

Final Project Proposal: Predicting Philadelphia Street Repaving in 2024

Using 2023 data to train the model, I will attempt to predict which streets in Philadelphia will be repaved in 2024. I will be using data that is common across 2022 and 2023.

The dependent variable for the model is whether or not the street was scheduled for repaving in 2023. This information comes from the [Philadelphia Streets Paving Program and List](#). The independent variables for the model include factors that could cause the wear and tear of the street or increase incentives to repave/restripe:

1. [311 Calls about Potholes](#)
2. [Whether Street is a Non-Thru Street for Trucks](#)
3. [Street Classification](#) (i.e. Major arterial, minor arterial, collector, etc)
4. [Number of Fatal Crashes](#) - OTIS often aligns their restriping/safety plans with streets that are about to be repaved
5. [Level of Bike Stress](#) - A classification scheme based on the estimated comfort of bicyclists in traffic (number of lanes, effective vehicle speed, and presence/type of bicycle facility)

This information will be spatially joined with the Philadelphia street network. By making a that uses 2022 data to make a model that predicts the 2023 paving schedule, I will apply it to the 2023 data to predict the 2024 paving schedule.

The project will meet the following requirements:

- Data is collected through a means more sophisticated than downloading
 - While the 2022 311 data is available via OpenDataPhilly, the 311 data for 2023 has to be accessed via API. I'll be using the data from 1/1/2023 to 12/1/2023.
- It combines data collected from 3 or more different sources.
 - The data will be coming from OpenDataPhilly, the 311 API, and DVRPC.
- The analysis of the data is reasonably complex, involving multiple steps (geospatial joins/operations, data shaping, data frame operations, etc).
 - I'll need to clean /trim data inputs and spatially join them to the street network
- You use an osmnx or pandana to perform an analysis of street network data
 - Depending on the depth of the street classification data, I may use osmnx as the framework for the street network data.
- You perform a machine learning analysis with scikit-learn as part of the analysis.
 - I will be developing a model (likely random forest) to predict whether or not a street will be repaved