Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Group Assignment

By

Deepika Bommareddy

Maximiliano Giorgi

Jesse Harold

Sanjay Mathew

Chenyu Wang

PROGRAMME NAME

Big Data & AI in Business

Module Leader:

Max Darby

Trinity Business School

TRINITY COLLEGE

UNIVERSITY OF DUBLIN

February 2025

Declaration

I declare that this work has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.
I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

Generative AI Declaration

Please choose A or B with regards to your use of ChatGPT & other generative AI tools in this project:

- [ ] **A. Nothing to declare.** I did not use ChatGPT or any other generative AI software. (see note)
- [x] **B. I used ChatGPT** or other generative AI software (see note)

NOTE:

- If you answer A and the corrector/supervisor, finds evidence that you have indeed used ChatGPT, this behaviour will be considered as unethical and you will be penalized accordingly with reference to the TCD policy on plagiarism.
- If you answer B, please clearly explain for which chapters or parts of your dissertation you used ChatGPT and how it helped you to improve your learning process within ethical guidelines. You may include your answer – 300 to 600 words approx.- in the appendix.

Signed: : Sanjay Mathew (24361567), Jesse Harold (24350882) , Maximiliano Giorgi (24352012), Krist Wang (24336364) , Deepika Bommareddy(24341243)

Date: 28/02/2025

Panda

Turning Waste into Wisdom: A Data-Informed Strategy for Smarter, Greener Waste Management

Deepika Bommareddy
Maximiliano Giorgi
Jesse Harold
Sanjay Mathew
Krist Wang

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

February, 2025

Table of Contents

About Panda 5

Business objectives 6

Data Architecture 7

- Data Source 8
- Data Ingestion 9
- Data Storage 10
- Data Analytics 12
- Data Integration 13

Strategy Implementation 14

Future considerations 15

References 17

Appendix 18

About Panda

Panda

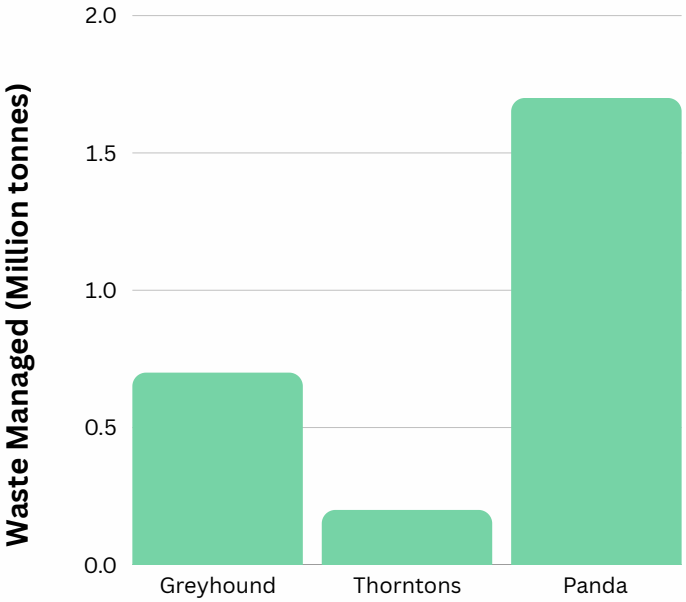
Panda Waste Management is the largest waste management provider in Ireland. Processing and managing 1.7 million tonnes of waste annually. Panda provides:

- Residential services: Bin collections for General, Recycling, and Organic Household waste, as well as access to 19 recycling centres for more specialised disposals.
- Commercial services: Tailored waste management for sectors including, but not limited to, hospitality, manufacturing, and property management. This includes skip hire, digital waste reports, and expert support.
- Circular economy initiatives: Panda's mixed recycling facility (MRF) enables the segregation, export, and repurposing of co-mingled waste from 400,000 homes in Ireland. Additionally, they compost 45,000 tonnes per year and convert non-recyclable commercial waste into Refuse-Derived Fuel (RDF), which can power 100,000 homes. (Panda Waste, n.d.)

Current technology

- AI / Computer Vision: Panda has installed four AI-powered robots in their MRF in Ballymount, Dublin, which maintain a 98% operational uptime and automate quality control by identifying and removing contaminants from recycling lines.
- RFID tagged bins and app: Panda has a mobile application that enables customers to view invoice payments, track bin weight history, and receive reminder alerts for collection days. This information is combined with RFID technology to provide the most accurate and up-to-date data.

(Graph presenting the 19 Panda facilities In Ireland)

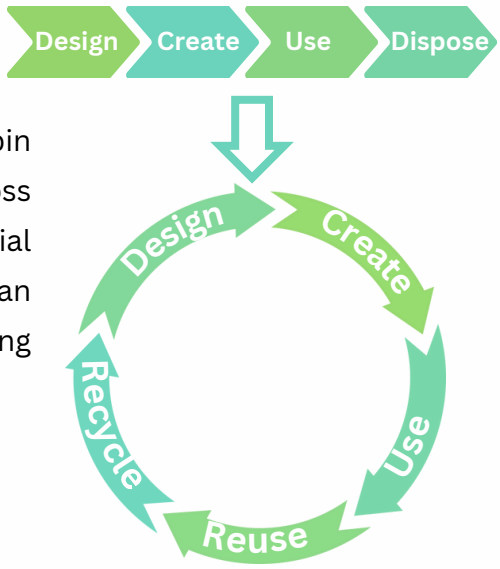


(Graph presenting Waste managed by Panda vs Competitors)

Objectives

Objective #1: Enhance Circular Economy Performance

Panda is at the forefront of sustainability in Ireland. The company has transformed its traditional linear "consume and dispose" business model into a circular economy model (Panda, 2024). By adopting the circular economy model, waste is no longer just discarded it becomes a resource. This shift is essential for Panda to achieve its goal of reaching Net Zero by 2040 (Panda, 2024). With the rapid advancement of big data, Panda has a significant opportunity to use computing power and data analytics to enhance its circular economy, viewing waste as a resource rather than a burden.



Big Data use:

Optimise resource extraction from waste by implementing smart bin devices on Panda-owned bins. These devices will be deployed across recycling, compost, and garbage bins for both municipal and commercial customers. An already existing waste sensor technology named Pello can be used. The big data from these sensors will improve waste sorting efficiency at recycling facilities and circular performance.

Objective #2: Optimising Fleet Operation

Improving fleet operations focuses on how Panda can boost its fuel efficiency when collecting bins from both private customers and businesses, aiming to reach the Net-Zero goal, whilst improving sustainability.

Big Data use:

Big Data enables route optimisation by analysing the location and fill levels of bins. This ensures that only bins with full bins are included in the truck's route, reducing unnecessary stops and fuel consumption. Moreover, fill levels can be used to predict which bins would be the heaviest, saving them for collection at the end of the journey to save fuel. By integrating location and fill data, Panda Waste can dynamically adjust routes as data is received, and ensure maximum efficiency.

Next to planning the optimal route based on bin payment, improving the drivers' fuel-efficient behaviour would also improve fuel efficiency. Companies like Amazon have integrated telematics installed in every active vehicle, which gives their drivers a score based on how fuel-efficient they have driven their vehicles (Breen, et al 2024). Big Data can monitor driver behaviour to give real-time feedback and an overall aggregate driver score after the trip.

Objectives

Objective #3: Improved Customer Experience

Panda has faced challenges in maintaining customer satisfaction and is at risk of losing customers to competitors. Panda's clients are often unhappy due to incorrect billing and frustration with unclear pricing.

Big Data use:

To enhance customer satisfaction, it is crucial that information remains symmetrical for both Panda and its clients to minimise miscommunication and service disputes. By creating dashboards, available to both the client and Panda, will improve information symmetry between the client and Panda. These dashboards will display Information such as waste composition or total waste weight and historical interaction.

Data Architecture

Introduction to Data Architecture

The framework of the data strategy focuses on using big data to improve operations and drive the business objectives. While balancing defensive and offensive strategies is important, the current objectives focus on an offensive approach as Panda strives for sustainable innovation. Although data protection is a key consideration, as information on personal waste could be classified as personal data. The overarching goal is to use big data to optimise and improve processes that allow for Panda to improve operational efficiency, sustainability and customer experience. These three key aspects are evident throughout the entire data architecture.

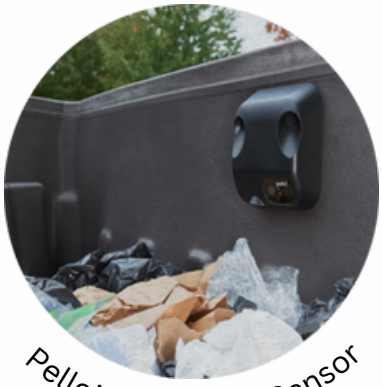
Data Architecture

Data Sources

Internal Data Sources:

1. IoT Fill Level Data:

The Pello system uses IoT weight sensors to monitor bin weight levels, providing streamed data. This supports our objectives by preventing waste overflow, which lowers the effectiveness of the circular economy. The sensors also help route optimisation by identifying the fullest bins for collection. They also improve customer experience through real-time waste tracking, and clarifying pickup charges.



Pello's Smart Bin Sensor

2. Imaging Data:

The Pello system uses imaging technology in bins to monitor contamination and improve waste sorting. By assessing contamination levels, it enhances the circular economy through better resource extraction in recycling. In Ireland, 100,000 tonnes of contaminated recyclable material end up in landfills annually (EPA, 2022). Panda can analyse this data to identify contaminated bins, prevent incorrect disposal and improve recycling efficiency.

3. Bin Location Data:

The real-time bin tracking system on the Pello device streams location data to customers and Panda's fleet. It supports Objective 2, optimises fleet operations with accurate bin locations for efficient pickups, and lets customers track their bins to prevent theft and unexpected events.

4. Telematics Data:

Telematics will be integrated into collection vehicles to gather data via cameras, GPS, and sensors, monitoring driver behaviours like speeding and fuel consumption. This system encourages fuel-conscious driving by providing real-time feedback and efficiency scores.

5. Customer data

Payment information and client interactions encompass all the data sources Panda gathers through its existing operations. Payment data will be used to track which bins are paid for, while client interaction data will be integrated into the client dashboard to enhance information symmetry.

External Data Sources:

1. Traffic Data:

Google Maps is an external source used to optimise the collection route based on real-time traffic updates. It will help optimise routes.

Data Architecture

Data Ingestion

All data must be processed to handle different data types and sources before storage. To ensure seamless ingestion and efficient storage, Apache Kafka will be deployed to handle the event-driven data for the users, where AWS IoT manages the connected IoT devices such as the bin sensors and vehicle telematics (AWS, nd). Kafka will act as the central event streaming platform, allowing real-time processing of customer and traffic events (Apache Kafka, nd).

Data from both services will be processed and enriched through the data journey, before being integrated with cloud storage and analytics following the next stages of the strategy.

Table 1: Data Source, Nature and Respective Objective

Data Source	Structured	Processing	Ingestion	Objective
Iot Sensor	Structured	Streamed	AWS IoT Core	1, 2, 3
Camera	Unstructured	Batched	Apache Kafka	1
Bin Location	Structured	Streamed	AWS IoT	2
Telematics	Structured	Streamed	AWS IoT	2
Customer Data	Structured	Batched	Apache Kafka	2, 3
Traffic Data	Unstructured	Batched	Apache Kafka	2

Data Architecture

Data Storage

For the data storage design, it is vital that three components, size, complexity and accessibility are taken into consideration.

Size:

Panda serves 300,000 municipal and residential customers, each with multiple bins, generating a vast amount of customer and sensor data (Panda, 2024). Managing all this data presents a significant storage challenge. The storage solution must be scalable and cost-efficient, allowing Panda to pay only for the data storage they use as they deploy smart waste sensors across Ireland.

Complexity:

The complexity of the data arises from several factors. First, structured and unstructured data is continuously streamed from each bin, generating large volumes that must be stored efficiently, while some components are processed in batches. Additionally, it is collected from multiple locations and must be centralised in a single repository for effective management and analysis.

Accessibility:

To support objectives, the data must be accessible to three key players in the waste management process: Fleet drivers, who track bin locations and fill levels. Waste management facilities, which need to assess the volume of waste to be processed. Customers, who benefit from insights into their waste production. Our data storage needs to allow for seamless access for all users, regardless of their role in the process.

Data Storage Solution - Centralised Cloud Storage:

Centralised cloud storage is the optimal solution for managing all three data storage considerations. It is well-suited for storing client account information, sensor data, and telematics. Customer payment data can be securely transferred over the internet from a user's account to the cloud, while telematics data collected locally in each vehicle can also be transmitted to the cloud for centralised processing as well as bin data. The recommendation is to use Azure Data Lake, which is a storage solution designed for big data analytics and allows data to store structured and unstructured data and supports batch, streaming and real-time processing. Azure Data Lake also supports automatic scaling and supports tiered storage, reducing costs for infrequent access (Azure, nd).

Table 2: Data Storage System Requirements

Storage System	Cloud Ceneralised	Scalable	Accessible	Supports Big Data	Supports Unstructured / Structured Data
Azure Data Lake	✓	✓	✓	✓	✓

Data Architecture

Data Analytics

Panda centralises waste collection data into a single source of truth (SSOT) to ensure data consistency, accessibility, and high-quality analytics. This structured approach enables the transformation of raw data into multiple versions of truth (MVOTs). This SSOT-MVOT model ensures that the source data and the transformed data for different business functions, such as waste processing, fleet management, and customer engagement are aligned (DalleMule & Davenport 2018). The collected data forms the foundation for descriptive, and predictive analytics.

Descriptive analytics provides an overview of waste collection and composition, including total refuse, recycling, compost volumes, and contamination levels detected through imaging. Additionally, location tracking ensures precise bin monitoring, while customer data verifies payments to determine collection eligibility. The integrated telematics are part of the descriptive analytics as well with relevant driver patterns relating to fuel efficiency. Meanwhile, predictive analytics enhances waste facility planning by estimating capacity needs, forecasting the distribution of waste types, anticipating highly contaminated bins, and optimising collection routes using location and sensor data. Furthermore, telematics data facilitates predictive driver scoring, ensuring ongoing improvements in fuel efficiency and driving patterns.

Table 3: Descriptive Analytics:

Descriptive Analytics	Data Source
Amount of Total Garbage	Iot Sensor Data
Amount of Total Recycling	
Amount of Total Compost	
Contamination of Each Type	Imaging
Bin Tracking	Location Tracker
Payment	Customer Data
Driver behaviour patterns	Telematics

Table 4: Predictive Analytics:

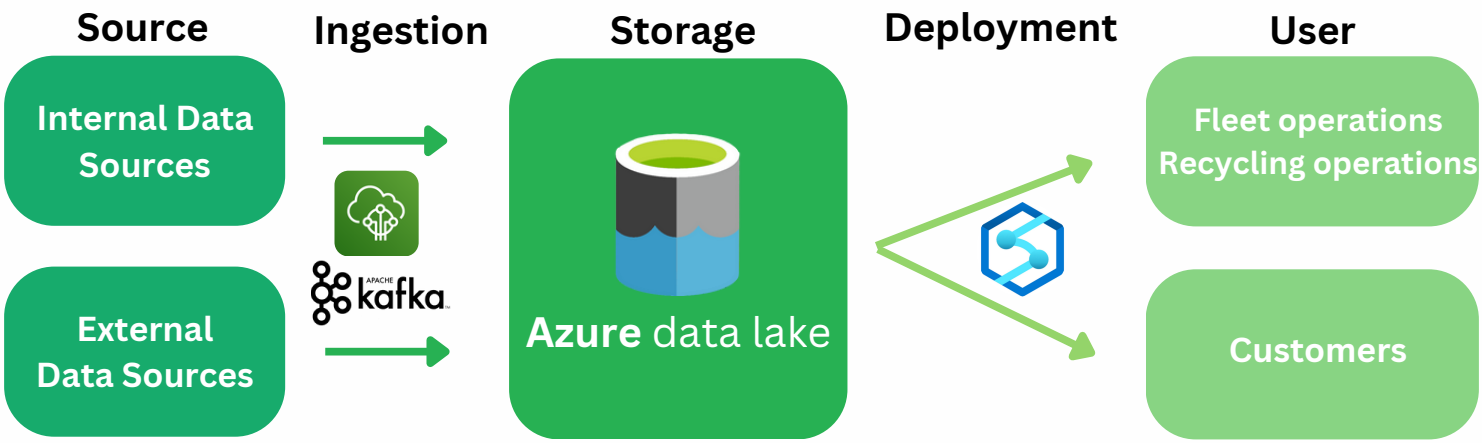
Predictive Analytics	Data Source
Estimate Capacity needed in Waste Facility	Iot Sensor Data
Predict the distribution of waste types	Location Tracker
Anticipate highly contaminated bins	Camera
Predict Optimal Routes	Location, Iot Sensor Data
Predict driver score/ relevant feedback	Telematics

Data Architecture

Data Storage

Once the data has been ingested and stored, the final step in the data storage process is to deliver it to the end users: Fleet Operations, Recycling Operations, and Customers.

The Azure Synapse Analytics allows for this to be accomplished. The cloud-based software integrates seamlessly with Azure Data Lake Storage (Microsoft, 2025). It enables querying, cleaning, and transforming data directly from the Data Lake, ensuring efficient processing and delivery. Additionally, it can be seamlessly integrated with Tableau to provide real-time dashboards, offering actionable insights to end users (Microsoft, 2025).



(Graph presenting Data Pipeline)

Data Governance

Strong data governance is essential for the success of Panda's data strategy. While the strategy carries relatively low risk and differs from industries such as banking and healthcare, where data is significantly more sensitive, it is still crucial to prioritise data governance. Poor practices could be exploited maliciously. That is why the strategy includes three key components of data governance:

Component 1:
Clear Data Use Policy:
All stakeholders are aware of what the data is being used for.

Component 2:
Data Retention Policy:
Data is retained only for as long as necessary.

Component 3:
Data Security & Privacy:
Data is securely Stored and Transmitted.

Data Architecture

Data Integration

Data Integration with AI:

The integration of an AI layer within Panda's data architecture facilitates real-time route prediction, demand forecasting, and operational optimisation. By analysing traffic patterns, bin payment status, and bin weight estimates, the AI system can dynamically optimise fleet routes, thereby reducing fuel consumption, improving delivery times, and enhancing overall efficiency. Additionally, the AI layer, powered by Azure Machine Learning (Azure ML), is a well-suited choice of AI tool since Azure Data Lake and Azure Synapse Analytics are already in use. The model trains predictive models to forecast peak waste collection periods, identify inefficiencies in collection routes and driver behaviour, and assess trends in waste composition. This AI-driven approach allows for real-time optimal route updates and ensures anomaly detection in drivers' fuel-efficient behaviour.

Moreover, Azure ML assists in predicting driver performance and recycling scores, which are displayed on operational dashboards for continuous strategy refinement with the tool Mentor.

Azure Machine Learning

mentor by edriving

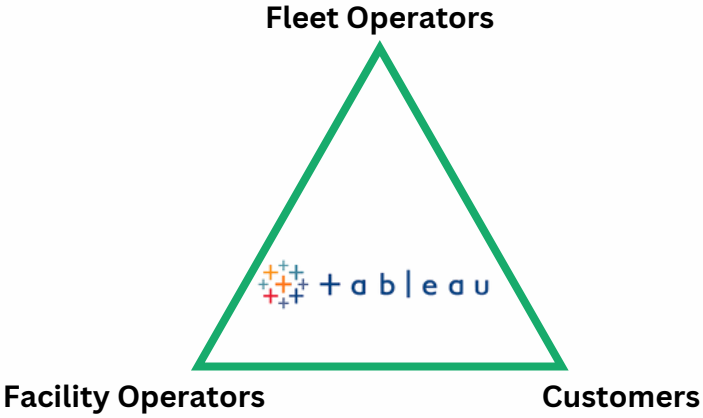
Data Integration with Dashboards:

As mentioned earlier, the three key players have distinct data needs to achieve our objectives, and each will interact with the dashboard in different ways:

Customers: The customer dashboard will provide insights into the amount of waste they generate with each pickup. This data will be integrated with existing billing information and account charges.

Fleet Operators: The fleet dashboard will display optimal routes for waste collection, improving efficiency and reducing operational costs.

Facility Operators: The facility dashboard will enable operators to monitor and estimate the volume of waste entering the facility. This will allow for better planning, resource allocation, and identification of contamination in bins.



Strategy Implementation

Step 1: Pilot Project

Action: Gradually implement telematics and bin sensors across select areas to test the system before full-scale deployment. Pick select regions in Ireland to implement the updated customer dashboard and update based on feedback.

Reason: Rolling out the solution incrementally allows Panda to measure its impact on KPIs such as recycling efficiency, fuel savings, and customer satisfaction while minimising initial costs and risks. As well the Azure Data Lake allows for this as storage scaling is a key feature.

Step 2: Data Literacy

Action: Train employees and customers on how to interpret and use new data insights.

Fleet operators: Understanding real-time route optimisation and driver scoring to improve efficiency.

Waste management: Interpret contamination levels to enhance sorting and recycling processes.

Customers: should understand how to navigate the new dashboard to track their waste usage and billing.

Step 3: Monitoring

Action: Establish clear performance metrics to evaluate the success of the data strategy:

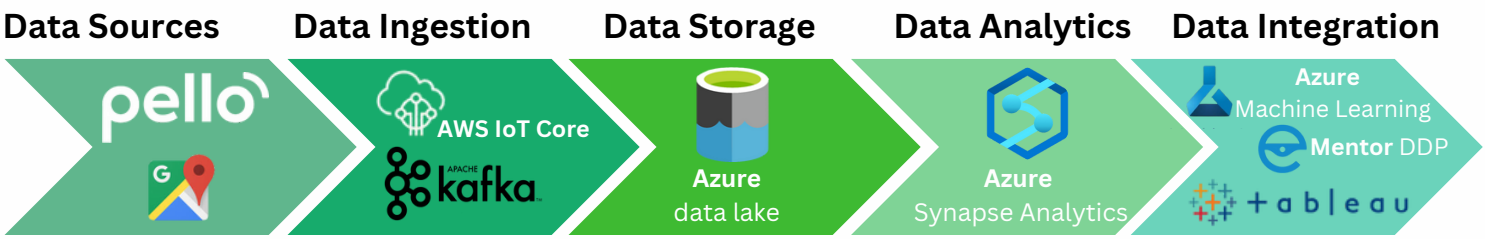
Key Metrics:

Circular Economy: Reduction in contamination levels, improved recycling rates.

Fleets Optimisation: Lower fuel consumption per bin, reduced route inefficiencies.

Customer Experience: Decrease in billing disputes, improved retention rates and growing customer base.

Required Tools



(Graph resenting the Required Tools)

Future Considerations

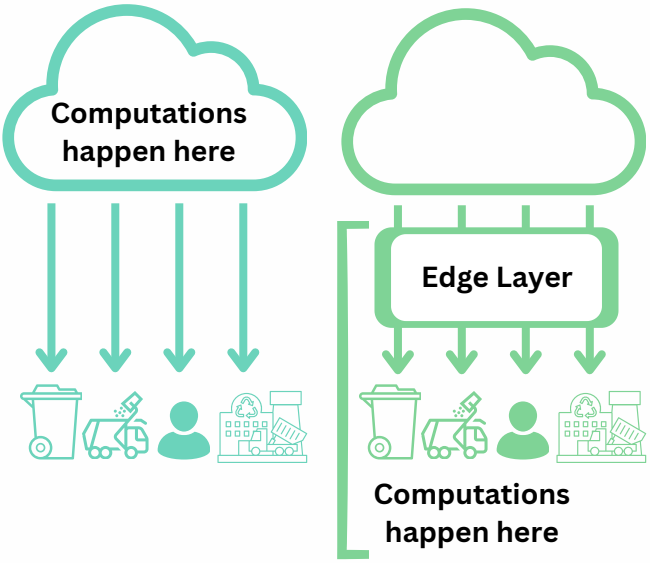
Panda Waste Management can future-proof its data strategy by integrating new data sources and emerging technologies.

Integration of new data sources

- For instance, using additional IoT sensors that can track a diverse range of waste and not just household waste, e.g Sensoneo's smart sensor, which can track electronic waste (Sensoneo, 2025), which is more difficult to identify for Pello sensors.
- Incorporating geospatial or weather data from satellite imagery and drones can enhance monitoring of illegal dumping, as shown in recent studies on drone-based environmental management (Levinson, Smith & Brown, 2025).
- Incorporating waste data and energy market data can provide Panda the opportunity to use AI driven dynamic pricing for selling Refuse-Derived Fuel (RDF) based on energy demand fluctuations, as well as Customers exceeding their waste allowance could face real-time, AI-driven pricing surcharges.

Incorporating additional data sources may result in high deployment and maintenance cost and may strain existing data infrastructure. One way to combat this would be to use edge processing instead of cloud processing to analyse data locally before sending only relevant insights to the cloud.
Dynamic RDF pricing can be pilot tested in smaller areas before being implemented in larger areas to see if it can work well. One possible issue of dynamic pricing is that it could disproportionately affect lower-income customers by increasing RDF prices in specific areas, e.g North Dublin vs South Dublin, which is displayed in the graph below

Cloud Computing vs Edge Computing



Waste Distribution & RDF Pricing In Dublin

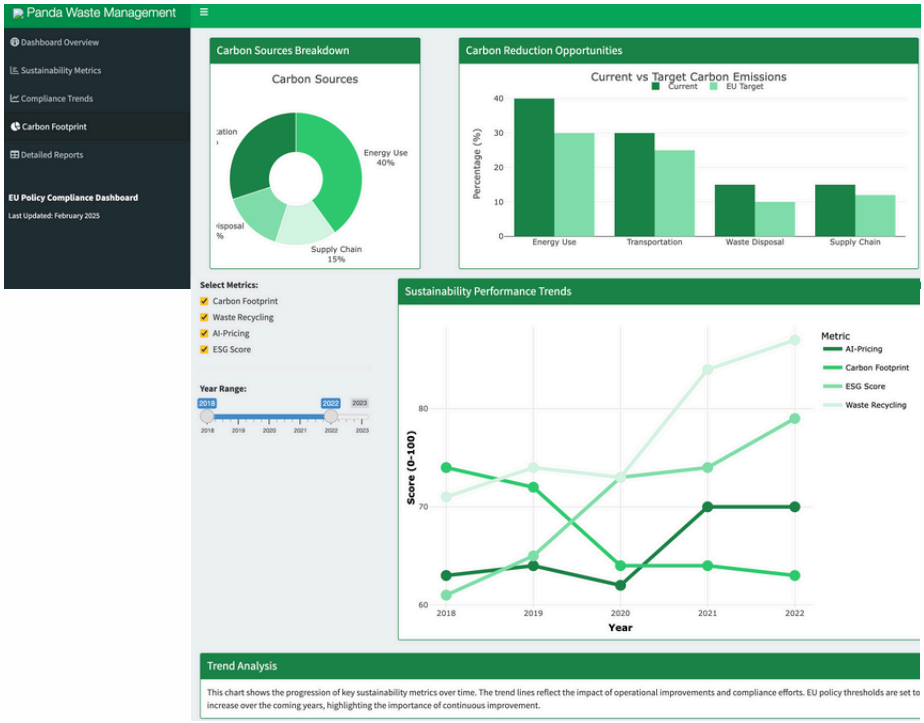
Future Considerations

Sustainability and ESG compliance

Panda must comply with international and regional sustainability policies that shape how businesses report and reduce environmental impacts. Key frameworks include:

- Task Force on Climate-related Financial Disclosures (TCFD) – Establishes guidelines for how businesses should disclose environmental risks and carbon reduction strategies. (Financial Stability Board, 2024).
- Corporate Sustainability Reporting Directive (CSRDs) – An EU directive that requires detailed, standardised ESG reporting for large companies operating in the EU (European Commission, 2024).
- Global Reporting Initiative (GRI) Standards – A widely used framework for tracking and reporting sustainability metrics, helping businesses demonstrate environmental responsibility and understand their impact on the economy.(Global Reporting Initiative, 2024)

To comply with these frameworks and maintain high standards, Panda must make significant investments in data collection, auditing, and reporting. Meeting evolving ESG requirements would require substantial funding, whether through in-house compliance teams or third-party services.
One approach to tackling this challenge is to collaborate with EU policymakers to align Panda's carbon tracking and sustainability initiatives with evolving climate policies. Creating a feedback loop would guarantee coordination between both parties, preventing Panda from perpetually playing catch-up.
Furthermore, Panda can develop real-time sustainability dashboards that furnish businesses and individuals with comprehensive reports on carbon impact and recycling efficiency, crucial metrics relevant to the frameworks discussed (refer to sample dashboards below). By providing tailored sustainability analytics, Panda can assist customers in monitoring their environmental contributions and achieving their ESG goals. This not only aids compliance with these frameworks but also bolsters customer trust and nurtures long-term partnerships.



(Potential Dashboard)

References

Amazon Web Services (AWS), n.d. AWS IoT Core – Connect Devices to the Cloud. Available at: https://aws.amazon.com/iot-core/

Apache Kafka, n.d. Apache Kafka. Available at: https://kafka.apache.org/ [Accessed 28 February 2025].
Breen, L. and Horton,, 2024. How MTData built a CVML vehicle telematics and driver monitoring solution with AWS IoT. Amazon Web Services. Available at: https://aws.amazon.com/blogs/iot/how-mtdata-built-a-cvml-vehicle-telematics-and-driver-monitoring-solution-with-aws-iot/

DalleMule, L. and Davenport, T.H., 2017. The 2 Types of Data Strategies Every Company Needs. Harvard Business Review. Available at: https://hbr.org/2017/05/whats-your-data-strategy

Environmental Protection Agency, 2011. Annual Environmental Report 2011. Available at: https://epawebapp.epa.ie/licences/lic_eDMS/090151b28043798e.pdf

Environmental Protection Agency, 2022. National Waste Statistics 2021. Cork: EPA.
European Commission, 2024. Corporate sustainability reporting. Available at: https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en

Financial Stability Board, 2024. Task Force on Climate-related Financial Disclosures (TCFD). Available at: https://www.fsb-tcfd.org/

Global Reporting Initiative, 2024. GRI Standards: Sustainability Reporting Framework. Available at: https://www.globalreporting.org/standards/

MDPI, 2022. The role of drones in monitoring illegal waste disposal: A review of recent studies. Available at: https://www.mdpi.com/2072-4292/14/16/3923

Microsoft, 2025. Azure Synapse Analytics | Microsoft Azure. Available at: https://azure.microsoft.com/en-us/products/synapse-analytics#Features

New York Post, 2025. Drones could one day monitor New Yorkers' garbage to help enforce new trash bin rules. Available at: https://nypost.com/2025/01/22/us-news/drones-could-one-day-monitor-new-yorkers-garbage-to-help-enforce-new-trash-bin-rules

O'Callaghan Moran & Associates, 2014. Environmental Impact Statement for Panda Waste Services. Cork: OCMA

Panda Waste Management, 2024. Internal Customer Churn Data. Dublin: Panda WM.
Recycleye, 2023. Case Study: AI Waste Sorting at Ballymount MRF. London: Recycleye Ltd.
Sensoneo, n.d. Top waste streams using smart sensors. Available at: https://sensoneo.com/top-waste-streams-using-smart-sensors

Switcher.ie, 2025. Price Comparison Report: Waste Collection Services. Dublin: Switcher Ltd.
Thorntons Recycling, 2024. Another landmark year for Thorntons Recycling. Available at: https://irishtrucker.com/News/14398

Appendix

AI Declaration

In this project, AI tools such as ChatGPT and Grammarly enhanced research, analysis, visualization, phrasing, and wording. ChatGPT was used to help find sources, explain sustainability concepts, and explore IoT-enabled waste tracking technologies. Additionally, AI assisted in writing the code for visualizations, including heatmaps and dashboards, which were created using R in RStudio. The use of AI streamlined data analysis, report structuring, and technical explanations, ensuring a more efficient and comprehensive approach to the project.