
# IOWA Case Competition - NovaSphere Logistics

## Optimizing NovaSphere Logistics: Data-Driven Insights for Cost & Sustainability



UCONN
SCHOOL OF BUSINESS

# Team Members

**Abishek Karnan Rajesh**
abishek.karnan_rajesh@uconn.edu

**Bharath Girirajan**
bharath.girirajan@uconn.edu

**Sudarsan Nullur Murali**
sudarsan.n_m@uconn.edu

**Tejasri Voota**
teja_sri.voota@uconn.edu



UCONN
SCHOOL OF BUSINESS

# Problem Statement & Business Case

NovaSphere Logistics is a mid-sized last-mile delivery service provider, specializing in e-commerce and retail deliveries across metropolitan areas. The company aims to deliver fast, reliable, and cost-effective logistics solutions while optimizing operational efficiency and sustainability.

**Key challenges include**

- High Costs → Increasing operational expenses.
- Delivery Delays → Late deliveries affecting customer satisfaction.
- High Carbon Emissions → Environmental concerns from inefficient logistics.



**Optimizing operations will lead to**

- Cost Reduction → Lower operational expenses & improved profitability.
- Faster Deliveries → Improved efficiency & customer satisfaction.
- Sustainable Logistics → Reduced carbon footprint & eco-friendly operations.



UCONN SCHOOL OF BUSINESS

# Dataset Overview - NovaSphere Logistics

**Dataset Summary**

The dataset consists of 350 rows and 5 key columns, capturing operational metrics for NovaSphere Logistics. It includes delivery time, package weight, cost, and emissions, providing insights into logistics efficiency and sustainability

**Key Features in the Dataset**

| | |
|---|---|
| Delivery_ID | Unique identifier for each delivery. |
| Delivery Time (minutes) | Time taken for each delivery, including delays |
| Package Weight (kg) | Weight of the package being delivered |
| Carbon Emissions (kg $CO_2$) | Environmental impact per delivery. |
| Cost ($) | Operational cost incurred for the delivery. |

**Data Quality Issues Identified Missing Values**: Package Weight (kg): 12 rows contain missing values.

**Incorrect or Outlier Entries**

Negative Delivery Time (-1 min): 15 row contained an invalid negative value.
Carbon Emissions = 9999 kg $CO_2$: 10 row had an unrealistic emission value.
Cost = $0: 7 rows had cost as zero, representing free deliveries due to loyalty points or other discounts[1].



[1] Information received from Lysanias that $0 cost mean loyaly points or discounts.

# Data Preprocessing - Ensuring Clean & Reliable Data

**Steps Taken for Data Cleaning & Preprocessing**

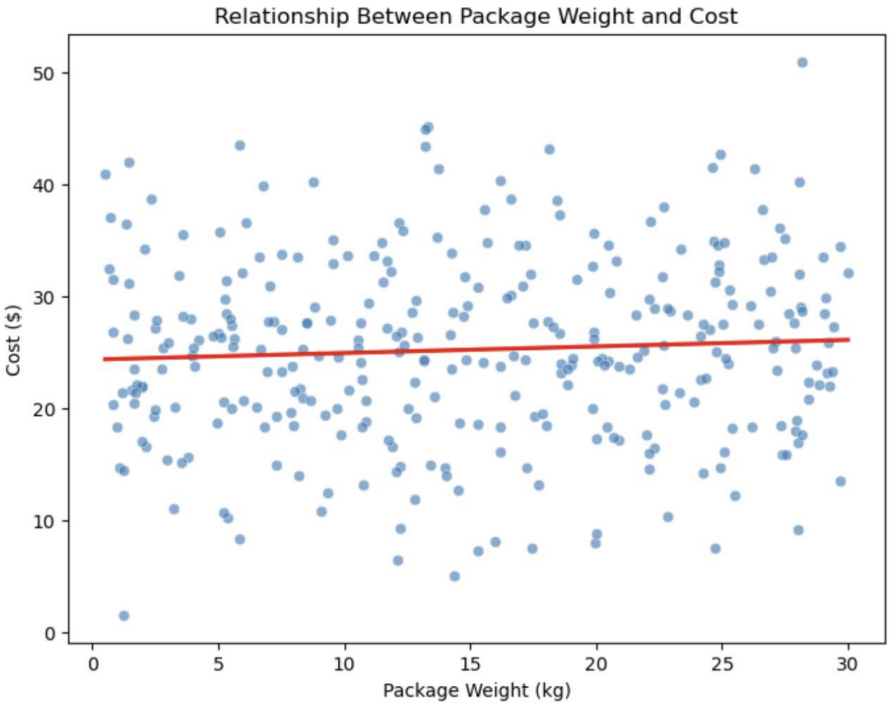
- **Handled Missing Values** → Removed 12 rows with missing package weight data.

- **Fixed Invalid Entries** →
  - Negative Delivery Time (-1 min): Removed 15 rows with incorrect values.
  - Carbon Emissions = 9999 kg $CO_2$: Replaced 10 rows with the mean (7.48) of valid emissions.

- **Removed Cost = $0 Rows** → 7 rows were dropped as they were outliers and did not contribute to cost analysis.

- **Final Cleaned Dataset** → After preprocessing, we were left with 317 rows of reliable data.

# Relationship Between Package Weight and Cost




- **Cost slightly increases with weight:**

0-10 kg: Average cost $24.72
10-20 kg: Average cost $25.24
20-30 kg: Average cost $25.69
The difference is small, meaning weight is not a strong driver of cost.

- **Variation in Costs Suggests Other Influencing Factors:**

Some lightweight packages (<10 kg) have high costs, indicating cost drivers beyond weight (indicating some external factors).
Some heavy packages (25-30 kg) have low costs, suggesting possible bulk shipment optimizations or discounts(indicating some external factors).

- **Insights from the Regression Line**

The red regression line in the scatter plot shows a slight positive correlation, meaning cost increases marginally with weight.
However, the spread of data points is large, confirming that package weight alone does not determine cost.



UCONN SCHOOL OF BUSINESS

# Relationship Between Delivery Time and Cost

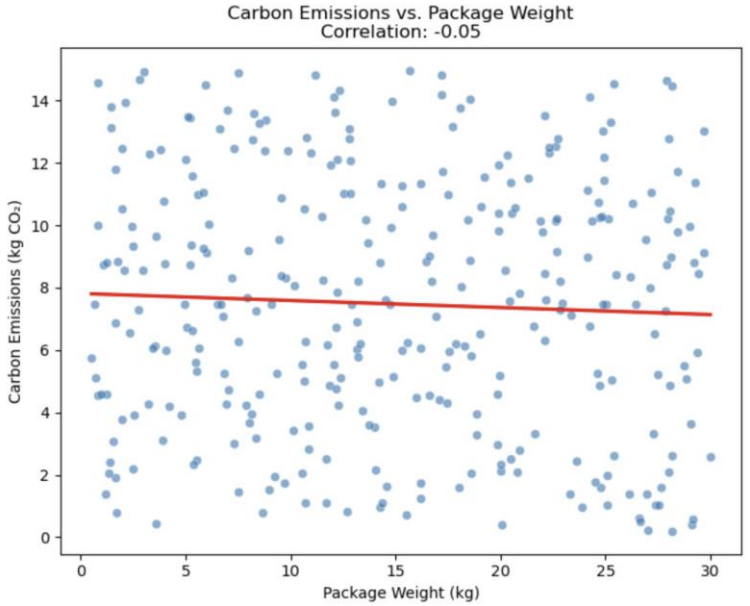

- **No Strong Correlation Between Delivery Time and Cost:**

The correlation coefficient is close to 0, confirming that delivery time does not directly impact cost.
The red regression line is almost flat, showing no clear upward or downward trend.

- **Longer Delivery Times Do Not Mean Higher Costs:**

Some short delivery times (<30 min) have high costs, while some longer deliveries (>90 min) have lower costs.
This suggests that cost is not determined by time alone but by other logistical factors.

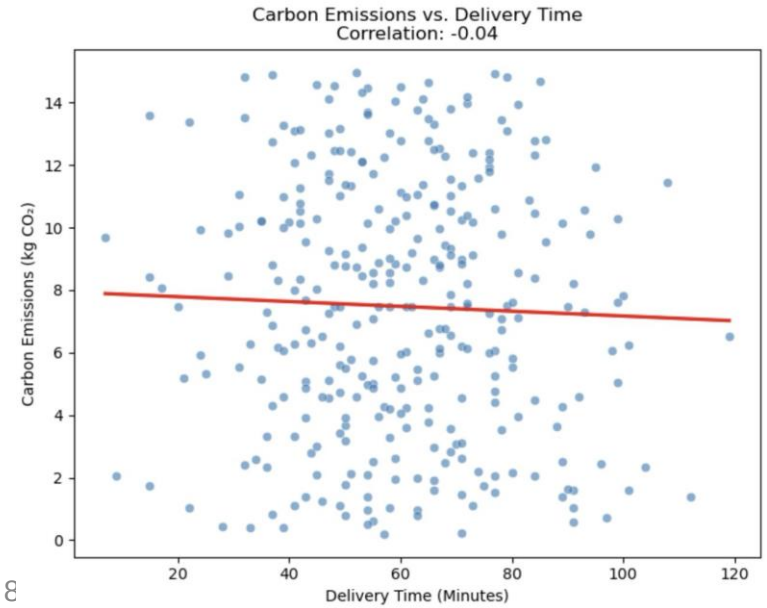
# Analysis of Carbon Emissions




- **Carbon Emissions vs Package Weight**

Correlation: -0.05 → Almost no relationship between package weight and carbon emissions.
The flat regression line confirms that package weight does not significantly impact emissions.

**Possible Reasons:**
Vehicles carry multiple packages, making individual package weight less relevant.
Route optimization & vehicle type (fuel efficiency) may play a bigger role.




- **Carbon Emissions vs. Delivery Time**

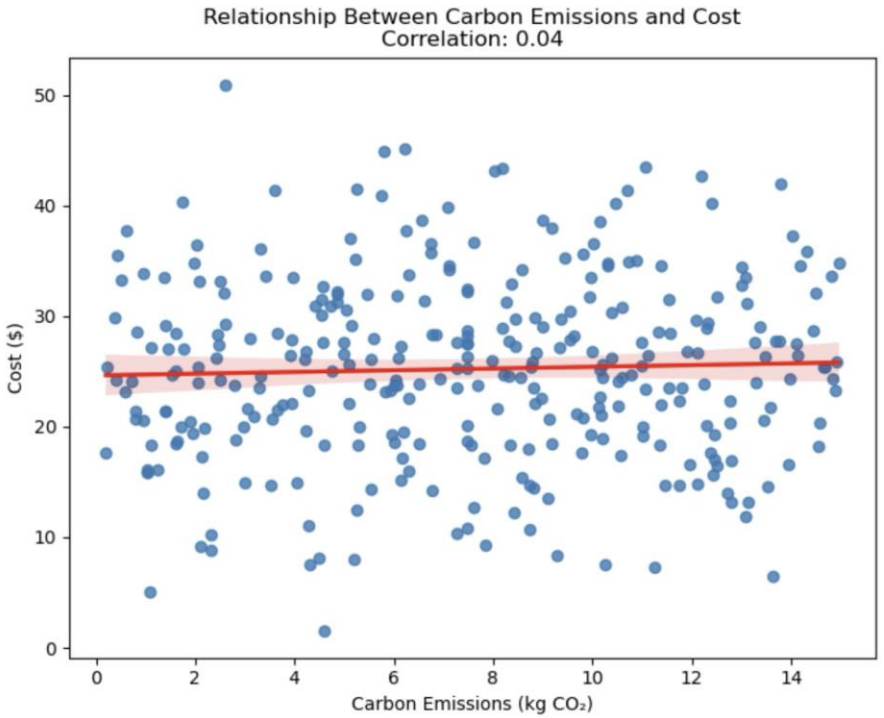
Correlation: -0.04 → No strong relationship between delivery time and carbon emissions.
The flat regression line suggests that longer delivery times do not necessarily lead to higher emissions.

**Possible Reasons:**
Traffic congestion, idling, or stop-and-go driving may affect emissions more than delivery duration.
Route type (highway vs. urban roads) could have a greater influence on emissions.



UCONN
SCHOOL OF BUSINESS

# Relationship Between Carbon Emissions and Cost




- **Weak Positive Correlation Between Carbon Emissions and Cost**

Correlation: 0.04, indicating a very weak relationship between emissions and cost. The red regression line is nearly flat, suggesting emissions do not strongly influence cost.

- **Wide Cost Variability Across Emission Levels**

Deliveries with low emissions (~2-4 kg $CO_2$) have a wide range of costs, from $10 to over $40.
Similarly, deliveries with high emissions (~12-14 kg $CO_2$) also have low and high cost variations.
This suggests that cost is influenced by other operational factors.



UCONN
SCHOOL OF BUSINESS

# Why Use LLMs to Analize Data?

**Automating Data Analysis with AI**

- Traditional data analysis requires manual SQL queries and expertise.
- LLMs enable natural language querying, making data insights accessible to non-technical users.

**Faster & Smarter Query Execution**

- The LLM-powered agent converts user queries into SQL dynamically.
- Eliminates the need for manual query writing, improving efficiency and accuracy.

**Intelligent Data Processing & Preprocessing**

- The system can identify missing values, date parsing, and type conversions.
- Ensures clean, structured, and optimized data for analysis.

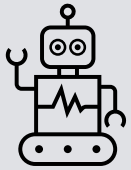
**Seamless User Experience**

- A Streamlit-powered UI enables real-time interaction with the LLM-based analyst.
- Users can upload datasets, ask questions, and get instant AI-generated responses.



UCONN
SCHOOL OF BUSINESS

# How the Problem is Solved & Technologies Used



**Technologies Used**

- LLM Model: OpenAI GPT-4 for natural language to SQL conversion.
- Database Engine: DuckDB for efficient query execution.
- Data Handling: Pandas for preprocessing and analysis.
- Web Interface: Streamlit for a user-friendly data exploration experience.

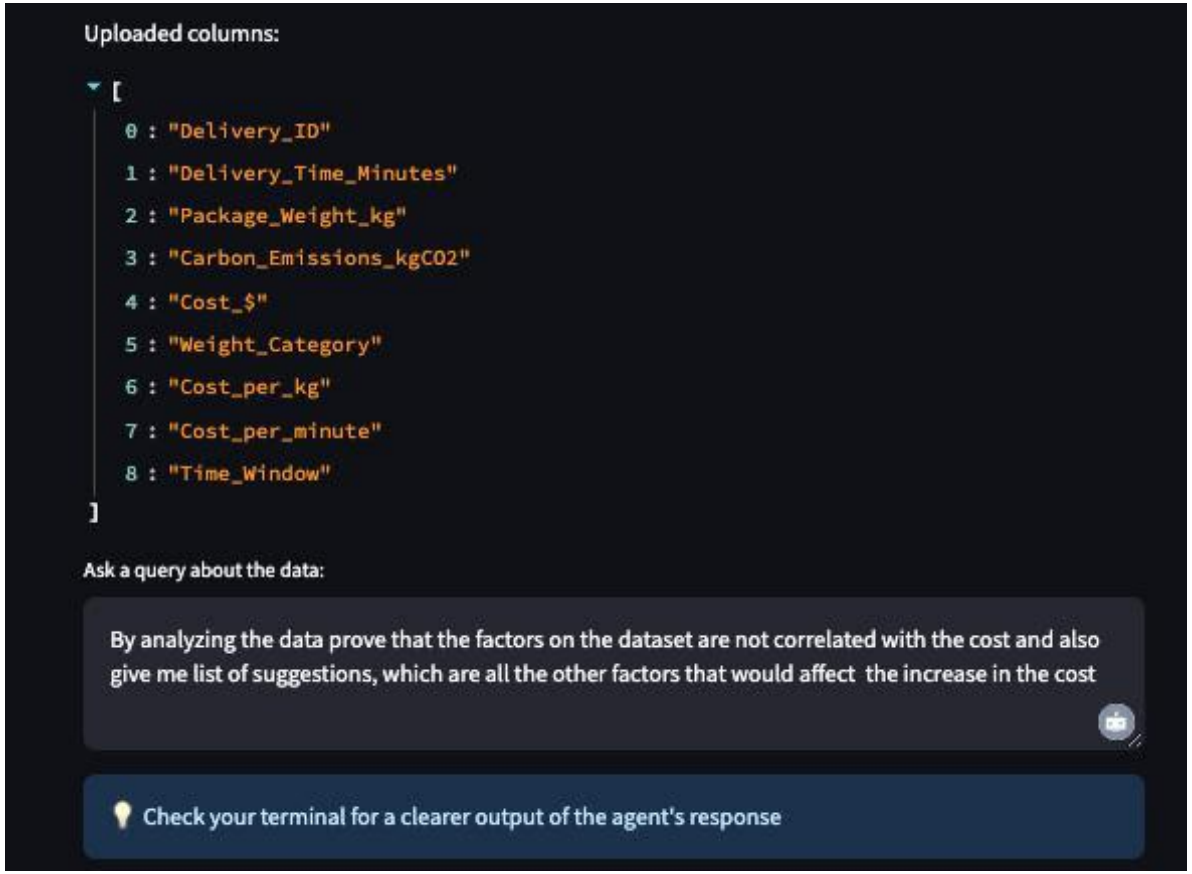
**Automated Query Processing**

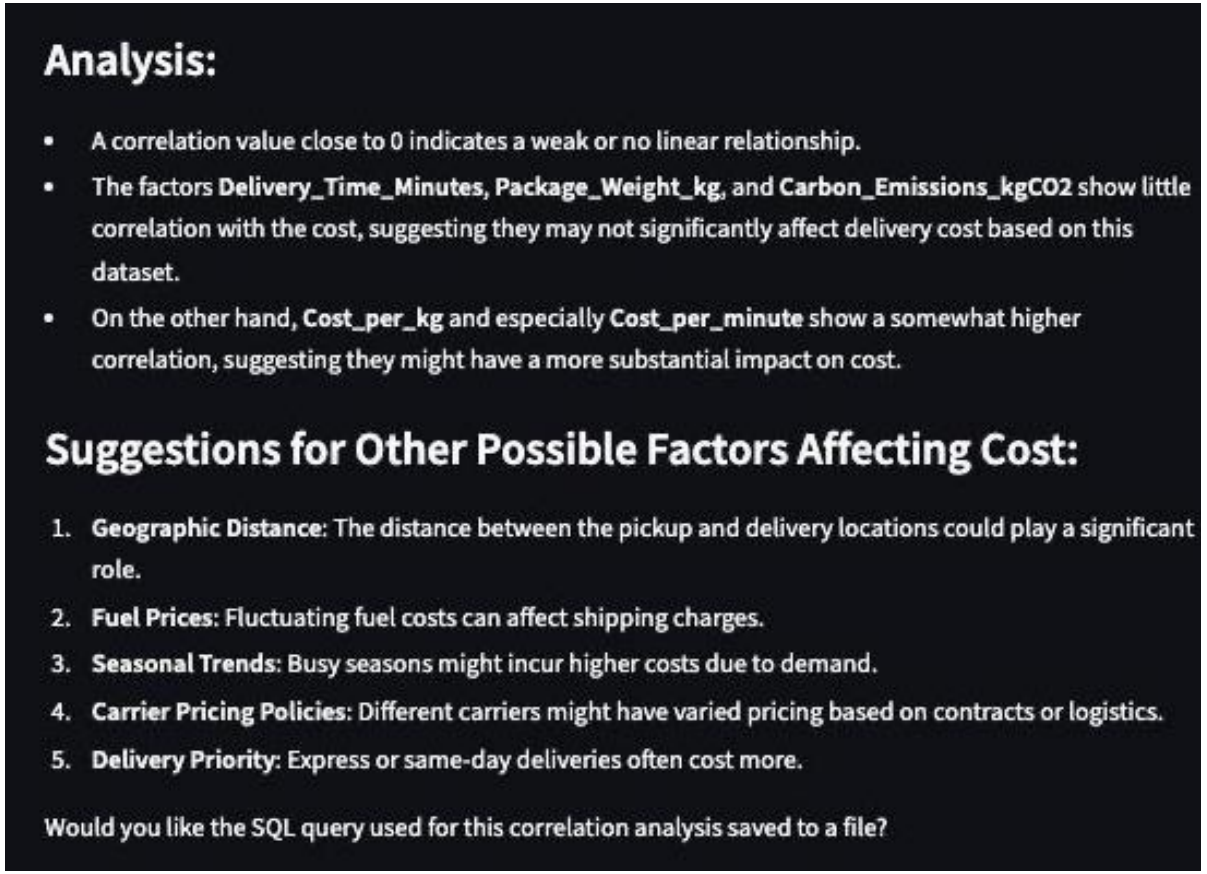
- User inputs a question → LLM converts it into SQL → DuckDB executes the query.
- AI interprets and presents results in a structured format.

**Scalability & Efficiency**

- Works with large CSV and Excel files with minimal computational overhead.
- Can be extended with more advanced AI-powered insights in the future.

# Insights Using Our Personalized Chatbot Agent




Our AI model independently confirmed that **cost is not significantly correlated with package weight, delivery time, or carbon emissions**. Instead, it identified **business-driven factors** like **geographic distance, seasonal demand, fuel prices, and delivery priority**, aligning with our manual analysis. This reinforces the **accuracy and reliability** of our findings.



UCONN SCHOOL OF BUSINESS

# The Bigger Picture – Understanding Cost Beyond the Data

**Potential Cost Optimization Factors (Reducing Operational Expenses & Increasing Profitability)**

- **Delivery Demand & Market Conditions**
  Peak seasons (holidays, sales events) drive higher labor and fuel costs due to increased demand.
- **Distance & Geographic Coverage**
  Longer routes, urban vs. rural delivery challenges, and cross-border logistics increase per-delivery cost.
- **Vehicle Type & Fleet Efficiency**
  Investing in fuel-efficient or electric vehicles reduces long-term fuel costs while improving sustainability.
- **Warehousing & Inventory Management**
  Decentralized warehousing allows for shorter delivery distances, reducing both cost and emissions.
- **Labor Requirements & Workforce Management**
  Hiring surges during peak seasons and overtime pay significantly impact total delivery costs.
- **Package Type & Handling Complexity**
  Fragile, oversized, or perishable goods require specialized handling, refrigeration, or custom packaging, which raises costs.
- **Technology & Automation in Logistics**
  AI-powered route planning and automated sorting hubs can cut costs by reducing inefficiencies.

# The Bigger Picture – Understanding Emissions Beyond the Data

**Carbon Emission Reduction Factors (Enhancing Sustainability & Environmental Compliance)**

- **Government Regulations & Sustainability Policies**
  Carbon tax policies, emission caps, and green incentives impact delivery operations and costs.
- **Route Optimization & Traffic Management**
  AI-powered logistics can reduce fuel consumption and idle time, lowering carbon emissions.
- **Seasonality & Weather Conditions**
  Extreme weather delays increase fuel consumption, while off-season deliveries may allow for eco-friendly slower shipping options.

**Key Takeaways from Our Analysis**

- No strong correlation was found between cost, delivery time, package weight, or carbon emissions.
- This suggests that other operational, market, and strategic business factors are influencing cost and emissions.
- Understanding these hidden drivers is crucial for NovaSphere Logistics to optimize pricing, efficiency, and sustainability.

# Thank You



UCONN
SCHOOL OF BUSINESS