Project Plan Presentation
Air Pollution Health Outcomes
Forecasting Tool
The Capstone Experience
Team Anthropocene Institute 1

Lindsey Boivin
Tate Bond
Hannah Francisco
Zhendong Liu
Lukas Richters

Department of Computer Science and Engineering
Michigan State University
Fall 2021
Functional Specifications

• There is a lack of precise data about air pollution and health effects on the individual
• This information is important for the public to know, as poor air quality can result in an increase in infant mortality, lung cancer, asthma hospitalizations, and other health issues
• Keep consumers, researchers, and politicians informed about air quality in their area when making decisions about housing, personal health, and public policy
• Designing a web app that uses machine learning to predict air quality information in each area, and displays that with a map overlay
Design Specifications

• User may also scroll through an interactive map overlay
• Search function and map will display pertinent information about air pollution and health risks in the area
• User can compare air quality in two different locations using easy to understand graphical representations
• Analytics page shows best and worst air quality areas in the US
• “The Zillow of air pollution”
Screen Mockup: Web Interface
Screen Mockup: Web Interface
Screen Mockup: Web Interface
Screen Mockup: Web Interface
Screen Mockup: Web Interface
Technical Specifications

• Machine learning with Python using Scikit-learn
• HTML, CSS, and JavaScript user interface
• Backend with Python using Flask
• Application hosted on an Apache Server on capstone iMacs
• SQLite database
System Architecture

Machine Learning Model

Backend

User Interface

SQLite

scikit learn

Flask

APACHE SOFTWARE FOUNDATION

HTML

CSS

JavaScript
System Components

• Hardware Platforms
  ▪ Hosting on Capstone iMac
  ▪ VM Fusion with Windows VM
  ▪ Python virtual environment

• Software Platforms / Technologies
  ▪ Scikit-Learn
  ▪ Flask/Apache
  ▪ HTML/CSS/JavaScript
  ▪ Google Maps API
  ▪ SQLite database
Risks

• Data Collection
  - There is lots of data to collect from many sources without a standard organization of information
  - Currently collecting more data and preprocessing it so that we have a single format for our data set
• Handling Errors with Live Sensor Data
  - We need to be able to update the database with the newest sensor data every hour, but sometimes the sensors fail to update their status with the API
  - Experiment with different ways to manage old data, i.e. assign lower weight to prediction with that value, or use the old data until a certain threshold time
• Making Predictions with Sparse Sensor Data
  - Some areas have very few sensors, so predictions will be less accurate
  - Potentially offset the error with accurate population or GDP data
Questions?