Wildfire Management Alberta

- *Kaixiang Ge*

1 INTRODUCTION

CASE STUDY

- In recent years, Alberta and the broader regions of Canada have experienced increasingly severe wildfires, significantly impacting residents and the environment.
- The increasing frequency and intensity of wildfires in Alberta underscores the urgent need for adaptive strategies and robust emergency responses to mitigate their impact on communities and the environment. By understanding trends and identifying relevant elements, we can develop effective management strategies that may include fire forecasts, preventive measures, and emergency response tools, and provide insights into policy options that can facilitate the implementation of wildfire management and mitigation programs.

Wildfire management in Alberta - Open Government

Alberta wildfire's seasonal statistics - Open Government

MISSION AND STAKEHOLDER

❖ **To enhance wildfire prevention and response strategies by leveraging comprehensive data analysis, enabling the Alberta Wildfire Management to better assess risks, allocate resources efficiently, and protect Alberta's communities and natural resources.**

✓ Alberta Wildfire Management Branch - Primary beneficiary and decision-maker

Improved wildfire prevention and response strategies, effective policy development and resource allocation for wildfire management.

✓ Local Communities - Residents and businesses in wildfire-prone areas

Increased safety and reduced risk of wildfires.

✓ Environmental Organizations

Protection of ecosystems and wildlife from wildfire damage.

2 DATA UNDERSTANDING

DATASET

- The dataset includes **25000+** records of wildfire incidents in Alberta over the past 18 years, offering a detailed and representative sample of the period.
- It comprises comprehensive information on various variables, including fire types, fire origin, size, general cause, specific reasons, activities, reaction timings of each incident, and the initial actions taken.
- The dataset is sourced from Alberta's wildfire management systems and weather monitoring agencies, ensuring a high level of accuracy and comprehensiveness.
- The data is openly accessible from Alberta open data - [Historical wildfire data : 2006 to 2023 - Open Government (alberta.ca)]

BUSINESS QUESTIONS

1. What are the most common ***causes*** of wildfires in Alberta over the 5 years?
2. Which ***areas*** are most susceptible to wildfires and why?
3. What are the ***main causes*** of wildfires in ***different areas***, and do they vary significantly?
4. What are the trends in ***response times*** and their effectiveness in mitigating ***wildfire damage***?
5. How do ***weather conditions, temperature and humidity influence*** the frequency and severity of wildfires?

3 DATA PREPRATION

DATA DICTIONARY

- There is 53 columns in the dataset.

Classification	Field Name	Explanation
Location	fire_location_latitude	Latitude of the ignition area.
Location	fire_location_longitude	Longitude of the ignition area.
Location	fire_position_on_slope	Position of the wildfire relative to the slope.
Basic	fire_origin	Ownership or administration of the land where the wildfire ignited.
Basic	fire_number	Unique identifier for each wildfire.
Basic	fire_name	Geographic name assigned to significant wildfires.
Basic	fire_type	Predominant fire behavior at initial assessment.
Cause	general_cause_desc	Classification of the wildfire cause.
Cause	industry_identifier_desc	Further detail of the industry involved in causing the wildfire.
Cause	responsible_group_desc	Recreational group responsible for causing the wildfire.
Cause	activity_class	Specific activity ongoing when the wildfire started.
Cause	true_cause	Specific reason for the wildfire start.
Weather	weather_conditions_over_fire	Weather conditions over the wildfire at initial assessment.
Weather	Temperature	Temperature at the wildfire site.
Weather	relative_humidity	Relative humidity at the wildfire site.
Weather	wind_direction	Wind direction at the wildfire site.
Weather	wind_speed	Wind speed at the wildfire site.
Time	fire_year	The wildfire year from April 1 to March 31 of the following year.
Time	fire_start_date	Date and time the wildfire started.
Time	discovered_date	Time the wildfire was first discovered.
Time	reported_date	Date and time the wildfire was first reported.
Time	dispatched_date	Date and time the first resource was dispatched.
Time	start_for_fire_date	Time and date the first resource started for the wildfire.
Time	assessment_datetime	Date and time the wildfire was initially assessed.
Time	ia_arrival_at_fire_date	Date and time the initial action resource arrived.
Time	fire_fighting_start_date	Date and time initial suppression began.
Time	first_bucket_drop_date	Date and time of the first bucket drop.

Classification	Field Name	Explanation
Time	bh_fs_date	Date and time the wildfire was determined to be at 'Being Held' status.
Time	uc_fs_date	Date and time the wildfire was determined to be 'Under Control'.
Time	to_fs_date	Date and time the wildfire was turned over to another agency.
Time	ex_fs_date	Date and time the wildfire was extinguished.
Time	uc_fs_date	Date and time the wildfire was determined to be 'Under Control'.
Time	to_fs_date	Date and time the wildfire was turned over to another agency.
Time	ex_fs_date	Date and time the wildfire was extinguished.
Burn Area	size_class	Grouping of wildfires based on final area burned.
Burn Area	current_size	Final area burned by the wildfire.
Burn Area	discovered_size	Size of the wildfire when first discovered.
Burn Area	assessment_hectares	Size of the wildfire at the time of assessment.
Burn Area	fire_fighting_start_size	Size of the wildfire at the start of suppression.
Burn Area	bh_hectares	Size of the wildfire at 'Being Held' status.
Burn Area	uc_hectares	Size of the wildfire at 'Under Control' status.
Burn Area	to_hectares	Size of the wildfire at turn over.
Burn Area	ex_hectares	Size of the wildfire at extinguishment.
Resource	det_agent	Detection agent that discovered the wildfire.
Resource	det_agent_type	General type of detection agent responsible for discovering the wildfire.
Resource	dispatched_resources	First Wildfire Management resource dispatched.
Resource	assesment_resource	Name of the person who first assessed the wildfire.
Resource	fuel_type	Predominant fuel type at initial assessment.
Resource	initial_action_by	Resource that took the first physical suppressive action.
Resource	ia_access	Method of access by the initial action resource.
Resource	bucketing_on_fire	Whether a helicopter with a bucket was used.
Resource	distance_from_water_source	Distance from water source for helicopter bucket.
Other	fire_spread_rate	Rate of spread of the wildfire at initial assessment.

DATA CLEANSING

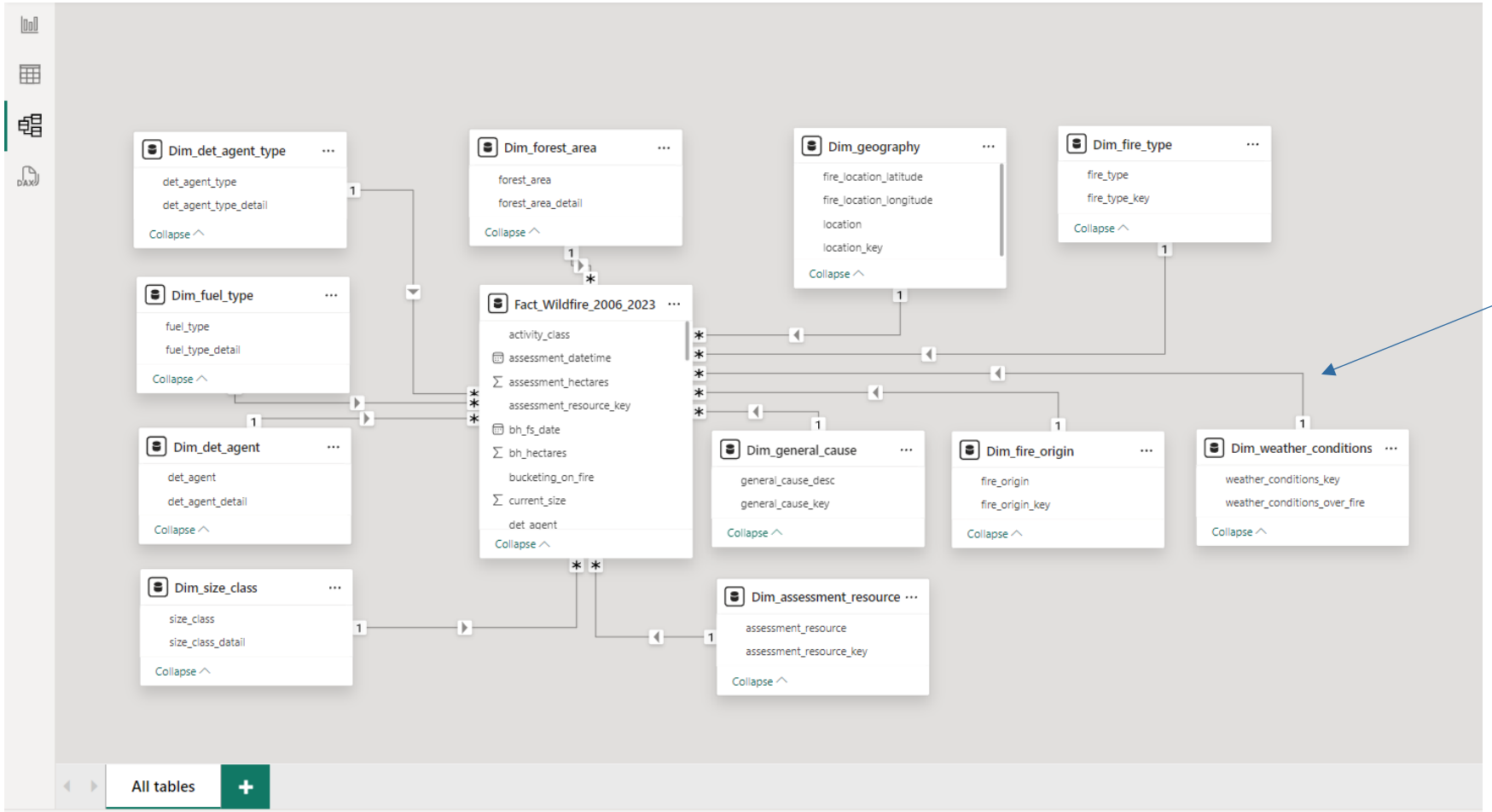
- Due to the large number of fields, combined with the need to solve the problem, keep 16 variables required for modeling as below.
- Variables with blue backgrounds are derived variables created to solve problems.

Classification	Field Name	Explanation
Location	fire_location_latitude	Latitude of the ignition area.
Location	fire_location_longitude	Longitude of the ignition area.
Location	**Forest_area**	The first letter from Fire_number and link to the area where the fire happened.
Basic	fire_origin	Ownership or administration of the land where the wildfire ignited.
Basic	**Fire_ID**	Fire_number is repeated in different year, combing fire_year and fire number to get the unique ID.
Cause	general_cause_desc	Classification of the wildfire cause.
Weather	weather_conditions_over_fire	Weather conditions over the wildfire at initial assessment.
Weather	temperature	Temperature at the wildfire site.
Weather	relative_humidity	Relative humidity at the wildfire site.
Weather	wind_direction	Wind direction at the wildfire site.
Weather	wind_speed	Wind speed at the wildfire site.
Time	fire_year	The wildfire year from April 1 to March 31 of the following year.
Time	reported_date	Date and time the wildfire was first reported.
Time	**Response_time**	Response_time = IF(NOT(ISBLANK(Fact_Wildfire_2006_2023[start_for_fire_date])) || NOT(ISBLANK(Fact_Wildfire_2006_2023[ia_arrival_at_fire_date])), DATEDIFF(Fact_Wildfire_2006_2023[reported_date], MIN(Fact_Wildfire_2006_2023[start_for_fire_date], Fact_Wildfire_2006_2023[ia_arrival_at_fire_date]), MINUTE), BLANK())
Burn Area	size_class	Grouping of wildfires based on final area burned.
Burn Area	current_size	Final area burned by the wildfire.

4 DATA MODELING

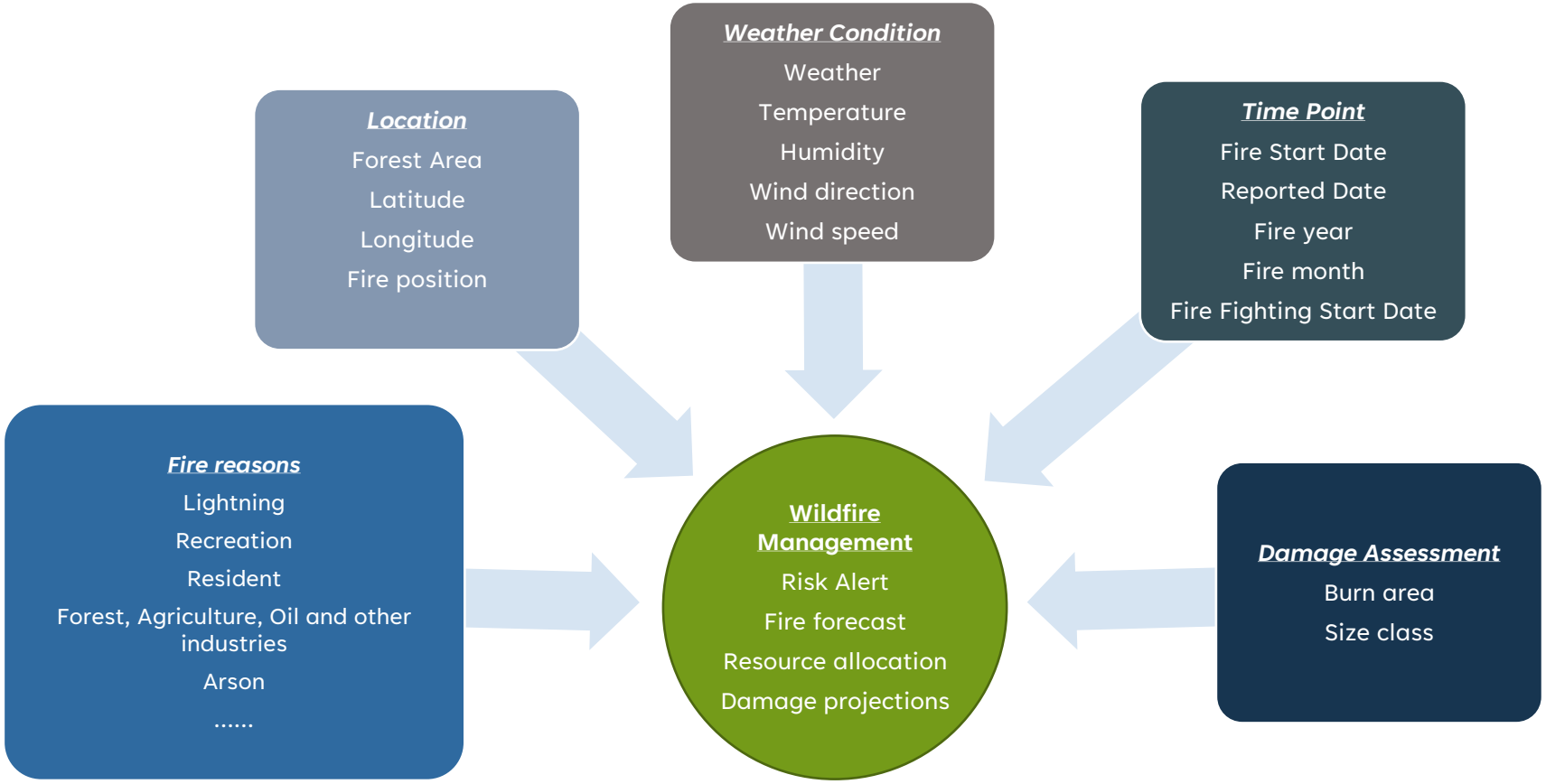
- Create Fact and Dimensions for the dataset
- Manage the relationship of these tables

DATA DIAGRAM



- Star
- 1 to many

CONCEPTUAL MODEL



5 PREDICTIVE ANALYTICS & VISUALIZATION

- DASHBOARD

Alberta Wildfire Infomation

Fire Year

2006 — 2023

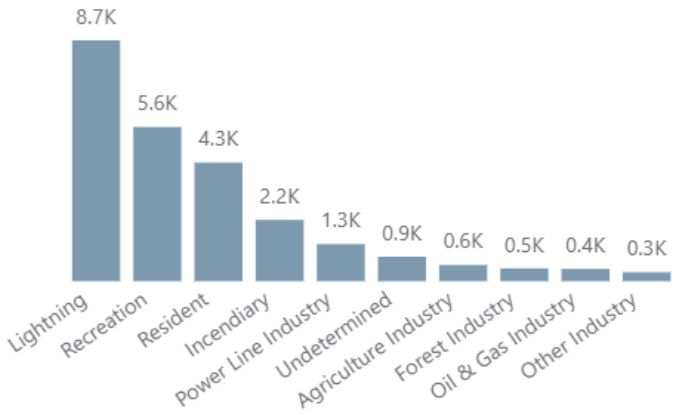
The number of wildfire

25.32K

Burn Areas

5.90M

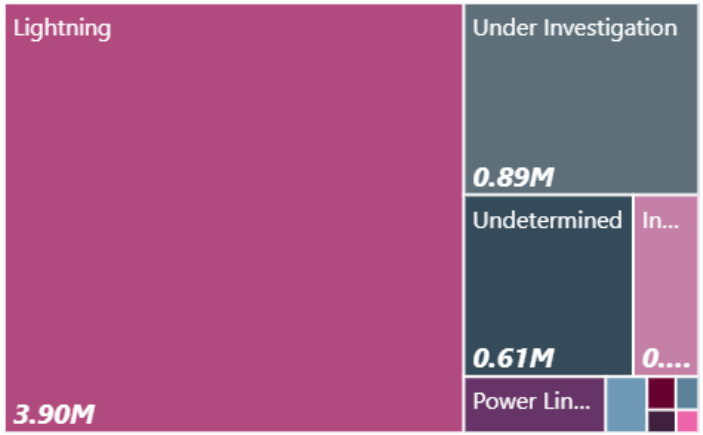
Top 10 Reason of Fire numbers



Response Time (Minutes)

Average of Response_time	Median of Response_time	Min of Response_time	Max of Response_time
71.12	3	0	41K

Top 10 Reason of Burn areas

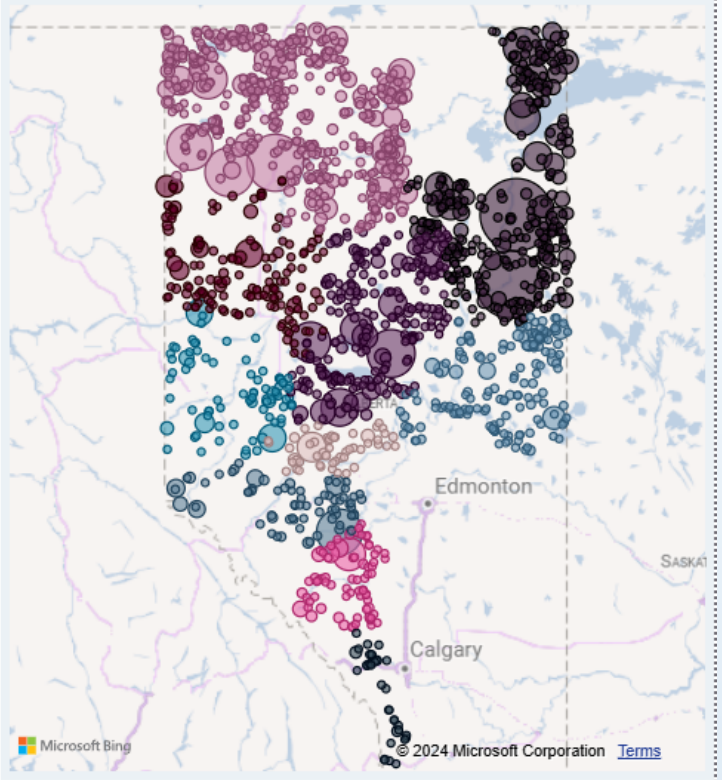


Weather of Wildfire Class

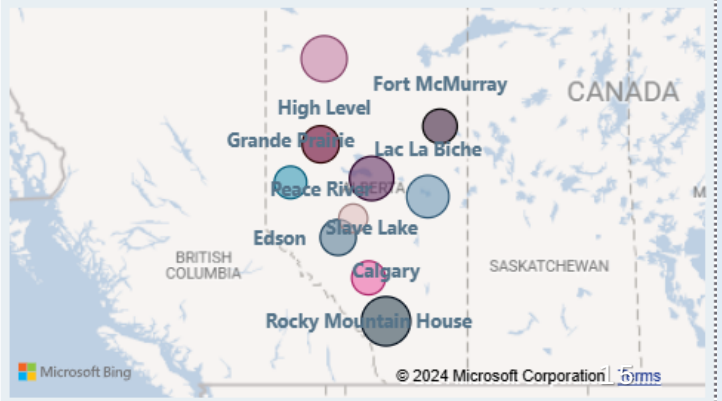
size_class_datail	CB Dry	CB Wet	Clear	Cloudy	Rainshowers
100.1-1000.0 ha	10.10%	4.33%	62.98%	22.36%	0.24%
10.1-100.0 ha	10.14%	3.66%	53.80%	30.14%	2.25%
1.1-10.0 ha	10.69%	5.26%	52.38%	29.19%	2.47%
0.11-1.0 ha	7.24%	6.41%	50.53%	32.80%	3.02%
0.01-0.1 ha	5.24%	7.07%	43.79%	38.78%	5.12%

0.7 hectares ≈ , 10 hectares ≈ × 14.3

Burn Areas of wildfires



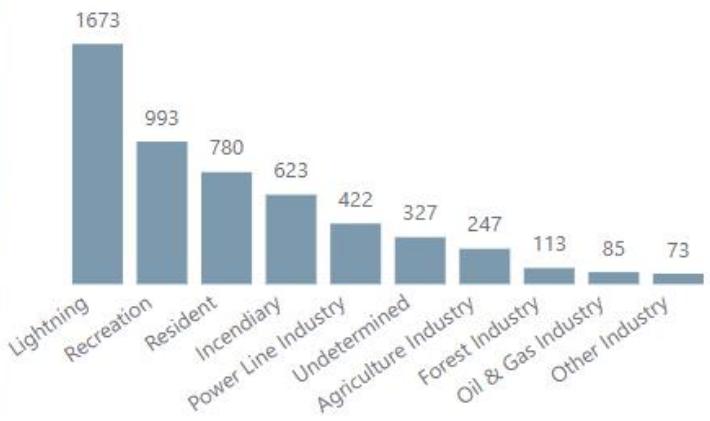
Fire counts of Area



➢ REPORT FOR QUESTION 1

1. What are the most common causes of wildfires in Alberta over the 5 years?

Top 10 Reason of Fire numbers



general_cause_desc	Count of Fire_ID
Lightning	1673
Recreation	993
Resident	780
Incendiary	623
Power Line Industry	422
Undetermined	327
Agriculture Industry	247
Forest Industry	113
Oil & Gas Industry	85
Other Industry	73
Total	**5336**

Fire Year

2019 — 2023

Top 10 Reason of Burn areas

general_cause_desc	Sum of current_size
Lightning	2,500,330.58
Under Investigation	404,106.24
Incendiary	223,984.41
Power Line Industry	49,983.13
Oil & Gas Industry	5,215.94
Recreation	3,885.74
Undetermined	2,554.01
Resident	1,081.23
Agriculture Industry	920.72
Railroad	90.24
Total	**3,192,152.24**

Conclusion

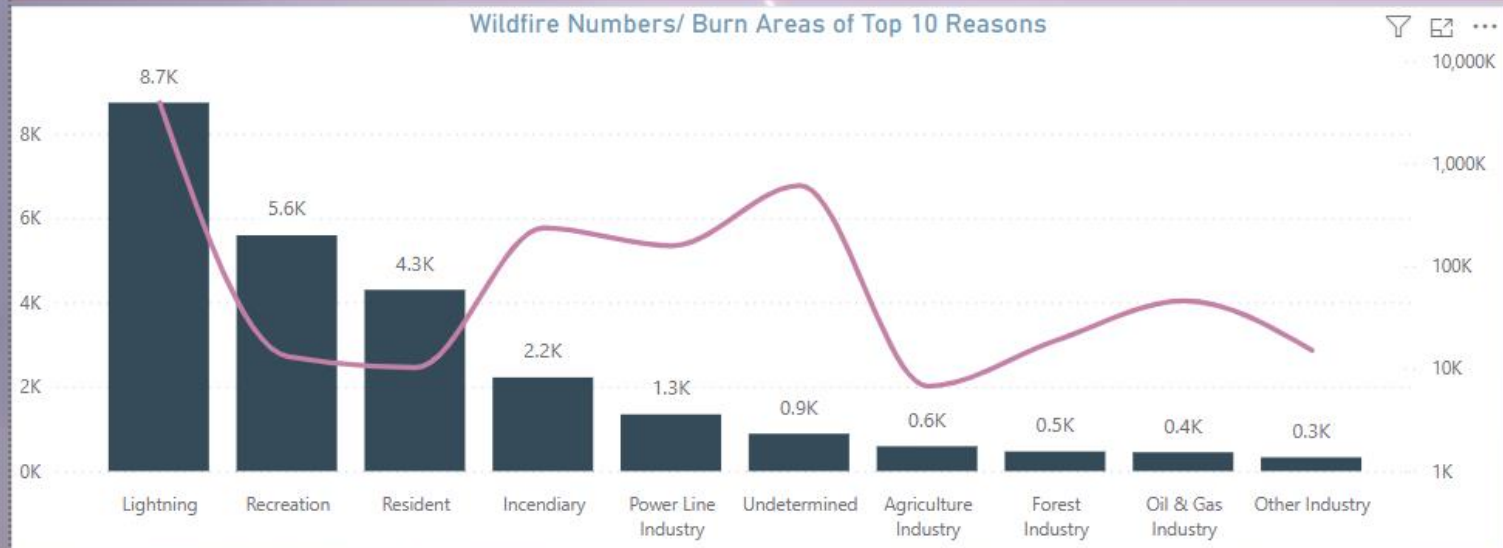
In the last 5 years, the most common reason to cause the wildfire is **'Ligthning','Recreation','Resident','Incendiary', and 'Power Line Industry'.**

These reasons cause the most biggest burn area in the last 5 years, is **'Lightning', *'Under Investigation'(Unknown)*, 'Incendiary', 'Power Line Industry', and 'Oil & Gas Industry', 'Recreation'** perspectively.

Overall, 'Lightning' casued the biggest number and burn area in AB.

➢ REPORT FOR QUESTION 2 & 3

2. Which areas are most susceptible to wildfires and why?
3. What are the main causes of wildfires in different areas, and do they vary significantly?



Fire counts of Area

forest_area_detail	Count of Fire_ID
Calgary	4275
High Level	3657
Slave Lake	3367
Lac La Biche	3057
Peace River	2251
Edson	2100
Fort McMurray	1902
Rocky Mountain House	1816
Grande Prairie	1652
Total	**25321**

Conclusion

Question 2:

Calgary *have* ***the most serious wilfdire situation*** *(* most wildfire numbers) caused by* ***'Recreation' and 'Resident'.***

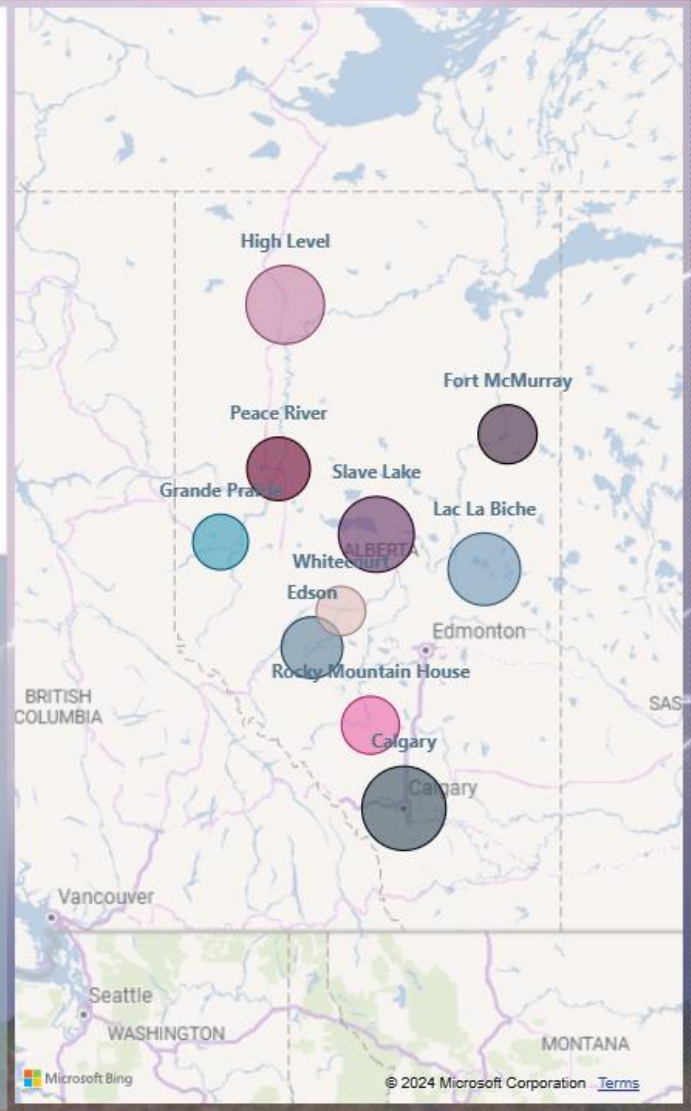
High Level *is the second rank of wildfire numbers in AB but totally different with Calgary, the top 1 reason of wildfire is* ***'Ligtning'*** *caused. 'Resident' and 'Incendiary' are in the top 3 reasons.*

Slave Lake *is similar with High level, the main reason of wildfire is* ***'Lightning'****.*

Lac La Biche *is interesting. The top 1 reason of wildfire is* ***'Incendiary'****, over the numbers of 'Lightning' caused.*

Question 3:

Calgary forest ares is the most susceptible to wildfires because of the recreation need.
Lightning caused wildfires are easiler to happen in the high altitude area.
Resident caused wildfires are average in the differnt foreat areas in Alberta.
Incendiary is gather in Lac La Biche and High Level.

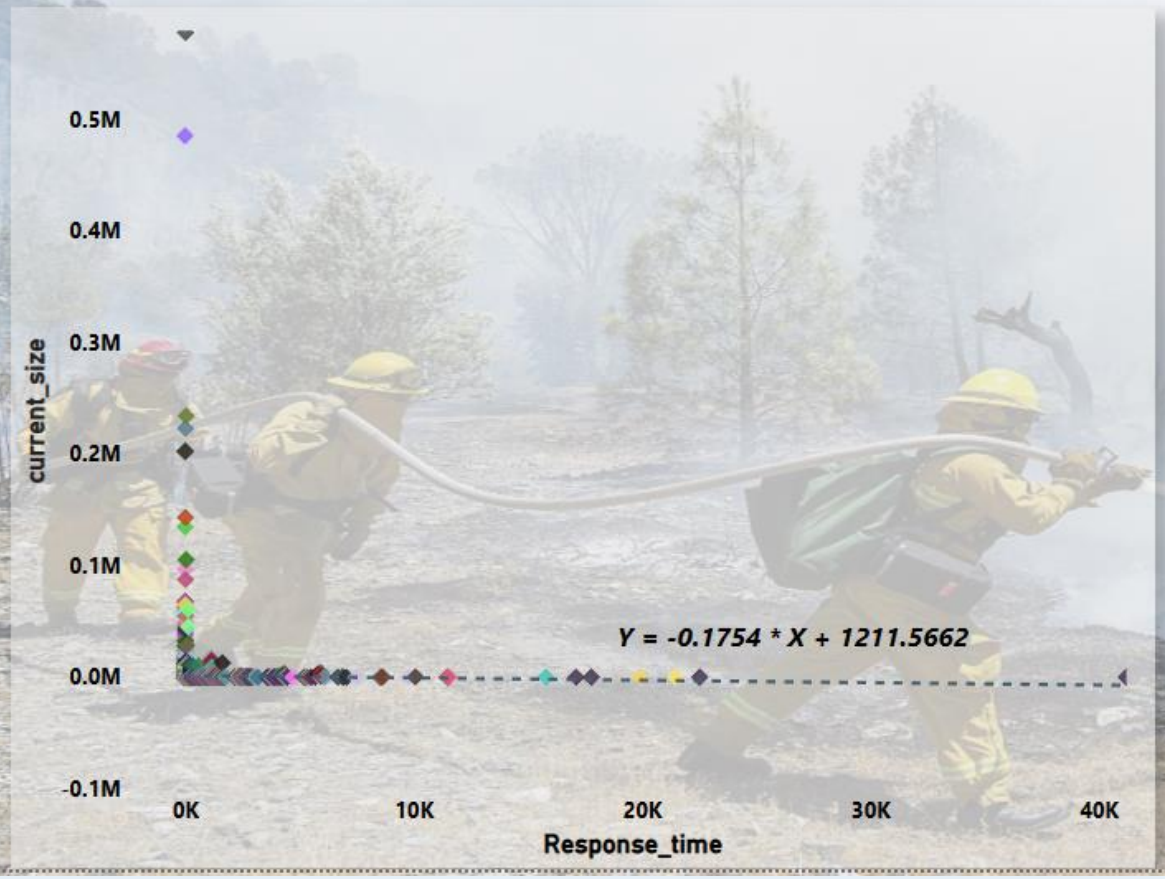


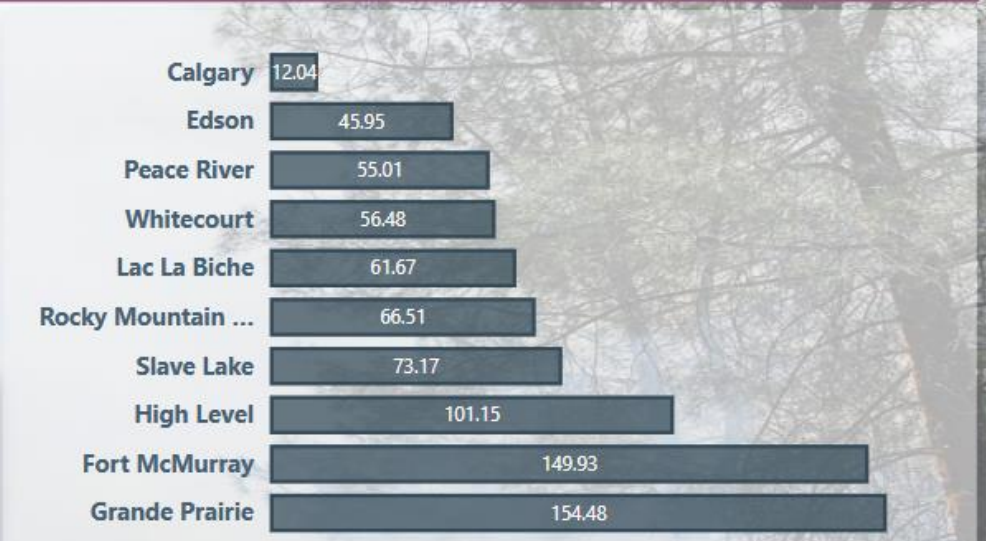
➢ REPORT FOR QUESTION 4

4. What are the trends in response times and their effectiveness in mitigating wildfire damage?

Response Time (Minutes)

Average of Response_time	Median of Response_time	Min of Response_time	Max of Response_time
71.12	3	0	41K





Conclusion

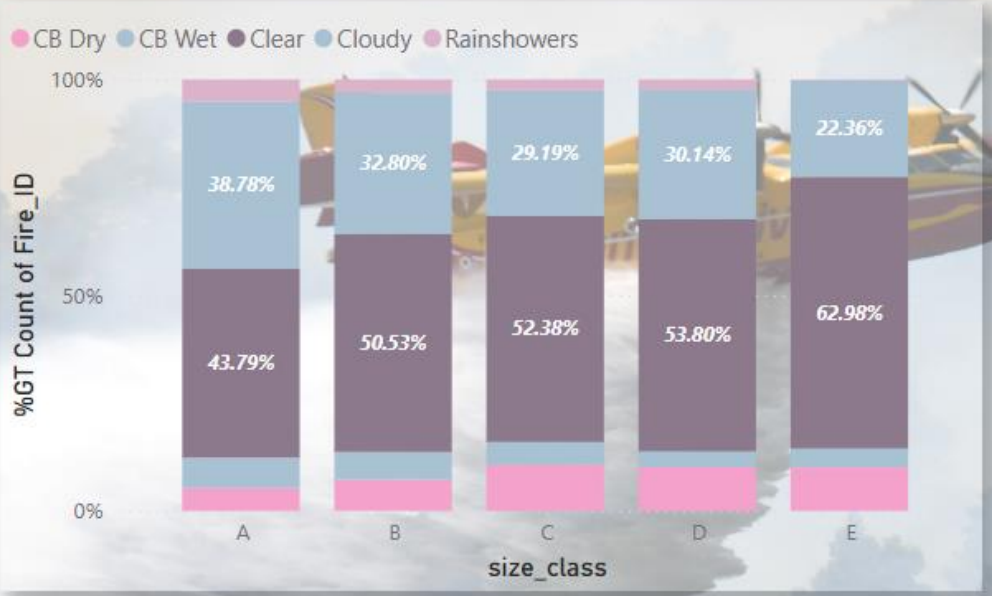
The scatterplot and the regression coefficients show that the slope is negative, then it proves that the response coefficients are very weakly related to the final area burned by the wildfire.

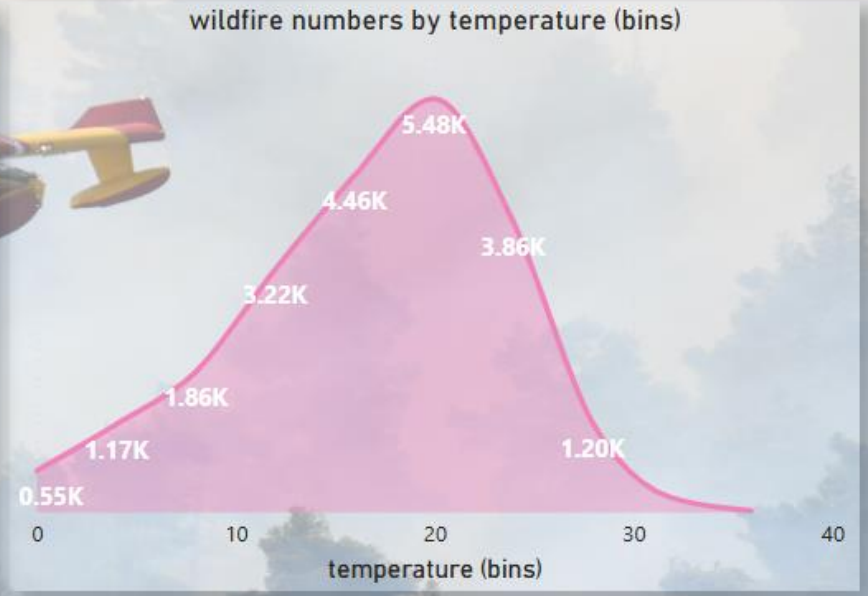
At the same time, there is still public demand for rapid response to wildfires. Therefore, by looking at response times in different forested areas, we found some differences between regions.

Comparing response times in different regions, the average response time in the Calgary region was 12 minutes, while the longest response time was 155.5 minutes, about 10 times that of the shortest region.

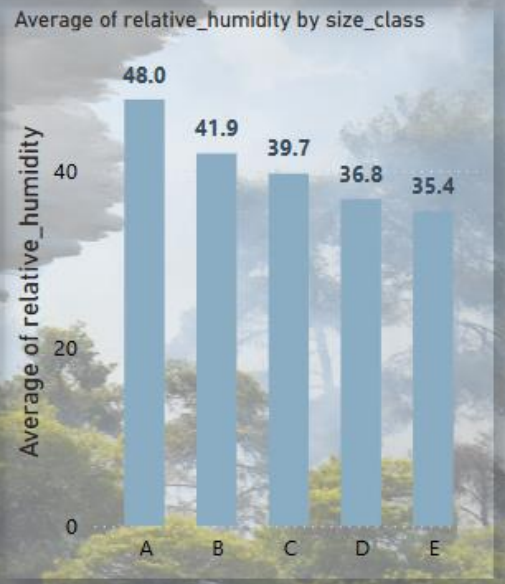
➢ REPORT FOR QUESTION 5

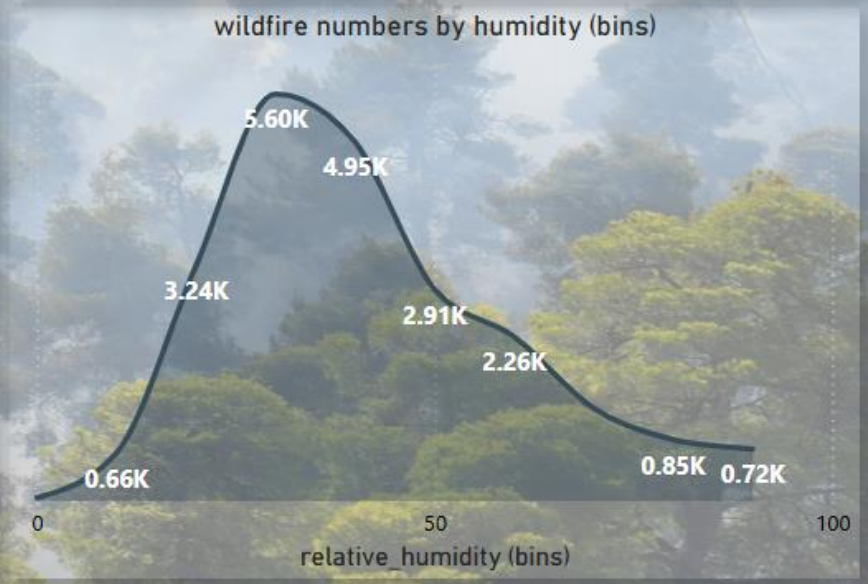
5. How do weather conditions, temperature and humidity influence the frequency and severity of wildfires?





size_class	size_class_datail
A	0.01-0.1 ha
B	0.11-1.0 ha
C	1.1-10.0 ha
D	10.1-100.0 ha
E	100.1-1000.0 ha
F	1000.1-5000.0 ha
G	5000.1+ ha





Conclusion

There is a correlation between both weather, temperature, and humidity and the number of wildfires and the area burned.

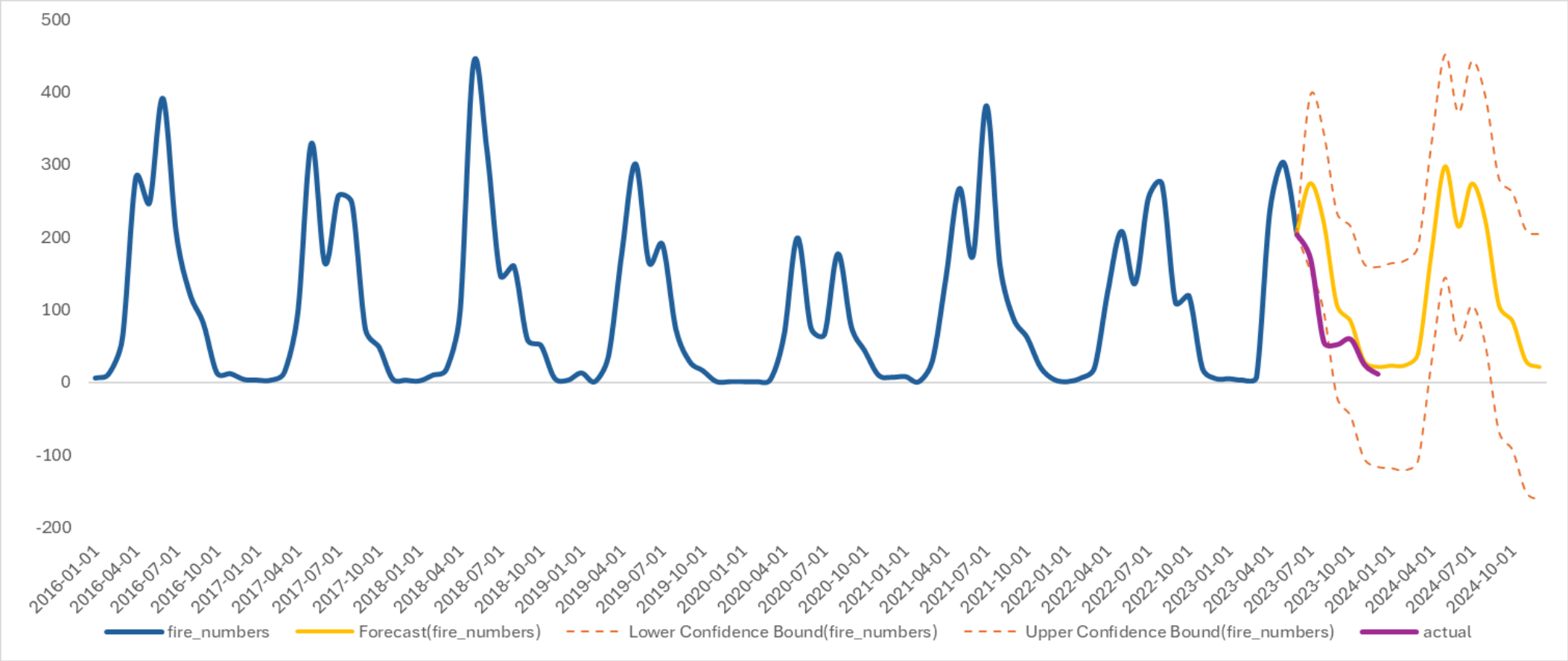
Therefore, consideration can be given to applying the correlation factors in subsequent modeling.

6 PREDICTIVE MODEL

TIME SERIES FORECAST

1) Through the analysis, we learned that wildfires occur seasonally, with a high incidence in the summer, so predicting the number of fires does not have a greater significance.
2) Through the prediction results of the time series model, although the actual number of fires on the test dataset is still within the 95% confidence interval of the predicted number, there is still some difference in the trend.

❖ In summary, we therefore believe that the time series model does not have a large significance in predicting the number of wildland fire occurrences.



CLASSIFICATION MODEL

- In solving the business questions, we found that many factors have a greater relationship to the final damage caused by the fire, such as the cause of the fire and the weather conditions at the time of the fire.
- Therefore, if at the early stage of a fire, we can predict the extent of the damage caused by the fire based on the available information, we can help the relevant departments to adjust their response strategies, such as resource preparation, equipment investment (e.g. helicopters, drones, etc.), and firefighters.

❖ **Set "size_class" as the Y variable.** This variable contains five classes by dividing the burned area of the fire, with A being the smallest and E being the largest, making the classification model well suited for this study.

A class = 0 to 0.1 ha

B class > 0.1 ha to 4.0 ha

C class > 4.0 ha to 40.0 ha

D class > 40.0 ha to 200 ha

E class > 200 ha

❑ **Select Top 9 features based on the business understanding.**

• Using ***Random Forest*** to get the importance of features.

```
Rank of Features:
1. Feature fire_location_latitude (0.18383832155888782)
2. Feature fire_location_longitude (0.150362033488555)
3. Feature relative_humidity (0.12239827194180786)
4. Feature temperature (0.110098739840662 49)
5. Feature wind_speed (0.08656205803791009)
6. Feature wind_direction (0.07687606207326275)
7. Feature forest_cause_combined (0.06109546976918299)
8. Feature reported_month (0.05841920878427568)
9. Feature weather_conditions_over_fire (0.04689692617491824)
10. Feature general_cause_desc (0.04588915061354143)
11. Feature forest_area (0.034179443652777 59)
12. Feature fire_origin (0.023384314064218 05)
```

❑ **Combined two features together to improve the importance**

BEST MODEL FOR PREDICTION

- Use the tools provided by ***Pycaret*** to find the classification algorithms that work best.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lightgbm	Light Gradient Boosting Machine	0.6848	0.7816	0.6848	0.6297	0.6425	0.2440	0.2605	
xgboost	Extreme Gradient Boosting	0.6784	0.7773	0.6784	0.6295	0.6441	0.2502	0.2600	
rf	Random Forest Classifier	0.6779	0.7696	0.6779	0.6193	0.6313	0.2161	0.2334	
gbc	Gradient Boosting Classifier	0.6768	0.0000	0.6768	0.6032	0.6191	0.1896	0.2147	
ada	Ada Boost Classifier	0.6690	0.0000	0.6690	0.5878	0.6000	0.1458	0.1760	
ridge	Ridge Classifier	0.6681	0.0000	0.6681	0.5613	0.5830	0.1036	0.1401	
et	Extra Trees Classifier	0.6676	0.7487	0.6676	0.6093	0.6249	0.2019	0.2157	
lr	Logistic Regression	0.6662	0.0000	0.6662	0.5739	0.5958	0.1315	0.1590	
dummy	Dummy Classifier	0.6635	0.5000	0.6635	0.4402	0.5293	0.0000	0.0000	
lda	Linear Discriminant Analysis	0.6619	0.0000	0.6619	0.5713	0.5969	0.1378	0.1608	
knn	K Neighbors Classifier	0.6533	0.7069	0.6533	0.6030	0.6177	0.1876	0.1953	
dt	Decision Tree Classifier	0.5982	0.6074	0.5982	0.6007	0.5994	0.1850	0.1850	
svm	SVM - Linear Kernel	0.5908	0.0000	0.5908	0.6145	0.5344	0.1317	0.1750	
nb	Naive Bayes	0.2839	0.6726	0.2839	0.6713	0.3815	0.0825	0.1159	
qda	Quadratic Discriminant Analysis	0.1611	0.0000	0.1611	0.1169	0.0909	-0.0025	-0.0107	

- Based on the better performance of the first two algorithm results, the best model parameters were measured for subsequent modeling.

	precision	recall	f1-score	support
0	0.74	0.91	0.81	3365
1	0.43	0.29	0.35	1282
2	0.36	0.02	0.03	261
3	0.30	0.04	0.08	69
4	0.40	0.19	0.26	88
accuracy			0.68	5065
macro avg	0.45	0.29	0.31	5065
weighted avg	0.63	0.68	0.64	5065

Best parameters for lighgbm found: {'colsample_bytree': 1.0, 'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 50, 'subsample': 0.8}

	precision	recall	f1-score	support
0	0.74	0.91	0.82	3365
1	0.45	0.32	0.37	1282
2	0.18	0.02	0.03	261
3	0.33	0.03	0.05	69
4	0.46	0.15	0.22	88
accuracy			0.69	5065
macro avg	0.43	0.28	0.30	5065
weighted avg	0.63	0.69	0.64	5065

Best parameters for xgboost found: {'colsample_bytree': 0.9, 'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 100, 'subsample': 0.9}

EVALUATE MODEL

- In the model generated by ***lightgbm***, we can see that the geographic location of the fire (latitude and longitude information) is ranked higher in importance and that the weather factor is relatively less important. Whereas in the model generated by ***Xgboost***, the influence of weather is more valued and there is no one item that is particularly weakened in the characteristics. We also concluded from our previous analysis that latitude has a correlation with fire severity, while there is no particularly significant relationship in longitude.

- Therefore, we believe that the predictive model generated by ***Xgboost*** is more in line with our knowledge of fire severity classifications in our study.

- LightGBM

```
[LightGBM] [Info] Total Bins 1688
[LightGBM] [Info] Number of data points in the train set: 67180, number of used features: 9
[LightGBM] [Info] Start training from score -1.609438
[LightGBM] [Info] Start training from score -1.609438
[LightGBM] [Info] Start training from score -1.609438
[LightGBM] [Info] Start training from score -1.609438
[LightGBM] [Info] Start training from score -1.609438
              precision    recall  f1-score   support

           0       0.76      0.81      0.79      3365
           1       0.44      0.30      0.35      1282
           2       0.16      0.13      0.14       261
           3       0.07      0.19      0.10        69
           4       0.16      0.38      0.22        88

    accuracy                           0.63      5065
   macro avg       0.32      0.36      0.32      5065
weighted avg       0.63      0.63      0.63      5065

Feature importances:
1. Feature wind_speed (1076)
2. Feature fire_location_latitude (1045)
3. Feature forest_cause_combined (988)
4. Feature fire_location_longitude (957)
5. Feature reported_month (950)
6. Feature relative_humidity (830)
7. Feature temperature (737)
8. Feature wind_direction (587)
9. Feature weather_conditions_over_fire (330)
```

- XGBoost

```
              precision    recall  f1-score   support

           0       0.77      0.80      0.79      3365
           1       0.42      0.32      0.36      1282
           2       0.15      0.14      0.14       261
           3       0.08      0.17      0.11        69
           4       0.18      0.35      0.24        88

    accuracy                           0.63      5065
   macro avg       0.32      0.36      0.33      5065
weighted avg       0.63      0.63      0.63      5065

Feature importances:
1. Feature wind_speed (0.16743145883083344)
2. Feature weather_conditions_over_fire (0.15818674862384796)
3. Feature reported_month (0.12257900089025497)
4. Feature fire_location_latitude (0.11092520505189896)
5. Feature forest_cause_combined (0.09450942277908325)
6. Feature temperature (0.09324748069047928)
7. Feature wind_direction (0.09277991205453873)
8. Feature relative_humidity (0.08365174382925034)
9. Feature fire_location_longitude (0.07668904960155487)
```

PREDICTION SYSTEM

- Create a python to read the excel file with input features and output the results, also save a result excel file.

```
PS C:\SAIT\2024 Spring - DA\PROJ406 - Data Analytics Capstone Project\model> & C:/Users/Kai_Benzi/AppData/Local/Programs/Python/Python312/python.exe "c:/SAIT/2024 Spring - DA/PROJ406 - Data Analytics Capstone Project/model/3.wildfire_size_forecast_xgboost.py"
  reported_month          forest_area  fire_location_latitude  fire_location_longitude   general_cause_desc weather_conditions_over_fire  temperature  relative_humidity wind_direction  wind_speed predicted_size_class
0         August            High Level               58.576545              -117.339872            Lightning                      Clear         28.0               15.0             NE        40.0                    D
1      September               Calgary               51.523369              -115.049229           Recreation                    CB Dry         19.0               30.0              S        10.0                    A
2           June          Lac La Biche               54.868917              -111.568655           Incendiary                    Cloudy         21.0               22.0        missing         6.0                    A
3           July  Rocky Mountain House               52.368423              -115.178479   Oil & Gas Industry                    Cloudy         12.0               33.0             SE        25.0                    A
4            May           Peace River               56.281744              -117.168802       Forest Industry                    CB Wet         30.0                5.0             SE        80.0                    E
5           June           Whitecourt               54.199491              -115.613844  Under Investigation                    CB Dry         22.0               15.0              S         5.0                    A
6       February                Edson               53.589251              -116.472135         Undetermined                    Cloudy         24.0                7.0             NW         9.0                    B
Predicted results have been saved to prediction_results_xgboost.xlsx
```

- To better understand and use the predictive model, we built a simple page that makes it easier for interested parties to use the predictive model.

A class = 0 to 0.1 ha

B class > 0.1 ha to 4.0 ha

C class > 4.0 ha to 40.0 ha

D class > 40.0 ha to 200 ha

E class > 200 ha

1) Using Html file to create a website as input.
2) Using Flask to read the information we input from the website and output the predictive result from the model.

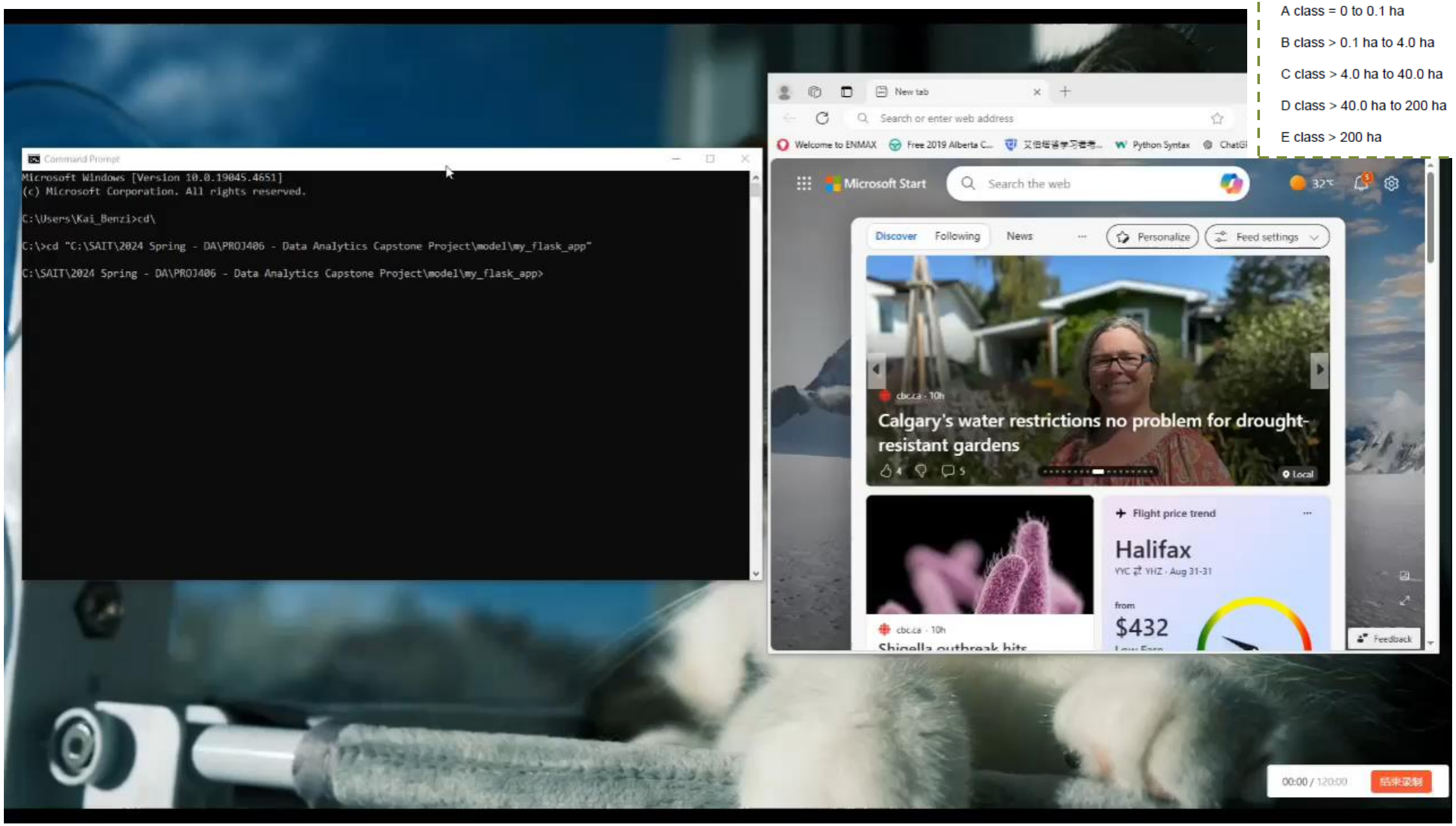
A class = 0 to 0.1 ha

B class > 0.1 ha to 4.0 ha

C class > 4.0 ha to 40.0 ha

D class > 40.0 ha to 200 ha

E class > 200 ha



7 CONCLUSION

WILDFIRE MANAGEMENT RECOMMENDATIONS

Based on predictive model

- ❑ Relevant departments can adjust their wildfire response strategies more effectively.
- ❑ Provide advance warning of the impact that fires that occur are expected to have on people's lives.

Based on analytical findings

- Wildfire response times and ultimate fire losses, while not strongly correlated, are still a topic of public concern, so further efficiency improvements are still needed for wildfire response in some areas.
- During the wildfire season, more frequent wildfire monitoring and warning is needed for weather conditions with low humidity, high temperatures, and no rainfall.
- There is a need for better management of the different causes of fires in different areas.

THANK YOU