Georgia Tech

CSE 6242 | Team 60 | Tox-e-mapper

Arya Kalappurayil, Cecili Poole, Diane Egret, Rhee Kang

Spring 2025

Section 1

Introduction

- Monitoring toxic chemical releases is vital for public health and sustainable urban development
- Existing tools like EnviroMapper lack advanced data visualization and analysis features
- This limits stakeholders' ability to track trends, assess regulatory impact, and make informed decisions
- The project aims to create an interactive platform to improve accessibility and usability of TRI data
- Enhanced tools will empower users to track pollution trends and support environmental risk mitigation

What's New

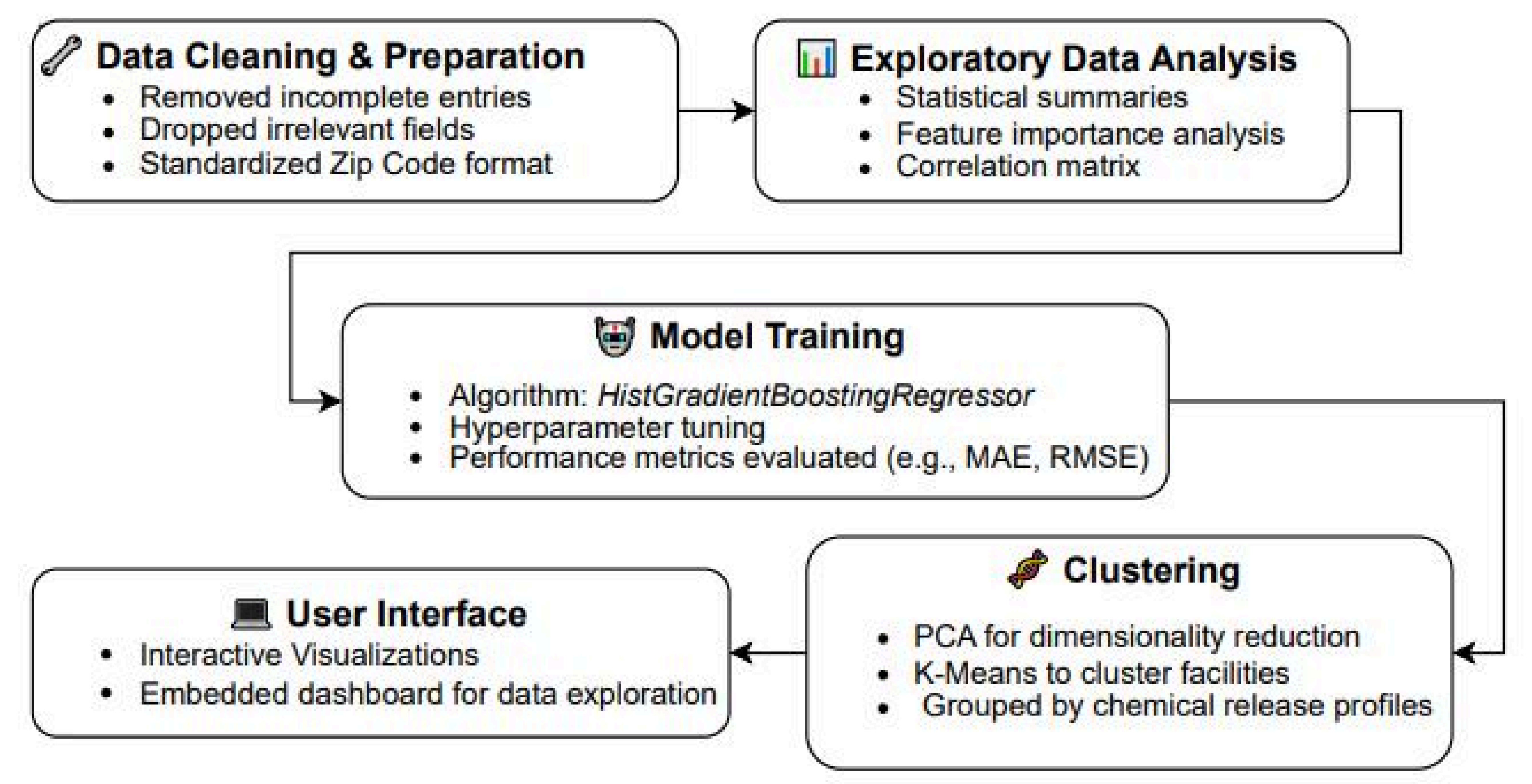
1. Dynamic Data Interactivity
2. Advanced Clustering Analysis
3. Geospatial Visualization

Section 2

Data

- Source: From the Toxics Release Inventory (TRI) provided by the Environmental Protection Agency (EPA)
- Size: 3 million records.
- Timeframe: Spans from 1987 to 2023
- Content: Includes chemical release quantities, facility information, and location data.

Methodology



Data Cleaning & Preparation
- Removed incomplete entries
- Dropped irrelevant fields
- Standardized Zip Code format

Exploratory Data Analysis
- Statistical summaries
- Feature importance analysis
- Correlation matrix

Model Training
- Algorithm: *HistGradientBoostingRegressor*
- Hyperparameter tuning
- Performance metrics evaluated (e.g., MAE, RMSE)

Clustering
- PCA for dimensionality reduction
- K-Means to cluster facilities
- Grouped by chemical release profiles

User Interface
- Interactive Visualizations
- Embedded dashboard for data exploration

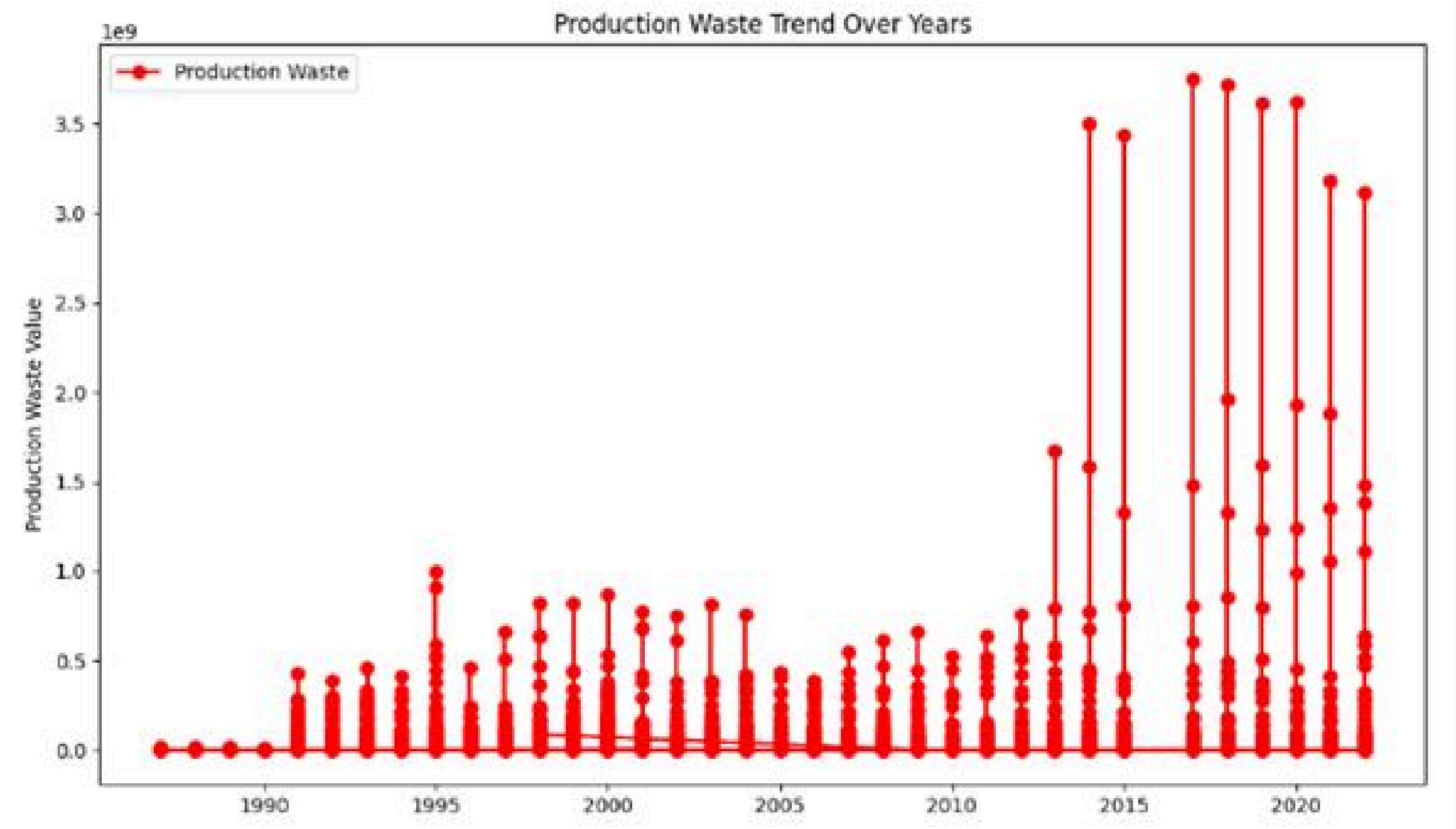


Fig1: EDA Result 1

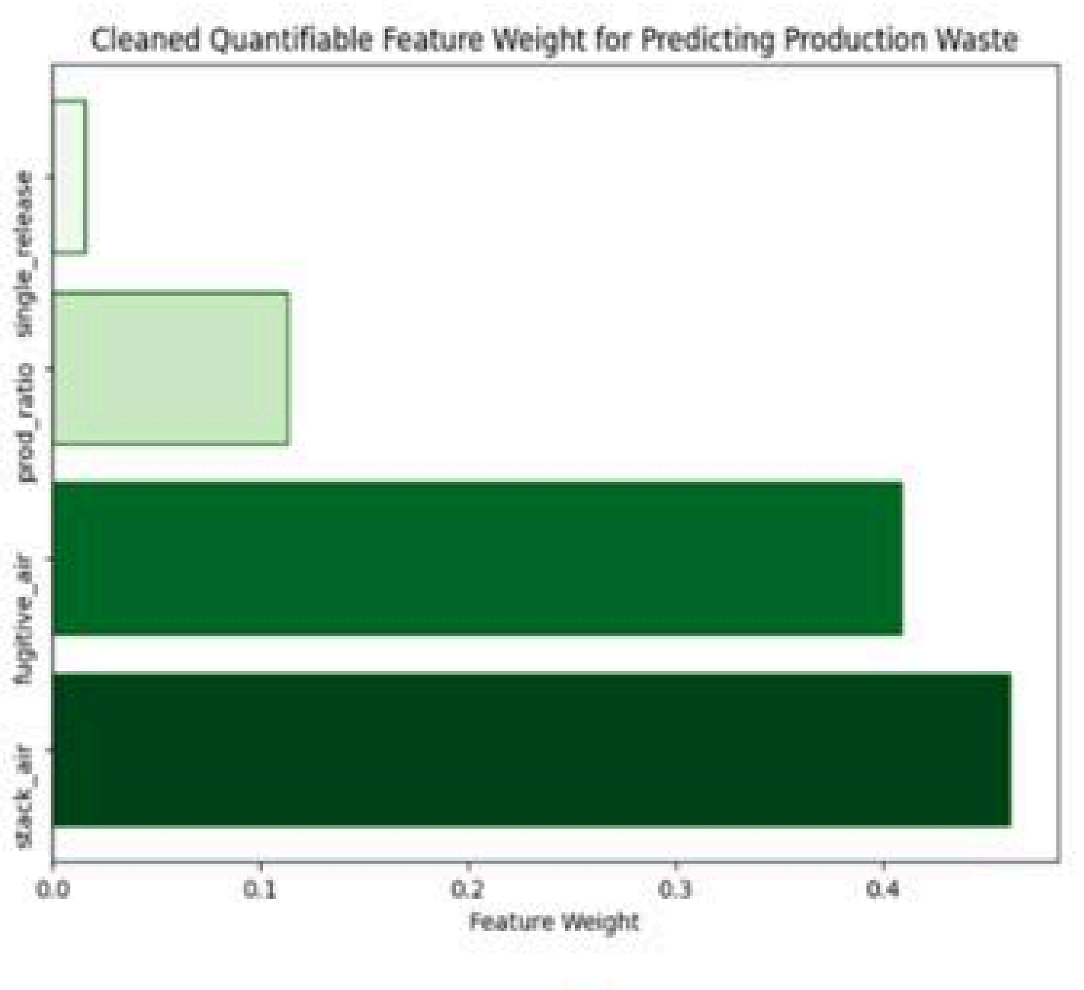
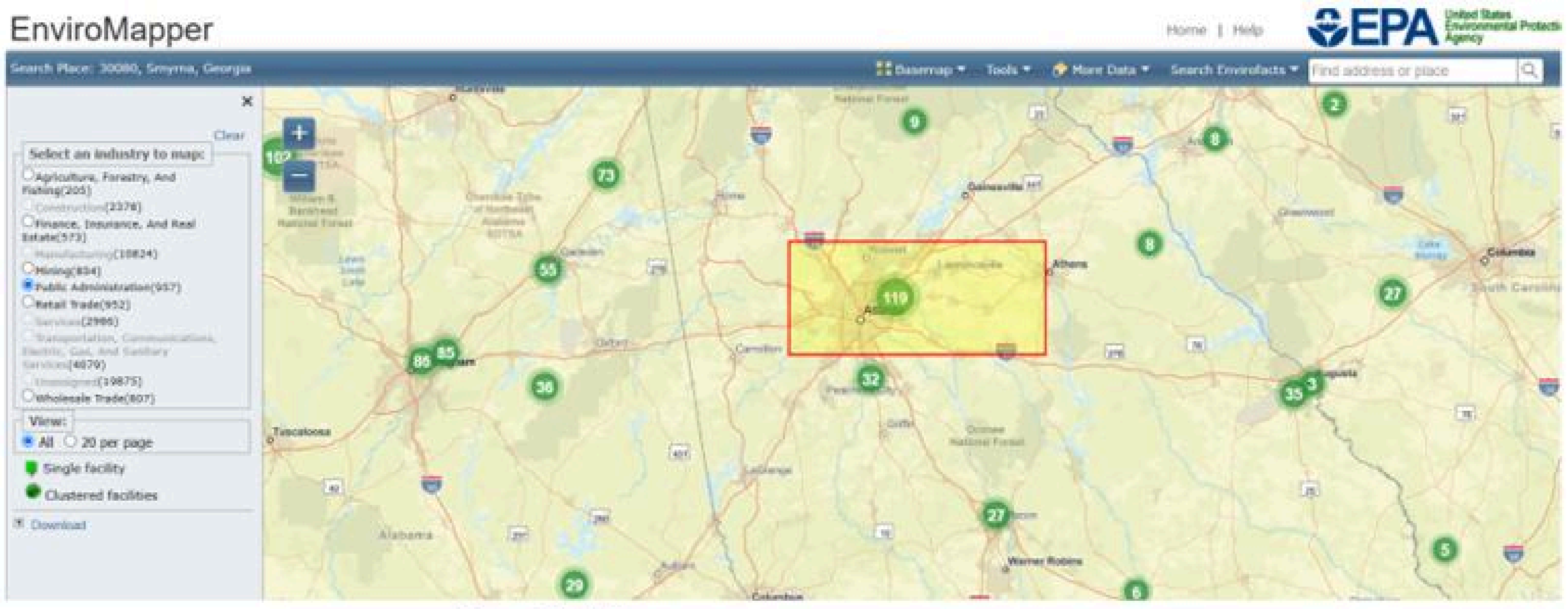


Fig2: EDA Result 2

Comparison : Enviromapper Vs Tox-e-mapper

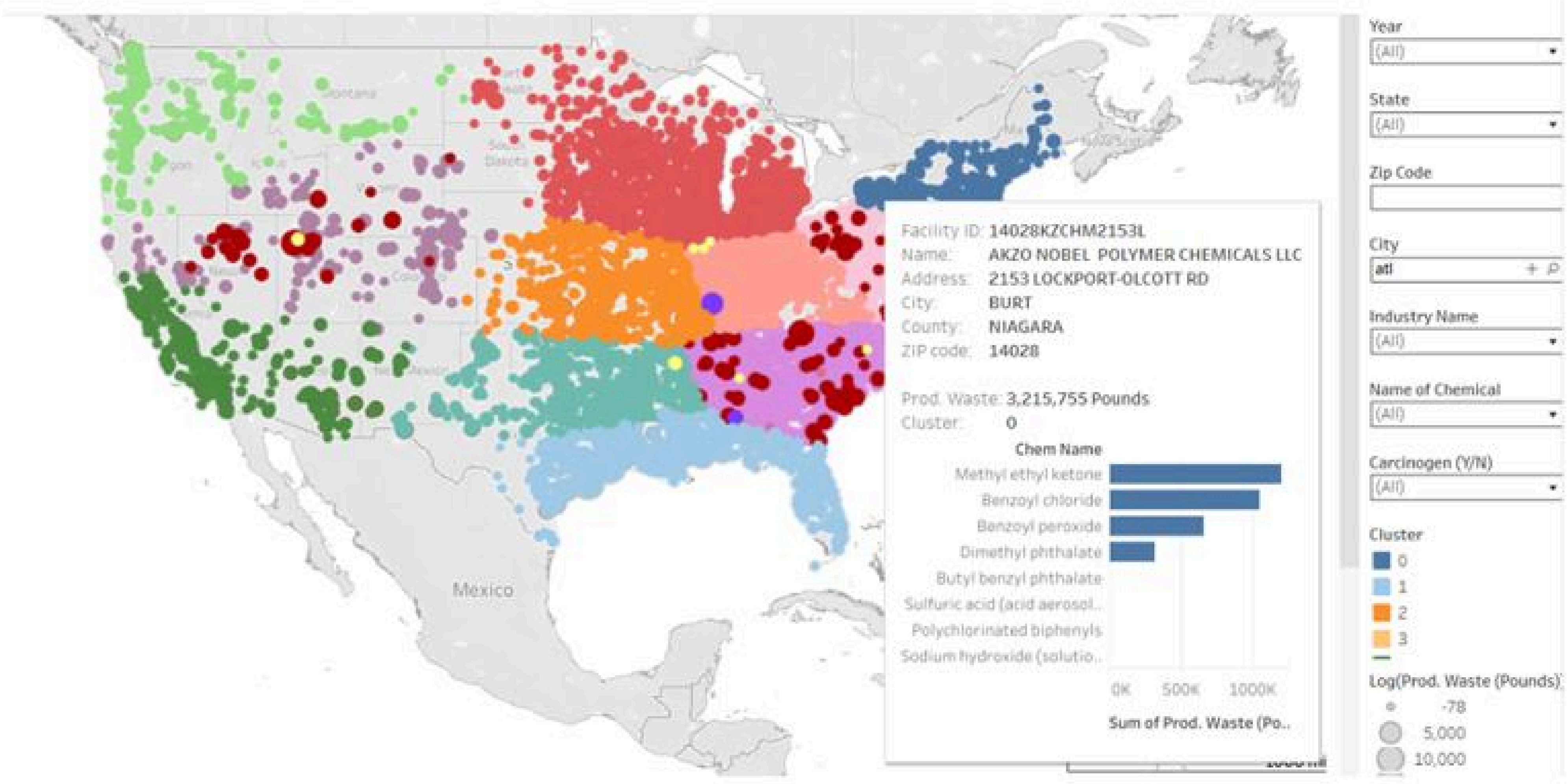

EnviroMapper

Search Place: 30080, Smyrna, Georgia

Select an industry to map:
○ Agriculture, Forestry, And Fishing(205)
○ Construction(2378)
○ Finance, Insurance, And Real Estate(573)
○ Manufacturing(18824)
○ Mining(834)
● Public Administration(957)
○ Retail Trade(952)
○ Services(2986)
○ Transportation, Communications, Electric, Gas, And Sanitary Services(4079)
○ Unassigned(19875)
○ Wholesale Trade(807)

View:
● All ○ 20 per page

Tox-E-Mapper


Facility ID: 14028KZCHM2153L
Name: AKZO NOBEL POLYMER CHEMICALS LLC
Address: 2153 LOCKPORT-OLCOTT RD
City: BURT
County: NIAGARA
ZIP code: 14028

Prod. Waste: 3,215,755 Pounds
Cluster: 0

Year: (All)
State: (All)
Zip Code:
City: atl
Industry Name: (All)
Name of Chemical: (All)
Carcinogen (Y/N): (All)

Evaluation

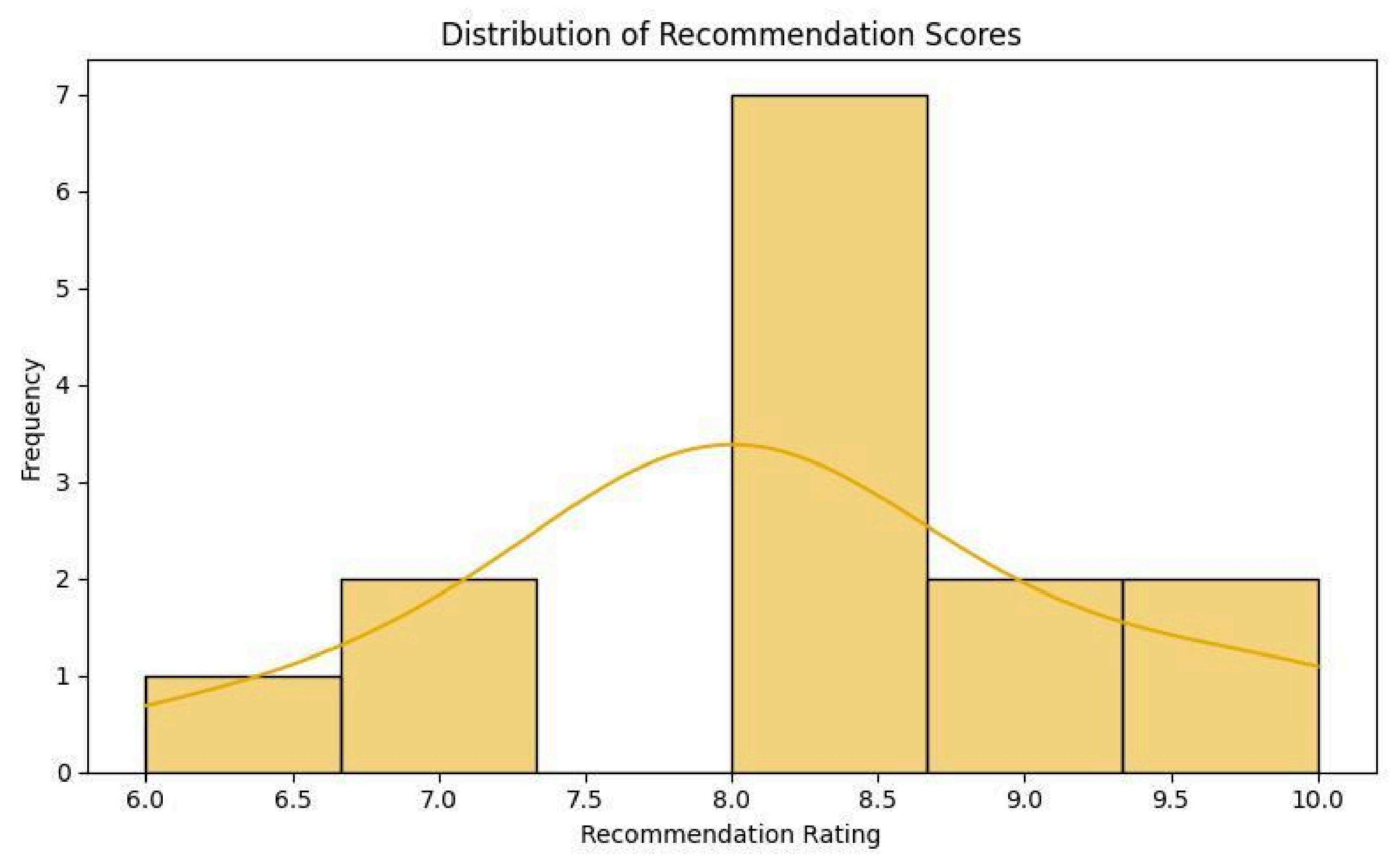
- Usability:
  User surveys compare Tox-E-Mapper and EnviroMapper in terms of task efficiency, satisfaction, and ease of use.
- Accuracy:
  - Optimal cluster count identified using the elbow method (inertia plot).
  - Prediction accuracy assessed using MAE and RMSE metrics.

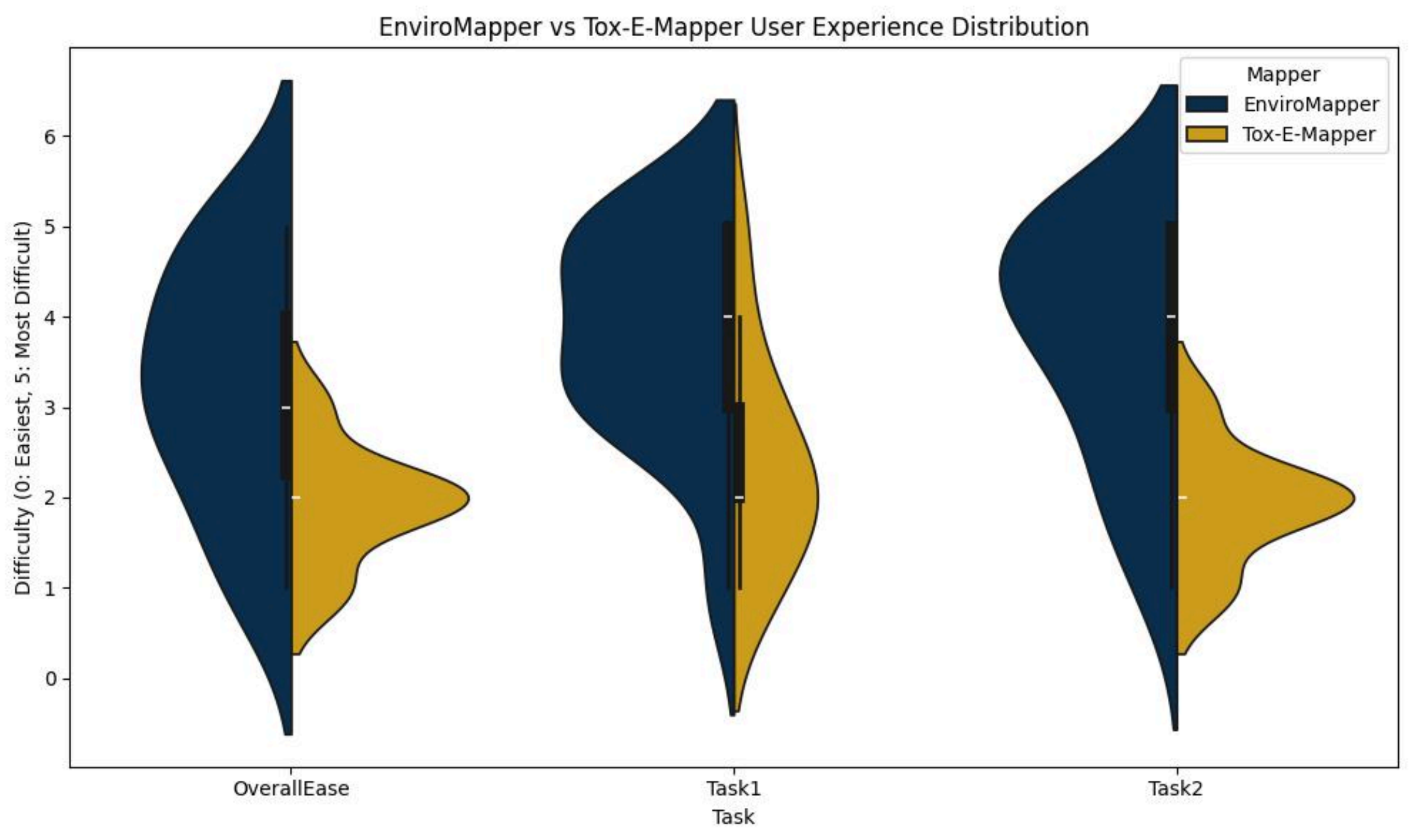
Results

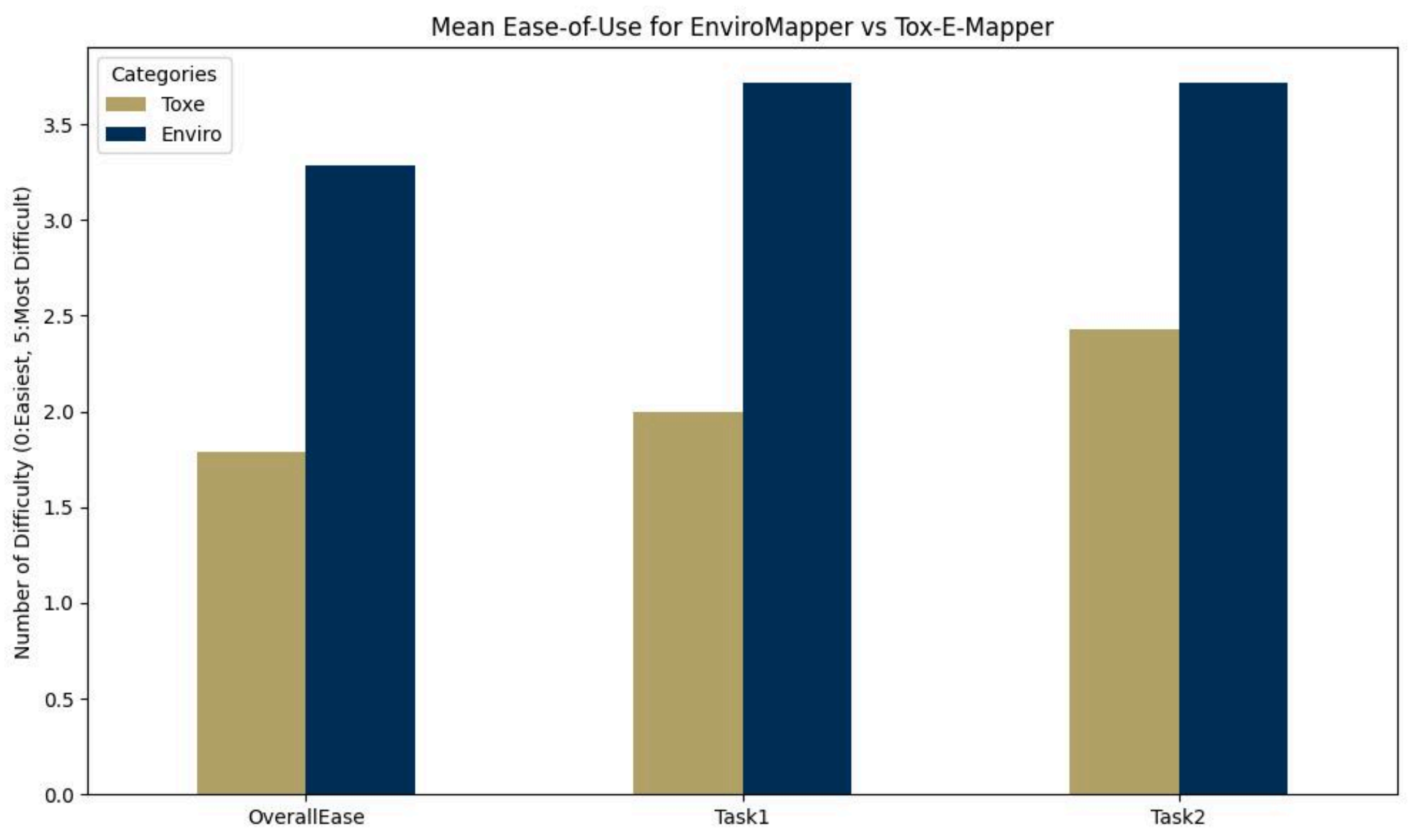
1. Sample Metrics from an experiment training model for forecasting with Histogram Gradient Boosting Regressor

Scoring Function	*MAPE*	*MAE*	*MSE*	*RMSE*
neg_mean_absolute_percentage_error	0.26	118481.36	54316723669698.88	7369988.04
neg_mean_squared_error	1.37e+20	143174.61	53606213011906.93	7321626.39
neg_root_mean_squared_error	0.26	118481.36	54316723669698.88	7369988.04
neg_mean_absolute_error	0.26	118481.36	54316723669698.88	7369988.04

2. User Survey results







Section 3

Limitations

1. Incomplete Dataset: Only includes government-reported industries.
2. Prediction Issues: Model often predicted zeros, limiting insights.
3. Clustering Time: Clustering process was computationally intensive and time-consuming.

FutureScope

1. Explore alternatives to forecasting, such as classification or anomaly detection, if data sparsity continues
2. Purchase a domain to publish the platform for public access.