YCBS 299 – 2025 autumn-Data Science Capstone Project

Business Problem Framing Report ---Team #1

Executive Summary

This project addresses the critical challenge of effective workload prediction at the fire station (caserne) level, which can provide important insights for resource allocation, scheduling, and public safety planning. By applying advanced data science methods—ranging from statistical models (ARIMA, Prophet) to machine learning (XGBoost, LSTM) and optimization techniques, this project will develop a data-driven predictive system that forecasts the monthly workload of each caserne.

---

1. Business Problem

The City of Montreal operates one of Canada's largest urban fire services, with 67 around fire stations ("casernes") distributed across the island. Each caserne handles a wide range of emergencies, from fires to medical interventions, rescue operations, hazardous material incidents, and false alarms. The workload faced by each caserne fluctuates significantly from month to month due to a variety of factors such as seasonal weather conditions, major public events, and demographic differences across boroughs, and random incident spikes. This unpredictability makes it challenging for SIM managers to ensure that resources—personnel, vehicles, and equipment—are allocated efficiently while still maintaining readiness for emergencies.

So the business problem that this project seeks to address is therefore: How can the Service de sécurité incendie de Montréal (SIM) anticipate the monthly workload of each caserne to support proactive staffing, scheduling, and resource allocation decisions?

Solving this problem would enable SIM to move from reactive to predictive operations, reducing costs associated with overtime and resource misallocation while improving public safety through faster, more reliable emergency response.

---

2. Business Problem Rationale

Accurately forecasting monthly workload at each Montreal fire stations (caserne) is relevant for not only operational efficiency, fiscal responsibility, but also a critical element of public safety. Fire departments worldwide face the challenge of balancing unpredictable demand with the need to maintain rapid response capabilities. Research has shown that the demand for fire and emergency services exhibits strong temporal patterns (daily, weekly, and seasonal), as well as sensitivity to external factors such as weather, population density, and socio-economic activity.

Spatial and spatio-temporal analyses of fire incidents further justify the fire station focus. Research applying spatial scan statistics, clustering, and geographically aware regression demonstrates that fire/exposure risk varies across neighborhoods and that local spatial context (building use, density, and previous incident history) meaningfully predicts incident occurrence. Using caserne-level, station-service polygons and incident geolocations (both available via Montreal’s open data portal) enables station-specific modeling rather than city-wide aggregates, making forecasts directly actionable for caserne managers.

From an operational perspective, better forecasts reduce two costly risks: under-staffing (which increases response time and endangers lives and property) and over-staffing (which raises overtime and idle-resource costs). Empirical studies in EMS and ambulance systems show that incorporating external predictors — weather, events, holidays, and local population metrics — materially improves predictive accuracy, enabling finer-grained deployment and dynamic reallocation strategies. These results directly map to the caserne prediction use-case: enriching historical incident counts with weather and contextual features is expected to yield more actionable monthly forecasts.

Finally, sound forecasting aligns with smart-city and evidence-based governance goals: it provides objective inputs for scheduling, budget planning, and long-term infrastructure decisions (e.g., whether to add or relocate a caserne). The open availability of incident and caserne data through the Ville de Montréal Open Data Portal provides a unique opportunity to implement such forecasting solutions. Leveraging these resources with advanced predictive modeling will allow the Service de sécurité incendie de Montréal (SIM) to transition from a reactive approach—responding as incidents occur—to a proactive strategy that anticipates demand and prepares accordingly. This not only enhances operational efficiency but also strengthens public trust in the city’s emergency preparedness.

---

3. State-of-the-Art Review

Historically, the workload forecasting for fire and emergency services has been addressed through three main approaches: (1) classical and modern time-series forecasting for station/area aggregated counts, (2) spatial and spatio-temporal hotspot models for high-resolution location-based risk, and (3) machine-learning and hybrid ensembles that combine exogenous predictors and temporal structure.

1) Time-Series & Statistical Forecasting

Classical time-series models such as ARIMA, SARIMA, and Exponential Smoothing (ETS) are widely used for forecasting aggregated station workloads, as they effectively capture trends and seasonality while remaining interpretable and robust. Extensions like dynamic regression and tools such as Prophet allow the inclusion of exogenous factors (e.g., weather, holidays), making them especially relevant for emergency demand forecasting. These models serve as reliable baselines in public-service applications, with Hyndman & Athanasopoulos' *Forecasting: Principles and Practice* offering a standard reference in this domain.

2) Spatio-Temporal Hotspot & Point-Process Methods

For high-resolution forecasting (hourly or daily demand in small areas), methods such as Spatio-Temporal Kernel Density Estimation (stKDE) and point-process models are commonly applied. These approaches generate spatial heatmaps of expected incidents by weighting past events according to their temporal relevance, making them effective at capturing both spatial sparsity and recurring seasonal cycles. Widely adopted in ambulance and EMS forecasting, stKDE and its extensions provide fast, interpretable forecasts that are especially useful for vehicle deployment and station coverage planning (Zhou & Matteson, 2015; Hu et al., 2018).

3) Machine Learning & Ensemble Methods

Recent studies increasingly apply machine learning models—such as Random Forest, XGBoost, and deep neural networks—to predict emergency incident counts using diverse features like weather, holidays, demographics, and land use. These approaches capture nonlinear relationships and often outperform classical time-series models when rich covariates and sufficient data are available. Tree-based ensembles are especially practical for monthly station-level forecasts, while deep learning models (e.g., LSTM, GRU) show promise for high-frequency data but require larger datasets and careful validation to avoid overfitting (Pirklbauer et al., 2019; Ku et al., 2024).

The state of practice often uses hybrid pipelines: a statistical model captures trend/seasonality while an ML model handles residual nonlinear effects and exogenous predictors. Ensembles or stacked models combine probabilistic intervals from time-series models with point estimates from ML to provide both accuracy and uncertainty quantification. Operational implementations emphasize rolling window validation that preserves temporal ordering, business-meaningful error metrics (MAE/RMSE and forecasting accuracy within a tolerance band), and interpretability for operations staff. Case studies that forecast station workloads (e.g., Amsterdam fire forecasting) show ensembles and model stacking that incorporate weather, weekday/holiday adjustments, and station-level heterogeneity perform well in practice.

---

4. Implementation Approach

To implement the solution pragmatically, this project will operate step by step: Data collection & preparation-> Exploratory analysis-> Model development ->Evaluation.

With combination of Alteryx and Python code for data ingestion, cleaning, processing, start from station-level monthly aggregation and baseline seasonal models (like ARIMA/Prophet) to set an interpretable benchmark, apply spatial insights to combine with station service-areas (maybe stKDE), then add exogenous variables (monthly weather summaries, holidays/events, population density) and compare tree-based ML (like XGBoost) to time-series with regressors, finally evaluate model by metrics like RMSE,MAE or workload classification accuracy.

Tableau will be used to visualize interactive dashboard and for stakeholder communication. These dashboards will display historical analysis, demand forecasts, possible gaps, and scenario comparisons in real time, allowing managers to quickly identify risks and make data-driven scheduling decisions.

Here is the “to do list” for the project implementation tracking, but it will probably be updated during project progression as request.

Tasks	**What**	**Tools**	**Target**	**Status**
Data collection	Incidents dataset	Searching engine,LLM,reference	17th,Sept.	done
	Fire station dataset			done
	Territoires adamin data			done
	Montreal deomgraphic data			in progress
	Montreal weather data			open
	Montreal season,festival data			open
Data preparation	Data cleaning	Alteryx,Python	24th,Sept.	in progress
	Data digest			in progress
	Data standarization			in progress
Exploratory	Data union/join	Alteryx,Python,Tableau	1st,Oct.	open
	Data aggregation			in progress
	Data visualization			in progress
Modelling	Statistical modelling	Python,Alteryx	22nd,Oct	open
	ML modelling			open
	Modeling ensembling and optimaztion			open
Evaluation	Model test	Python,Alteryx	22nd,Oct.	open
	Model selection			open
Final Report	Report and presentation	MS office	3rd,Nov.	open

5. References

- The Service de sécurité incendie de Montréal at a glance
- Spatio-temporal analysis of fire incidences in urban context(case studies e.g., Chhetri et al., Mashhad/Ardabil studies)
- P.Villoria H.,et al (2023). Time series forecasting methods in emergency contexts
- R.Justin M.,et al (2021).Predicting emergency medical service call demand: A modern spatiotemporal machine learning approach
- Ville de Montréal. (2025). *Open Data Portal.* https://donnees.montreal.ca
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice.* https://otexts.com/fpp3/
- Zhou, Z., & Matteson, D. (2015). *Predicting Ambulance Demand: a Spatio-Temporal Kernel Approach.* https://arxiv.org/abs/1507.00364
- Pirklbauer, K., et al. (2019). Predicting the category of fire department operations.
- Ku, C. Y., et al. (2024). Predictive Modeling of Fire Incidence Using Deep Neural Networks (MDPI Fires)
- Legemaate, G. (2019). Applied urban fire department incident forecasting (case study: Amsterdam)
- Reference housing districts: https://www.donneesquebec.ca/recherche/dataset/vmtl-quartiers?utm_source=chatgpt.com
- Census population: https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/download-telecharger.cfm?Lang=E