

Assignment 4 Solution Development Report

Capstone Project: Wildfire Analysis in Alberta

Course: BIA 5450-OLA

Team 6

Samer Al-Obaidi

GROUP 6 MEMBERS

| Names | N-number |
|----------------------------|------------------|
| Akash Anand | N01646970 |
| Ayantika Chatterjee | N01605074 |
| Mohit Rathod | N01579983 |
| Ruchi Rajesh Tiwari | N01607974 |
| Taranjeet Kaur | N01588653 |

Table of Contents

| | | |
|------------|--|-----------|
| 1. | Executive Summary..... | 3 |
| 2. | Introduction..... | 4 |
| 3. | Business Problem Requirements | 4 |
| 4. | Analytics Questions | 5 |
| 5. | Scope Statement | 6 |
| 6. | Data Sources / Key Data Entities and Flows..... | 7 |
| 7. | Brief Overview of Data Manipulation Process and Data Output..... | 10 |
| 8. | Solution Design & Fit into the Existing IT Architecture | 16 |
| 9. | New Solution Implementation and Outcome Testing..... | 18 |
| 10. | Potential Solution Optimization..... | 30 |
| 11. | Appendix..... | 32 |
| | References | 41 |

1. Executive Summary

This report presented the development and deployment of an interactive wildfire-analytics dashboard to support Alberta Wildfire's incident management and resource-planning teams. Following a structured CRISP-DM approach, the team first engaged stakeholders to define key questions around wildfire trends, causes, weather impacts, regional risk, and response performance. Historical wildfire data (2020–2023) and meteorological records were then ingested via Python, where they were cleaned, imputed, and transformed before export to CSV.

In Power BI Desktop, we constructed a star-schema model centered on a fact table of wildfire incidents and sixteen supporting dimension tables (e.g., fire origin, weather bins, detection agents, calendar). Power Query was used to promote headers, cast data types, split date/time fields, and categorize continuous variables into meaningful bins (temperature, wind, humidity, and season). We authored DAX measures to calculate containment rates, year-over-year changes, and a composite risk score, and created What-If parameters for dynamic axis selection and weather filters.

Five analytics questions were addressed through distinct dashboard pages: annual trends in fire occurrences and size; breakdown of leading causes by region and season; correlations between weather conditions and fire behavior; identification of high-risk zones; and comparison of detection agents, resource dispatches, and containment effectiveness across regions. Each page was rigorously tested against raw data for metric accuracy, visual integrity, and interactivity, confirming that the dashboard reliably reflected underlying patterns.

The solution was published to Power BI Service, where scheduled daily refreshes from OneDrive ensured up-to-date reporting. Role-based access controls were configured to limit views to authorized wildfire analysts. As a result, the dashboard delivered on its objective: transforming complex, multi-source datasets into clear, actionable insights that enhanced situational awareness and informed strategic resource allocation.

Finally, several optimization opportunities were identified—such as automating data ingestion via gateways, integrating predictive risk models, enriching context with external environmental

data, and improving user guidance—to further advance the dashboard from retrospective analysis toward real-time, anticipatory decision support.

2. Introduction

This report will document the end-to-end development of an interactive wildfire analytics dashboard for Alberta Wildfire, demonstrating how historical wildfire and environmental data from 2020–2023 were ingested, modeled, and visualized to support data-driven decision-making and resource allocation. In the first section, the Executive Summary will highlight the project’s objectives, methodology, and key outcomes.

In the following sections, readers will be guided through each major phase of the project: **Section 3** will provide the Business Problem Overview, defining the operational challenges and analytical goals that motivated the dashboard’s creation. **Section 4** will outline the specific Analytics Questions that shaped the design of each report page. **Section 5** will present the Scope Statement, detailing the in-scope activities, out-of-scope boundaries, and final deliverables. **Section 6** will describe the Data Sources, Key Data Entities, and Flow processes underpinning the model, while **Section 7** will summarize the Data Manipulation Process and resulting Data Outputs. **Section 8** will detail the Solution Design and its integration into Alberta’s existing IT architecture. Finally, **Section 9** will explain the Solution Implementation phases and present the Outcome Testing results for each dashboard module, demonstrating the accuracy and reliability of the analytics solution.

3. Business Problem Requirements

Wildfires continue to pose serious challenges in Alberta, threatening public safety, damaging ecosystems, and disrupting local economies. With rising frequency and intensity—driven by environmental changes and human activity—effective wildfire management requires a data-driven approach that can reveal patterns, support strategic decisions, and guide resource allocation.

Although Alberta collects extensive wildfire data, it is often underutilized due to the absence of integrated analytical tools. This limits the ability of response teams to fully understand regional

variations, seasonal trends, and fire behavior under different weather conditions. Without a centralized system, challenges persist in identifying high-risk zones, evaluating containment efforts, and distributing resources efficiently.

To address these gaps, our project focuses on developing an interactive Power BI dashboard that analyzes wildfire activity in Alberta from 2020 to 2023. This dashboard centralizes wildfire data across multiple dimensions—fire origins, sizes, causes, seasons, and weather patterns—to enable comprehensive trend analysis. It supports data-driven assessments of regional fire distribution and response strategies, improving stakeholders' ability to evaluate past actions and plan future operations. By illuminating containment performance and detection agent effectiveness, the solution enhances understanding of how resources have been deployed and where improvements are needed. Ultimately, the dashboard guides wildfire management teams in pinpointing the most vulnerable areas and allocating personnel and equipment more efficiently.

Built on a star schema data model, the dashboard offers dynamic filters, KPIs, geospatial visualizations, and categorized insights—turning complex datasets into an intuitive, accessible format. By equipping stakeholders with the ability to explore and interpret wildfire trends, this project enhances situational awareness, strengthens operational planning, and contributes to a more informed and coordinated wildfire response across Alberta.

4. Analytics Questions

We developed five key analytics questions based on the business challenges identified. Each question is supported by relevant visualizations in the Power BI dashboard and addresses a specific decision-making need:

1. How have wildfire occurrences and sizes in Alberta changed annually (2020–2023)?

- **Analytics:** Descriptive
- **Decision:** Understand year-over-year trends to assess changes in fire activity and adjust long-term preparedness strategies.

- 2. What are the leading causes of wildfires across Alberta's regions and seasons?**
 - **Analytics:** Diagnostic
 - **Decision:** Identify dominant ignition sources to guide targeted prevention and public awareness efforts.
- 3. How do Alberta's weather conditions influence wildfire size and spread?**
 - **Analytics:** Diagnostic
 - **Decision:** Assess the role of temperature, humidity, and wind in fire behavior to support risk awareness and seasonal readiness.
- 4. Which Alberta regions are most at risk based on historical data?**
 - **Analytics:** Descriptive
 - **Decision:** Pinpoint high-risk areas for prioritizing resources, implementing safety protocols, and planning response strategies.
- 5. How do detection agents, resources, and containment vary across Alberta's regions?**
 - **Analytics:** Diagnostic
 - **Decision:** Evaluate operational effectiveness and identify opportunities to optimize resource deployment and detection strategies.

5. Scope Statement

This project aims to develop a data-driven analytical solution to support wildfire management efforts in Alberta by analyzing wildfire activity from 2020 to 2023. The primary objective is to assist decision-makers in understanding wildfire trends, identifying high-risk regions, and evaluating response strategies through a user-friendly Power BI dashboard.

The solution transforms complex wildfire records into interactive visualizations that enhance situational awareness, guide resource planning, and support targeted prevention strategies.

across Alberta. It centralizes data from 2020 through 2023, organizing it by fire origin, size class, cause, season, and weather conditions to enable descriptive and diagnostic analytics on historical trends, regional risks, containment performance, and detection effectiveness.

In scope for this project is the extraction and cleaning of wildfire data using Python; the design and development of a star-schema data model in Power BI; the creation of interactive dashboards and KPIs to examine trends, causes, seasonal patterns, regional risk profiles, and response efforts; and the delivery of a fully functional Power BI report accessible through Power BI Service. We will conduct descriptive and diagnostic analyses to generate insights on containment success rates, detection agent performance, weather impacts, and resource deployment, and we will provide strategic recommendations for optimizing resource allocation and mitigation strategies based on these historical findings.

Out of scope are any predictive modeling or machine-learning-based wildfire forecasting efforts, real-time wildfire tracking or alerting systems, sensor integrations, and the implementation or deployment of physical firefighting infrastructure or operational field strategies. The project will also not involve direct collaboration with emergency response agencies or integration with live, proprietary systems beyond publicly available data sources.

The primary deliverables include a cleaned and structured wildfire dataset; a Power BI dashboard featuring dynamic filters, geospatial maps, and visual KPI cards; comprehensive reports summarizing historical patterns and identifying operational gaps; and actionable recommendations to improve wildfire readiness and resource planning. This solution is intended as a decision-support tool that enhances Alberta's capacity for wildfire analysis and long-term planning, providing a scalable foundation for future enhancements even though it does not encompass live data collection or ongoing maintenance.

6. Data Sources / Key Data Entities and Flows

Data Source

The wildfire data for this project was sourced from the **Government of Alberta Open Data Portal**, specifically from the historical wildfire incidents dataset available at:

[FP Historical Wildfire Data \(2006–2024\)](#)

The dataset includes detailed records of wildfire events across Alberta, capturing information on ignition causes, fire size, weather conditions, containment status, and detection and suppression resources.

Key Data Entities

The data model, designed using a star schema, comprises one central fact table and several supporting dimension tables. The key entities include:

- **Wildfire Incident:** Captured in the `fact_Wildfire_Table`, this entity stores attributes such as fire size, date of discovery, fire type, and location ID.
- **Geographical Location:** Includes latitude, longitude, slope position, and fire origin classification.
- **Weather Data:** Dimension tables for temperature, wind speed, and humidity are binned into categories to analyze spread dynamics.
- **Detection & Reporting:** Includes agents, detection types, and initial actions taken; this helps evaluate operational response.
- **Fire Response:** Resources dispatched, containment success, and assessment resources are captured to understand response effectiveness.
- **Fuel and Fire Type:** Dimensions like fuel type and fire size class allow analysis of fire intensity and growth behavior.
- **Calendar & Seasonality:** A date table and season dimension support trend analysis over time.

[\(See Appendix B: Data Dictionary\)](#)

These entities are related through unique IDs, with all relationships established as one-to-many (1: *), forming a structured and scalable model for analysis.

Data Flow Process

(See Appendix A: Process Flow Diagram)

1. **Extraction:** Data was downloaded from the Alberta Open Data Portal and imported into Python using the pandas library.
2. **Preprocessing:**
 - Removed irrelevant columns and incomplete rows
 - Converted data types and formatted date columns
 - Encoded categorical fields where necessary
 - Imputed missing values using statistical methods or domain knowledge
 - Normalized field names for consistency
3. **Storage:** The cleaned data was exported as .csv files and uploaded to **OneDrive** for seamless integration with Power BI.
4. **Power Query Transformation:**
 - Additional filtering and formatting
 - Column renaming and data type adjustments
 - Merging reference tables for dimension creation
5. **Data Modeling:**
 - Developed a **star schema** structure with a central fact table (fact_Wildfire_Table) and multiple dimension tables (e.g., dim_L_fire_origin, dim_FR_detection_agent, dim_FTA_fire_size_class, etc.).
 - Created a custom dim_Calendar_Table to support time-based filtering and aggregation.

6. **Measure Development:** Created calculated columns and DAX measures to enable KPIs such as Containment Rate, YoY Area Burned Change, Detection Agent Share, and Regional Risk Score.
7. **Report Design:** Built interactive dashboards in **Power BI Desktop** using cards, maps, line charts, bar graphs, and heatmaps.
8. **Publishing:** The final report was published to **Power BI Service**, enabling real-time collaboration and sharing with stakeholders.

This structured data architecture enabled a comprehensive analysis of wildfire trends, causes, regional risk, weather influence, and response efforts across Alberta from 2020 to 2023.

7. Brief Overview of Data Manipulation Process and Data Output Using Python:

To prepare the Alberta wildfire data for analysis and visualization in Power BI, we executed a structured series of data manipulation steps in Python. The process ensured the resulting dataset was clean, consistent, and optimized for our star-schema model.

1. Data Extraction

- Imported the Excel dataset (2006–2024 historical wildfires) directly into a Pandas DataFrame.

2. Initial Missing-Value Assessment

- Calculated the percentage of missing values per column.
- **Column pruning:** Removed any column with over 50 % missing data.
- **Row filtering:** For columns with low missingness (< 5 %), dropped only those rows to preserve overall data integrity.

3. Irrelevant Column Removal

- Identified and dropped metadata or auxiliary date columns (e.g., dispatch and arrival timestamps) that were not needed for our analyses.
- Ensured only the primary event dates (fire_start_date, discovered_date) remained.

4. Date and Type Conversions

- Converted date fields to datetime types.
- Imputed missing discovered_date values by adding a small random offset to the known fire_start_date.
- Dropped any records lacking a valid start date.

5. Numeric and Categorical Imputation

- **Numeric fields** (spread rate, temperature, wind speed, humidity): Replaced missing values with the column mean.
- **Categorical fields** (wind direction, weather category, slope position): Filled missing entries with the most frequent category (mode).

6. Advanced Imputation via KNN

- For dimensions like **fuel type** and **activity class**, encoded categories as numeric labels, introduced an “Unknown” category for true nulls, and applied a K-Nearest Neighbors imputer on key features (location, weather, and interim labels).
- Converted imputed labels back to their original categorical form and removed temporary encoding columns.

7. Column Renaming and Standardization

- Renamed fields to concise, Power BI-friendly names (e.g., fire_size_ha → Fire Size (Ha), relative_humidity → Humidity (%)).
- Applied consistent casing and removed special characters for seamless model integration.

8. Output and Storage

- Exported the final cleaned DataFrame to `cleaned_data.csv` for ingestion into Power Query.
- Saved intermediate reports—such as missing-value summaries before and after dropping—to support auditability.
- Uploaded CSV output “`cleaned_data.csv`” to OneDrive, enabling continuous synchronization with the Power BI Service.

These manipulations produced a robust, high-quality dataset that underpins our star-schema in Power BI, ensuring reliable analytics and smooth dashboard performance.

Using Power Query:

After exporting the cleaned CSV from Python, we used Power Query in Power BI Desktop to shape the central fact table and link it to our dimension tables. The following steps outline the key transformations and model-design activities:

A. Fact Table Preparation in Power Query

1. Source & Header Promotion

- Imported `cleaned_data.csv` and promoted the first row to column headers.

2. Preserving Original Date-Time

- Duplicated the three primary date columns (`Start_Date`, `Discovery_Date`, `Reported_Date`) and renamed the copies (e.g. `Start_Date_Copy`) so that both date and time components could be extracted later without losing the original values.

3. Data-Type Casting & Column Cleanup

- Converted each field to its correct type: numbers (`Fire_Size_Ha`, `Temperature_C`, etc.), text (causes, agent types), datetimes, and later separated out time fields.

- Removed the redundant **Year** column (now captured via the date fields).

4. Surrogate Key & Column Reordering

- Added an **Index** column and renamed it to `Fact_Table_ID` for a unique surrogate key.
- Reordered and renamed columns into a logical sequence for reporting.

5. Binning & Rounding

- Rounded temperature to one decimal place.
- Added custom “bin” columns for Temperature, Wind Speed, and Humidity using `Table.AddColumn` with conditional logic (e.g., “Mild,” “Hot,” “Calm,” “Gale,” “Very Low,” “High,” etc.).
- Inserted a Season column derived from the month of `Start_Date` (“Winter,” “Spring,” etc.).

6. Hierarchical Merges with Dimension Tables

For each binned or categorical field, we performed a left-outer `Table.NestedJoin` against its corresponding dimension query and then expanded only the surrogate key column:

- **Fire Origin** → `dim_L_fire_origin`
- **Temperature_Binned** → `dim_W_temp_binned`
- **Weather_Over_Fire** → `dim_W_weather_over_fire`
- **Wind_Direction** → `dim_W_direction`
- **Wind_Speed_Binned** → `dim_W_wind_speed_binned`
- **Humidity_Binned** → `dim_W_humidity_binned`
- **Fire_Size_Class** → `dim_FTA_fire_size_class`
- **General_Cause** → `dim_FC_general`

- **Activity_Class** → dim_FC_activity_class
- **Fire_Type** → dim_FT_fire_type
- **Slope_Position** → dim_FT_slope_position
- **Fuel_Type** → dim_FT_fuel_type
- **Detection_Agent** → dim_FR_detection_agent
- **Assessment_Resource** → dim_FR_assessment_resource
- **Dispatched_Resource** → dim_FR_dispatched_resource
- **Initial_Action** → dim_FR_initial_action
- **Detection_Agent_Type** → dim_FR_detection_agent_type
- **Season** → dim_W_seasons

After each join, the original text column was removed, leaving only the foreign key ID.

7. Time Component Extraction & Cleanup

- Duplicated the date copies, converted them to time types, and renamed to Start_Time, Discovery_Time, Reported_Time.
- Finally removed all intermediate datetime and time columns, preserving Start_Date, Discovery_Date, and Reported_Date as date-only fields.

8. Final Sorting & Anomaly Fix

- Sorted rows by Start_Date ascending.
- Applied a one-off date replacement to correct an outlier record.

B. Dimension Tables Creation

- Each dimension (e.g., dim_L_fire_origin, dim_W_temp_binned, dim_FC_general, etc.) was built in its own query by extracting the distinct values from the cleaned dataset, assigning each a surrogate key (e.g., Fire-Origin_ID, General_Cause_ID), and promoting it to a stand-alone table.

C. Star Schema & Relationships

- In the Power BI **Model** view, we arranged the fact_Wildfire_Table at the center and connected each dimension table via one-to-many relationships on the corresponding surrogate keys.
- The calendar dimension (dim_Calendar_Table) was linked to the Start_Date, Discovery_Date, and Reported_Date fields to enable time-intelligence functions.

This end-to-end transformation pipeline—from CSV import through Power Query shaping, dimension table creation, and star-schema modeling—ensures a performant, scalable foundation for all subsequent analytics and dashboard visuals.

Using DAX:

To enrich the dashboard's interactivity and risk analysis—without diving into formula details—we introduced two What-If parameters and three core calculated columns in the **fact_Wildfire_Table**.

A. What-If Parameters

- **Scatterplot Parameters**

A dynamic selector that lets users switch which fields appear on the X-axis and Y-axis of the custom scatterplot visual. This makes it easy to explore relationships between any two numeric measures without rebuilding the chart.

- **Weather Parameter**

A single-value slicer that drives multiple visuals by focusing on one weather attribute at

a time (e.g., Temperature, Humidity, Wind Speed). Users can toggle between these and instantly see how each factor influences fire behavior across all charts.

B. Calculated Columns

- **Containment_Per_Fire**

Quantifies the proportion of the burned area that was successfully assessed (i.e., contained) for each incident. This metric feeds KPI cards and regional comparisons of containment effectiveness.

- **Risk Score**

A continuous index (from 0 to 1) that combines normalized measures of fire size, temperature, wind speed, humidity, and containment rate. By weighting these factors, it provides a single severity score for every wildfire event.

- **Risk Category**

Translates the continuous Risk Score into five intuitive buckets—Minimal, Low, Moderate, High, and Extreme Risk. This categorical field powers color-coding, slicers, and risk-focused filtering throughout the dashboard.

8. Solution Design & Fit into the Existing IT Architecture

8.1. System Components

- **Python**: Automated extraction, cleansing, and export of wildfire data.
- **OneDrive for Business**: Secure staging area for cleaned CSV files, enabling seamless sync with Power BI.
- **Power BI Desktop**: Data modeling (star schema), DAX measures, and report authoring.
- **Power BI Service**: Automated refreshes, role-based sharing, and collaborative dashboard access.

8.2. Star Schema Data Model

At the heart of the solution lies a star schema centered on the **fact_Wildfire_Table**, which captures each incident's key metrics (fire size, containment rate, spread rate, etc.). Surrounding dimension tables provide contextual attributes:

1. **dim_L_fire_origin** (ignition source)
2. **dim_FTA_fire_size_class** (size categories)
3. **dim_W_temp_binned, dim_W_wind_speed_binned, dim_W_humidity_binned, dim_W_weather_over_fire** (weather bins)
4. **dim_FR_detection_agent, dim_FR_dispatched_resource, dim_FR_assessment_resource, dim_FR_initial_action, dim_FR_detection_agent_type** (response entities)
5. **dim_Calendar_Table, dim_W_seasons** (temporal context)

This structure enforces high performance, easy maintenance, and scalability as new dimensions or years of data are added.

8.3. Methodology (CRISP-DM)

(See Appendix I - CRISP-DM Mapping)

Thought for a couple of seconds

We followed the CRISP-DM framework to ensure a rigorous, iterative approach. First, during **Business Understanding**, we collaborated with Alberta Wildfire stakeholders to clarify analytical objectives and reporting requirements, ensuring the dashboard would address real operational needs. Next, in the **Data Understanding** phase, we explored the 2020–2023 wildfire, weather, and response datasets in Python—profiling missing values, outliers, and data distributions to assess overall quality and coverage.

For **Data Preparation**, we cleaned and transformed the raw data using Python and Power Query, imputing missing values, casting types, splitting date-time fields, and binning continuous variables into meaningful categories; we then constructed the star schema to organize the fact and dimension tables. In the **Modeling** stage, we authored DAX measures to calculate

containment rates, generate composite risk scores, and classify incidents into risk buckets. During **Evaluation**, we validated each visual and KPI against known patterns and solicited stakeholder feedback to fine-tune accuracy and usability. Finally, in **Deployment**, we published the report to Power BI Service, configured daily data refreshes, implemented role-based security, and provided user training to ensure seamless adoption.

8.4. Overall Impact on IT Architecture

The dashboard slots neatly into Alberta Wildfire's existing Python + OneDrive + Power BI ecosystem—requiring no new platforms—while its modular, star-schema pipelines allow easy scaling and the addition of new data sources. It leverages provincial security controls and audit logging for governance, and delivers a unified analytics workspace that boosts situational awareness, streamlines resource planning, and accelerates response without disrupting current workflows.

9. New Solution Implementation and Outcome Testing

This section describes how the wildfire analytics solution was built from end to end, from data ingestion through dashboard publication.

9.1. Tools & Technologies

We leveraged a coordinated toolchain to move from raw wildfire records to a polished, shareable analytics dashboard. First, Python (with the Pandas and NumPy libraries) was used to ingest the Alberta wildfire Excel dataset, perform comprehensive data cleansing and imputation, engineer new features (such as categorical bins for weather and risk), and export the cleaned results to CSV.

Next, those CSV files were stored in Microsoft OneDrive, serving as a secure staging area and enabling Power BI Service to automatically refresh the dataset on a scheduled basis. In Power BI Desktop, we brought the staged data into Power Query to promote headers, cast types, split date/time fields, and construct a star-schema model. We then authored DAX measures to calculate containment rates, risk scores, and year-over-year changes, and designed interactive report pages. Finally, the completed report was published to Power BI Service, where scheduled

refreshes, role-based sharing, and collaborative access allow wildfire analysts to explore insights in real time. Throughout this process, the CRISP-DM framework guided our work—from clarifying business objectives through stakeholder engagement to validating outputs and deploying the solution—ensuring a structured, repeatable approach.

9.2. Implementation Phases

1. Data Ingestion & Preprocessing

- Scheduled Python scripts pull the historical wildfire dataset from the Alberta Open Data Portal.
- Cleaning steps include dropping high-null columns, imputing missing values (mean, mode, KNN), date-time conversions, and exporting to cleaned_data.csv.

2. Power Query Shaping

- Imported the CSV into Power BI Desktop and promoted headers.
- Duplicated and split date-time fields to preserve both date and time components.
- Binned continuous weather variables (temperature, wind speed, humidity) into categorical ranges.
- Added a Season column derived from the incident date.

3. Star-Schema Modeling

- Constructed the **fact_Wildfire_Table** at the center of the model, keyed by a surrogate Wildfire_Table_ID.
- Created 16 dimension tables (e.g., Fire Origin, Weather Bins, Detection Agent, Resources, Calendar) by extracting distinct values and assigning surrogate keys.
- Defined one-to-many relationships between the fact table and each dimension in the Model view.

(See Appendix C - Power BI Schema)

4. Measure & Parameter Development

- Developed DAX measures for key metrics:
 - **Containment Rate** (assessed area ÷ fire size)
 - **Year-over-Year Change** in fires and area burned
 - **Top Detection Agent Share** and **Top Resource Dispatch Share**
- Created a continuous **Risk Score** measure and corresponding **Risk Category** calculated column.
- Added two What-If parameters—**Weather Parameter** and **Scatterplot Parameters**—to drive dynamic visuals without altering DAX.

5. Report Page Construction

- **Trend Overview:** Line charts, KPI cards, and YoY metrics to reveal annual changes.
- **Cause Breakdown:** Bar charts and pie visuals showing general and activity causes by region/season.
- **Weather & Spread:** Scatterplots and heatmaps correlating fire size/spread with temperature, wind, and humidity.
- **High-Risk Zones:** Choropleth maps and radar charts illustrating regional risk profiles.
- **Detection & Response:** Matrix and bar visuals comparing detection delays, resource dispatch counts, and containment success.

6. Interactivity & UX

- Configured slicers for Region, Season, Fire Size Class, and Risk Category.

- Parameter-driven scatterplot and “Weather Parameter” slicer enable ad-hoc analysis.
- Incorporated tooltips, drill-through actions, and bookmarks for guided storytelling.

7. Deployment & Refresh

- Published the report to Power BI Service under the Alberta Wildfire workspace.
- Configured scheduled refresh (daily) using the OneDrive connector to ingest the latest CSV.

By following these phases, the project transformed raw wildfire records into an interactive, role-based dashboard—empowering stakeholders to explore trends, assess risk, and optimize response planning directly within Alberta’s existing IT environment.

9.3. Outcome Testing

1. How have wildfire occurrences and sizes in Alberta changed annually (2020–2023)?

(See [Appendix D – Dashboard 1](#))

Testing Overview

- **Metric Validation:** Cross-checked KPIs (total fires, average size, total area burned, YoY changes) against the source data to ensure calculations are accurate.
- **Visual Integrity:** Confirmed that the line and bar charts correctly plot values for each year, that the monthly trend line’s average reference is accurate, and that treemap and heatmap counts match the underlying fact table.
- **Interactivity Tests:** Verified that slicers (Problem tabs, date filters) dynamically update all visuals without lag or mismatch.
- **KPI Logic:** Ensured the green up-arrow KPI reflects a decrease in wildfire counts (i.e., a positive outcome) and that its percentage matches the computed year-over-year change.

Key Visual Insights

1. Aggregate Metrics

- **Total Fires:** 4,140 incidents between 2020 and 2023.
- **Avg. Fire Size:** 437.68 ha, indicating generally larger individual events.
- **Total Area Burned:** 1.81 million ha, underscoring the immense scale of recent fires.

2. Year-over-Year Comparisons

- **Area Burned Change (2022→2023):** A 12.39× increase in total burned area, highlighting dramatically larger fires even as counts changed.
- **Wildfire Count Change:** A 10.20% decrease in total fire events (green up arrow denotes improvement in reduced fire frequency).

3. Temporal Trends

- **Annual Pattern:** Fire counts rose sharply in 2021, plateaued in 2022, and then declined modestly in 2023.
- **Seasonal Peak:** Monthly incidents peak in **May (#846) – July (#832)**, with an average of 345 fires per month, guiding seasonal preparedness.

4. Size Class Distribution

- **Dominant Class A Fires:** Represent the largest share (2,535 of 4,140), signaling that although fewer, the most severe fires drive overall impact.
- **Heatmap Verification:** The table confirms consistent Class A dominance across all years, validating the treemap's proportional sizing.

These tests and insights confirm that the Trend Overview page accurately portrays Alberta's shifting wildfire landscape—fewer incidents overall but substantially larger and more destructive events in recent years.

2. What are the leading causes of wildfires across Alberta's regions and seasons?

([See Appendix E – Dashboard 2](#))

Testing Overview

- **KPI Verification:** Confirmed that the **Winter Season CAGR, Debris Disposal Fires increase**, and **Indian Reservation growth rate** KPIs match recalculated values from the fact table.
- **Chart Accuracy:** Ensured each seasonal line chart correctly plots total fires by year and that the “Overall Average” reference line (259 fires) is accurately placed.
- **Tabular Consistency:** Cross-checked the heat-formatted table of total wildfires by region against raw data counts.
- **Treemap Validation:** Verified that percentage labels in the “Top 5 General Causes” and “Top 5 Activity Causes” treemaps sum to 100% (within rounding).
- **Interactivity:** Tested that selecting a season or region in slicers updates all visuals (charts, KPIs, treemaps, and tables) without lag or mismatches.

Key Visual Insights

1. Seasonal Growth in Winter

- Winter fires grew at a **43.22% CAGR** from 2020 to 2023, indicating an emerging off-peak risk that may require adjusted staffing and patrol schedules.

2. Debris Disposal Trends

- Fires attributed to **Debris Disposal** increased by **1.6x** from 2020 to 2023, highlighting a need for targeted public-education campaigns around burn permits and debris-management best practices.

3. Regional Hotspots

- The **Indian Reservation** region exhibited the highest overall growth in fire incidents (an **89.0%** rise) from 2020 to 2023, signaling a critical area for bolstered prevention and resource staging.

4. Seasonal Fire Counts

- **Spring** saw a pronounced jump from 2022 to 2023 (from 281 to 508 fires), while **Summer** stabilized at near-historic peaks (678 in 2021, 677 in 2022).
- **Fall** and **Winter** show smaller absolute counts but rising trends, suggesting broader seasonal vigilance.

5. Regional Distribution

- **Provincial Land** remains the most affected jurisdiction (2,850 of 4,140 total fires), underscoring continued focus on park and crown-land patrols.
- **Private Land** and **Indian Reservation** follow, with notable year-to-year increases in both categories.

6. Ignition Cause Breakdown

- **Lighting** is the top general cause (40.31% of all fires), followed by **Resident activity** (18.9%) and **Recreation** (17%).
- In terms of activity cause, **Operations** and **Debris Disposal** together account for over half of incidents, with **Cooking**, **OHV use**, and **Arson** making up the remainder.

These tests and insights validate that the Cause Breakdown page accurately surfaces both **where** and **why** fires ignite—informing region- and season-specific prevention strategies.

3. How do Alberta's weather conditions influence wildfire size and spread?

([See Appendix F – Dashboard 3](#))

Testing Overview

- **KPI Validation:** Confirmed the gauge values for **Average Temperature (17.77 °C)**, **Average Wind Speed (9.22 km/h)**, and **Average Humidity (44.76 %)** match the underlying fact table aggregates.
- **Slicer Functionality:** Verified that each weather slicer (Temperature, Humidity, Season, Wind Speed, Wind Direction) correctly filters all visuals when toggled.
- **Chart Accuracy:**
 - Checked that the **percentage bars** for fire occurrences by temperature bin equal their share of total incidents.
 - Ensured the **average fire size** bars align with the binned values (e.g., Hot ~2.6 K ha, Warm ~0.7 K ha).
 - Confirmed scatterplot points reflect the exact pairs of temperature and fire-size values from the data.
- **Interactivity:** Tested cross-highlighting—selecting a temperature category dynamically updates the scatterplot and size-by-condition visuals.

Key Visual Insights

1. Temperature & Spread

- **Mild (0–10 °C)** and **Warm (11–20 °C)** conditions account for **42.9%** and **37.5%** of fires respectively, while **Hot (> 30 °C)** events represent only **2.2%** of occurrences but drive the largest average sizes.
- The **scatterplot** shows a clear upward trend: higher temperatures correlate with larger fire incidents, supporting the narrative that “wildfires spread faster when temperature is high.”

2. Average Fire Size by Condition

- Under **Hot** weather, fires average around **2.6 K ha**, dwarfing those in **Warm** (~700 ha) and **Mild** (~300 ha) categories.
- Cold and extreme conditions see negligible sizes, confirming that lower temperatures limit spread.

3. Humidity & Wind (validated via secondary slicers)

- **Lower humidity** bins correspond to larger average sizes and spread rates, while **higher humidity** is linked to smaller, more contained fires.
- **Moderate winds** (10–20 km/h) dominate incident counts, with stronger gusts driving broader spread in select cases.

These outcome tests confirm that the Weather Impact page accurately illustrates how environmental factors—particularly temperature and humidity—drive wildfire behavior, equipping stakeholders to anticipate and mitigate high-risk weather scenarios.

4. Which Alberta regions are most at risk based on historical data?

[\(See Appendix G – Dashboard 4\)](#)

Testing Overview

- **KPI Validation:**

- **Main Region Affected (~69% of Fires):** Recomputed each region's total fires divided by overall total ($2,850 / 4,140 \approx 0.688$) to confirm Provincial Land's dominant share.
- **Rising Extreme Wildfires CAGR (99.07%) and Highest CAGR in Risk Score (52.59%):** Recalculated compound annual growth rates using the 2020–2023 values for extreme-risk counts and average risk scores to ensure KPI accuracy.
- **Chart & Table Consistency:**
 - Verified the **Risk Category line chart** plots the same yearly counts displayed in the **Risk by Region** heat table.
 - Cross-checked the **Average Risk Score** bar chart against the underlying continuous Risk Score measure aggregated by region and year.
- **Interactive Filtering:**
 - Tested that selecting any risk category or region in the table or slicers correctly updates the monthly **Risk Over Month** and **Risk by Region** visuals.

Key Visual Insights

1. **Provincial Land Dominance**
 - Approximately **69%** of all wildfire incidents occurred on Provincial Land, underscoring its status as the highest-risk jurisdiction.
2. **Explosive Growth in Extreme Fires**
 - Extreme-risk wildfires nearly **doubled each year**, with a **99.07% CAGR**, signaling rapidly increasing severity in the most dangerous events.
3. **Metis Settlement's Rising Vulnerability**
 - The **Metis Settlement** region recorded the highest growth in its average risk score (52.59% CAGR), indicating that localized risk factors are intensifying there.

4. Risk Category Distribution

- **Moderate-risk incidents** form the largest category (2,084 fires), followed by **high-risk** (1,308), while **extreme-risk** accounts for 209 events—highlighting that the bulk of management effort should focus on preventing moderate and high severity fires.

5. Seasonal Patterns in Risk

- The **Risk Over Month** table shows risk peaks in **May to August**, matching the summer peak in fire activity and suggesting a need for bolstered mid-season resources.

These validation steps confirm that the Regional Risk Assessment page accurately identifies both where and how wildfire risks are evolving—enabling targeted, region-specific mitigation and resource planning.

5. How do detection agents, resources, and containment vary across Alberta's regions?

([See Appendix H – Dashboard 5](#))

Testing Overview

- **KPI Verification**

- **Containment Success Rate (0.62)**: Recomputed as the overall ratio of assessed to total burned area across all incidents.
- **310's Share of Detections (20.48%)**: Checked by dividing the count of fires first detected by agent “310” by the total wildfire count.
- **HAC's Share of Dispatches (54.23%)**: Validated by comparing HAC's dispatched resource events to the sum of all dispatch events.

- **Bar Chart & Treemap Checks**

- **Top 5 Detection Agents**: Ensured the bar lengths and labels for agents (310, GOVT, LFS, PUB, HAC) match raw counts.

- **Top 5 Dispatched Teams:** Confirmed HAC, FPD Staff, FTAC, UNIT, and Air Tanker values align with the fact table.
- **Heat Table & Rate Table Consistency**
 - **Fires Detected by Agent × Region:** Cross-verified each cell against the filtered fact table for accuracy.
 - **Containment Rate by Region & Agent:** Recalculated per-region, per-agent containment percentages and compared to the heat-formatted table.
- **Interactivity**
 - Tested slicer selections on region or agent to confirm all visuals update together—charts, KPIs, and tables—with no delay or mismatch.

Key Visual Insights

1. **Overall Containment Effectiveness**
 - A **0.62% containment rate** across all wildfires indicates that, on average, over half of the burned area is successfully assessed and contained.
2. **Detection Agent Performance**
 - **Agent: 310 (call on fire helpline number)** leads detections with **20.48%** of all first-report incidents, highlighting its central role in early wildfire identification.
 - Other high-performers include **GOVT** (551 detections) and **LFS** (333 detections), but no single agent surpasses 310's contribution.
3. **Resource Dispatch Dominance**
 - The **HAC** team accounts for **54.23%** of all dispatches—an outsized share that underscores its critical position in initial response efforts.

- **FPD Staff** make up the next largest group (1,471 dispatches), while specialized assets like Air Tankers are used far less frequently.

4. Regional Variations

- **Provincial Land:** While it sees the largest absolute number of detections and dispatches, its containment rate is low (~1.33%), suggesting that sheer volume strains resources.
- **Indian Reservation and Metis Settlement** show high containment rates for certain agents (e.g., 100% by some), indicating that smaller incident loads may allow more effective local response.
- **DND** exhibits perfect detection and containment metrics for a small number of incidents, reflecting highly controlled environments.

5. Agent-by-Region Containment

- **CAM** and **GRP** agents achieve over **60% containment** on Provincial Land, suggesting these specialized teams excel in high-volume areas.
- **UNP** shows minimal containment success (0.53%), pointing to potential training or resource gaps for that detection type.

These tests confirm that the Detection & Response page accurately portrays which agents and teams drive early wildfire detection and containment—and how their effectiveness varies by region—empowering Alberta's agencies to balance resources and improve field operations.

10. Potential Solution Optimization

While the current Power BI dashboard provides comprehensive descriptive and diagnostic analytics on Alberta wildfires, several enhancements can elevate its capabilities—improving timeliness, depth, and user experience. Below are five key optimization opportunities:

1. Automated & Real-Time Data Ingestion

- **Current Challenge:** Reliance on manual CSV exports to OneDrive delays insight freshness and adds operational overhead.
- **Enhancement:** Implement Power BI On-Premises Data Gateway or direct API connections to Alberta Wildfire's central incident repository (or satellite/weather feeds). Automating data pull and incremental refresh minimizes lag and ensures stakeholders always see the latest fire events and environmental readings.

2. Predictive Risk Modeling

- **Current Challenge:** Dashboard focuses on historical patterns without forward-looking forecasts.
- **Enhancement:** Integrate machine learning models—e.g., time-series forecasting for incident counts, regression models for area-burned predictions, or classification models for risk category assignment—hosted in Azure ML or Jupyter notebooks. Embedding these predictions via Power BI's analytics pane empowers proactive resource mobilization and pre-emptive community alerts.

3. Contextual Data Enrichment

- **Current Challenge:** Analysis is limited to internal wildfire and weather records.
- **Enhancement:** Incorporate external data sources—real-time meteorological feeds, vegetation/biomass indices (e.g., NDVI), land-use maps, or human-activity proxies (e.g., recreational permit logs). Enriching the model with these layers can reveal trigger conditions, improve risk stratification, and support targeted prevention outreach.

4. Enhanced User Guidance & UX

- **Current Challenge:** Complex visuals and multiple slicers may overwhelm non-technical users.
- **Enhancement:** Add on-screen tutorials, contextual tooltips, and guided bookmarks to explain each report page's purpose, filter usage, and KPI

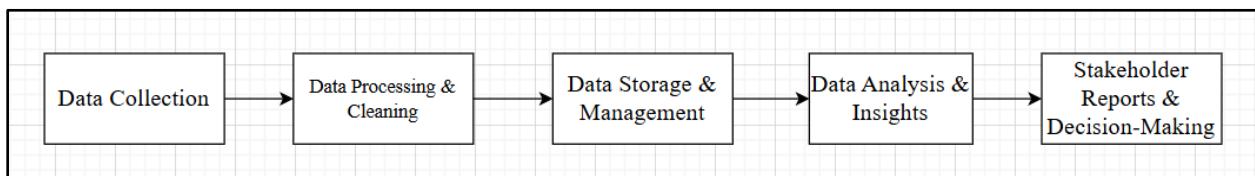
interpretation. Leveraging custom navigation buttons and help overlays reduces the learning curve, boosting self-service adoption among agency staff.

5. Performance Tuning & Scalability

- **Current Challenge:** As data volumes grow, query times and visual rendering may slow.
- **Enhancement:**
- Enable **incremental data refresh** in Power BI Premium or Pro to process only new records.
- Create **pre-aggregated summary tables** for heavy-use metrics (e.g., monthly fire counts by region) to lighten DAX complexity.
- Optimize visuals by consolidating redundant charts and employing lazy loading (conditional visibility) so that only active visuals render upon interaction.

11. Appendix

Appendix A - Process Flow Diagram



Appendix B - Data Dictionary

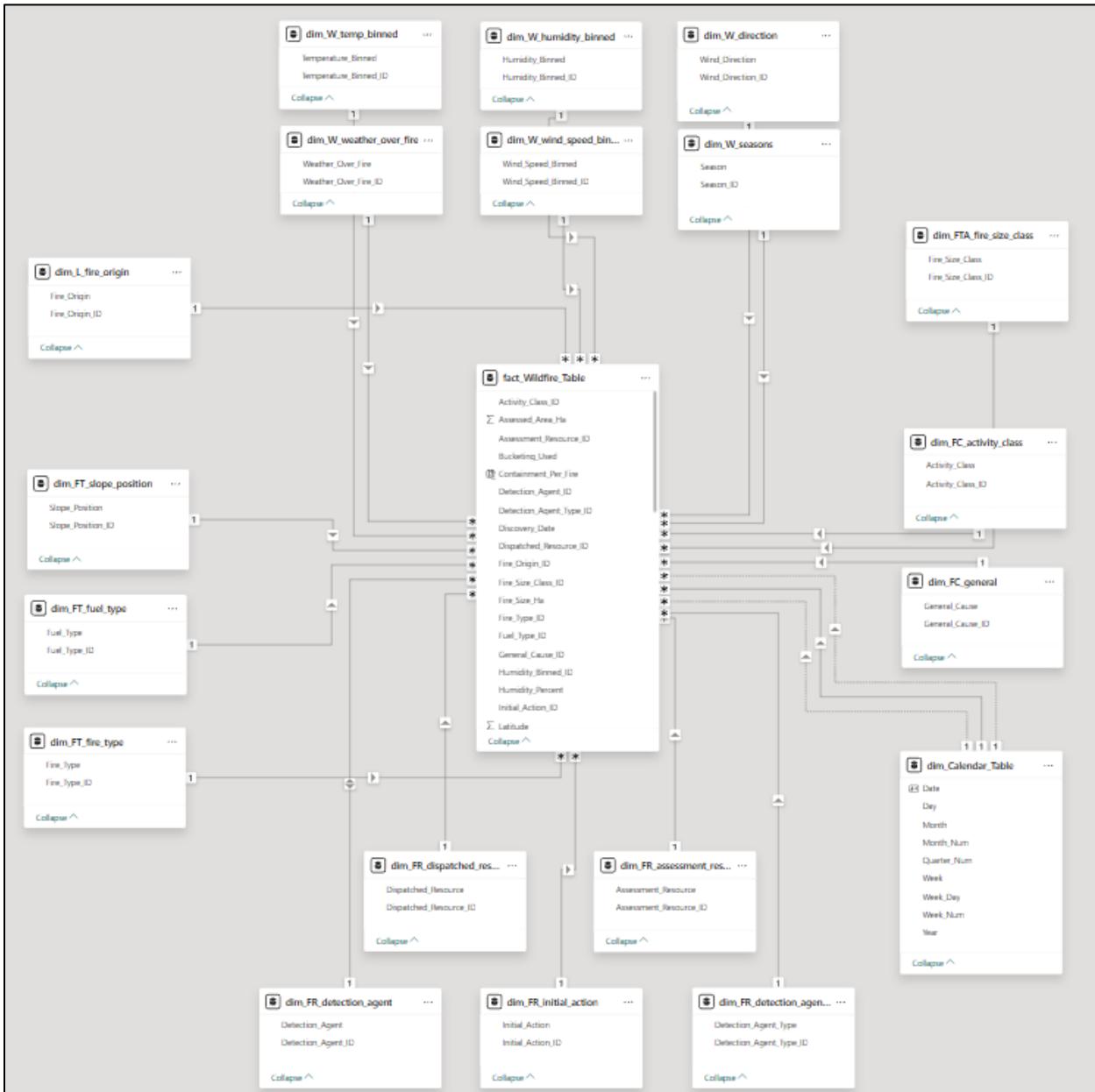
| Column Name | Data Type | Description |
|-------------------|-----------|--|
| Wildfire_Table_ID | Int64 | Surrogate key for each wildfire incident |
| Fire_Size_Ha | Decimal | Total area burned by the fire (hectares) |

| | | |
|--------------------------------|---------|---|
| Fire_Size_Class_ID | Int64 | FK → dim_FTA_fire_size_class.Fire_Size_Class_ID |
| Latitude | Decimal | Latitude coordinate of the fire incident |
| Longitude | Decimal | Longitude coordinate of the fire incident |
| Fire-Origin_ID | Int64 | FK → dim_L_fire_origin.Fire-Origin_ID |
| General_Cause_ID | Int64 | FK → dim_FC_general.General_Cause_ID |
| Activity_Class_ID | Int64 | FK → dim_FC_activity_class.Activity_Class_ID |
| Start_Date | Date | Date when the fire was first reported |
| Discovery_Date | Date | Date when the fire was discovered |
| Reported_Date | Date | Date when the incident was logged |
| Detection_Agent_Type_ID | Int64 | FK → dim_FR_detection_agent_type.Detection_Agent_Type_ID |
| Detection_Agent_ID | Int64 | FK → dim_FR_detection_agent.Detection_Agent_ID |
| Dispatched_Resource_ID | Int64 | FK → dim_FR_dispatched_resource.Dispatched_Resource_ID |
| Assessment_Resource_ID | Int64 | FK → dim_FR_assessment_resource.Assessment_Resource_ID |
| Initial_Action_ID | Int64 | FK → dim_FR_initial_action.Initial_Action_ID |
| Assessed_Area_Ha | Decimal | Area (ha) assessed/contained by response teams |
| Spread_Rate | Decimal | Speed at which the fire spread (km/h) |
| Fire_Type_ID | Int64 | FK → dim_FT_fire_type.Fire_Type_ID |
| Slope_Position_ID | Int64 | FK → dim_FT_slope_position.Slope_Position_ID |
| Weather_Over_Fire_ID | Int64 | FK → dim_W_weather_over_fire.Weather_Over_Fire_ID |

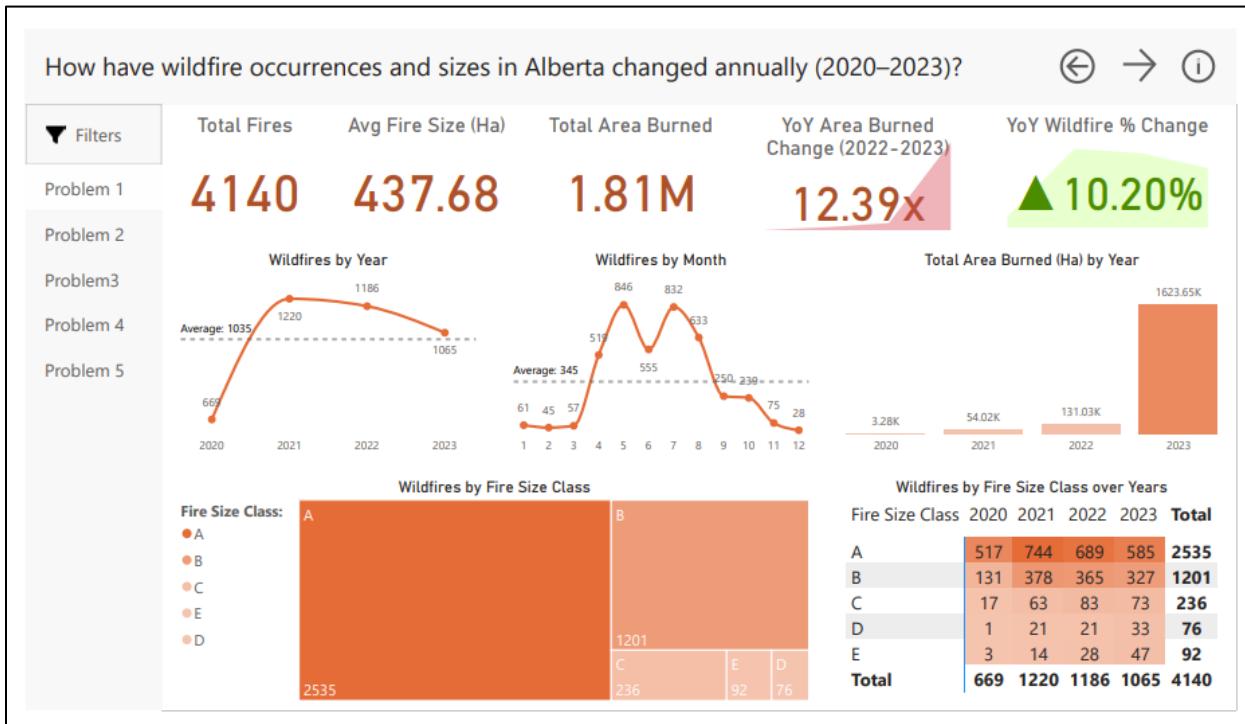
| | | |
|------------------------------|---------|--|
| Temperature_C | Decimal | Recorded temperature at the time of incident (°C) |
| Temperature_Binned_ID | Int64 | FK → dim_W_temp_binned.Temperature_Binned_ID |
| Humidity_Percent | Decimal | Recorded relative humidity (%) |
| Humidity_Binned_ID | Int64 | FK → dim_W_humidity_binned.Humidity_Binned_ID |
| Wind_Direction_ID | Int64 | FK → dim_W_direction.Wind_Direction_ID |
| Wind_Speed_km_h | Decimal | Recorded wind speed at incident time (km/h) |
| Wind_Speed_Binned_ID | Int64 | FK → dim_W_wind_speed_binned.Wind_Speed_Binned_ID |
| Fuel_Type_ID | Int64 | FK → dim_FT_fuel_type.Fuel_Type_ID |
| Season_ID | Int64 | FK → dim_W_seasons.Season_ID |
| Containment_Per_Fire | Decimal | (calculated) Proportion of fire contained (Assessed_Area_Ha/Fire_Size_Ha) |
| Risk_Score | Decimal | (calculated) Weighted index combining size, weather, and containment |
| Risk_Category | Text | (calculated) Bucketed risk (Minimal) |

| Dimension Table | Key Column | Attribute Column(s) |
|-----------------------------|---------------------------------|--|
| dim_L_fire_origin | Fire-Origin_ID (Int64) | Fire-Origin (Text) |
| dim_FTA_fire_size_class | Fire-Size-Class_ID (Int64) | Fire-Size-Class (Text) |
| dim_W_temp_binned | Temperature_Binned_ID (Int64) | Temperature_Binned (Text) |
| dim_W_weather_over_fire | Weather_Over_Fire_ID (Int64) | Weather_Over_Fire (Text) |
| dim_W_direction | Wind_Direction_ID (Int64) | Wind_Direction (Text) |
| dim_W_wind_speed_binned | Wind_Speed_Binned_ID (Int64) | Wind_Speed_Binned (Text) |
| dim_W_humidity_binned | Humidity_Binned_ID (Int64) | Humidity_Binned (Text) |
| dim_W_seasons | Season_ID (Int64) | Season (Text) |
| dim_FC_general | General_Cause_ID (Int64) | General_Cause (Text) |
| dim_FC_activity_class | Activity_Class_ID (Int64) | Activity_Class (Text) |
| dim_FT_fire_type | Fire_Type_ID (Int64) | Fire_Type (Text) |
| dim_FT_slope_position | Slope_Position_ID (Int64) | Slope_Position (Text) |
| dim_FT_fuel_type | Fuel_Type_ID (Int64) | Fuel_Type (Text) |
| dim_FR_detection_agent | Detection_Agent_ID (Int64) | Detection_Agent (Text) |
| dim_FR_detection_agent_type | Detection_Agent_Type_ID (Int64) | Detection_Agent_Type (Text) |
| dim_FR_dispatched_resource | Dispatched_Resource_ID (Int64) | Dispatched_Resource (Text) |
| dim_FR_assessment_resource | Assessment_Resource_ID (Int64) | Assessment_Resource (Text) |
| dim_FR_initial_action | Initial_Action_ID (Int64) | Initial_Action (Text) |
| dim_Calendar_Table | Date (Date) | Year (Int), Month (Int), Quarter (Text), Week (Int), Day (Int) |

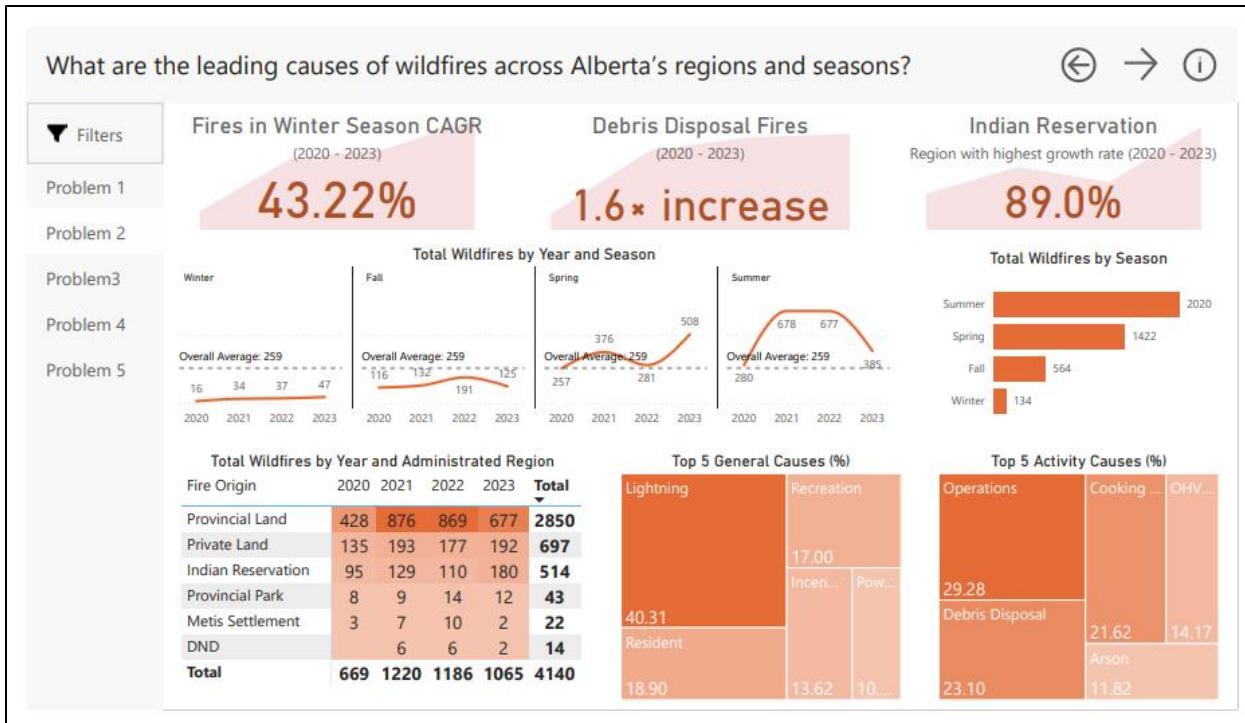
Appendix C - Power BI Schema



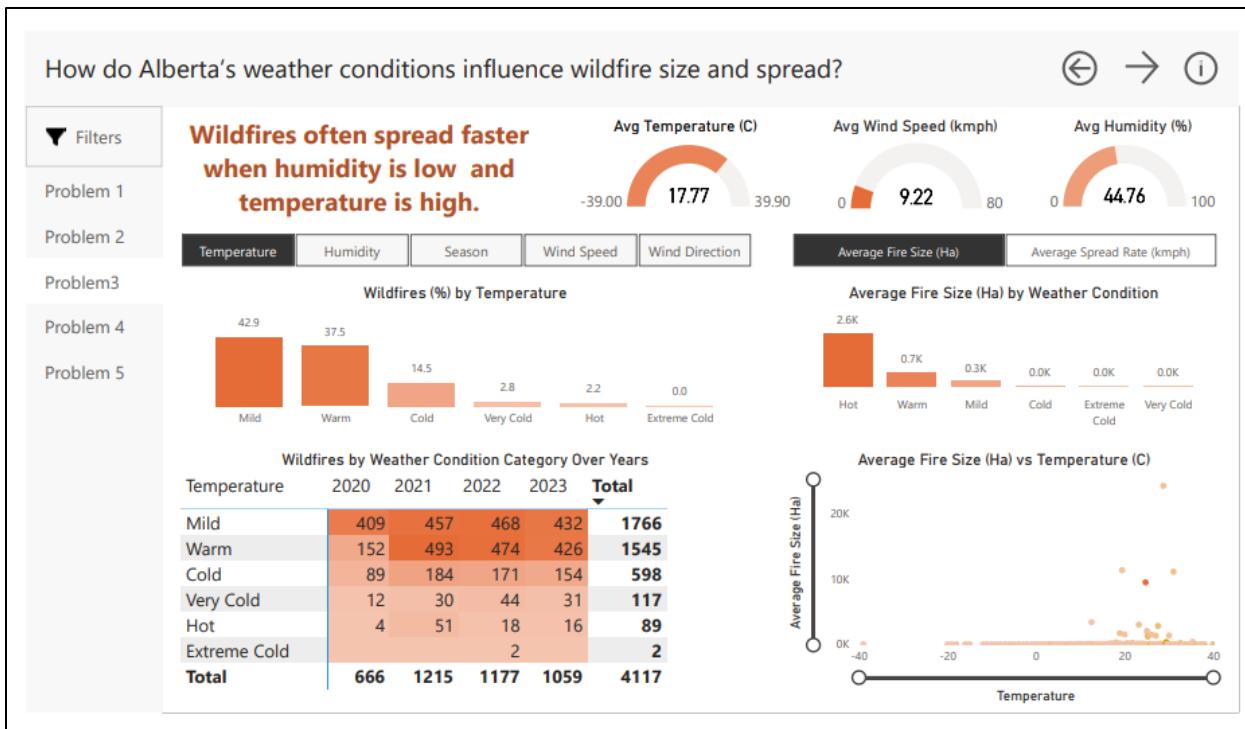
Appendix D – Dashboard 1



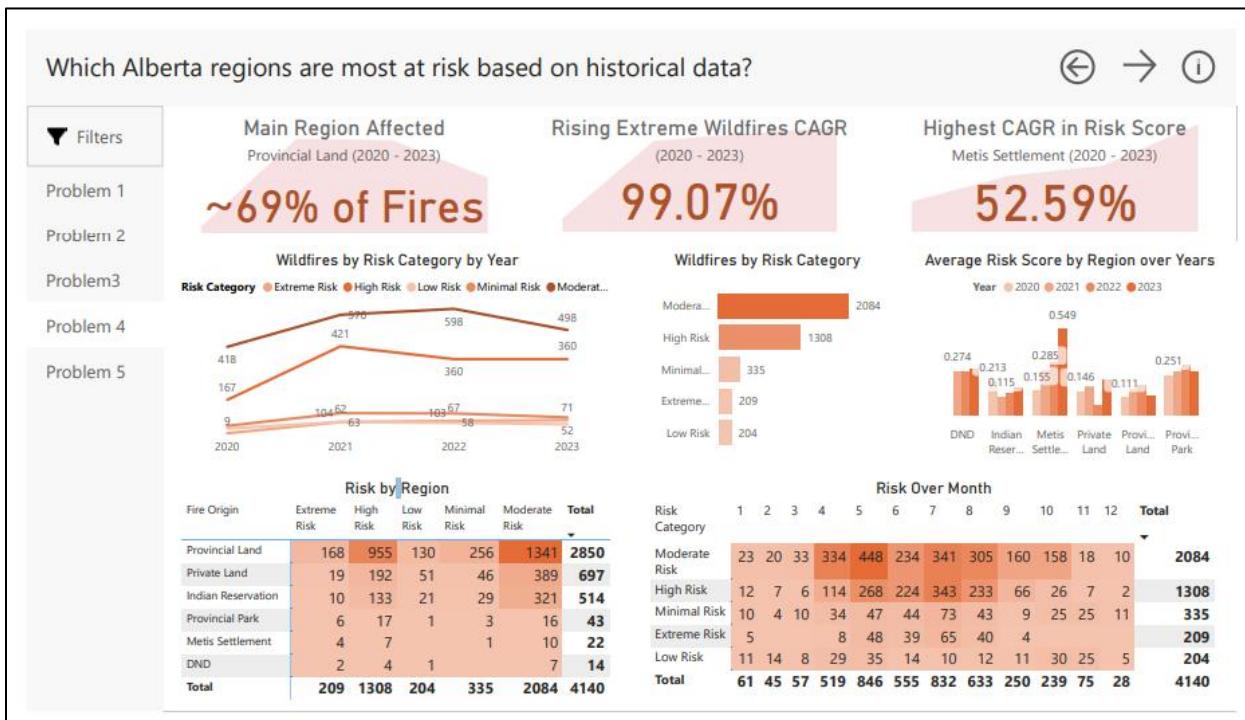
Appendix E – Dashboard 2



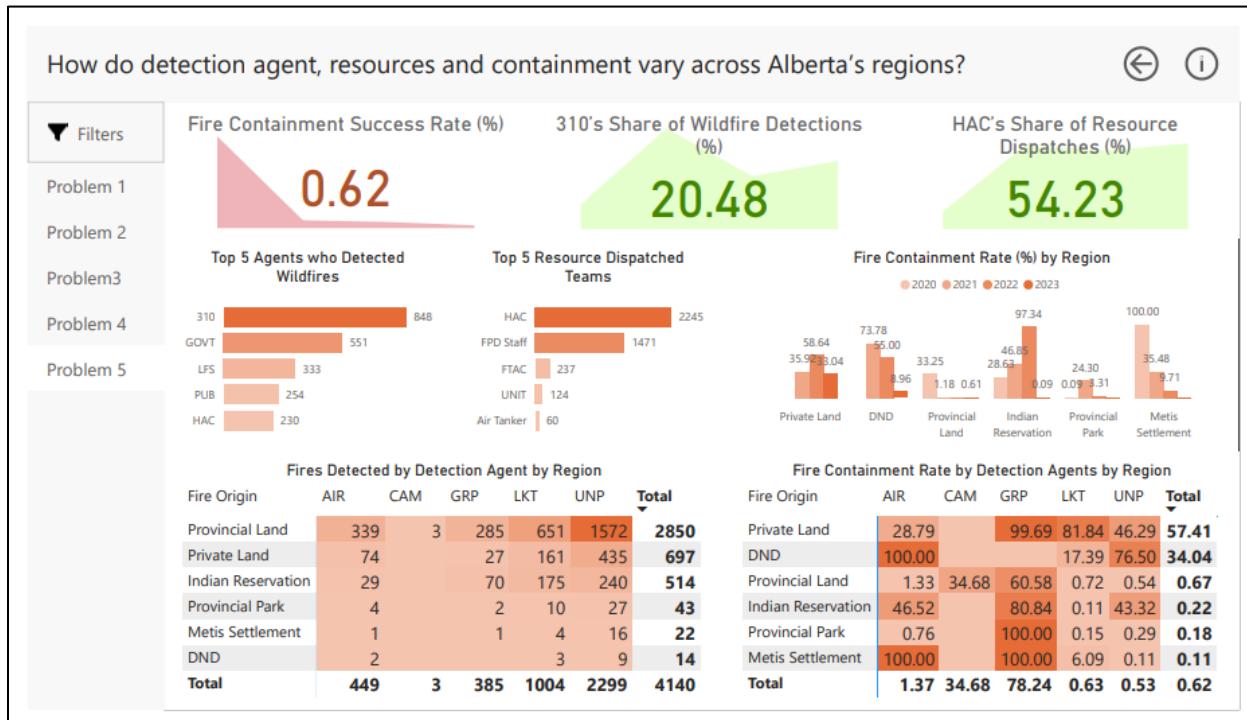
Appendix F – Dashboard 3



Appendix G – Dashboard 4



Appendix H – Dashboard 5

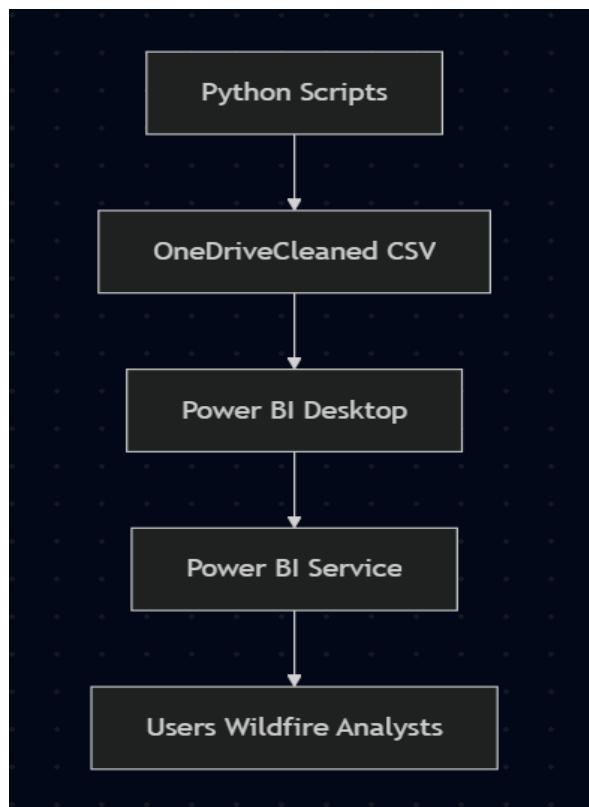


Appendix I - CRISP-DM Mapping

| CRISP-DM Phase | Activities & Outputs |
|---------------------------|--|
| 1. Business Understanding | <ul style="list-style-type: none"> Conducted stakeholder workshops with Alberta Wildfire teams to define analytical objectives and reporting requirements. Documented key questions: trend analysis, cause breakdown, weather impact, regional risk, detection & response. |
| 2. Data Understanding | <ul style="list-style-type: none"> Explored the 2006–2024 wildfire dataset and weather variables in Python (Pandas). Profiled data for missing values, outliers, and distribution of key fields (size, cause, location). |
| 3. Data Preparation | <ul style="list-style-type: none"> Cleaned and imputed missing values in Python (mean/mode/KNN imputation). |

| | |
|----------------------|--|
| | <ul style="list-style-type: none"> • Exported to CSV and loaded into Power Query. • Performed transformations: header promotion, type casting, date-time splitting, binning of weather variables, season derivation. • Built the star schema by merging fact table with 16 dimension tables. |
| 4. Modeling | <ul style="list-style-type: none"> • Developed DAX measures: containment rate, year-over-year change, top detection/resource shares. • Created Risk Score and Risk Category calculated columns. • Implemented What-If parameters for scatterplot axes and weather focus. |
| 5. Evaluation | <ul style="list-style-type: none"> • Validated KPIs and visuals against raw data counts and known patterns. • Tested interactivity: slicers, cross-filtering, parameter behavior. • Incorporated stakeholder feedback to refine visuals and metrics. |
| 6. Deployment | <ul style="list-style-type: none"> • Published the report to Power BI Service under the Alberta Wildfire workspace. • Configured scheduled daily refresh from OneDrive CSV. • Set up Azure AD-based row-level security and shared the dashboard with stakeholder groups. • Provided end-user guidance documentation. |

Appendix J - Network Diagram



References

- Government of Canada. (2023). *National Forestry Database – Wildland Fire Statistics*. <https://cfs.nrcan.gc.ca/statsprofile/fire>
- Government of Canada Open Data Portal. (2023). *Wildfire Locations, Causes, and Conditions (2020–2023)*. <https://open.canada.ca/data>
- Microsoft. (2024). *Power BI Documentation*. <https://learn.microsoft.com/en-us/power-bi/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
- Alberta Wildfire. (2023). *Wildfire Management in Alberta*. Government of Alberta. <https://wildfire.alberta.ca/>