Forward College

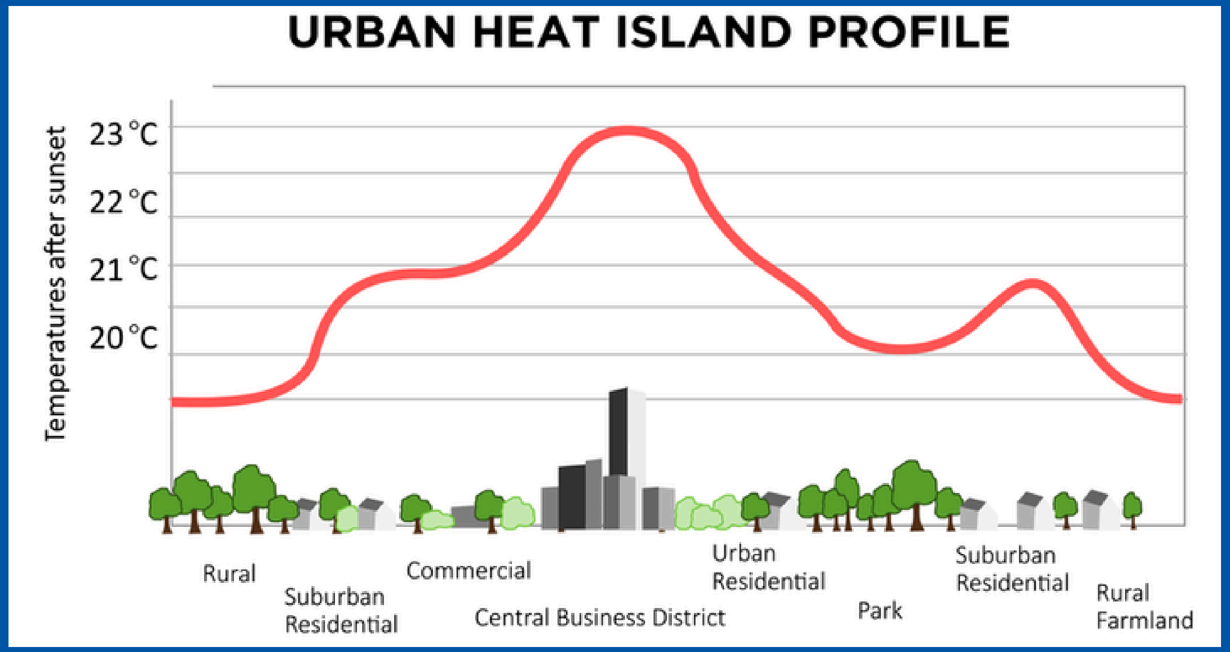
Heat Control Predictor

Using Machine Learning to assess heat-related mortality in cities

Tengku Hamza bin Tengku Mohd Uzaini Final Year Project

Problem Statement

- Urban heat islands increase heat-related mortality in major citites; policymakers lack predictive tools



- Who is this app targeted for?
  - City planners
  - Public health authorities
  - Citizens in dense urban areas

- Impact: Without intervention, heat stress can worsen, leading to higher mortality and economic burden

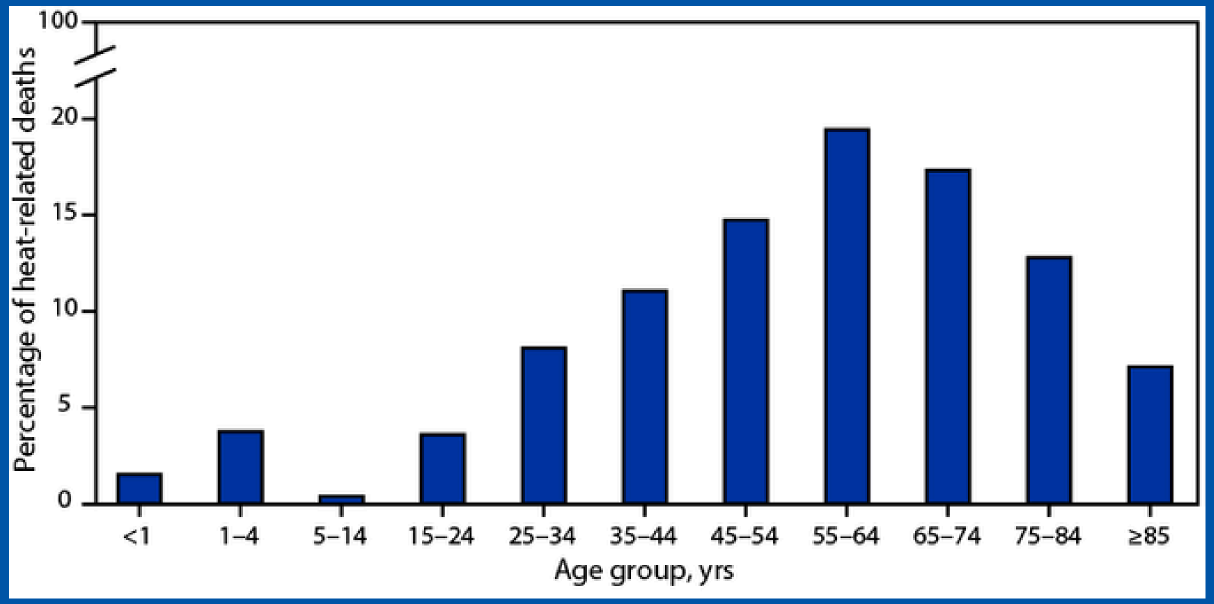
Data Overview

- Source: Kaggle
- Granularity: One row = one city observation
- Size: ~120k rows × 13 columns
- Target Variable: Health Impact (Mortality Rate per 100k)

	City Name	Latitude	Longitude	Elevation (m)	Temperature (°C)	Land Cover	Population Density (people/km²)	Energy Consumption (kWh)	Air Quality Index (AQI)	Urban Greenness Ratio (%)	Health Impact (Mortality Rate/100k)	Wind Speed (km/h)	Humidity (%)	Annual Rainfall (mm)	GDP per Capita (USD)
0	City_236	40.014907	135.759794	1657.234222	32.140727	Industrial	638.140184	29866.608120	90.967582	51.032061	20.060893	0.017909	46.681317	1750.054664	13551.38192
1	City_487	-81.752906	67.784550	1781.007943	28.199772	Water	2757.814606	25461.567500	121.919061	17.819991	45.591306	1.585266	43.291975	758.591768	41967.28373
2	City_21	20.126899	33.924075	3140.598901	11.492930	Water	6020.462986	2539.737270	169.190188	51.045248	10.525874	8.614523	76.935296	2494.912602	17335.37251
3	City_216	-47.308667	154.638241	992.282813	34.909265	Green Space	9491.952711	32146.724390	90.989624	19.211930	49.900393	0.297596	42.739059	1762.646698	31400.53605
4	City_292	-83.425194	31.018268	597.192562	28.465786	Water	5191.476501	2513.126338	92.082516	51.178231	21.993127	19.375498	43.509082	2134.723878	27399.94789

Objectives & Key Questions

- Objectives:
  - Predict urban heat-related mortality using city-level data
  - Understand which urban features increase risk
  - Provide actionable insights for urban planning
- Analytical Questions:
  - How does population density, greenness, and energy consumption affect mortality?
  - Which cities are most vulnerable to heat stress?

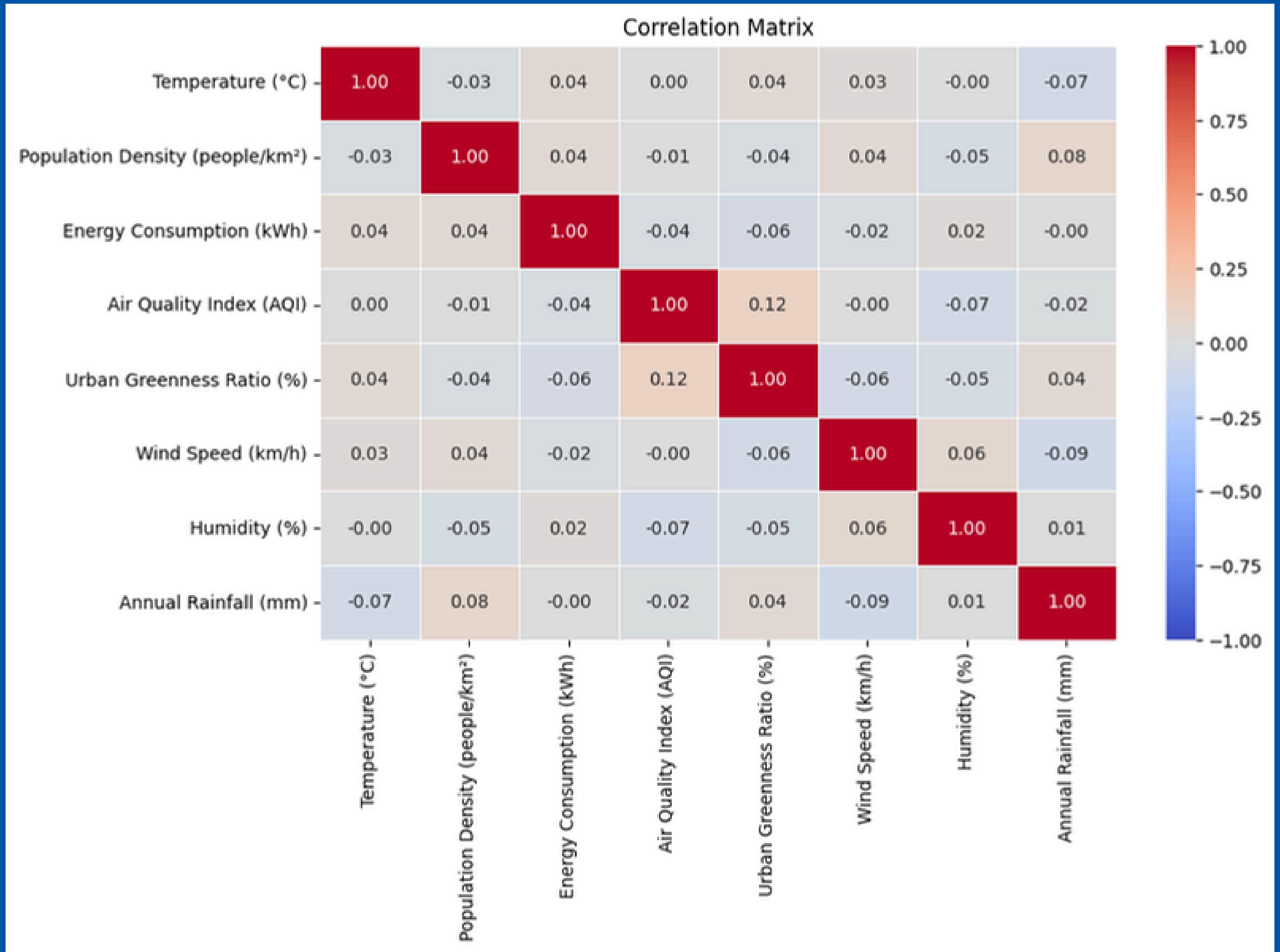


Methodology Overview

- Pipeline:
  - Data Preprocessing
  - Exploratory Data Analysis
  - Feature Engineering
  - Model Training
  - Model Tuning
  - Deployment

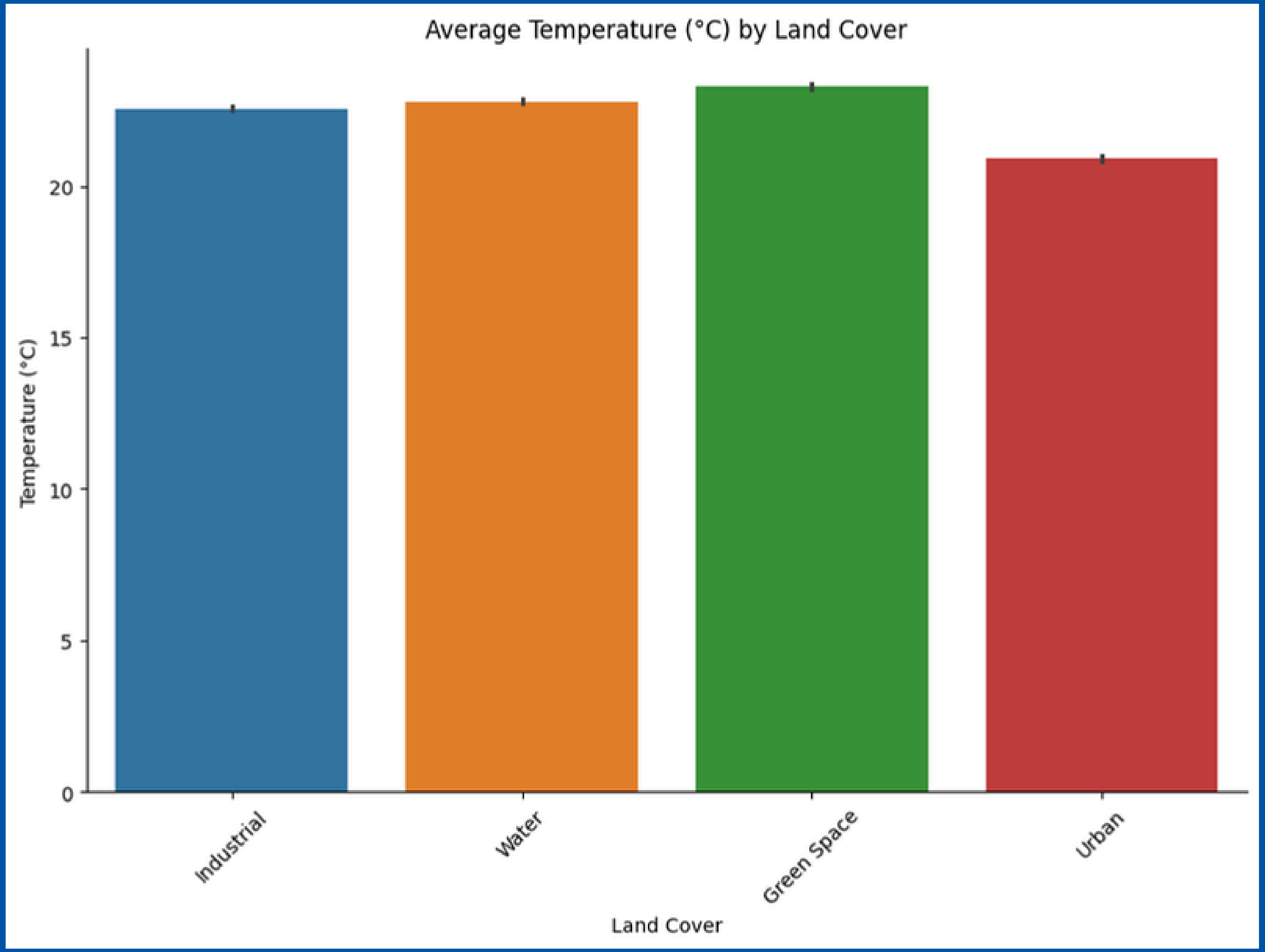
- Key Steps:
  - Handle missing data
  - Create derived features
  - Scale numeric features

EDA Key Findings (1)



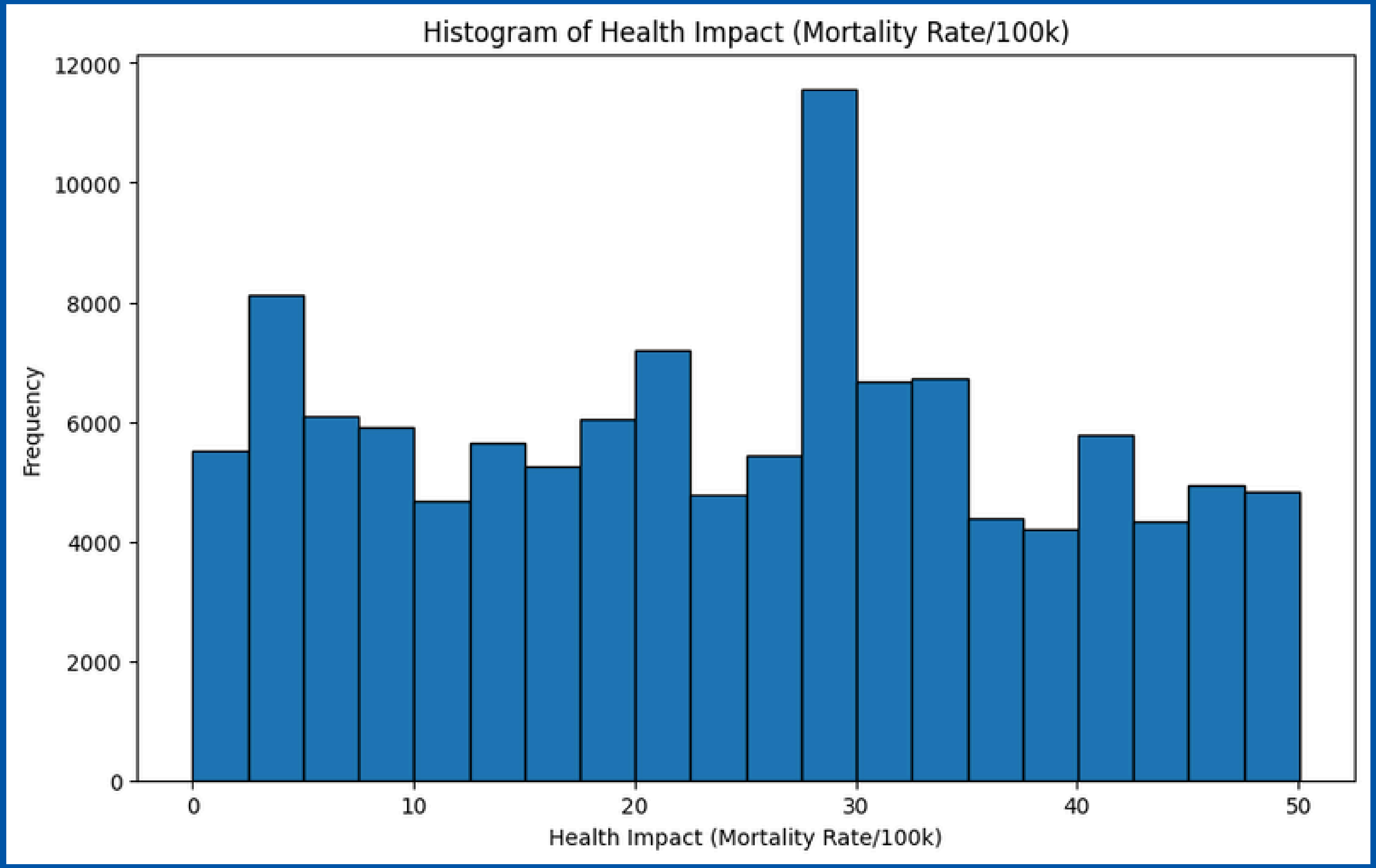
There appears to be no major correlation which means that the features are not strongly related to each other. A linear regression model would not be a good fit for this dataset.

EDA Key Findings (2)



The data is mostly even, however greener land cover has a higher temperature than urban land cover. This is unusual because urban areas are usually warmer than vegetated areas due to the urban heat island effect. The higher temperature in greener areas in this dataset might be due to some bias, so the Urban Greenness Ratio will be a better indicator.

EDA Key Findings (3)



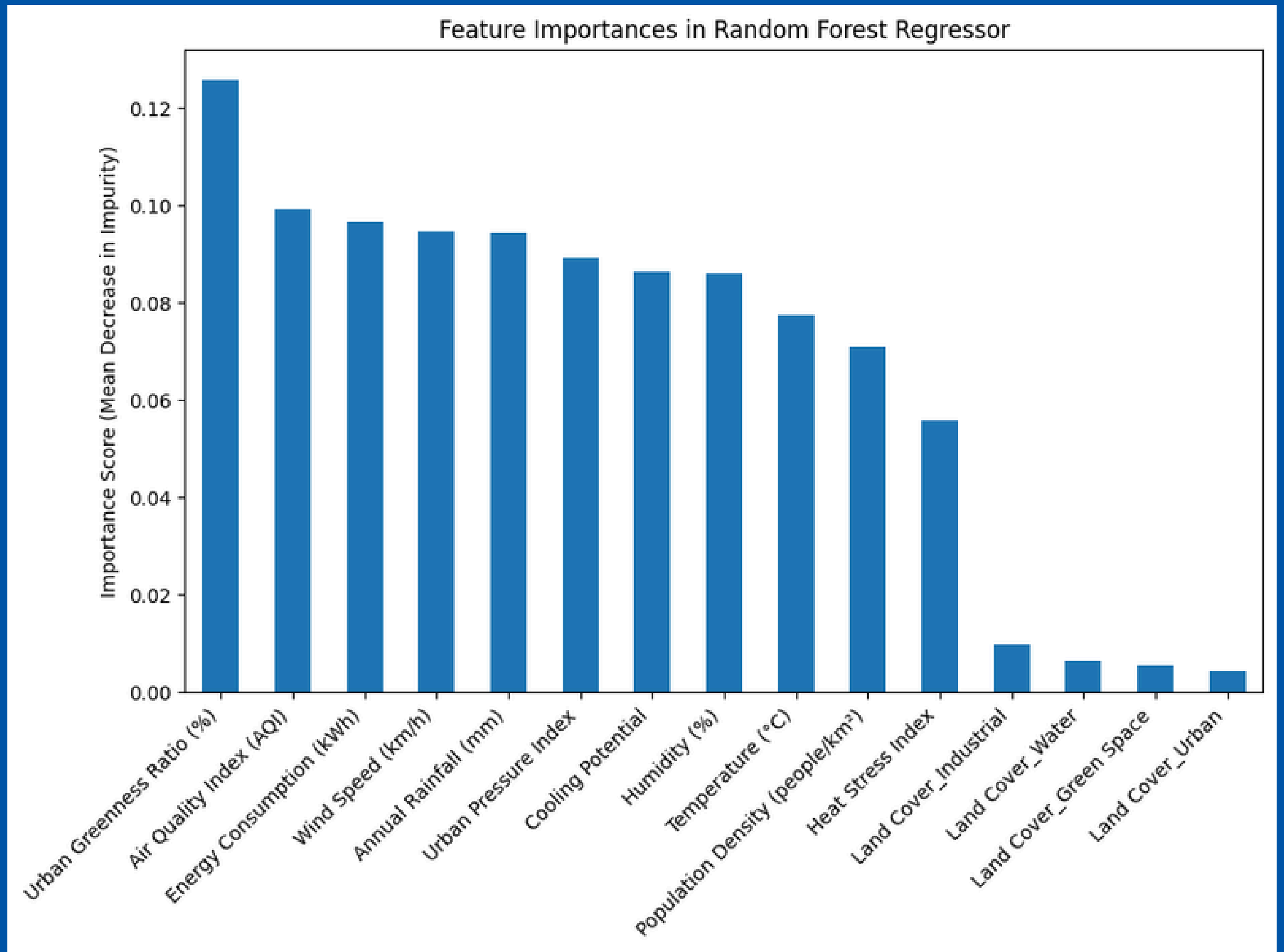
The data is also mostly even, however there is a spike in frequency at around 30 deaths. This is most likely due to it being the average among the dataset, so that's where we will set the threshold for health impact value interpretations.

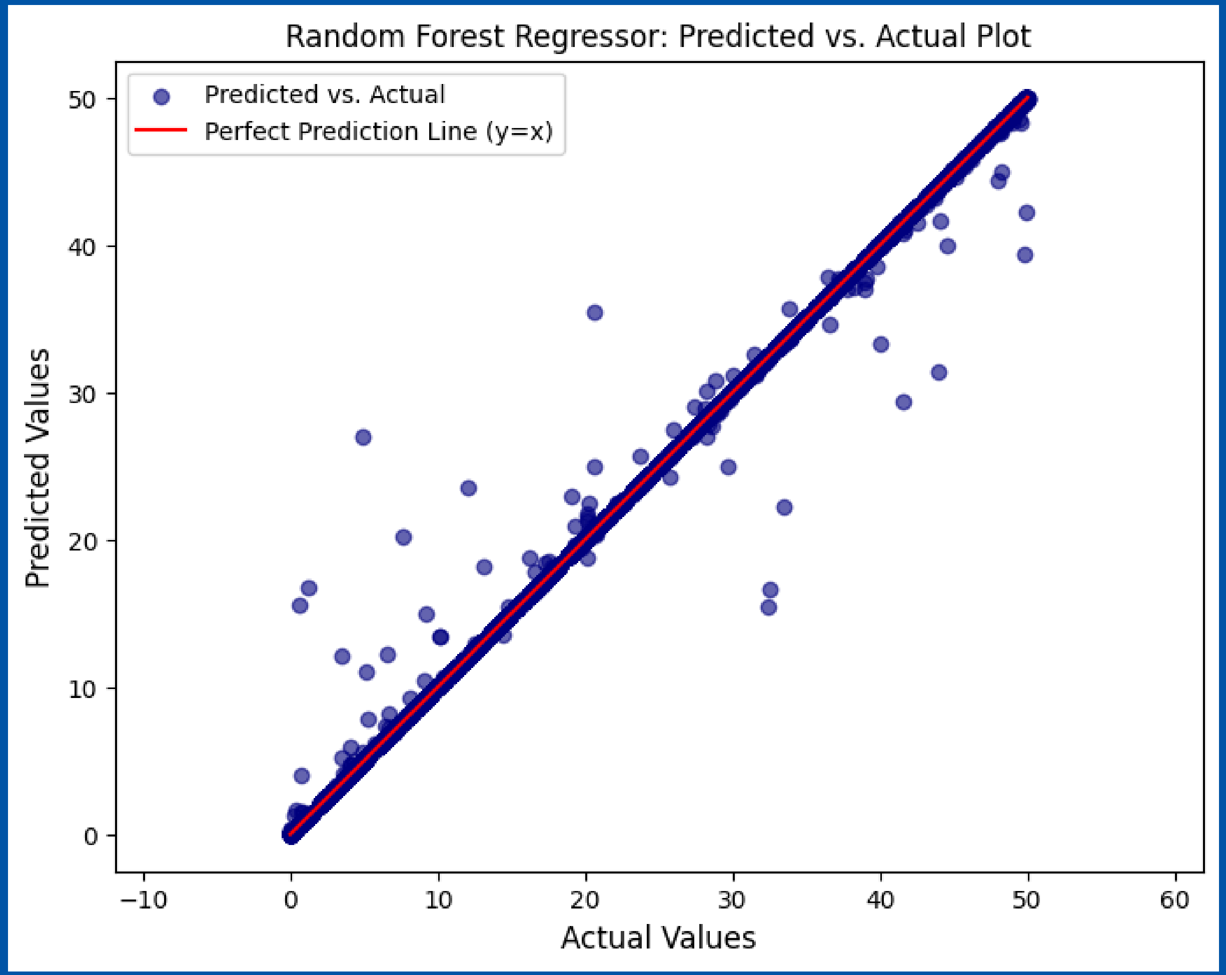
Modeling Approach

- Algorithms:
  - Tuned Random Forest Regressor as the prediction model
  - Baseline Linear Regressor and Gradient Boosting Regressor for comparison
- Validation:
  - 80/20 train-test split
- Feature Engineering:
  - Scaling numeric columns
  - One-hot encoding for land cover
  - Derived features (Heat Stress Index, Urban Pressure Index, Cooling Potential)

Results and Evaluation

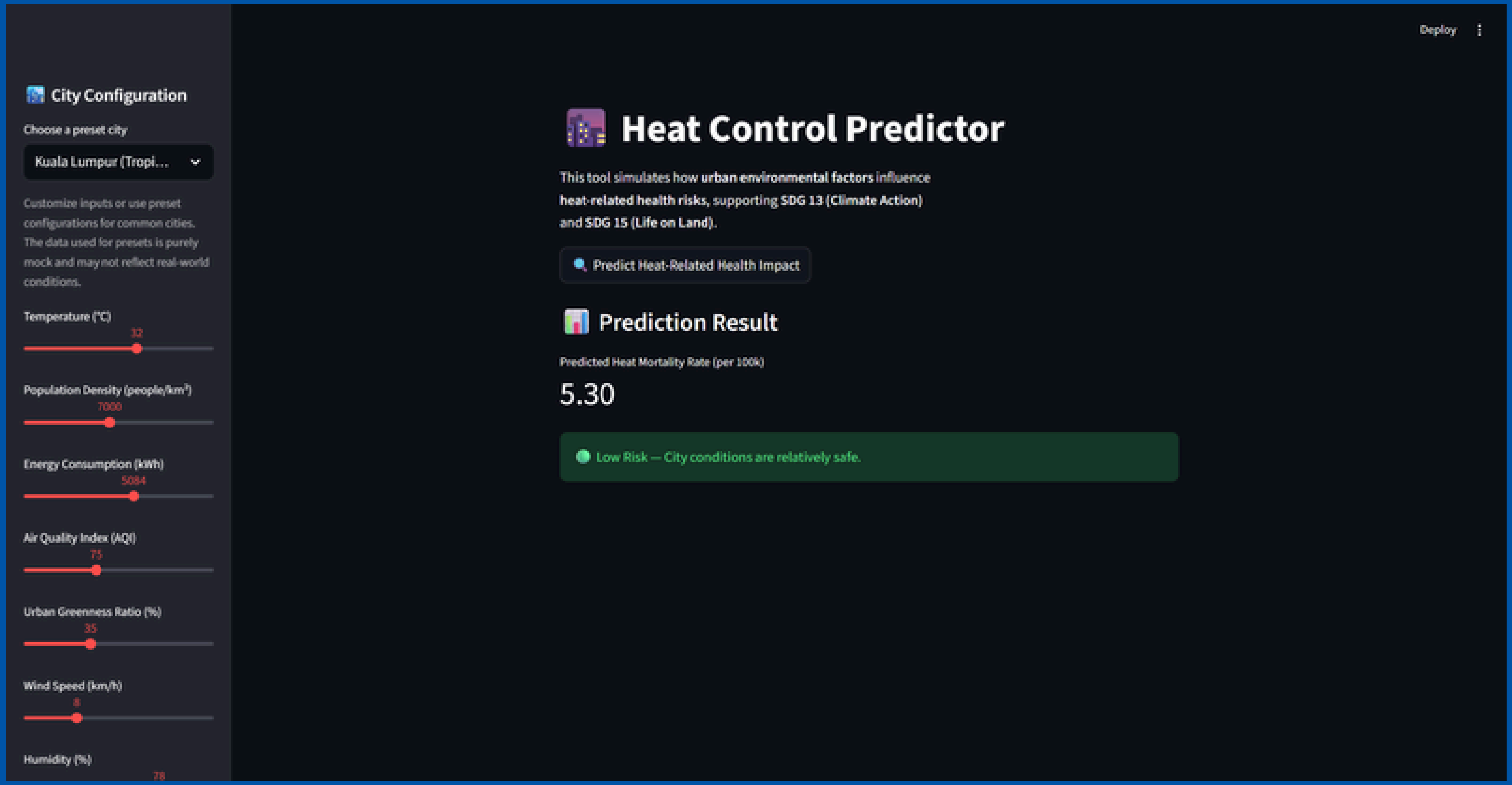
- Metrics:
  - MAE: 0.03
  - MSE: 0.13
  - RMSE: 0.36
  - R^2: 1.00
- Insights:
  - Model predicts mortality accurately; feature importance shows greenness, air quality index and energy consumption are key drivers





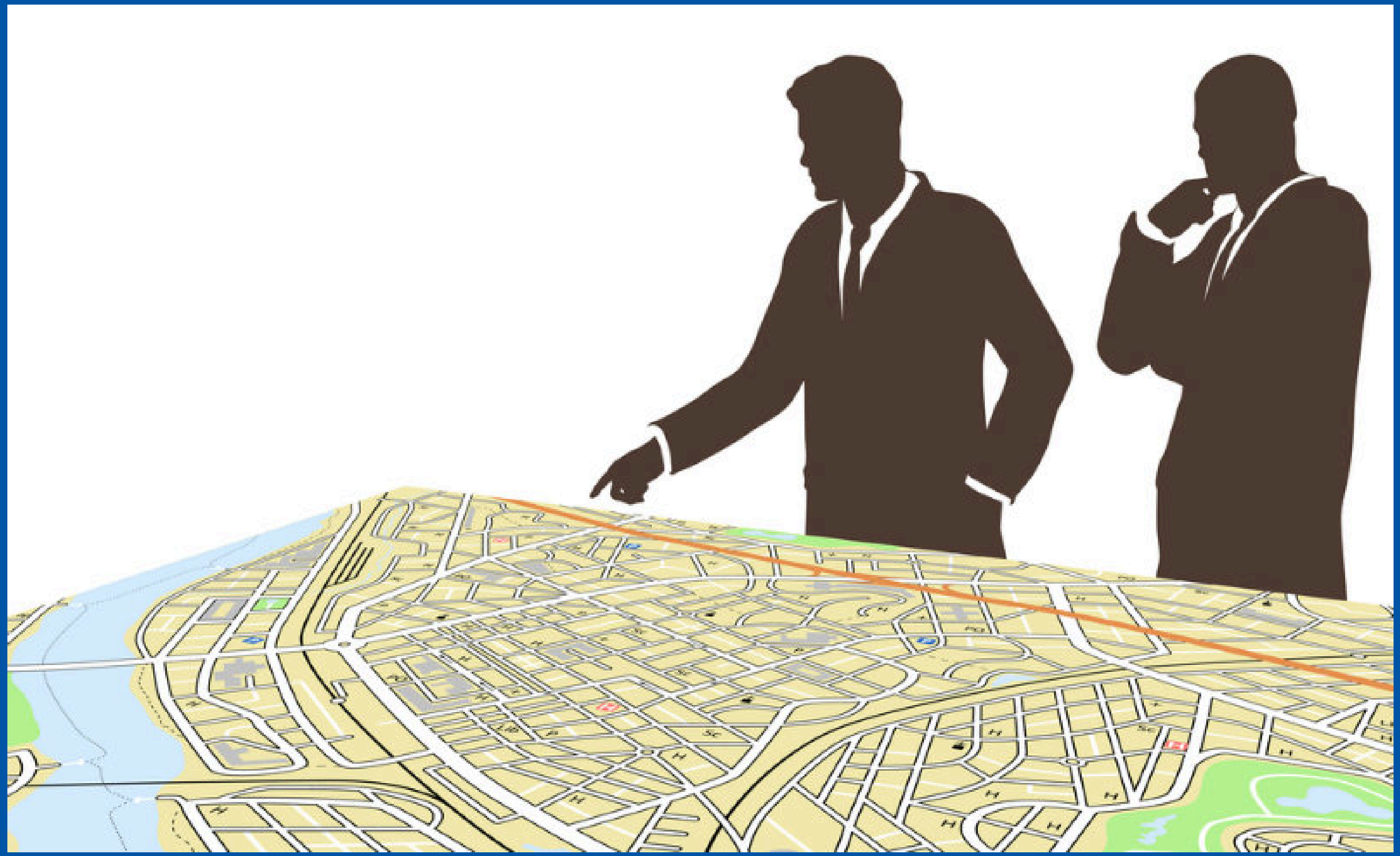
Project Demo

- Flow:
    - Input variables (AQI, Temperature, etc) or choose a preset
    - The model predicts the health impact mortality from inputted values
    - The model comes up with a prediction as well as an interpretation (Low Risk, Moderate Risk, High Risk)



Measures Of Success

- Target Metric:
  - RMSE < 0.5 ✅ achieved
- Business KPI:
  - Prioritize interventions in cities with highest predicted mortality and vulnerability



Challenges and Limitations

- Large joblib model files
    - GitHub push issue, a new repository had to be created
- Long tuning time
    - Tuning took over 5 hours, a heavy load on my laptop

Future Work & Recommendations

- Test additional models (XGBoost, Deep Learning)
- Add geospatial analysis or satellite imagery for inputs

Full Tech Stack

- Language:
  - Python

- Infrastructure:
  - Jupyter Notebook
  - Github
  - Streamlit

- Libraries:
  - Pandas
  - Scikit-learn
  - Joblib
  - Streamlit

That’s All, Thank You