## Final Proposal: Informing Fire Safety Inspections

#### *Creating a model that assists fire inspectors in prioritizing non-residential buildings based on fire risk*

#### Problem Statement

The Peoria Fire Department (PFD) conducts regular inspections of non-residential buildings to assess whether they are in compliance with local fire safety code. These inspections are meant to catch fire code violations before they result in a building fire and put the lives of building occupants in danger.

The PFD does not currently have a mechanism or process for prioritizing fire inspections of non-residential buildings. In addition, inspectors currently face a substantial backlog of buildings that have not been recently inspected.

We will merge data sets that include various characteristics of non-residential buildings with incidents of fire to see whether there are any relationships between these characteristics and a fire incident. The goal of our data science model will be to predict where fires may be most likely to occur in the future, establishing a more refined prioritization model for the PFD.

#### Hypothesis

Risk of fire increases as year built increases, as date from last fire inspection increases, as square footage increases, and as certain code violation counts increase.

#### Assumptions

* Fire Inspections that are completed are complete with quality data.
* The inspectors will have to continually update the Occupancy Vulnerability Assessment Profile (OVAP) data
* OVAP scores will be able to calculate risk according to the below list:
  + Maximum: 60 or more
  + Significant: 40 to 59
  + Moderate: 15 to 39
  + Low: 14 or less
* Based on risk, number of commercial occupancies in the City of Peoria, and the number of inspections that can be completed each year calculate a process for assigning inspections on a rotation basis so that every commercial occupancy is inspected over a period of 3-5 years.
* Due to the following data items being incomplete or unaccessable some information may need to be estimated:
  + Software changes at the Fire Department
  + Incomplete commercial occupancy listings
  + Errors in OVAP scoring
  + Changes in occupancy without correct permitting
* There isn’t a large number of commercial fires to compare inspection data to.
* Residential homes are not inspected unless multiple apartments.

#### Goals:

1. Identify to what degree various building characteristics impact the fire risk of the building.

2. Building a predictive model based on a number of parameters including home age, location, code violation history, that assigns a risk assessment to each non-residential building in the City of Peoria.

3. Rank fire districts by number of buildings that have buildings with an elevated risk of fire, to help better allocate fire inspector resources.

#### Success Metrics:

1. Efficacy of our model to ‘successfully’ highlight properties or zones at highest risk of fire
   1. Successful if the slate of properties our model predicts/prioritizes actually return an OVAP score, close to what we predicted given the data we have from 2017, upon human inspection.
2. PFD adopts recommended process for prioritizing fire inspections of non-residential buildings
   1. Establish timelines for inspections by priority
   2. Improved communication and coordination across companies and shifts
   3. Posted schedule of fire inspections online
3. Fewer fires
4. Improved coordination of code enforcement and fire inspections between PFD and City of Peoria Community Development
5. [MAYBE SOMETHING ABOUT COMMUNICATING RISK FACTORS WITH COMMERCIAL REALTORS/BUYERS/OWNERS]
6. Maintain comprehensive, up-to-date listings of the commercial properties

#### Risks:

We probably won't be able to replicate the Pittsburgh case study exactly with the time that we have. We may just be able to get all the data sets that we need, clean them, organize them, and try creating a basic (imperfect) model to start with and learn from.

A good model that improves performance can cause us to begin collecting biased data. For example, if the model initially identifies that buildings with deep fryers are more likely to have incidents of fires, those types of buildings would be inspected more often. If fire incidents at these buildings go down, our model would begin to tell inspectors that those types of buildings are less likely to cause fires, even while that may not be true. (This is a consideration that should impact how our model is inserted into the inspection process. We might be able to avoid this issue by using our model to complement inspectors current process, rather than replace it.)

#### Limitations:

1. Possible Limitation: As mentioned above, the project we have in mind for our predictive model may be hindered most heavily by time. For example, our model would need to be re-evaluated and re-adjusted multiple times before implementation as a full replacement could happen. If inspectors plan to use this model to supplement their efforts then decreasing bias, among other factors, will be our top priority to ensure reliable and consistent results.

Approach to Limitation: Our approach to this limitation is to create a predictive model for our project that will operate as a complimentary resource for inspectors in hopes of strengthening the efficiency of the inspection process.

2. Possible Limitation: Another major limitation would be the size and scale of our training data set. For example, if we train our model from inspections collected only in summer months of 2015 then faults and errors will arise most definitely when we apply the model moving forward.

Approach to Limitation: Our solution is to identify a temporal range for our training set that is fair and reasonable, meanwhile also providing the highest accuracy for predictive analysis in our model.

#### Data Sets Used

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Source** |
| Fire Incidents (2010-2017) | Catalogues all fire incidents to which the PFD was called. *Why: Fire Incidents are what we will be trying to predict and therefore will be used to evaluate the accuracy of our model.* | Peoria Fire Department |
| Fire Inspections  (2010-2017) | Catalogues inspection results for non-residential buildings *Why: How long ago an inspection occured may impact the risk of a fire in the near future.* | Peoria Fire Department |
| Building Characteristics | Catalogues several characteristics of all buildings, such as year built, square footage, etc. *Why: Some building characteristics may impact the risk of a fire in the future. For example, older buildings may contain knob-and-tube electrical wiring, which is less safe than modern electrical wiring.* | Peoria County GIS Department / City of Peoria Information Systems Department |
| Code Violations  (Dates?) | Catalogues ordinance violations found when inspecting buildings *Why: Code violations may point to issues that could cause a fire, such as exposed electrical wiring.* | City of Peoria Community Development Department |

#### Expected Challenges

#1. Scaling Data Collection / Access

In keeping the scope small, within Peoria, we don’t anticipate any trouble acquiring or accessing relevant data. However, if we ever look to scale, acquiring testing data from other communities may prove more challenging.

#2. Data Uniformity

In collecting data from a couple of different partners, we anticipate a need to do some data cleaning to ensure databases are formatted similarly and line up for valid analyses.

#### Case Studies

**New Orleans** - predicted fire risk to help their Fire Department prioritize where to install smoke detectors

**Article**: <https://datasmart.ash.harvard.edu/news/article/predicting-fire-risk-from-new-orleans-to-a-nationwide-tool-846>  
**Report**: <http://nola.gov/performance-and-accountability/nolalytics/files/full-report-on-analytics-informed-smoke-alarm-outr/>   
**Code** (*written in R, a different programming language, but the logic would still help us re-create it in Python*): <https://github.com/cno-opa/smoke-alarm-outreach>  
**Data Used**: American Housing Survey (available), American Census Survey (available to us), Property Data (available to us), Fire Incident Data (available to us)

**Pittsburg** - predicting fire risk to help Fire Department prioritize fire safety inspections

**Article**: <http://www.govtech.com/data/Pittsburgh-Uses-Data-to-Predict-Fire-Risk.html>

**Report**: <http://michaelmadaio.com/Metro21_FireRisk_FinalReport.pdf>   
**Code**: <https://github.com/CityofPittsburgh/fire_risk_analysis>   
**Data Used**: Property Data (available to us), Permits, Licenses and Inspections Data (available to us), Fire Incident Data (available to us)

Notes:

* Need to separate out types of fire incidents (concerned about building fires, not - for example - controlled burnings or overheated motors) and weigh them according to priority
* Model needs to take into account not only risk of fire, but likelihood of loss of life in said fire (movie theatre vs. drive-up atm)
* Parcels vs. addresses - fires are probably logged at an address, but a single address may be related to several parcels
* Our model accuracy should be assessed by how many true positives (predicted fire and there was a fire) *and* how few false negatives it guesses (predicted no fire and there was a fire)