**Table of Contents**

[1. Introduction 2](#_Toc185253558)

[2. Methodology 3](#_Toc185253559)

[2.1. Data Analysis with Python 3](#_Toc185253560)

[2.1.1. Description of Dataset 3](#_Toc185253561)

[2.1.2. Data cleaning, preparation, and feature engineering. 3](#_Toc185253562)

[2.1.3. Data Segmentation and Feature Engineering 3](#_Toc185253563)

[2.1.4. EDA to identify trends, anomalies, and correlations. 3](#_Toc185253564)

[2.1.5. Python-based tools and workflows for visualization and exploration 4](#_Toc185253565)

[3. Results 4](#_Toc185253566)

# Introduction

* Purpose, Scope, Objectives, and Structure of the Document

This document serves as an end-of-year review of waste shipments at the facility, providing a comprehensive analysis of waste generation, handling, and shipment activities over the past year. By leveraging advanced data analysis techniques and tools, the review aims to extract actionable insights to enhance operational efficiency, ensure regulatory compliance, and support long-term sustainability initiatives. It consolidates key findings from the annual waste shipment data to guide decision-making processes and establish benchmarks for future improvements.

The scope of this review encompasses all waste shipments over the past year, including hazardous waste, non-hazardous waste, universal waste, TSCA-regulated waste, used oil, and recyclable materials. This review evaluates data trends, waste stream variability, and shipment volumes, while identifying anomalies and compliance risks. The primary objectives of this review are to analyze waste shipment patterns to better understand operational and logistical efficiencies, evaluate adherence to federal and state regulations such as RCRA and TSCA, and develop data-driven recommendations to optimize waste management practices. The analysis also explores opportunities for waste minimization and long-term environmental sustainability by identifying inefficiencies and improving resource allocation.

This document is structured to provide a logical flow from data analysis to actionable recommendations. The first section introduces the purpose, scope, and objectives of the review to establish its context and relevance. The second section outlines the methodology, detailing the dataset preparation, exploratory data analysis, and advanced analytical methods applied during the review. The third section presents key findings, including trends in waste shipment volumes, waste stream variability, and compliance risks. The fourth section focuses on recommendations derived from the analysis, offering strategies to improve operational efficiency, strengthen compliance, and support sustainability initiatives. The appendices include supplementary materials such as Python code examples, detailed notes on analytical methods, and a glossary of technical terms to enhance understanding and reproducibility of the analysis.

This structure ensures the document remains professional, concise, and aligned with the end-of-year review’s objective of evaluating waste shipments while offering insights that can drive strategic improvements in waste management practices.

# Methodology

### Data Analysis with Python

### Description of Dataset

The dataset consists of annual waste shipment records, encompassing various waste streams such as hazardous waste, universal waste, TSCA-regulated waste, and recyclable materials. Key columns in the dataset include attributes such as the on-site origin of the waste (waste\_area), the mass of the waste (QTY\_kg), the waste Type (categorizing waste as hazardous or non-hazardous), and TSDF (identifying the Treatment, Storage, and Disposal Facility). Additional columns such as Group provide operational context, linking waste streams to specific activities like remediation, laboratory operations, or O&M activities. This dataset serves as the foundation for identifying shipment patterns, assessing compliance performance, and deriving operational insights.

The dataset includes over 20 key variables, representing waste movements across the facility and destinations over the past year. Its structure supports a wide range of analyses, from compliance tracking and anomaly detection to shipment optimization. A thorough understanding of these columns and their relationships is essential for performing meaningful data exploration and deriving actionable insights.

### Data Cleaning, Preparation, and Feature Engineering.

The raw dataset required extensive preprocessing to ensure accuracy and consistency for downstream analysis. Missing values were identified and addressed through imputation or removal, depending on the column's significance. Infinite and erroneous values were replaced with NaN, and invalid entries in categorical columns such as Type and TSDF were corrected using consistent naming conventions. Dates, initially stored in varying formats, were uniformly reformatted to %Y-%b-%d for compatibility with time-series models.

Quantitative variables such as QTY\_kg were converted to standard units where necessary, enabling consistent comparisons. Derived columns, such as monthly aggregates of waste shipments, were created to facilitate trend analysis. These cleaning and standardization steps ensured that the dataset met the high-quality standards required for technical and regulatory applications.

### Data Segmentation and Feature Engineering

To enable granular analysis, the data was segmented by key attributes, including waste type, TSDF, and operational grouping. Hazardous waste streams, such as LDR, Lab Packs, Waste Fuel, and ACT, were isolated using the Type column and combined into a unified dataset for hazardous waste analysis. TSDF-based segmentation allowed for facility-specific evaluations, enabling insights into routing and disposal efficiency. Further segmentation by Group identified subcategories such as remediation waste, laboratory waste, and specialized streams like asbestos and hydrazine.

Feature engineering enhanced the dataset’s analytical depth by creating new variables. Examples include time-based features (e.g., shipment frequency by quarter) and waste density calculations. These engineered features supported more robust exploratory analysis and advanced modeling.

### EDA to identify trends, anomalies, and correlations.

EDA was conducted to identify trends, anomalies, and correlations within the dataset, providing a foundation for deeper statistical analysis. Summary statistics revealed aggregate waste volumes and shipment frequencies, segmented by Type, waste\_area, and TSDF. Seasonal trends were identified by grouping shipments by month, exposing peaks and troughs in waste generation. Anomalies, such as unexpected TSDF assignments or outlier values in QTY\_kg, were flagged for further investigation.

Visualizations played a critical role in the exploratory process. Time-series plots highlighted temporal patterns in waste generation, while correlation heatmaps identified relationships between variables. Horizontal bar charts were used to assess the distribution of EPA codes, providing insights into the compliance profile of shipments. This step established a clear understanding of the data’s structure and variability, guiding subsequent analytical methods.

### Python-based tools and workflows for visualization and exploration

Python served as the primary tool for data preparation, exploration, and visualization, leveraging its versatility and computational efficiency. Key libraries included pandas for data manipulation, numpy for numerical operations, and matplotlib and seaborn for static and interactive visualizations. Data cleaning tasks, such as handling missing values and standardizing columns, were automated using custom functions to streamline preprocessing. These scripts ensured reproducibility and consistency across the analysis.

Python-based visualizations provided stakeholders with actionable insights into shipment trends and compliance risks. For example, time-series plots of QTY\_kg highlighted seasonal variations, while bar charts visualized TSDF-specific waste volumes. The workflow’s scalability allowed for iterative analysis, ensuring that all data-driven insights were grounded in a technically rigorous process.

# Results

# Key Metrics

* Export Volumes (QTY\_kg, Container Volume)
* Categories (Type, EPA\_Codes).
* Shipment Frequency
* TSDF usage.

# Waste Stream Analysis by Type and Origin

* Breakdown by waste\_area, Group, and Shipping\_Name.
* Variability in hazardous vs. non-hazardous waste streams.

# Trends and Patterns

* Annual trends in waste volumes and shipment frequencies.
* Insights into variability by waste stream (Type, EPA\_Codes, waste\_area).

# Anomalies and Compliance Risks

* Identification of irregular shipments or outlier data points.
* Implications for regulatory compliance and risk management.

# Operational Impacts and Efficiency

* Waste handling and disposal pathway optimization.
* Opportunities for container utilization and shipping efficiency.

**Recommendations for Improvement**

* Enhancing Operational Efficiency
  + Reducing redundant shipments and optimizing routes.
  + Improving container use and shipment planning.
* Strengthening Compliance and Risk Management
  + Addressing anomalies and improving data accuracy.
  + Proactively managing regulatory reporting requirements.
* Opportunities for Waste Minimization and Sustainability
  + Increasing recycling rates and reducing hazardous waste volumes.
  + Aligning operations with corporate sustainability goals.

**Appendices**

* Bayesian Statistics
* Mixed-Effects Models
* Machine Learning
* Time-Series Analysis
* Comparative Analysis of Analytical Methods
* Python Code Examples for Waste Shipment Analysis
* Glossary of Terms and Acronyms

**Advanced Analytical Methods**

* Bayesian Statistics for uncertainty management and projections.
* Time-Series Analysis for forecasting trends and seasonal variations.
* Machine Learning for predictive insights and anomaly detection.
* Mixed-Effects Models for variability across waste streams and origins.