Business Problem Framing Report

YCBS 299 - Data Science Capstone Project

Alex Djoboulian, Kritika Gupta, Daya Nambi and Rajesh Kamaraj

Team #4

Business problem

Context

- Montreal has **332,000 buildings** and a surface area of **499.3 km²**.
- Fire safety is handled by **67 fire stations** with **2,694 full-time staff**.
- The city conducts a maximum of **7,500 inspections annually**, which only covers **2.3% of total buildings per year**.
- With current resources, it would take **44.26 years** to inspect all buildings once.

Current Challenges

1. **Impossible Scale of Inspections -** The number of buildings far exceeds the inspection capacity, while traditional inspection methods cannot scale.
2. **Limited Workforce -** Workforce constraints with retirements leading to loss of institutional knowledge.
3. **Urban Growth -** Large numbers of new constructions increase fire risks and add to the inspection backlog.
4. **High Population Density -** With **4,106 inhabitants/km²**, fire risks have serious consequences on public safety.
5. **Rising Complexity -** Firefighters need to manage multiple systems: dispatch (RAO), scheduling, equipment tracking.

Problem statement:

–Predict high fire risk areas by month in the city of Montreal

Business problem rationale

Consequences of Inaction

- **Public Safety Risk:** Fire safety inspections and systems (e.g. sprinklers, compliance with fire codes) are associated with significant reductions in fire severity and casualty rates. For example, properties *with more complete fire safety* systems in Surrey, BC *saw up to ~59% lower fire severity*. City of Surrey
    - The 2023 Old Montreal fire (7 fatalities) exposed code-compliance problems and triggered a criminal investigation; media and docs reported prior fire-code issues at the building. (https://en.wikipedia.org/wiki/2023_Old_Montreal_fire)

- **Community displacement & service disruption** - Multi-unit fires in dense Montreal boroughs routinely displace dozens of residents and disrupt nearby services (recent examples in Montréal-Nord and Villeray).

- **Huge Economic Costs:** Fire losses—including property damage, insurance payouts, and public costs—are significant. The *Canadian Fire Statistics* reports large losses across provinces annually. Report Fire Losses in Canada - Canadian Fire Marshalls Council Also, a study estimated Canada's total cost of fire (including health, property, etc.) is much larger than commonly assumed. NRC Publications

- **Threats to business continuity -** Fires cause downtime, loss of assets, disruptions to operations. If buildings aren't properly inspected, preventive issues are more likely to go unnoticed, increasing risk of major losses

- **Wildfire Costs & Climate Impact**: Wildfire events impose huge financial burdens—e.g. Statistics Canada reported $945 million in insured damages in certain BC regions in 2023 from wildfires. Statistics Canada

Together, these consequences highlight the urgency of adopting predictive, data-driven inspection methods to reduce fire risk, protect residents, and safeguard Montreal's long-term resilience.

State-of-the-art review

Predicting fire risk is a complex challenge because it is shaped by building design, human behavior, and urban conditions. Around the world, researchers have shown that moving beyond traditional inspection cycles and using **data-driven methods**—such as machine learning, deep neural networks, BIM models, and geospatial analysis—can significantly improve the accuracy of identifying which buildings or areas are most at risk.

1. Machine learning based risk analysis and predictive modeling of structure fire-related casualties

   https://www.sciencedirect.com/science/article/pii/S2666827025000283

   This study analysed over 48,000 reported structure fire incidents in Oregon that occurred from January 2012 through August 2023. The dataset includes 2136 fires that led to civilian casualties including 317 confirmed fatalities. Bagged decision tree classifiers with random forest algorithm were used to quantify the importance of factors related to socioeconomic conditions, population characteristics, structural and behavioral incident details, and local infrastructure on the severity of injuries. Our results show that the age of victims, fire service response times, and availability of working smoke or fire detectors were among the most important parameters for predicting fatal outcomes of structure fires. Furthermore, a predictive Bayesian regularized neural network ensemble classifier was developed to model the severity of casualties and project a spatial risk classification on the census block level. **The network model achieves a prediction accuracy of 92.5 % for the classification of structural fire-related casualty severities.** With information aggregated to the census block scale and information related to specific fire incidents removed, the retrained model based solely on spatially available data reaches an 87.6 % severity classification accuracy. As the first statewide analysis of its kind, our spatial assessment provides a useful tool for resource allocation, risk factor reduction, and safety education efforts targeted to reduce the number of serious injuries or fatalities from structure fires.

2. Predictive Modeling of Fire Incidence Using Deep Neural Networks

   https://www.mdpi.com/2571-6255/7/4/136

   Developed and tested a **deep neural network (DNN)** model to estimate fire incidence likelihood. The model incorporated **demographic, architectural, and economic variables** (e.g., population density, building use, construction type, and socioeconomic indicators).The DNN model significantly outperformed traditional approaches. Achieved a **high coefficient of determination ($R^2 \approx 0.89$)**, indicating strong explanatory power. Model also showed **low RMSE (root mean square error)**, reflecting accurate predictions with minimal error spread. Results demonstrated that **building characteristics (age, type, and density)**, as well as **economic activity levels**, were among the most influential predictors of fire risk.

3. Data-Driven Structural Fire Risk Prediction for City Properties

   https://ojs.aaai.org/index.php/AAAI/article/view/30325

   This research contributes to the advancement of fire inspection practices by providing a data-driven approach. Future directions for this research include investigating alternative machine learning models to potentially improve prediction accuracy and robustness. Additionally, exploring suitable imputation techniques for missing data in the fire incident dataset can enhance the quality of predictions. Since **Random Forest worked best among the three models in the reduced data, combining appropriate imputation techniques with Random Forest could also result in higher performance**. Furthermore, extending the analysis to predict other categories of fires, such as the severity or cause of fires, can provide additional insights and aid in developing targeted prevention strategies. It is also worthwhile to consider alternative evaluation metrics that capture the specific needs and priorities of fire departments. By adopting the proposed and future directions, the utility of fire risk prediction can be enhanced, ultimately leading to improved public safety and the prevention of fire incidents.

4. Use of geospatial information and convolutional neural networks to predict the fire risk for the non-residential buildings -

   https://www.sciencedirect.com/science/article/abs/pii/S2352938521000069

   The article below uses geospatial information and convolutional neural networks to predict the fire risk for the non-residential buildings. The research used both tabular and image data to build a multi-modal model. The data source was the London Fire Bridge dataset. They used a convolutional neural network for the image data and a multi-layer perceptron for the tabular data. **The ROC AUC score of the model achieved 0.8195.** When analyzing the features, the researchers discovered that **building use was the greatest factor for fire risk.**

5. From Occurrence to Consequence: A Comprehensive Data-driven Analysis of Building Fire Risk

   This large-scale study analyzed **over one million fire incident reports** combined with **socioeconomic and building-level data** to predict both the **occurrence** of fires and their **consequences** (e.g., injuries, spread, and severity). Using advanced data-driven methods, the researchers identified that **building safety features**—such as smoke detectors, sprinklers, and alarms—were among the strongest predictors of reduced harm. The analysis also highlighted how **socioeconomic conditions** and **building characteristics** shape fire risk across communities.
   Scale: One of the largest fire risk studies to date, leveraging >1M incidents.
   Predictive Focus: Looked beyond occurrence to also model **severity of outcomes**.
   Findings: Buildings with safety systems had significantly lower spread and casualty rates.
   Policy Implication: Demonstrates that predictive models can inform **targeted inspections**, resource allocation, and **code-enforcement strategies** to reduce harm.

6. Use of BIM (Building Information Modeling) technology to assess fire risk in buildings -

   https://www.sciencedirect.com/science/article/abs/pii/S0360132321005898

The next article below uses **BIM (Building Information Modeling) technology** to assess fire risk in buildings. The study focuses on **three frameworks; fire risk factor, probability of fire, and the consequence of fire.** The researchers built a hierarchical system with 3 levels. The first is the target layer which evaluates the fire risk. The second is the Criterion layer which factors groups in probabilities . The third is the index layer where they take measurable variables into account such as the number of sprinklers. The research used Analytic Hierarchy Process (AHP) to determine the highest risk factors for fire. The score for each factor was based on actual building data and the scores were standardized so that the results would be comparable for different buildings. In order to calculate the fire risk index, a weight sum model method was implemented. The model was then validated by comparing results to known buildings that are safe, which would confirm the realistic result of the model. The results proved that the model was more efficient than other traditional methods.

7. Designing an indicator system to assess urban fire risk -

   https://www.sciencedirect.com/science/article/abs/pii/S0379711225002000

   This article below designed an indicator system to assess urban fire risk. The research claims that the fire risk involves the behavior of humans, social systems and the environment. For the method of study, they used statistical and spatial analysis to see how the indicators correlate with historical fire data and the urban spatial features. The study also used geospatial methods such as hotspot analysis to see the patterns geographically for fire risk relative to the indicators. To measure the results of the study, the researchers used Analytic Hierarchy Process to assign numerical weights for each indicator. Another measurement is a numerical fire risk score for each urban unit. For the validation of their system, they compared the predicted high risk areas to the historical fire hotspots and the results showed a 75%. The correlations for each metric ranged from 0.3 to 0.6, and the district scores for each area in the region ranged from .2 to .72.

   From the Fire Risk Prediction Presentation by Ville de Montréal Jan 16, 2024 Martin-Guy Richard CISO - Chief Information Security Officer. *Below are Key takeaways from the reports that's helpful in our study -*

8. Predictive Modeling of Building Fire Risk Metro21: Smart Cities Initiative

   http://michaelmadaio.com/Metro21_FireRisk_FinalReport.pdf

   Fire departments cannot inspect every property annually, and legacy systems prioritize inspections based on permits or rules—not actual fire risk. A risk-based, data-driven approach is essential to focus limited resources on the highest-risk buildings. Data-driven Predictive Model: By combining historical fire incidents, property assessments, and code violations, the Metro21 team ***built a predictive model that correctly identified ~55% of actual fire incidents within a 6-month window***—compared to just 0.21% accuracy if guessing randomly. The model outputs property-level risk scores (1–10), delivered via dashboards and interactive maps. This helps inspectors and chiefs target high-risk buildings, integrate risk scores into planning, and ultimately improve public safety

9. Fire Risk Prediction Models (Probability approach to risk-based inspections)

   https://omfpoa.com/wp-content/uploads/2021/04/OMFPOA-Fire-Risk-Prediction-Model-1.pdf

   This OFMEM model shows us a strong template for Montreal: by combining our building registry, past fire calls, and socio-economic/geospatial features, we could similarly generate risk scores per building, allowing us to target inspections much more effectively. ***Their validation strategy (training/test split, use of recall/AUC) also gives us confidence about evaluating our models*** in a way that reflects actual performance, not just optimistic numbers.

10. A Building Fire Risk Prediction Validation Project

    https://fireunderwriters.ca/assets/img/FUS_Building_Fire_Risk_Validation_Project.pdf

    The study validated machine-learning models for building fire risk in Vancouver and New Westminster, using datasets such as fire incidents, inspections, property assessments, census, crime, and building footprints. Models trained on 2013–2016 data predicted ~70% of 2017 fires with acceptable false positive rates (23–29%). Importantly, risk-based targeting nearly tripled severe violations discovered in early inspections (from 15% to 44% in Vancouver). Influential features included building area, land value, prior incidents, employment income rates, and presence of sprinklers. The project shows how Canadian municipalities can use existing datasets to priori tize inspections, reduce fire risk, and align with "smart cities" approaches.

11. Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta

    https://poloclub.github.io/polochau/papers/16-kdd-firebird.pdf

    The Firebird project (Atlanta, 2016) showed how data-driven models can transform fire inspections by combining diverse datasets—fire incidents, property records, business licenses, crime, and demographics—to both predict fire risk and identify previously unlisted inspectable properties. Using machine learning (SVM, Random Forest), ***Firebird achieved about 71% true positive rate with ~20% false positives, a strong benchmark*** compared to random selection. Importantly, the system didn't just produce scores; it delivered interactive maps and dashboards that allowed fire officials to prioritize inspections geographically and by property type. For Montreal, this demonstrates that integrating multiple local datasets can help uncover overlooked buildings, focus limited inspection resources, and make the results directly actionable for city planners.

Conclusion

After reviewing these studies, our guiding theses for moving this project forward are:

The literature shows that fire risk prediction is most effective when it combines diverse datasets, leverages advanced modeling techniques, and produces outputs that are actionable for city operations. These insights provide us with clear directions for how to design, validate, and operationalize our model for Montreal.

1. **Fire risk is multi-dimensional** — shaped by building features, human behavior, socio-economic factors, and urban context.
2. **Data-driven methods consistently outperform traditional inspection cycles**, which are too resource-heavy and rule-based.
3. **Machine learning and ensemble models (Random Forest, Neural Nets)** achieve high predictive accuracy (70–92%) in different regions.
4. **Socioeconomic and demographic variables matter** (income, density, crime rates, population age) alongside building characteristics.
5. **Building safety systems (sprinklers, alarms, detectors)** strongly reduce fire severity and are top predictors across studies.
6. **Geospatial analysis (hotspot mapping, parcel data, GIS)** helps visualize and prioritize risks geographically.
7. **Risk scores and dashboards** make predictions actionable for inspectors, turning analytics into operational decision tools.
8. **Canadian pilots (Vancouver, New Westminster)** prove local feasibility—risk-based targeting tripled severe violations found.
9. **Validation strategies (train/test splits, AUC/recall metrics)** are essential to ensure realistic performance in imbalanced datasets.
10. **Montreal can replicate this by integrating open datasets** (incidents, property registry, census, crime, fire stations) to build predictive models that focus inspections where they matter most.