Proposal: Pothole Risk Map & Predictor

Project Title and Summary

Title: Predictive Heatmap for Syracuse Pothole Risk Using Citizen Complaints and Road Infrastructure Data

Summary:

This project aims to design and deploy an interactive, data-driven heatmap that visualizes pothole risk zones across Syracuse, NY, with the goal of informing both public awareness and city-level infrastructure planning. Using historical 311 service request data from the City of Syracuse's Open Data portal—specifically filtering for pothole-related reports—we will conduct a comprehensive temporal and spatial analysis of complaint patterns, repair timelines, and recurrence rates. By incorporating additional factors such as seasonal effects, neighborhood-level density, and road infrastructure features (where available), we intend to construct a predictive model that highlights areas at high risk for pothole formation.

The core deliverable will be a web-based interface or dashboard that overlays these risk scores on an interactive Syracuse map, allowing residents to understand the state of their neighborhood roads and enabling city officials to prioritize inspections, maintenance, and resource allocation. The project will also explore the use of large language models (LLMs) to classify and summarize complaint texts for urgency or sentiment, helping identify areas where residents report more severe or repeated issues. By making this risk model transparent and publicly accessible, the tool can foster greater accountability and collaboration between the city government and its residents.

Problem Statement :

Syracuse's aging road infrastructure, compounded by harsh winter conditions and fluctuating freeze-thaw cycles, has resulted in a persistent pothole problem that significantly impacts both quality of life and public safety. Each year, the city receives thousands of pothole-related complaints via its 311 service request platform. These complaints represent valuable crowdsourced insights into the condition of city streets—but currently, they are reactive in nature and largely inaccessible to the public in a meaningful format.

While the city collects and stores this data, there is no existing tool that visualizes long-term trends, identifies high-risk streets or neighborhoods, or predicts where potholes are most likely to reoccur. This absence of predictive infrastructure planning leads to inefficiencies in maintenance scheduling and prioritization. Additionally, many residents feel disconnected from the process, unaware of whether their complaints are addressed, and lacking visibility into broader infrastructure efforts.

This project seeks to address this disconnect by turning passive complaint data into an active planning tool. By analyzing historical 311 reports alongside spatial and seasonal patterns, we aim to build a predictive layer on top of existing records—highlighting pothole-prone zones and enabling the Department of Public Works to move from a reactive to a proactive maintenance strategy. The visualization will also serve as a civic engagement tool, empowering residents to see how infrastructure issues evolve over time and how their input contributes to solutions. Ultimately, this tool aspires to improve the transparency, efficiency, and equity of road maintenance efforts in Syracuse.

Data Sources

- **311 Service Requests (Cityline)**
    - Fields: date, location (lat/lng), category ("pothole"), status
    - Completeness: Excellent (over 10,000 entries)
    - Updated: Frequently
- **[Snow Plow Tracking / Winter Road Maintenance (if available)]**
    - May provide weather-adjusted pothole risk
- **Optional External Data**:
    - NOAA Weather API for Syracuse
    - Google Maps traffic overlays (for road usage)
    - Syracuse neighborhood shapefiles for visualization

Technical Approach

- **Data Cleaning & Aggregation**:
    - Filter 311 dataset for "pothole" complaints
    - Normalize location data to neighborhood zones
    - Engineer features: seasonal tag, complaint frequency, fix status lag
- **LLM Augmentation**:
    - Use GPT-4 to auto-classify complaint texts (e.g., urgency, sentiment)
    - Use LLMs to validate predicted risk explanations
- **Predictive Model**:
    - Build risk scores per neighborhood using historical trends
    - Optional: train simple time-series or ML models (logistic regression, XGBoost)
- **Visualization**:
    - Use Folium or Plotly to create an interactive heatmap
    - Deploy as static HTML or Streamlit app

Deliverable

- 📊 **Interactive Web Map** highlighting pothole-prone streets or areas
- 🧠 Risk model trained on real 311 history
- 📄 PDF report explaining insights
- Optional: mini chatbot to query pothole trends by neighborhood

Success Criteria

- 311 pothole data correctly mapped and cleaned
- At least 5 visualizations generated during exploration phase
- Functional map showing real-time or historical risk zones
- Project usable by city staff or public
- README + TECHNICAL.md + METHODOLOGY.md documentation present

Timeline (Weeks 1–10)

Week	Goal
1–2	Dataset exploration + final proposal
3–4	Data cleaning + EDA + risk feature design
5–6	Risk model v1 + initial map prototype
7–8	Add LLM layers + refinement
9–10	Final map polish + report + deploy site
11–12	Documentation, testing, submit to city

⚠️ *Risks & Mitigations*

- **Incomplete Location Data** → fallback to ZIP/neighborhood-level mapping
- **LLM hallucinations in classification** → validate manually + prompt-tuning
- **Deployment issues** → static map + GitHub Pages as fallback to Streamlit