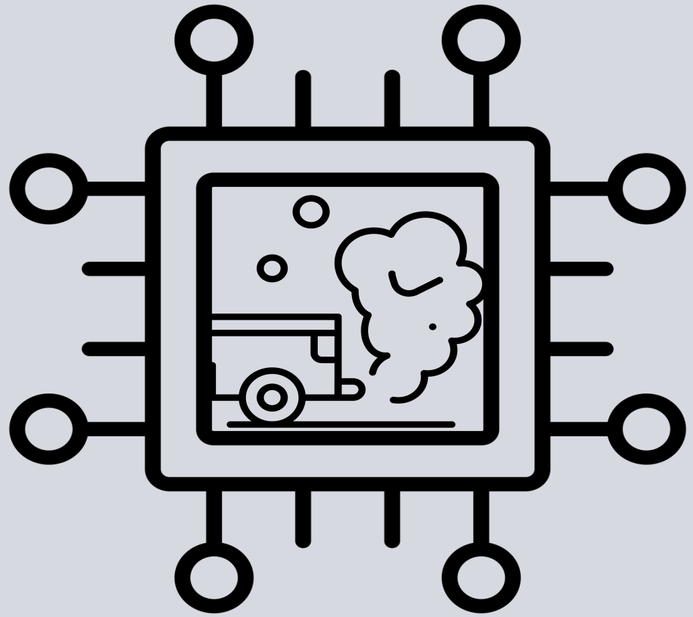
AI 4 AQ

Artificial Intelligence for Air Quality

A Case Study in the Salt Lake Valley



MIDS Spring 2024

Silas Gifford, Jared Feldman, Steven Tseng, Claira Kauffmann

Introduction

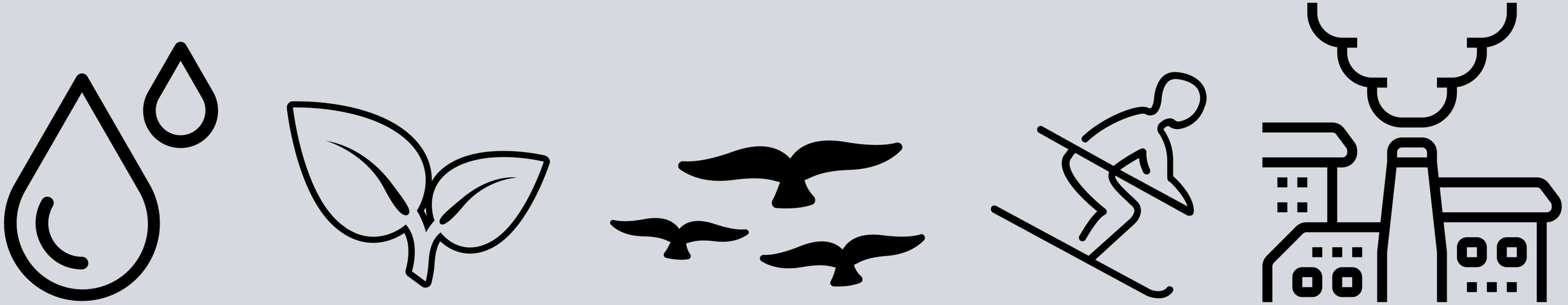
Silas Gifford

Jared Feldman

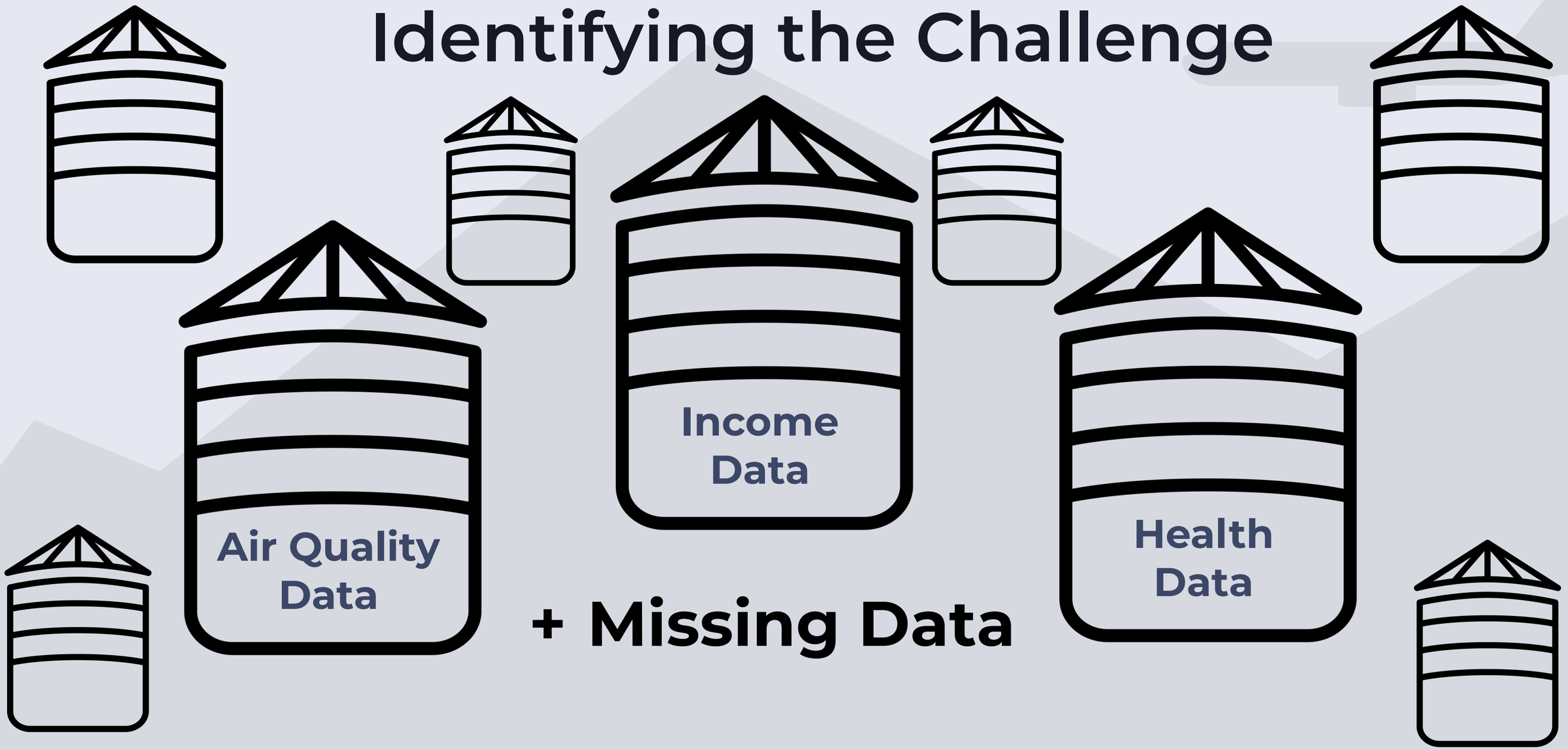
Claira Kauffmann

Steven Tseng

Understanding the Problem



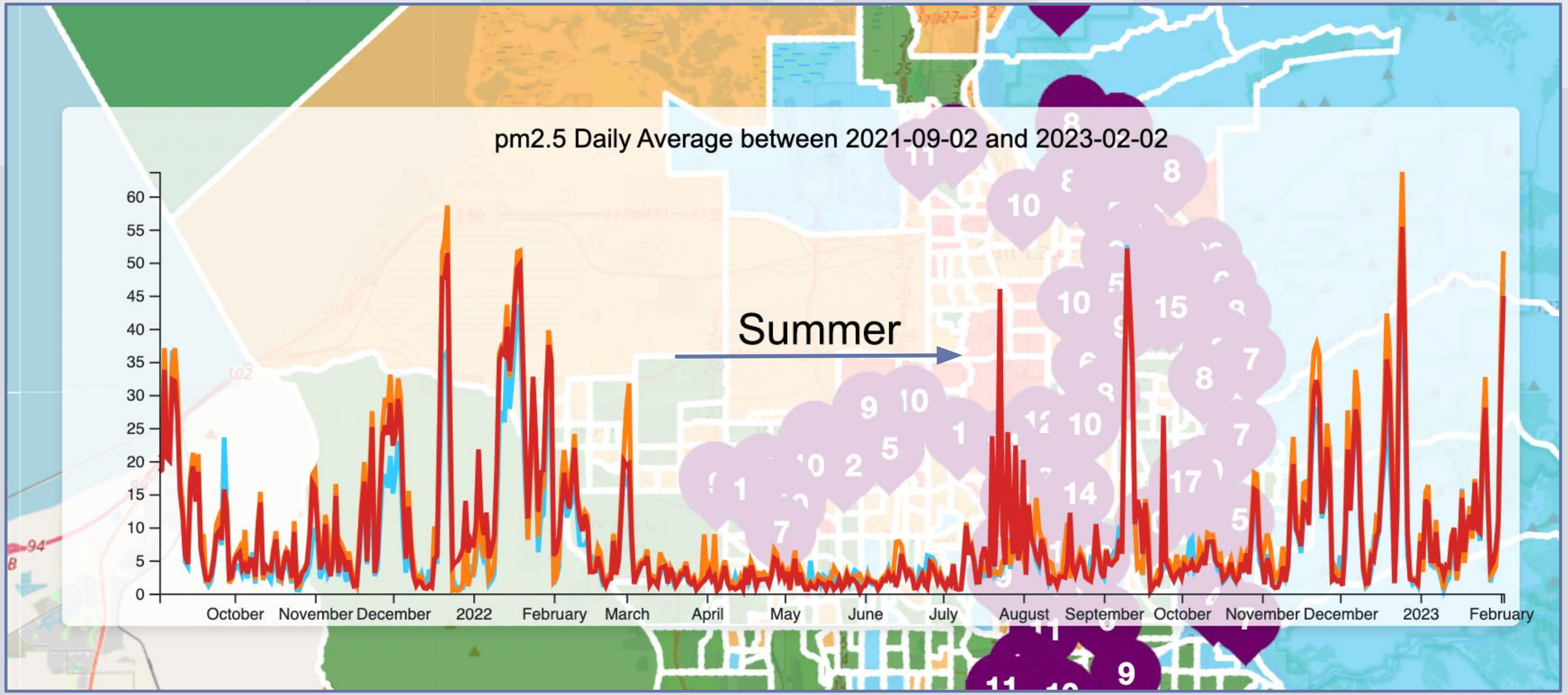
Identifying the Challenge



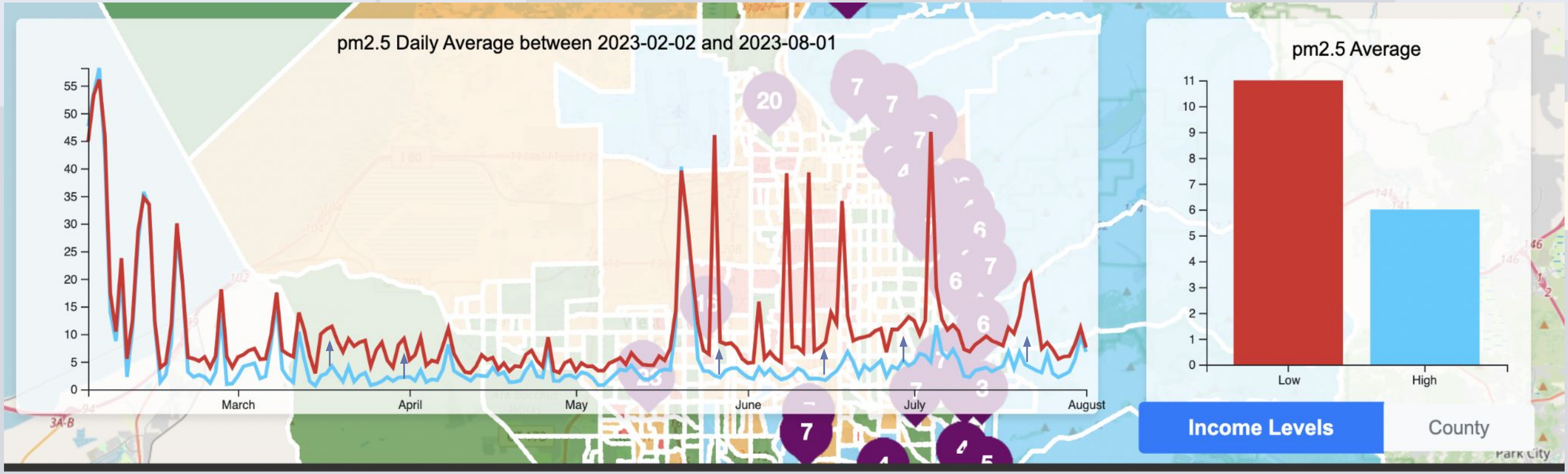
+ Missing Data

Demo

Dashboard Insights



Dashboard Insights



User Feedback

“The air quality is sometimes so bad that I can’t see the mountains that are less than a mile from my house. This dashboard highlights areas where there is better air quality, areas where it might be possible to go for a run on an otherwise bad air day.”

- Amanda Apgar, Local Salt Lake City resident

Technical Approach

Datasets

Source	Data Elements	Date Range	Granularity
PurpleAir	PM 2.5, PM 10	2016 - Present	Single sensor locations; daily
Department of Housing and Urban Development, Low to Moderate Income Population	Mapping, Income Level	2023	Census Block Groups
Medicare Claims	Count of diagnoses	5% sample 2016-2018 and 2020-2022, full data for 2019	Hospitals; weekly
Open-Meteo	Wind speed and direction	2016 - Present	Daily
OpenStreetMaps	Mapping	Current	All of the region of interest

Data Pipeline

Open-Meteo

OpenStreetMap

Low to Moderate Income Populations by Block Group

FEATURE

Medicare

PurpleAir

GitHub

Docker

CSV

JSON

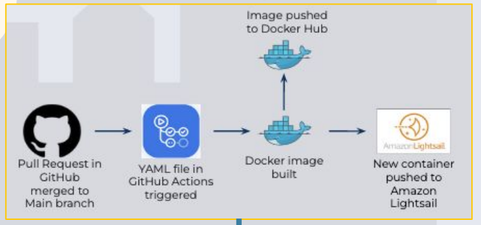
AWS S3

cron

AWS Lambda

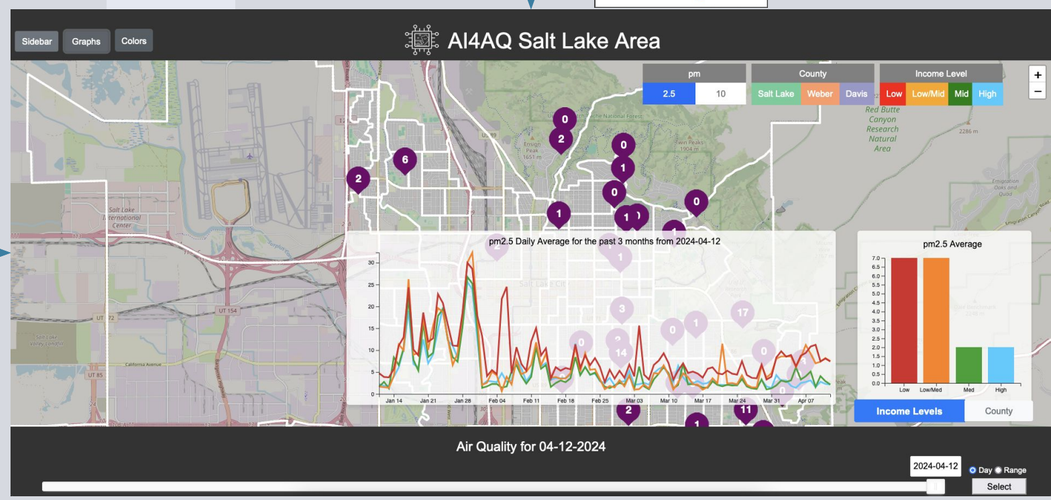
AWS Secrets Manager

CI/CD Pipeline



Leaflet

AmazonLightsail



Flask

JavaScript

HTML

SQLite

Modeling

Architectures:

1. Baseline A: County Average
2. Baseline B: Nearest Neighbor Sensor
3. XGBoost
4. Feedforward Neural Network
5. Feedforward + CNN features

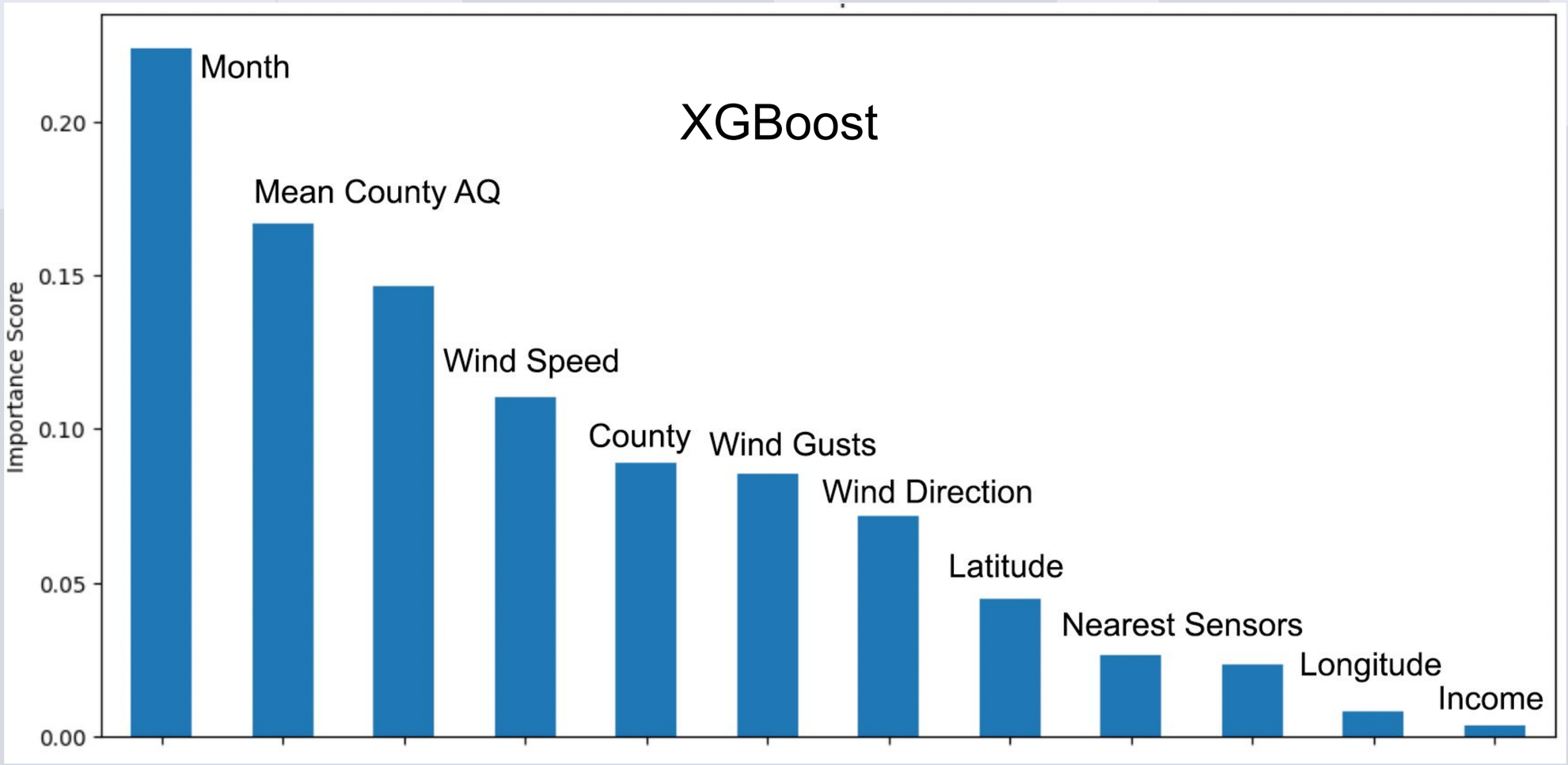
Features:

month, average PM 2.5, county, wind speed, wind gusts, latitude, wind direction, longitude, income category, nearest PM 2.5, nearest PM 10

Image Feature:

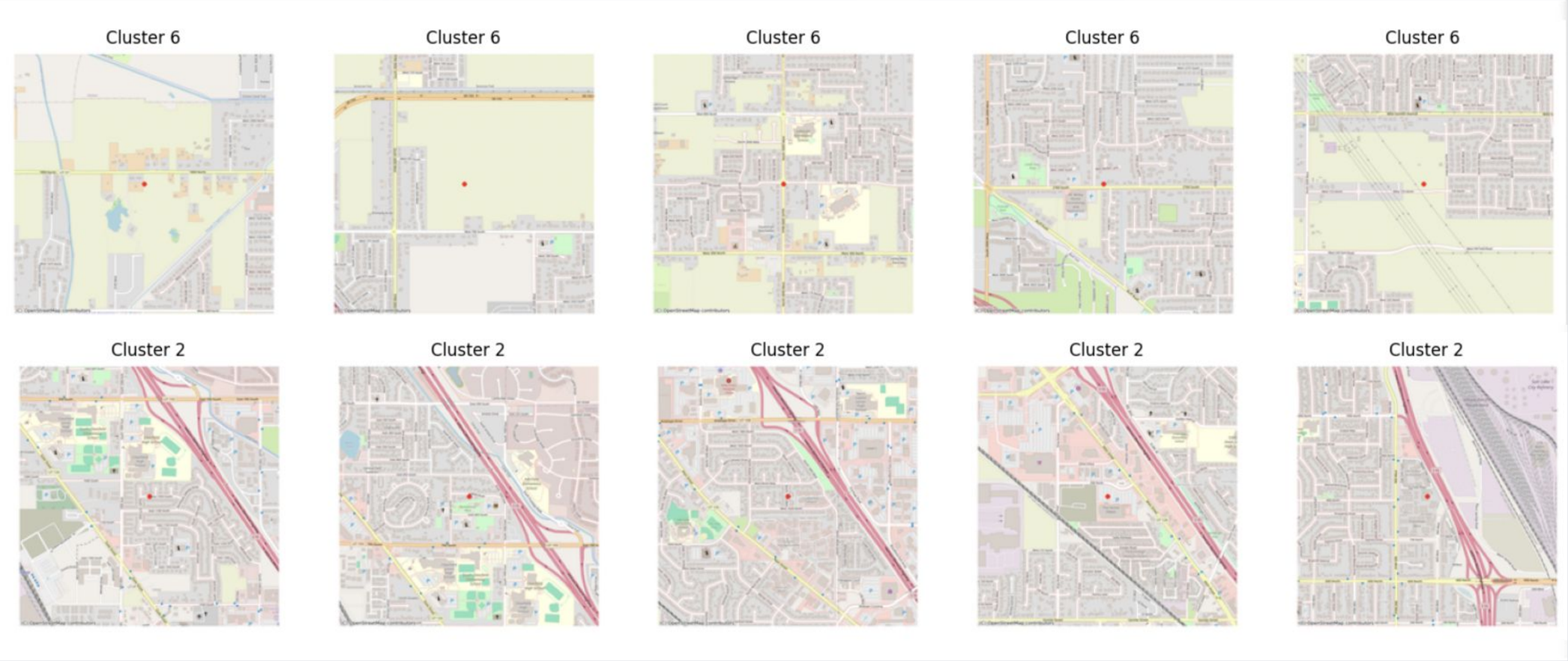
census tract map images

Feature Importance



Unsupervised CNN and k-means

912 overlapping map tiles clustered into 10 groups.



CNN Features + Feedforward

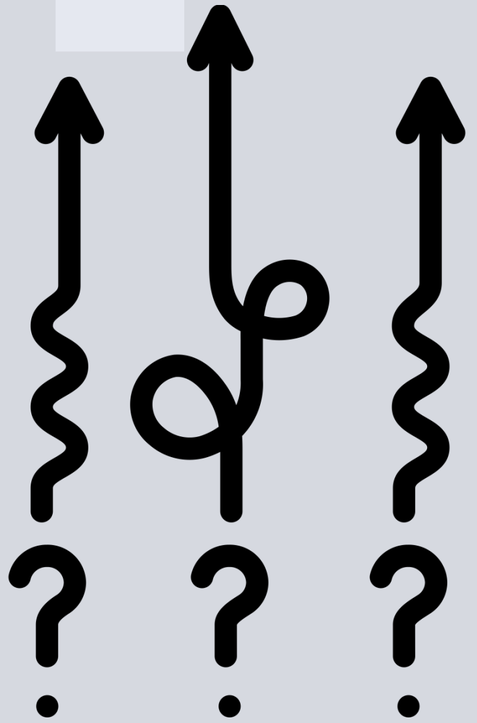
Layer (type)	Output Shape	Param #
input_layer_3 (InputLayer)	(None, 38)	0
dense_4 (Dense)	(None, 128)	4,992
dropout_1 (Dropout)	(None, 128)	0
dense_5 (Dense)	(None, 64)	8,256
dropout_2 (Dropout)	(None, 64)	0
dense_6 (Dense)	(None, 32)	2,080
dense_7 (Dense)	(None, 1)	33

Evaluation

	Mean Absolute Error	**Root Mean Squared Error**	**R-Squared**
Baseline - County Average	53.3397	80.5263	-0.07415
Baseline - Nearest Sensor	40.5587	230.4030	-0.02297
XGBoost	6.6377	12.0719	0.5166
Feedforward	4.6647	10.6322	0.6018
Feedforward + CNN features	**2.1563**	**8.7303**	**0.7311**

Technical Challenges

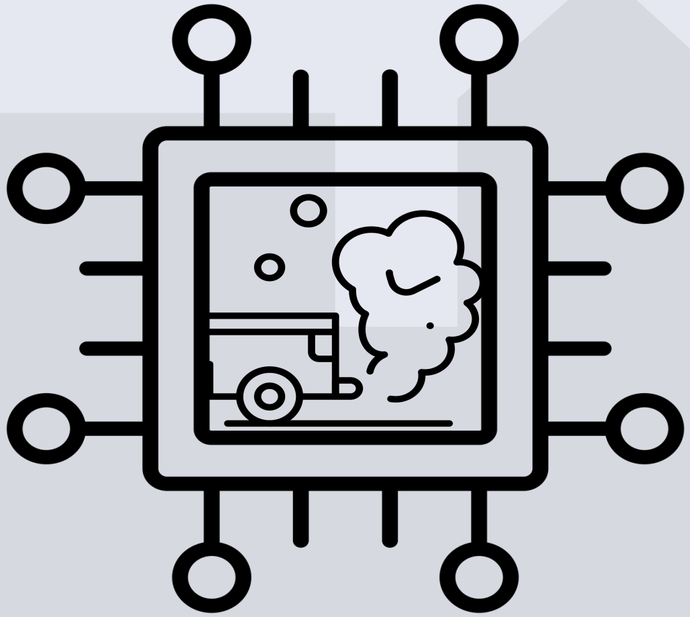
1. Data Limitations
   a. insufficient data
   b. inaccurate data
2. Deployment
   a. application
   b. models



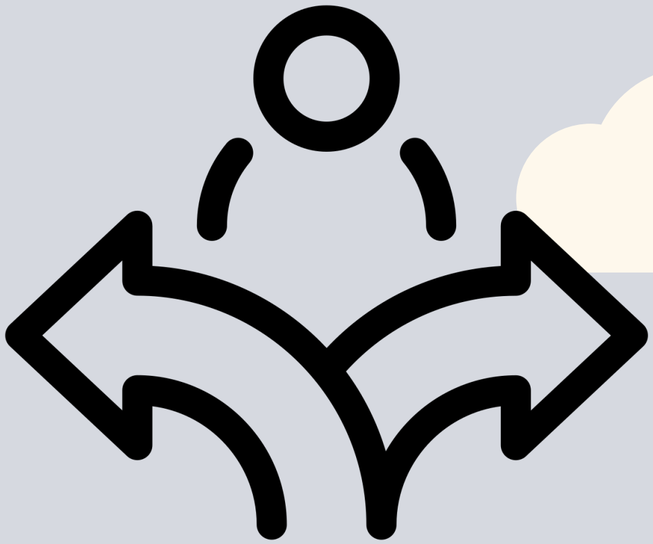
Next Steps

- Expand our dashboard beyond the Salt Lake Valley
- Establish a reliable network of PM 10 measurements
- Further modeling
  - could include physical models in addition to machine learning techniques as introduced with CMAQ-CNN
  - include additional historic data, such as historic OpenStreetMap tiles

Project Mission



AI4AQ helps address the air quality crisis in Salt Lake City. With machine learning and the combination of disparate data sources, our dashboard allows residents to gather information on their neighborhood's air quality and its effects, even if there is a gap in sensor coverage.

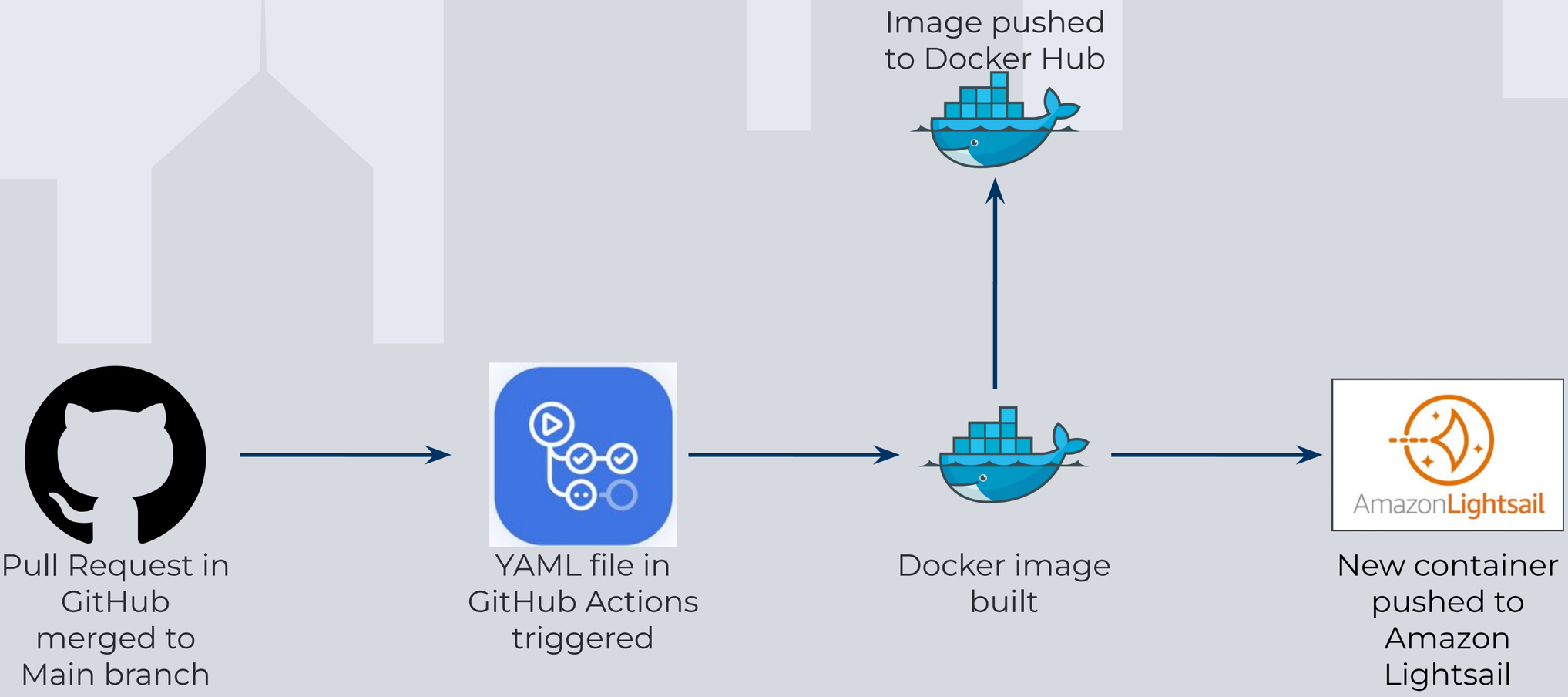


Appendix

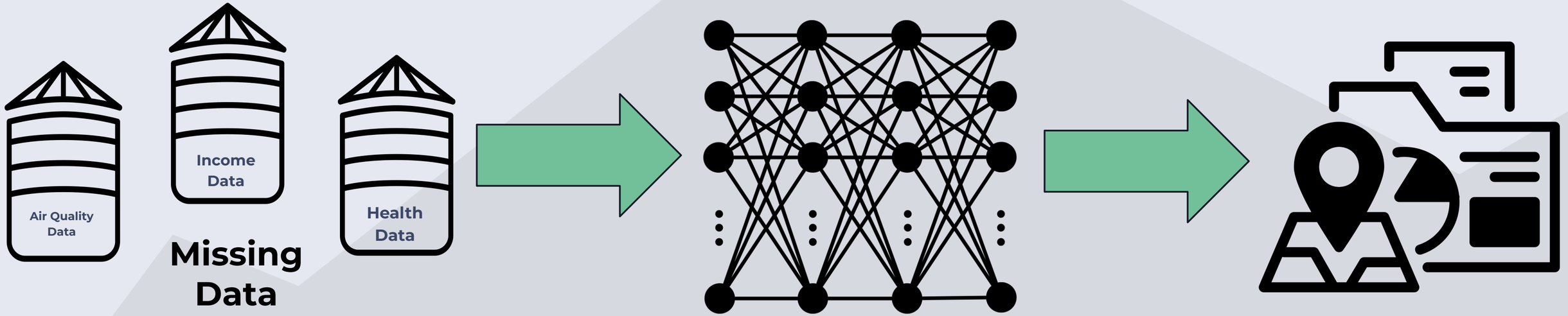
Acknowledgements

- Abdul Alfozan, Privacy and Ethical Review
- Rachel Edie and Chris Pennell, Utah Department of Air Quality
- Amanda Apgar, Local Residents who reviewed the dashboard
- Korin Reid, Medicare Data Access & General Advisement
- Joyce Shen, General Advisement

CI/CD Pipeline



Our Solution



AI4AQ helps address the air quality crisis in Salt Lake City. With machine learning and the combination of disparate data sources, our dashboard allows residents to gather information on their neighborhood's air quality even if there is a gap in sensor coverage.