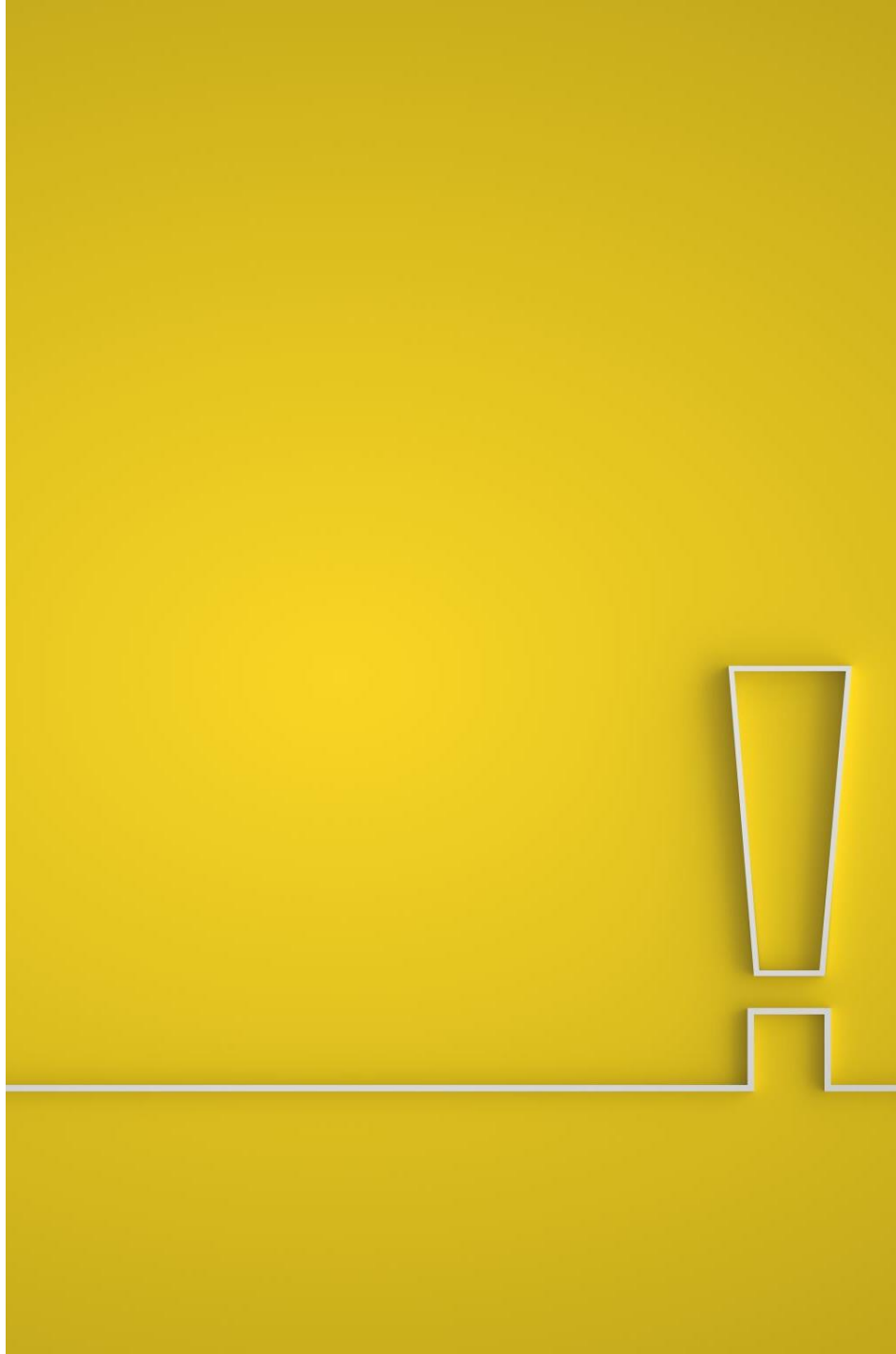
High Fire Risk Areas in Montreal Prediction

- YCBS 299 - TEAM 1
- AHMED IBRAHIM,
- EUNSEO LEE,
- PRADIP KUMAR,
- PRANAVKUMAR PATHAK

Business Problem

- Safety of people living in Montreal.
- Predict areas with high fire risk in the coming months
  - Improve inspection plans
  - Allocate resources
  - Ensure proper maintenance
- Minimize impact of fire



Timetable

Data Exploration

Feature Engineering

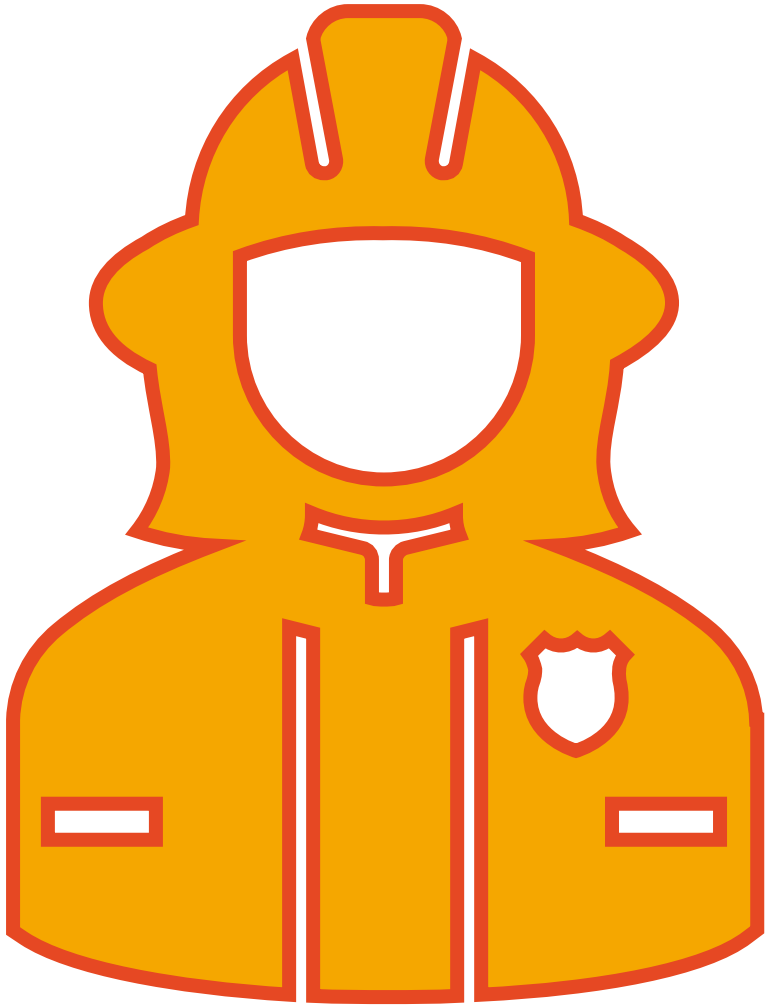
Models

Model Performance and Evaluation

Conclusion

Datasets

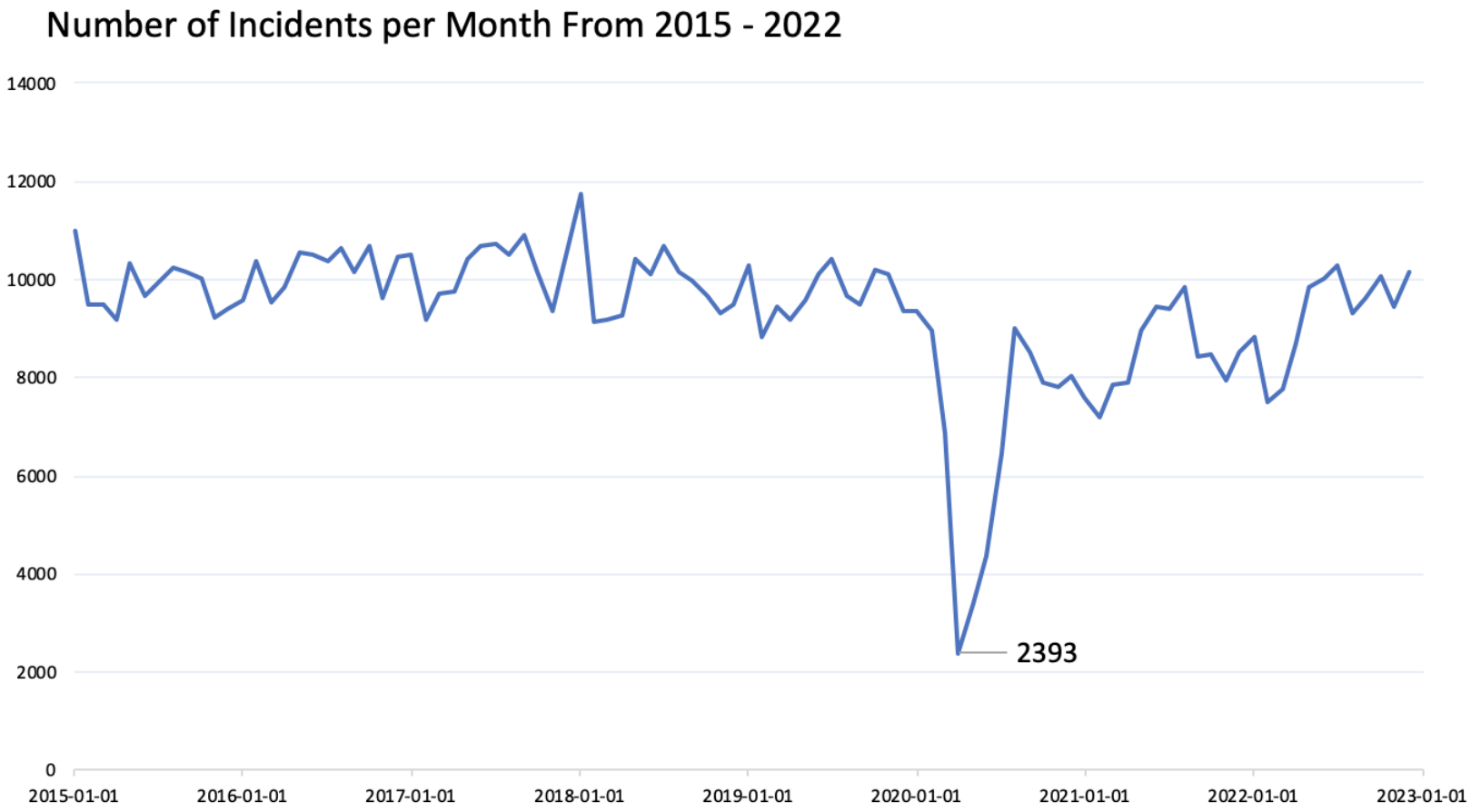
- Montreal Boroughs & Affiliated Cities,
- Crimes,
- Incidents,
- Property Assessments,
- Demographics,
- Weather.



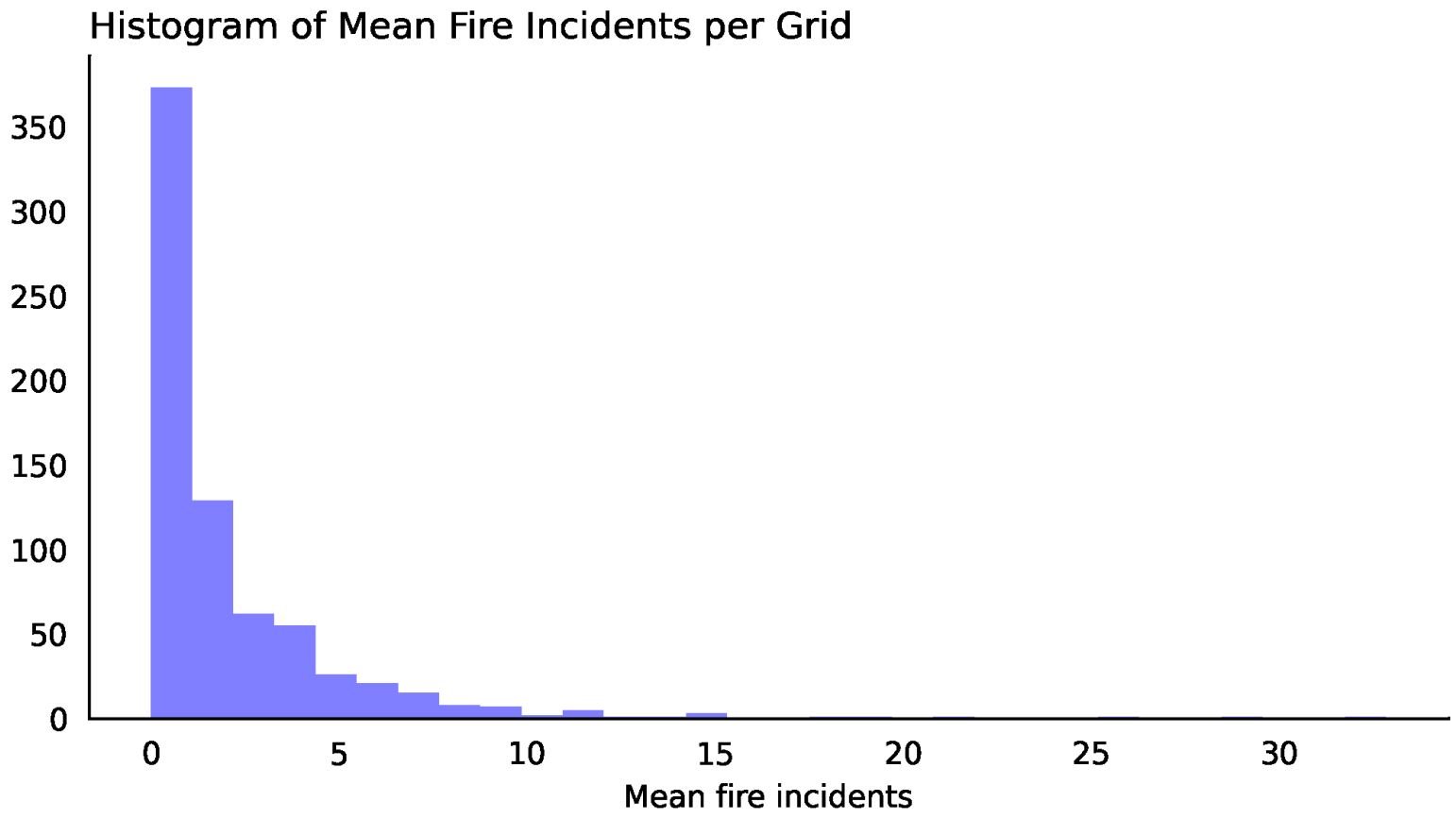
Pattern

Number of incidents range between **7000 to 12000**

Exception at 2020 - **dropping to 2393**



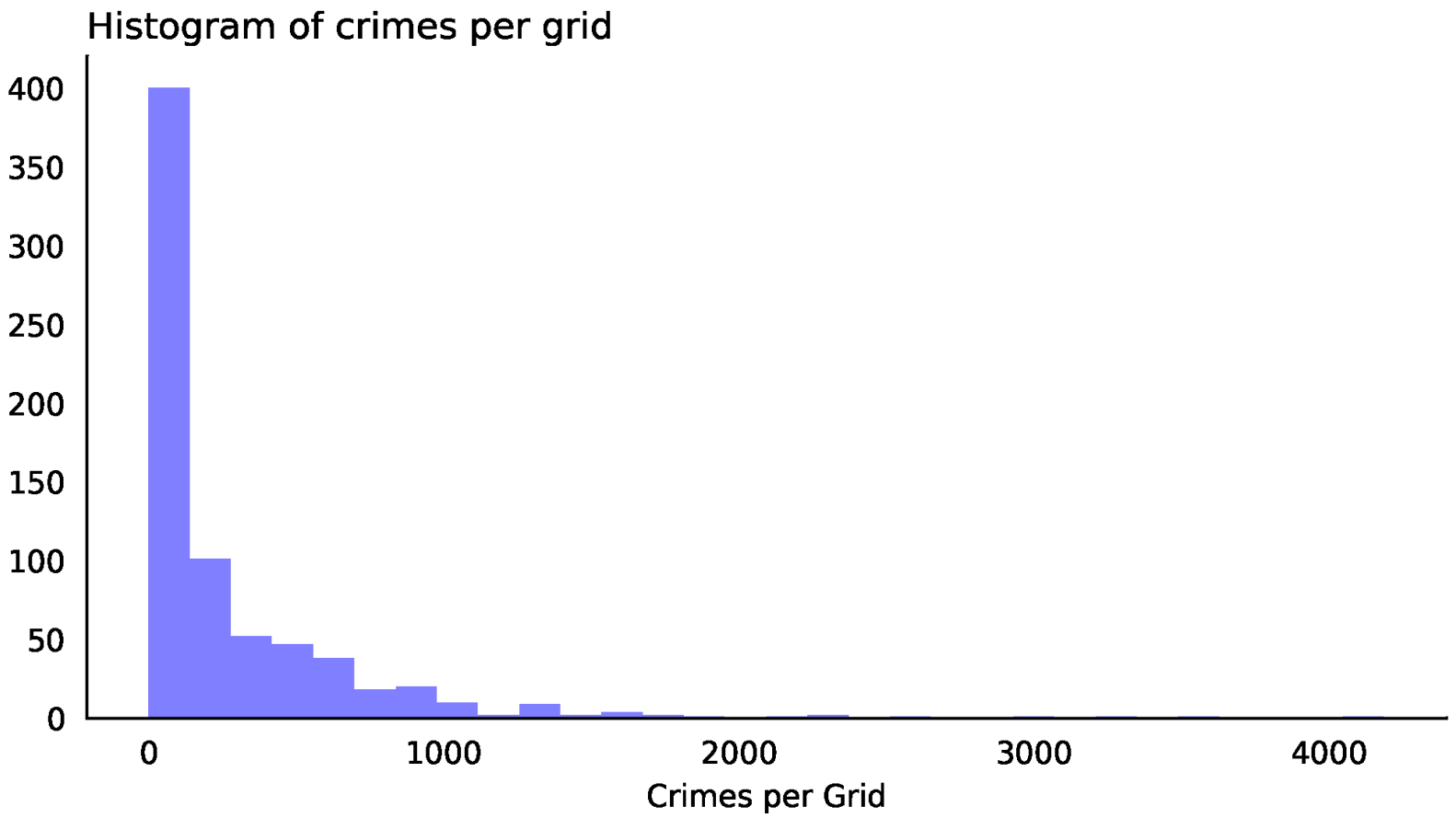
Distribution - Incidents



Distribution of average fire incidents that have occurred since 2015 – 2022 per grid

Data is skewed to the left - shows most of areas had **close to 0-3 fire incidents** over the 8 years

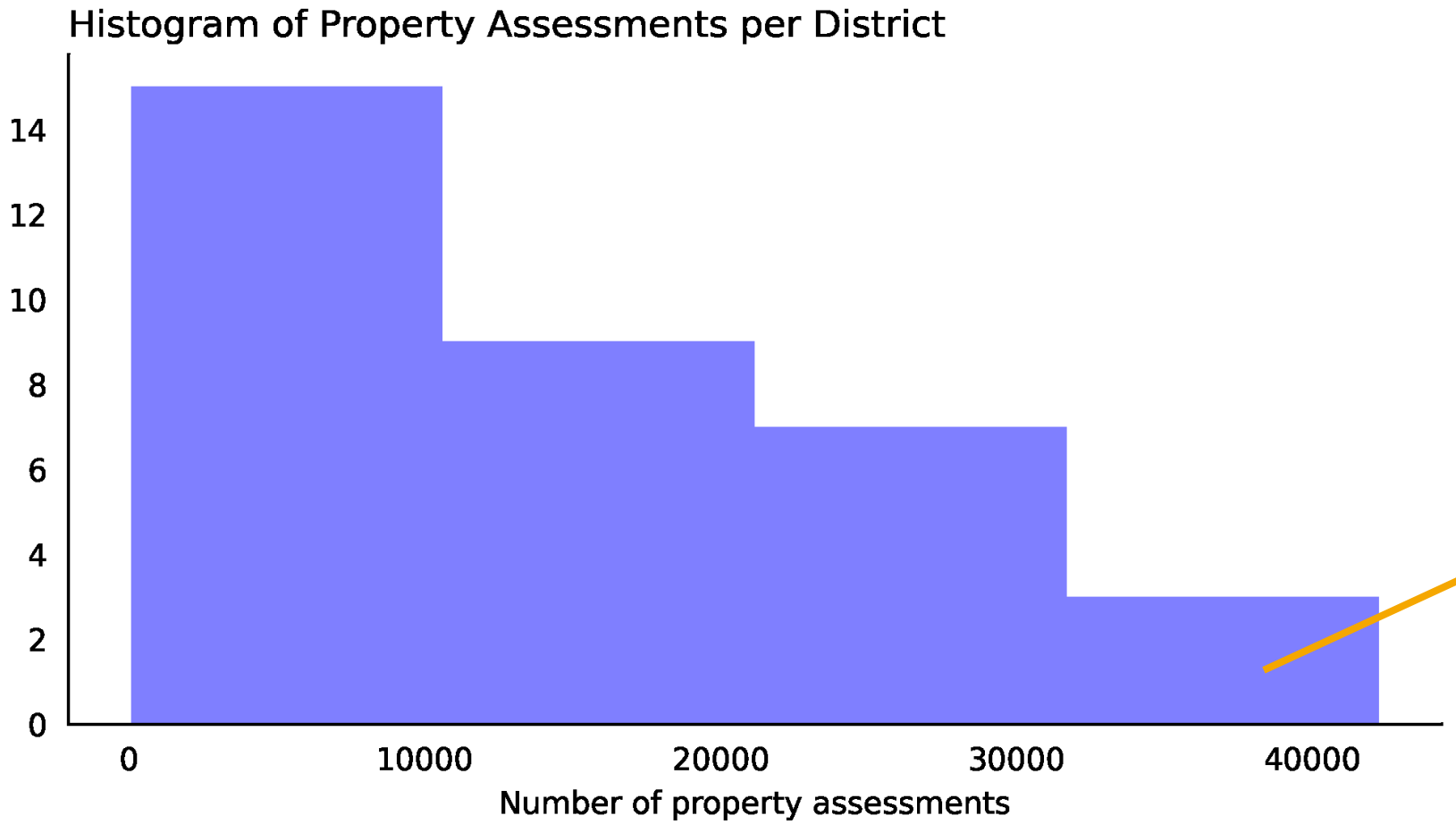
Distribution - Crime



Distribution of number of criminal acts that have occurred since 2015 – 2022 per grid

Data is skewed to the left - shows most of areas had **close to <130 crimes** over the 8 years

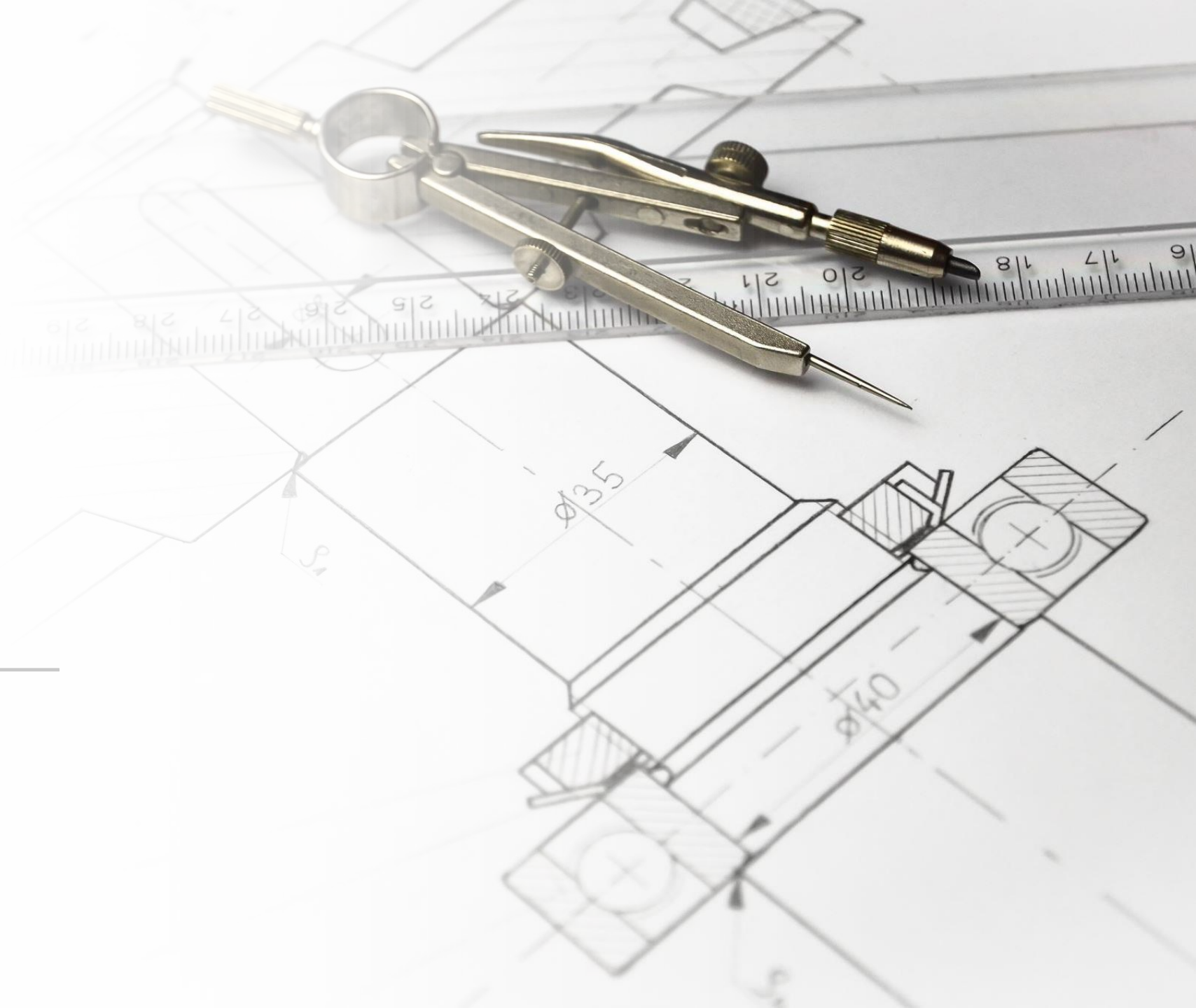
Distribution – Property Assessment



Distribution of property assessments per borough/cities in Montreal from 2015 to 2022.

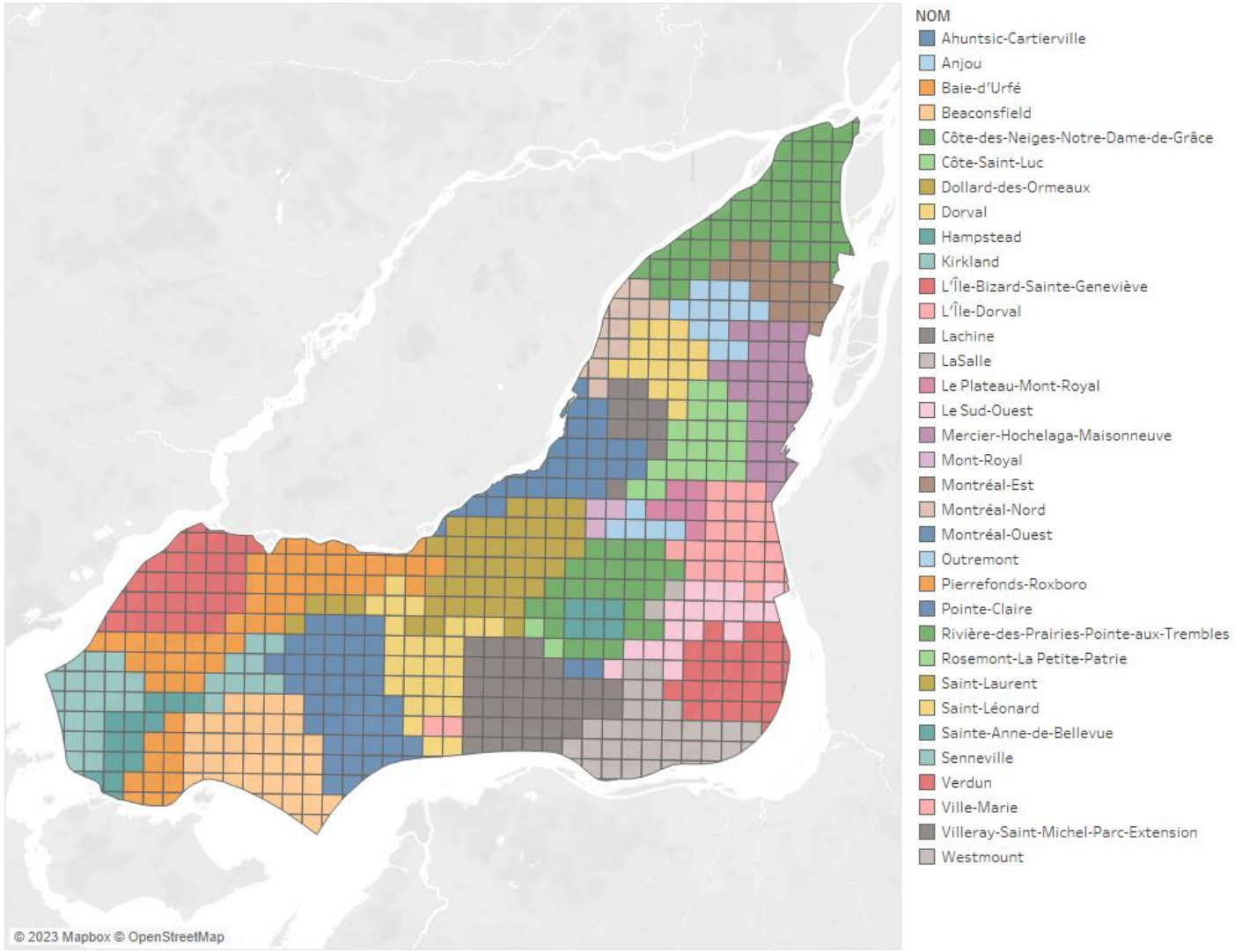
Ville-Marie
Rivière-des-Prairies-Pointe-aux-Trembles
Rosemont-La Petite-Patrie
Mercier-Hochelaga-Maisonneuve

Feature Engineering



Grid Strategy

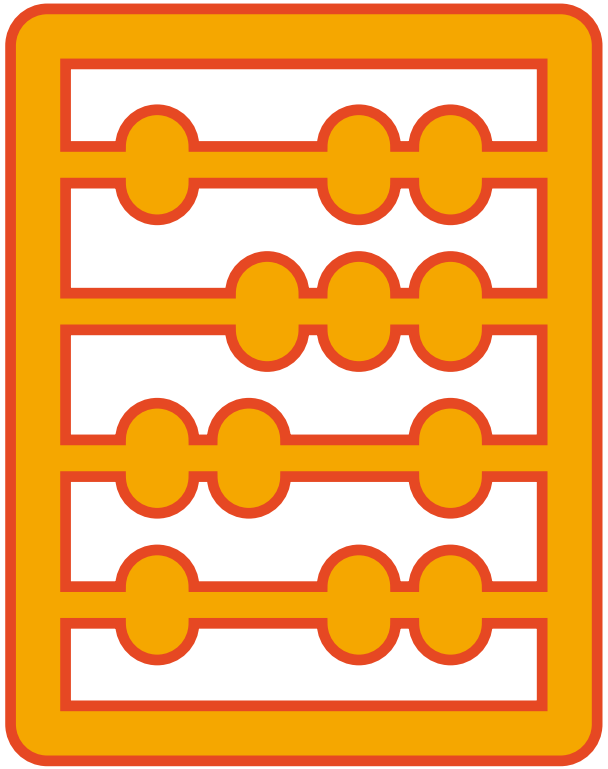
The Montreal shapefile is tessellated into square grids (~1 km^2). Each grid is assigned a unique ID, district, and area.



Data Integration & Aggregation

- Montreal segregated **714 grids** (1km^2)

- All features are integrated per grid
  - Temporal: **month/quarter**
  - Static: **boroughs and cities**



Feature generation

Datasets

- Montreal Boroughs & Affiliated Cities,
- Crimes,
- Incidents,
- Property Assessments,
- Demographics,
- Weather.

Group by
grid ID,
month,
district ….

Aggregate using
mean, sum, count,
max ….

Features

- Sum of units deployed,
- Crime count,
- Population density,
- mean temperature,
- Precipitation sum,
- ….

Fire-Risk Score

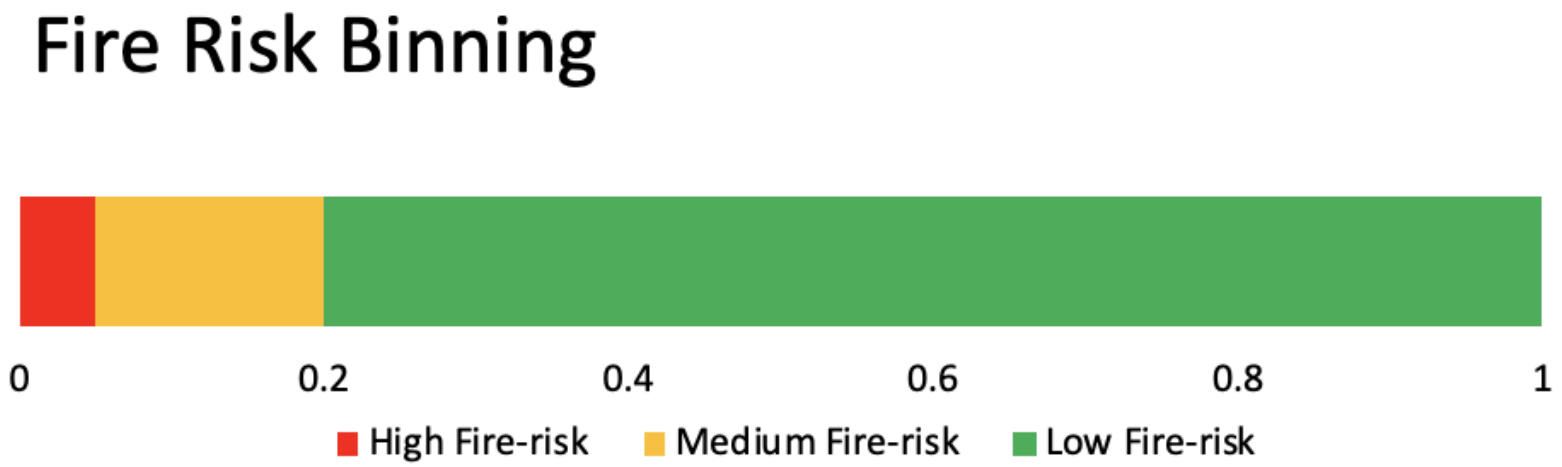
How do we classify fire-risk?

1. Count monthly fire incidents and average quarterly count per grid
2. Rank grids per month
3. Handle tie breaker using quarterly rank

Predicting Target

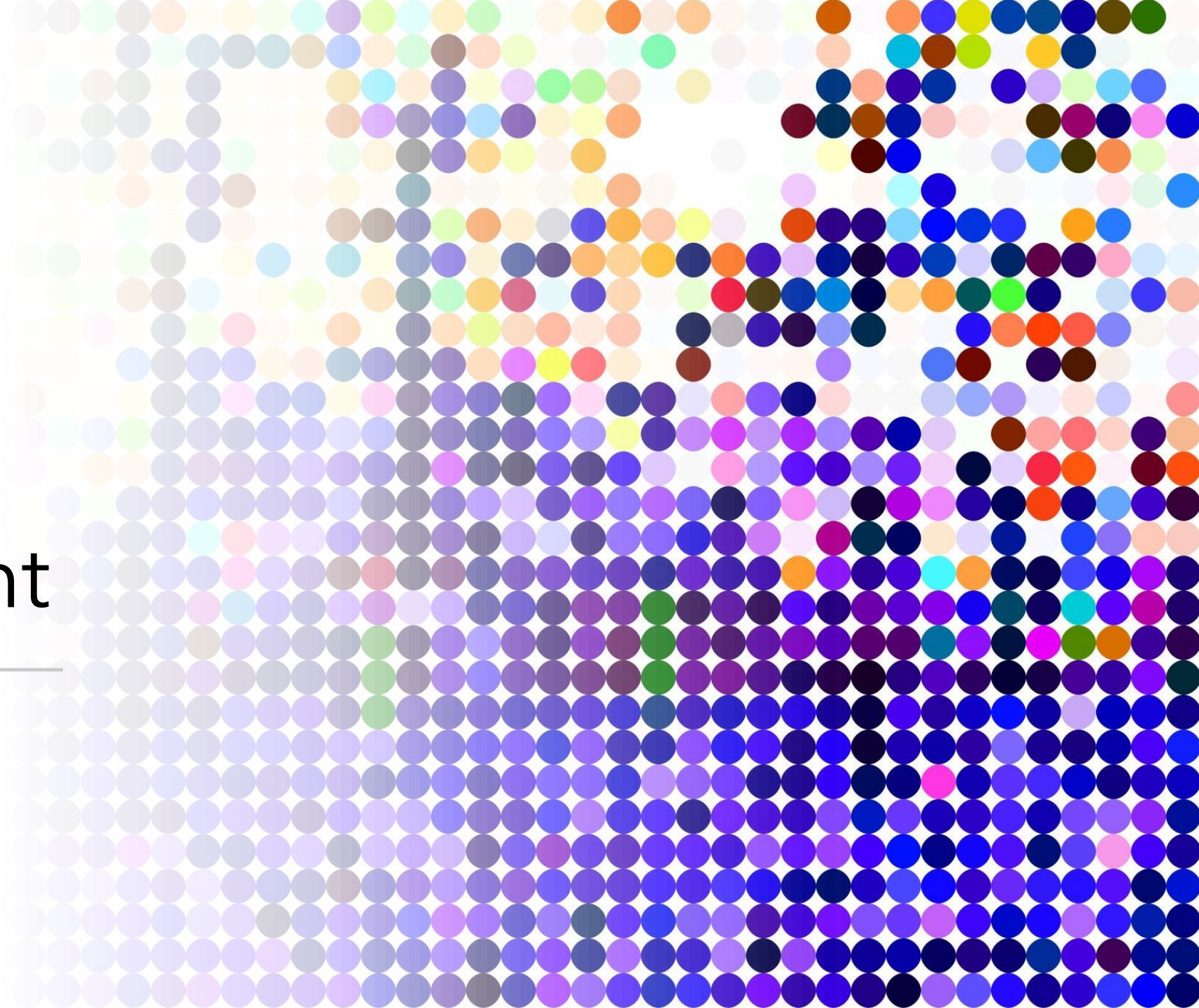
Classification

- High Fire-risk – 5%
- Medium Fire-risk – 15%
- Low Fire-risk – 80%



Definition of each class: High, Medium, Low

Model Development



Models

Comparing ensemble models

Decision Tree (baseline)

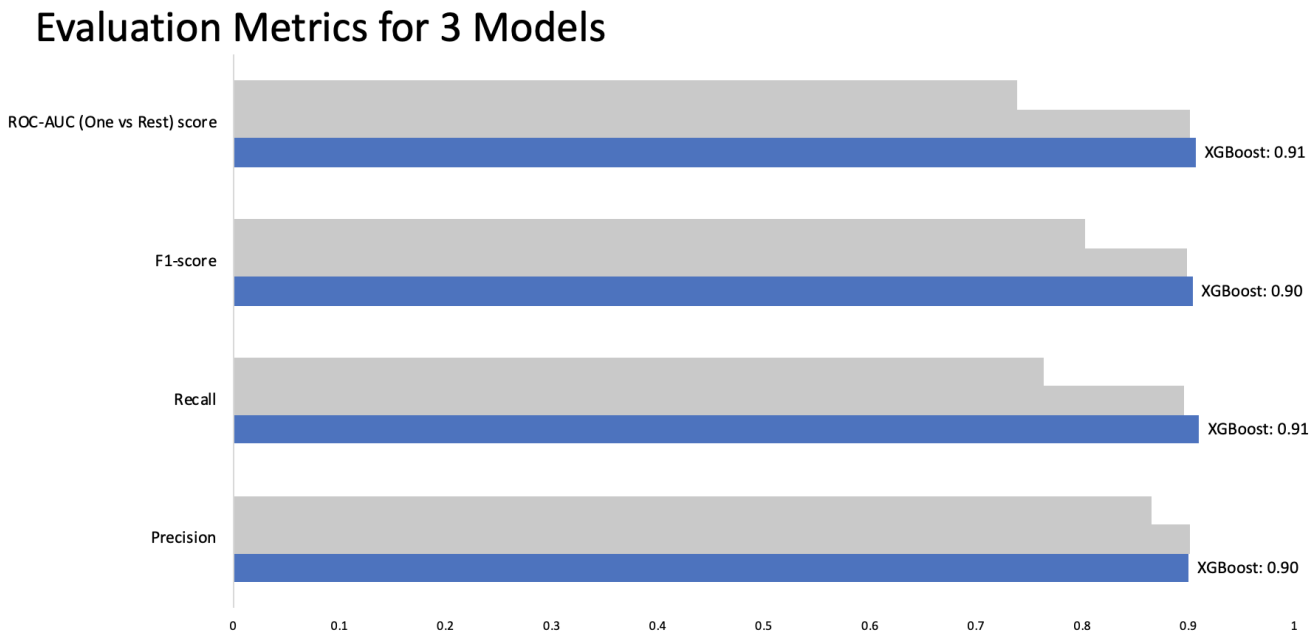
Random Forest

XGBoost

Train/test split

Train: 2015-01 to 2022-07 (91 months)

Test: 2022-08 to 2022-12 (5 months)



Model evaluation

- Comparison of model performances

- **XGBoost** had highest evaluation scores for most metrics

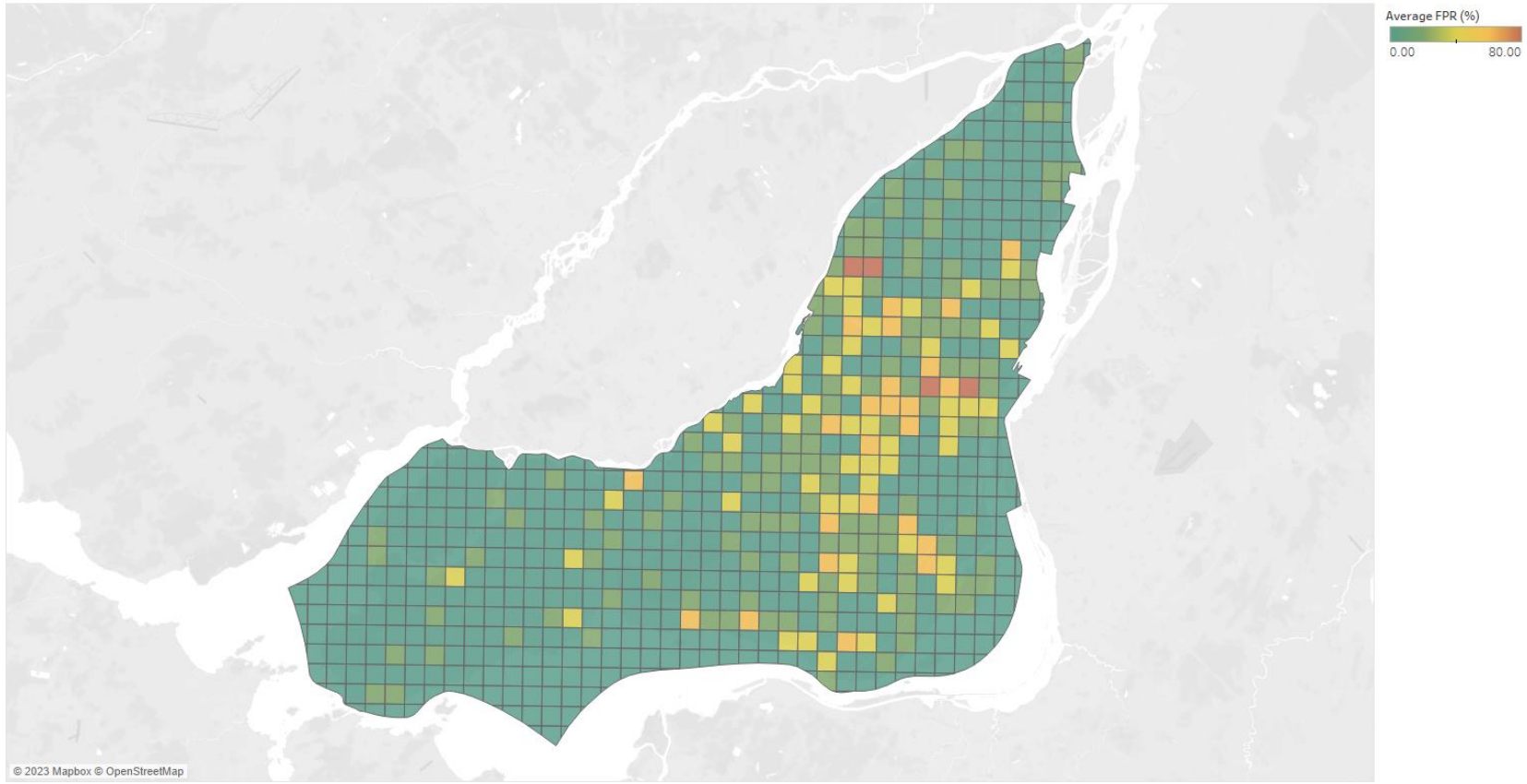
XGBoost Evaluation

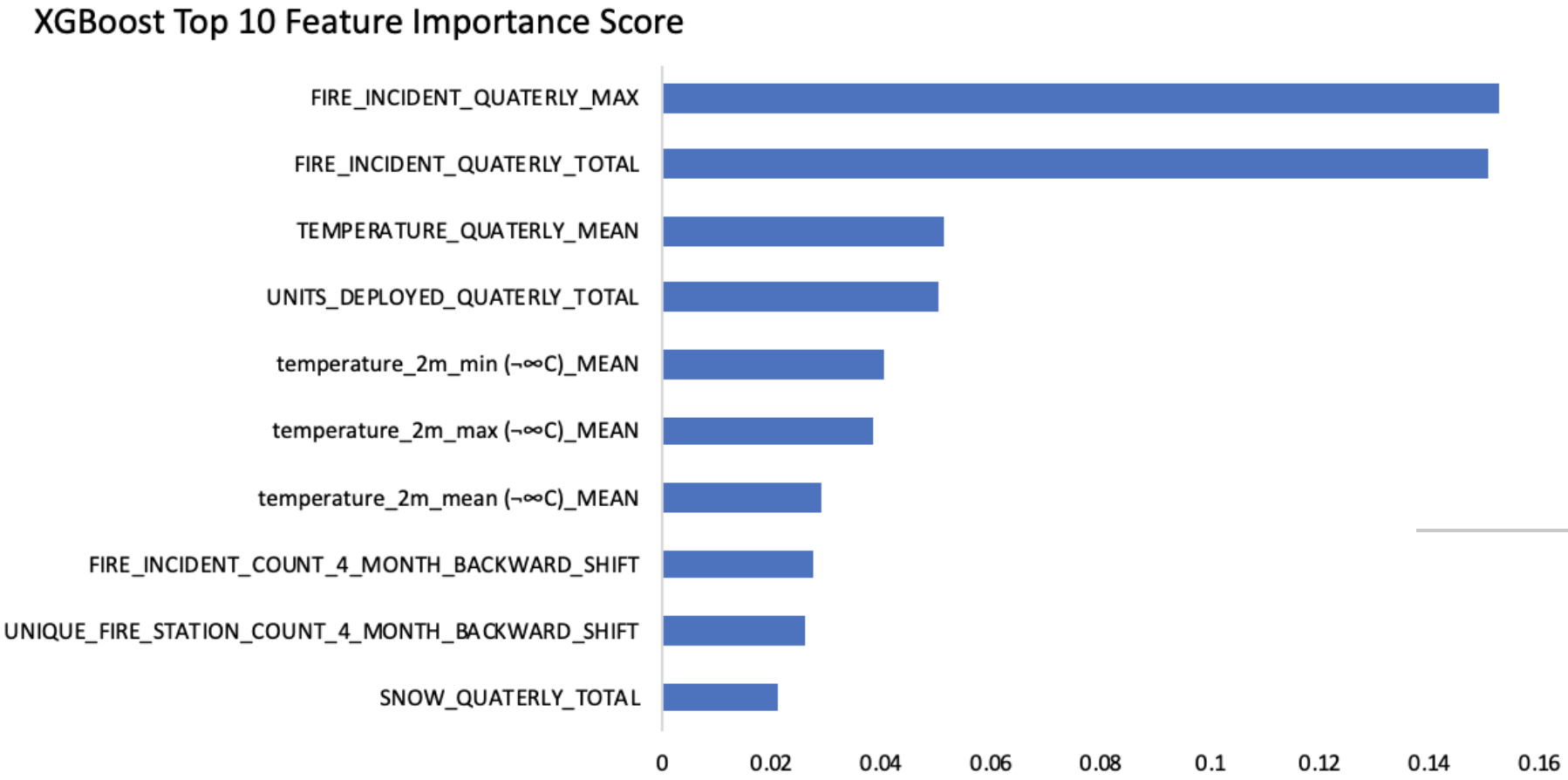
On average between classes, **89%** of the cases can be accurately classified

- **87% of the high fire-risks** can be accurately predicted
- **30% of the medium fire-risks** can be accurately predicted
- **97% of the low fire-risks** can be accurately predicted

Precision	Recall	F1-Score
90%	91%	90%

Average False Positive Rate for Testing Phase (5 months)

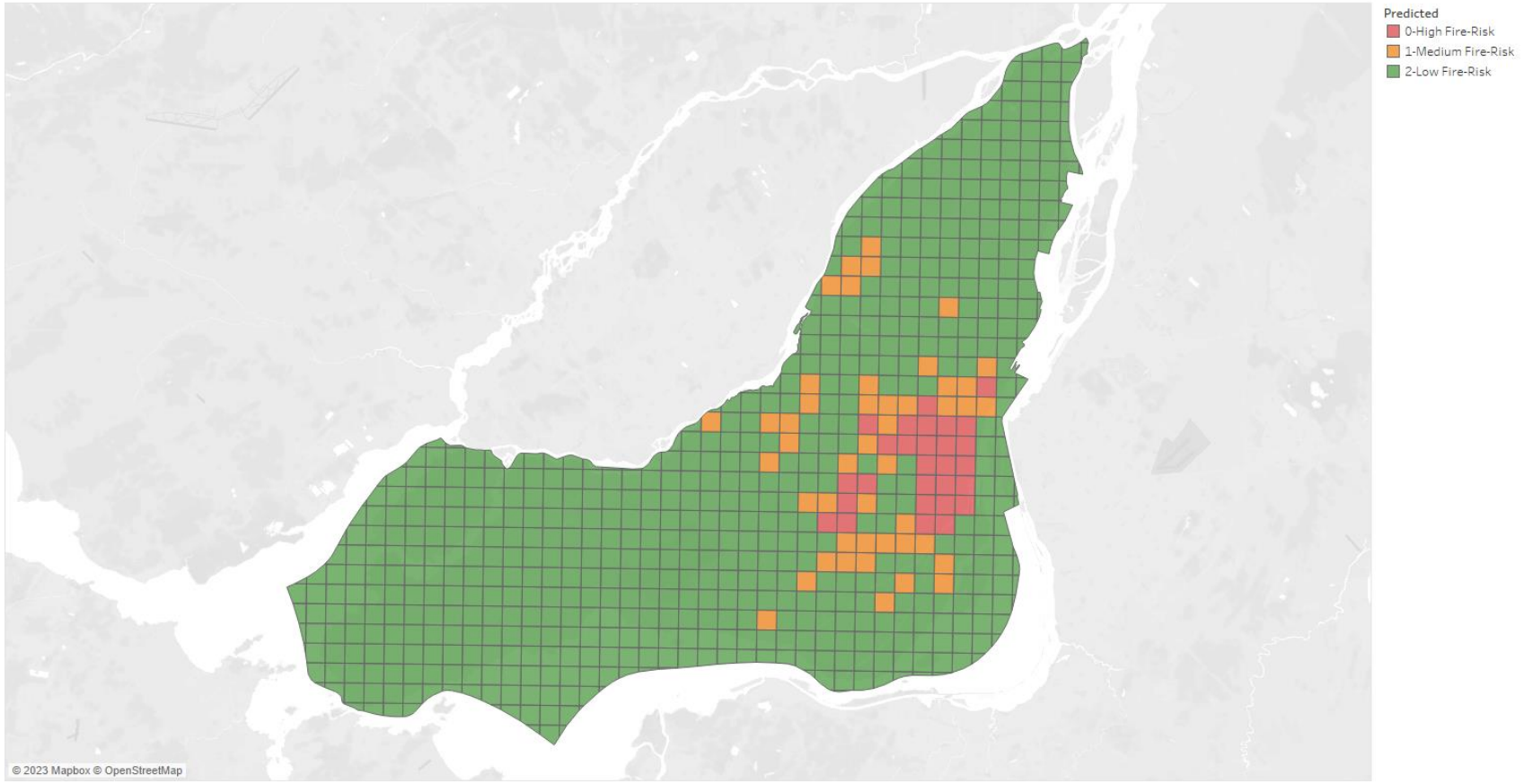




Top 10 Important Features

Top 10 features that impacted the XGBoost model the most

Prediction Map for February 2023



Recommendation

XGBoost Model for predicting high fire-risk areas in Montreal for the upcoming month

Focus area: Ville-Marie

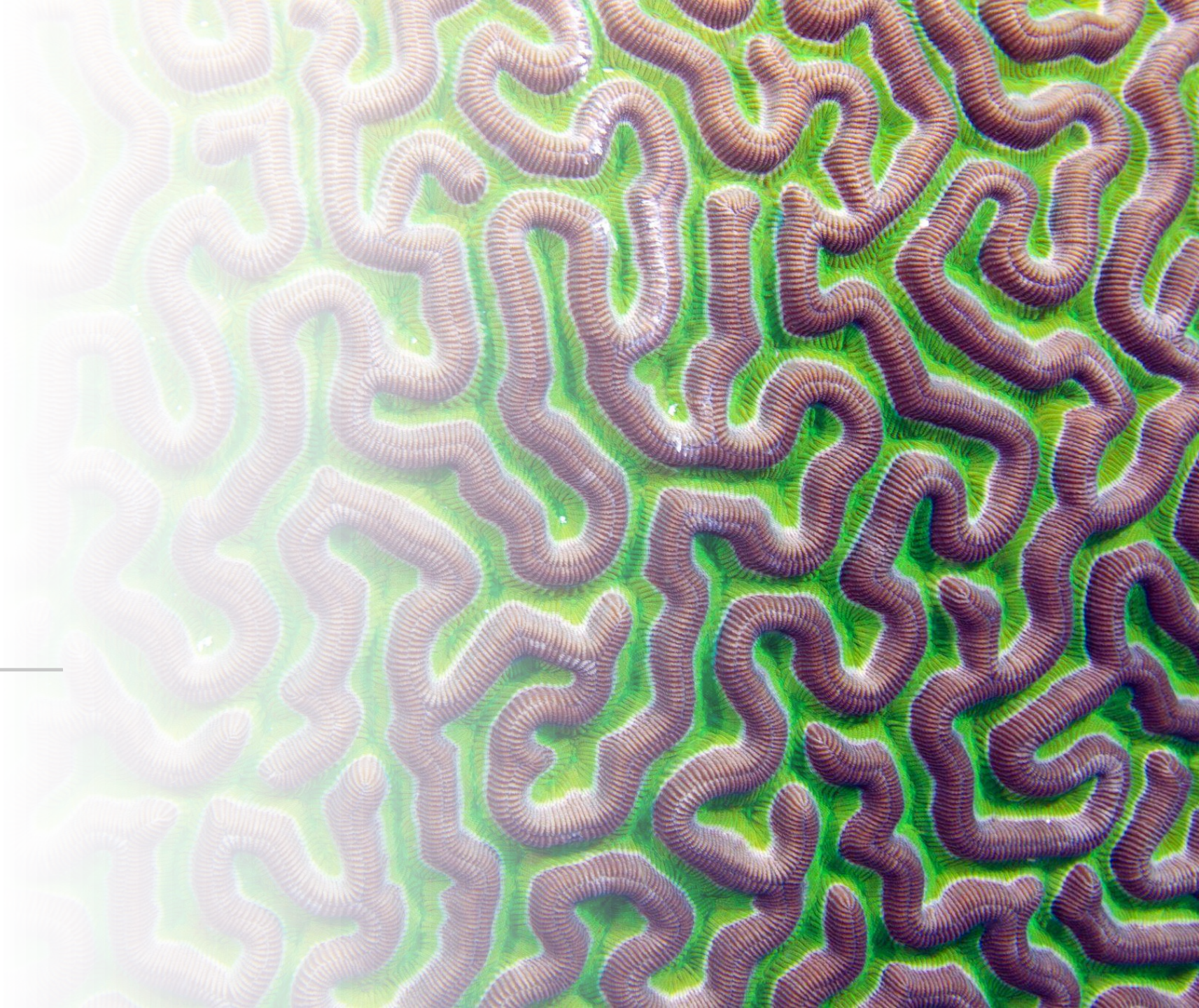
Extending it to forecast for longer period ahead

Strategy plan

Raise awareness in focus areas

Q&A

Appendix



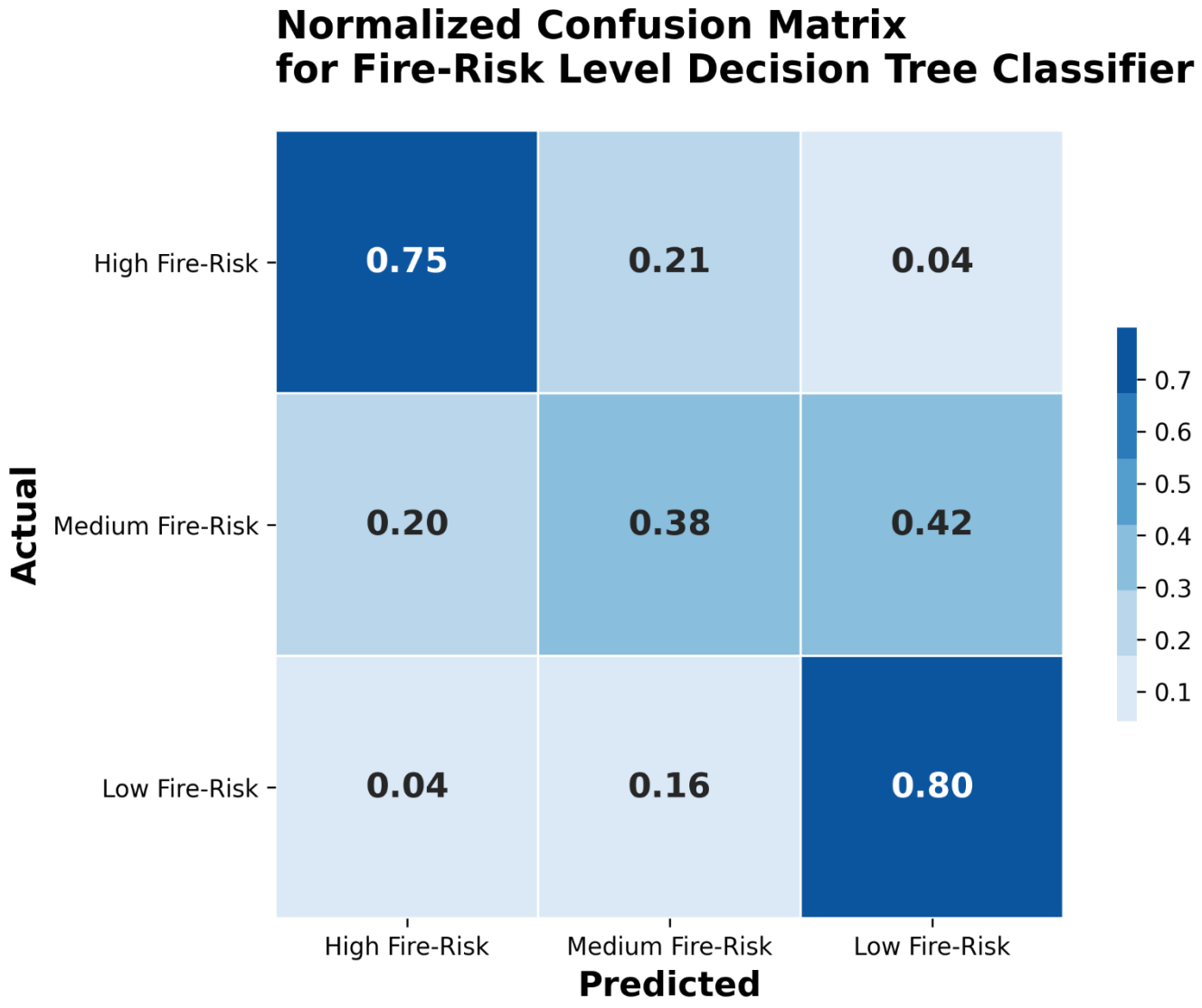
Appendix – Fire-risk levels

Classification	Level
High Fire-Risk	0
Medium Fire-Risk	1
Low Fire-Risk	2

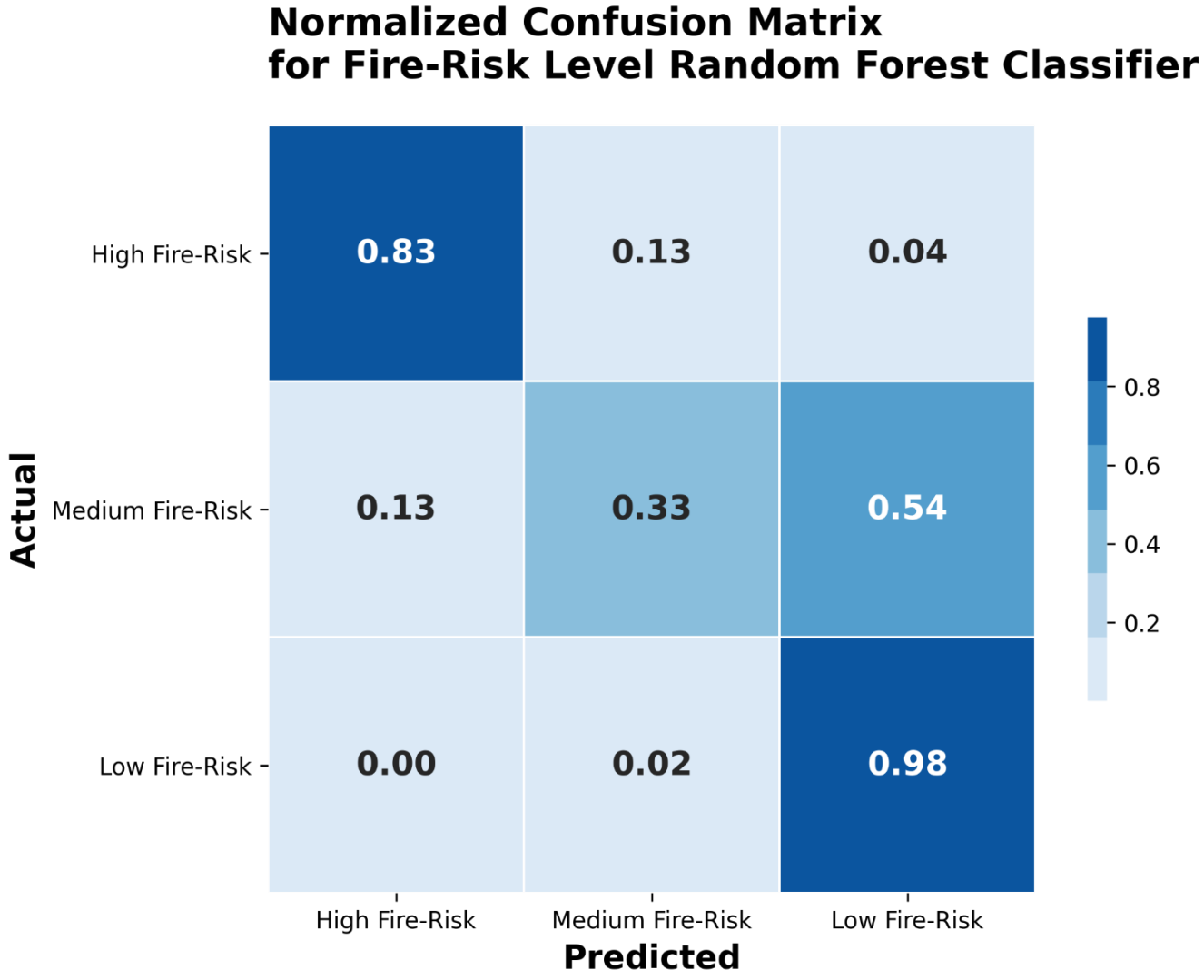
Appendix – Performance evaluation comparison between models

Model	Precision	Recall	F1-score	AUC
Decision Tree	0.87	0.76	0.80	0.74
Random Forest	0.90	0.90	0.90	0.90
XGBoost	0.90	0.91	0.90	0.91

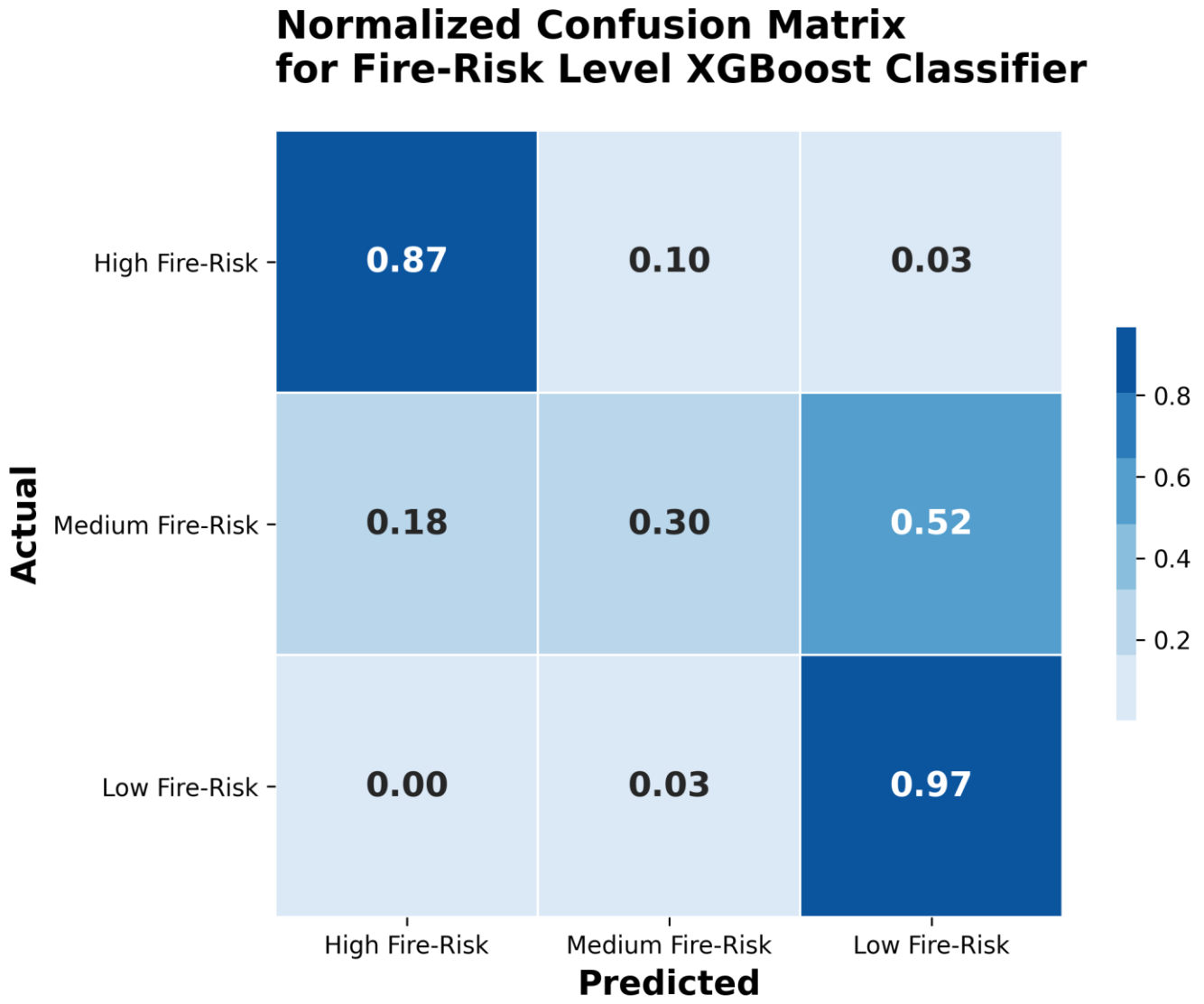
Appendix – Confusion Matrix (Decision Tree)



Appendix – Confusion Matrix (Random Forest)



Appendix – Confusion Matrix (XGBoost)



Appendix – List of columns for Montreal shapefile

- MUNID: Identifier for the administrative division of municipalities in Quebec, (MAMROT).
- CODEID: Unique identifier.
- CODEMAMROT: Identifier for the administrative division - unique identifier with MAMROT district code as prefix.
- NOM: Name of the administrative division - as described by the Quebec Government Toponymy Commission.
- TYPE: Type or entity of the administrative division - e.g., Borough, Associated City.
- ABREV: Abbreviation for the definition of boroughs and associated cities.
- NUM: Internal alphanumeric identifier (geomatics).
- AIRE: Official non-calculated area in square meters.
- PERIM: Official non-calculated perimeter in meters.
- GEOM: Administrative division geometry formatted according to the Well-known text standard.

Appendix – List of columns for fire incidents dataset

- INCIDENT_NBR: Unique ID for incident.
- CREATION_DATE_TIME: Timestamp of incident.
- DESCRIPTION_GROUPE: Grouping of intervention types into 6 categories: *Building Fires, Other Fires, Non-Fire, Fire Alarms, First Responders, False Alerts/Cancellations*.
- INCIDENT_TYPE_DESC: Detailed incident type.
- CASERNE: Number of the fire stations responsible for the area where the event occurred.
- NOM_VILLE: Name of the city where the incident occurred.
- NOM_ARROND: Name of the borough where the incident occurred.
- DIVISION: SIM division responsible for the area where the event occurred.
- LONGITUDE, LATITUDE: Geographic location of the event after obfuscation at an intersection according to the WGS84 geodetic reference.
- NOMBRE_UNITES: Number of vehicles deployed to respond to the event.

Appendix – List of columns for criminal incidents dataset

- CATEGORIE: Nature of the event. 6 categories include: break and enter, theft from motor vehicle, motor vehicle theft, mischief, robbery, and criminal offense causing death.
- DATE: Timestamp of criminal event.
- QUART: Time of day when the event was reported to the SPVM. Options include day (8:01 a.m. and 4:00 p.m.), evening (4:01 p.m. and midnight) and night (12:01 a.m. and 8:00 a.m.).
- PDQ: Number of the police station covering the area where the event occurred.
- X: Geospatial position according to the MTM8 projection (SRID 2950).
- Y: Geospatial position according to the MTM8 projection (SRID 2950).
- LATITUDE: Geographic location of the event after obfuscation at an intersection according to the WGS84 geodetic reference.
- LONGITUDE: Geographic location of the event after obfuscation at an intersection according to the WGS84 geodetic reference.

Appendix – List of columns for property assessment dataset

- ID_UEV: Unique system identifier.
- CIVIQUE_DEBUT: Civic number (range - start).
- CIVIQUE_FIN: Civic number (range - end).
- NOM_RUE: Street name.
- SUITE_DEBUT: Unit number (apartment or local).
- ETAGE_HORS_SOL: Maximum number of floors:
- If the UEF includes a single building: Number of floors of the building.
- If the UEF includes multiple buildings: Number of floors of the building with the most floors (maximum).
- NOMBRE_LOGEMENT: Number of housing units.
- ANNEE_CONSTRUCTION: Year of construction.
- CODE_UTILISATION: CUBF coding.
- LETTRE_DEBUT: First letter of the apartment.
- LETTRE_FIN: Last letter of the apartment.
- LIBELLE_UTILISATION: CUBF description.
- CATEGORIE_UEF: Unit evaluation category (Regular or Condominium).
- MATRICULE83: Roll number (NAD83 MT8 geospatial system).
- SUPERFICIE_TERRAIN: Land area for property assessment purposes (square meters).
- SUPERFICIE_BATIMENT: Building floor area, i.e. gross floor area corresponding to the sum of the areas of each of the whole floors of the main building and, if applicable, those of the attic, integrated garage and integrated greenhouse (square meters).
- NO_ARROND_ILE_CUM: Borough identifier (MAMROT reference identifier).
- MUNICIPALITE: Internal municipality identifier.

Appendix – List of columns for population dataset

- CODEMAMROT: Identifier for the administrative division - unique identifier with MAMROT district code as prefix.
- NOM: Name of the administrative division - as described by the Quebec Government Toponymy Commission.
- YEAR: Yearly timestamp.
- POPULATION: Total number of individuals.

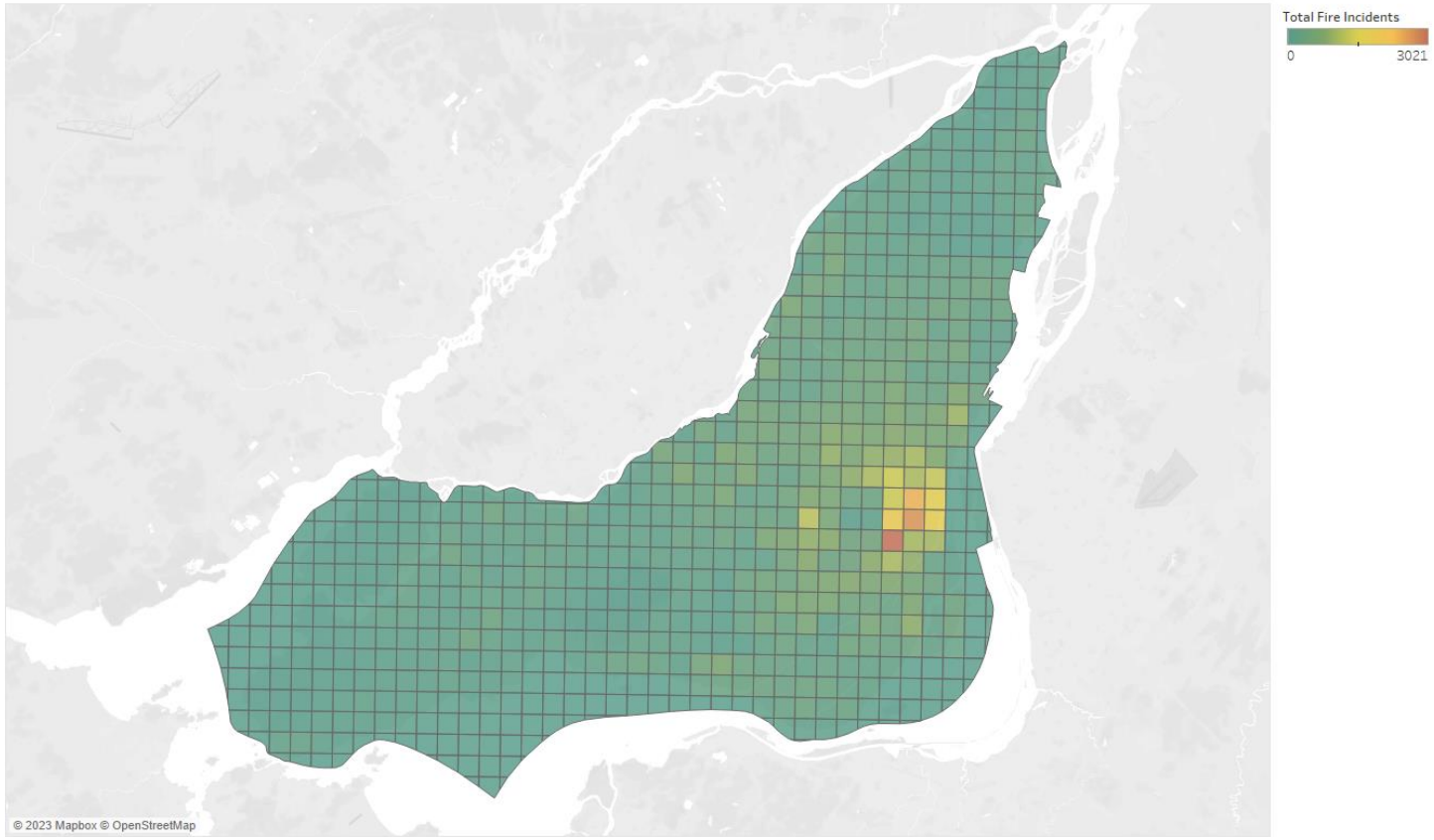
Appendix – List of columns for weather dataset

- Time: UNIX timestamp.
- Temperature_2m_max (°C): Maximum temperature for the day.
- Temperature_2m_min (°C): Minimum temperature for the day.
- Temperature_2m_mean (°C): Mean temperature for the day.
- Shortwave_radiation_sum (MJ/m²): The sum of solar radiation on a given day in Megajoules.
- Precipitation_sum (mm): Sum of daily precipitation (including rain, showers, and snowfall).
- Rain_sum (mm): Sum of daily rain.
- Snowfall_sum (cm): Sum of daily snowfall.
- Windspeed_10m_max (km/h): Maximum wind speed on a day.
- Windgusts_10m_max (km/h): Maximum wind gusts on a day.
- Winddirection_10m_dominant (°): Dominant wind direction.
- Wt0_fao_evapotranspiration (mm): Daily sum of ET_0 Reference Evapotranspiration of a well watered grass field.

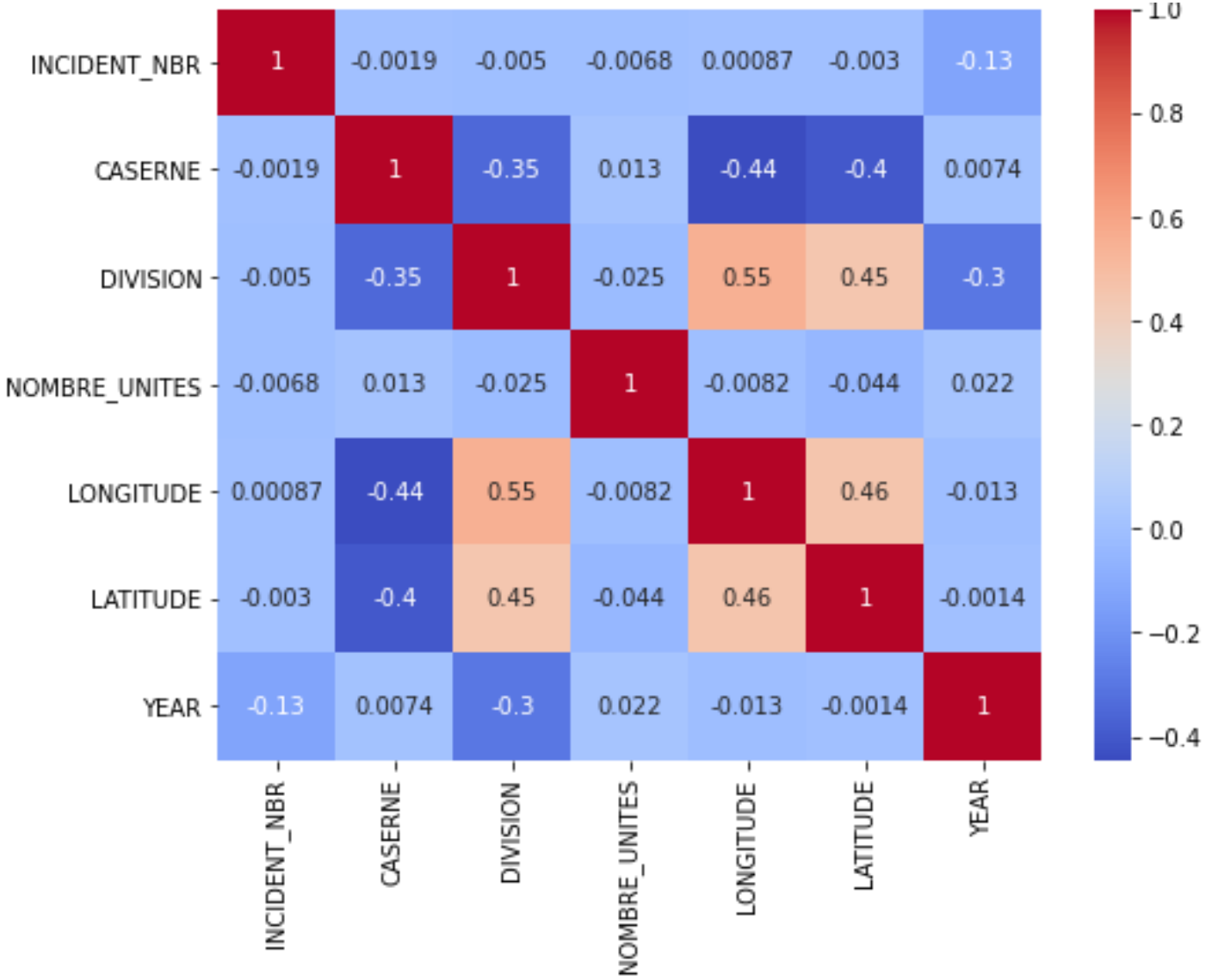
Decision Tree	Random Forest	XGBoost
max_depth=100, min_samples_split=2, min_samples_leaf=1, criterion='gini', max_features='sqrt', splitter='random', random_state=42	n_estimators=900, max_depth=100, min_samples_split=5, min_samples_leaf=2, max_features='sqrt', class_weight='balanced', criterion='gini', bootstrap=True, oob_score=True, random_state=42, n_jobs=-1	num_class=3, learning_rate=0.1, max_depth=60, n_estimators=600, subsample=0.6, colsample_bytree=0.6, eval_metric='aucpr', objective='multi:softprob', tree_method='gpu_hist', gpu_id=0,

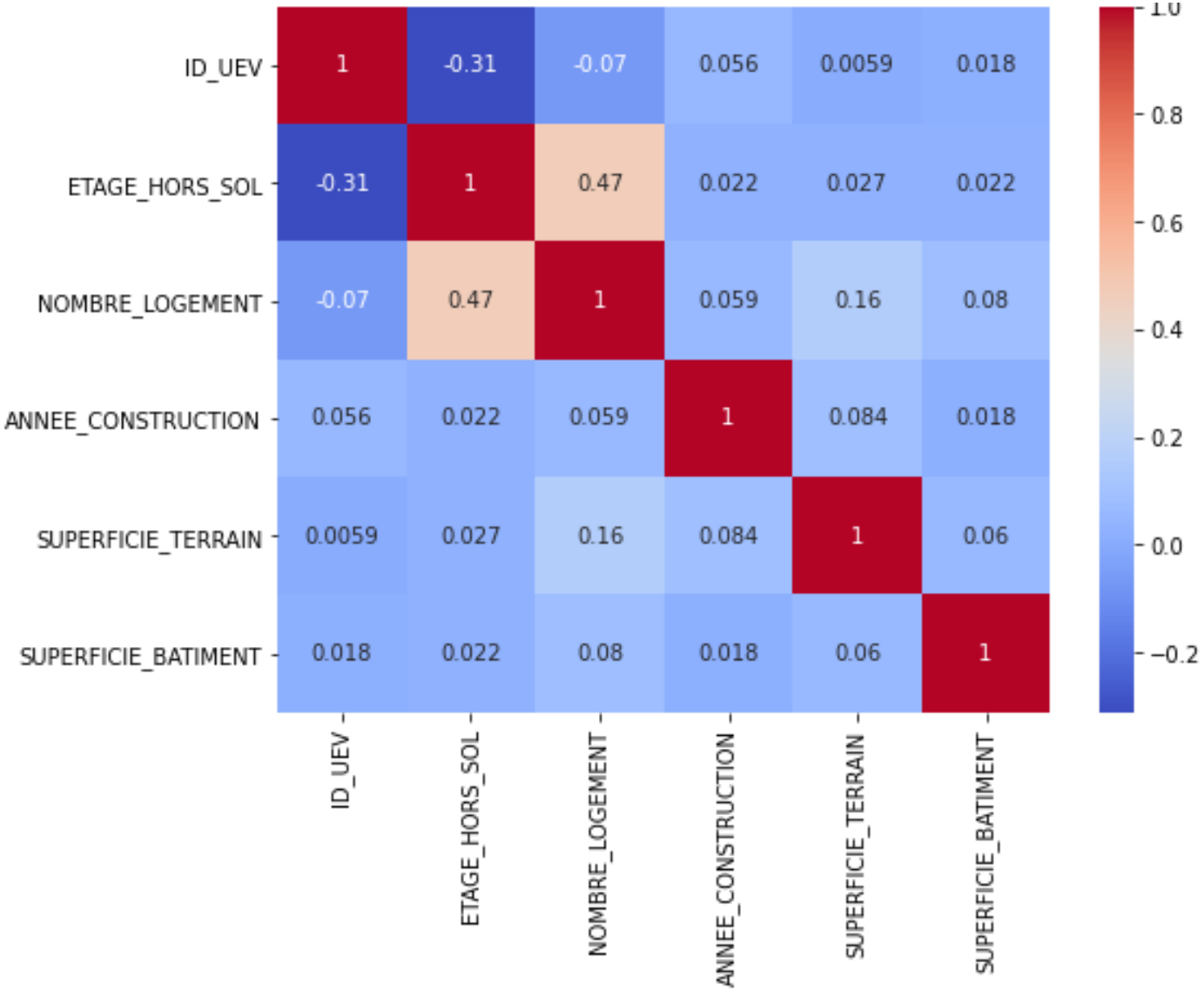
Appendix – List of hyperparameters obtained through grid search for each model.

Appendix – Fire incidents per grid on the Montreal shapefile



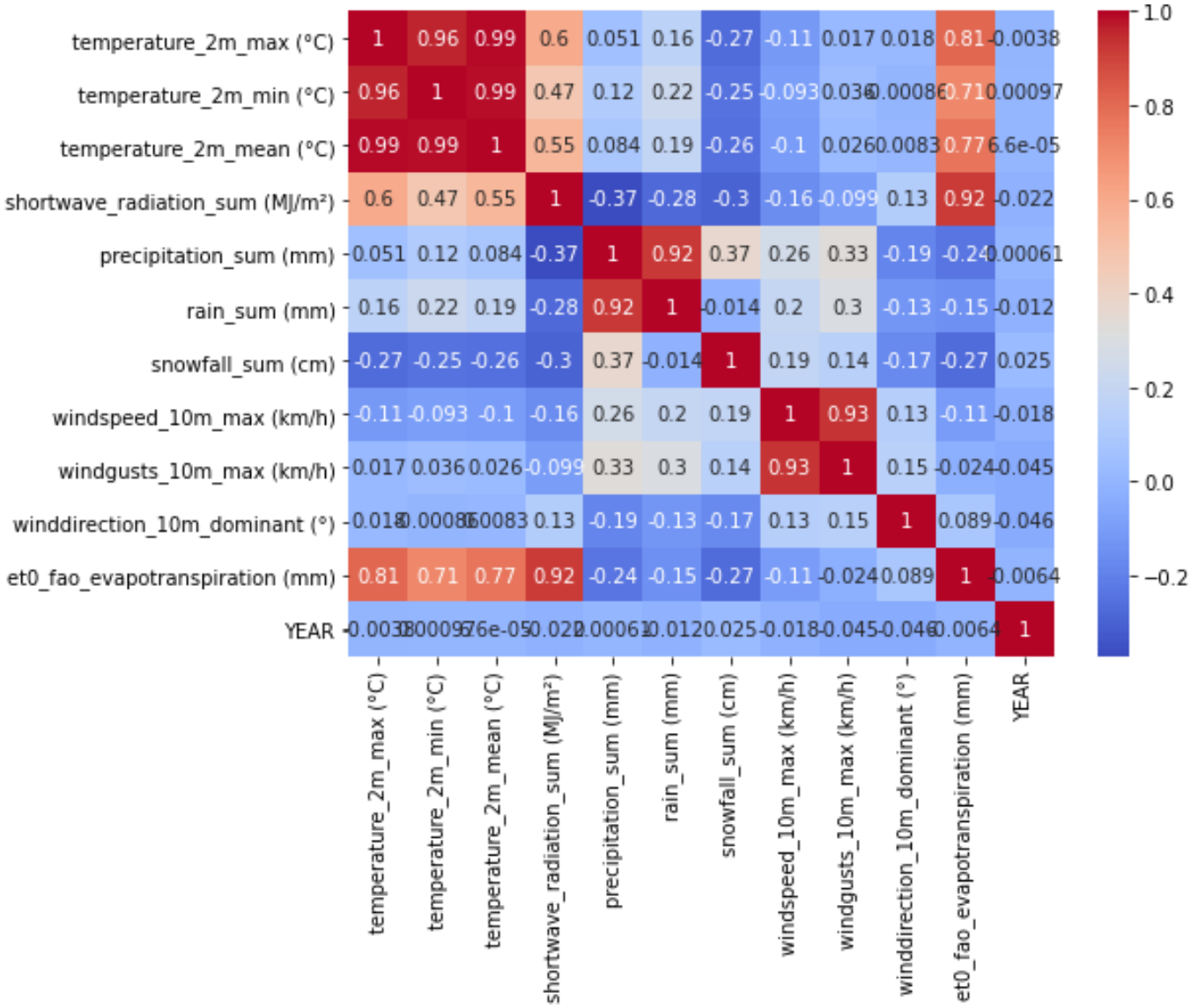
Appendix – Correlation matrix for numerical features from fire incidents dataset



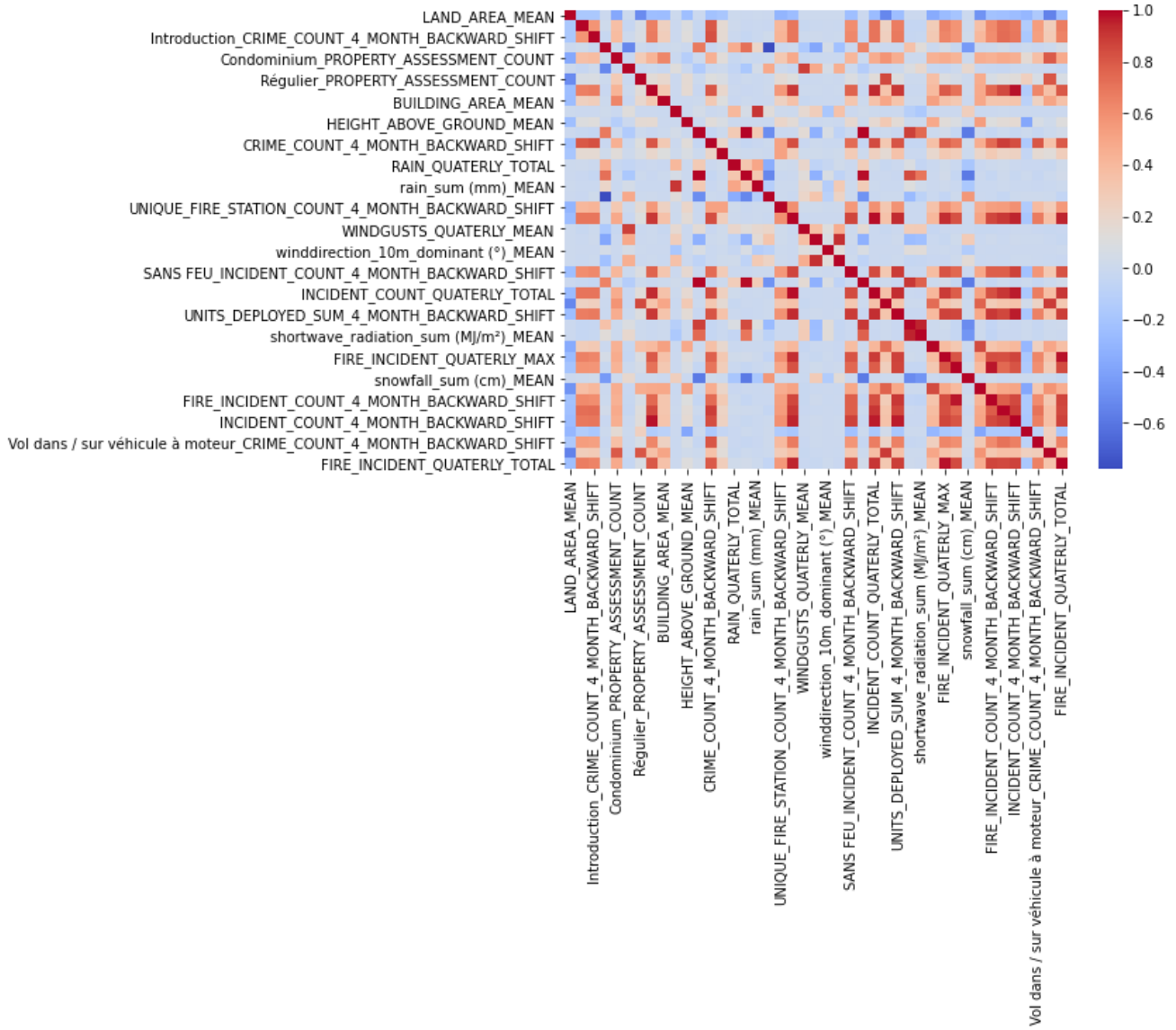


Appendix – Correlation matrix for numerical features from property assessment

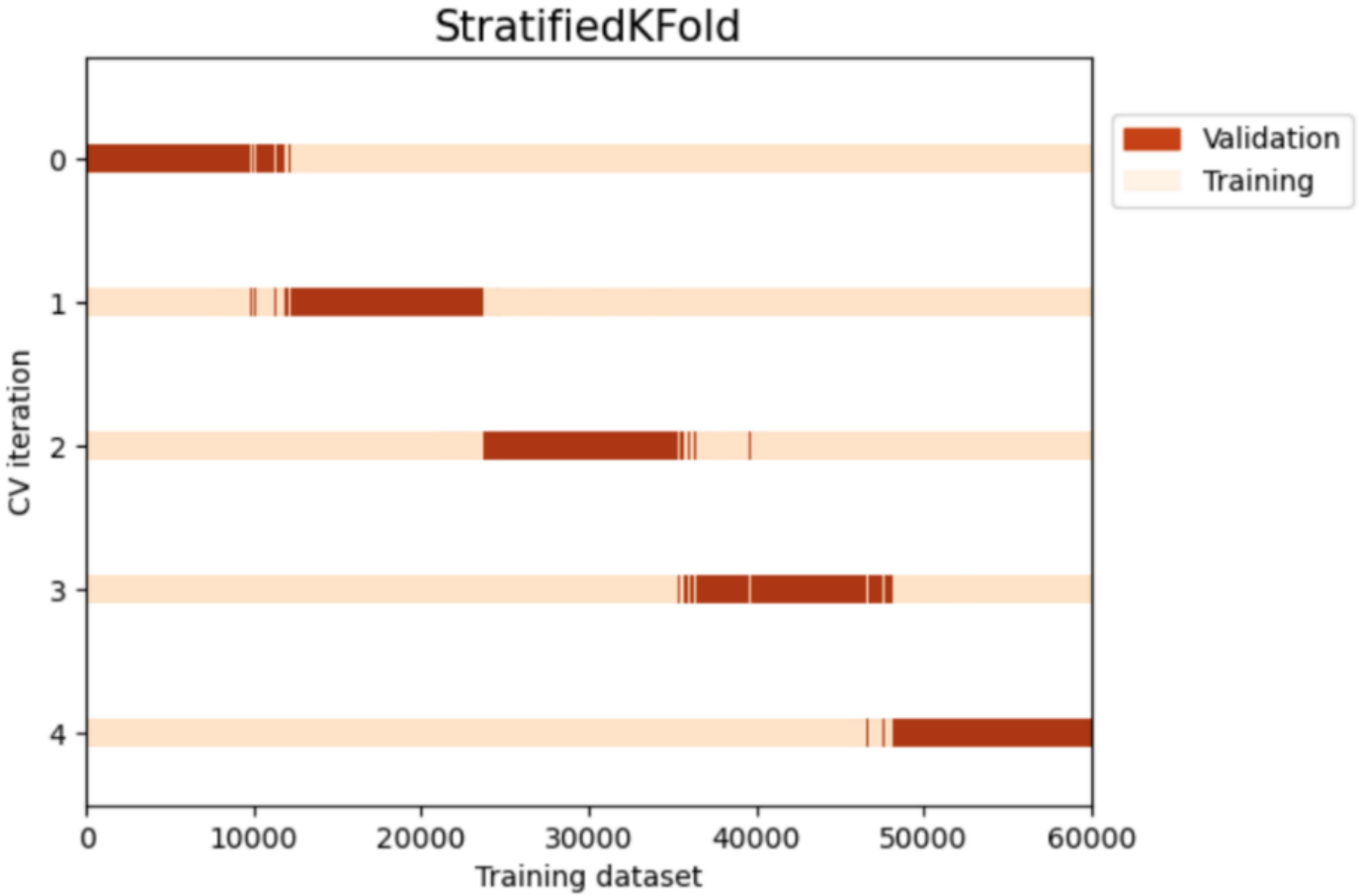
Appendix – Correlation matrix for numerical features from the weather dataset



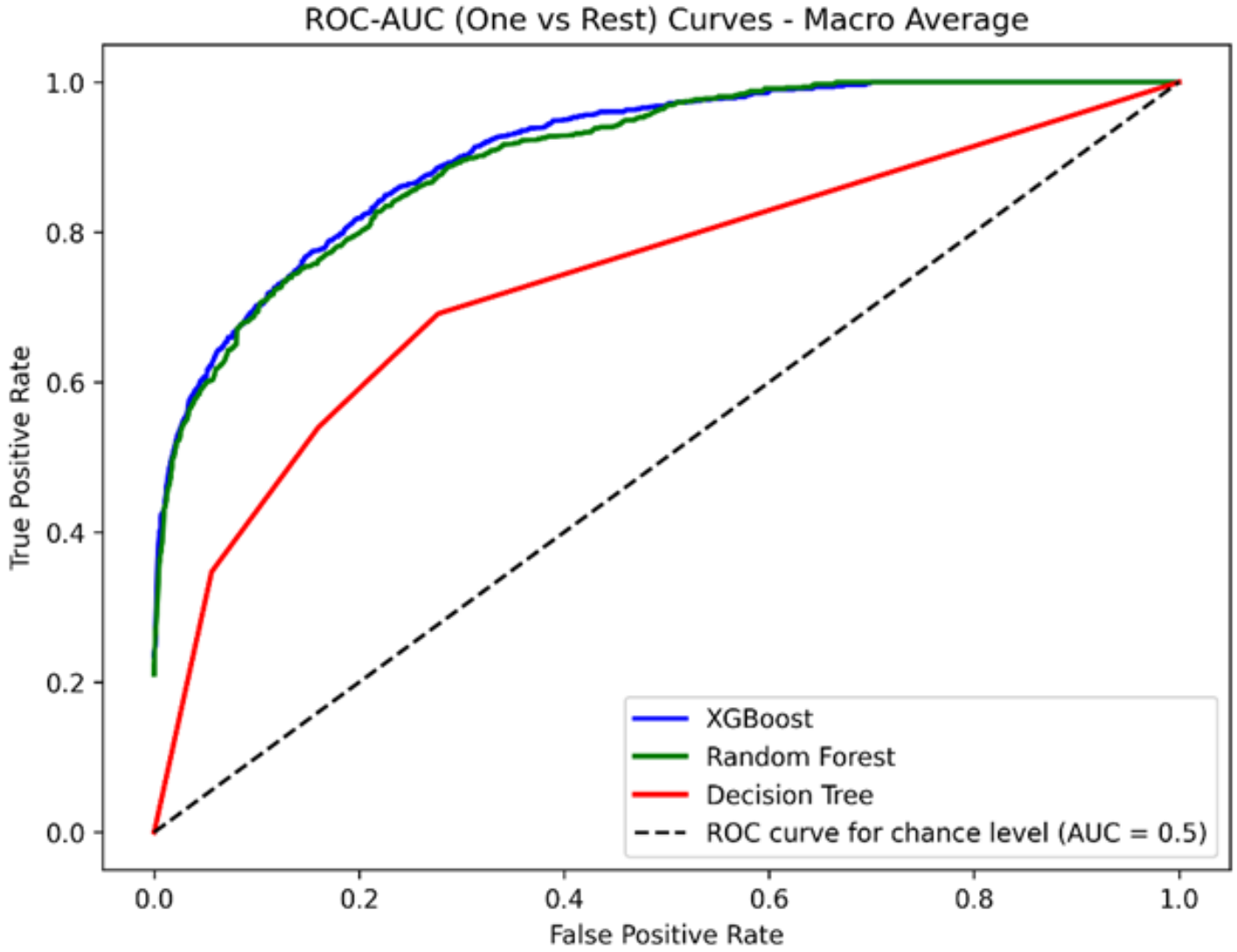
Appendix – Correlation matrix for filtered features



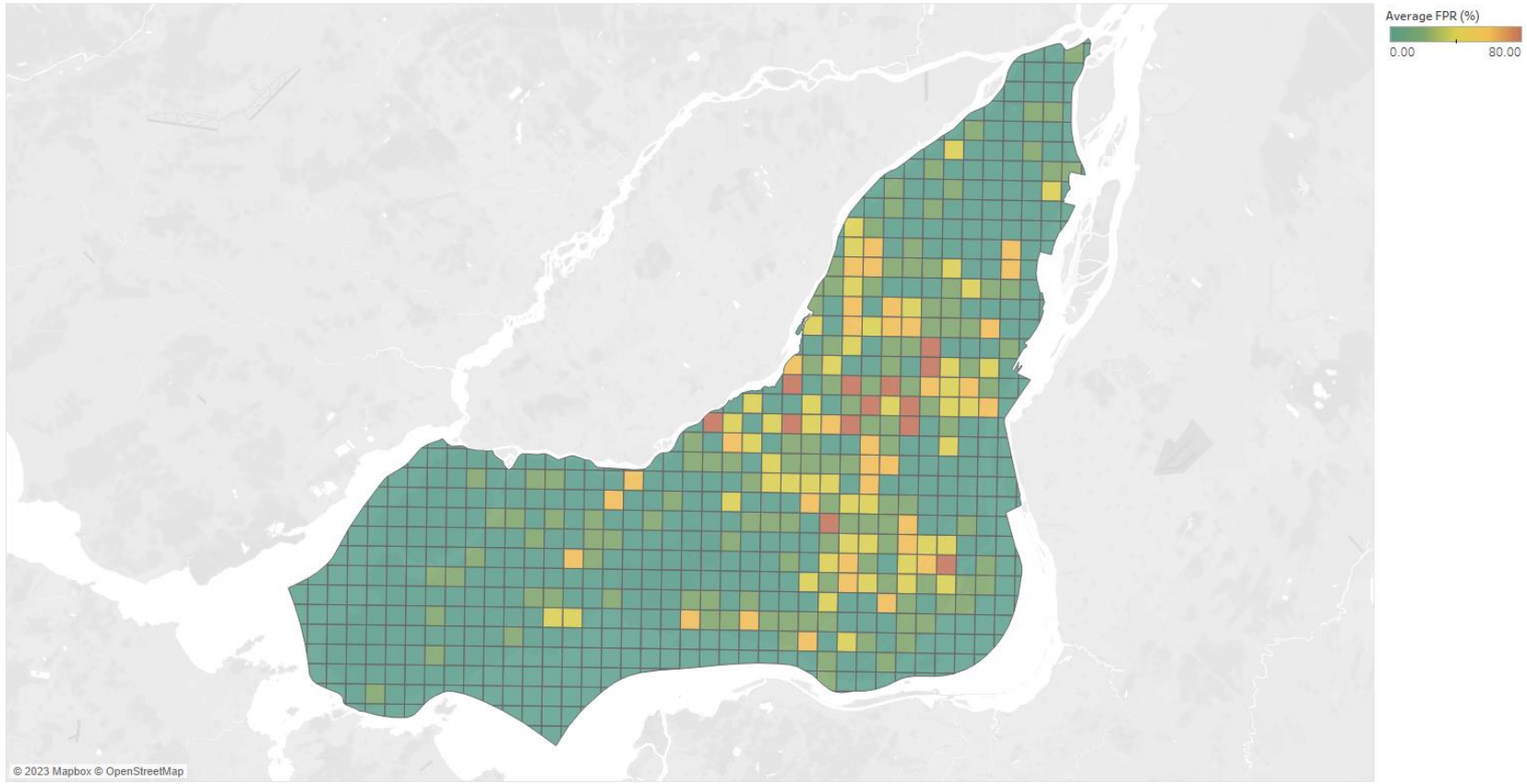
Appendix – 5 iterations of stratified cross validation with training dataset



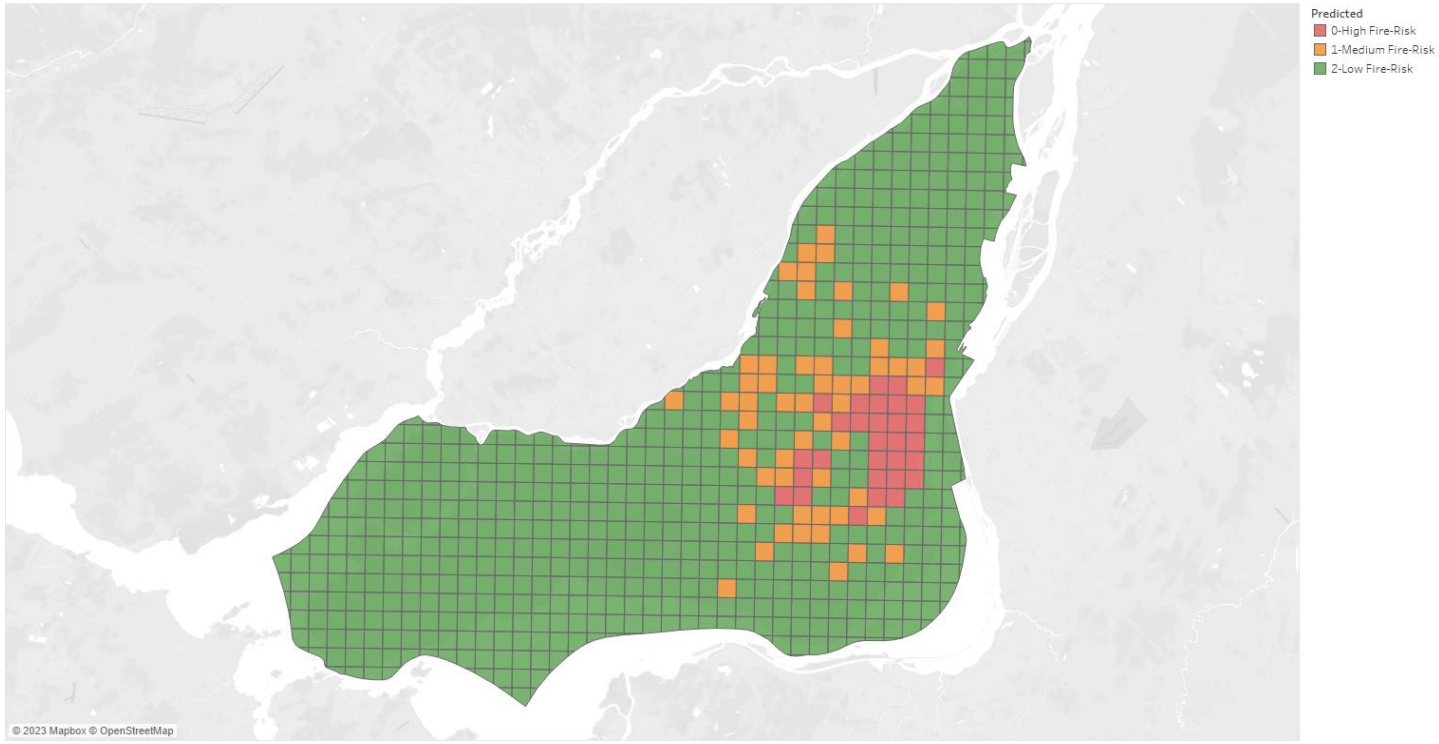
Appendix – ROC-AUC curve (One vs Rest) for XGBoost, random forest and decision tree classifiers

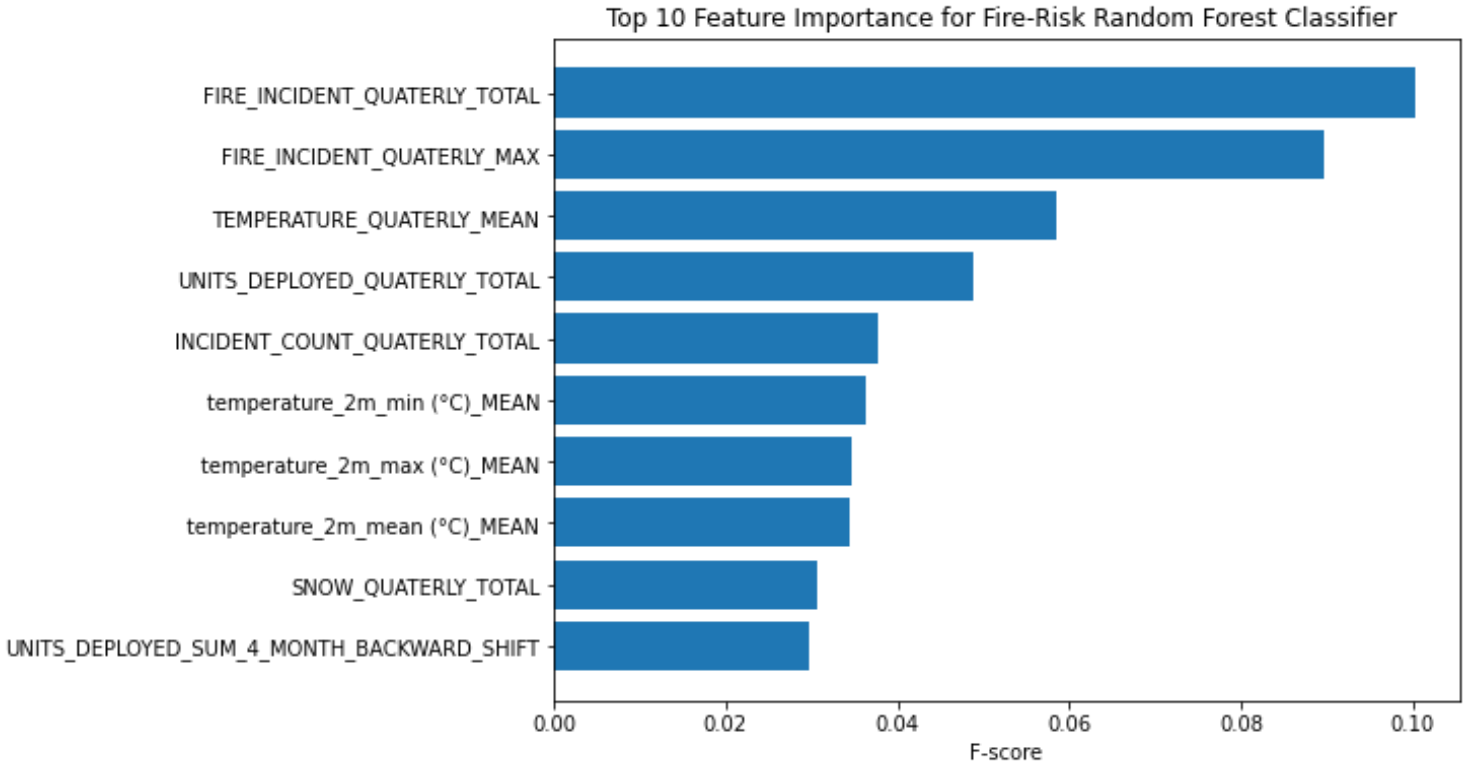


Appendix – Average false positive rate for fire-risk prediction of Montreal using a random forest classifier



Appendix – Predicted fire-risk levels for the city of Montreal for February 2023 using random forest classifier





Top 10 features by feature importance for the random forest classifier