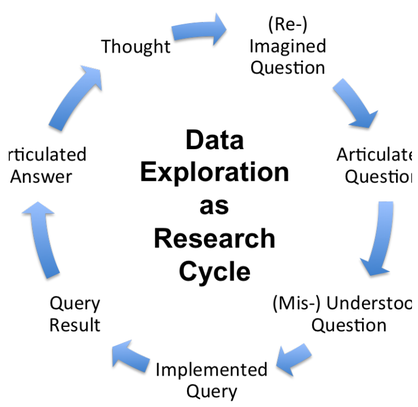
Project Mission

The project **"National Predictive Analysis of Toxic Releases"** aims to develop predictive analytics focused on forecasting on forecasting the release of toxic chemicals across the United States. Using historical data from the U.S. Environmental Protection Agency's (EPA) Toxics Release Inventory (TRI) from 2013 to the present, the system will analyze patterns of chemical releases across industries, geographic regions, and specific pollutants. The goal is to provide insights into future trends in toxic emissions, helping policymakers, environmental agencies, and industries make informed decisions to mitigate the environmental impact of hazardous chemicals. This project also aims to promote sustainability by identifying opportunities for reducing toxic releases and complying with regulatory requirements.

Introduction



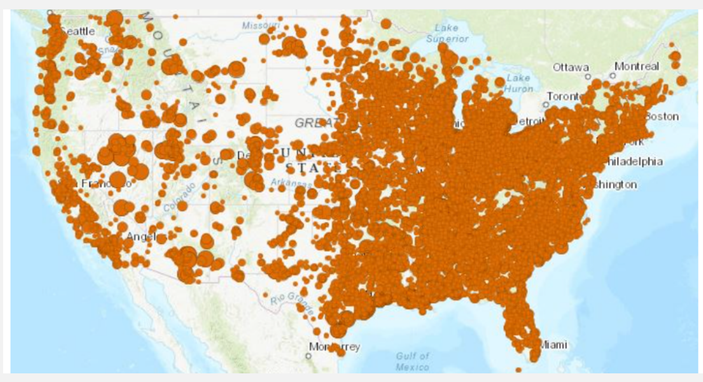
Data Exploration

Models Implemented

Conclusion

Project Link

Introduction



Industrial facilities in United States

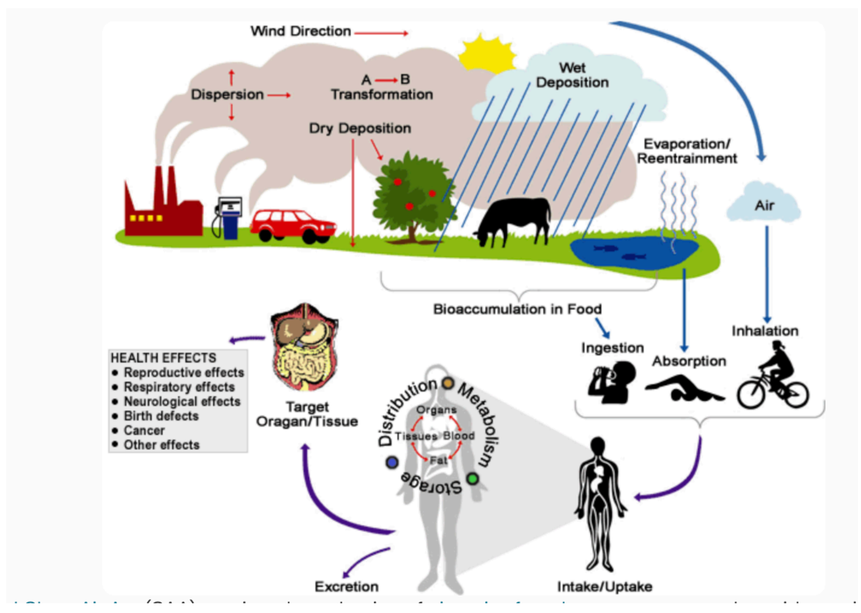
Toxic releases from industries pose very dangerous environmental and public health risks. The U.S. EPA's Toxics Release Inventory (TRI) has tracked these releases since 1987, but there is a need for predictive tools to forecast future trends and guide decision-making. The **"Toxics Releases National Predictive Analysis"** project aims to develop a system that uses historical TRI data to predict future toxic release patterns across industries and regions.

Industrial activities contribute significantly to economic growth but often come with environmental and public health consequences, particularly from the release of toxic chemicals into air, water, and soil. These toxic releases pose severe risks to ecosystems and human health, including long-term exposure effects such as respiratory diseases, cancer, and contamination of natural resources. Recognizing these challenges, the U.S. Environmental Protection Agency (EPA) established the Toxics Release Inventory (TRI) program in 1987. The TRI provides a comprehensive database tracking the management and release of toxic chemicals by industries, enabling transparency and informed decision-making.

There is a growing need to enhance its value by developing predictive tools capable of forecasting future toxic release patterns. Such tools can empower stakeholders—including government agencies, environmental organizations, and industries—with actionable insights to anticipate risks, improve compliance, and adopt proactive environmental management strategies.

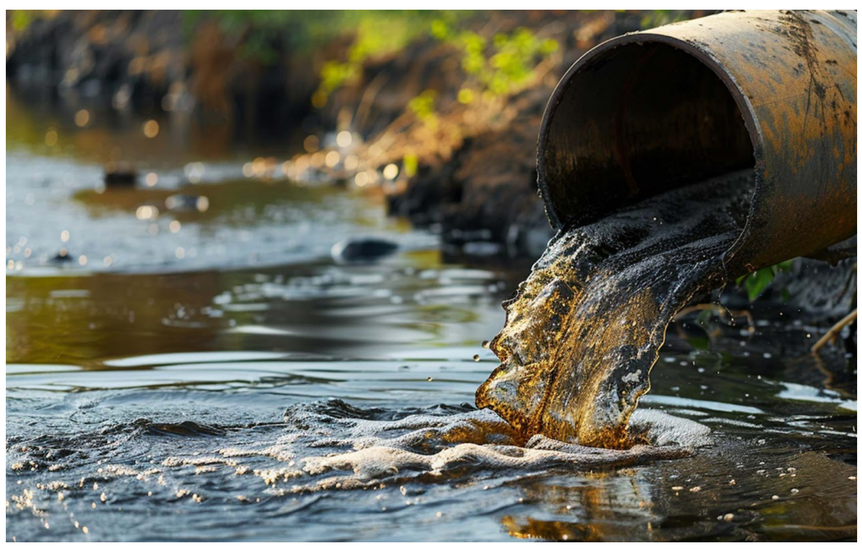
Why is it important?

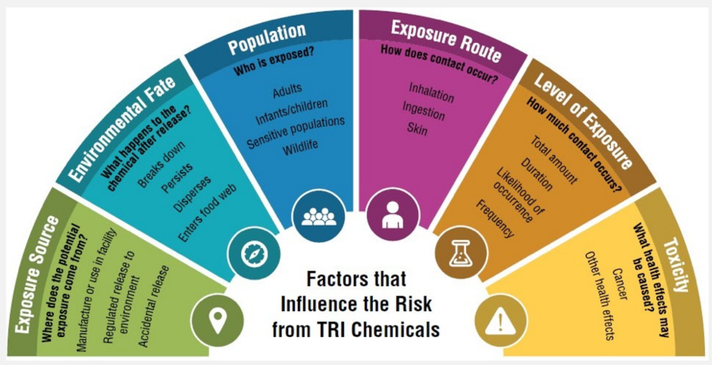
The **"Toxics Releases National Predictive Analysis" project** is crucial for several reasons. It protects public health by predicting future toxic chemical releases and enabling timely interventions to mitigate risks. By addressing environmental hazards such as air and water pollution, the project helps safeguard ecosystems and biodiversity from long-term damage. Predictive insights empower governments and regulatory bodies to create more effective, forward-looking environmental policies, ensuring proactive risk management rather than reactive responses. Communities benefit significantly from reduced exposure to hazardous chemicals, leading to healthier living conditions and improved quality of life. For industries, the project provides a roadmap to align operations with environmental regulations, avoid penalties, and enhance corporate sustainability initiatives. It also encourages innovation in cleaner technologies and sustainable practices to minimize toxic releases. By identifying high-risk regions and industries, the system supports targeted interventions and efficient resource allocation. Furthermore, the project facilitates collaboration among stakeholders, fostering a shared commitment to environmental protection and public health. Ultimately, this initiative helps balance industrial development with ecological preservation, ensuring a healthier and more sustainable future for all.



Who does it affect?

The release of toxic compounds poses significant risks to various populations, particularly disadvantaged communities residing near industrial facilities and chemical plants. These toxic substances can contaminate vital natural resources, including water, air, and soil, resulting in environmental degradation and serious health issues for living beings. Prolonged exposure to such hazardous materials can lead to severe diseases, such as cancer, respiratory illnesses, and skin conditions, especially in areas located in close proximity to industrial operations.Additionally, it can disrupt weather patterns, contribute to water contamination, and exacerbate air pollution. Wildlife is also adversely impacted, as contaminated water sources can lead to a decline in biodiversity and alterations to ecosystems.Farmers and agricultural workers exposed to these toxic substances face heightened risks, and those who consume contaminated food supplies are similarly affected. Furthermore, children may experience malnutrition and developmental issues as a result of exposure to these hazardous conditions.





Current Efforts and Remaining Challenges:

Accurate predictions regarding the quantity and type of chemicals released are essential for enforcing stricter limits and industrial standards for emissions. Furthermore, only industrial emissions are measured, leaving other chemical sources, such as pharmaceuticals,. Studies examining the impact of chemical dispersion on climate change are still incomplete, and while these factors may not have an immediate impact on the environment, they could pose serious threats in the future.Public campaigns focused on environmental safety have also raised awareness among the public about the dangers of toxic releases. The predictive analysis of toxic releases plays a crucial role in addressing these issues by utilizing data-driven approaches to anticipate chemical emissions and their potential environmental impacts. By analyzing historical data and identifying patterns in chemical usage, industries can better understand their emissions profiles and implement targeted strategies to reduce hazardous releases. Furthermore, predictive models can help forecast potential future emissions based on current operations and planned changes, allowing for proactive measures to mitigate risks.As the field of predictive analysis grows, these existing gaps provide opportunities to forecast toxic substance releases and mitigate their impact before they threaten environmental and human health.This process will lead us towards a safer and more sustainable future.

Key questions to be addressed:

1. Which industries contribute the most to toxic chemical releases across the country?
2. What are the most common toxic chemicals released into air, water, and soil, and which media (air, water, or soil) is most affected?
3. How have toxic release levels changed over the past decade in industries like manufacturing, mining, and chemical processing?
4. What is the state with the highest total amount of chemical releases, and how do these releases compare to other states?
5. Which state has the highest number of chemical release facilities?
6. Based on current trends and historical data, what is the probability of a chemical release happening in 2024?
7. Is there an increase in the number of new chemicals being released each year?
8. What is the most common method used to release chemicals from industrial facilities? Are chemicals more often released directly into the environment, or are they treated before being disposed of or transferred?
9. Are the chemicals being released from industrial facilities today more carcinogenic or harmful to human health than those released in the past? What is the trend in the toxicity and carcinogenicity of these chemicals over the years?
10. How much of the released chemicals are transferred outside of the facility? Has this amount increased or decreased over the years?

Project Link

Data Exploration

DATA COLLECTION

The dataset was collected from the U.S. Environmental Protection Agency (EPA), which provides data from the Toxics Release Inventory (TRI). The TRI is a resource for learning about toxic chemical releases and pollution prevention activities reported by industrial and federal facilities. This dataset spans several years and includes comprehensive records on the release of toxic chemicals into the environment, providing insights into pollution patterns across various industries and regions. The data collected from TRI will be used to analyze trends, evaluate environmental risks, and develop predictive models for future toxic waste releases.

DATA EXPLORATION

- Analyzed and identified the attribute types.For Instance
  1. Categorical: Chemical Name, State, City
  2. Numerical: Release Quantity, Latitude, Longitude
- Used Q-Q plot to check normality in the distribution of total release quantities.

DATA CLEANING

- **Removing Duplicates**: Eliminated duplicate entries to ensure each record is unique and accurate.
- **Handling Missing Data**:
  - *Categorical Columns*: Missing values replaced with "Unknown."
  - *Numerical Columns*: Missing values filled with the column mean.
  - *Carcinogen Column*: Missing values classified as carcinogenic or non-carcinogenic by reviewing similar chemicals.
- **Renaming Columns**: Simplified column names for better readability (e.g., "8.1A - ON-SITE CONTAINED" to "ON-SITE CONTAINED").
- **Standardizing Units**: Converted all measurements to pounds to maintain consistency in numerical analysis.
- **Quality Checks**: Ensured consistency in categorical data (e.g., state and chemical names) and corrected anomalies or typos.

DATA REDUCTION

- PCA was performed to identify key factors affecting release quantities, reducing the dimensionality of the data.
- We removed columns that were unnecessary for our analysis, reducing the dataset from 122 to around 19 columns.
- Additionally ,We combined several columns that contained redundant or related information into a single column for clarity.
- For example, we merged columns representing different aspects of chemical release data into a respective column.

DATA TRANSFORMATION

- **Label Encoding**: Categorical columns, such as 'CARCINOGEN,' were label-encoded to make the dataset suitable for analysis.
- **Time Series Stationarity**:
  - Differencing was applied to the 'TOTAL RELEASES' column to remove trends and seasonality, ensuring stationarity for time series modeling.
  - The stationarity of the processed data was validated using the Augmented Dickey-Fuller (ADF) test, confirming its readiness for further analysis.

STATISTICAL SUMMARY

- **Min Values**: All columns have a minimum of 0, indicating some records report no activity in these categories.
- **25th Percentile**: Many records report 0 values, showing no activity in these categories for a significant portion of the dataset.
- **Max Values**: Extremely high maximum values suggest a few facilities report exceptionally large amounts.
- **Standard Deviation**: High variability in the data due to some facilities reporting disproportionately large values.
- **Median**: Lower than the mean, indicating a skewed distribution caused by a few very high values.

BEFORE AND AFTER DATA PREPROCESSING

```
print("Total number of rows and columns before exploration: ",data.shape)
data.head()
```

Total number of rows and columns before exploration: (806162, 122)

	1. YEAR	2. TRIFD	3. FRS ID	4. FACILITY NAME	5. STREET ADDRESS	6. CITY	7. COUNTY	8. ST	9. ZIP	10. BIA	...	113. 8.2 – ENERGY RECOVER ON	114. 8.3 – ENERGY RECOVER OF	115. 8.4 – RECYCLING ON SITE	116. 8.5 – RECYCLING OFF SIT	117. 8.6 – TREATMENT ON SITE
0	2018	28655MLDDP213RE	1.100004e+11	MOLDED FIBER GLASS NORTH CAROLINA	213 REEP DR	MORGANTON	BURKE	NC	28655	NaN	...	0.0	0.0	0.0	0.0	0.0
1	2018	91744MNTXN13300	1.100005e+11	MAINTEX INC	13300 E NELSON AVE	CITY OF INDUSTRY	LOS ANGELES	CA	91746	NaN	...	0.0	0.0	0.0	0.0	0.0
2	2018	15690SSVND130LI	1.100144e+11	ATI FLAT ROLLED PRODUCTS HOLDINGS LLC	130 LINCOLN AVE	VANDERGRIFT	WESTMORELAND	PA	15690	NaN	...	0.0	0.0	8535863.0	0.0	2735204.0
3	2018	5320WCHRTR37WMI	1.100420e+11	CHARTER WIRE LLC	3700 W. MILWAUKEE ROAD	MILWAUKEE	MILWAUKEE	WI	53208	NaN	...	0.0	0.0	0.0	16.0	0.0
4	2018	0808WPNDRL51SHA	1.100702e+11	PANDROL	501 SHARPTOWN RD	SWEDESBORO	GLOUCESTER	NJ	8085	NaN	...	0.0	0.0	0.0	3571.0	0.0

5 rows × 122 columns

```
df1=pd.read_csv("/content/combined_tri_data_2014_2023.csv",low_memory=False)

null_counts = df1.isnull().sum()
null_counts = null_counts[null_counts > 0]  # Filter columns with at least one null value

# Display the counts
print("Count of columns with null values before exploration : ",null_counts.count())
```

Count of columns with null values before exploration : 25

RAW DATA BEFORE PREPROCESSING

(674021, 19)

	YEAR	FACILITY NAME	CITY	ST	INDUSTRY SECTOR	CHEMICAL	CARCINOGEN	WATER	OTHER DISPOSAL	ON-SITE RELEASE TOTAL	OFF-SITE RELEASE TOTAL	TOTAL TRANSFER
0	2017	DRS NIS LLC	DALLAS	TX	Computers and Electronic Products	Methanol	NO	0.0	0.0	11436.00	0.0	22113.00
1	2017	AMERICAN ELECTRIC POWER KAMMER/MITCHELL PLANT	MOUNDSVILLE	WV	Electric Utilities	Selenium compounds	NO	170.0	5.0	17200.00	2.0	2.00
2	2017	GEORGIA-PACIFIC LLC	TAYLORSVILLE	MS	Chemicals	Formaldehyde	YES	0.0	0.0	23640.00	5.0	948.00
3	2017	HITACHI ASTEMO AMERICAS INC. - BEREA KY	BEREA	KY	Transportation Equipment	Certain glycol ethers	NO	0.0	0.0	24674.45	0.0	7192.35
5	2017	SIGECO A B BROWN GENERATING STATION	MOUNT VERNON	IN	Electric Utilities	Polycyclic aromatic compounds	YES	0.0	0.0	1.00	0.0	0.00

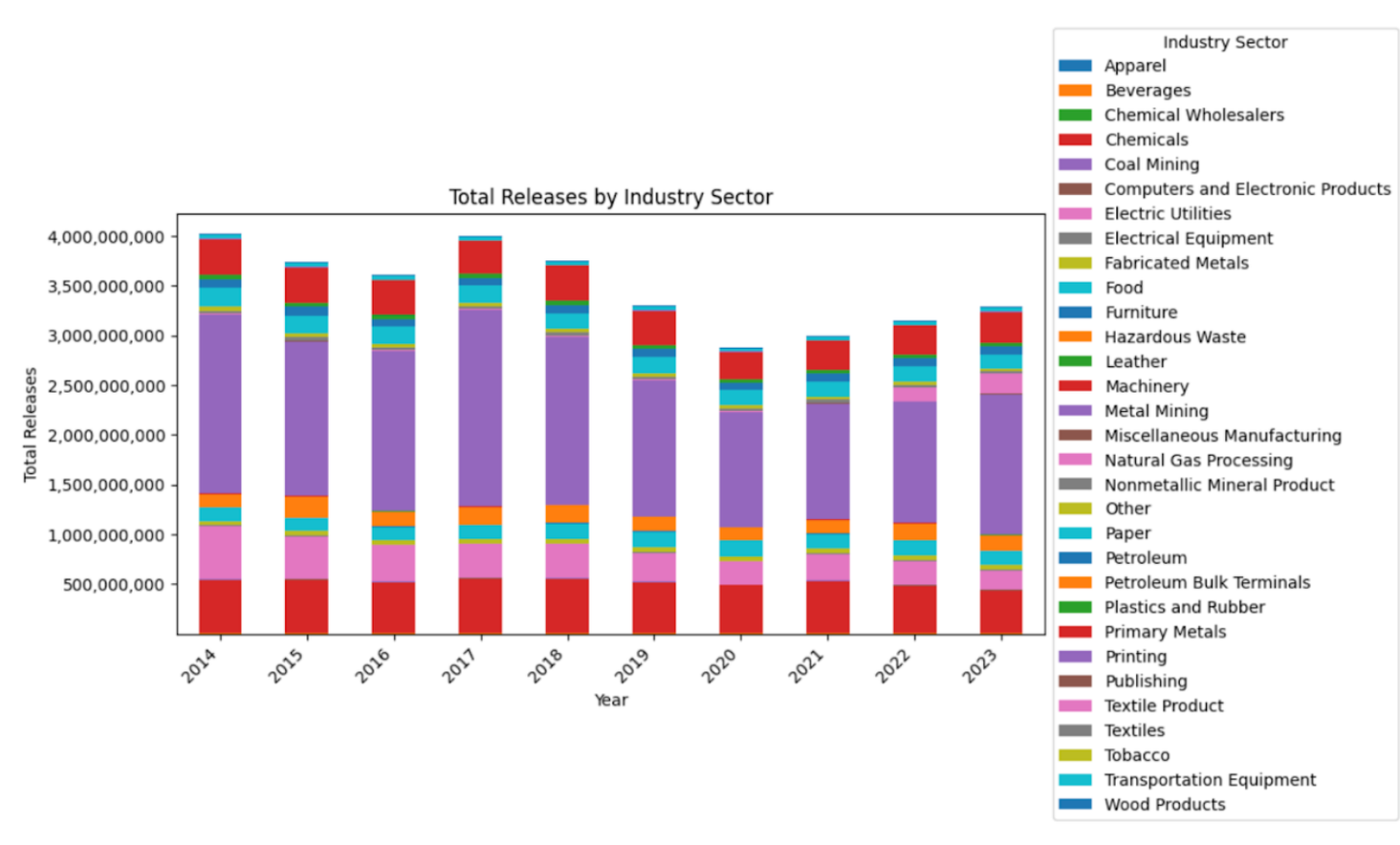
```
null_counts = df.isnull().sum()
null_counts = null_counts[null_counts > 0]  # Filter columns with at least one null value

# Display the counts
print("Count of columns with null values after exploration : ",null_counts.count())
```

Count of columns with null values after exploration : 0

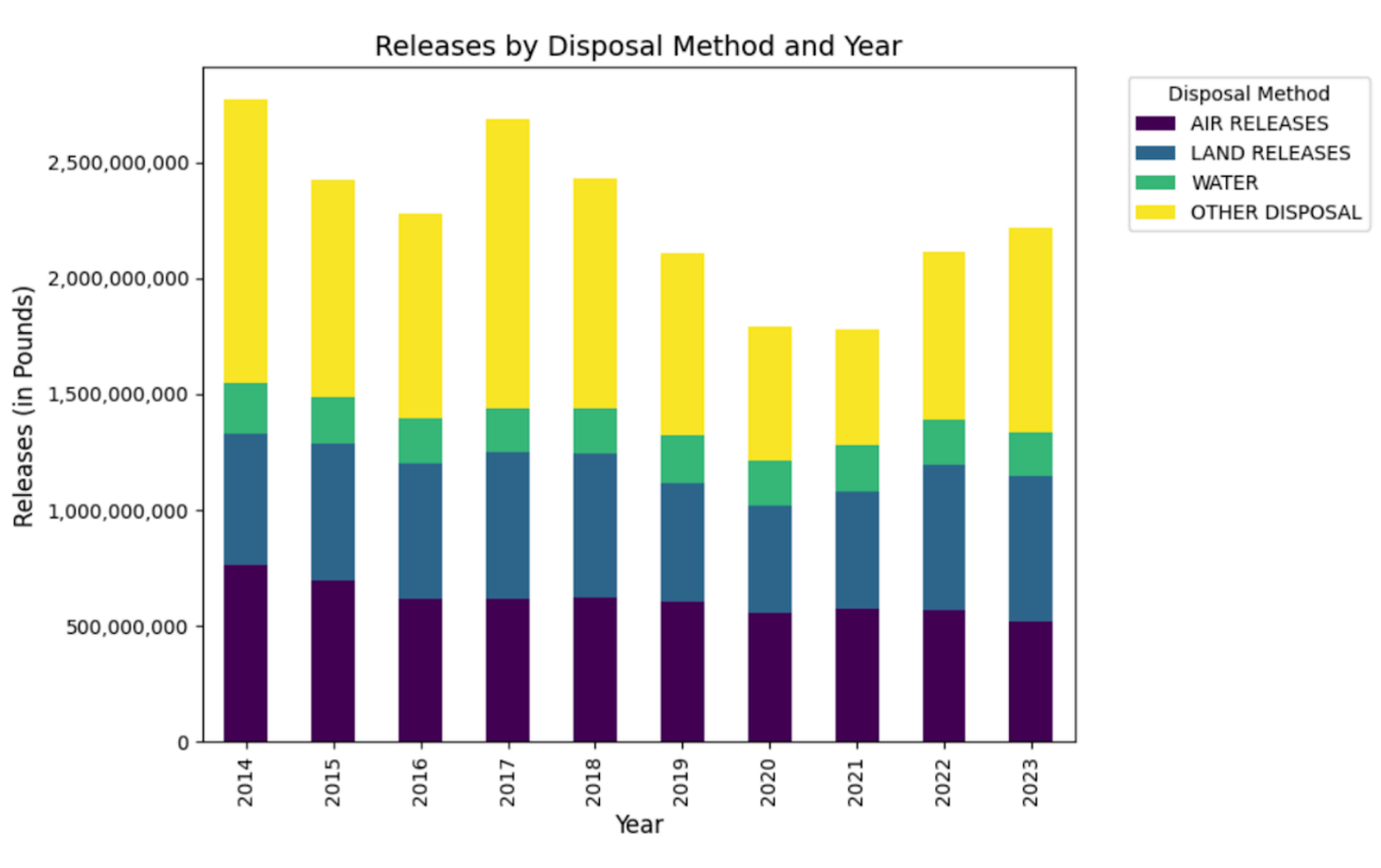
DATA AFTER PREPROCESSING

DATA EXPLORATION



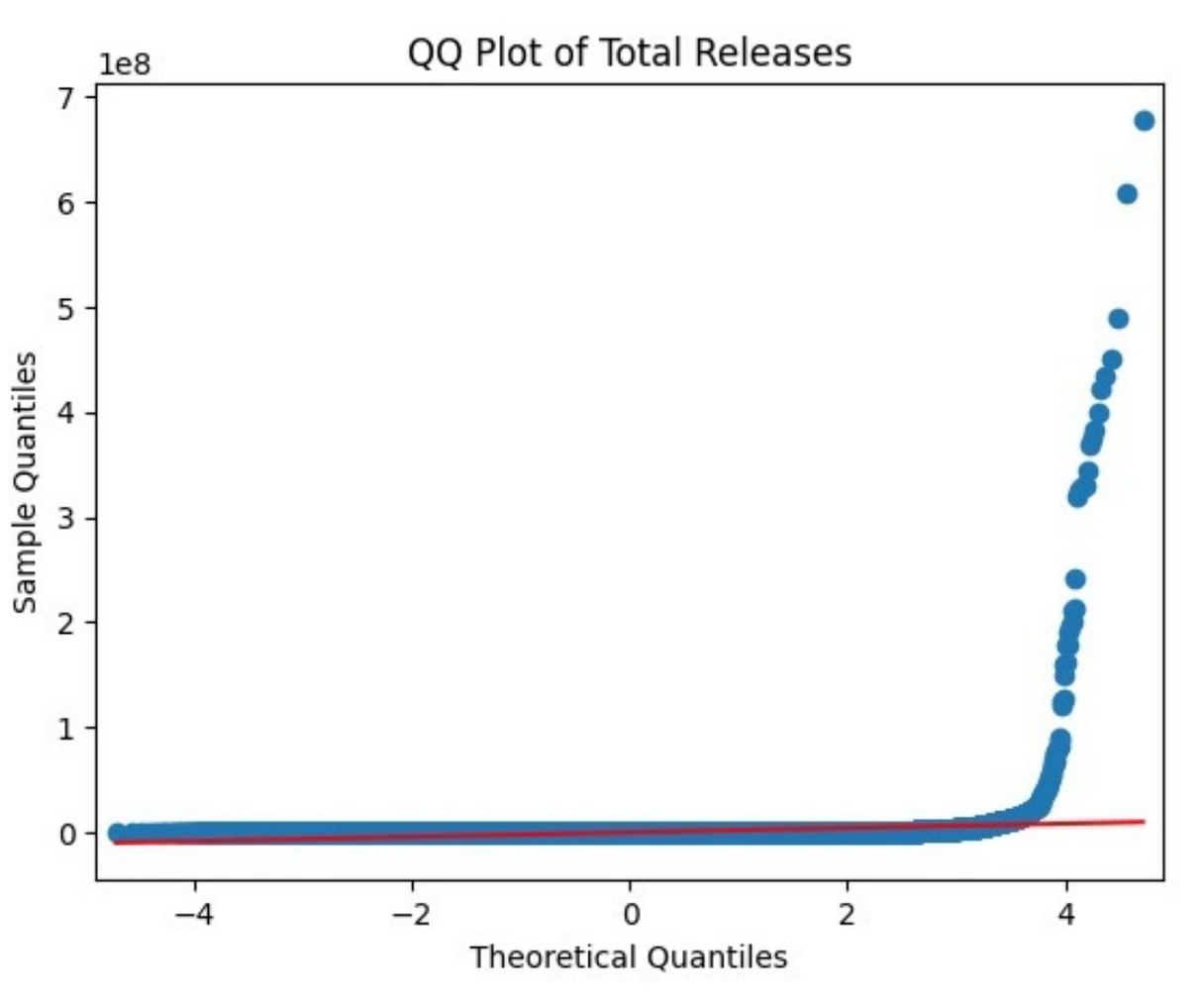
Total Releases by Industry

The total releases vary from year to year but are consistently high, ranging from 2.5 billion to 4 billion pounds.Different industry sectors contribute varying proportions to the total releases, with certain sectors like **Chemicals**, **Coal Mining**, and **Petroleum** showing significant shares across multiple years.The overall trend appears stable, with no dramatic increase or decrease across these years, although minor fluctuations can be observed.



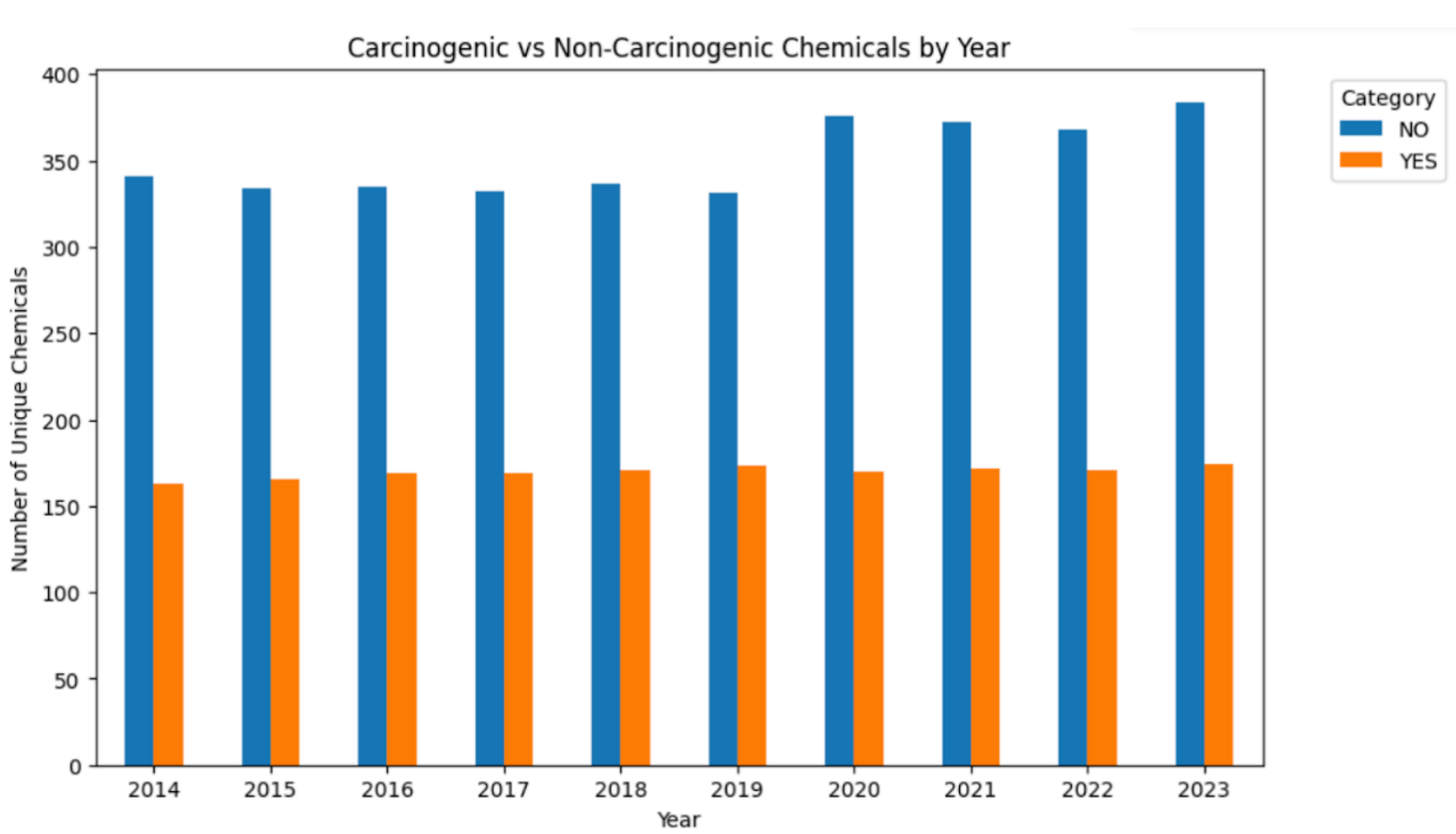
Total Releases by Disposal Method

The disposal method contributing the most to releases is Other Disposal, which consistently takes up the largest portion of the stacked bars.Air Releases have a steady but somewhat lower contribution in comparison to Other Disposal and Land Releases, though it remains substantial.Land Releases and Water Disposal also account for a significant portion of the total releases, with Water Disposal slightly increasing in recent years.Trends: From 2014 to 2023, the relative proportions of each disposal method appear stable, with Other Disposal continuing to dominate. There is no drastic shift in disposal methods over the years.



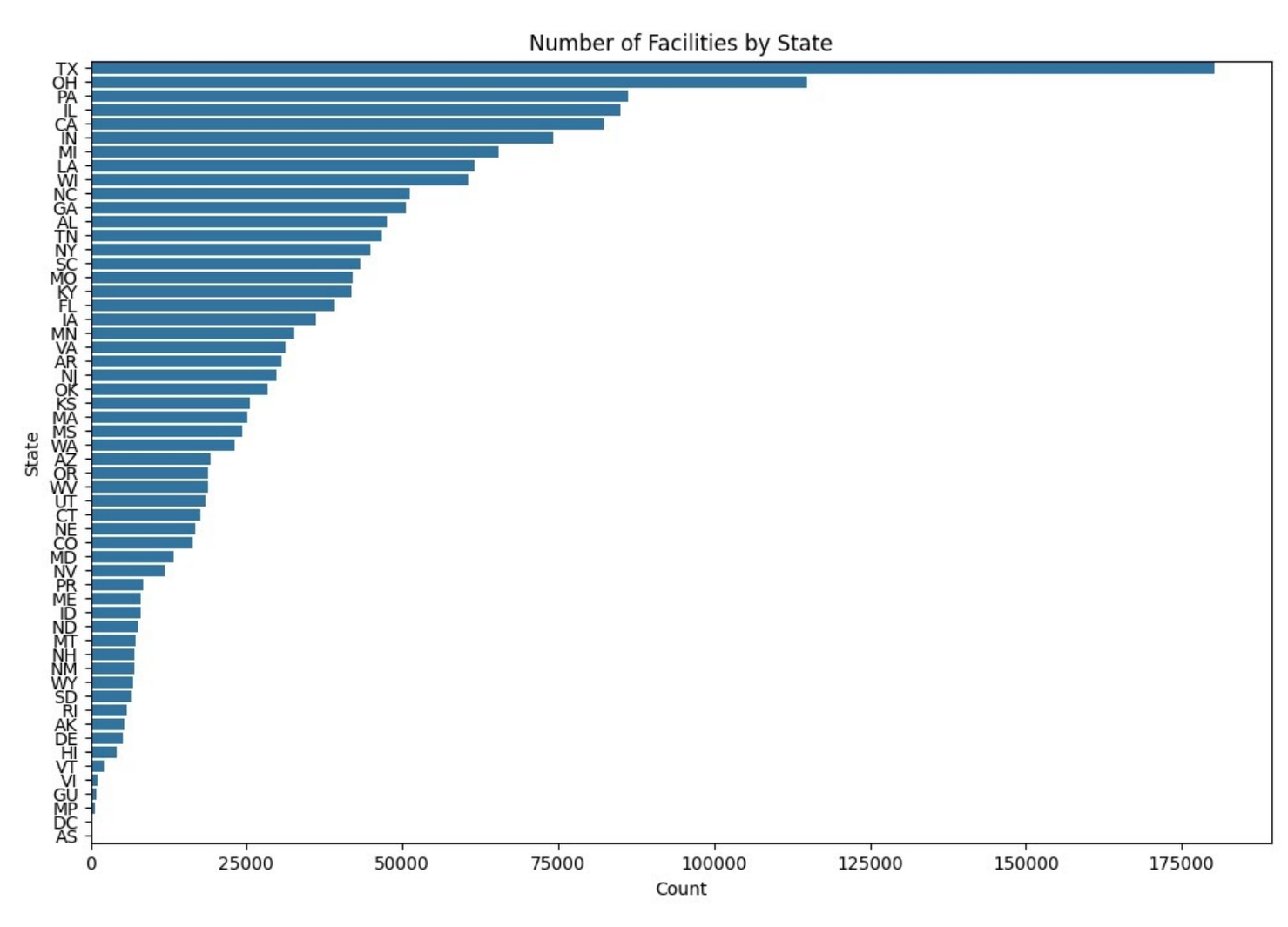
Q-Q Plot for Total Release vs Normal Distribution

The QQ plot reveals that total chemical releases deviate from a normal distribution.Most points deviate from the red line, especially on the right side, indicating a few very high values (outliers) are affecting the data.



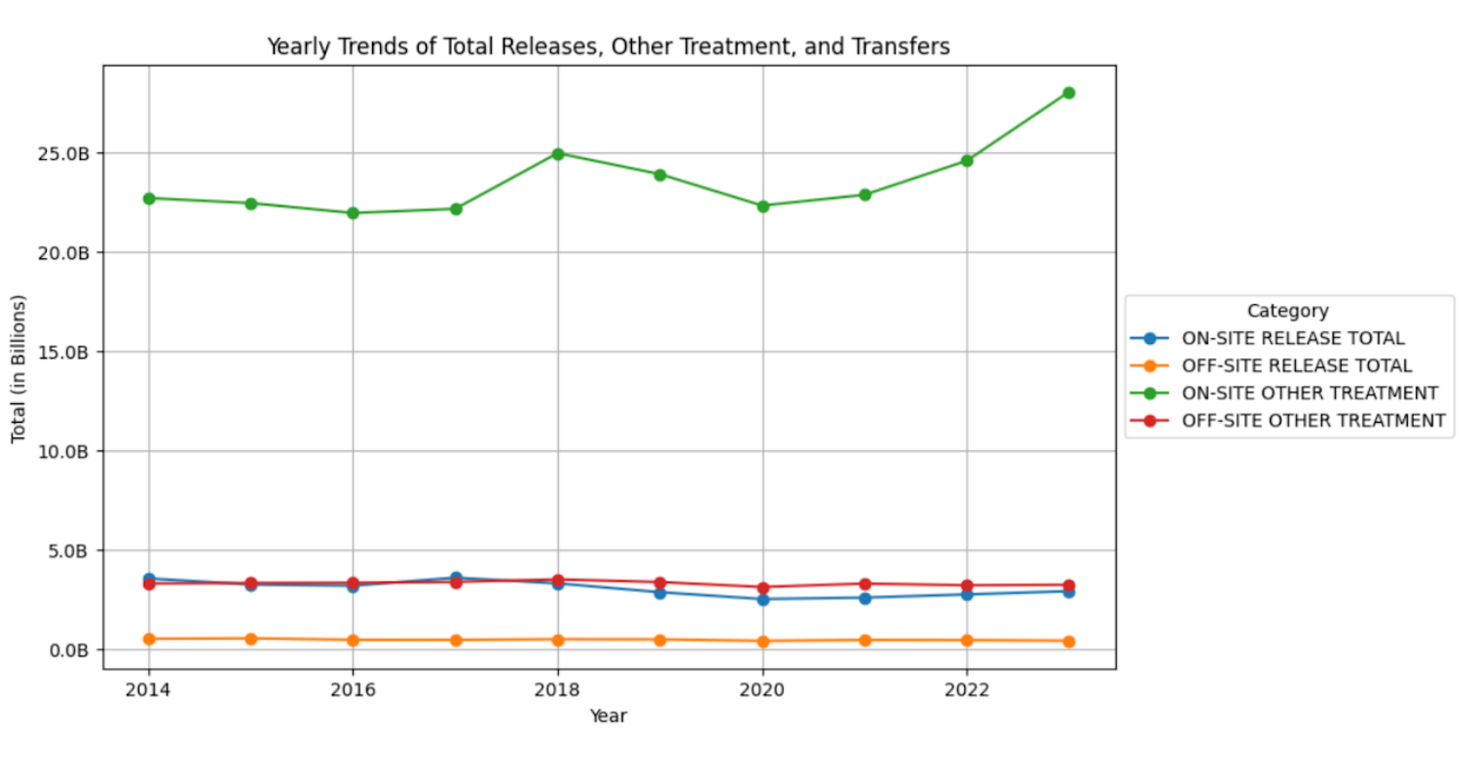
Release of Carcinogenic Chemicals

This bar chart compares the number of carcinogenic chemicals ("YES") and non-carcinogenic chemicals ("NO") across different years from 2014 to 2023.The count of non-carcinogenic chemicals (blue) is consistently higher than that of carcinogenic chemicals (orange) across all years.There appears to be little variation in the number of chemicals in each category from year to year.The count remains stable, ranging close to 350 chemicals each year. There is no apparent trend of increase or decrease over the years.On average, non-carcinogenic chemicals are more than twice as prevalent as carcinogenic ones.



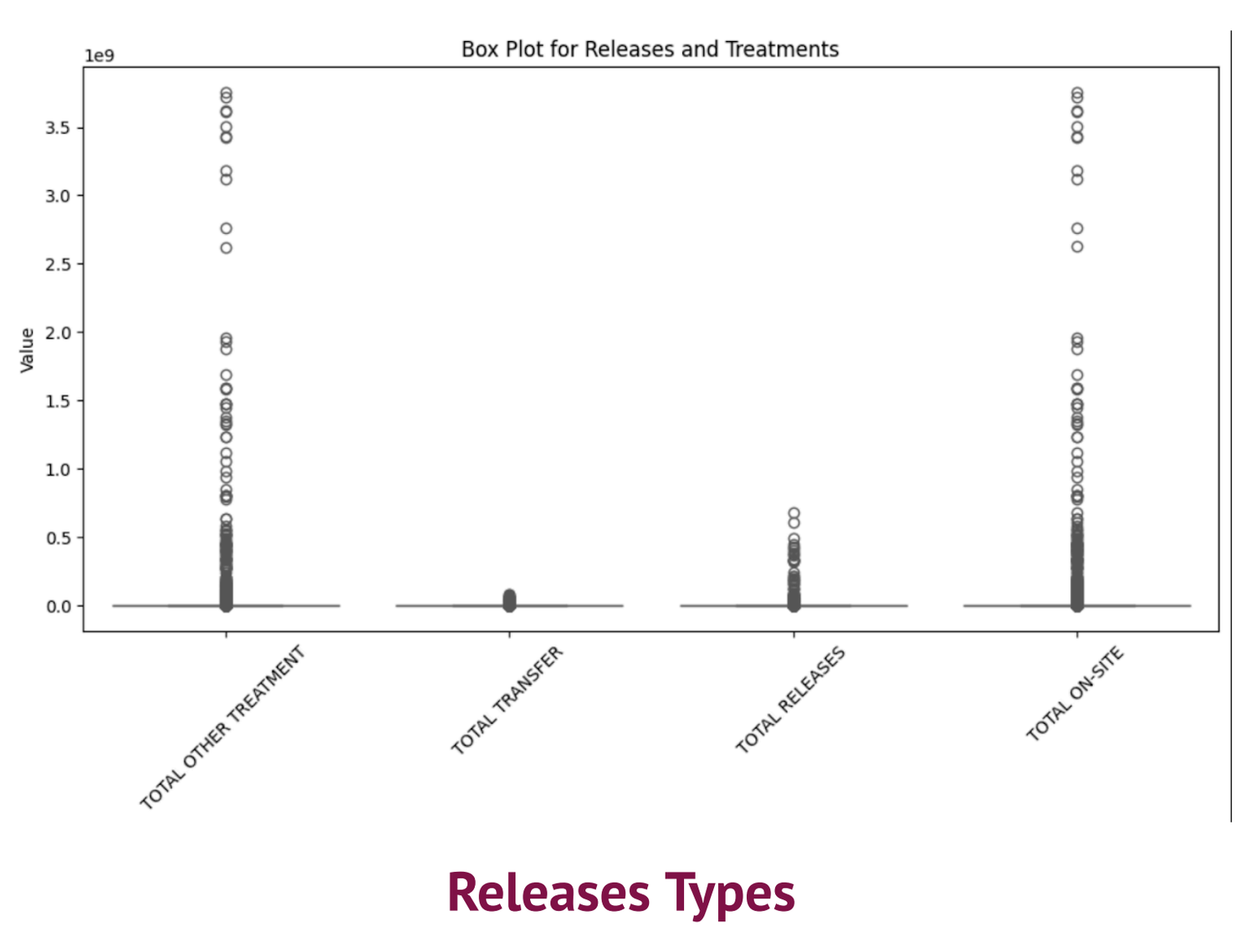
Barplot of releases by State

This chart displays the total number of facilities across different states. Texas has the highest number of facilities, while the average across other states is around 10,000 facilities per state.



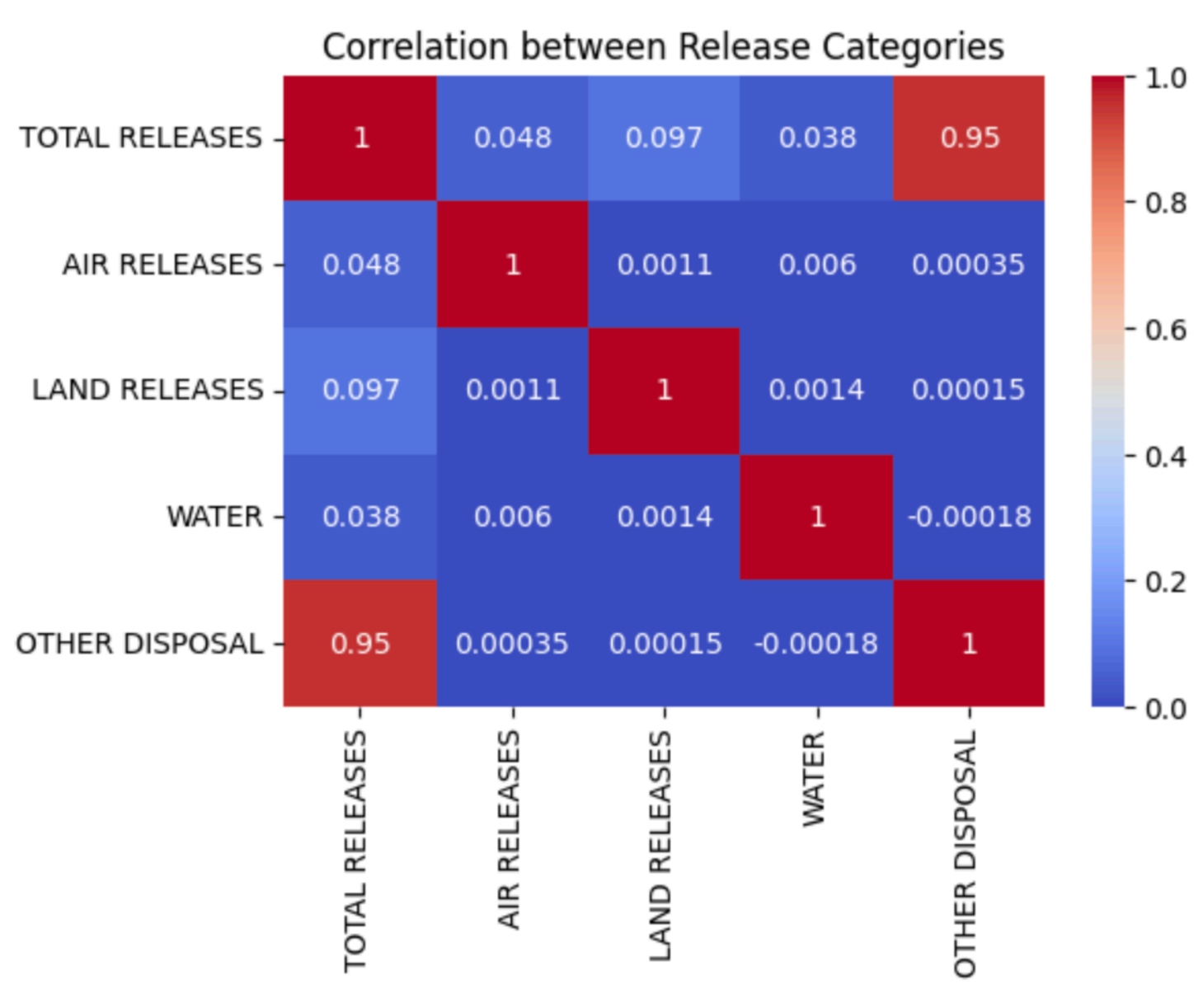
Yearly Trends

The on-site other treatment (green line) is significantly higher compared to other categories and has shown a steady increase over the years, particularly from 2020 to 2022.The on-site release total (blue line) and off-site other treatment (red line) remain relatively stable and much lower than the on-site other treatment.Off-site release total (orange line) is the lowest among all categories and does not exhibit any significant change over time.



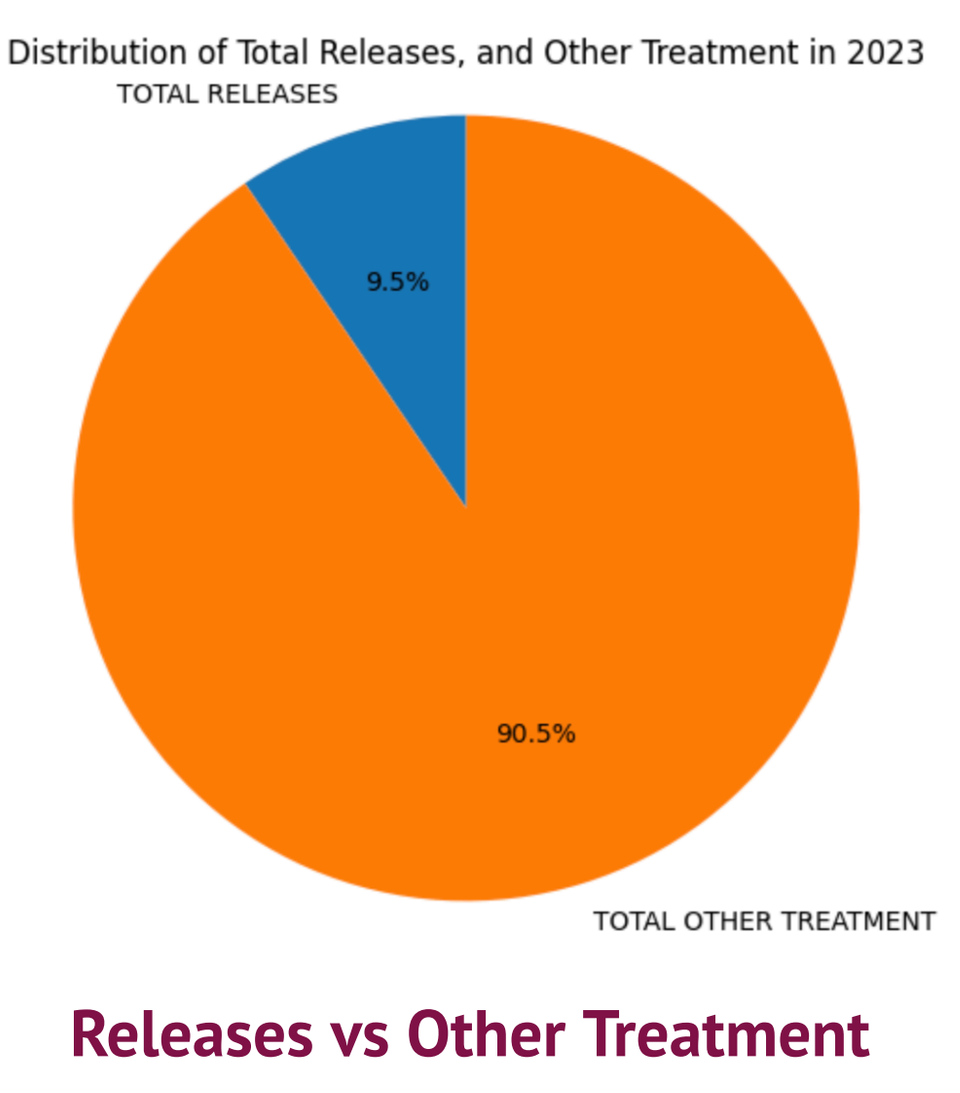
Releases Types

Most values in each category are concentrated near the lower range, with a few extreme outliers significantly inflating the scale.The extreme values suggest that a small number of incidents or sites contribute disproportionately to overall totals in some categories.The central tendency (medians) in all categories is close to zero, emphasizing that most values are small compared to the few outliers.



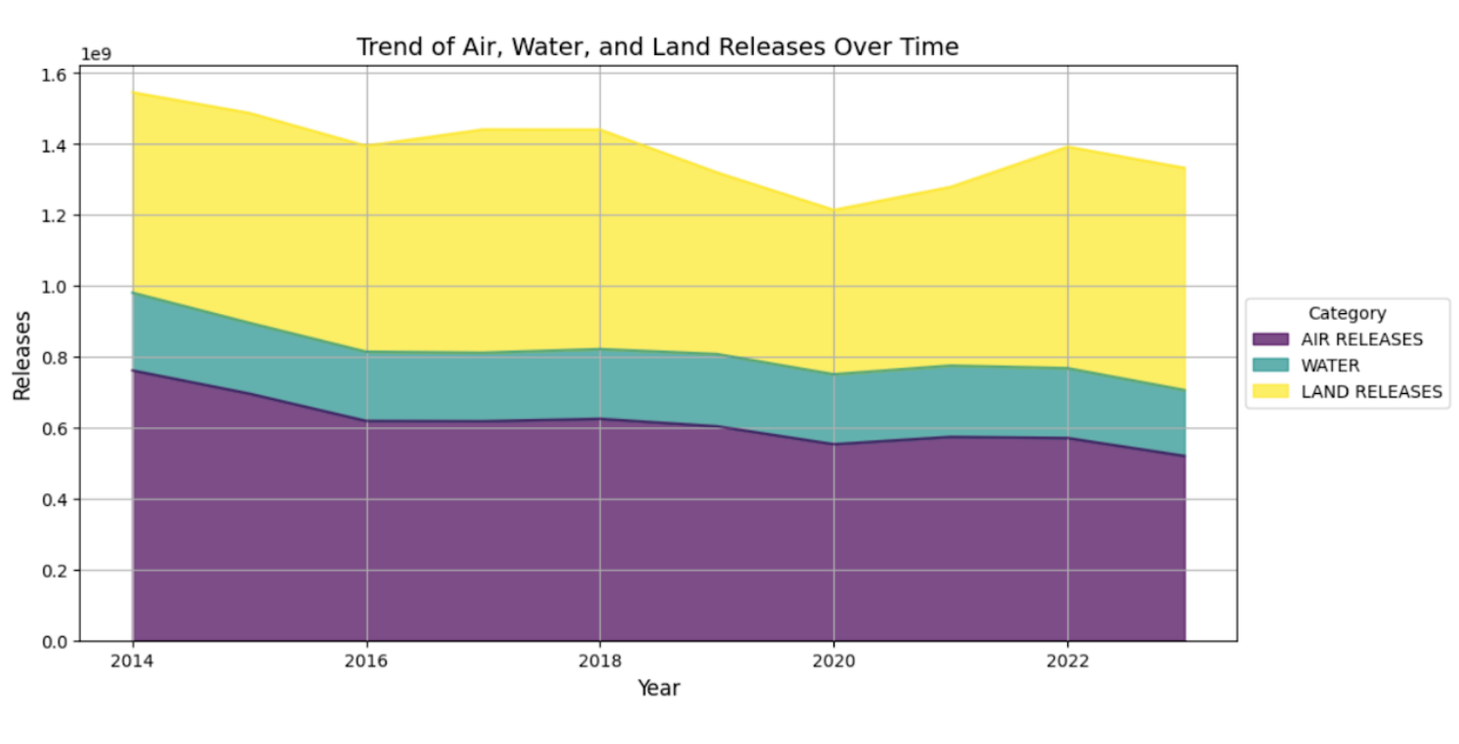
Correlation between Release Categories

The heat map displays the pairwise correlation between five variables: Total Releases, Air Releases, Land Releases, Water Releases, and Other Disposal.The strength of correlation is colour-coded, with red representing strong positive correlations and blue representing weak or no correlations.The is a strongest correlation between Total Releases and Other Disposal (0.95).Indicates that "Other Disposal" is the major driver of overall releases.Minimal correlation between the other categories (e.g., air, land, water releases) suggests that their trends are largely independent of each other.



Releases vs Other Treatment

This chart shows the proportions of Total Releases versus Other Treatment in 2023.Total Releases indicating that only a small portion of waste or pollutants is being directly released.Other Treatment, which implies the majority of waste or pollutants undergo treatment instead of being directly released.A high proportion of waste being treated reflects a strong commitment to environmental sustainability and pollution management.It indicates effective waste treatment systems or policies in place, leading to reduced direct environmental releases.

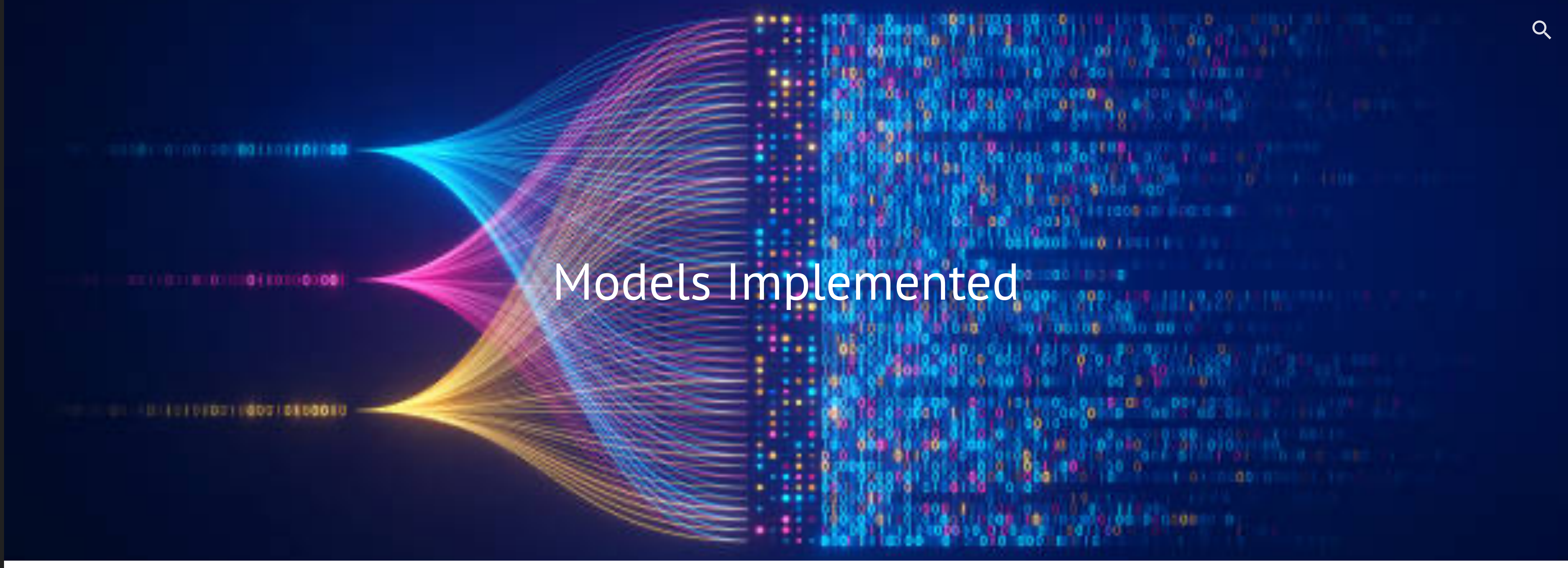


Area chart for Air,Land and Water Release

Air Releases have a significant downward trend from 2014 to 2023 and it reflects success in reducing airborne pollutants.Water Releases are relatively stable, showing slight fluctuations without any dramatic changes and indicates either consistent efforts or stagnant focus on water pollution control.Land Releases are the largest contributor throughout the years.It has initially declined between 2014 and 2018 but increased again from 2019 onward.

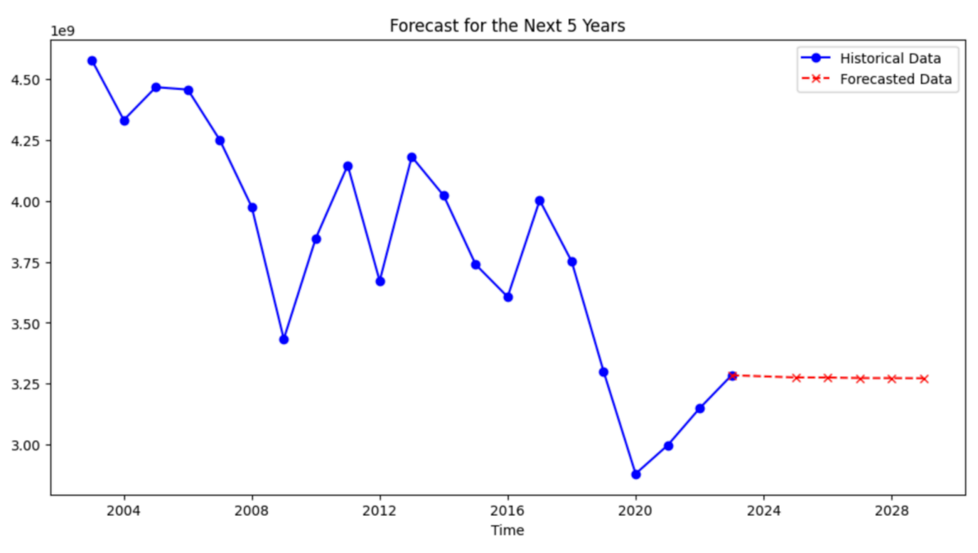
Project Link

National Predictive Analysis of Toxic Releases



Models Implemented

Our predictive analysis utilized advanced models to forecast total chemical releases and related metrics, providing actionable insights into environmental trends and industrial practices.

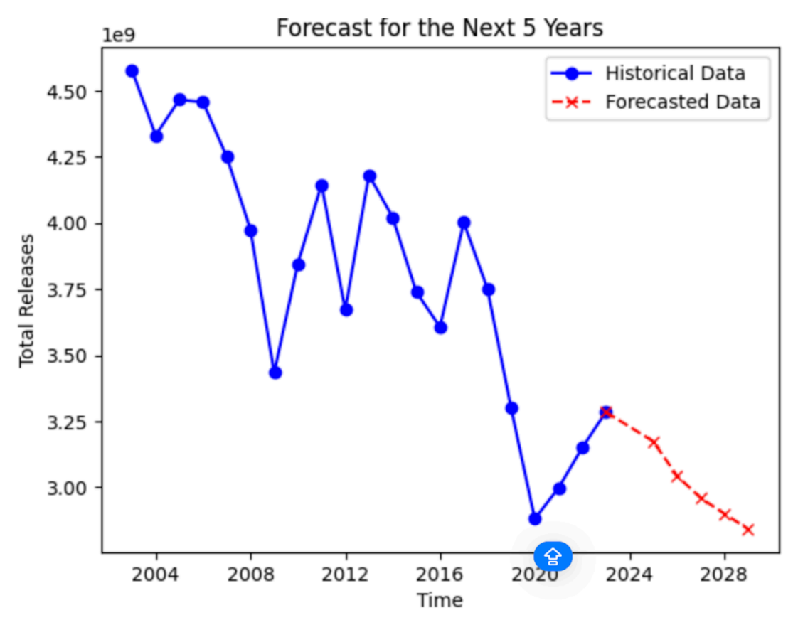


SARIMAX

Designed for time series data, SARIMAX captures trends, seasonality, and incorporates external factors, making it ideal for forecasting with seasonal patterns.

BEST RMSE: 1.12e8

FORECAST (2024) : 3,275,649,000

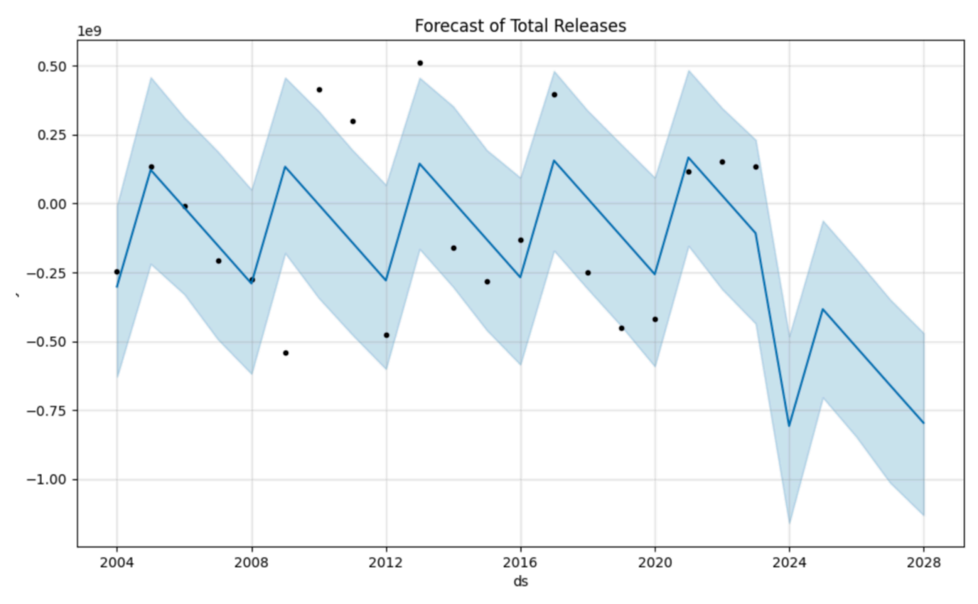


ARIMA

Best suited for stationary time series data, ARIMA models trends and patterns effectively, but it is limited to data without seasonal components.

BEST RMSE: 2.51e8

FORECAST (2024) : 3,173,345,000

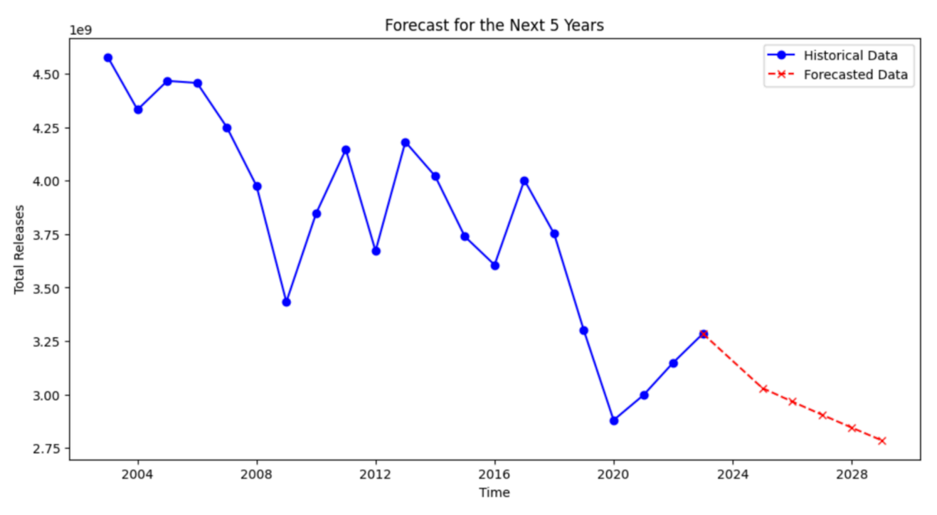


PROPHET

An intuitive forecasting tool designed for time series with trends and seasonality, Prophet is user-friendly but less suited for complex datasets.

BEST RMSE: 5.68e8

FORECAST (2024) : 3,275,649,000

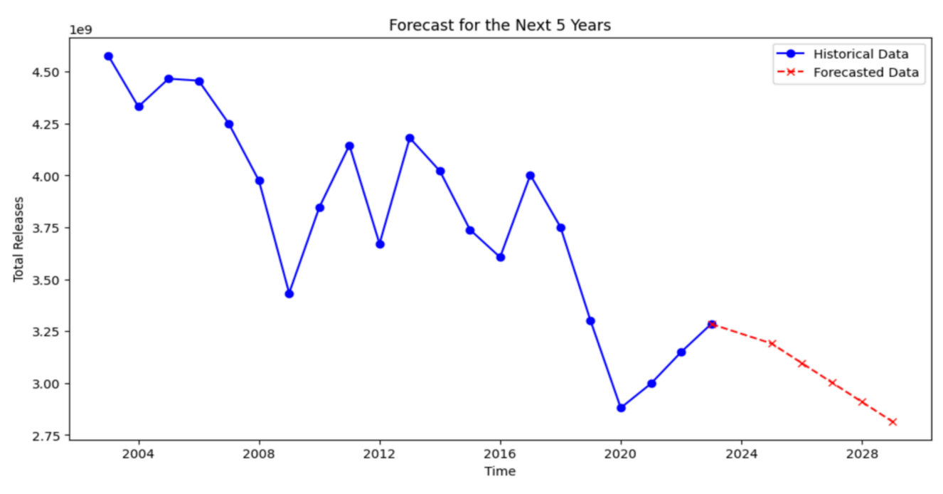


EXPONENTIAL SMOOTHING

This method uses weighted averages to forecast time series with seasonal trends, offering simplicity and reliability for straightforward datasets.

BEST RMSE: 4.10e8

FORECAST (2024) : 3,030,442,000

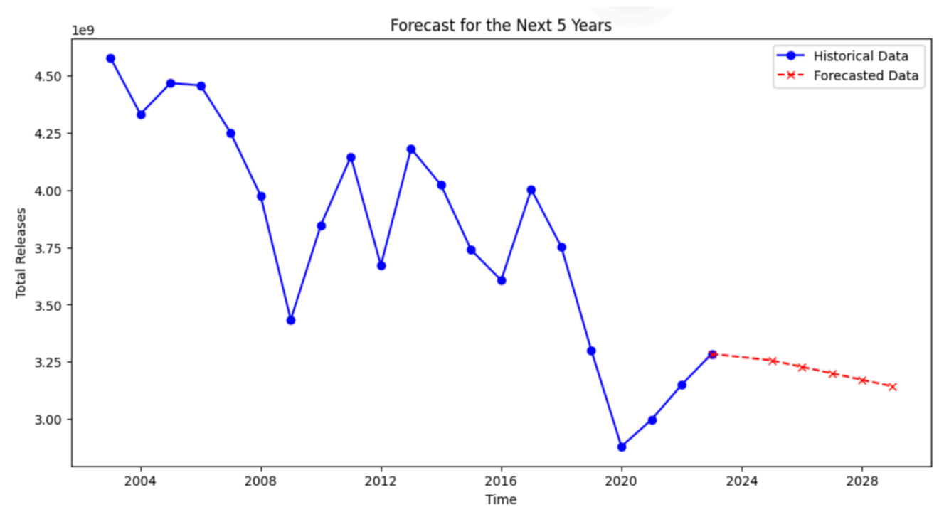


XGBREGRESSOR

Known for its high accuracy, XGBoost is a robust model for capturing non-linear patterns in structured/tabular data.

BEST RMSE: 2.40e8

FORECAST (2024) : 3,190,811,000

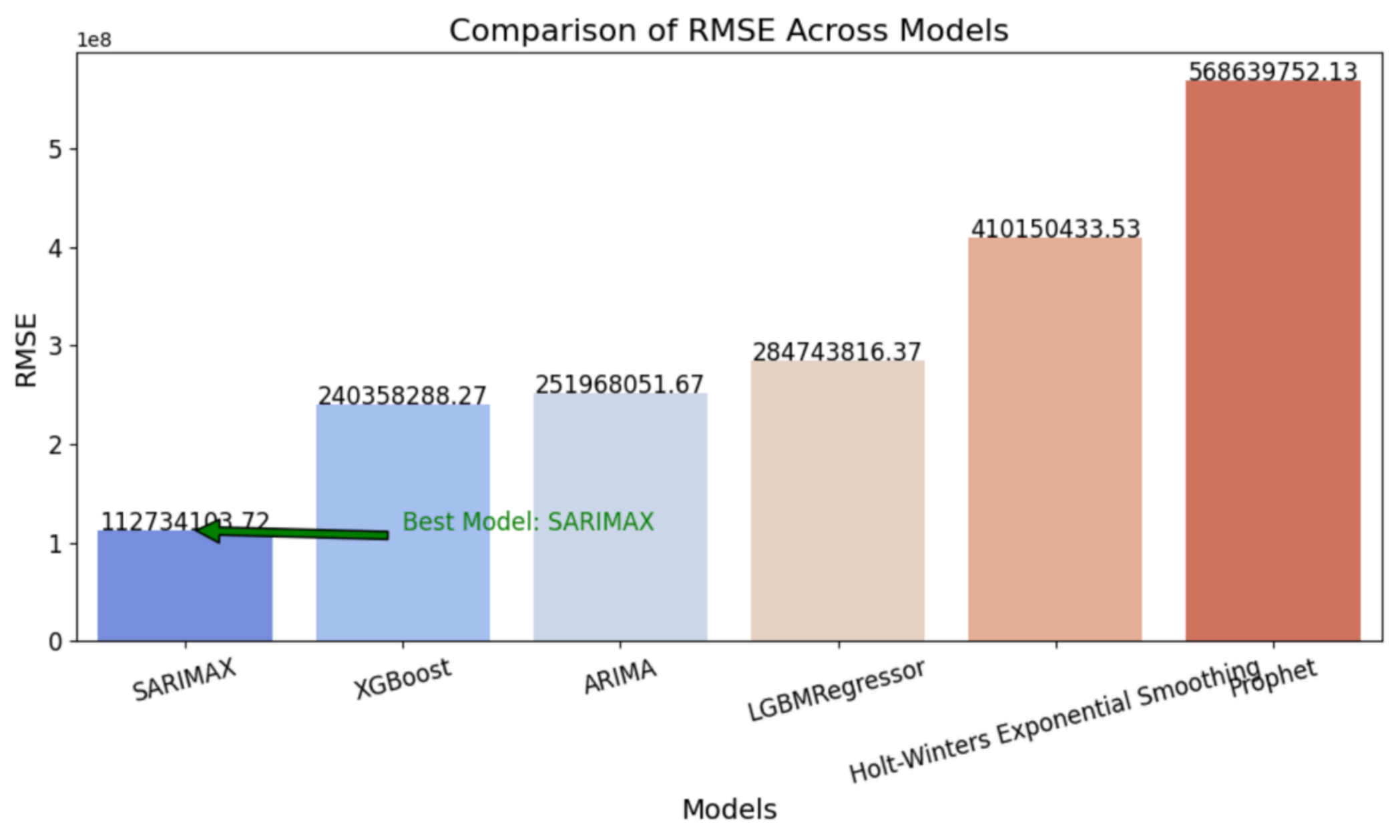


LGBMREGRESSOR

A fast and efficient machine learning model that excels in handling large datasets with complex relationships, making it a top choice for regression tasks.

BEST RMSE: 2.84e8

FORECAST (2024) : 3,256,304,000



RMSE Comparison Report:

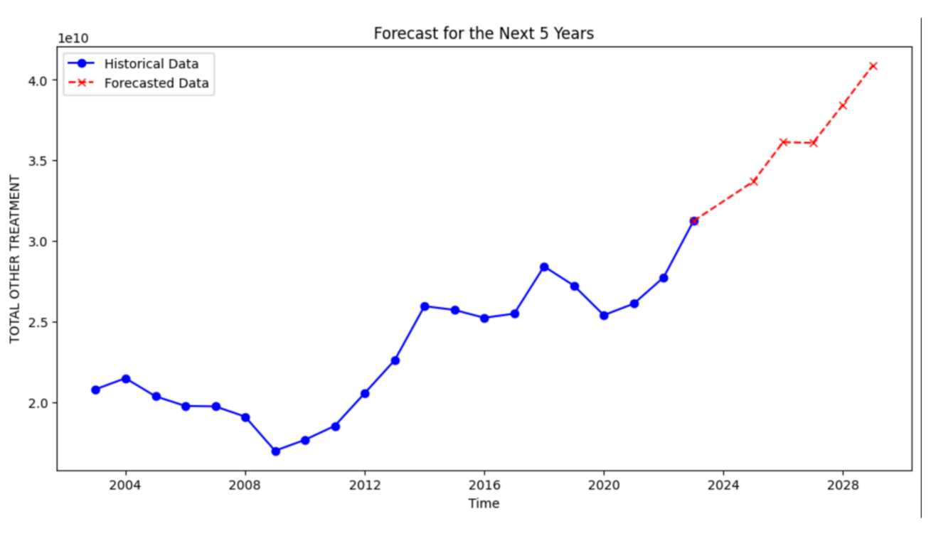
The graph shows RMSE values for various models applied to the dataset. SARIMAX emerged as the best model with the lowest RMSE (1.12e8), outperforming others due to its ability to handle temporal dependencies and seasonality effectively.

Other models like XGBoost (2.40e8) and ARIMA (2.51e8) performed moderately, while Prophet and Holt-Winters struggled due to the dataset's high variability and lack of consistent seasonality.

Recommendation:

SARIMAX is the most suitable model for forecasting in this case, given its superior performance. Further exploration of feature engineering and ensemble techniques could improve other models.

Prediction for Total Treatment , On-site Releases, Off-site Releases :

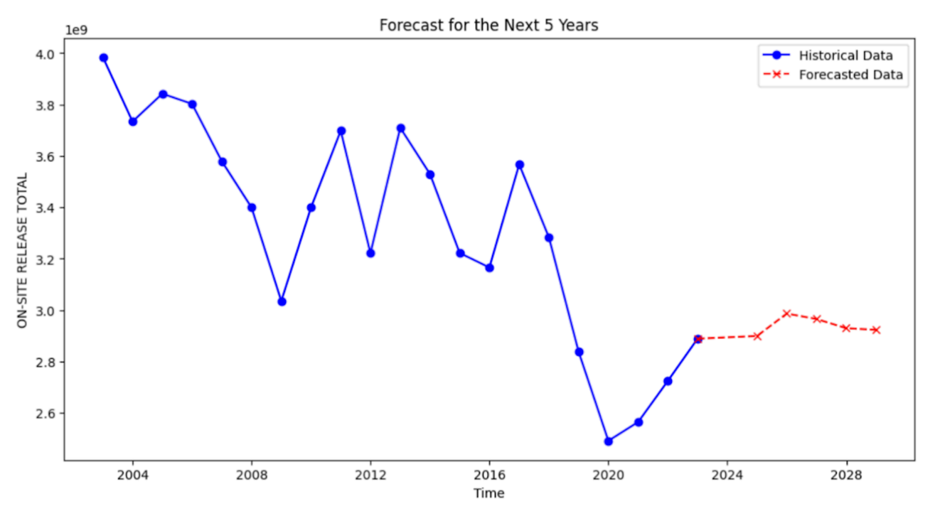


TOTAL TREATMENT

BEST MODEL : XGBOOST

BEST RMSE: 7.56e8

FORECAST (2024) : 33,690,530,000

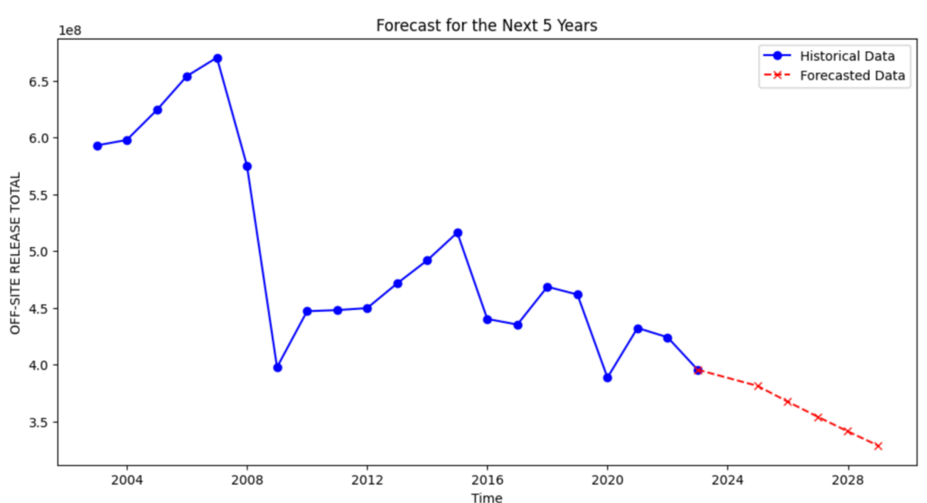


ON-SITE RELEASES

BEST MODEL : SARIMAX

BEST RMSE: 1.25e8

FORECAST (2024) : 2,899,630,000



OFF-SITE RELEASES

BEST MODEL : EXPONENTIAL SMOOTHING

BEST RMSE: 2.39e8

FORECAST (2024) : 381,149,200

DATASET COMPARISON

(674021, 19)

	YEAR	FACILITY NAME	CITY	ST	INDUSTRY SECTOR	CHEMICAL	CARCINOGEN	WATER	OTHER DISPOSAL	ON-SITE RELEASE TOTAL	OFF-SITE RELEASE TOTAL	TOTAL TRANSFER
0	2017	DRS NIS LLC	DALLAS	TX	Computers and Electronic Products	Methanol	NO	0.0	0.0	11436.00	0.0	22113.00
1	2017	AMERICAN ELECTRIC POWER KAMMER/MITCHELL PLANT	MOUNDSVILLE	WV	Electric Utilities	Selenium compounds	NO	170.0	5.0	17200.00	2.0	2.00
2	2017	GEORGIA-PACIFIC LLC	TAYLORSVILLE	MS	Chemicals	Formaldehyde	YES	0.0	0.0	23640.00	5.0	948.00
3	2017	HITACHI ASTEMO AMERICAS INC. - BEREA KY	BEREA	KY	Transportation Equipment	Certain glycol ethers	NO	0.0	0.0	24674.45	0.0	7192.35
5	2017	SIGECO A B BROWN GENERATING STATION	MOUNT VERNON	IN	Electric Utilities	Polycyclic aromatic compounds	YES	0.0	0.0	1.00	0.0	0.00

Raw Data Overview Prior to Model Implementation

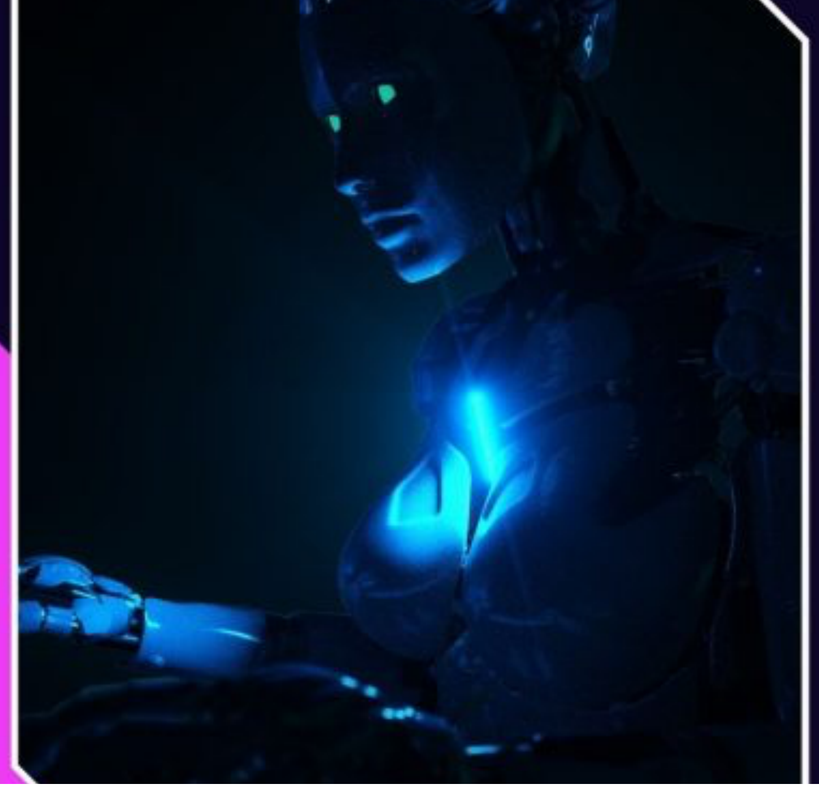
This data is collected from the Environmental Protection Agency (EPA) and has undergone preprocessing to address missing values, outliers, and duplicates before being transformed for modeling.

YEAR	TOTAL RELEASES	TOTAL RELEASES_diff	TOTAL OTHER TREATMENT	TOTAL OTHER TREATMENT_DIFF	ON-SITE RELEASE TOTAL	ON-SITE RELEASE TOTAL_diff	OFF-SITE RELEASE TOTAL	OFF-SITE RELEASE TOTAL_diff
2003-01-01	4.577736e+09	NaN	2.082006e+10	NaN	3.984669e+09	NaN	5.930665e+08	NaN
2004-01-01	4.331598e+09	-2.461371e+08	2.150193e+10	6.818729e+08	3.733736e+09	-2.509334e+08	5.978628e+08	4.796295e+06
2005-01-01	4.466363e+09	1.347647e+08	2.038164e+10	-1.120291e+09	3.842127e+09	1.083911e+08	6.242363e+08	2.637354e+07
2006-01-01	4.456620e+09	-9.743215e+06	1.978378e+10	-5.978555e+08	3.802475e+09	-3.965203e+07	6.541451e+08	2.990881e+07
2007-01-01	4.249822e+09	-2.067978e+08	1.975672e+10	-2.705810e+07	3.579399e+09	-2.230759e+08	6.704232e+08	1.627805e+07
2008-01-01	3.975131e+09	-2.746915e+08	1.912316e+10	-6.335600e+08	3.399738e+09	-1.796604e+08	5.753921e+08	-9.503105e+07
2009-01-01	3.433271e+09	-5.418592e+08	1.700656e+10	-2.116604e+09	3.035633e+09	-3.641058e+08	3.976387e+08	-1.777534e+08
2010-01-01	3.846183e+09	4.129120e+08	1.768271e+10	6.761522e+08	3.399231e+09	3.635981e+08	4.469527e+08	4.931394e+07
2011-01-01	4.145959e+09	2.997759e+08	1.854835e+10	8.656400e+08	3.697988e+09	2.987571e+08	4.479715e+08	1.018822e+06
2012-01-01	3.671236e+09	-4.747233e+08	2.057344e+10	2.025089e+09	3.221549e+09	-4.764391e+08	4.496872e+08	1.715735e+06
2013-01-01	4.181040e+09	5.098038e+08	2.260975e+10	2.036308e+09	3.709514e+09	4.879647e+08	4.715263e+08	2.183903e+07
2014-01-01	4.021577e+09	-1.594629e+08	2.596807e+10	3.358321e+09	3.529757e+09	-1.797565e+08	4.918198e+08	2.029354e+07
2015-01-01	3.738664e+09	-2.829132e+08	2.573684e+10	-2.312315e+08	3.222431e+09	-3.073263e+08	5.162329e+08	2.441314e+07
2016-01-01	3.606244e+09	-1.324192e+08	2.524564e+10	-4.912040e+08	3.166080e+09	-5.635045e+07	4.401642e+08	-7.606879e+07
2017-01-01	4.002350e+09	3.961060e+08	2.550476e+10	2.591292e+08	3.567115e+09	4.010345e+08	4.352357e+08	-4.928495e+06
2018-01-01	3.751576e+09	-2.507742e+08	2.843077e+10	2.926006e+09	3.283036e+09	-2.840784e+08	4.685398e+08	3.330419e+07
2019-01-01	3.299707e+09	-4.518688e+08	2.724179e+10	-1.188981e+09	2.838031e+09	-4.450050e+08	4.616760e+08	-6.863813e+06
2020-01-01	2.880432e+09	-4.192754e+08	2.541154e+10	-1.830254e+09	2.491901e+09	-3.461300e+08	3.885307e+08	-7.314538e+07
2021-01-01	2.997275e+09	1.168425e+08	2.612259e+10	7.110544e+08	2.565049e+09	7.314750e+07	4.322256e+08	4.369498e+07
2022-01-01	3.149024e+09	1.517495e+08	2.775309e+10	1.630495e+09	2.725020e+09	1.599714e+08	4.240038e+08	-8.221852e+06
2023-01-01	3.284579e+09	1.355554e+08	[illegible]	3.492993e+09	2.889134e+09	1.641139e+08	3.954452e+08	-2.855854e+07

Dataset Transformation For Development And Forecasting

This dataset has been transformed with the necessary changes, such as differencing, to ensure the data is stationarized and make it suitable for predicting chemical releases.

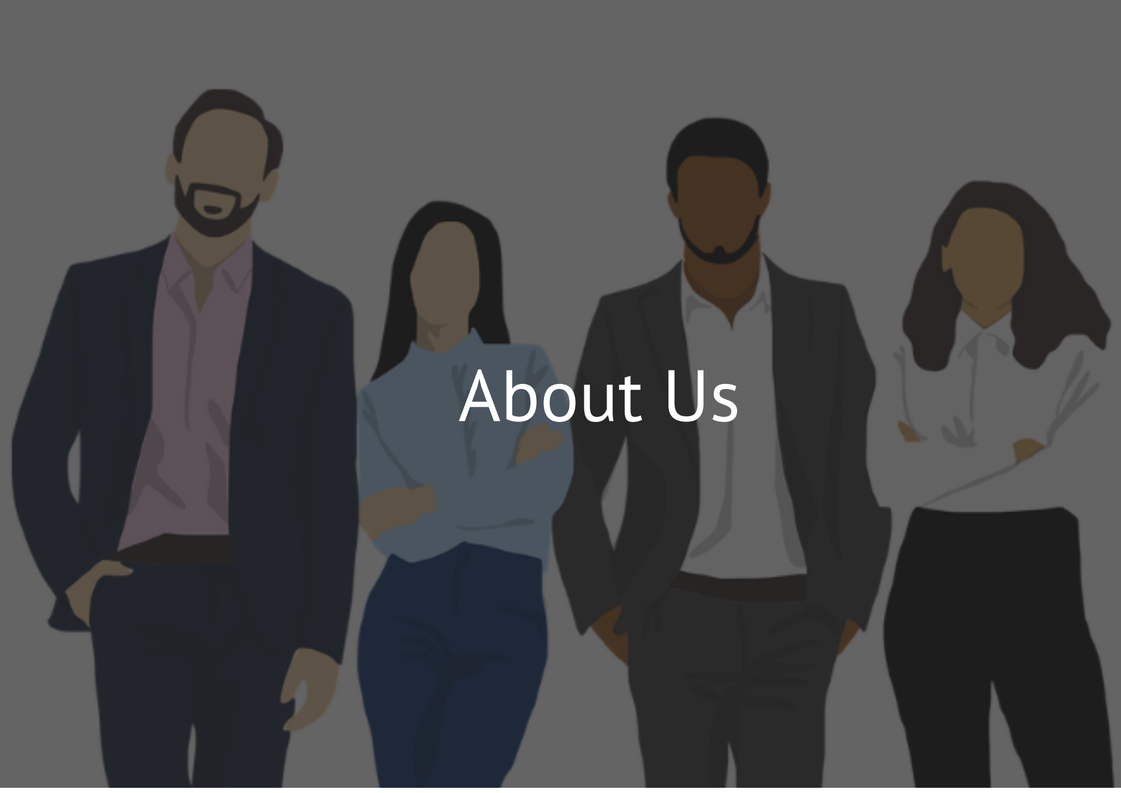
Project Link



Conclusion

Thankyou for being with us(.)

Project Link



About Us

Sathish Kumar Prabaharan

I am **Sathish Kumar Prabaharan**, an aspiring **Data Scientist** currently pursuing my Master's in Data Science at the **University of Colorado, Boulder**. With a strong foundation in **Artificial Intelligence (AI)** and **Machine Learning (ML)**, I am passionate about leveraging cutting-edge technologies to develop innovative solutions that address real-world challenges in the tech industry. My experience spans across **software engineering**, AI model development, and project collaboration, all driving my commitment to advancing AI and ML applications.

sapr5159@colorado.edu

+1 (303) 847-2533

Madhumitha Somasundaram

I am **Madhumitha Somasundaram**, an aspiring Data Scientist currently pursuing my master's in Data Science at the **University of Colorado Boulder.** With a strong foundation in developing web and mobile applications, I am passionate about transitioning into data science. My technical background has given me hands-on experience in software development, problem-solving, and building scalable solutions, which I now aim to combine with advanced data science techniques to drive insights and innovations.

maso2929@colorado.edu

+1 (303) 731 9251

Project Link