Proceedings of the 16th International Conference on Industrial Engineering and Operations Management
Ankara, Türkiye, June 23–25, 2026

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AN16.20260205
Published: June 23, 2026

Data Driven Approach to Achieve Operational Excellence: A Case Study of a City Call Center

Houssam El Ouaid, Hamza Guessab and Loubna Benabbou
Management Science Departement
Université du Québec à Rimouski
Lévis, Canada
houssam.elouaid@uqar.ca, hamza.guessab@uqar.ca, Loubna_Benabbou@uqar.ca

Raul Flores and Cynthia Lavoie
Ville de Lévis
Lévis, Canada
rflores@ville.levis.qc.ca, cynthia.lavoie@ville.levis.qc.ca

Abstract

Artificial intelligence has emerged as a strategic lever for transforming public organizations, particularly in their pursuit of operational excellence. In a context characterized by increasing citizen expectations and growing process complexity, municipalities are required to rethink their management practices through data valorization and intelligent automation. The present study presents a structured reflection on the optimization of the 311 call center of the City of Lévis. We discuss the application of the Digital DMAIC methodology, progressing systematically from process definition through data-driven analysis to enable the selection of the most critical key performance indicator. This KPI is then leveraged to train predictive models that inform and support the solution ultimately implemented.

Keywords
Artificial Intelligence, Predictive Modeling, Call Center Optimization, Digital DMAIC, Machine Learning

1. Introduction
Operational excellence was and still is a philosophy of organizations focused on continuous improvement throughout the entire organization to meet customer expectations and reach an optimal operational state. The main idea of operational excellence is to create lean operations throughout the organization (Le, 2025). Lean is the key concept, which explains why many companies have adopted a quality system known as Lean Six Sigma by integrating Lean manufacturing and Six Sigma (Antony et al., 2019; Liu et al., 2020; Pope et al., 2023; Kumar et al., 2023). Lean Six Sigma (LSS) is an organization-wide initiative aimed at achieving zero defects and reducing variability in manufacturing and process industries (Mittal et al., 2023). It also provides a comprehensive toolkit DMAIC (Define, Measure, Analyze, Improve, and Control), that enables organizations to optimize processes, eliminate defects, reduce costs, and enhance customer satisfaction (Kumar et al., 2023).

As previously mentioned, quality improvement has always been and remains the primary factor for guaranteeing operational excellence within an organization, or more specifically within a process. Montgomery defines quality improvement as "the reduction of variability in processes and products" (Montgomery, 2013). Quality has a cost; therefore, variability is the main factor that can either increase or decrease it. Consequently, statistical approaches can be used to quantify variability and attempt to reduce it, through methods such as Statistical Process Control (Woodcock, 2025), Design of Experiments (Lamidi et al., 2024), and advanced control charts designed to monitor complex quality profiles (Capezza et al., 2024). In the service industry, as in our project, processes often rely on structured data and relatively limited datasets. In contrast, big data analytics encompasses a broader range of data types, including unstructured and semi-structured data from multiple sources (Brown et al., 2024). Big Data is commonly described through five key characteristics: the large volume of data generated, the accuracy of the collected data, velocity, value, reflecting the business value derived from data, and the variety of data sources and formats (Kumar et al., 2021). Unlike statistical methods, machine learning and deep learning algorithms generate hypotheses that are consistent with data, and they have been widely applied across various domains in the literature, including demand planning (Tettey et al., 2025; Li et al., 2025), inventory management (Barros et al., 2026; Singh et al., 2024), transportation and logistics (Deniz and Ozceylan, 2023; Chen and Hu, 2025), and production systems (Schlaepfer et al., 2026; Melnychenko, 2024; Veldkamp and Chung, 2024). A considerable body of recent research has focused on the positive impact of big data on continuous improvement (Brochado et al., 2025; Horani et al., 2025). However, although the link between big data and continuous improvement has been established, its practical implementation remains sporadic and largely unstructured (Horani et al., 2025).

In this paper, we propose the Digital DMAIC, also called DDMAIC, which is a data-driven, DMAIC continuous improvement approach based on data valorization models (data analytics, machine learning) to improve the quality of the call center and enhance the client experience by achieving operational excellence. The study is conducted within the City of Lévis, and more specifically within the 311 service. To achieve this objective, the approach is structured around four main steps: (i) understanding and diagnosing the existing process; (ii) identifying the major KPIs and the variables causing variability in these KPIs; (iii) using machine learning algorithms to predict the selected KPI; and (iv) proposing improvement actions. In the next section, we present the Digital DMAIC approach.

2. Methodology

We propose the Digital DMAIC approach, referred to as DDMAIC, which follows the classical DMAIC framework while being driven by data processed through machine learning algorithms rather than traditional statistical methods. Essentially, DDMAIC retains the five phases of the DMAIC cycle: Define, Measure, Analyze, Improve, and Control, while integrating data valorization principles at each stage. The approach begins with the Define and Measure phases, which focus on collecting, cleaning, and preprocessing large volumes of raw data to obtain a clear and accurate understanding of the process under study and to improve the quality of the data used. The core of the DDMAIC approach lies in the Analyze phase, which relies on machine learning models, particularly deep learning algorithms, that are evaluated and compared to identify the most effective one. This phase incorporates variable selection and feature engineering techniques to enhance the explanatory power and robustness of the model. The predictive results obtained are then leveraged in the Improve phase to reduce process variability. Finally, the Control phase focuses on monitoring the process using an interactive dashboard (Figure 1).

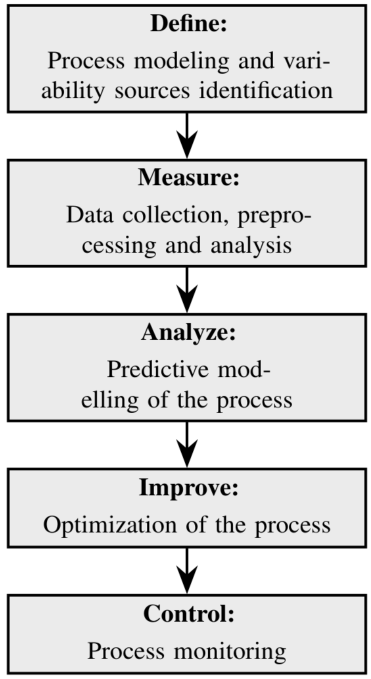


Figure 1. Digital DMAIC approach

The DDMAIC methodology demonstrated its strengths within the 311 call center of the City of Lévis. Three main factors motivated the choice to undertake this project within this institution: (i) the call center is characterized by a high level of instability and is continuously subject to significant variability; (ii) citizen satisfaction and service quality are directly influenced by fluctuations in call volume; and (iii) the skill level and qualifications of City of Lévis employees were particularly conducive to the digitalization of processes.

The DDMAIC project began with a kickoff meeting that brought together all project stakeholders, coordinated by the innovation and performance manager. During this meeting, the overall framework of the project was defined, critical factors that could influence its success or failure were identified, and a roadmap for the upcoming months was established. Subsequently, several digitalization initiatives were implemented to prepare the internal environment and ensure stakeholder involvement.

In the following sections, we present the different stages of Digital DMAIC application within our project through Define, Measure, Analyze, Improve and Control, powered by data processing tools.

3. Case Study: Digital DMAIC for Predictive Performance Management in the 311 Call Center of the City of Lévis

3.1 Define Phase: Process Modeling and Critical-to-Quality Requirements Identification

The 311 call center of the City of Lévis is a multi-channel operational system that handles citizen requests through telephone calls, web forms, and social media platforms. Each of these services presents constraints in terms of demand management, response time, and service quality. As a result, the 311 call center knows a significant variability in the workload. This variability directly affects operational performance and renders the workforce planning and service level management more complex.

To better understand the operating logic of the three operational service channels of the 311 call center, a three-level process modeling approach was adopted. At the first level, a high-level black-box model was developed to represent the overall functioning of the 311 service and its main interactions with the external environment, without entering into operational details. At the second level, this global model was decomposed into three distinct black-box subprocesses corresponding to the three main service channels. This modeling was complemented by the use of SIPOC representations to formalize the core service processes and identify key inputs, outputs, and stakeholders associated with each channel. Based on this process analysis, Critical-To-Quality (CTQ) (Pyzdek and Keller, 2014) elements were defined to translate citizen expectations into measurable performance indicators.

The CTQ framework made it possible to associate each service requirement with one or more Key Performance Indicators (KPIs), ensuring a clear alignment between user expectations, operational objectives, and performance measurement. The resulting set of indicators provides a structured basis for the subsequent measurement and analysis phases, by translating qualitative service requirements into quantifiable variables. At the end of the Define phase, and after a series of meetings with 311 supervisors, five preliminary indicators were retained: average call handling time, call abandonment rate, first-call resolution rate, service level achieved within two minutes, and percentage of emails processed.

These indicators are intentionally defined at a preliminary level and are further evaluated and refined during the Measure phase to retain a limited number of key variables for in-depth analysis.

3.2 Measure: Indicator Selection and Variability Assessment

During the Measure phase, the five indicators initially defined in the previous phase were examined both quantitatively and qualitatively. The objective of this step was to refine the set of KPIs and retain only those most relevant for the subsequent Analyze phase. Among the five KPIs identified previously, four were related to telephone call service, while one was associated with the web forms service. An analysis of historical request volumes revealed a substantially higher frequency of telephone calls compared to web form submissions. Owing to the low occurrence rate of the latter, the corresponding indicator was excluded from further analysis, as its limited volume reduced its statistical relevance. Consequently, the Measure phase focuses primarily on the telephone call channel, which represents the dominant source of workload variability within the system.

Following this channel selection, four indicators remained: average call handling time, call abandonment rate, first-call resolution rate, and service level achieved within two minutes.

To understand the causes behind the variability of these KPIs, a cause-and-effect analysis was conducted using Ishikawa diagrams. A diagram was developed for each indicator in order to structure potential sources of variability in a systematic manner. Based on brainstorming sessions with 311 supervisors, the identified causes were grouped according to the 5M framework: workforce, methods, systems, data and information flows, and operational environment.

This qualitative analysis provided a structured understanding of the main factors influencing the variability of the selected performance indicators. In particular, the cause-and-effect analysis revealed that the factors affecting the service level achieved within two minutes largely overlap with those influencing the other three indicators, namely average call handling time, call abandonment rate, and first-call resolution rate.

As a result, the service level achieved within two minutes did not introduce additional explanatory information beyond that already captured by the remaining indicators. Consequently, this indicator was excluded from further analysis, allowing the analytical scope to be refined to three Key Performance Indicators. The Measure phase therefore enabled a focused transition to the Analyze phase, which concentrates on the quantitative examination of these factors in order to identify the most influential indicator. This indicator subsequently serves as the core variable for the implementation of artificial intelligence–based forecasting and decision-support models.

3.3 Analyze: Selection of the Target Variable and Predictive Modelling of the Process

The objective of this step is to choose the variable that will be at the core of the AI analysis and also the best predictive algorithm for this variable variation. This step contains four substeps: (i) applying machine learning algorithms, (ii) feature selection, (iii) feature engineering, and (iv) predictive model selection.

3.3.1 From Performance Indicators to Analytical Variables

At the end of the Measure phase, three performance indicators were retained for the quantitative analysis phase: average call handling time, call abandonment rate, and first-call resolution rate.

To clearly understand the operational meaning of these indicators, their underlying formulations were examined in collaboration with 311 supervisors. This analysis revealed that the selected KPIs are not independent quantities but rather ratios constructed from several operational variables:

$$\mathrm{KPI}_1: \text{ Average Call Handling Time} = \frac{\text{Total Handling Time}}{\text{Number of Answered Calls}}$$

$$\mathrm{KPI}_2: \text{ Call Abandonment Rate} = \frac{\text{Number of Abandoned Calls}}{\text{Number of Calls Offered}}$$

The third indicator, First-Call Resolution Rate, is derived from several heterogeneous operational variables related to request management and service processing activities.

A closer examination of the indicator formulations shows that KPI1 and KPI2 are derived from a limited and well-defined set of operational variables, namely total handling time, number of answered calls, number of abandoned calls, and number of calls offered. In contrast, the third indicator, first-call resolution rate, is computed from a larger number of heterogeneous variables, some of which are only partially available over the four-year historical dataset used in this study.

Consequently, the analysis strategy was refocused from the KPI level to the variable level. The study therefore retained the two most structurally robust indicators and concentrated the quantitative analysis on their four underlying variables. These variables constitute the fundamental drivers of call center performance and are fully supported by consistent historical data. In the following step, these variables are analyzed quantitatively in order to identify the most informative variable, which is subsequently selected as the core target for predictive modeling.

3.3.2 Data Characterization and Selection of the Target Variable

This subsection aims to characterize the operational data retained for the analytical phase and to justify the selection of the target variable used for predictive modeling. Following the transition from performance indicators to analytical variables, the analysis focuses on the four operational variables underlying the retained KPIs, namely total call handling time, number of answered calls, number of abandoned calls, and number of calls offered.

a. Description of the Analyzed Variables

The dataset used in this phase consists of historical call center data collected over a four-year period, with a regular temporal granularity. The retained variables describe different aspects of call center activity: (i) number of calls offered represents the total incoming demand addressed to the call center; (ii) number of answered calls reflects the effective service capacity and agent availability; (iii) total call handling time captures the overall workload generated by processed calls; and (iv) number of abandoned calls is associated with service congestion and queue dynamics.
These variables were selected because they constitute the fundamental operational drivers of call center performance and are directly involved in the computation of the retained KPIs.

b. Comparative Analysis of Temporal Dynamics

A first exploratory analysis was conducted by comparing the temporal evolution of the main call-related variables.

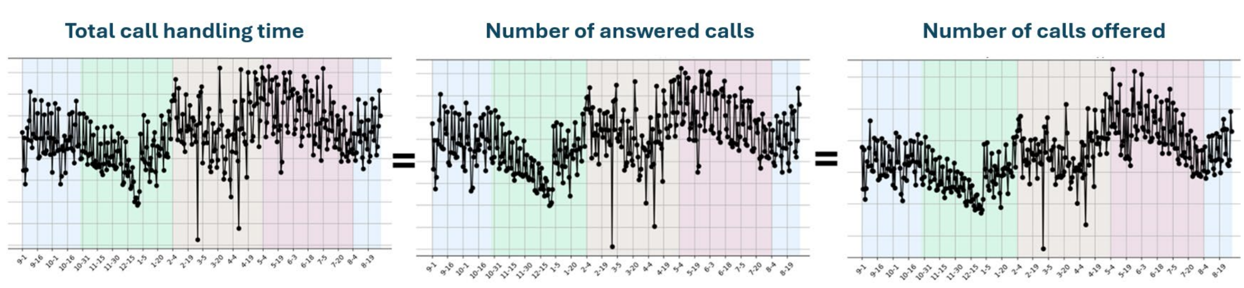


Figure 2. Comparison of the daily trends of total call handling time, number of answered calls, and number of calls offered

Figure 2 illustrates the daily trends of the total call handling time, the number of answered calls, and the number of calls offered.

The results reveal a strong similarity in their temporal patterns. These variables exhibit comparable seasonal fluctuations, with coincident peaks and troughs across time. Periods of increased call volume are systematically associated with a rise in the number of answered calls and an elongation of the total handling time.
This behavior can be explained by the direct operational relationships linking these variables. The number of calls offered reflects the incoming demand, which directly conditions the number of calls that can be answered given agent availability. In turn, the total call handling time increases proportionally with the volume of calls processed and their average duration. The observed coherence between these variables suggests the presence of shared temporal structures that are particularly suitable for predictive modeling.

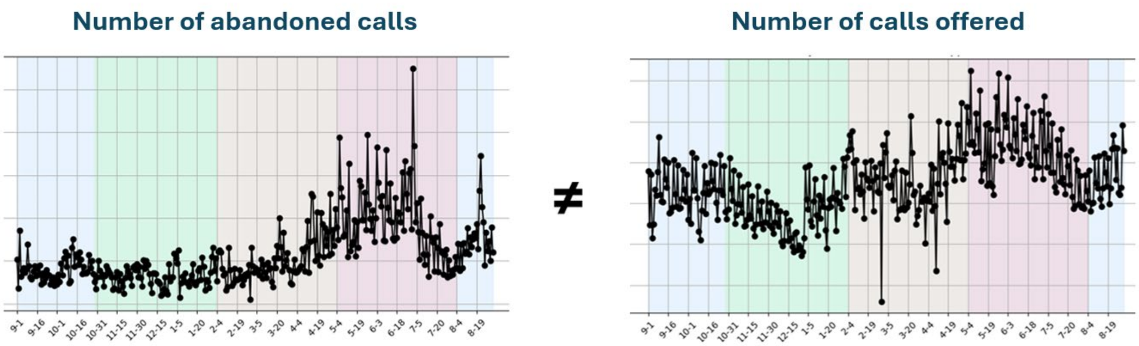


Figure 3. Comparison of the daily trends of number of abandoned calls and number of calls offered

In contrast, Figure 3 compares the temporal evolution of the number of abandoned calls with that of the number of calls offered. Unlike the previous variables, the abandonment series displays a markedly different and more irregular behavior. Although abandonment may increase during high-demand periods, its evolution is less stable and less predictable over time.

c. Variability and Operational Sensitivity Analysis
To further support these observations, a variability analysis was conducted. The number of calls offered, number of answered calls, and total call handling time show relatively smooth dynamics dominated by demand-driven fluctuations. Their variability is primarily explained by external demand patterns and recurrent temporal effects.
By contrast, the number of abandoned calls exhibits higher volatility and irregularity. This variability is influenced not only by call volume but also by several internal and controllable operational factors, such as agent availability and staffing levels, queue management policies, efficiency of internal procedures and scripts, distribution of breaks and work schedules.

As a result, abandonment is more sensitive to short-term organizational decisions and operational constraints, which reduces its suitability as a primary target for long-horizon predictive modeling.

d. Correlation Structure and Analytical Implications
A correlation analysis conducted between the retained variables confirms these findings. Strong correlations are observed between the number of calls offered, the number of answered calls, and the total call handling time, indicating that these variables evolve coherently and are driven by common demand-related dynamics.

Conversely, the correlation between the number of abandoned calls and the other variables is weaker and less stable, reflecting its dependence on additional operational factors beyond call volume alone.

This correlation structure reinforces the analytical relevance of focusing on demand-oriented variables for predictive purposes.

e. Justification for the Selection of the Target Variable

Based on the combined results of the temporal, variability, and correlation analyses, the number of calls offered was selected as the target variable for predictive modeling.

This choice is justified by several key arguments: (i) the number of calls offered represents the primary expression of incoming demand, independently of internal operational constraints; (ii) it exhibits strong temporal regularities and stable seasonal patterns that are well suited for data-driven forecasting; (iii) it acts as an upstream driver of other performance-related variables, including answered calls and total handling time; and (iv) predicting call demand provides direct decision-making value for resource planning, workforce allocation, and capacity management.

Consequently, the analytical strategy adopted in this study focuses on modeling the number of calls offered as the core variable for predictive analysis.

3.3.3 Predictive Modelling of the Process

The objective of this step is to develop a predictive modelling framework capable of accurately forecasting the variability of incoming demand at the 311 call center. Based on the analytical results obtained in the previous phases, this step focuses on predicting the number of calls offered, identified as the most informative target variable. The predictive modelling phase is structured into four substeps: (i) application of machine learning and deep learning algorithms adapted to time-series data, (ii) feature selection based on statistical relevance and temporal dependencies, (iii) feature engineering to enrich the input variables, and (iv) predictive model evaluation and selection.

a. Feature Engineering for Predictive Modelling

To well prepare for the prediction phase, a dedicated feature engineering step was conducted to improve the model's ability to learn from the underlying demand dynamics. This work focused on transforming raw inputs into more informative representations, thereby enhancing the quality of the information provided to the learning algorithms. As highlighted in the literature, feature engineering represents a critical stage in the machine learning pipeline, as it strongly influences model performance by transforming raw data into meaningful features, ultimately improving both predictive accuracy and learning efficiency (Rella, 2025).

a.1. Temporal Variables

Temporal variables were generated to capture the seasonal effects observed in the exploratory analysis. We begin with five binary variables: (i) day of the week (Monday to Friday), representing differences in activity between weekdays; (ii) time of day, a categorical variable indicating the time of the observation, adapted to the granularity used (60 minutes); (iii) month of the year, a numeric variable from 1 to 12 capturing seasonal effects; (iv) day of the year, a numeric variable from 1 to 365 (or 366 for leap years); (v) week of the year, a variable from 1 to 53 indicating the calendar week; (vi) holiday, a binary variable indicating whether the date is a holiday in Quebec; and finally, two binary variables, (vii) day before and (viii) day after a holiday, indicating whether the observation corresponds to the day before or the day after a holiday.

a.2. Meteorological Variables

Meteorological data were obtained from the official website of Environment and Climate Change Canada (Environment and Climate Change Canada, 2025) and cover the period from January 1, 2019, to April 30, 2025, with a daily granularity. After data preprocessing and date standardization, the retained variables are: (i) maximum temperature (°C), reflecting extreme weather conditions that may influence mobility patterns and citizen service requests; (ii) minimum temperature (°C), capturing the effects of periods of intense cold, often associated with incidents related to freezing, heating, or infrastructure; (iii) total precipitation (mm), referring to the sum of rainfall and snowfall, relevant for calls related to road conditions, snow removal, and municipal services; and (iv) snow on the ground (cm), which is an indicator of persistent winter conditions that may increase the demand for municipal service interventions.

a.3. Lagged Variables

To exploit the intrinsic temporal dependence of call time series, we created time-lagged variables such as the number of calls offered on the previous day represented by *Calls offered at* $t-1$, and the number of calls offered two to five days prior to the current observation represented by *Calls offered at* $t-2$ *to* $t-5$.
These variables are essential for time series models, as they provide a history of past values that enables the capture of dependency patterns between consecutive days.

b. Model Evaluation Metrics

To evaluate the performance of the predictive models, two main metrics were retained: the Mean Absolute Error (MAE) and the Weighted Absolute Percentage Error (WAPE). These indicators make it possible to measure the discrepancy between predicted values and actual values, each providing a complementary perspective on model accuracy.

b.1. Mean Absolute Error (MAE) (Willmott and Matsuura, 2005)

The MAE corresponds to the average absolute error between the observed values and the predicted values:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where:

- y_i: actual value at time i,
- $\hat{y}_i$: predicted value at time i,
- n: total number of periods.

The lower the MAE value, the more accurate the model. For example, an MAE of 2 indicates that, on average, the model makes an error of 2 calls per time slot.

b.2. Weighted Absolute Percentage Error (WAPE)

The WAPE represents the average absolute error expressed as a percentage of the actual volume:

$$\mathrm{WAPE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n} y_i}$$

where:

- y_i: actual value at time i,
- $\hat{y}_i$: predicted value at time i,
- n: total number of periods.

The WAPE is often preferred to the MAE because it:

- expresses the average error as a percentage, facilitating interpretation;
- is more robust to large variations in call volumes;
- is less sensitive to extreme values;
- adapts better to highly unbalanced series (low-demand time slots);
- does not depend on the unit of measurement (unlike the MAE).

In addition to its interpretability, the choice of the Weighted Absolute Percentage Error (WAPE) is supported by the literature, particularly in call center forecasting contexts. Previous studies have shown that absolute percentage-based metrics are more appropriate when evaluating models on aggregated call volumes, as they reduce scale-related distortions and better reflect operational impacts. In particular, the WAPE has been shown to be robust to large variations in call volumes and to avoid the scaling issues associated with absolute-error-based criteria, which can bias model evaluation and lead to misleading conclusions in staffing and demand estimation problems (Ding et al., 2015).

c. Presentation of Predictive Models

Several families of predictive models were mobilized to anticipate the variability of the number of calls offered to the 311 call center of the City of Lévis. These models, drawn from both classical machine learning and deep learning, were selected for their ability to handle temporal data and to capture the complexity of the observed phenomena.

c.1. Ensemble Methods

Four ensemble methods were retained: (i) Random Forest (Breiman, 2001) is chosen for its robustness to noise and its ability to reduce variance by aggregating multiple decision trees; (ii) XGBoost (Chen and Guestrin, 2016) is used for its strong predictive performance through sequential learning and built-in regularization; (iii) CatBoost (Prokhorenkova et al., 2018) is well suited to the efficient treatment of categorical variables, reducing the need for extensive encoding; and (iv) AdaBoost (Freund and Schapire, 1997) is evaluated for its capacity to iteratively correct prediction errors by reweighting difficult observations.

These models are particularly relevant in a context of tabular and heterogeneous data, such as call history and meteorological variables.

c.2. Deep Learning Models for Time Series

To better capture complex temporal dependencies, recurrent neural network architectures were also integrated: (i) LSTM (Long Short-Term Memory) (Hochreiter and Schmidhuber, 1997), capable of memorizing long-term temporal dependencies; (ii) GRU (Gated Recurrent Unit) (Cho et al., 2014), lighter than LSTM, with comparable computational efficiency and performance; and (iii) BiLSTM (Bidirectional LSTM) (Schuster and Paliwal, 1997), used to explicitly model both past and future dependencies within a sequence.

These deep learning models were applied using an hourly temporal granularity (1 hour), which was selected as a trade-off between capturing meaningful temporal patterns and ensuring model stability and interpretability. Focusing on a single temporal resolution allows for a consistent comparison between modeling approaches and facilitates the identification of the configuration providing the best predictive performance for the 311 call center.

d. Modeling Results

The predictive models were evaluated using an hourly temporal granularity (1 hour), which was retained for the final analysis in order to ensure methodological consistency and operational feasibility. Model performance was assessed using two complementary error metrics: the Mean Absolute Error (MAE) and the Weighted Absolute Percentage Error (WAPE). Table 1 reports these two metrics for the seven models, ranked from the highest to the lowest error, with the retained model (BiLSTM) highlighted (Table 1).

Table 1. Predictive model performance at hourly granularity.

Model	**MAE (1h)**	**WAPE (1h)**	**Rank**
Ensemble methods (tabular learners)			
AdaBoost	7.24	23.58%	7
CatBoost	6.79	22.11%	6
XGBoost	6.70	21.83%	5
Random Forest	6.66	21.69%	4
Deep learning (recurrent neural networks)			
LSTM	6.44	19.34%	3
GRU	5.96	17.88%	2
BiLSTM	4.42	13.36%	1

To complement the tabular synthesis, Figure 4 compares the models in terms of WAPE and makes the gap between the two families immediately visible: all three recurrent neural networks fall below the best ensemble method, and the retained BiLSTM model reduces the WAPE by about 38% relative to the best ensemble (Random Forest).

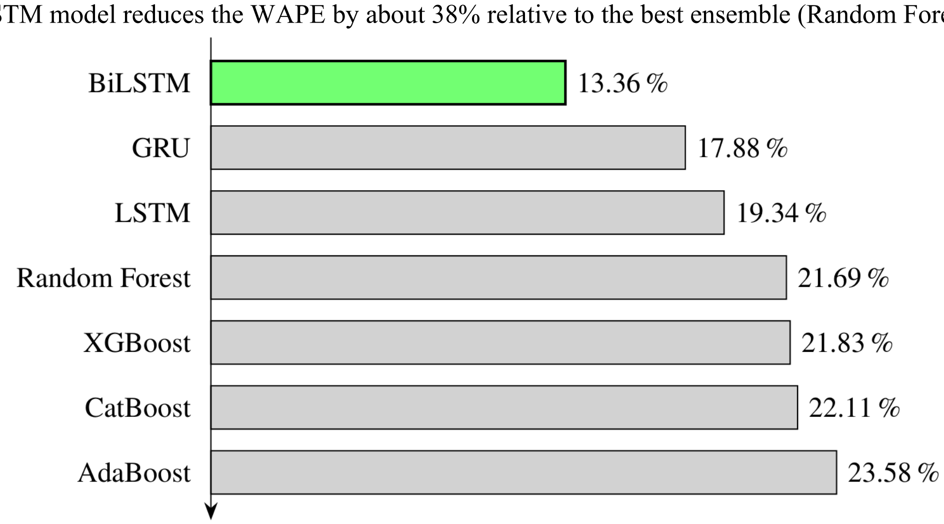


Figure 4. Comparison of the predictive models by WAPE (lower is better).

A first observation is that the two metrics are perfectly consistent: the MAE and the WAPE rank the seven models in exactly the same order, which reinforces the robustness of the comparison and indicates that the conclusions are not an artefact of the chosen error measure.

Two clearly separated performance regimes emerge. The four ensemble methods form a homogeneous group, with WAPE values confined to a narrow band between 21.69% (Random Forest) and 23.58% (AdaBoost), and MAE values between 6.66 and 7.24 calls per hourly slot. Within this group, Random Forest is the most accurate and AdaBoost the least, the latter being penalized by its sequential reweighting scheme, which is known to be sensitive to the noise and outliers that characterize a volatile call-arrival process. The three recurrent neural networks form a second, distinctly more accurate group: LSTM (19.34%), GRU (17.88%), and BiLSTM (13.36%) all outperform the best ensemble method, confirming that the temporal structure of the demand is better captured by sequence models than by purely tabular learners.

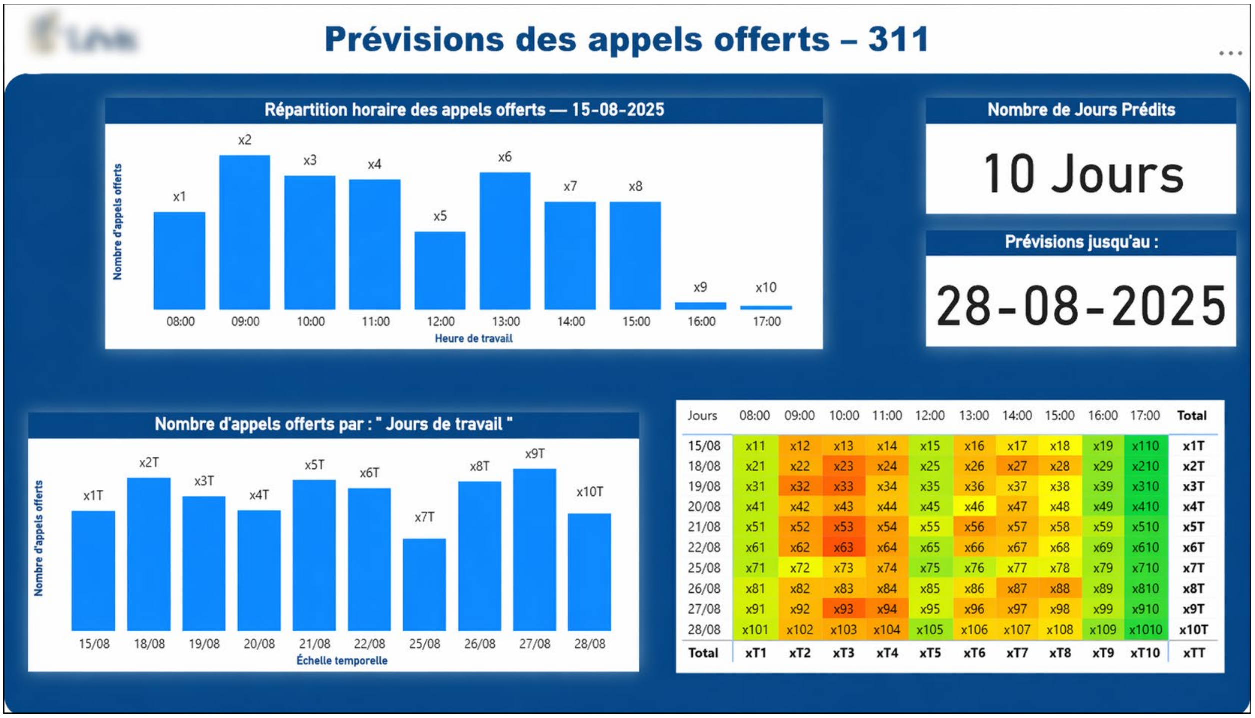
The advantage is most pronounced for the bidirectional architecture. By processing each training sequence in both the forward and backward directions, the BiLSTM exploits the local context surrounding every time step and thereby captures the short-term dependencies of the call-arrival process more effectively than its unidirectional counterparts. Relative to the best ensemble method (Random Forest), the BiLSTM reduces the WAPE from 21.69% to 13.36%, an absolute gain of 8.33 percentage points and a relative improvement of about 38%, while the MAE drops from 6.66 to 4.42 calls, a reduction of roughly 34%. In operational terms, a WAPE of 13.36% means that, on average, the hourly forecast deviates by approximately 13% from the realized call volume, while the associated MAE indicates a typical error of only about four to five calls per hourly slot, a level of accuracy that is directly actionable for short-term staffing and scheduling decisions.

It is important to note that finer temporal granularities were not retained for the final analysis due to computational constraints and operational considerations. Training models at higher resolutions significantly increased data volume and computational complexity, limiting their practical applicability. Consequently, the hourly granularity was selected as an appropriate compromise between predictive accuracy and feasibility.
In this context, the BiLSTM model emerged as the most effective configuration, outperforming all other approaches on both metrics with an MAE of 4.42 calls and a WAPE of 13.36%. This model was therefore retained as the final predictive solution for the 311 call center.

3.4 Improve: Decision-Support Implementation and Performance Monitoring

Following the very successful results of our BiLSTM model, in the Improve phase we aim to materialize these predictive results through an interactive dashboard intended for supervisors of the 311 call center.

The solution consists of a dashboard developed using Power BI, which consolidates call forecasts over a 10-business-day horizon into a single screen. The dashboard is divided into several visual representations (as shown in Figure 5): in the upper-right corner, the forecast horizon expressed in business days; a chart illustrating the intraday distribution of calls by hour; another showing the total daily call volume forecast for each day; and finally, a heat matrix summarizing all this information and highlighting critical periods. This visual hierarchy facilitates immediate and intuitive use for decision-making.



Jours	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	Total
15/08	x11	x12	x13	x14	x15	x16	x17	x18	x19	x110	x1T
18/08	x21	x22	x23	x24	x25	x26	x27	x28	x29	x210	x2T
19/08	x31	x32	x33	x34	x35	x36	x37	x38	x39	x310	x3T
20/08	x41	x42	x43	x44	x45	x46	x47	x48	x49	x410	x4T
21/08	x51	x52	x53	x54	x55	x56	x57	x58	x59	x510	x5T
22/08	x61	x62	x63	x64	x65	x66	x67	x68	x69	x610	x6T
25/08	x71	x72	x73	x74	x75	x76	x77	x78	x79	x710	x7T
26/08	x81	x82	x83	x84	x85	x86	x87	x88	x89	x810	x8T
27/08	x91	x92	x93	x94	x95	x96	x97	x98	x99	x910	x9T
28/08	x101	x102	x103	x104	x105	x106	x107	x108	x109	x1010	x10T
Total	xT1	xT2	xT3	xT4	xT5	xT6	xT7	xT8	xT9	xT10	xTT

Figure 5. Interactive Power BI dashboard: forecasts of offered calls (311)

This tool supports early planning for the allocation of material and human resources by forecasting peak periods in call volume (Haddad et al., 2025). Moreover, the Power BI interface facilitates a managerial transition toward a more data-driven organizational culture (Saidur, 2025), by transforming predictive results into a tool for collective learning and continuous improvement (Makhloufi et al., 2023).

3.5 Control: Performance Monitoring and Statistical Process Control

To conclude our DDMAIC approach, we move to the Control phase, where we aim to ensure that the solutions implemented remain effective over time. During this phase, we verify the accuracy of our machine learning model, which predicts the number of calls offered.

The monitoring plan is based on the daily updating of actual data and their comparison with the predicted values, with the monitoring of error indicators such as the MAE, whose purpose is to detect unusual variations on a weekly basis, and the WAPE, to provide a view over a wider horizon, in addition to the gap between the predicted and actual values, without forgetting to retrain the model regularly to follow changes in the process (Msakni et al., 2023).
We use control charts here, a statistical process control tool for the continuous monitoring of a process to identify the causes of variability and make adjustments (Tran, 2021). We tried to create two types of control charts with control limits of ±2σ and ±3σ: one that analyzes each hour of the day with parameters specific to each hour, and another that groups all hours of the day with parameters specific to the day. The analysis we performed covers the period between August 19 and September 17.

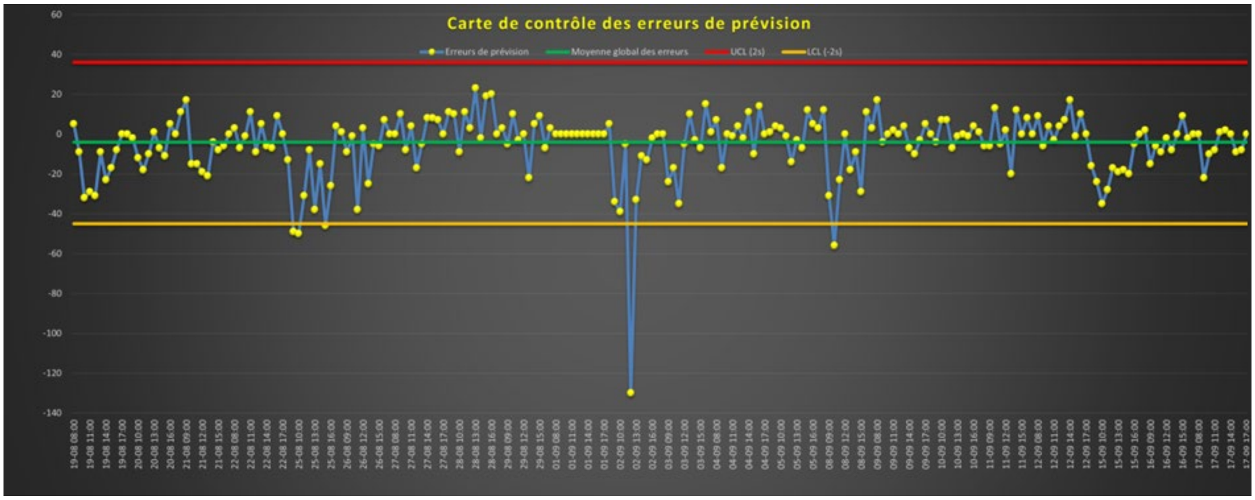


Figure 6. Global forecast error control chart with bounds at ±2σ

The results (Figure 6) showed anomalies in the period from late August to early September. Following meetings with supervisors at the 311 call center, they were able to link the increase in calls during this period to the start of the school year. This explanation not only demonstrates that the detection system is working well but also highlights the importance of contextualizing each anomaly. This will lead us to integrate variables into our model that explain seasonal events, adjust the model's training window, and refine the temporal variables to better capture hourly variability.

4. Conclusion

This paper presents the application of artificial intelligence to improve operational management in the call center of the City of Lévis. The DMAIC methodology enabled the structuring of this operation from problem definition to process control.

The Analyze phase confirmed the complexity of our data, influenced by temporal variables and others related to the environment of the City of Lévis. We were able to use several machine learning models, ranging from ensemble models to deep learning models such as BiLSTM, which provided the best performance.

Achieving excellent results was not sufficient without proposing an interactive dashboard, which can provide an anticipatory view of call volume and enable planners to anticipate, in advance, the material and human resources required to deliver the best possible quality of service.

Beyond these results, several perspectives emerge to further strengthen the link between predictive modelling and operational excellence. A natural next step is to measure the impact of the forecasts on the operational performance of the 311 service, by quantifying their effect on the key performance indicators identified throughout the DDMAIC cycle, in particular the call abandonment rate and the service level achieved within two minutes. This impact should be considered through the lens of resource planning and optimization, by translating the predicted call volumes into proactive staffing and scheduling decisions aimed at reducing citizens' waiting time. A further line of work consists of analyzing the underlying causes of incoming calls in order to anticipate demand at its source and act preventively rather than reactively, thereby closing the loop between prediction, resource optimization, and continuous improvement.

Finally, this work shows how effective use of data can facilitate digital transformation and ensure operational excellence in any context, whether industrial or service-based.

Acknowledgements

This work was supported by the Research Chair in Artificial Intelligence for Supply Chains at the Université du Québec à Rimouski (UQAR). The authors also gratefully acknowledge the support of the City of Lévis.

References

Antony, J., Lizarelli, F. L., Fernandes, M. M., Dempsey, M., Brennan, A., and McFarlane, J., A study into the reasons for process improvement project failures: Results from a pilot survey, *International Journal of Quality and Reliability Management*, vol. 36, no. 10, pp. 1699–1720, 2019.

Barros, J., Silva, N., Gonçalves, J. N. C., Santos, M. Y., and Costa, C., A machine learning-based framework for predicting supply delay risk using big data technology, *International Journal of Data Science and Analytics*, vol. 21, no. 1, p. 71, 2026.

Breiman, L., Random forests, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

Brochado, A. F., Rocha, E. M., and Pimentel, C., A big data–driven system for smart continuous improvement, *Discover Applied Sciences*, vol. 7, p. 1213, 2025.

Brown, W., Wilson, G., and Johnson, O., Understanding the role of big data analytics in enhancing customer experience, *Preprint*, 2024.

Capezza, C., et al., An adaptive multivariate functional EWMA control chart for monitoring complex quality profiles in real manufacturing processes, *Journal Title Unavailable*, 2024.

Chen, N. and Hu, Y., The critical role of digital infrastructure in inequality: New evidence from urban–rural consumption in China, *International Review of Economics and Finance*, vol. 104, p. 104766, 2025.

Chen, T. and Guestrin, C., XGBoost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., Learning phrase representations using RNN encoder–decoder for statistical machine translation, *Proceedings of EMNLP*, pp. 1724–1734, 2014.

Deniz, N. and Ozceylan, E., A bibliometric and social network analysis of data-driven heuristic methods for logistics problems, *Journal of Industrial and Management Optimization*, vol. 19, no. 8, pp. 5671–5689, 2023.

Ding, S., Koole, G., and van der Mei, R. D., On the estimation of the true demand in call centers with redials and reconnects, *European Journal of Operational Research*, vol. 246, no. 2, pp. 646–658, 2015.

Environment and Climate Change Canada, Historical climate data, *Environment and Climate Change Canada Database*, https://climate.weather.gc.ca, 2025.

Freund, Y. and Schapire, R. E., A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

Haddad, T., Jabr, N., and Haddad, A., Simulation-based decision support for call centre staffing optimisation and performance improvement, *DMAME Journal of Industrial Engineering*, 2025.

Hochreiter, S. and Schmidhuber, J., Long short-term memory, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

Horani, O. M., Al-Adwan, A. S., Al-Rahmi, W. M., and Alkhallifah, A., Enabling innovation with big data analytics capabilities: The moderating role of organizational learning culture, *Data Science and Management*, 2025.

Kumar, K. S. A., Hababa, S. M., Worku, B., Tadele, G., Mengistu, Y. G., and Prasad, A. Y., Big data characteristics, classification and challenges: A review, *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 12, pp. 4236–4243, 2021.

Kumar, S., Sarulkar, P., and Yadav, V. K., Operational excellence of the steel industry using the Lean Six Sigma approach: A case study, *International Journal of Quality & Reliability Management*, 2023.

Lamidi, K., et al., A comprehensive overview of design of experiments applications in quality control and assurance, *Journal Title Unavailable*, 2024.

Le, L., Operational excellence through Lean manufacturing and certification standards, *ResearchGate Preprint*, 2025.

Li, J., Yu, Q., Liu, X., Zhang, H., and Zhang, Z., From spatiotemporal charging patterns to localized supply–demand planning for electric vehicle infrastructure, *Transportation Research Part D: Transport and Environment*, vol. 149, p. 105058, 2025.

Liu, G., Agostinho, F., Duan, H., Song, G., Wang, X., Giannetti, B. F., Santagata, R., Casazza, M., and Lega, M., Environmental impacts characterization of packaging waste generated by urban food delivery services: A big-data analysis in the Jing–Jin–Ji region (China), *Waste Management*, vol. 117, pp. 157–169, 2020.

Makhloufi, L., Vasa, L., Rosak-Szyrocka, J., and Djermani, F., Understanding the impact of big data analytics and knowledge management on green innovation practices and organizational performance: The moderating effect of government support, *Sustainability*, vol. 15, no. 11, p. 8456, 2023.

Melnychenko, O., Utilising AI models to analyse the relationship between battlefield developments in the Russian–Ukrainian war and fluctuations in stock market values, *Forum Scientiae Oeconomia*, vol. 12, no. 4, pp. 83–98, 2024.

Mittal, A., Gupta, P., Kumar, V., Al Owad, A., Mahlawat, S., and Singh, S., The performance improvement analysis using Six Sigma DMAIC methodology: A case study on an Indian manufacturing company, *Heliyon*, vol. 9, no. 3, p. e14625, 2023.

Montgomery, D. C., Introduction to Statistical Quality Control (7th ed.), *Wiley*, 2013.

Msakni, M. K., Risan, A., and Schütz, P., Using machine learning prediction models for quality control: A case study from the automotive industry, *Computational Management Science*, vol. 20, p. 14, 2023.

Pope, J., Boyle, J., and Worrall, M., Airway management in paediatric emergencies outside of an intensive care setting: A quality improvement project using Lean/Six Sigma methodology, *Archives of Disease in Childhood – Education and Practice Edition*, pp. 1–4, 2023.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A., CatBoost: Unbiased boosting with categorical features, *Advances in Neural Information Processing Systems*, 2018.

Pyzdek, T. and Keller, P., The Six Sigma Handbook (4th ed.), *McGraw-Hill Education, New York, NY, USA*, 2014.

Rella, B. P. R., The role of feature engineering in machine learning: Techniques, challenges, and automation with data engineering, *Iconic Research and Engineering Journals*, vol. 8, no. 10, pp. 805–823, 2025.

Saidur, M. J. I. Md., AI-enhanced business intelligence dashboards for predictive market strategy in U.S. enterprises, *International Journal of Business and Economics Insights*, pp. 603–648, 2025.

Schlaepfer, B., Langer, J., Erni, S., and Colombo, V., A new approach for the field detection of sleep bruxism based on inertial sensor data and machine learning classification, *Scientific Reports*, vol. 16, no. 1, p. 443, 2026.

Schuster, M. and Paliwal, K. K., Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

Singh, A. K., Simha, J. B., and Agarwal, R., Prediction of intermittent demand occurrence using machine learning, *EAI Endorsed Transactions on Internet of Things*, 2024.

Tettey, A., Pham, H., Chen, H., Wu, D., and Wooley, A., A decision support framework for electric demand planning in distribution systems during extreme minimum temperatures, *Energy Reports*, vol. 14, pp. 1137–1148, 2025.

Tran, K. P., Introduction to control charts and machine learning for anomaly detection in manufacturing, *Machine Learning for Industrial Applications*, Chapter, 2021.

Veldkamp, L. and Chung, C., Data and the aggregate economy, *Journal of Economic Literature*, vol. 62, no. 2, pp. 458–484, 2024.

Willmott, C. J. and Matsuura, K., Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Climate Research*, vol. 30, pp. 79–82, 2005.

Woodcock, A., Re-establishing control limits in statistical process control to improve quality monitoring and variation detection in real-time process data, *Journal Title Unavailable*, 2025.

Biographies

Houssam El Ouaid is currently pursuing a Master's degree in Project Management at the Université du Québec à Rimouski (UQAR). His research interests include data science, operational excellence, machine learning applications in public service management, demand forecasting, and data-driven decision-making.

Hamza Guessab is currently pursuing a Master's degree in Project Management at the Université du Québec à Rimouski (UQAR). His research interests include data science, operational excellence, machine learning applications in public service management, demand forecasting, and data-driven decision-making.

Loubna Benabbou is a Full Professor and holds the Research Chair in Artificial Intelligence for Supply Chains in the Department of Management Sciences at Université du Québec à Rimouski (UQAR). She is an Affiliated Professor at Mila and at the Department of Operations Research and Computer Science at Université de Montréal. She is also a member of IVADO, GERAD, CIRRELT and ID. Her research focuses on developing and applying machine learning and operations research methods to support data-driven decision-making, with a particular emphasis on supply chain decarbonization and resilience. She holds an industrial engineering degree from École Mohammadia d'Ingénieurs, as well as an MBA and a Ph.D. in Machine Learning and Optimization from Université Laval.

Raul Flores Innovation and Performance Advisor at the City of Lévis.

Cynthia Lavoie Assistant Director of the Architecture and Digital Transformation Department at the City of Lévis.