
Housing units	Year built, number of rooms, district, persons per housing unit	Used to estimate family size and potential number of children
Population	Year, births, net migration, age groups, district, household structure	Captures demographic trends and cohort sizes
Urban development projects	Project name, district, units, estimated delivery year	Captures future housing availability and uncertainty
Population projections	Scenario (low/medium/high), year, age group	Used to test scenarios in the forecasting model
Childcare	Number of funded preschool care places, number of funded after-school care places, occupancy per district	Captures availability of childcare, which may influence whether families stay in Lausanne and children's school attendance patterns

At the physical level, the models do not require specialized hardware and can be run on a standard computer. All raw datasets are stored securely on the City of Lausanne server with restricted access. The data are first explored and cleaned in RStudio, and relevant subsets are then exported to Python (using Conda environments or Google Colab) for modeling, scenario analysis, and visualization. The final outputs include forecast tables, scenario analyses, and plots, which will be used to support planning decisions by the School Department.

8 Documentation

All scripts will be thoroughly commented to ensure they can be easily understood and replicated for similar projects. I will structure the code so that it is simple to adapt to different datasets or to include new variables in future years. Each section of the script will include clear explanations of the steps, the purpose of the calculations, and any assumptions made, making it easy for others to modify and reuse the workflow.

9 Risks

Several factors could affect the accuracy of the forecasts. Missing or inconsistent data from school enrollments, housing, or population sources, delays in urban development projects, unexpected shifts in migration, birth rates or childcare, or the absence of important variables could lead to suboptimal model performance. Statistical or machine learning models may also fail to capture all complex interactions between housing, demographics, and school enrollments. To reduce these risks, data will be carefully cleaned and validated, scenario analyses will be used, and multiple modeling approaches will be compared. Even so, some uncertainties may remain, which could affect the precision of the results and require extra time for adjustments. Inaccurate forecasts could cause the School Department to make suboptimal decisions about buildings or teacher allocation, potentially leading to overcrowded classrooms, wasted resources, or extra costs. Careful monitoring of the data and models is therefore essential to ensure reliable and useful projections.

10 Conclusions

This project forecasts the number of students in Lausanne schools by combining historical enrollments, demographic data, housing trends, and childcare availability. The model uses features such as AgeScol, births, net migration, housing units, and preschool and after-school care to capture factors influencing student numbers. Both statistical and machine learning methods, including linear regression, Random Forest, and Gradient Boosting, are applied, with performance evaluated using MAE, RMSE, and R^2. Scenario testing explores uncertainties in housing delivery, migration, and childcare, providing insights into how different assumptions may affect enrollments. The results, presented in tables and plots, support planning decisions for classrooms, teacher allocation, and resource management. While uncertainties remain, continuous data updates and scenario analyses help maintain reliable and actionable forecasts. Overall, this project demonstrates how integrating multiple data sources with data science methods can provide practical guidance for school planning in Lausanne.

Acknowledgements

I would like to thank Patrick Florio, head of the economic and statistic office of the City of Lausanne, for helping me gain a deeper understanding of this complex topic.

Statement

„Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls die Arbeit als nicht erfüllt bewertet wird und dass die Universitätsleitung bzw. der Senat zum Entzug des aufgrund dieser Arbeit verliehenen Abschlusses bzw. Titels berechtigt ist. Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbstständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die schriftliche Arbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.“

Date: 05.10.2025 Signature(s): Mélisande Rappo

References and Bibliography

[1] Statistique Vaud, Projections scolaires à Lausanne, internal report, 2015.

[2] City of Lausanne, School Department, School enrolment data, internal dataset, 2024.

[3] City of Lausanne, Lausanne Family Office, Public childcare and after-school care data, internal dataset, 2024.

[4] City of Lausanne, Office d'appui économique et statistique, Portrait statistique de Lausanne, 2025. Available at: https://www.lausanne.ch/officiel/statistique/portrait-statistique.html

[5] Statistique Vaud, Perspectives 2021–2040: résultats régionaux. Available at: https://www.vd.ch/etat-droit-finances/statistique/statistiques-par-domaine/01-population/perspectives-demographiques

[6] Office fédéral de la statistique (OFS), Statistique des bâtiments et des logements (StatBL). Available at: https://www.bfs.admin.ch/bfs/en/home/statistics/buildings-housing.html