**82 Below Project Outline**

## Team Information

### Team Members and contact information:

* **Thailer Simmons** – [*tsimmons1@hawk.iit.edu*](mailto:tsimmons1@hawk.iit.edu)
* **Sebastian Buzenas** – [*sbuzenas@hawk.iit.edu*](mailto:sbuzenas@hawk.iit.edu)
* **Pranav Kuchibhotla** – [*pkuchibhotla@hawk.iit.edu*](mailto:pkuchibhotla@hawk.iite.du)
* **Gavin Coffer** – [*gcoffer@hawk.iit.edu*](mailto:gcoffer@hawk.iit.edu)
* **Obaid Bin-Mahfoudh** – [*obinmahfoudh@hawk.iit.edu*](mailto:obinmahfoudh@hawk.iit.edu)

### Team name:

* 82 Below

### Motto:

* Visualize, Innovate, Eliminate

### Logo:



### Deliverable Sharing Permissions:

* The project deliverable may be shared to everyone, both in and out of the Illinois Tech community.

## Problem Summary

### A. Problem we are addressing. Links & citations, context, why problem is important, the benefit we would be providing

We are addressing the lack of clarity of visible information regarding lead pipeline data in Chicago.

**Problem:**

Chicago's outdated lead service lines pose a serious public health risk, particularly affecting the vulnerable populations such as children. Despite the significant impact this causes it is hard to find clear information regarding one's risk for lead exposure, both at home and beyond. Our project focuses on mapping and analyzing lead contamination using geospatial data to visualize and quantify the impact of lead poisoning in different Chicago neighborhoods.

### B. Why the problem is important to business, govt, industry, civic interest

1. High lead levels in drinking water are linked to cognitive and developmental issues, especially in children.
2. Chicago has one of the highest numbers of lead service lines in the U.S., hence requiring urgent policy intervention.
3. Understanding the distribution of lead contamination helps inform the public and policymakers for targeted action.

### C. Project Sponsor, Subject Matter Expert:

Our subject matter expert is Elin Betanzo. She is the founder and CEO of Safe Water Engineering LLC, a company aimed to identify, analyze, and combat contamination in drinking water through engineering and policy consulting.

### D. Brief explanation of our goals: preliminary, main, aspirational

#### Preliminary:

1. Collect and clean relevant data:
   1. Service line inventory data: A dataset containing addresses and their service line composition information (lead, not lead, etc.). It will display information on both the private and public side.
   2. Address type: A dataset containing a list of addresses in Chicago and what type of property is residing there i.e.: Residential, commercial, school, etc.
2. Generate basic visualizations (heatmaps) to identify high-risk areas.

#### Main:

1. Develop a geo-spatial analysis model to predict and label lead contamination risks.
2. Cross-reference lead levels with socio-economic factors like income levels, school locations, and more.
3. Create a website and dashboard that will present this information in an easy to consume way.
4. Provide ease of access to related datasets.

#### Aspirational:

1. If we can successfully map out lead risk factors from service lines, we could extend this to risk factors for other forms of lead contamination such as paint or other dangerous contaminants.
2. Creating a tool for Chicago residents and policy makers to learn and take action against lead.
3. Additions to the website so it has elevated visuals with animations that will elevate the delivery of the information that we want to present.

### E. likely outcome of our work, potential impact

1. Advanced data-driven insights on lead contamination levels and trends for Chicago neighborhoods
2. Community Awareness
3. Potential for Policy Influence

Supporting Detail

### Data sources needed, form and structure of data

* 1. Lead pipeline data (CoC, EPA)
     1. .csv format
     2. Contains an address and its associated lead service line information, e.g. if its lead or not both on public and private sides.
  2. Address longitude and latitude data
     1. .csv form
     2. Contains the latitude and longitude of a given address in Chicago
  3. Schools associated with a given address.
     1. .csv form
     2. Contains schools attending to a given address. Hopefully elementary, middle, and high school. Potentially daycares and pre-schools although that is not set for a given address.
  4. Demographic Data (Census & income reports)
     1. .csv form
     2. Contains information regarding a neighborhood's general racial makeup and income levels
  5. Labeling addresses: residential, schools, restaurant, commercial etc.
     1. .csv form
     2. Contains information on what type of property is at a given address i.e.: school, residential, restaurant etc.
  6. Lead-related hospitalizations
     1. .csv form
     2. Contains information about lead-related patient intakes at given hospitals
  7. Policy insights
     1. List form
     2. Current policies and what it means
     3. Past policies that might have prevented effective changes.

### Software tools that have been used/may be useful. Models, algorithms, AI, methods

1. Data Processing tools: Python (libraries: pandas, NumPy)
2. Geo Analysis and Visualization: Python (Geopandas), ArcGIS
3. Maybe ML models
4. Website frameworks: React
5. GitHub for coding collaboration and website hosting.

### Preliminary goals

1. Compile and clean datasets.
   1. Associate all Chicago addresses with a water service line
   2. Associate all Chicago addresses with a type (residential, school, commercial, restaurant, etc.)
   3. Associate all residential Chicago addresses with attending schools
2. Generate initial visualizations to assess contamination hotspots

### Main goals

1. Compile data for Chicago addresses and assess general lead risk factors for each address and neighborhood
2. Map the compiled data
3. Create a website dashboard that allows for easy viewing and access to this information

### Aspirational goals

1. Add additional lead risk factors to the compiled data, to include soil lead levels, potential existence of lead paint in homes, and age of homes to create a more holistic view of the potential for lead contamination in each community.
   1. Such a goal will require more datasets involving the additional variables
   2. A sort of “risk index” will have to be determined that accurately reflects show how each risk factor affects a community. Is lead in water more dangerous than lead in soil?
2. The visual appeal of the website is a final step that we can take to make the website more attractive and clickable. While we are most concerned with the presented data being accurate and easy to access, a website that catches attention will aid in spreading the message of the importance of lead contamination awareness. The more people are aware of the issue, the more people will be able to buy personal water filters for their home or will be able to lobby or protest for change in their area.

### Are our goals reasonable within the timeframe/resources available?

1. Our team should be able to complete the main goals laid out to their full extent. After compiling the data, most of our time is going to be spent creating the website. Any time devoted after creating a foundational website will be spent adding features and innovating it.

### Resources/methodologies team needs to better master to be successful. Tutorials and stuff

1. **Methodologies:**
   1. GIS and mapping techniques for spatial analysis
   2. Use of Excel (or other spreadsheet editing applications) to scrub data for our specific variables of interest
   3. Use of GIS software to enrich feature layers with CSV datasets gathered from the City of Chicago
   4. Website development frameworks and software. Both front-end and back-end if applicable.
   5. Statistical random sampling.
2. **Resources:** Access to accurate and recent lead contamination datasets from local government databases.

### Current plans

1. Policy legislation
   1. Holes in current policies
   2. Mayoral ignorance?
   3. Identify gaps in current policies addressing lead contamination
   4. Assess city wide efforts and potential improvements
2. Maps:
   1. Current addresses with lead pipelines
   2. Lead levels vs income
   3. Mapping lead pipelines to the locations
   4. Schools & restaurant exposé
   5. Lead contamination vs. income levels
   6. Identifying areas at highest risk
   7. Schools & restaurants at risk
3. Additional Information:
   1. Locality based prediction
   2. Predict future contamination risk based on historical trends
   3. Integrate socio-economic factors for targeted interventions