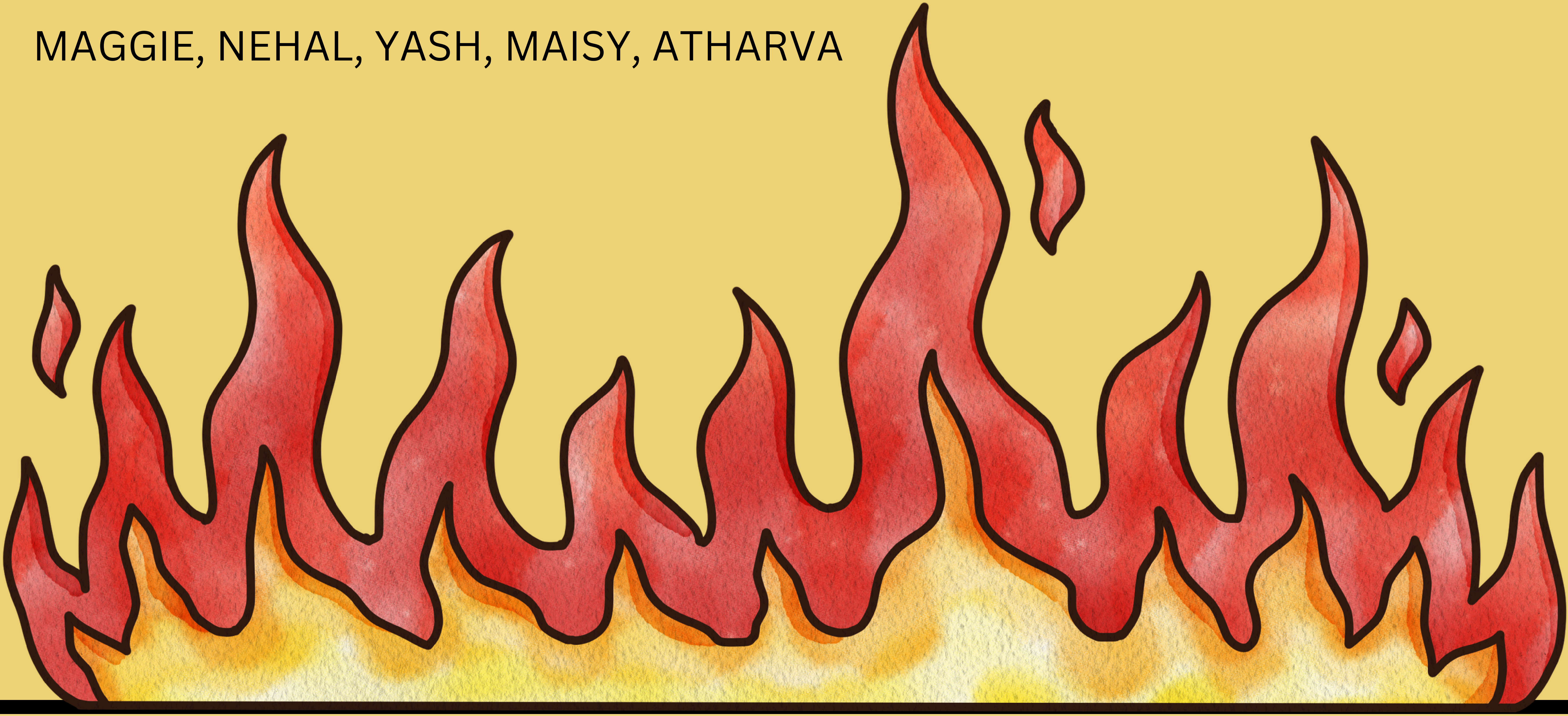
# TORONTO FIRE INCIDENTS

MAGGIE, NEHAL, YASH, MAISY, ATHARVA



# TABLE OF CONTENT



# BACKGROUND

| | |
|---|---|
| Significant Human Impact | Since 2010, Toronto has experienced 32,902 fire incidents, resulting in 1,796 fatalities and displacing 4,986 individuals. |
| Substantial Economic Losses | These incidents have led to financial losses totaling approximately $909.6 million. |
| Need for Enhanced Prevention | The considerable human and economic toll underscores the urgent need for effective fire prevention strategies. |

# VALUE PROPOSITION

**Fire incidents Prevention**

Analyzing Fire incidents to identify causes and recommending best practices to address vulnerabilities and reduce fire occurrences in the future.

**Estimating Dollar Loss**

Utilizing machine learning algorithms to predict estimated dollar losses enables insurance companies and emergency services to allocate resources more effectively, ensuring timely interventions and minimizing financial impacts.

# EXECUTIVE SUMMARY

- Models Built: Random Forest and Gradient Boosting

- Top 5 features influencing Estimated_Dollar_Loss, according to RF, are:
  - ‘Number_of_responding_personnel' - Number of TFS responding personnel
  - ‘Response_Time_Seconds' - the duration, measured in seconds, between when the fire is reported and when the first responders arrive at the scene
  - ‘Extent_Of_Fire' - OFM Extent Of Fire code and description
  - ‘Status_of_Fire_On_Arrival' - OFM Status of Fire On Arrival code and description
  - ‘TFS_Firefighter_Casualties' - Count of TFS casualties

- RF is the best at predicting ‘Institution/School’ & ‘Hazardous Materials/Gas’ types of fire

# DESCRIPTION OF THE DATA

32,093 INCIDENTS * 46 FEATURES

PLACE OF FIRE

- BUILDING_STATUS
- PROPERTY_USE
- INITIAL_CAD_EVENT_TYPE
- INTERSECTION
- LATITUDE
- LONGITUDE

FIRE STATION

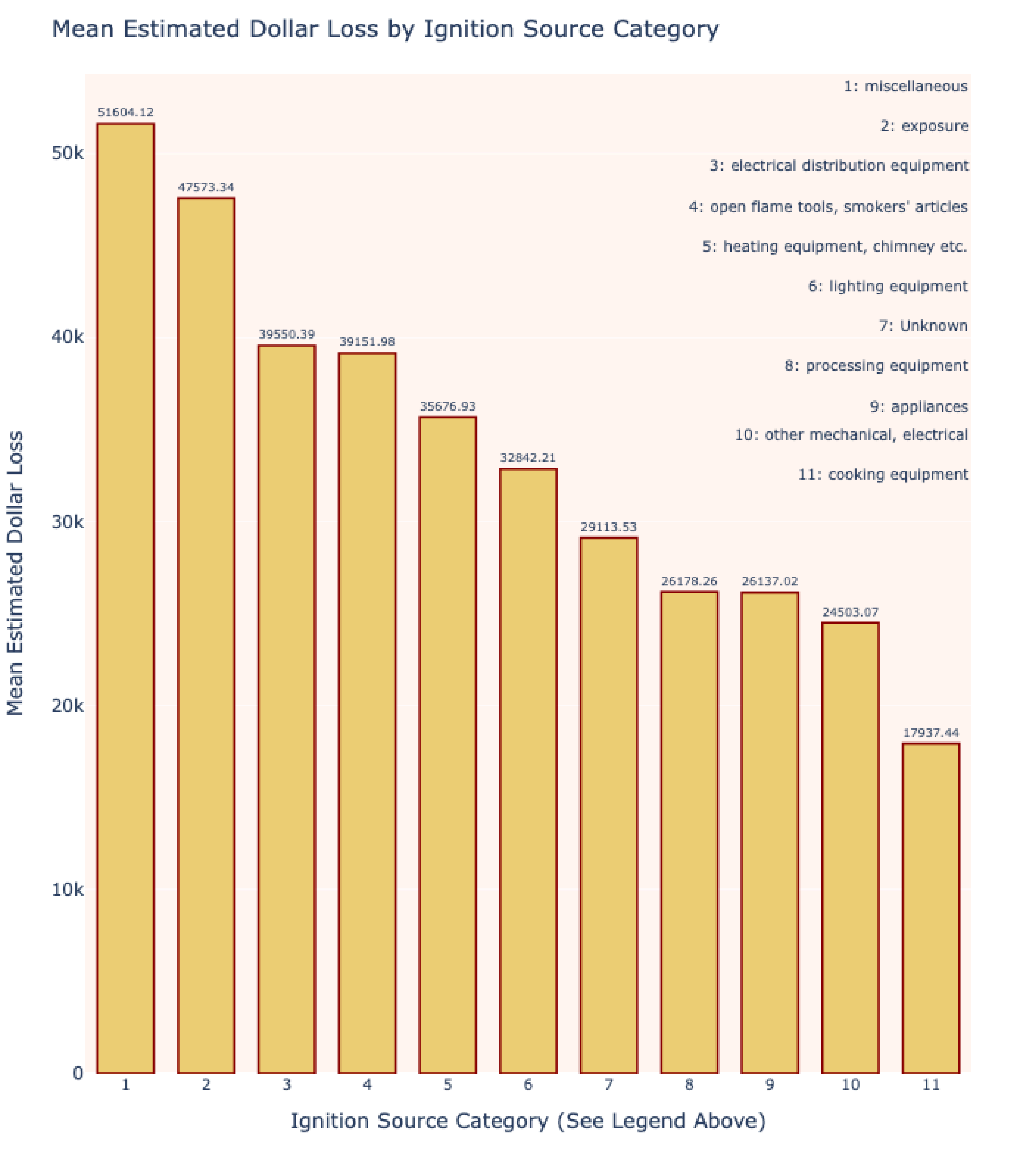
- INCIDENT_NUMBER
- INCIDENT_STATION_AREA
- INCIDENT_WARD
- TFS_ALARM_TIME
- TFS_ARRIVAL_TIME
- TFS_FIREFIGHTER_CASUALTIES

# DESCRIPTION OF THE DATA

## FIRE INCIDENT DETAILS

- AREA_OF_ORIGIN
- POSSIBLE_CAUSE
- BUSINESS_IMPACT
- CIVILIAN_CASUALTIES
- COUNT_OF_PERSONS_RESCUED
- ESTIMATED_NUMBER_OF_PERSONS_DISPLACED
- EXPOSURES
- SMOKE_SPREAD
- MATERIAL_FIRST_IGNITED
- METHOD_OF_FIRE_CONTROL
- NUMBER_OF_RESPONDING_APPARATUS
- NUMBER_OF_RESPONDING_PERSONNEL
- SPRINKLER_SYSTEM_OPERATION
- SPRINKLER_SYSTEM_PRESENCE
- STATUS_OF_FIRE_ON_ARRIVAL

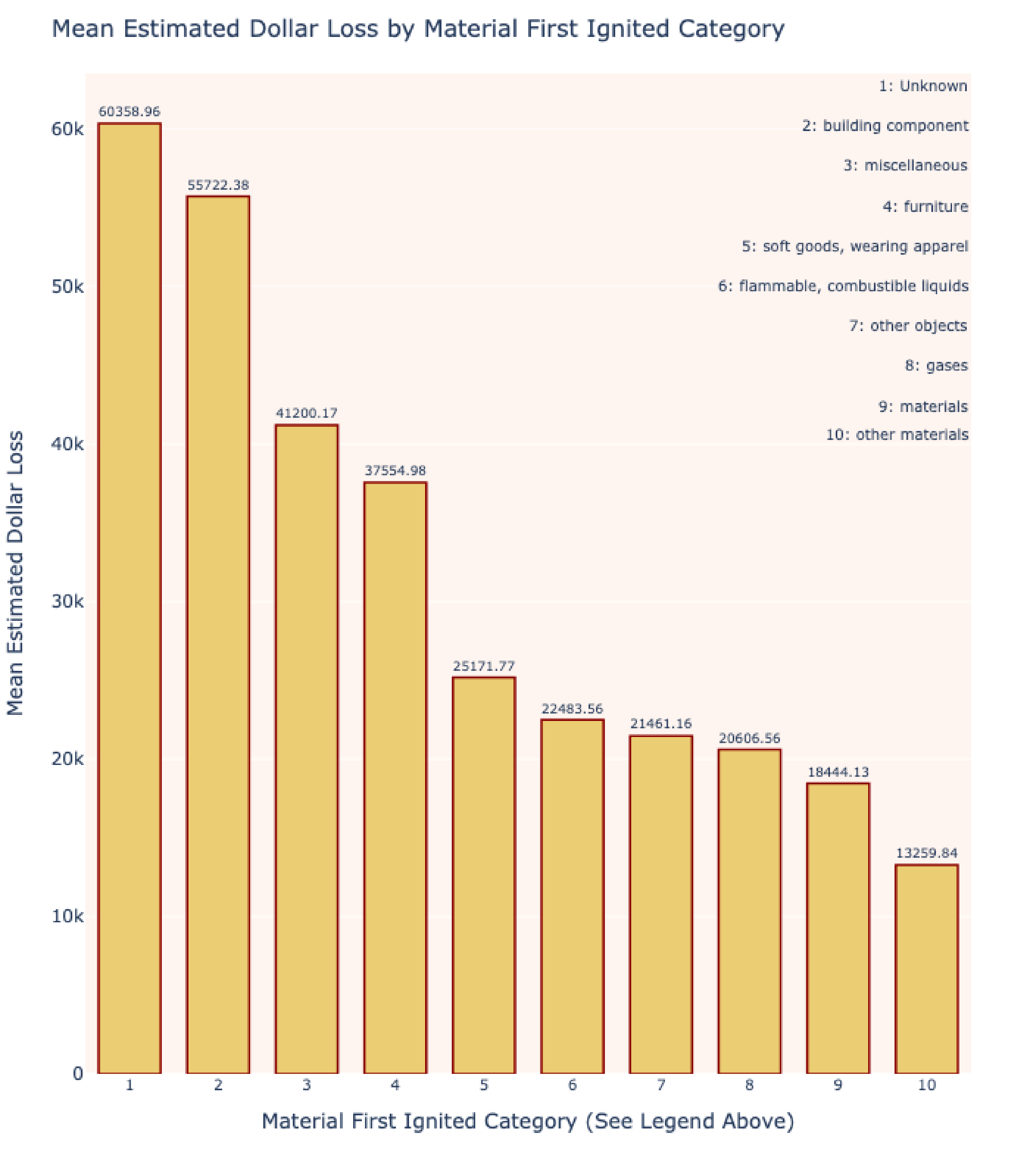
# EDA INSIGHTS



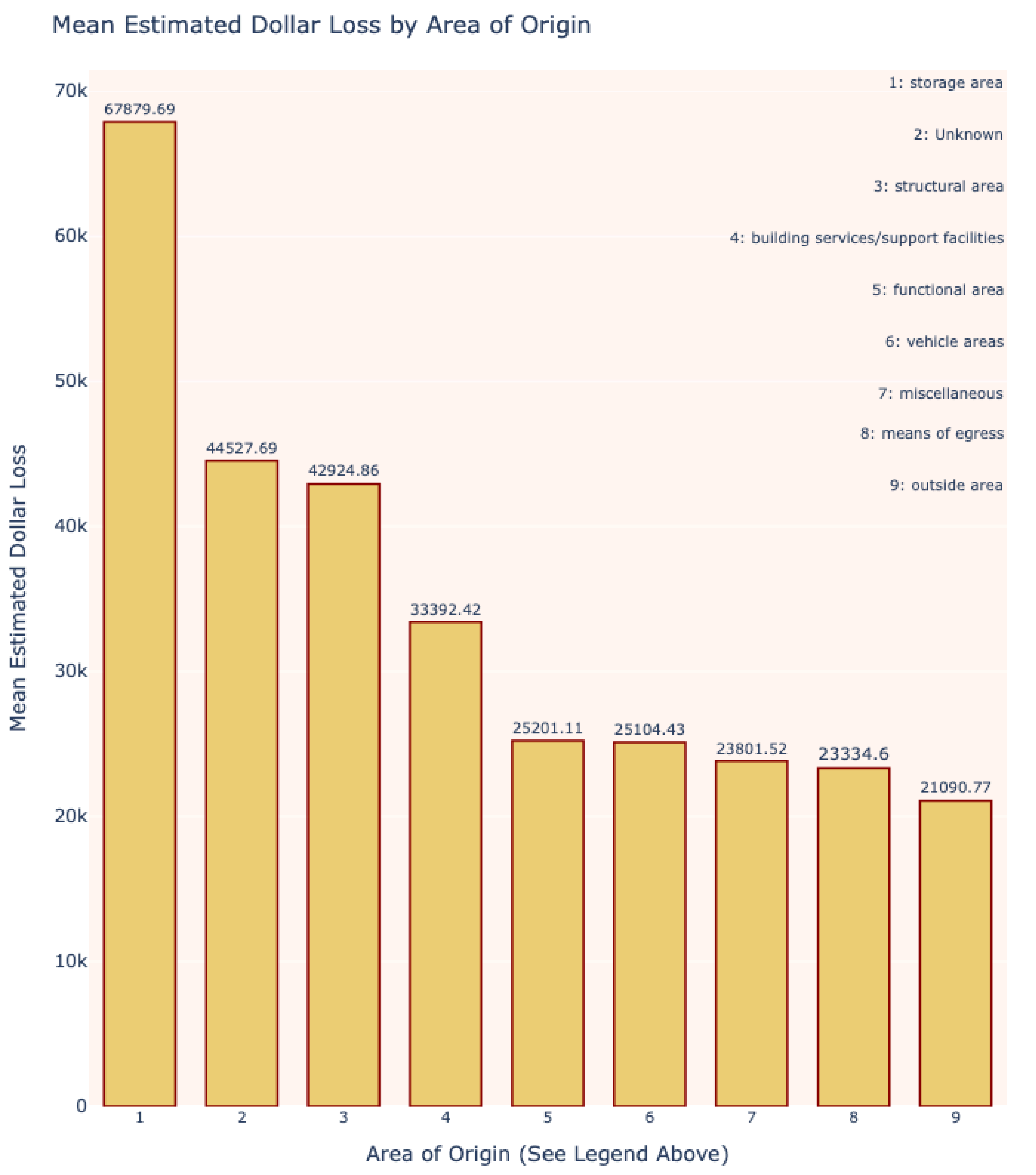

# Mean Estimated Dollar Loss by Ignition Source

- What is in miscellaneous?
- Exposures include forest fires
- Heating equipment

# Mean Estimated Dollar Loss by Material First Ignited

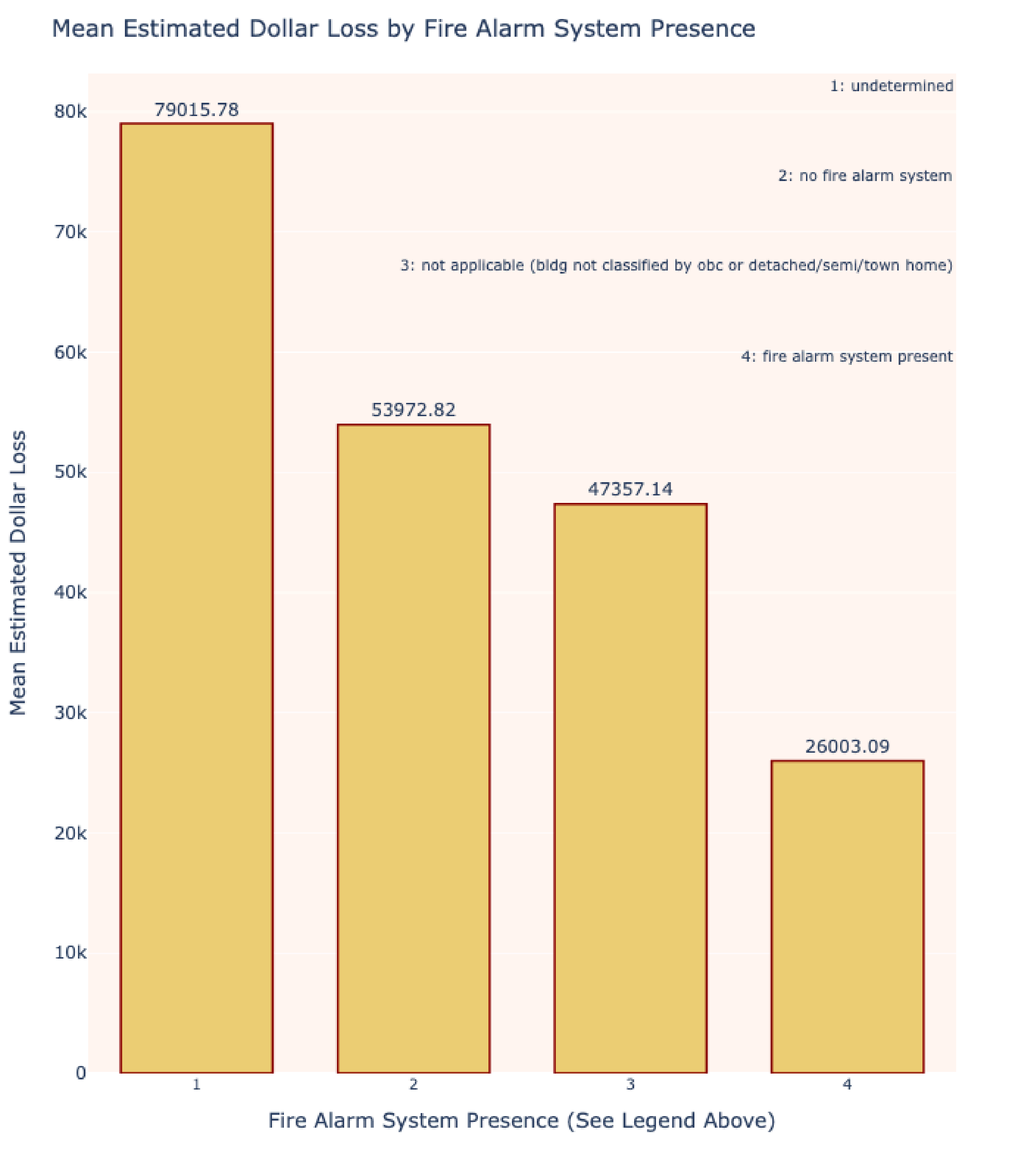



- We're not safe indoors now ???
- Miscellaneous strikes again!



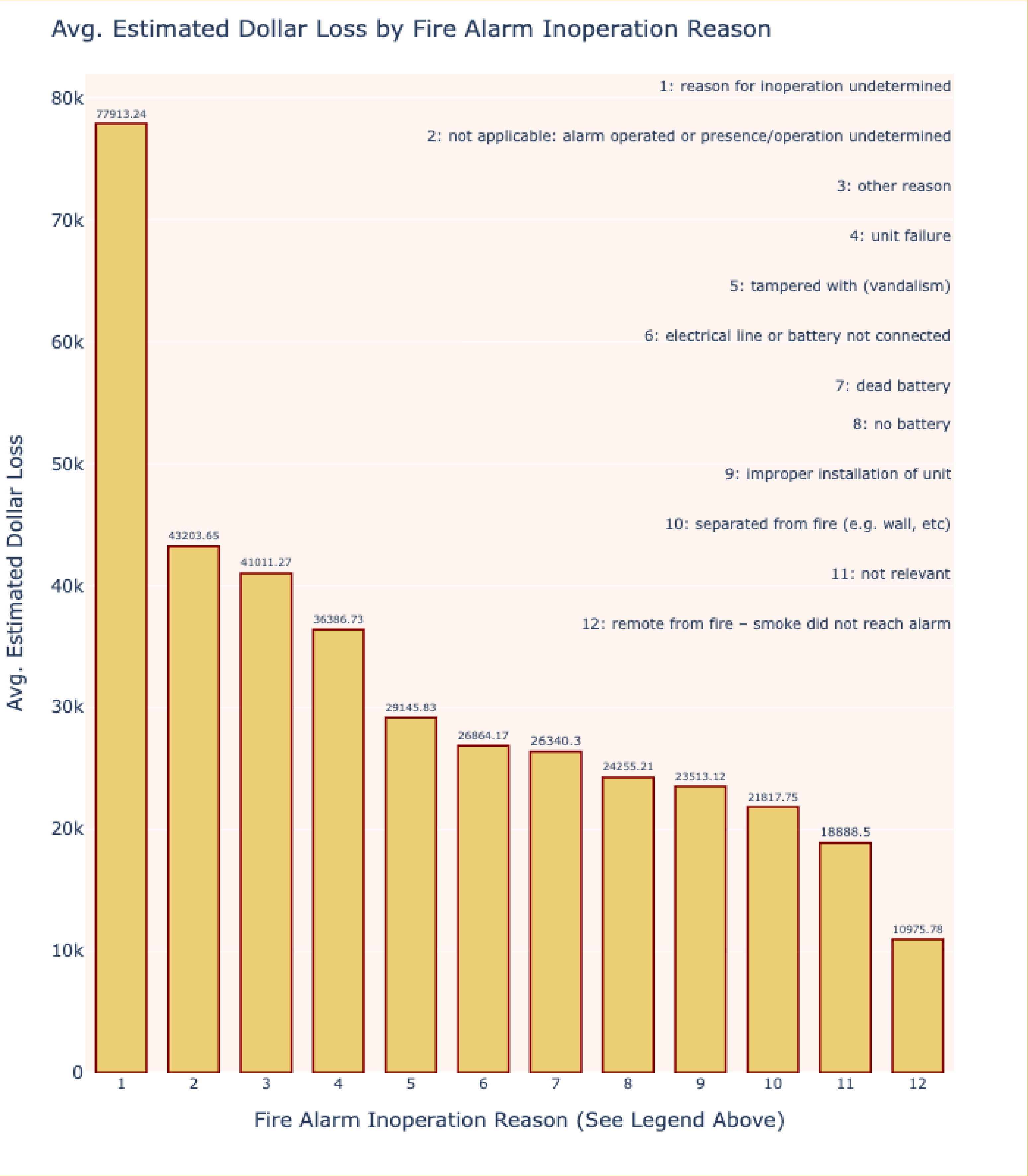

# Mean Estimated Dollar Loss by Area of Origin

- Storage Area stands out
- Capital Intensive structure



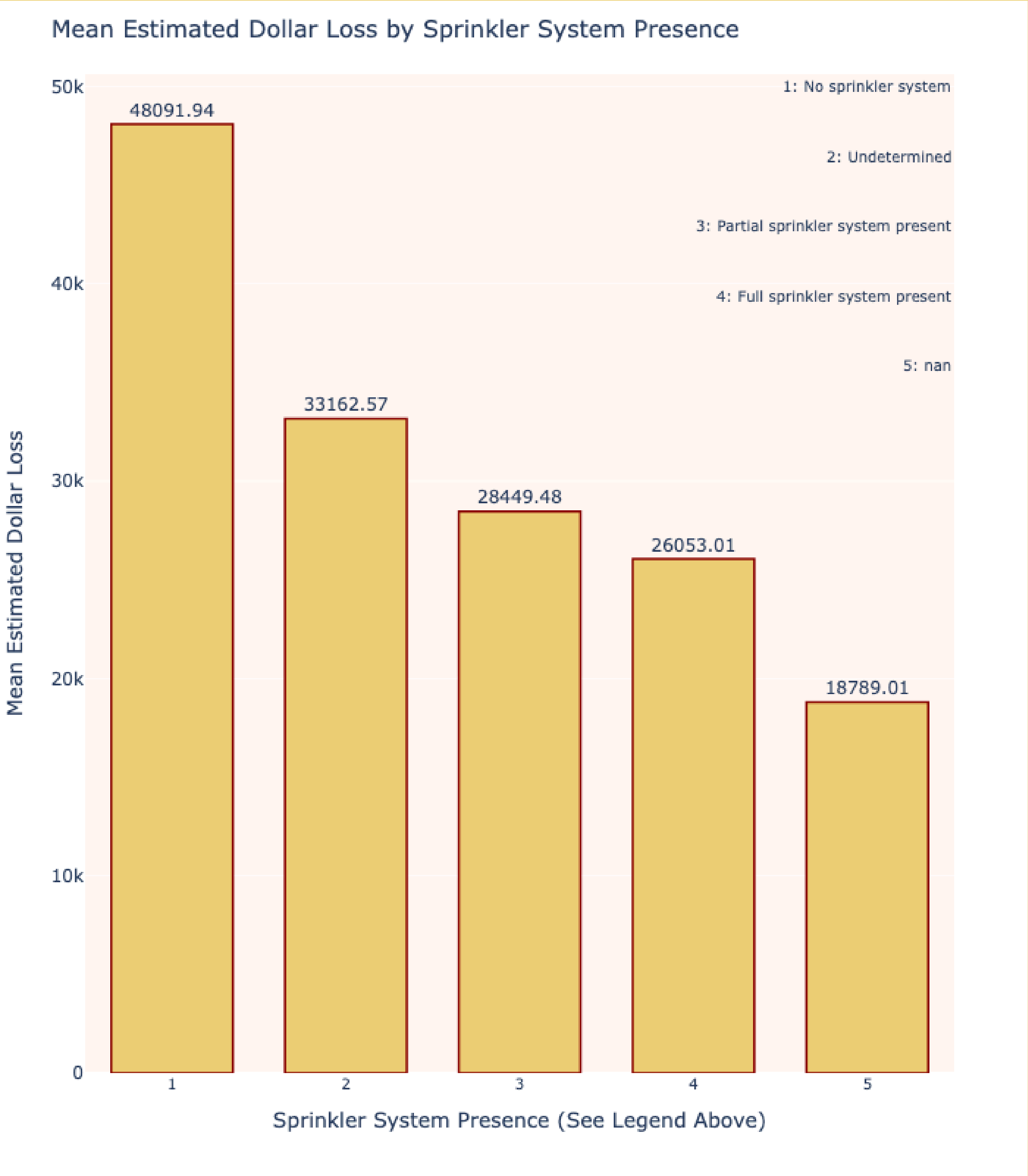

# Mean Estimated Dollar Loss by Fire Alarm System Presence

- Undetermined means something!
- Stark contrast in fire alarm presence



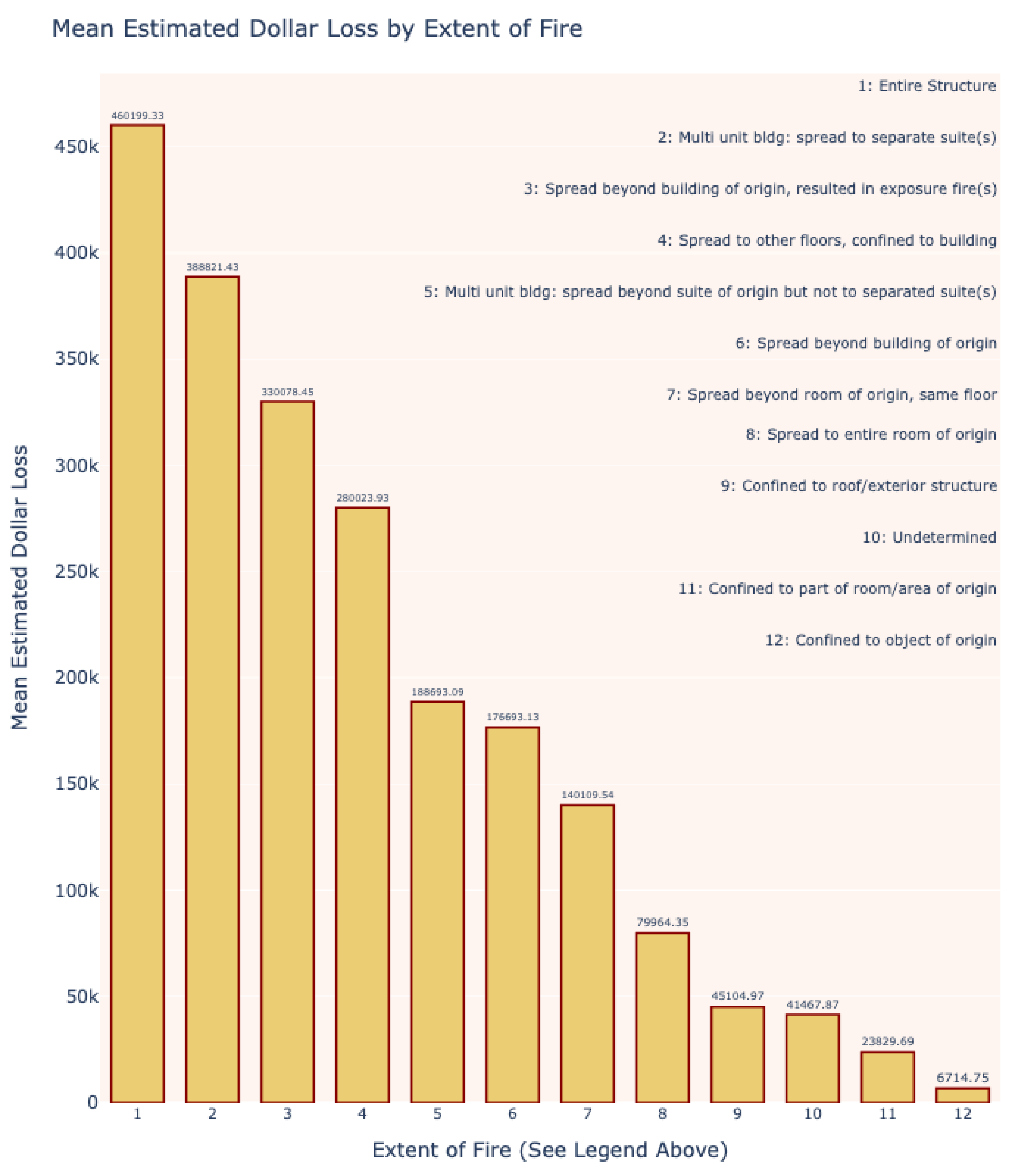

# Mean Estimated Dollar Loss by Fire Alarm Inoperation Reason

- Are fire alarms really effective?
- Not having a fire alarm detect a fire, is still better than not having an alarm system at all!




# Mean Estimated Dollar Loss by Sprinkler System Presence

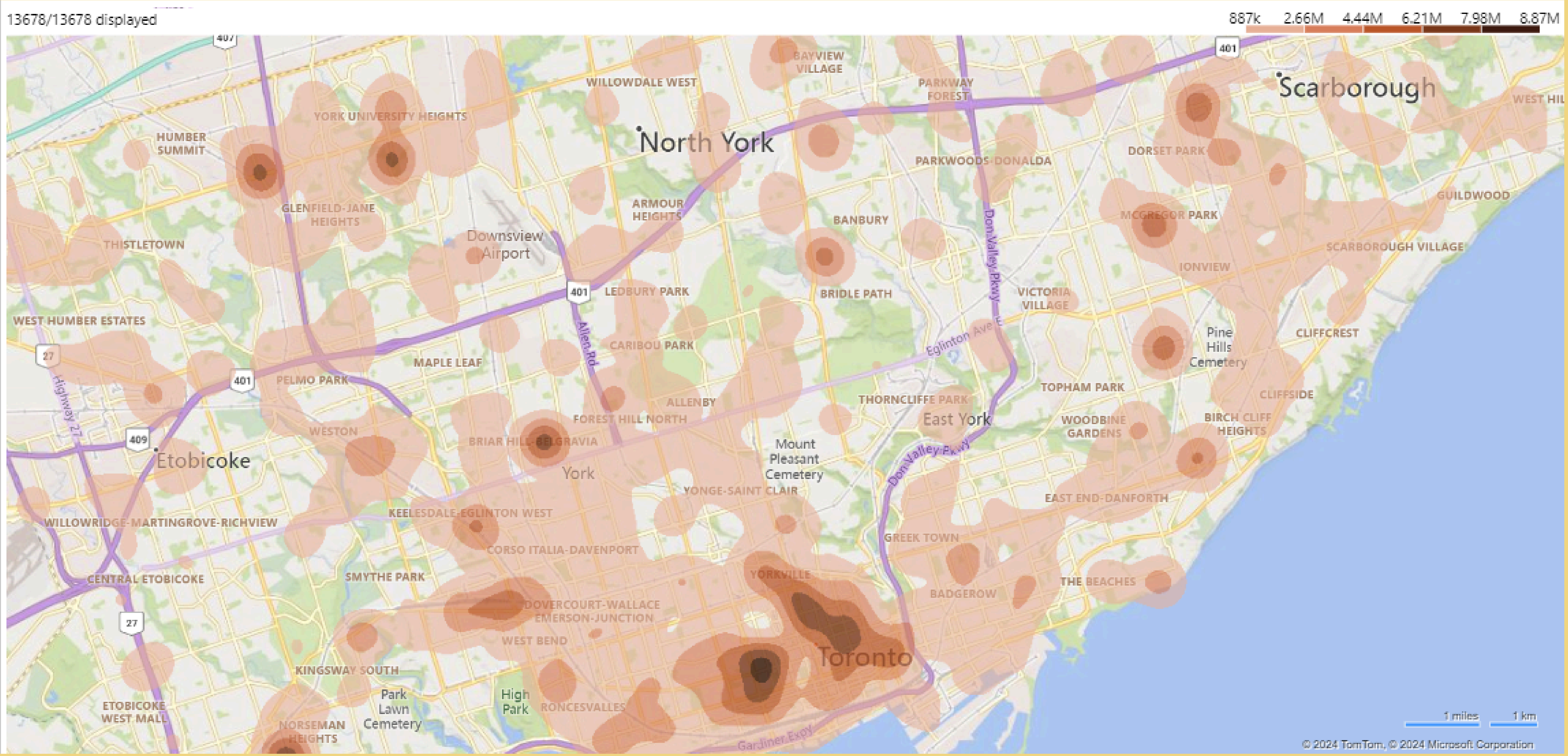
- Prevention is better than cure
- Risk assessment makes intuitive sense



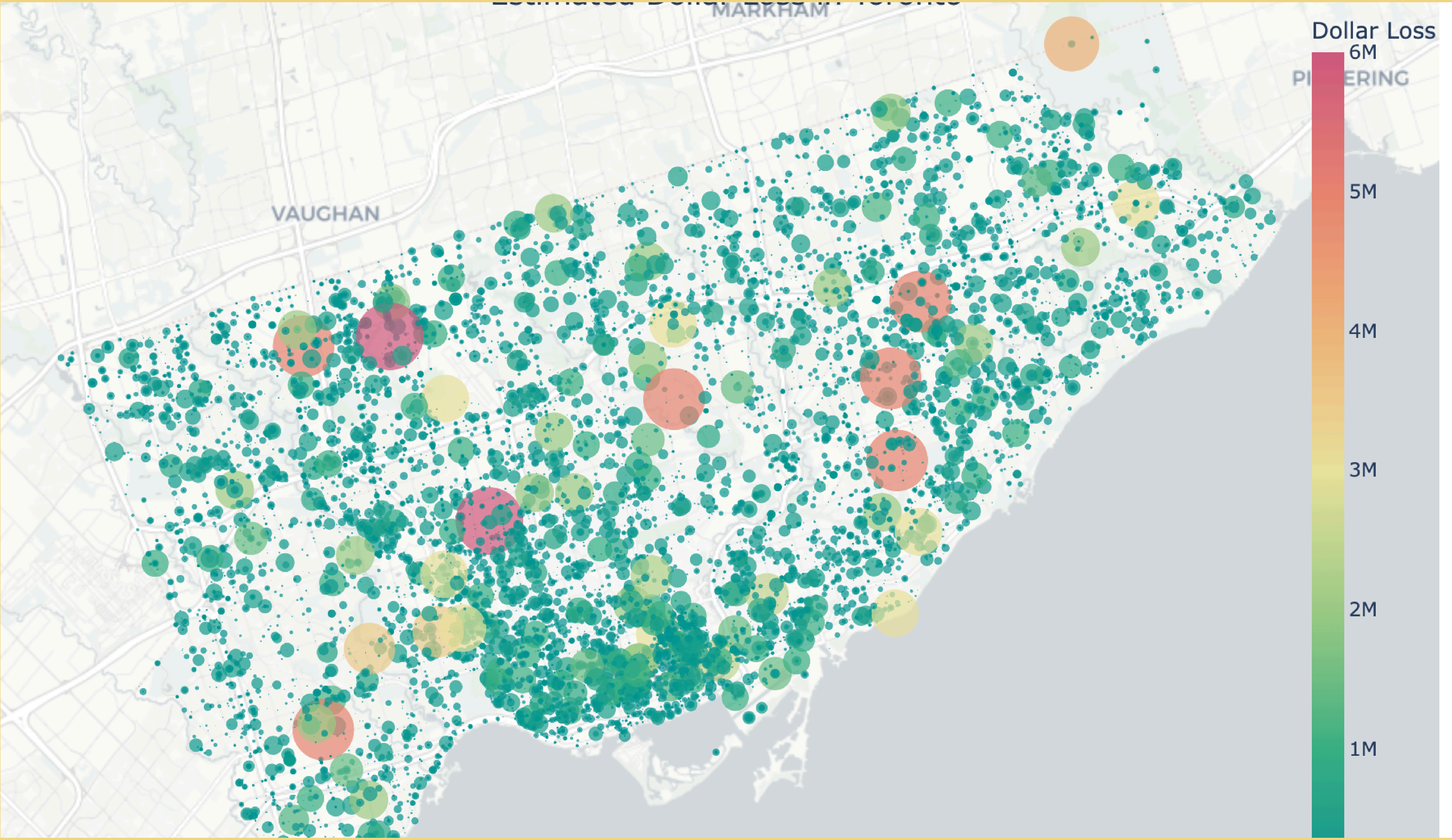

# Mean Estimated Dollar Loss by Extent of Fire

- Follows a clear downward trend with average estimated dollar loss.
- As far as the fire spreads, the more the loss is.

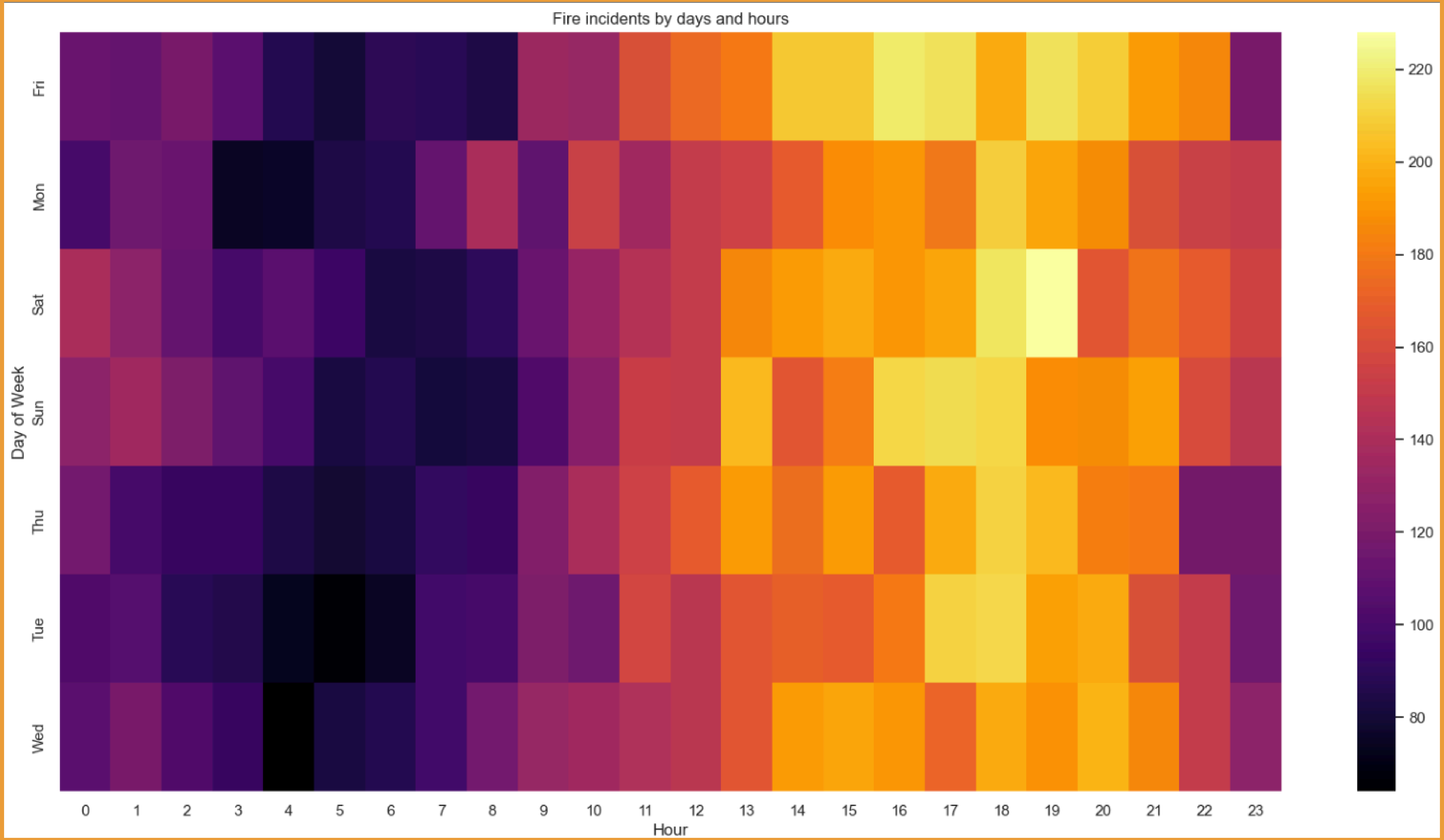
# HEATMAP OF INCIDENT FREQUENCY



# HEATMAP OF EST. DOLLAR LOSS

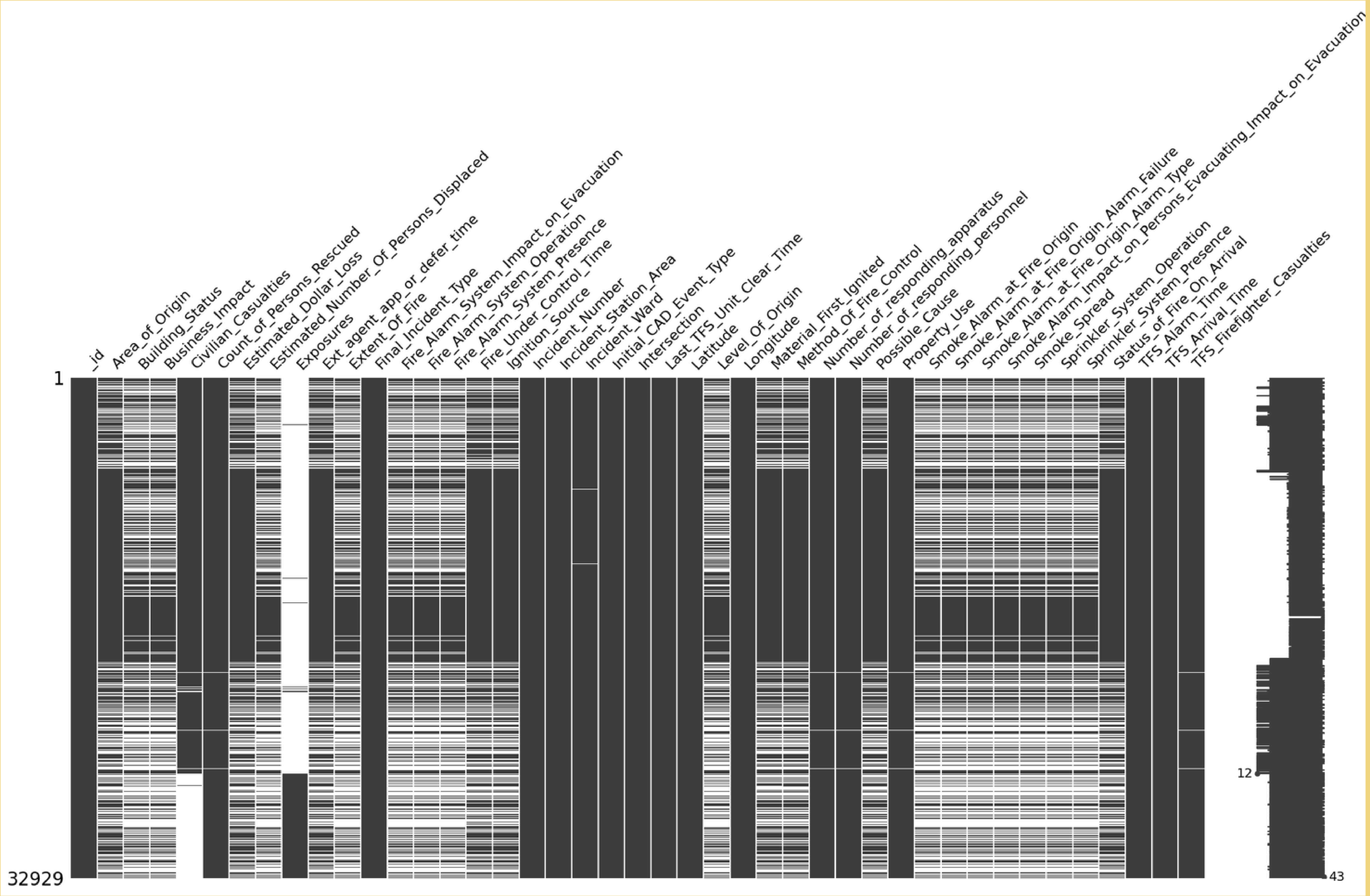


# HEATMAP OF FIRE INCIDENTS



# DATA PREPROCESSING

# Handling Missing Values




Removing Irrelevant coloumns

Removing NA values of the Estimated dollar loss col.

Imputing the data for NA values of the other columns

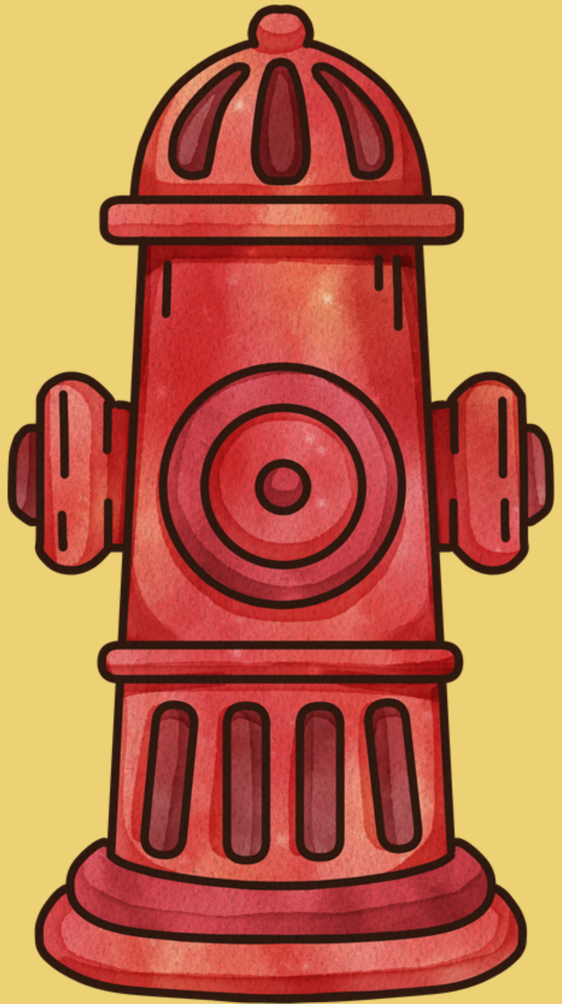
Removing outliers

# Handling Categorical data

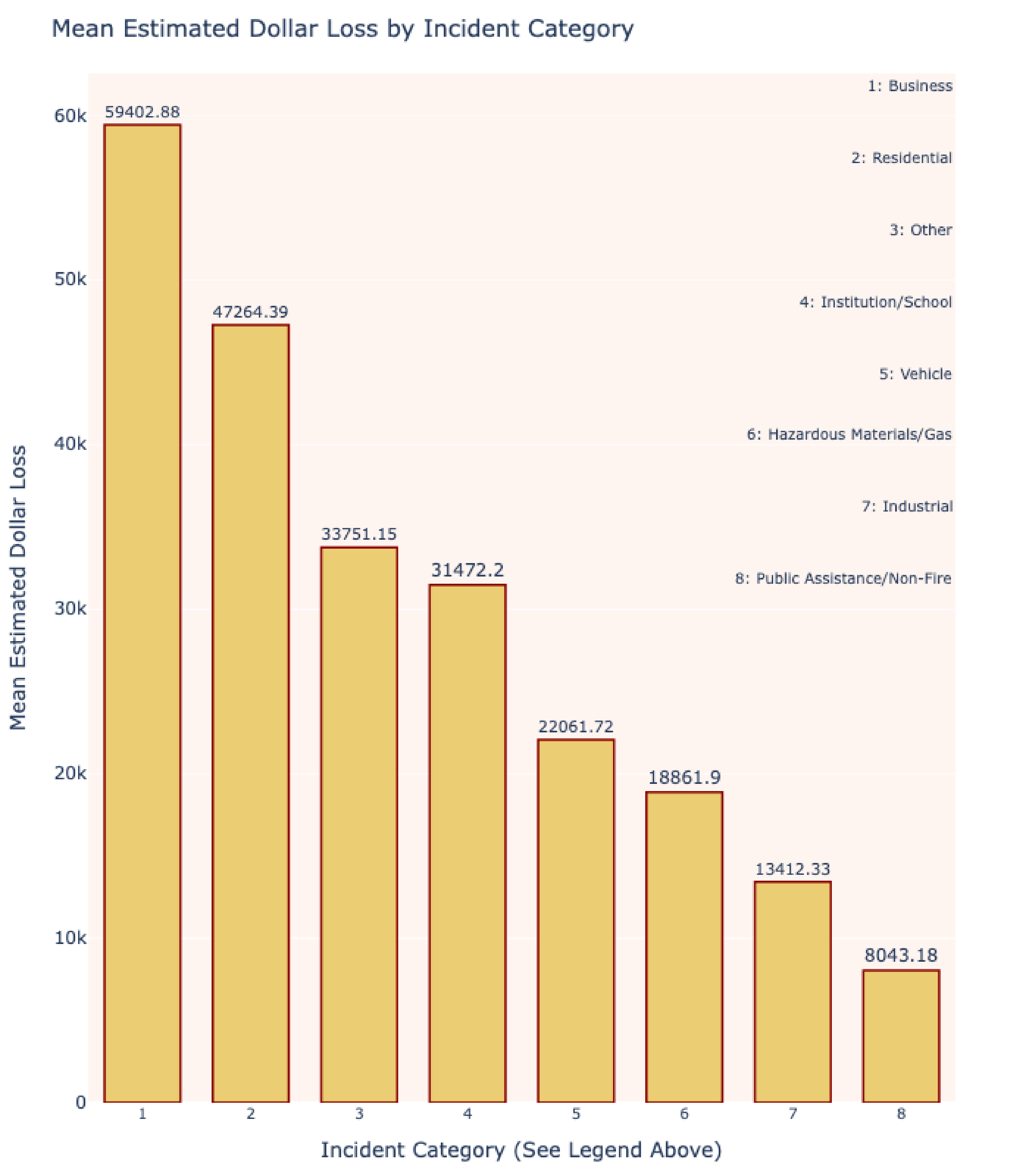
- Data Composition: The dataset predominantly consists of categorical variables, with each fire type (e.g., household fires) associated with its own set of categorical features, and a limited number of numerical columns.

- After cleaning the dataset and using hot encoding (dummifying) categorical variables: 700 columns * 23000 rows

- Using the Standard Incident Report Codes List of Ontario, we merged categorical variables into meaningful broader groups.



# Feature Engineering of the Categorical Variables

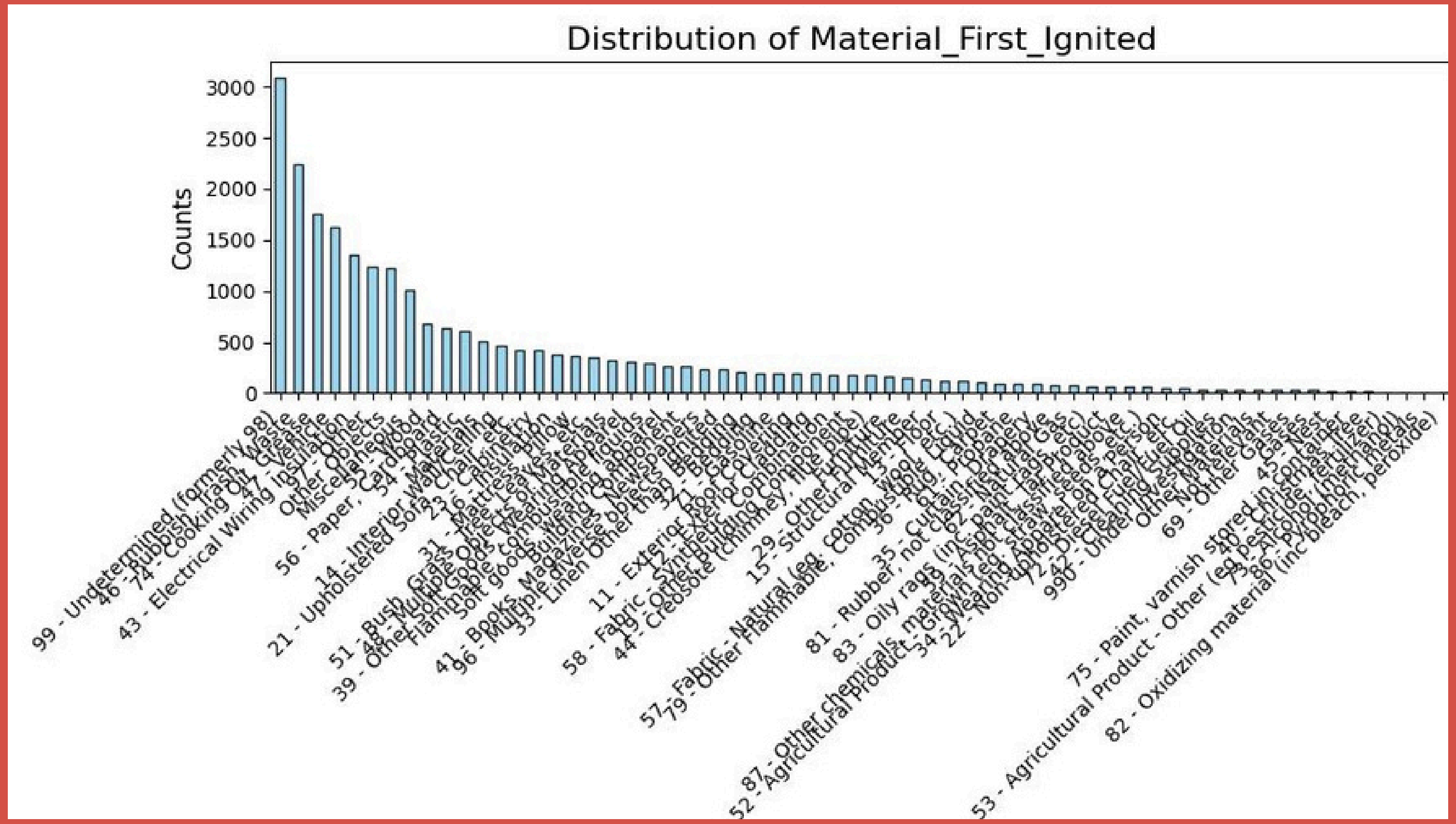


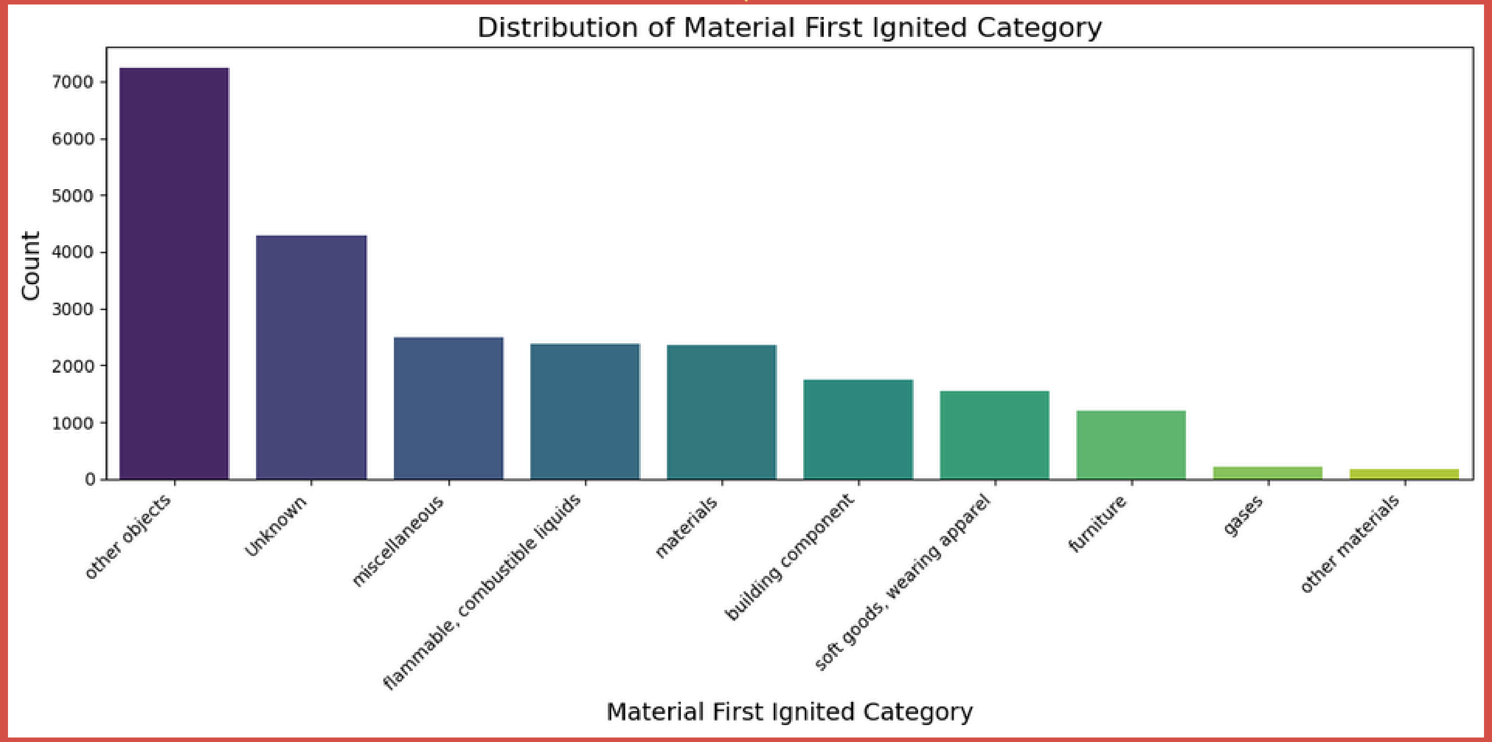

| | Initial_CAD_Event_Type |
|---|---|
| 0 | Vehicle Fire |
| 1 | Fire - Grass/Rubbish |
| 2 | Fire - Highrise Residential |
| 3 | Fire - Commercial/Industrial |
| 4 | Fire - Residential |
| 5 | Vehicle Fire |
| 6 | Fire - Residential |
| 7 | Alarm Highrise Residential |
| 8 | Alarm Commercial/Industrial |
| 9 | Vehicle Fire |

...

| Incident_Category |
|---|
| Other |
| Vehicle |
| Residential |
| Business |
| Industrial |
| Public Assistance/Non-Fire |
| Institution/School |
| Hazardous Materials/Gas |

# Feature Engineering of the Categorical Variables




```
Material_First_Ignited_dict = {
  "building component": [
    "11 - exterior roof covering",
    "12 - exterior cladding (excluding roof)",
    "13 - floor",
    "14 - interior wall/ceiling",
    "15 - structural member",
    "16 - insulation",
    "19 - other building component",
    "building component"
  ],
  "furniture": [
    "21 - upholstered sofa, chair, etc.",
    "22 - non-upholstered chair, etc.",
    "23 - cabinetry",
    "29 - other furniture",
     "furniture"
  ],
  "soft goods, wearing apparel": [
    "31 - mattress, pillow",
    "32 - bedding",
    "33 - linen other than bedding",
    "34 - wearing apparel on a person",
    "35 - curtain, drapery",
    "36 - rug, carpet",
    "39 - other soft goods, wearing apparel",
    "soft goods, wearing apparel"
  ],
  "other objects": [
    "40 - christmas tree",
    "41 - books, magazines, newspapers",
    "42 - cleaning supplies",
    "43 - electrical wiring insulation",
    "44 - creosote (chimney, flue pipe)",
    "45 - nest",
    "46 - rubbish, trash, waste",
    "47 - vehicle",
    "48 - multiple objects or materials",
    "other objects"
  ],"75 - stored paint, varnish",
    "79 - other flammable, combustible
liquid",
    "flammable, combustible liquids"
  ],
```

# PREDICTING MODELS

# Machine Learning Models to Predict ‘Estimated dollar Loss’

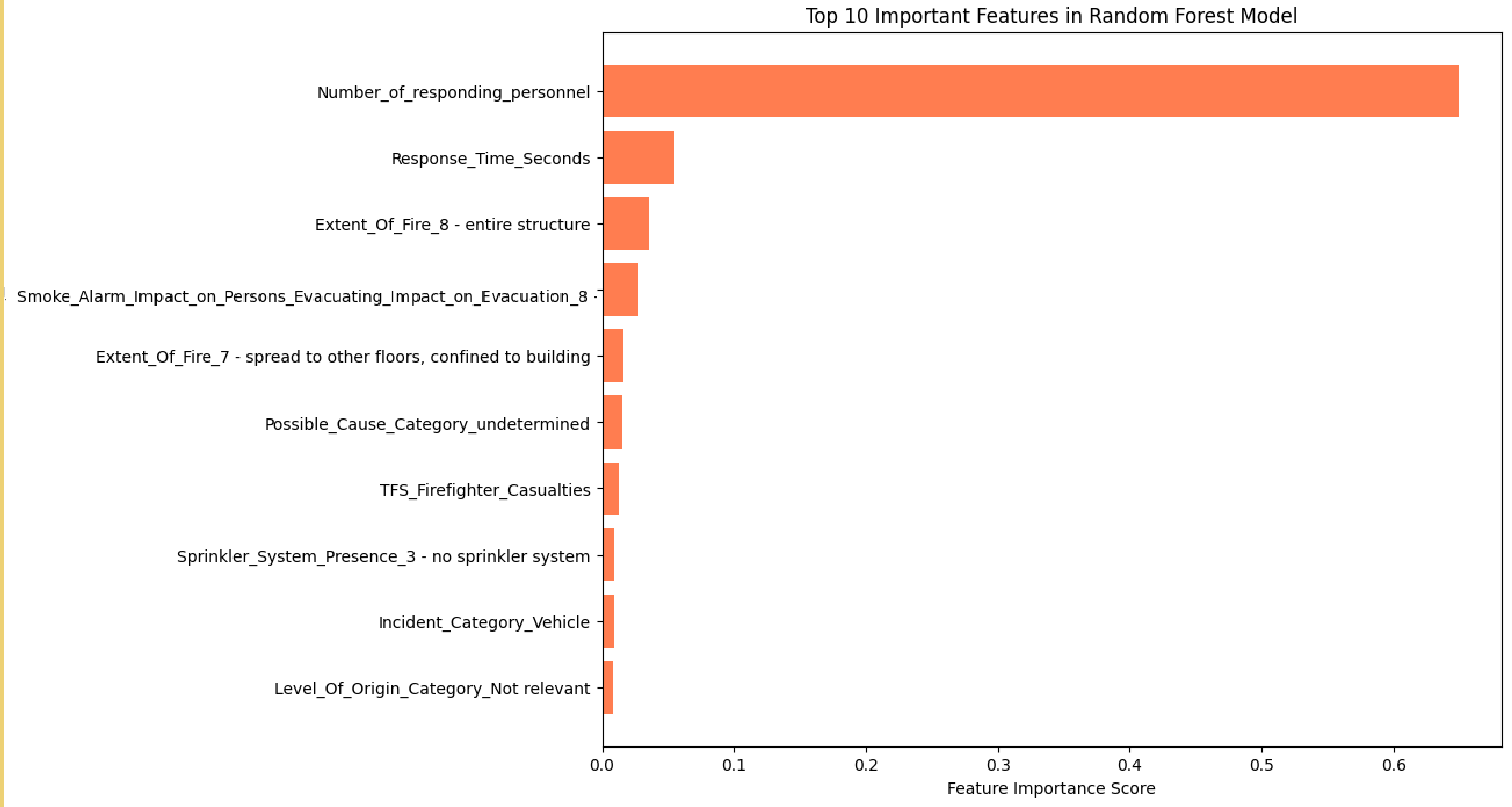
- Random Forest (Parallel)
- Gradient Boosting (Sequential)

Reason for choosing these models:

- Robust to feature scaling, no need to normalize or standardize the input features.
- Can handle categorical variables (though pre-processing may vary by implementation).
- Calculate feature importance, allowing insight into which variables contribute most to predictions.

# Random Forest results

Best Parameters: {'max_depth': 50, 'min_samples_leaf': 5, 'min_samples_split': 3, 'n_estimators': 300}




Cross-Validation RMSE: 101,866.911

Test RMSE: 145,965.201

Normalized CV RMSE Range: 0.020

Normalized CV RMSE Mean: 3.550

# Predicting Loss by Fire Type

| Incident_Category | CV_RMSE | Normalized_CV_RMSE_Range | Normalized_CV_RMSE_Mean | Test_Set_RMSE |
|---|---|---|---|---|
| Vehicle | 52009.9 | 0.017337 | 2.35747 | 80806.93 |
| Other | 140095.0 | 0.023492 | 4.17629 | 127837.788 |
| Residential | 96595.3 | 0.027599 | 2.04372 | 167138.565 |
| Business | 193123.0 | 0.038625 | 3.25107 | 306014.215 |
| Public Assistance/Non-Fire | 26445.3 | 0.132226 | 3.28791 | 23410.489 |
| Industrial | 42632.5 | 0.142108 | 3.17861 | 30550.319 |
| Institution/School | 20019.2 | 0.008008 | 0.63609 | 403018.61 |
| Hazardous Materials/Gas | 7090.87 | 0.023636 | 0.37594 | 111711.235 |

- We can see that the model is much better at predicting the loss of these types of fires: ‘Public Assistance’, ‘Institution/School’ & ‘Hazardous Materials/Gas’
- For the ‘Other' category, which contains heterogeneous data points, the cross-validation error is large.

# Random Forest with FI & Log-Trsnf

- Looked at the feature importance
- Retrain model RF Model using top 25 important features
- Compared the results with the original model.
- Best Parameters:
  {'max_depth': 10, 'min_samples_leaf': 4,
  'min_samples_split': 2, n_estimators': 250}

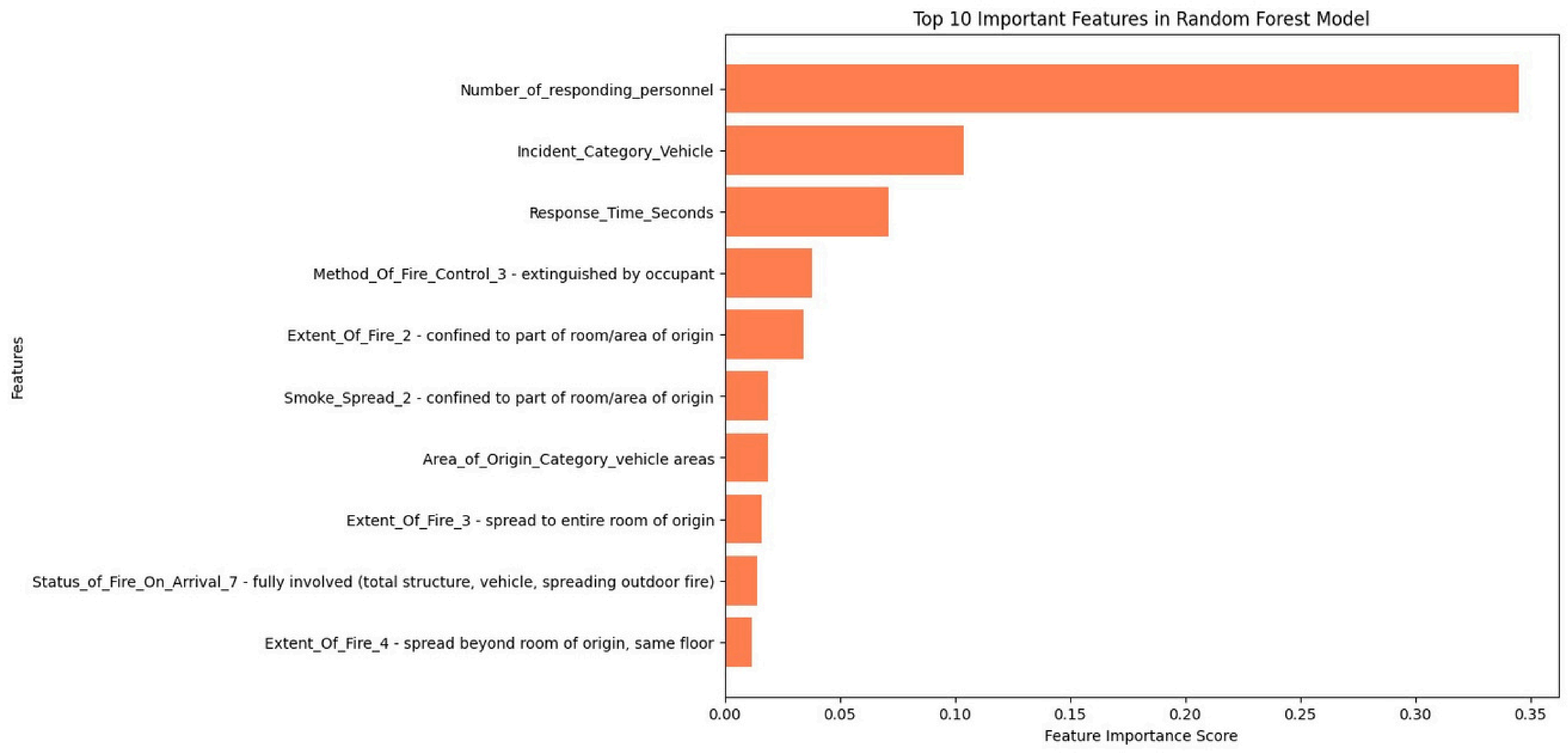
Cross-Validation
RMSE:
85,231.863

- Afterwards, we applied a log transformation to the "Estimated_Dollar_Loss" column to stabilize variance and normalize the distribution
- Trained a Random Forest model using the transformed data to predict the log-transformed dollar loss

Test RMSE:
105,291.394

# RF with FI & Log-Trsnf

Best Parameters: {'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 200}




CV RMSE: 92,893.09

Test RMSE: 81,351

Normalized CV RMSE Range: 0.114

Normalized CV RMSE Mean: 0.178

# Gradient Boosting with FI

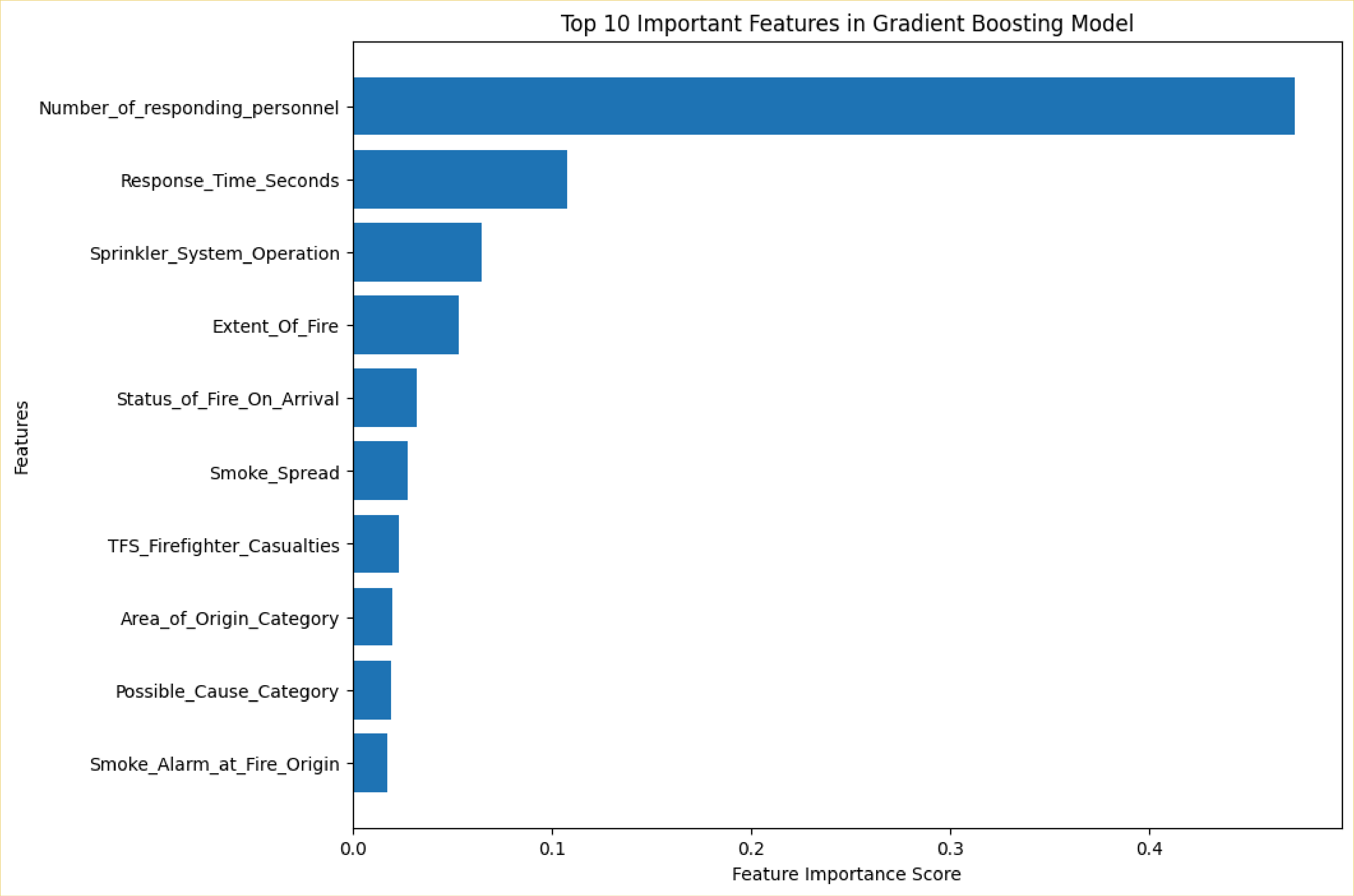
- Applied a log transformation to the "Estimated_Dollar_Loss" column to stabilize variance and normalize the distribution.

- Trained GB model using top 25 features obtained earlier.

- Compared the results with the original RF model and RF with Feature Importance.

Best Parameters: {'max_depth': 10, 'min_samples_leaf': 3, 'min_samples_split': 4, 'n_estimators': 150, learning_rate = 0.01}

# Gradient Boosting Results




Cross-Validation RMSE: 77,278.82

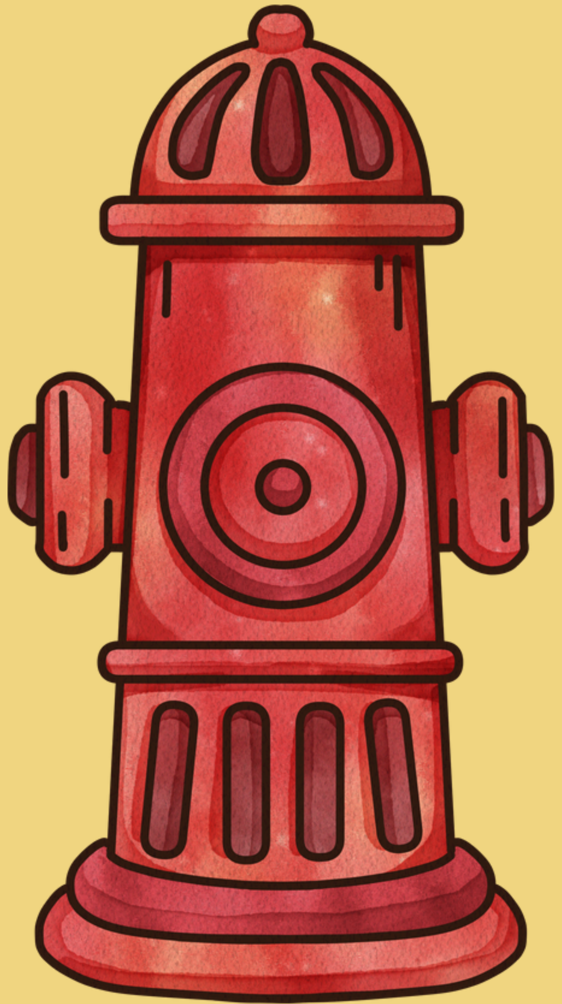
Test RMSE: 87,365.308

# All Model Results

| MODEL | CV RMSE | Test RMSE |
|---|---|---|
| Random Forest | 101,866.911 | 145,965.201 |
| Random Forest with FI | 85,231.863 | 105,291.394 |
| Random Forest with FI and Log- transformation | 81,351.98 | 92,893.09 |
| Gradient Boosting with FI | 77,278.82 | 87,365.308 |
| Gradient Boosting with FI and Log-transformation | 65,233.27 | 73,234.97 |

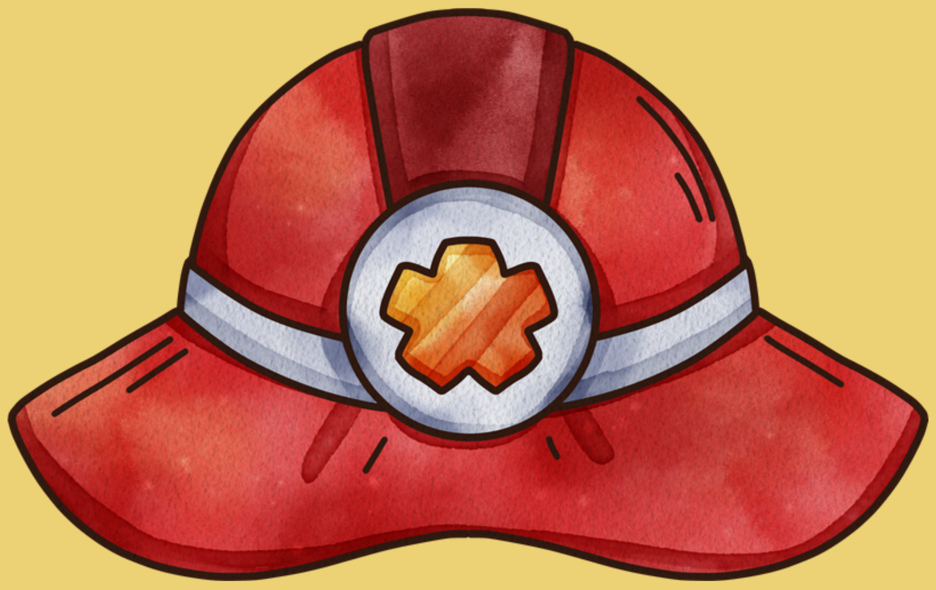
# Business Suggestions

- Have Carbon Dioxide Extinguishers at our places, especially for the fires due to electric equipments.
- Fully functional Sprinkler system should be present especially in the unattended areas like storage area, building area/support facilities, building vehicle areas, etc.
- Faster response times are strongly associated with reduced damage, highlighting the importance of efficient emergency services and rapid mobilization.

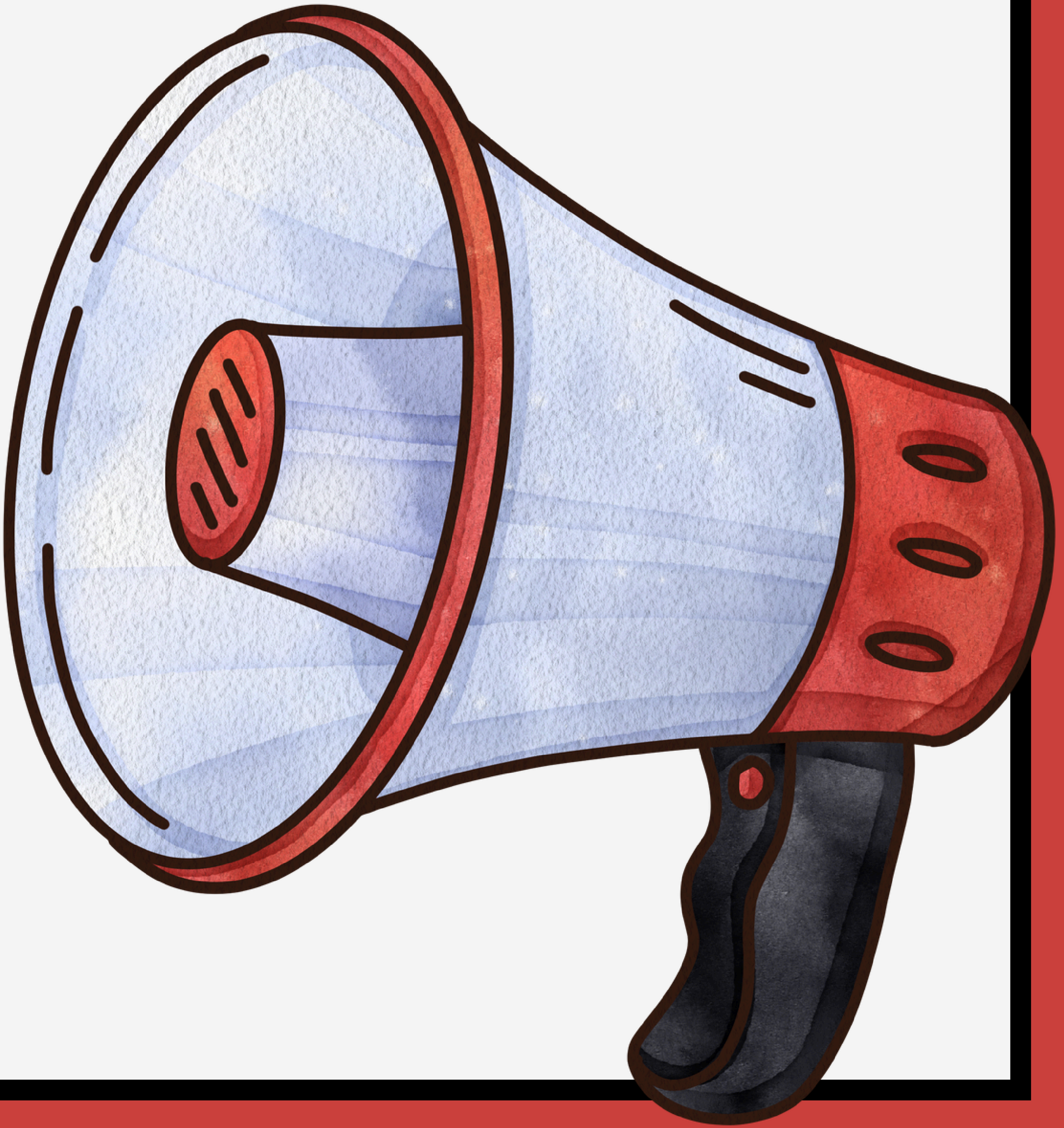


# References

- Dataset:
  https://open.toronto.ca/dataset/fire-incidents/
- Standard Incident Report Codes:
  https://www.toronto.ca/ext/open_data/catalog/data_set_files/ofmcodes2009.pdf

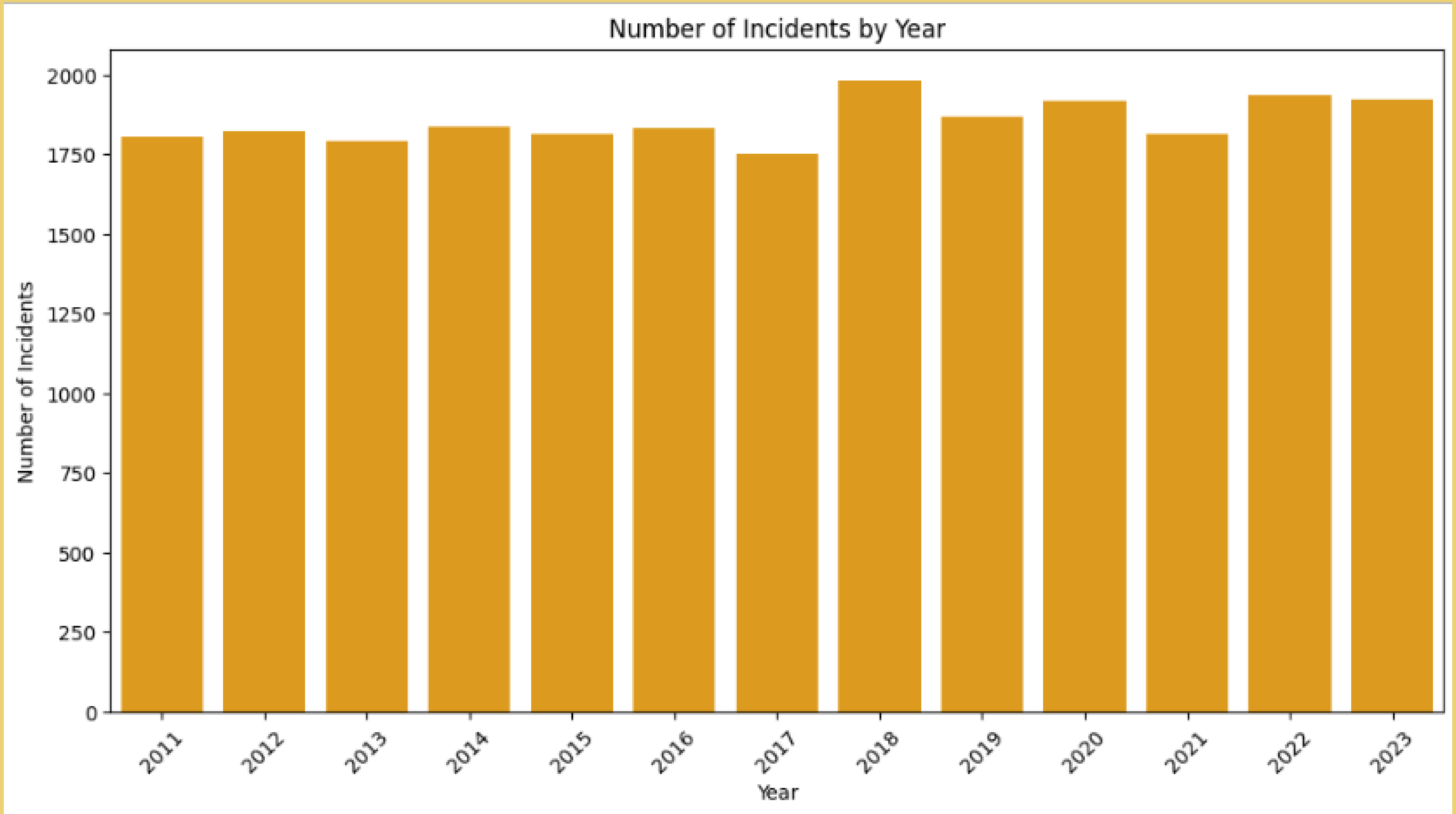


# Questions or Comments?

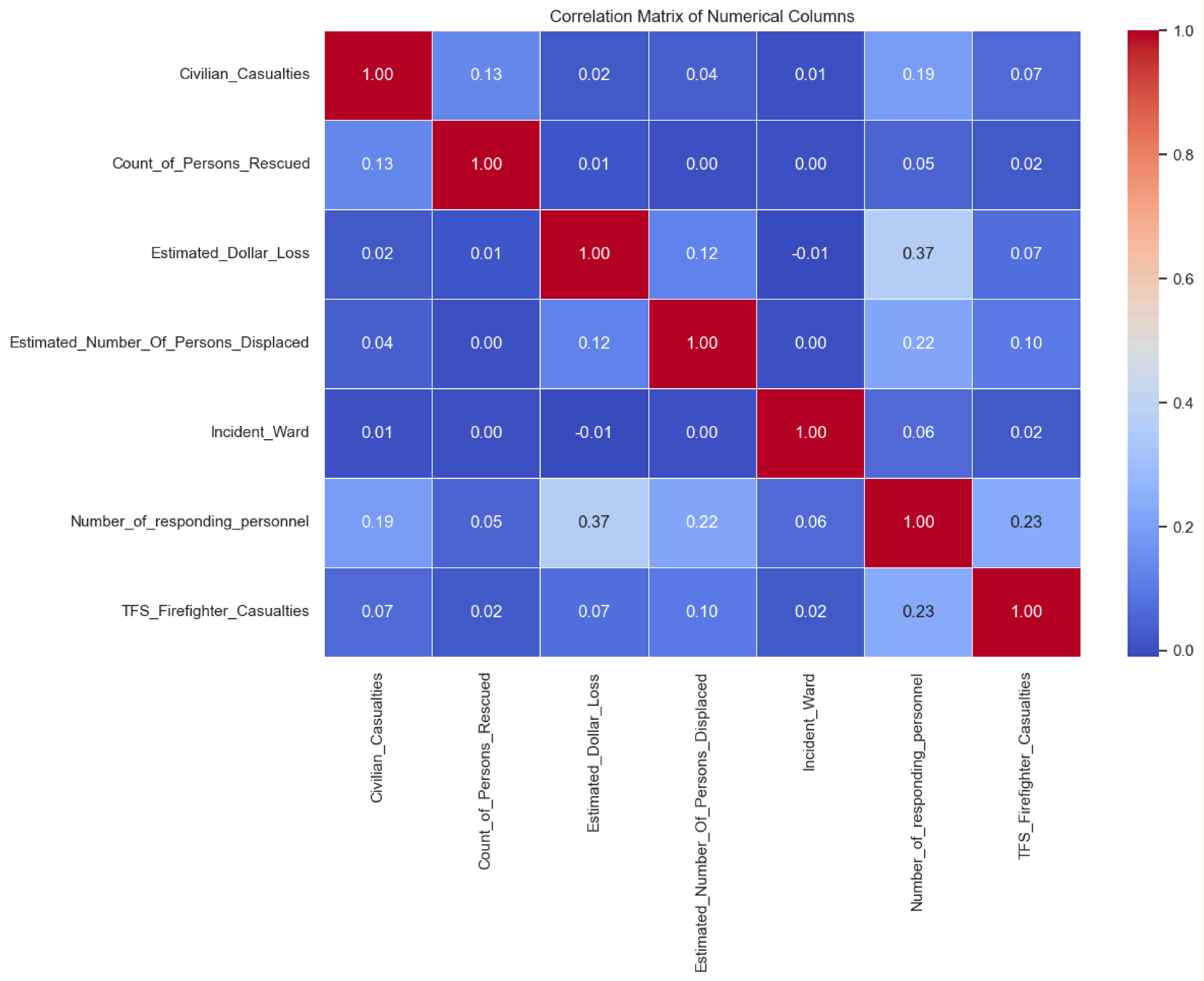


# Appendix

# NUMBER OF FIRE INCIDENTS



# Correlation Matrix for some variables




- Certain variables like Latitude and Longitude are collinear to incident ward, so we dropped them.
- Number of responding personnel stands out as a good predictor for estimated dollar loss.
- Estimated number of people displaced also has a fair positive correlation with number of responding personnel, because more personnel are required to deal with more destructive fires.

# Progress in our model performances

```
Best Parameters: {'max_depth': 15, 'min_samples_leaf': 6, 'min_samples_split': 3, 'n_estimators': 200}
Best Cross-Validation RMSE: 125302.88121948186
Feature Importances:
                                   Feature  Importance
15            Number_of_responding_apparatus           0
16           Number_of_responding_personnel           0
18                             Property_Use           0
28                    Response_Time_Seconds           0
4                            Extent_Of_Fire           0
1                            Area_of_Origin           0
26              Status_of_Fire_On_Arrival           0
27               TFS_Firefighter_Casualties           0
19                Smoke_Alarm_at_Fire_Origin           0
12                          Level_Of_Origin           0
17                           Possible_Cause           0
23                             Smoke_Spread           0
9                    Fire_Under_Control_Time           0
8                 Fire_Alarm_System_Presence           0
11                  Last_TFS_Unit_Clear_Time           0
10                          Ignition_Source           0
13                    Material_First_Ignited           0
6       Fire_Alarm_System_Impact_on_Evacuation           0
3                Ext_agent_app_or_defer_time           0
0                        Incident_Category           0
```

- Many scattered categorical variables.
- Used Ordinal encoding for categorical variables