That’s an excellent and very realistic set of considerations — and you’re thinking exactly like a good **team coordinator** should.

Given your constraints:

* 🧍‍♀️ 4 team members
* ⏰ 3 weeks total
* 💼 All with full-time jobs
* 💻 Working remotely / limited meeting overlap
* 🎓 Project is part of an after-work **training course**
* 🗂️ The goal is **a meaningful but feasible “fire-risk readiness” project** using the **SIM Interventions** dataset

Let’s make this **manageable**, **collaborative**, and **aligned with course expectations** — not overloaded with unnecessary data or complexity.

**🔹 1. Adjusted Dataset Plan (Keep It Simple, Yet Insightful)**

Given your 3-week window and remote setting, you should **limit yourself to 2–3 datasets total**:

| **Dataset** | **Why Keep It** | **Type of Insight It Adds** |
| --- | --- | --- |
| **Main:** SIM Fire Interventions (from Open Canada) | Core dataset – defines your dependent variable (incidents) | Foundation for the whole analysis |
| **+1:** Building / Property Dataset (Montréal open data) | Adds physical risk factors (age, type, density) | Enables normalization or exposure understanding |
| **+2:** Census / Demographics Dataset (Statistics Canada) | Adds social & human context (population, dwellings) | Helps interpret spatial fire risk differences |

Optional (only if time allows): simple **weather summary** (e.g., average temp/humidity per month) from Environment Canada — **but treat it as a stretch goal.**

👉 **So total: 2 core datasets + 1 optional (3 max).**

**🔹 2. Brainstorming Topics for the First Team Meeting**

Below are discussion points your team can use to guide a **brainstorming session (60–90 mins)**:

**🔸 A. Problem Definition & Scope**

* What is the **main question** we want to answer? (e.g., “Which areas of Montréal have higher fire risk and why?”)
* Should our **target variable** be:
  + (a) classification: low/medium/high risk per grid
  + (b) forecasting: next-month risk per grid
* What **deliverable** is most valuable:
  + Dashboard / map visualization?
  + Report + presentation?
  + Predictive model (simple but interpretable)?

**🔸 B. Data Strategy**

* Which **additional datasets** can we realistically manage in 3 weeks?
  + Prioritize availability, cleanliness, and ease of join (e.g., both have coordinates or postal codes).
* What **spatial unit** to use: 500 m grid or neighborhood polygons?
* Decide **software stack**: Python (Pandas, GeoPandas, Scikit-learn), QGIS, or Google Colab?

**🔸 C. Team Logistics**

* How many times per week can we meet live? (e.g., twice a week + async communication)
* Which platform for collaboration: Google Drive / GitHub / Slack / MS Teams?
* How to document each step to avoid confusion (e.g., shared notebook, version control).

**🔸 D. Modeling & Output**

* How simple or complex should our model be? (Maybe logistic regression or decision tree — not deep learning.)
* What metrics to evaluate? (Accuracy, recall for “high-risk” areas)
* How to visualize results to make them **actionable** for fire authorities?

**🔹 3. Suggested 3-Week Plan with Activities & Milestones**

| **Week** | **Focus** | **Key Activities** | **Expected Deliverables** |
| --- | --- | --- | --- |
| **Week 1: Data Understanding & Planning** | Define scope, download data, align on methodology | - Kickoff meeting (set goals, roles) - Explore SIM data structure & quality - Decide on supplementary datasets (max 2–3) - Define grid / spatial unit - Assign roles | - Project scope document - Data samples - Clear roles table |
| **Week 2: Data Processing & Feature Engineering** | Cleaning, joining, creating features | - Download & clean additional data - Join datasets (by coordinates or polygons) - Create features (incident rate, building density, population density) - Draft simple EDA visuals - Agree on target variable definition (classification or forecasting) | - Clean joined dataset - Feature list - First visualizations |
| **Week 3: Modeling, Evaluation, and Presentation** | Build model + summarize findings | - Train/test split, build simple model (logistic regression, decision tree) - Evaluate & interpret results - Create visual output (heatmap of risk zones) - Write summary report & prepare presentation | - Risk map / model output - Short report - Team presentation |

**🔹 4. Example Task Distribution (Balanced and Collaborative)**

Each person has ownership *and* participates jointly in the “key shared tasks” during meetings.

| **Role** | **Main Responsibility** | **Works Independently On** | **Joins Team For** |
| --- | --- | --- | --- |
| 🧭 **Team Lead / Coordinator** | Organize schedule, oversee deliverables, integrate work | Project timeline, task coordination | Kickoff, weekly check-ins |
| 🧹 **Data Engineer / Integrator** | Download, clean, join datasets | Data preprocessing scripts (Pandas, GeoPandas) | Defining schema, merging data |
| 📊 **Data Analyst / Visualizer** | Explore data, create charts/maps | EDA, visualization (Matplotlib, Seaborn, Folium) | Feature design, interpreting results |
| 🤖 **Modeler / Evaluator** | Build and evaluate model | Baseline model, metrics, classification results | Deciding target variable and thresholds |

✅ **Team-wide meetings** (1–2 hrs each, weekly) should focus on:

* Agreeing on how to define the target variable (Week 1)
* Reviewing joined dataset and early visuals (Week 2)
* Reviewing model results and presentation draft (Week 3)

Everything else (data cleaning, plotting, writing) can be done asynchronously, shared via notebooks or a shared Drive folder.

**🔹 5. Example Brainstorm Outputs to Aim For**

By the end of the first meeting, your notes could include:

* ✅ Problem statement: “Predict high-risk fire zones in Montréal using intervention and building data.”
* ✅ Datasets: SIM Interventions + Building data + Census (optional weather).
* ✅ Target variable: 3-level fire risk classification (Low / Medium / High).
* ✅ Model type: Decision Tree or Logistic Regression.
* ✅ Visualization: Montréal heatmap showing risk levels per grid.
* ✅ Roles: Each team member assigned and timelines set.

Would you like me to provide a **sample agenda and discussion guide** (with suggested time splits and guiding questions) for your **first 60–90 minute online team meeting**?  
That would help you structure the conversation efficiently and ensure every member contributes meaningfully.

Perfect 👍 — below is a **structured brainstorming and project plan** tailored to your situation:

* 4 team members
* fully online coordination
* only 3 weeks to deliver
* using the *SIM Fire Intervention dataset (Montreal)*
* goal: create a model that helps authorities **forecast or classify fire risk** (to improve preparedness)
* constraint: limited time and coordination → **simplify dataset scope (max 2–3 complementary datasets)**

**🔹 1. Overall Project Vision**

**Goal:** Build a simple, interpretable model that helps fire departments and city authorities **predict or classify fire risk zones** in Montreal — either by *risk level* (High / Medium / Low) or *forecasted number of incidents next month*.

**Outcome:** Deliver a compact prototype dashboard or notebook visualization that highlights risk zones and gives insights into contributing factors (e.g., past incidents, building density, land use, etc.)

**🔹 2. Recommended Datasets (Maximum 3 Additional)**

| **Dataset** | **Why Include It** | **How It Helps Target Variable** |
| --- | --- | --- |
| **SIM Fire Incidents (main)** | Base dataset with all fire intervention details | Used to compute historical fire frequencies → core for classification/forecasting |
| **Building Footprints / Usage Dataset** (e.g., from Données Montréal) | Adds context: residential vs. industrial, age, area | Key predictor of risk — older/dense buildings = higher risk |
| **Zoning / Land Use Data** | Helps cluster regions by permitted activities | Refines fire likelihood (industrial zones vs. parks) |
| **Weather or Climate Data (optional)** | Only if easily accessible (temp, humidity) | Can be used for short-term forecast-type prediction |

→ Keep to **3 datasets total** (main + 2 others) to stay realistic within 3 weeks.

**🔹 3. Suggested Plan Structure (by Week)**

**Week 1 – Exploration & Planning**

**Objectives:**

* Understand data structure and agree on project scope.
* Decide final target variable (classification vs. forecasting).
* Assign team roles.

**Team Meeting #1 (Kickoff)**

* Everyone reviews the main dataset before the meeting.
* During meeting:
  + Agree on **target variable definition**.
  + Decide **grid or polygon size** (e.g., 500m × 500m).
  + List final 2 other datasets.
  + Assign **roles and responsibilities** (see below).

**Between meetings:**

* Each person prepares one small deliverable (EDA, dataset cleaning, map preview).

**Week 2 – Modeling & Integration**

**Objectives:**

* Clean and merge datasets.
* Create features and target variable.
* Prototype model(s).

**Team Meeting #2 (Data Review & Modeling)**

* Review merged dataset.
* Discuss correlation and potential features.
* Decide simple baseline model (e.g., Random Forest for classification or linear regression for forecasting).

**Between meetings:**

* Each member develops their part (see roles below).
* Shared GitHub or Google Drive folder for code + notes.

**Week 3 – Evaluation & Presentation**

**Objectives:**

* Validate model performance.
* Interpret results.
* Create a short report or dashboard.

**Team Meeting #3 (Final Integration)**

* Combine all outputs (visuals, metrics, maps).
* Discuss insights and policy recommendations.
* Prepare slides/notebook for submission.

**🔹 4. Suggested Role Distribution (4 Members)**

| **Role** | **Key Tasks** | **When to Collaborate** |
| --- | --- | --- |
| **Data Engineer** | Data cleaning, merging datasets, coordinate shared code notebook | Week 1–2 |
| **Data Scientist (Modeling)** | Create features, build baseline model, tune parameters | Week 2 |
| **Data Analyst / Visualization Lead** | Visualize EDA, risk map, model results | Week 2–3 |
| **Project Coordinator / Presenter** | Organize meetings, integrate results, finalize report/presentation | All weeks |

✅ Everyone joins during **key decision points** (e.g., target variable definition, model selection, interpretation) to ensure full team input.

**🔹 5. Brainstorming Discussion Topics for Team**

Here are **brainstorming ideas/questions** to discuss collaboratively during meetings:

**🔸 Target Definition**

* Should we predict *fire occurrence next month* or *risk level (low/med/high)*?
* Which time frame makes sense given the data frequency (monthly, seasonal)?
* How do we define “high risk”? (e.g., top 20% of grid cells by historical frequency)

**🔸 Data & Features**

* Which additional datasets are feasible within time?
* Should we spatially aggregate (500m grid) or use administrative zones?
* What engineered features could help (e.g., avg. past fires, nearby hydrants, population density)?

**🔸 Modeling**

* Should we start with a simple baseline model (e.g., logistic regression or random forest)?
* How to evaluate (accuracy, F1-score, RMSE, etc.)?
* How to visualize model outputs on a city map?

**🔸 Collaboration & Tools**

* Shared code via GitHub or Google Colab?
* How to handle time zone or meeting scheduling conflicts?
* One person merges everyone’s work — or each commits to the repo?

**🔹 6. Deliverables Summary**

| **Deliverable** | **Description** | **Owner** |
| --- | --- | --- |
| **Data Summary Notebook** | EDA and data cleaning report | Data Engineer |
| **Model Notebook** | Model development and evaluation | Data Scientist |
| **Visualization Dashboard / Notebook** | Risk map or time-based forecast visualization | Data Analyst |
| **Final Report / Presentation** | Summary of findings, recommendations | Project Coordinator |

Would you like me to follow up with a **sample timeline + meeting agenda templates** (for 3 weeks, with estimated hours and milestones) to help the team coordinate efficiently despite limited availability?