

DS at Scale Capstone: Blight in Detroit

Problem Description

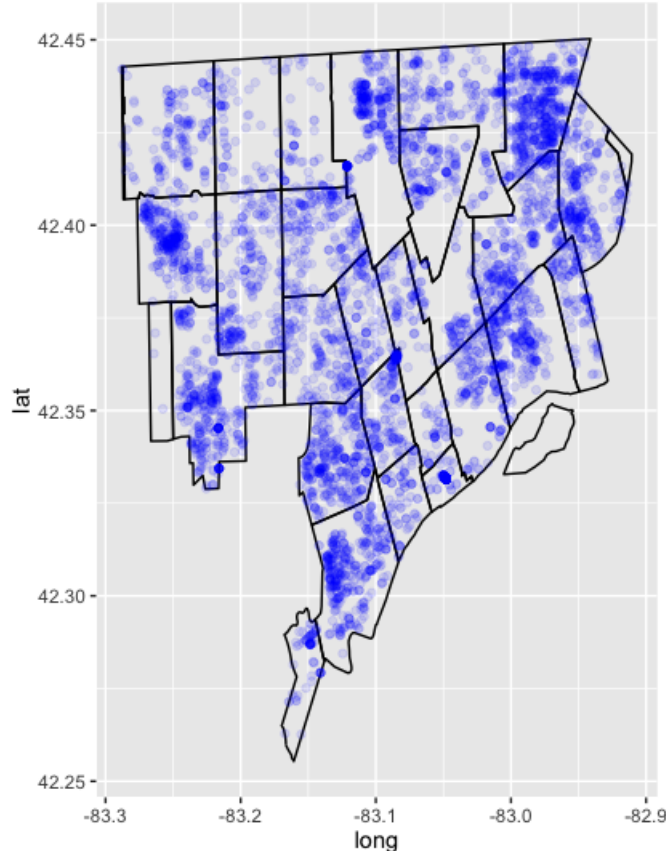
The purpose of this project was to build a binary classification model of blighted buildings in Detroit. The goal was to predict if a building was blighted or not using publicly available data.

The datasets used in the project included parcel information, permit violations, municipal service requests, crime, and demolitions in the Detroit area. The response variable was generated by considering a building as “blighted” if at least one ‘Dismantle’ demolition permit was associated with that building. The remaining datasets were used to create features to help predict if a building was blighted or not.

Data description

The data for this project came from five different sources: detroit-demolition-permits.txv, detroit-311.csv, detroit-crime.csv, detroit-blight-violations.csv, and Parcel_Points_Ownership.csv. These files contain information about demolition permits, crime reports, blight violations, 311 calls, and parcel information respectively; and information about how they can be obtained is listed in the appendix below.

In all of these data files incidents were recorded along with a latitude and longitude (lat/lon) where the incident occurred. As part of the data cleanup process, incidences with lat/lon well outside of the Detroit area were discarded. Namely, records where the latitude was less than 42.25 or greater than 42.5 were discarded; and records where the longitude was less than -83.3 or greater than -82.9 were discarded. Similarly, records where lat/lon was missing were discarded. A plot showing the locations of the blighted buildings after this data cleanup is shown below.



Methods

Deriving Buildings

The Parcel Points Ownership dataset was used as the ground truth for what defined a building. This dataset provides the locations as lat/lon pairs and addresses of properties in Detroit.

In all the datasets, the lat/lon pairs were converted to Universal Transverse Mercator (UTM), which changes the lat/lon pairs to points on a two-dimensional Cartesian grid. Because Detroit is a relatively small geographic area, little relative geographic fidelity is lost through this conversion.

Following the generation of UTM coordinates, records in the demolition permits, crime reports, blight violations, and 311 calls datasets were associated with the same building identifiers as those in the parcel points ownership dataset if the record's geographic coordinates were within 30 meters of the building's location. If a record fell within the extents of multiple buildings, that is, if the record was within 30 meters of the center of multiple buildings in the parcel points dataset, that record was assigned to the closer building.

Appendix

Data

Data for 311 calls, blight violations, crime, and demolitions is available at the Data Science at Scale github page: https://github.com/uwescience/datasci_course_materials/tree/master/capstone/blight

Parcel data is available from the Detroit Open Data Portal here: <https://data.detroitmi.gov/Property-Parcels/Parcel-Points-Ownership/eijm-6nr4/data#Export>