**Code for Charlottesville Fire Risk Modeling Data – 11/7/2019**

This document describes the source data for Fire Risk Modeling. It describes two files: PIN\_Summary.csv and GPIN\_Summary.csv. This is a minimum viable source for modeling, containing information on the use code risk, year built, and square footage of the structure. Additional information about the features is available can be added to the data in the future.

This file was built from the following files:

Real Estate (Residential Details) – City of Charlottesville

Real Estate (Commercial Details) – City of Charlottesville

City Land UseCodes Exclusions for Company Inspections.xlsx – Charlottesville Fire Department

Note: The Real Estate (Residential Details) and the Real Estate (Commercial Details) have different structures and contents.

[Full production methods will be described in the future. The files were standardized, combined, and selected fields extracted. There are some questions about the data that will need to be discussed with the City of Charlottesville staff.]

**PIN\_Summary\_20191107.csv**

This is a summary file at the PIN data level. The Output file contains the following fields (selected from the Residential and Commercial files):

**GPIN**. The Geographic Parcel Identification Number. This should be an integer.

PIN. The Parcel Identification Number. This is a text string.

**UseCode**. The Use Code assigned by Neighborhood Development Services. This is a text string. It can have one of the following values:

**YearBuilt**. The Year the structure on the PIN was built. This is an integer.

**Area\_SqFT**. The Square Footage of the structure on the PIN. This is an integer.

For Commercial records, this was the GrossArea.

For the Residential Records, this was calculated as follows:

UnfinishedBasement = Basement - FinishedBasement,

Area\_SqFT = SquareFootageFinishedLiving + UnfinishedBasement

**Source**.This was the source of the data, from the Residential or the Commercial file. This is a text string.

**Risk**. The risk score or the UseCode assigned by the Charlottesville Fire Department (Lucas Lyons). This has a value from 0 – 3, where 0 represents the lowest risk and 3 represents the highest risk.

**Zoning**. This is the zoning classification associated with the GPIN.

Note: There are NA values spread throughout the files.

**GPIN\_Summary\_20191107.csv**

This is a summary file aggregated at the GPIN data level. It was built from the PIN\_Summary.csv data.

**GPIN**. The Geographic Parcel Identification Number. This should be an integer. There are some NA values.

**PIN\_ALL**. A pipe delimited list of all the PINs in the GPIN. This is a text field.

**PIN\_Count**. The number of PINs associated with the GPIN. This is an integer.

**Oldest\_Structure**. This is the earliest date associated with a PIN in the GPIN. This is an integer.

**Largest\_Structure\_SqFT**. This is the largest structure (in sq feet) associated with a PIN in the GPIN. This is an integer.

**UseCode\_All**. This is a list of all the use codes of the PINs associated with the GPIN. This is a pipe-delimited text field.

**Max\_UseCodeRisk**. This is the maximum value of the use codes of the PINs associated with the GPIN. This has a value from 0 – 3, where 0 represents the lowest risk and 3 represents the highest risk.

**Source\_All**. This is a list of all the sources (residential or commercial) for the PINs associated with the GPIN. This is a text field.

Hopefully this is a start. It is clear that the field names should be changed to be more consistent. Also, should check with the City of Charlottesville staff to ensure the data are interpreted correctly. The basic information will remain the same, but there will be changes in the future.

Doug Caldwell

[caldwellinva@verizon.net](mailto:caldwellinva@verizon.net)

11/7/2019