WILDFIRES IN ALBERTA:

A DATA-DRIVEN ANALYSIS

- Saubhagya Pandey (230939)

INTRODUCTION:

Alberta grapples with the ever-present threat of wildfires. These unpredictable blazes leave a trail of devastation, impacting landscapes, communities, wildlife, and the economy.

Accurate wildfire prediction is crucial. Early knowledge of potential fire size allows the Alberta Fire Department to prioritize resources, initiate evacuations, and implement preventative measures.

To address this critical need, our project, accepted by the Alberta Fire Department, leverages AI and ML to develop wildfire size class prediction models for different Alberta regions. This report utilizes initial fire condition data to build these models.

The following sections explore key questions:

- **Vulnerability:** Which Forest Service Areas (FSAs) are most susceptible and why?
- **Causes:** What triggers wildfires in these vulnerable regions?
- **Impact:** How will wildfires affect populations, particularly indigenous communities?

We'll then delve into the methodologies used, including data acquisition, vulnerability assessment, and the chosen AI/ML models.

METHODOLOGY:

Collecting and Preprocessing Data:

1. I collected Data from various sources-
   - Fire_data: From the link provided by the ICG (contains information about wildfires)
   - Fsa_data: From statistique Canada (contains location information of FSA regions)
   - Fsa_names: from other online sources (contains information about names of FSA regions)
   - Csd_data: from statistique Canada (contains location information of Census SubDivisions)
   - Indi_data: from Canada census website (information about indigenous population in CSDs)
2. Then I dropped the information which was not useful for our purpose and Handled missing values for remaining information.
3. I integrated the FSA regions and CSDs with the main data (fire_data) using the lattitudes and longitudes information using GeoPandas library
4. Then I handled missing values using appropriate methods.

FSA Vulnerability and Reason Assessment:

1. For finding most vulnerable FSA regions, I have used simply the frequency of wildfires in each FSA region.
2. For finding the reasons for Wildfire, again used the frequency of reasons for wildfires.
3. For finding the reasons for Wildfire with Highly Burnt Areas, I filtered the dataset to keep only those rows which contain a fire size which is in top 25%ile of the data and then found the frequency of reasons.

Population Vulnerability Assessment:

1. Used those regions where fire size was large (top 25%ile) and reason for fire was natural (lightning) as well as the fuel type (vegetation) was highly favorable for the wildfire to spread (Matted Grass, Standing Grass, Boreal Spruce, Boreal Mixedwood-Green).
2. Used those CSD regions where the fires were most frequent.

Wildfire Size Class Prediction:

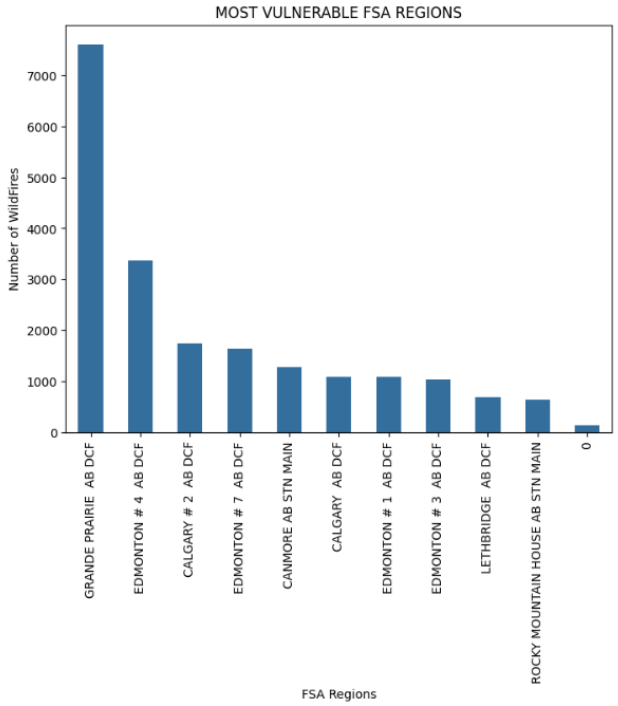
Chosen Model: Random Forest Regression (scikit-learn) for robustness, scalability, and interpretability in wildfire size class prediction.

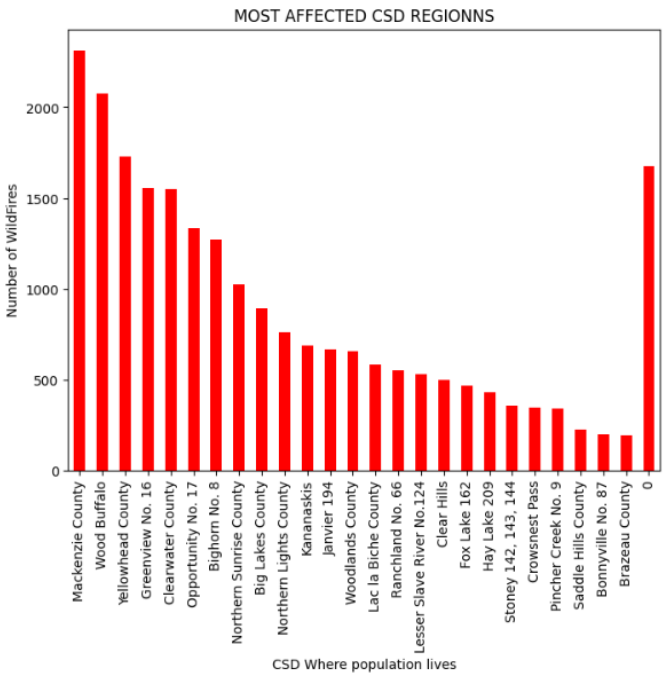
Training & Evaluation:

- **Preprocessing:** Cleaned & engineered data (pandas), scaled data (StandardScaler), encoded categorical features (OneHotEncoder) for model compatibility.
- **Train-Test Split:** Divided data (train_test_split) for model building (training) and generalizability evaluation (testing).
- **Model Training:** Built Random Forest Regression model (RandomForestRegressor) with hyperparameter optimization for best performance.
- **Model Evaluation:** Evaluated model performance on testing set using MSE & R^2 (scikit-learn metrics) to assess prediction accuracy.

EXPERIMENTS AND RESULTS:

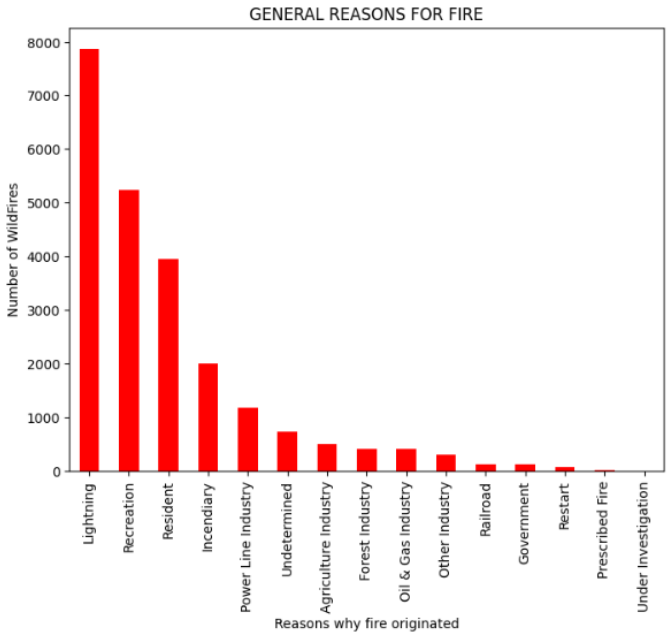
Most Affected FSA and CSD Regions

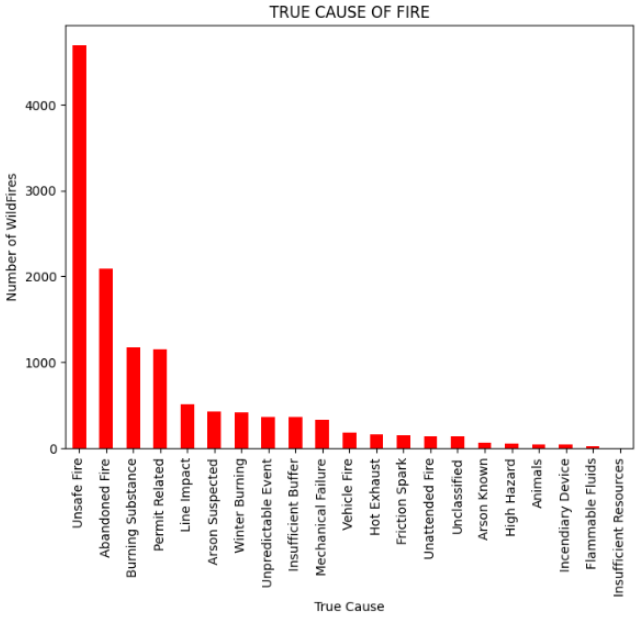




1. Most fires occurred in the FSA region: GRANDE PRAIRIE AB DCF
2. The number of fires in CSD regions decrease almost linearly when we go from most frequent to less frequent ones.

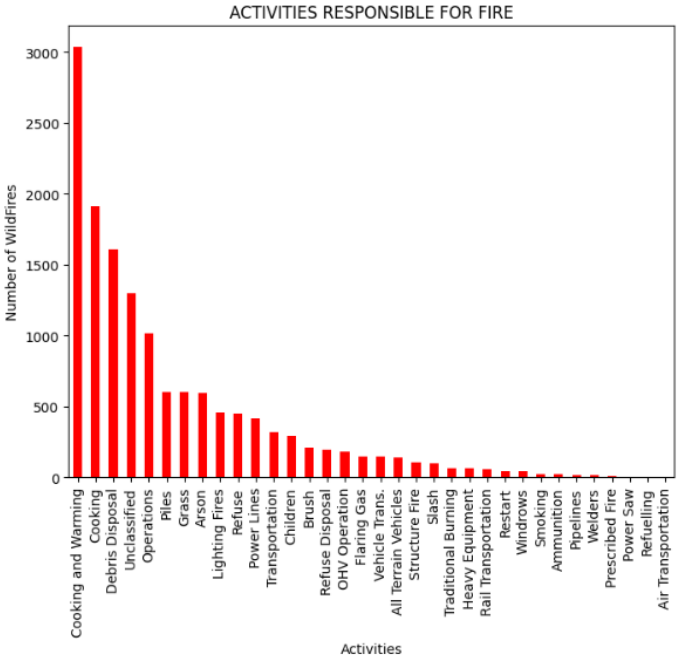
Main Reasons for Fire:

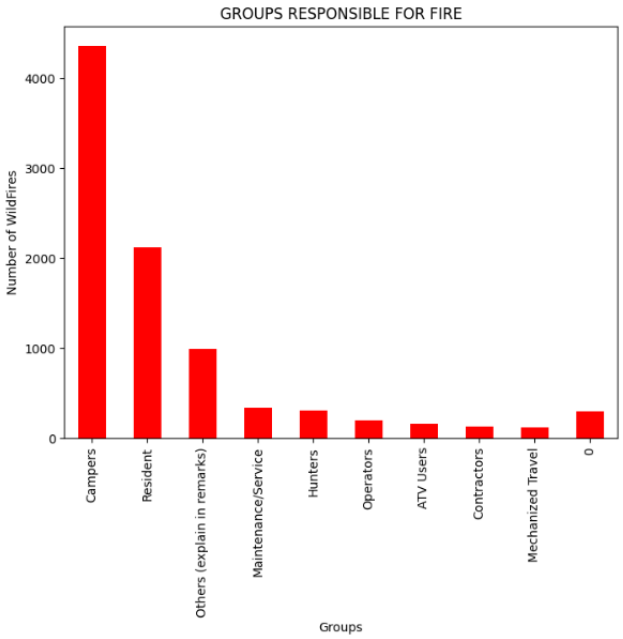




*Main reasons for Fire are basically Lightning , Recreation, and Resident cause unsafe fire.

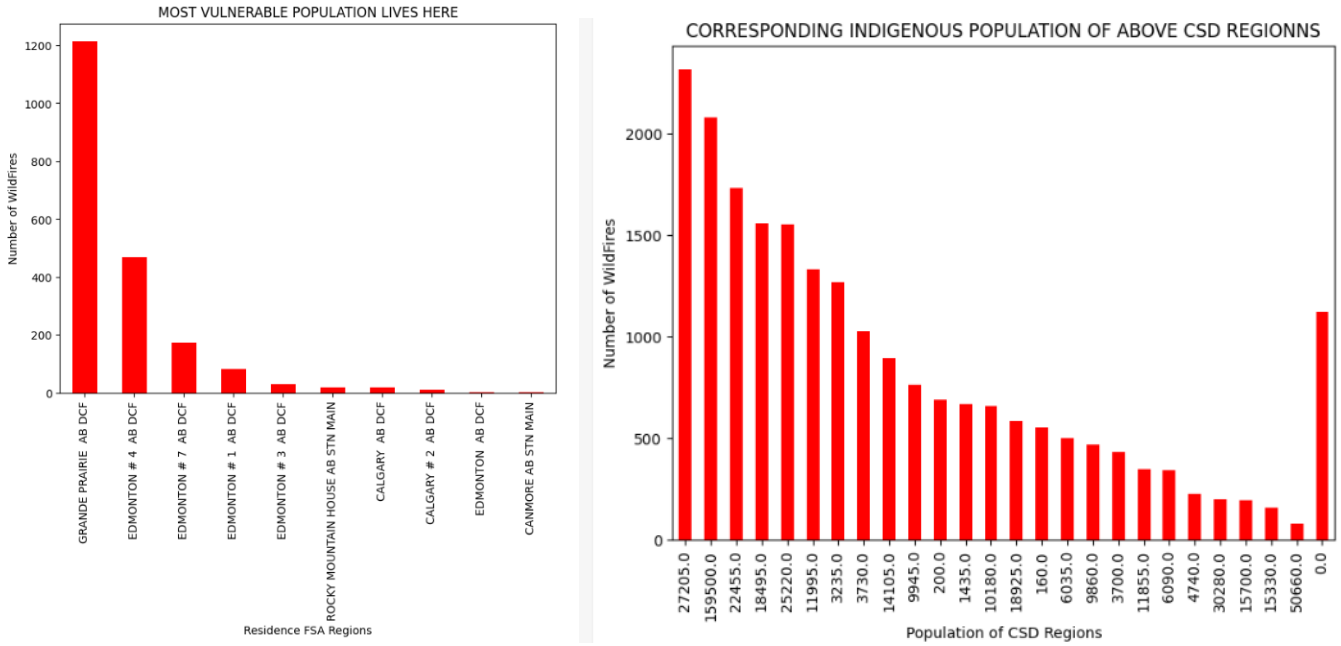
Activities and Groups Responsible for Fire





1. Main groups responsible for the fires are
   - Campers
   - Residents
2. Main activities done by these groups which caused fires are:
   - Cooking
   - Warming
   - Debris disposal

Most Vulnerable Populations to Fire



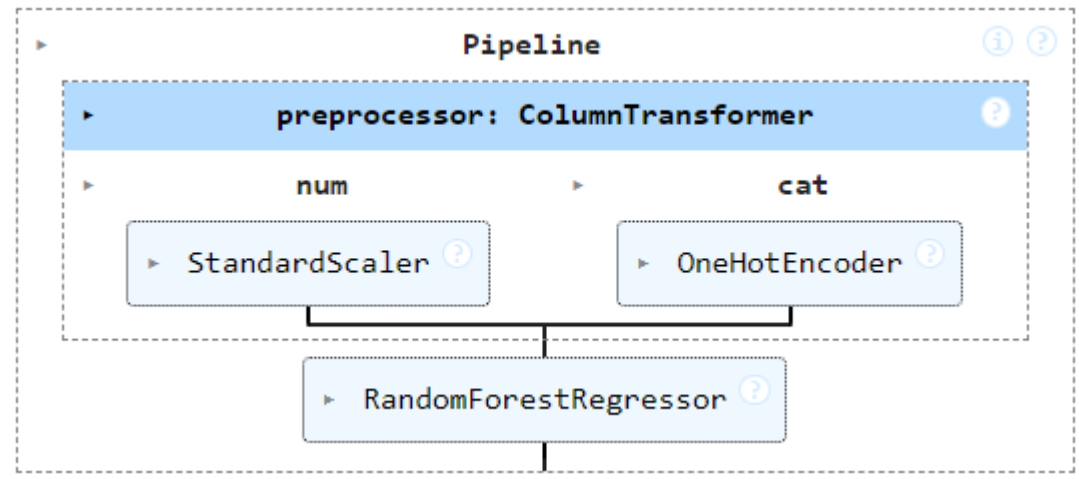
GRAPH 1: Shows those FSA regions where fire size was large (top 25%ile) and reason for fire was natural (lightning) as well as the fuel type (vegetation) was highly favorable for the wildfire to spread (Matted Grass, Standing Grass, Boreal Spruce, Boreal Mixedwood-Green).

GRAPH 2: Shows indigenous populations of those CSD regions where the fires were most frequent.

```
INDIGENOUS POPULATION MOST VULNERABLE TO WILDFIRE =  295885
```

(Sum of Indigenous Populations of Top 10 CSD Regions)

ML Model used for Predicting the Fire Size:



NOTE: Due to limitations of feature engineering and lack of quality of data, as well as due to the fact that the size of wildfires depend on a lot of unpredictable circumstances, the model gives a high error. But this surely can be improved if enough time is provided to me.

```
Mean Squared Error: 21694134.506138425
R-squared: -0.6738449419526171
```

CONCLUSION:

- **Key Findings:**
    - **Most Vulnerable FSAs:** Grande Prairie, Alberta, experienced the highest frequency of fires.
    - **Primary Causes:** Lightning, recreational activities, and unsafe practices by residents (e.g., cooking, warming fires).
    - **Responsible Groups:** Campers and residents were the main groups responsible for starting fires.
    - **Vulnerable Populations:** Indigenous communities in high-frequency fire regions, especially in areas with favourable wildfire spread conditions (e.g., matted grass, boreal spruce).
- **Model Effectiveness:**
    - **Model Used:** Random Forest Regression.
    - **Performance:** Initial models faced challenges but provided a foundation for understanding wildfire prediction complexities.
    - **Metrics:** Model performance indicated limitations, highlighting the need for further refinement.
- **Project Value:**
    - **Strategic Aid:** Supports Alberta Fire Department in wildfire mitigation and preparedness by identifying vulnerable FSAs, primary causes, and at-risk populations.
    - **Resource Allocation:** Facilitates informed decision-making for resource prioritization.
- **Limitations and Future Improvements:**
    - **Data Limitations:** Faced challenges with data availability and quality.
    - **Future Enhancements:**
        - Incorporate real-time weather data and satellite imagery.
        - Improve feature engineering techniques for better model accuracy.
        - Focus on dynamic data sources for more responsive predictive models.

References:

1. Statistics Canada website
2. Sklearn Documentation
3. Canada Census Reports
4. Chat-Gpt and Google Gemini

THANK YOU for Reading!!!

An analysis by Saubhagya Pandey (230939 - IIT KANPUR) 😊