ANALYSIS REPORT

Transition to in-house delivery

NOVEMBER 2024

TEAM FLY

Beata Faitli, Chiara Di Fiore, Guillermo Ysusi, Jasmine Pank, Liliana Mardirossian, Youpeng Sun

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report evaluates the transition from an outsourced logistics model to an in-house delivery system, focusing on financial, operational, and strategic aspects. Key objectives include reducing costs, leveraging tax incentives, and optimising route planning to enhance efficiency.

An in-house model can deliver an 8.8% cost saving within a 75-mile radius, supported by electric vehicle (EV) adoption and government grants. While the 20% cost-saving target remains unmet, the model offers significant non-financial benefits, such as greater operational control, predictable costs, and improved customer service through consistent next-day delivery. Tax incentives like the First Year Allowance and fuel duty exemptions further enhance its financial viability.

The proposed fleet meets next-day delivery demands effectively, even during peak periods. EV adoption aligns with sustainability goals, boosting environmental branding and cutting carbon emissions. Trial of route planning software is recommended to further maximise fleet efficiency, along with setting KPIs.

Beyond a 75-mile radius, in-house delivery is less cost-effective, prompting consideration of alternative growth strategies. Long-term opportunities included regional expansions.

REPORT

Background/Context

- The objective of this analysis is to determine whether transitioning to an in-house delivery model is commercially and operationally viable compared to the current outsourced logistics services.
- Key analytic questions:
	- What are the comparative costs and operational efficiencies of outsourced versus in-house logistics?
	- What combination of vehicles will optimise in-house delivery efficiency while minimising operational costs and staffing needs?
	- What route planning strategies can further reduce costs and environmental impact?
	- What would be the associated capital investment and potential tax savings?^{[1](#page-3-2)}

Project Development Process

Data Cleaning Process

The data cleaning process involved steps for checking missing values, duplicates and addressing incomplete data issues. Ways to merge the three sheets were evaluated. Unnecessary columns were removed from the DataFrame, and columns were renamed to ensure consistency.

● **Missing Values:** missing values were handled with targeted approaches. Missing despatch dates were filled with random values based on delivery proportions, while missing delivery costs were excluded and weights calculated. A random forest model was used to address the high volume of missing values in 'CubicMPerUnit' to preserve data variance.

 1 The 5 Whys and Fishbone diagram (see Appendix 1). SWOT, PESTLE and competitor (see Appendix 2).

- **Incomplete Postcode Data:** to address incomplete postcode data, we generated a new DataFrame from Ordnance Survey Code-Point Open data and validated the provided Postal Sectors against this official dataset. This allowed us to calculate each postal sector centroid's geographical coordinates.
- **DataFrame Merge**: The order detail variables were aggregated based on their OrderID. This approach ensured that the merging process did not generate duplicates. After evaluating various merging methods, a left join was selected to maintain flexibility for future analysis. This choice was particularly beneficial for preserving data with unmatched rows that could have become relevant later.

Exploratory Data Analysis

- Initial data exploration involved descriptive and basic visualisations. Outlier analysis was conducted using boxplots and descriptive statistics. However, following a thorough examination of the data and discussions with Altura, many identified outliers were determined to be valid and were subsequently included in the final analysis.
- Excluded data points: weekend dispatch days, cancelled/ deleted orders, samples, direct orders dispatched by suppliers directly and orders above 15 pallets. This resulted in 5% of data points unsuitable for route planning. However, the costs of these points were manually added to the final monthly costs.
- A cumulative analysis revealed that only 18.5% of orders fell within a 50-mile radius, prompting an extension to 75 miles, covering 33.7% of orders. Beyond this, our analysis highlighted that returns diminish. See our long-term recommendations below for potential strategic in-house delivery expansion options.
- See [Appendix 3](#page-25-0) for detailed insights into the data cleaning and analysis.

Data Visualisation

Visualisations were designed to convey insights to technical and business stakeholders, balancing functionality, accessibility and aesthetics.

Types of Visualisations

● *Bar graphs:* used for categorical data to highlight magnitudes.

Line graphs: Depict trends, such as the cumulative percentage of orders by distance.

- *Maps:* Intuitively display geographic data such as delivery routes and charging station locations.
- Tables: Present multiple data fields concisely where visualisation isn't effective.

Colour Scheme

- Green, aligned with Altura branding, is the primary colour, paired with purple for contrast and readability.
- A green-purple gradient minimises visual clutter, except in maps where distinct colours are required to easily identify categories.
- Grey replaces purple in dual-category bar graphs to emphasise green, highlighting key data.

Accessibility

● In the graph below, the colour palette was chosen for suitability with colour blindness, as colour is a key element of the map.

● In other graphs, alternative methods such as varying bar sizes or adding labels ensured colour wasn't the sole differentiator.

Size & Layout

● Horizontal or vertical bars were selected to optimise slide layout in the presentation, with data labels replacing x-axes for clarity.

Technical Overview of The Code

Preferred Tools and Workflow Development:

The primary tool for the analysis was Python, chosen for its robust data processing, modeling, and visualisation capabilities. While R was initially considered for generating DataExplorer reports and exploratory analysis, its integration with Python proved time-consuming. Consequently, R was used only in the early stages, and Python was used exclusively for the main analysis.

Tableau was employed for preliminary data verification and visualisation. It helped ensure data accuracy and provided insights into daily order volumes and delivery locations. Although creating a Tableau dashboard was not within the project scope, these visualisations informed Pythonbased analysis and grounded the team's direction.

A modular approach was adopted to streamline the workflow. Smaller workflows were created for specific processes such as fleet optimisation, geolocation, and competitor analysis. Where appropriate, these were later integrated into the main process workflow for a cohesive analysis. Code blocks that were lengthy to run were kept in the separate supplementary workflow. The team initially used Google Colab for collaborative coding but transitioned to an offline workflow due to execution time issues. Strict version control, daily updates, code block numbering and a shared GitHub repository ensured seamless collaboration and documentation.

Key Analytical Techniques:

- **Fleet-optimisation:** Fleet optimisation minimised operational costs while meeting delivery demands within the **50-mile and 75-mile radii**. The approach used:
	- **Linear Programming (LP):** Optimised continuous variables such as delivery distance and vehicle utilisation.
	- **Mixed-Integer Programming (MIP):** Handled discrete decisions, including fleet composition and vehicle assignment.
	- Key constraints included vehicle capacity, driver shift limits, and delivery timeframes. Seasonal demand fluctuations were accounted for in cost projections. Refer to Appendix 3 and the final presentation for a detailed breakdown of the constraints.
- **Route Optimisation:** To enhance delivery efficiency, clustering techniques grouped orders within the **50- and 75-mile radii**, aligning vehicle capacity with geographic demand. Manual adjustments accounted for major road access, HGV restrictions, and EV charging network to ensure route accuracy.
	- **HGV Routing:** Prioritised highways and motorways to reduce transit times due to local restrictions. Minimised HGV usage overall, focusing on bulky items aligned with their capacity.
	- **EV Charging Analysis:** Infrastructure analysis showed sufficient charging stations for electric VANs but limited support for LGVs and HGVs. No additional adjustments were needed for VAN routes within the radii.

This ensured compliance with road restrictions and operational efficiency while supporting sustainability goals. See **Section 6.5 of the Main Workflow Notebook** for further details.

- **Vehicle loading algorithm:** This algorithm was not integrated into the main workflow due to time constraints and scope requirements.
- **In-House Fleet Expansion:** Stakeholder feedback emphasised optimising operations within the 50- and 75-mile radius, as expanding beyond 75 miles from a single site was deemed cost-inefficient. Long-term strategies suitable for Altura were explored: The

main approach included in the shortened report is **Establishing New Sites**: Using geospatial and heatmap analyses to identify high-demand regions for additional hubs.

● **Competitor Analysis**: The competitor analysis involved data scraping to extract customer sentiment. While not a primary focus of the presentation or report, it offered valuable insights into review trends and areas of dissatisfaction among customers, supporting the evaluation of potential improvements for Altura's operations.

Please refer to Appendix 3. for a detailed description of the coding steps and methods.

Patterns, Trends, and Insights

1. Cost Savings Opportunity with Larger Delivery Radius:

Despite falling short of the 20% savings target, the in-house model delivers significant benefits in cost control, brand visibility, and customer experience.

An 8.8% cost saving is evident within a 75-mile delivery radius.

In-house costs would remain predictable and stable compared to fluctuating third-party rates.

2. Strong Demand for 2-4 Day Delivery and Potential to Offer Free/Low-fee Next Day Delivery

79% of customers opt for a free 2–4-day delivery, while 12% choose next working day delivery, highlighting the importance of fleet readiness for rapid turnaround times.

Orders placed before 16:30 are consistently dispatched the same day, demonstrating efficiency and enabling next-day delivery.

3. Technological Integration in Logistics & Fleet Optimization:

While charging infrastructure for electric vans is well-established, the network for electric LGVs and HGVs needs further development to support broader adoption.

The recommended fleet will include **electric** vans for boxed items and **fuel** LGVs/HGVs for palletised or bulky items. LGVs will handle narrow city routes, while HGVs will focus on motorway-accessible locations.

BP Pulse supplies nearly a third of charging stations, ensuring coverage along key delivery routes. Its monthly subscription lowers charging costs, while 150kW fast-charging boosts efficiency and maximises delivery time.

Vehicle Costs Considerations: One-off purchase costs are £66k for vans, £145k for LGVs, and £230k for HGVs. EVs qualify for a £5,000 grant per van and seven tax savings. Staffing includes 1.5 drivers per vehicle and one manager per 15 drivers. On-site charging stations, costing £20,000 each, can support two vans simultaneously.

Investing in route planning software will be needed to ensure efficiency and minimise costs and operational strain.

Optimised Fleet Strategy for a 50-mile Radius - Dividing into five strategic regions based on clustering and major road access. Recommending 17 vehicles (8 vans, 6 LGVs, 3 HGVs) to efficiently manage demand, including next-day delivery and seasonal surges.

4. Financial Advantages of Tax Incentives:

Although fleet ownership requires significant capital, it provides tax savings^{[2](#page-13-0)}, builds equity, and stabilises costs. A phased 5-year vehicle replacement cycle maximises resale value and ensures continuity. Monitoring EV advancements will support future fleet expansion in line with Altura's growth goals.

 2 Please refer to the appendix for details on the tax savings.

5. Strategic Future Expansion Opportunities & KPIs:

75-Mile Radius Expansion Strategy: Divided into 9 regions, doubles customer reach and capacity, requiring 29 vehicles (16 vans, 7 LGVs, 7 HGVs).

Strategic focus on expanding into high-demand hubs like London and Manchester-Leeds.

Leveraging KPIs such as on-time delivery and cost-per-mile will ensure continuous performance improvement and accountability.

Overall, we recommend starting with a 50-mile radius for in-house delivery and shortly after expanding to a 75-mile radius to further enhance savings while providing a more sustainable option for your customers and allowing future growth of the company with controlled costs.

APPENDIX

Appendix 1 - Problem Analysis

5 Whys Analysis

Problem Statement: Is it commercially and financially viable to replace the current model of outsourced freighting with third-party carriers and bring this function in-house?

- **1.** Why consider replacing outsourced freighting with in-house delivery?
	- Outsourced freighting could be costly or inefficient, assuming yearly and seasonal increases in third-party logistics (3PL) costs.
	- The company might be paying high fees to 3PL providers or encountering delays and quality issues that impact customer satisfaction. The current high OTIF (Ontime-in-full) score suggests the relatively good performance of the current model.
- **2.** Why are outsourced freighting costs or inefficiencies a problem?
	- They may be affecting profitability and operational control.
	- Outsourcing comes with fixed rates, additional surcharges, and less flexibility, making it challenging to adjust costs during fluctuating demand. Altura may also need more control over delivery times or customer experiences.
- **3.** Why do profitability and operational control matter in this decision?
	- Lower costs and better control over logistics could improve customer service and overall financial performance.
	- Bringing logistics in-house allows for more flexibility in managing routes, optimising deliveries, and responding to customer needs, improving service levels and lowering overall costs per delivery.
- **4.** Why would in-house delivery offer more flexibility and lower costs?
	- In-house delivery enables better route optimisation, complete control over vehicle management, and the ability to use cost-efficient solutions (e.g., specific vehicle types and optimised routes).

- Altura could leverage route-planning software, dynamic scheduling, and more efficient fuel use. Additionally, fixed costs could be amortised (spread) over time with owned vehicles, leading to potential cost savings.
- Altura could claim tax benefits on investments such as purchasing vehicles, equipment, or other capital assets through Capital Allowances (E.g., Annual investment allowance, writing down allowance, first-year allowance, enhanced capital allowance for low-emission, energy-efficient, or electric vehicles). Fleet management systems, warehousing, and equipment required for delivery operations may also qualify for tax benefits.
- In comparison, the cost of hiring 3PL providers is usually classified as operating or business expenses. These costs are fully deductible from profits before calculating corporate taxes. While Altura will not get capital allowances, they can reduce their taxable profits by deducting the total amount of third-party logistics fees as part of their operating costs. If the company is VAT-registered, it may be able to reclaim the VAT paid on third-party logistics services, depending on the nature of the transactions.
- **5.** Why would these improvements lead to commercial and financial viability?
	- If the initial capital investment and operational costs for an in-house logistics system are lower than ongoing outsourced fees, the in-house model would be more financially viable in the long term.
	- Through better cost control, flexibility, and potential improvements in customer satisfaction, Altura could reduce overall logistics costs and improve profit margins, making the switch to in-house delivery commercially and financially sound.

Refined Problem Statement: (remains similar to original problem statement)

Is it commercially and financially viable for Altura Packaging to transition from its current model of outsourced freighting to an in-house logistics operation?

The 5 Whys analysis shows that the core issues driving the decision are **costs, control, and flexibility**. Supposing the company can reduce costs and improve operations by managing logistics internally (e.g., through fleet optimisation, technology, and strategic vehicle use), bringing the function in-house may be financially viable. However, a thorough cost-benefit analysis of initial

investment and ongoing expenses (vehicles, staffing, fuel, maintenance) is essential to confirm this viability.

Fishbone Problem Analysis

Problem Statement: Is it commercially and financially viable to replace the current model of outsourced freighting with third-party carriers and bring this function in-house?

- **Trend:** Estimated third-party delivery service cost increase of 5-10% per year due to fuel and transportation costs, labour costs and automation investments.
- **Current State:** Current monthly cost of third-party delivery is approximately 750K. Further seasonal fluctuation needs to be accounted.
- **Target Effect**: 20% saving on overall delivery costs. Approximately 150K per month reduction in delivery cost to 600K per month.
- **Six Key Areas of Focus:**
	- **Costs**
		- **Third-party Delivery:**
			- Rising costs Estimated 5-10% annual increase
			- Variable costs estimated 10-20% surcharge during peak periods
			- No Capital Investment
			- Business expenses are deductible from profits before calculating corporate taxes
		- **In-house Delivery:**
			- In-house operation costs could also increase annually for similar reasons as for third-party providers
			- The analysis will explore how new in-house delivery would accommodate peak seasons and fluctuating demands.
			- High Initial Capital Investment analysis to determine exact costs
			- Potential Tax Benefits Annual Investment Allowance, etc
			- Ongoing operational costs Staffing, insurance, fuel, maintenance, etc
	- **Operations**
		- **Third-party Delivery:**

- 99% OTIF (On-Time-in-Full) very high current standards from third-party services
- Less control potential delays and inconsistencies
- Peak-period management
- No scope for scaling
- **In-house delivery:**
	- Analysis to investigate how to maintain or further improve high standards of OTIF
	- Greater control can improve efficiency
	- Customisation route management and scheduling could reduce costs
	- Scalability and expansion potential for end-to-end delivery for customers, potential to expand delivery for other products
- **Technology**
	- **Third-party delivery:**
		- Dependence on third-party technology
		- Further questions for stakeholders include how delivery targets are ensured and what systems or processes are in place to monitor the success of each delivery (e.g., real-time tracking, proof of delivery, or customer feedback).
	- **In-house delivery:**
		- Potential for customisable Tech Integration with existing Altura systems, such as inventory management.
		- Investments in Technology: AI route optimisation, fleet management to improve efficiency
		- Analysis to determine the cost of technological system investments

○ **Staffing**

- **Third-party delivery:**
	- No additional staffing requirements other than order management and follow-up
	- Limited flexibility
- **In-house delivery:**

- Hiring and training costs
- Continued staffing costs
- Greater flexibility

○ **Customer Impact**

- **Third-party delivery:**
	- Reported 99% OTIF suggests high customer satisfaction further investigate positives and negatives on next stakeholder meeting
	- Limited Control Over Service Quality
- **In-house delivery:**
	- Potential for Improved customer experience
	- Consistency
	- Potential for improved customer loyalty
	- Potential for end-to-end customer journey if expanding on offering delivery for existing customers

○ **Sustainability**

- **Third-party delivery:**
	- Limited Sustainability Options Sustainability may not be a thirdparty priority
	- No Direct Control Over Environmental Impact
- **In-house delivery:**
	- Potential to invest in eco-friendly fleet electric or hybrid
	- Green Logistics Strategy route optimisations, eco-friendly delivery strategy

Following the problem analysis, key areas were identified, leading to the development of targeted analytical questions. These questions explore potential solutions, focusing on cost, operational efficiency, sustainability, etc. Additionally, metrics and key performance indicators (KPIs) were proposed to monitor and measure progress effectively, enabling a comprehensive analysis and informed decision-making.

Analytical questions:

.

● **Cost Savings and Financial Viability**

- a. What are the current costs of outsourced freighting (annual fees, per-delivery charges, additional surcharges)?
- b. What are the projected initial capital costs for setting up an in-house logistics operation (vehicles, technology, warehousing, staffing)?
- c. How do the ongoing operational costs of in-house logistics (fuel, maintenance, staffing, insurance) compare to the costs of outsourced freighting?
- d. How would the potential tax savings from capital allowances (Annual Investment Allowance, Writing Down Allowances) impact the overall financial viability of the in-house model?
- e. At what point (break-even analysis) would the initial investment in an in-house fleet be recouped through operational cost savings?
- f. Potential metrics:
	- i. Cost per Delivery
	- ii. Cost per Mile or Kilometre
	- iii. Total Logistics Costs
	- iv. Return on Investment (ROI)
	- v. Tax Savings from Capital Allowances

Operational Efficiency

- a. How would route optimization and fleet management improve delivery efficiency compared to the current outsourced model (e.g., time saved, distance reduced, cost per delivery)?
- b. What impact would in-house logistics have on delivery time and reliability, and how would this compare to the performance of 3PL providers?
- c. What is the utilisation rate of vehicles and fleet capacity under an in-house model compared to outsourced freighting? To consider fluctuations of in demand-periods, vs quiet periods
- d. Potential metrics:
	- i. On-Time Delivery Rate
	- ii. Average Delivery Time
	- iii. Fleet Utilisation Rate
	- iv. Fuel Efficiency (MPG or L/100KM)
- **Staffing requirements**

- a. How many drivers, logistics coordinators, and support staff would be required to operate the in-house fleet?
- b. What are the projected staffing costs (salaries, training, and benefits) for running in-house logistics?
- c. How would staffing requirements fluctuate with seasonal demand or growth in business volume?
- d. Potential Metrics:
	- i. Driver Utilisation Rate
	- ii. Labor Costs per Delivery
	- iii. Training Time and Costs

Customer service impact

- a. How would in-house logistics affect customer satisfaction in terms of delivery times, flexibility, and service reliability?
- b. What is the current level of customer dissatisfaction related to outsourced deliveries (e.g., late shipments, order tracking issues)?
- c. Can in-house logistics offer value-added services such as more flexible delivery times or real-time tracking that could enhance customer experience?
- d. Potential metrics:
	- i. Customer Satisfaction Score (CSAT)
	- ii. Delivery issue rate (late delivery/wrong item/damaged goods)
	- iii. First Attempt Delivery Success Rate

● **Operational flexibility**

- a. How much more flexible would in-house logistics be compared to outsourced freighting in terms of scaling up or down during demand fluctuations?
- b. What operational risks (e.g., vehicle breakdowns, staffing shortages) might arise from bringing logistics in-house, and how would they be mitigated?
- c. Potential metrics:
	- i. Capacity Utilisation during off-peak/peak times
	- ii. Scalability metric response to fluctuations
	- iii. Risk management track downtime/breakdowns/maintenance of fleet
- **Sustainability goals**

- a. What is the environmental impact of switching to in-house logistics (e.g., fuel consumption, emissions), and how does this compare to third-party providers?
- b. Can the company adopt more sustainable practices, such as using electric or fuel-efficient vehicles, and what are the potential cost savings or additional investments required for these practices?
- c. How would this shift contribute to Altura's long-term sustainability targets (e.g., carbon footprint reduction)?
- d. Potential metrics:
	- i. Carbon Emissions per Delivery
	- ii. Fuel Consumption per Mile/Kilometre
	- iii. Percentage of Electric/Hybrid Vehicles

Stakeholder analysis

As part of the problem analysis, a detailed overview of key stakeholders was conducted to identify their roles and influence. This is to guide communication and management efforts throughout the project, ensuring that stakeholder needs and expectations are consistently addressed. Refer to the image below for the visual representation of the stakeholder analysis.

Appendix 2 - Market Research

Comprehensive SWOT, PESTLE, and Competitor Analyses were conducted as part of the project. These sections have been excluded from this version of the report to ensure client confidentiality is maintained.

Appendix 3 - Detailed Coding and Analysis Steps

Overall Analysis Approach

- **Preferred Tools**:
	- The primary tool used for the analysis was **Python**, as it provided robust support for data processing, modelling, and visualisation. While R was initially considered as a complementary tool, integrating R within the Python environment proved time-consuming and impractical across team resources. Consequently, the team decided to use Python exclusively for the main analysis.
	- **Use of R** R was utilised in the early stages of the project for specific tasks: Generating DataExplorer reports for the initial and merged datasets and conducting basic exploratory analysis. However, sections requiring R were either limited or commented out, as running R code without extensive installations posed challenges.
	- **Use of Tableau:** After data cleaning, the final CSV files were loaded into Tableau for preliminary analysis. While creating a Tableau dashboard was not within the project scope as discussed with Altura. By leveraging Tableau for data verification and initial exploratory visualisations, the team ensured that the dataset was ready for deeper analysis while maintaining alignment with Altura's project goals. The tool was utilised for the following purposes:
		- **Data Verification:** Tableau was used to verify the accuracy and consistency of the cleaned data. Daily order volumes and delivery locations were visualised to ensure alignment with expectations.
		- **Insights and Visualization:** Provided brief insights into overall order distribution and the consistency of delivery locations across different days. It helped the team to understand that the orders are spread highly in terms of geography and there is no significant change of customer locations over time. These visualisations helped ground the team's approach and informed key directions for the Python-based analysis.
		- Limited Role: While Tableau offered valuable initial insights, it was not used for further analysis or incorporated into the final submission. The primary analysis and visualisations were completed using Python.

- **Workflow Development:** To ensure efficiency and consistency, the team adopted a **modular approach** - creating a main workflow for integrated results and small separate workflows for individual processes. At the final stage, these smaller notebooks were combined into a cohesive submission workflow for assignment purposes. This modular approach enabled efficient collaboration and allowed different team members to focus on their areas of expertise, ensuring a streamlined and organised reporting structure.
	- A **main process workflow** was created to integrate key insights and adapt code blocks from various smaller notebooks. Beata Faitli took a lead to integrate insights from costing and smaller workflows into the main workflow process.
		- Initially, the team utilised **Google Colab** for collaborative coding and development. While Colab provided a convenient, shared environment for team collaboration, execution times grew lengthy as the code base expanded. This led to delays and inefficiencies, particularly for large sections of code.
		- **Transition to Offline Workflow:** To address these challenges, the team transitioned to an offline workflow.
		- **Version Control:** Strict daily version control was implemented to track progress and prevent conflicts. Each team member worked on assigned sections of the code in offline environments.
		- **Code Organization:** Code blocks were numbered by sections to clearly delineate tasks and maintain consistency. A Google Sheet document was used to track active sections and assigned tasks, ensuring visibility across the team. At the end of each day, team members merged their work into a unified notebook. Saved the updated version in a shared folde**r** for easy access by all members. Backed up the notebook on a private GitHub repository to ensure the integrity and safety of the project files.
		- **Outcome:** This structured approach allowed the team to: Maintain efficient collaboration despite working offline. Track progress systematically and avoid duplicating efforts. Save time by merging completed sections into a cohesive notebook daily. This process ensured

that the team could work efficiently on a growing codebase while maintaining clear documentation and progress tracking.

- The **smaller workflows** included:
	- **Fleet Optimization Analysis:** Independent analysis by Youpeng Sun, whose work on fleet optimization and postcode handling was integrated into the main process workflow.
	- **Postcode Cleaning and Geocoding:** A separate workflow developed by Guillermo Ysusi was developed to clean postcodes and acquire coordinates using Ordnance Survey open-source data. This step was necessary due to the partial postcodes provided but would not be required by Altura in future iterations.
	- **Competitor and Sentiment Analysis:** Performed as a separate workflow by Chiara Di Fiore

List of data files provided for the main process workflow notebook:

- **1. TEAM_FLY_LSE_EP_Assignment3_Altura_orders_data_original.xlsx**
	- **a.** Description: The original dataset provided by Altura, containing order details.
	- **b.** Note: This file is password-protected and required for initial data exploration and analysis.

2. **TEAM_FLY_LSE_EP_Assignment3_co_ordinates.csv**

- **a. Description:** A dataset containing derived latitude and longitude coordinates for customer locations.
- **b. Source:** Created using Ordnance Survey Open Data for geocoding partial postcodes.
- 3. **TEAM_FLY_LSE_EP_Assignment3_bulky_item_list.csv**
	- **a. Description:** A list of product IDs for items over the 80 kg weight limit that did not trigger warnings for palletized delivery.
	- **b. Source:** Extracted from Altura's website.
- **4. TEAM_FLY_LSE_EP_Assignment3_PostcodeArea_list.csv**
	- **a. Description:** A reference file listing UK postcode areas.
	- **b. Source:** Derived from Wikipedia

https://en.wikipedia.org/wiki/List_of_postcode_areas_in_the_United_Kingdom

- **5. TEAM_FLY_LSE_EP_Assignment3_Altura_annual_financial_statement.csv**
	- **a. Description:** Altura's financial data, sourced directly from their website.
	- **b. Use:** Provides insights into annual turnover, operational expenses, and financial trends.

List of data files provided for the secondary process notebook:

1. TEAM_FLY_LSE_EP_Assignment3_Complete_Data_Final.csv

- **a. Description:** This is the merged and wrangled complete data set.
- **b. Source:** This is generated by the main process workflow.
- **c. Use:** We use this for mapping to ensure the information matches.

2. TEAM_FLY_LSE_EP_Assignment3_Customers_UpperPostcodes.csv

- **a. Description:** This is the customers' data set where the provided postcodes had been previously transformed into upper letters.
- **b. Use:** This was used for validation against the Ordnance Survey data.

3. TEAM_FLY_LSE_EP_Assignment3_Eastings_Northings_Unique.csv

- **a. Description:** This includes all the postal sectors in the Ordnance Survey data with their corresponding Eastings and Northings codes.
- **b. Source:** This file is produced by the Secondary Notebook.
- **c. Use:** As it takes too long to run, the cell was deactivated and the resulting file is provided.

4. TEAM_FLY_LSE_EP_Assignment3_National_ChargePoint_Registry.csv

- **a. Description:** This includes all the information for each EV charge point provided by the National Charge Point Registry.
- **b. Source:** This was downloaded from the gov.uk website.

5. TEAM_FLY_LSE_EP_Assignment3_Altura_reviews_12_months_20241031.csv

- **a. Description:** This includes Altura's reviews on Trust Pilot related to deliveries.
- **b. Source:** This was scraped from Trust Pilot.

6. TEAM_FLY_LSE_EP_Assignment3_competitor_reviews_12_months_20241031.csv

- **a. Description:** This includes Altura's competitor reviews on Trust Pilot related to deliveries.
- **b. Source:** This was scraped from Trust Pilot.

List of Jupyter notebooks provided for the submission:

- **TEAM_FLY_LSE_EP_Assignment3_main_process_workflow.ipynb**
	- a. Description: This is the main notebook used for end-to-end analysis. It integrates results and insights from all other areas of work, including Costing analysis, Postcode analysis, Fleet optimization. This notebook Provides a consolidated view of the analysis and serves as the primary workflow for submission.

● **TEAM_FLY_LSE_EP_Assignment3_Youpeng_Sun_fleet_analysis.ipynb**

- a. **Description:** Independent analysis carried out by Youpeng Sun.
- b. Use of the postcode area list was incorporated into the main workflow.
- c. **Key notes**: Incorporates a vehicle loading optimization algorithm that Altura may wish to explore further. This algorithm was not integrated into the main workflow due to time constraints and scope requirements.

● **TEAM_FLY_LSE_EP_Assignment3_Secondary_Notebook.ipynb**

- a. **Description:** This notebook integrates all side workflows into one submissionready file. Includes: Postcode validation, Interactive Mapping, EV Charging Networks Evaluation, and Sentiment Analysis.
- b. **Key Notes:** These workflows were kept separate from the main workflow to reduce execution time and simplify the loading of the main notebook.

Key Analysis Steps

1.Key libraries used during the analysis (Refer to section 2. Of the Main workflow workbook)

- The project utilised a range of Python libraries, grouped into categories based on their functionality. Below are the major libraries and their primary purposes:
	- **Data Manipulation and Analysis**
		- **Pandas:** For handling and processing structured data, including data cleaning, transformation, and analysis.
		- **NumPy:** For numerical computations and handling multi-dimensional arrays.
	- **Visualisation libraries:**
		- **Matplotlib:** For creating static visualisations like scatter plots, bar charts, and histograms.
		- **Seaborn:** For advanced statistical plotting and enhancing the readability of visualisations.
		- **WordCloud:** For generating word clouds to visualise frequently mentioned products or terms.
	- **Statistical Analysis and Modeling**

- **Scipy:** For statistical computations, including hypothesis testing and data distribution analysis.
- **Sklearn:** Used for machine learning tasks like: Random Forest Regression to handle missing data. K-Means clustering for customer segmentation. Model evaluation metrics (e.g., RMSE, R²).

○ **Mapping and Geospatial Analysis**

- **Geopy:** For calculating distances between geographical coordinates.
- **Shapely:** For handling geometric objects, such as polygons, used in geospatial analyses.
- **Geopandas:** For geospatial data manipulation and analysis.
- **Folium:** For interactive maps, clustering points, and heatmap visualisations.
- **Fleet Optimization**
	- **Pulp:** For implementing Linear Programming (LP) and Mixed-Integer Programming (MIP) to optimise fleet composition and routing.
- **File Handling and Utilities**
	- **msoffcrypto:** For handling password-protected Excel files.
	- **OS, SYS, Logging, Subprocess, Warnings:** For system-level operations and logging progress

This combination of libraries allowed for robust data manipulation, statistical analysis, machine learning, geospatial processing, and optimization modeling, ensuring the analysis was comprehensive and well-supported by industry-standard tools.

2. Data Cleaning and Validation (Section 3. of the Main process workflow workbook)

● **Initial Dataset Overview:**

- The dataset included 32,063 orders, 23,237 customers, and 76,573 detailed orders.
- \circ Basic exploratory analysis was conducted using Python functions like .info(), .describe(), .isna(), and R's DataExplorer to validate data integrity and identify missing values or anomalies.
- **Handling Missing Values:**
	- **Key Columns:**
		- Missing **DeliveryCost** values (2.87%) were primarily from deleted or cancelled orders, which were saved separately for review.
		- LineWeight was calculated as PackWeight * Packs to address missing entries.

- For **CubicMPerUnit**, a **Random Forest model** was used to predict and fill 4,016 missing values, chosen for its high accuracy given the volume of missing data.
- **Delivery Dates:**
	- Missing dispatch dates were imputed using random values based on observed proportions of same-day, next-day, and within 7 days.
- **Postcode Data Cleaning and Geolocation:**
	- Utilised **Ordnance Survey Code-Point Open Data** to validate and enhance geolocation accuracy: (Refer to supplementary workflow workbook)
		- Corrected formatting errors and matched postcodes incrementally.
		- Missing values were cleaned, leaving 170 unmatched rows, which were exported for review.
		- Appended Easting and Northing data, later converted to latitude and longitude using pyproj.

3. Data Integration, Transformation and enriching the data (Section 3. of the Main process workflow workbook)

- **Merging Data:**
	- Initially merged three datasets into a **76,000-row DataFrame**, but this complexity led to challenges in cost analysis.
	- Opted to **aggregate order details** into a summarised format, simplifying the dataset for analysis while maintaining key attributes.
	- Confirmed that all aggregated orders assumed single shipments for initial modeling, with plans to validate this with Altura.
- **Generating csv files:**
	- To maintain transparency and consistency, the workflow saved CSV files at every stage where data was excluded or transformed:
	- Excluded data (e.g., incomplete orders, unmatched postcodes) was saved for potential further exploration by Altura.
	- Each step of the analysis generated intermediate CSV files to document changes and track the evolution of the dataset.
- **Enriching data:**
	- Data enrichment included incorporating supplementary files for enhanced analysis:
		- **Postcode data** from Ordnance Survey for geolocation.
		- Postcode Area codes and names from Wikipedia resource
		- **National EV charging points** for sustainability and fleet planning insights. (used in the supplementary workflow)
		- **Altura's financial data** for cost analysis.
		- **Unpalletized bulky items** to refine product classifications.
		- Costing analysis results incorporated into the fleet optimisation

- **Straight-line distance** from Altura's eCommerce location was calculated for use in delivery distance analysis, adding a valuable metric for route optimization
- As part of the data enrichment process, the team explored various approaches to address incomplete postcodes. One method involved creating dummy postcodes by randomly generating letters to complete missing entries. However, this approach proved ineffective due to the inherent complexity and structure of UK postcodes, which require specific patterns for accuracy and validation. As a result, this method was excluded from the final submission to maintain data integrity and focus on more reliable geolocation techniques, such as utilising the Ordnance Survey dataset for validation and filling gaps**.**

4. **Exploratory Data Analysis** *(Section 4. of the Main process workflow workbook)*

This section focuses on exploring the dataset to derive insights into delivery costs, customer behaviour, and operational metrics. Key subtopics include:

- Descriptive statistics to analyse delivery costs, product values, and customer delivery charges.
- Exploration of dispatch timelines, order packaging dimensions, weights, and palletization.
- Examination of relationships between variables using correlation and scatter plots.
- Identification of high-cost deliveries segmented by courier type and delivery profiles of customers based on their geographical distribution and preferences.
- Trend analysis of orders over time to uncover seasonality and growth patterns.
- Visual summaries of product-related data: word clouds and other methods to visualise the most common products.

5. Seasonality and turnover projections: *(Section 5. of the Main process workflow workbook)*

- This section addresses the financial implications of delivery costs and revenue projections while accounting for seasonal trends. As explained by Altura during our meetings, Altura experiences a seasonal increase of 20% in busy months, from September to January.
- **Turnover Estimation**: Estimation of annual turnover based on order volumes, product values, and delivery charges.

6. Fleet Optimization and Cost Analysis: *(Section 6. of the Main process workflow workbook)*

- The workflow integrates analysis results presented during both the **initial pitch** and the **final presentation**. In the initial pitch, two distinct fleet models were proposed: one based on traditional fuel and another incorporating electric vehicles (EVs).
- For the final presentation, the approach was updated based on **stakeholder feedback**. This included a deeper investigation into the **EV charging infrastructure** (detailed in the

supplementary analysis Jupyter notebook and presentation slides). The projections were revised from long-term 15-year estimates to a **practical monthly cost breakdown**, excluding the integration of in-house and third-party elements for handling remaining orders.

- The focus shifted to a like-for-like cost comparison for the 50- and 75-mile radii, while excluding orders beyond these distances, which will remain managed exclusively by third-party providers. This refined approach highlighted that the initial recommendations were likely overly optimistic, potentially overlooking key cost factors due to the complexities of a hybrid in-house and third-party model.
- **Goal:** Minimise operational costs while meeting delivery demands for two key radii: 50 miles and 75 miles. This required balancing vehicle utilisation, delivery schedules, and staffing constraints.
- **Approach:** Linear and Mixed-Integer Programming (LP/MIP) We used Linear Programming (LP) and Mixed-Integer Programming (MIP) models to optimise the fleet effectively:
	- **Linear Programming (LP):**
		- Focused on optimising **continuous variables** such as:
			- Delivery distances to minimise fuel and maintenance costs.
			- Vehicle utilisation to ensure efficient use of the fleet.
		- Provided a cost baseline by minimising operational expenses.
		- **Staffing and Vehicle Costs:** Optimised staffing levels and fleet usage to balance costs.
		- **Seasonal Adjustments:** Accounted for fluctuations in demand throughout the year, ensuring scalability and responsiveness.
	- **Mixed-Integer Programming (MIP):**
		- Designed to handle **discrete decisions** for real-world applications, including:
			- Assigning specific vehicle types (**VANs**, **LGVs**, **HGVs**) to routes based on order requirements.
			- **Fleet Composition:** Determined the required number of vehicles for each radius:
				- 1. **VANs** for boxed goods.
				- 2. **LGVs** for palletized deliveries and box overflow.
				- 3. **HGVs** for bulky items and large palletized orders.
			- **Route Feasibility:** Ensured delivery routes complied with time constraints, maximising efficiency within operational limits.
	- **Key Constraints:** To ensure operational feasibility and compliance, the following constraints were applied:
		- **Vehicle Capacity:** Weight and volume limits were tailored to each vehicle type to meet regulatory and logistical requirements.
		- **Driver Shift Limits:** Delivery schedules adhered to legal driving hours of 9–10 hours/day, ensuring compliance and safety.

- **Delivery Timeframes:** Factored in 5–7 hours for active deliverie**s** after accounting for travel time to and from the hub.
- Delivery time configuration checked if next-day delivery or 2-4 day delivery is more cost-effective

Refer to the following presentation slides for key details on the fleet-optimisation:

Driver per Vehicle Ratio:

- O UK Shift Standards: Drivers in the UK are typically limited to 9 hours per day, with a maximum of 10 hours twice a week (Source - https://www.gov.uk/drivers-hours/eu-rules) This limit means one driver per vehicle may not suffice.
- O Recommended Ratio: Based on these guidelines, our recommendation is 1.5 drivers per vehicle. This takes into account driver distribution across the fleet and factors in holiday and rest days.
- O Rationale: This approach balances operational needs with compliance, ensuring continuous coverage and minimizing disruptions.

• Manager to Driver Ratio:

- O Industry Insight: A typical manager-to-driver ratio in logistics ranges from 1:10 to 1:20. This ratio supports effective oversight of driver performance, regulatory compliance, and operational efficiency.
- O Benchmark: The National Private Truck Council (NPTC) recommends a 1:15 ratio in private fleet operations (Source verification pending due to access restrictions).
- O Our Approach: Based on these insights, we've selected a 1:15 ratio to balance supervision needs and maintain operational effectiveness.

Maximum stops per vehicle

- Maximum shift time: 9 hours (max 10, but only to be utilised for unforeseen delays)
- Break time: 1 hour (incorporating breaktime EV charging for VANs)
- Maximum drive time: 8 hours
- Time per stop = Drive time between stops and off-loading time
- Added Drive Time: For the 75 mile radii we reduced the number of stops due to the increase drive time between stops.
- Route planning software: With the use of the software Altura could obtain more accurate values

Vehicle Type & Fleet Capacity

*Refer to appendix

Cost Minimisation

*maintenance costs are typically lower for EVs **Refer to appendix

CONSIDERATIONS

EV Benefits:

- EV grant: £5k per Van, up to £315k
- 7 types of tax reduction incentives**

Staffing:

- 1.5 drivers per vehicle (£33k salary)
- 1 manager per 15 drivers (£48k) salary)**

On-Site Charging:

- £20k per charging station
- 2 vans per station

7. Analysis of in-house fleet expansion: (Section 6. of the Main process workflow workbook)

- **Refinement Based on Stakeholder Feedback:** For the final presentation, less emphasis was placed on expanding the in-house fleet model to larger radii beyond 75 miles, as this was deemed less practical under current constraints. The focus instead shifted to optimising operations within the 50- and 75-mile radii, while still considering strategic long-term opportunities.
- The notebook explored key approaches to expanding the in-house fleet:
	- **Exploring Potential New Sites:**
		- i. Analysed the feasibility of establishing new regional delivery sites based on demand density, geographical coverage, and operational viability.
		- ii. Used geospatial analysis to identify high-demand regions that could benefit from localised delivery hubs, minimising travel time and costs.

8. Secondary processes notebook

● **Validating the postal sector.** To validate the postal sector, we downloaded the Code-Point Open dataset from the Ordnance Survey and merged the multiple CSV files into a single one. We looked for the provided postal sector in the official dataset. If this was not found, we proceeded to eliminate the spaces and execute a progressive validation. We confirmed the first character of the postal sector existed in the official dataset, then the second, then the third, until there was no coincidence. Then, we added a space after the last coincidence and tried to validate further characters. We kept as many characters as matched the Ordnance Survey data.

For each postal sector, we extracted the Eastings and Northings values, from which we calculated the geographical coordinates.

- **Mapping the orders.** Once the dataset was enriched with the geographical coordinates, we mapped the data to understand the orders' distribution better. We used the folium library to generate interactive maps. Several maps were created, for example direct deliveries or deliveries within a user-established radius for user-defined dates.
- **Analysing the electric charging networks.** To revise the feasibility of electric vehicle charge points and the potential best option for Altura, we used the National Charge Point Registry.

We analysed the networks with the most charge points within a 75-mile radius of the distribution centre. We mapped the six largest networks based on the proportion of charging stations they have within the determined area.

● **Sentiment Analysis** We scraped Trust Pilot reviews related to Altura's deliveries and one of Altura's main competitors.

We cleaned the data by removing punctuation and special characters, tokenising the text, removing stopwords, and applying lemmatization.

With clean delivery-related reviews, we could generate word clouds, check for the most common words, calculate the polarity and sentiment, and analyse the resulting categories.

.

Appendix 4 - Patterns, Trends, and Insights

Tax Reduction Incentives (in Addition to Government Grants and Subsidies)

Upfront Benefits

- First Year Allowance (FYA): is specific to energy-efficient, or environmentally beneficial assets. Investing in low-emission/electric vans and trucks may qualify for 100% first-year capital allowances, allowing Altura to deduct the full purchase cost in the first year, enhancing cash flow benefits. There is no upper limit on this.
- Annual Investment Allowance (AIA): Altura can claim 100% of the cost of qualifying electric commercial vehicles/machinery, up to a $£1$ million. This allows Altura to offset the full cost of electric fleet vehicles against profits in the first year.

Ongoing Benefits

- **Fully electric vehicles are exempt from paying Vehicle Excise Duty (VED), saving around £260 per year per vehicle** (compared to petrol/diesel vehicles) for standard commercial vans and up to **£500 per year per vehicle for lorries.**
- **Electricity used to charge electric vehicles is exempt from the UK's fuel duty**. Diesel/petrol vehicles, in contrast, are subject to fuel duty (currently at **52.95 pence per litre**), significantly reduces the cost of running an electric fleet.
- There are enhanced **tax deductions for the costs associated with battery leasing or replacing EV batteries**, which may further reduce operating costs.
- Electric vehicles are currently exempt from the London Congestion Charge, **a daily congestion charge of £15 per vehicle**.

Employee Benefits

.

● Altura employees who have access to electric company vans (zero emission) for personal use, will not be required to pay Benefit-in-Kind tax.

Route Optimisation Software

- Excluded user-based pricing models as they become costly with team growth.
- Open Door Logistics offers a free, open-source option but does not offer real-time tracking and has a less intuitive interface.
- Zeo Route Planner balances usability, price, and essential features like real-time route optimisation.
- eLogii is better for long-term expansion due to its customisable and scalable plans. Further trials and NDA discussions are needed to determine the best fit for Altura.

Appendix 5 -References

John Ozuysal (2023) 15 KPIs for Logistics Every Logistic Manager Should Measure, <https://www.datapad.io/blog/kpis-for-logistics>

Abby Jenkins (2023) The Essential Logistics KPIs & Metrics You Need to Track [https://www.netsuite.com/portal/resource/articles/inventory-management/logistics-kpis](https://www.netsuite.com/portal/resource/articles/inventory-management/logistics-kpis-metrics.shtml)[metrics.shtml](https://www.netsuite.com/portal/resource/articles/inventory-management/logistics-kpis-metrics.shtml)

ATRI (American Transportation Research Institute) (2023) An Analysis of the Operational Costs of Trucking: 2023 Update [https://truckingresearch.org/wp-content/uploads/2023/06/ATRI-](https://truckingresearch.org/wp-content/uploads/2023/06/ATRI-Operational-Cost-of-Trucking-06-2023.pdf)[Operational-Cost-of-Trucking-06-2023.pdf](https://truckingresearch.org/wp-content/uploads/2023/06/ATRI-Operational-Cost-of-Trucking-06-2023.pdf)

DRfreight (2023) Mastering logistics Benchmarking, A comprehensive guide to drive performance and continuous improvement [https://dfreight.org/blog/logistics-benchmarking-a](https://dfreight.org/blog/logistics-benchmarking-a-comprehensive-guide/)[comprehensive-guide/](https://dfreight.org/blog/logistics-benchmarking-a-comprehensive-guide/)

UK Government - Drivers' hours,<https://www.gov.uk/drivers-hours/eu-rules>

List of Postcode areas in the UK https://en.wikipedia.org/wiki/List_of_postcode_areas_in_the_United_Kingdom

- MathWorks (Mixed-Integer Linear Programming (MILP) Algorithms) -

https://www.mathworks.com/help/optim/ug/mixed-integer-linear-programming-algorithms.html - Gurobi Optimization (Mixed Integer Programming Basics)

https://www.gurobi.com/resources/mixed-integer-programming-mip-a-primer-on-the-basics/

- Springer Nature Link (Linear and Mixed Integer Programming)

https://link.springer.com/chapter/10.1007/978-3-642-55309-7_30

- Aalto University (Mixed Integer Linear Programming $(MILP) -$

https://mycourses.aalto.fi/pluginfile.php/2164112/mod_resource/content/2/PIEO23_3_MILPMO DELS.pdf

- Springer Science (Mixed Integer Linear Programming in Process

Scheduling: Modeling, Algorithms, and Application) - https://lin.engin.umich.edu/wpcontent/uploads/sites/551/2021/10/Mixed-Integer-Linear-Programming-in-Process-Scheduling.pdfs)