By Dr. Mu Mu, Prof. Rashid, Aria Banazadeh, Jatin Arora, Ryan Gichuru and Shazab Hassan

Summary

This report analyses and optimises delivery consolidation for GXO. It combines solutions from multiple team members, highlighting key methodologies and findings. The project aims to reduce transport costs and improve environmental standards.

Project Analysis Report

Analysis and Optimisation for Delivery Consolidation

1. Introduction

Project Overview

This project aims to identify opportunities for consolidating deliveries based on specified criteria, enabling cost savings, time efficiency, and environmental benefits. By joining separate consignments into single deliveries where feasible, the solution seeks to optimise logistics, reduce transport expenses, and lower carbon emissions.

Data Provided

There were 2 datasets provided:

1. sys\_order\_consolidation: this contained main information about each order and it was divided into 2 excel spreadsheets (order\_header and order\_line);

The metadata of this dataset is:

|  |  |
| --- | --- |
| **Order\_header** | |
| **Column Name** | **Description** |
| consignment | A unique identifier for a group of orders that are shipped together as a single load. |
| order\_id | A unique identifier for each individual order placed by a customer. |
| ship\_by\_date | The date by which the order must be shipped to ensure timely delivery. |
| deliver\_by\_date | The date by which the order must be delivered to the customer. |
| order\_date | The date when the order was placed by the customer. |
| ship\_dock | The specific dock or location from which the order will be shipped. |
| name | The name of the customer or recipient of the order. |
| town | The town or city where the order will be delivered. |
| postcode | The postal code of the delivery address. |
| shipment\_group | A grouping identifier used to categorise shipments, possibly based on delivery routes or regions. |
| customer\_id\_slim | A streamlined or simplified identifier for the customer, used for internal processes. |
| customer\_id | A unique identifier for each customer in the system. |
| carrier\_id | The identifier for the carrier responsible for transporting the order. |
| v\_wcs\_status | Status code used in the warehouse control system to indicate the current state of the order. |
| status | General status of the order, such as ‘Complete’, 'Released', 'Allocated', etc. |
| instructions | Special instructions or notes related to the order, such as handling instructions or delivery preferences. |
| is\_vas | A flag indicating whether the order includes Value Added Services (VAS), such as special packaging or additional services. |

|  |  |
| --- | --- |
| **Order\_line** | |
| **Column Name** | **Description** |
| order\_id | A unique identifier for each individual order placed by a customer. |
| customer\_id\_slim | A streamlined or simplified identifier for the customer, used for internal processes. |
| customer\_id | A unique identifier for each customer in the system. |
| sku\_id | A unique identifier for each stock-keeping unit (SKU), representing a specific item. |
| line\_id | A unique identifier for each line item within an order. |
| config\_id | Configuration identifier, possibly indicating specific configurations or variants of products. |
| description | A textual description of the item or product. |
| qty\_ordered | The quantity of the item ordered by the customer. |
| expected\_weight | The expected weight of the ordered quantity. |
| cpl | Cases per layer, indicating how many cases can fit in a single layer on a pallet. |
| lpp | Layers per pallet, indicating how many layers can be stacked on a single pallet. |
| cpp | Cases per pallet, indicating the total number of cases that can fit on a single pallet. |
| pallet\_type | Type of pallet used for shipping the order. |
| max\_support | Maximum supportable weight or quantity for the pallet or shipment configuration. |
| v\_pallets\_dec | A decimal value representing the volume of pallets required for the shipment. |
| full\_pallets | The number of full pallets needed for the shipment. |
| full\_layers | The number of full layers needed for the shipment. |
| loose\_cases | The number of cases that do not fit into full pallets or layers, remaining as loose cases. |
| min\_pallets\_to\_fulfil | The minimum number of pallets required to fulfil the order. |

The usefulness of this dataset was that it allowed us to properly filter and combine the orders into their consignments so that we analyse them as groups instead of individual orders.

1. address\_with\_geo: this one contained the geo location for the delivery of each order.

The metadata of this dataset is:

|  |  |
| --- | --- |
| **Address\_with\_geo** | |
| **Column Name** | **Description** |
| CUSTOMER\_ID | A unique identifier for each customer. |
| POSTCODE | The postal code of the customer's address. |
| HOME | The full address of the customer's home location. |
| PC TRIMMED | A trimmed version of the postal code, possibly used for grouping or matching purposes. |
| HOME TRIMMED | A trimmed version of the home address, possibly used for grouping or matching purposes. |
| COUNTRY | The country where the customer resides. |
| NAME | The name of the customer. |
| ADDRESS1 | The first line of the customer's address. |
| ADDRESS2 | The second line of the customer's address, if applicable. |
| TOWN | The town or city where the customer resides. |
| COUNTY | The county where the customer resides. |
| HOME EASTING | The easting coordinate of the customer's home location (used in mapping and GIS applications). |
| HOME NORTHING | The northing coordinate of the customer's home location (used in mapping and GIS applications). |
| DEST EASTING | The easting coordinate of the destination location (used in mapping and GIS applications). |
| DEST NORTHING | The northing coordinate of the destination location (used in mapping and GIS applications). |
| DISTANCE (METERS) | The distance between the home and destination locations, measured in meters. |
| DISTANCE (KM) | The distance between the home and destination locations, measured in kilometers. |

The value of this dataset lies in its ability to facilitate an alternative method for consolidating consignments. Instead of strictly requiring consignments to be for the same customer, this method utilises geographical proximity, such as consignments being within 30 miles of each other. This approach also considers other location-based factors or drive time between locations, providing more flexibility and efficiency in identifying consolidation opportunities.

1. Analysis Process

Data Preparation

Proper data preparation is essential for ensuring accurate and reliable analysis. The preparation of the data was divided into five sections, each playing a critical role in cleaning and preparing the data for further analysis. Below are the steps we followed, along with the reasoning behind each step and how they contribute to effective data preparation:

1. **Datatype Check:**
   1. **Description:** This step involves checking the datatype (string, integer, datetime…) of every column from both datasets and make sure that they are all of the right types, for example:
      * numerical values must be Integer type
      * names/addresses/etc.. must be String type
      * time and date of delivery values must be of the Datetime type
   2. **Importance:** Ensures that data is stored in a format that allows for appropriate and efficient processing and analysis.
   3. **Impact:** not having the right Datatype for certain columns can lead to **errors** when performing analysis, especially when trying to perform calculations or trying to create insightful visualisations.
   4. **Observations/results:** after performing the Datatype check, all the columns match the right datatype according to the requirements.

**A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated**

Order\_line DataType and Null value check

Geo\_data DataType and Null value check

A screenshot of a computer

Description automatically generated

Order\_header DataType and Null value check

1. **Null Value Check:**
   1. **Description:** This step involves identifying any missing values in the Datasets and decide whether to drop those values or keep them.
   2. **Importance:** If they Null values are not handled it can lead to bias in decision making and results, or it can lead to error during analysis.
   3. **Impact:** Depending on the nature of the Null values in certain columns, the decision of how to handle these values can greatly affect the results of the analysis. For example: Dropping too many Rows of data unnecessarily can decrease the likelihood of finding any consolidation opportunity at the end or retaining rows with insufficient data for analysis can lead to inaccurate or incomplete results.
   4. **Observations/results:** The results of this check are as follows:
      1. **Order\_header:** In this data there are:
         * 45 data points which have null instruction value
         * 7 data points with missing postcode
         * 14 datapoints with missing town information
         * 1 datapoint with missing carrier ID

The decision taken in this scenario are:

* Null Instructions, Postcodes, and Towns:
  + The missing instruction values were deemed redundant and thus ignored.
  + Missing postcodes and town information were found in the ‘Geo\_data’ dataset, so these missing values were also ignored.
* Missing Carrier ID:
  + Since the carrier ID is crucial for the consolidation process and no alternative source for this data was available, the row with the missing carrier ID was dropped.
    1. **Order\_line:** There were no missing values in this data.
    2. **Geo\_data:** In this data there are:
       - 4 data points which have null Address1 value
       - 733 data points which have null Address2 value
       - 39 data points which have null Town value
       - 2451 data points which have null County value

The decision taken in this scenario are:

* All the values which are missing can be acquired from looking at the Easting and Northing values in the same dataset if needed, hence there was further action was taken.

1. **Integrity Check:**
   1. **Description:** This step involves if the snippets datasets that we were give are accurate, reliable and consistent between them. This is done by running various tests(functions) to understand the level of likeness that these datasets have between them. Here is a quick breakdown of those functions:
      1. Integrity check 1: Verify that the number of **unique customers** match between the order datasets
      2. Integrity check 2 and 3: Verify that all the **customer ID**’s in ‘Order\_header’ are the present in ‘Order\_line’ and vice-versa.
      3. Integrity check 4 and 5: Verify that all the **customer ID**’s are in ‘geo\_data’ are all present in ‘Order\_line’ and ‘Order\_header’ datasets.
      4. Integrity check 6: Verify **all order ID**’s in ‘order\_header’ are present in ‘order\_line’
   2. **Importance:** If this step is not carried out, we won’t be able to determine if the datasets are interconnected, specifically if they contain data about the same customers and orders. Without it, we wouldn't be able to confirm the relationships between the datasets, which could result in errors during the consolidation process and potentially lead to zero opportunities being identified.This is also crucial for ensuring that there are no outlier, duplicates or inconsistencies between the data.
   3. I**mpact**: By ensuring that the data are consistent between the dataset we can draw accurate and reliable conclusions. This step helps in identifying and correcting discrepancies that could skew analysis results. The integrity checks will confirm if all relevant data points are present across the datasets, facilitating accurate data consolidation and analysis. This improves the quality of the insights derived from the data, leading to more effective decision-making and operational efficiency. Specifically, it ensures that consolidation opportunities are correctly identified, thus optimizing logistics processes and potentially reducing costs.
   4. **Observations/results:** The integrity tests ran and checked the presence of consignments and orders in the line-item table and customers in the ‘geo\_data’ and vice versa. 5/6 tests came back negative(Only Integrity check 1 was positive) as there were several consignments which did not have any orderliness and many customers’ addresses were not present in ‘geo\_data’. This concludes the result that the data that we were given was a random slice from a larger dataset, hence the rows which failed the integrity tests have little to no effect on the analysis and can be ignored in this instance but when we are running the code in production with full dataset, integrity test should be closely followed, so as to point out any data inconsistency and prevent orders from erroring in later stages.
2. **Dropping Columns:**
   1. **Description:** This step includes removing any columns of data which are unnecessary for our analysis during the consolidation process, or contain any redundant information. For example: After merging the datasets, because the information about the destination is present in multiple of the datasets, it’s reasonable to drop the columns and only keep one version. Similarly, because there both the slim version and the complete version of the Customer\_ID and Order\_ID are present, only keeping either version can be acceptable for our analysis and consolidation process.
   2. **Importance:** Dropping the unnecessary columns can help create more clarity when analysis the data as less columns are present, and the focus will remain on the most important columns for our consolidation process.
   3. **Impact:** By having less columns to conduct the analysis and process to find consolidation will be much less time consuming and require less computational power as there will be less data work with, making it easier to draw meaningful insights.
   4. **Observations/results:** Here is a general overview of the columns that were dropped and the reasoning behind them:
      1. **Duplicates:** Certain columns were present in multiple datasets, for example the information regarding the delivery location were present in both in the ‘Order\_header’ and ‘Geo\_data’, hence the columns in one dataset was dropped.
      2. **Redundant Data:** The columns containing redundant data that will not be relevant to the data analysis or to the consolidation process itself. For example: The description column, the instructions column, etc… .
3. **Conversions:**
   1. **Description:** In this phase, we ensure that all necessary data is in the correct format for analysis. This involves converting data into appropriate units of measurement or formats that are compatible with the analytical techniques being used.
   2. I**mportance:** This step ensures that the data is in the correct format for accurate analysis. Different analytical techniques and tools often require data in specific formats or units. For example, certain analyses might need measurements in metric units, or geospatial analyses might require coordinates in longitude and latitude instead of Easting and Northing.
   3. **Impact**: By converting the data into the right formats, we will ensure that the analysis process later will run smoothly and accurately. This step minimizes any errors that can later on such wrong incompatible data for the technique.
   4. **Observations/results:** The only Columns that need to be converted are the Easting and Northing columns from the **‘**geo\_data**’**. This conversion is necessary because the APIs used to calculate the routing of consignments require location data in longitude and latitude format. Reporting delivery destinations using longitudes and latitudes is recommended since they are a global standard, provide greater precision, and integrate more easily with most routing APIs.

Tools Used

Throughout the project, the programming language used to perform the analysis and implement the algorithms for finding consolidation opportunities was Python, executed within a Jupyter notebook environment. Python was chosen due to its powerful libraries for data analysis, such as Pandas and NumPy, as well as its flexibility in handling different data formats and integration capabilities with various APIs.

Project Structure:

The project adopted a modular structure to ensure organized, maintainable, and reusable code. This approach involved separating different parts of the code into relevant files, each serving a specific purpose. The modular structure enhances clarity and allows for easy updates and debugging.

File Structure

A screenshot of a computer program

Description automatically generatedThe directory structure of the project is as follows:

Diagram of File Structure

Below is a description of the most relevant parts of this structure:

* Data Folder: This folder contains all the dataset.
* Helpers Folder: This folder contains helper.py, a python file that has helper functions that support the main analysis, such as data preprocessing, conversion, analysis and functions for the overall consolidation processes.
* analysis.ipynb file: The Jupyter notebook where the main analysis and algorithms are implemented. This notebook calls functions from helper.py and processes the data.
* requirements.txt: Lists all Python dependencies required to run the project, ensuring that anyone who wants to reproduce the analysis can easily set up the same environment.

Code Documentation

All functions in the helper.py file and the Jupyter notebook are properly documented. This includes descriptions of each function's purpose, the parameters they accept, and the return values and after each process/step is carried out the reasoning for the decisions taken are documented in the jupyter notebook. Proper documentation ensures that the code is understandable and maintainable, facilitating collaboration and future updates.

Visual Analysis

In this section, we provide an overview of the visual analyses conducted, highlighting their purpose and significance. Additionally, we propose further visualisation analyses that we would undertake if given more time to work on this project.

The first few visualisations that were conducted revolve around getting a deeper understanding of the columns, and the patterns linked between them. This entails creating various Bar charts, Histograms, and Scatter Plots, here are some examples:

Distribution Of Full\_Pallets:

Purpose and Rationale: This visualization aims to understand the distribution of the full pallet across different orders. It helps in identifying the average pallet size and understanding the load on each truck.

Significance: Knowing the distribution of full pallets is crucial for logistics planning. It allows us to predict the average load on each truck and manage resources effectively.

A graph of distribution of pallets

Description automatically generated

**Next Steps:**

* Create a more detailed graph showing the distribution of full pallets across the orders.
* Superimpose this data on the orders by date to better understand the load pattern and improve resource allocation.

Distribution of Ship By Date:

Purpose and Rationale: This visualization shows the distribution of shipment dates, helping us understand when most orders are scheduled.

Significance: Identifying peak shipping dates allows for better scheduling and resource management, ensuring that peak times are well-staffed and equipped.

A group of graphs and diagrams

Description automatically generated

Next Steps:

* Analyse the correlation between shipping dates and order volumes.
* Implement predictive analysis to forecast future shipping peaks and optimise staffing and resource allocation.

Count of Statuses in Order Header

Purpose and Rationale: This bar chart visualises the various statuses of orders in the order header dataset, such as "Complete," "In Progress," "Picked," etc.

Significance: Understanding the distribution of order statuses helps in monitoring workflow efficiency and identifying potential bottlenecks in the order processing pipeline.

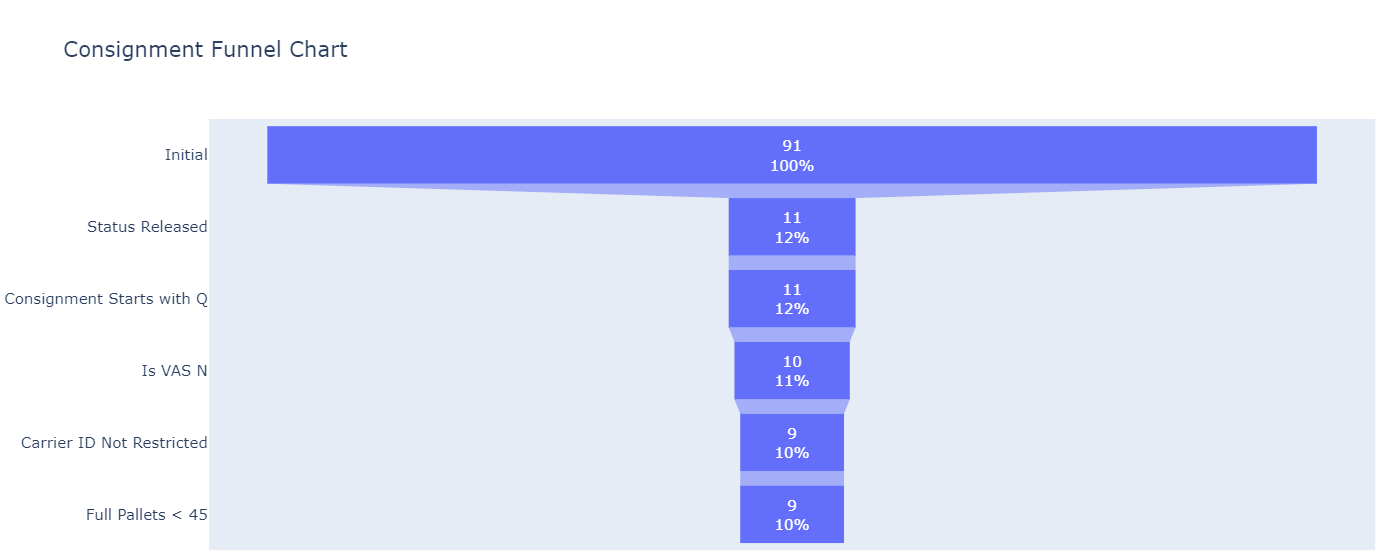
A graph of a bar graph

Description automatically generated

Funnel Chart of Filtered Consignments by Hour

Purpose and Rationale: The funnel chart visualises how the number of total consignments decreases as each filter is applied. This helps in understanding the impact of various constraints on the consignment pool.

Significance: The funnel chart is crucial for identifying the most restrictive filters and understanding their impact on the number of eligible consignments. This helps in fine-tuning filters to balance efficiency and practicality.



## Further Visualisation Analyses

Given more time, the following visualisation analyses would be conducted to further enhance our understanding and optimisation of the logistics process:

Number of Orders/Consignments by Hour

Purpose and Rationale: A visualisation showing the number of orders and consignments coming through the warehouse hourly throughout the day. This helps identify peak hours for order processing.

Significance: This is crucial information for the company as it allows for better workforce management. Understanding peak and low activity hours helps in planning shifts and manpower allocation. For example, during peak hours, the company can offer extra-time to the previous shift workers to handle the increased workload. Conversely, during low activity hours, the company can reduce the number of employees to optimise labour costs.

Next Steps:

* Implement heat maps to visualise the density of orders throughout the day.

Consolidation Opportunities by City

Purpose and Rationale: This visualisation aims to identify cities where there are significant opportunities for consolidating consignments. This helps in understanding regional demand and optimizing delivery routes.

Significance: Identifying cities with high consolidation opportunities allows for strategic planning in logistics, reducing transportation costs, and improving delivery efficiency.

Next Steps:

* Create Bar chart showing consolidation opportunities in each city.
* Analyse the relationship between city consolidation opportunities and delivery times to optimise routes further.

1. Methodology

Criteria for Consolidation

The methodology used for the consignment consolidation process are the ones that were instructed to us in the task presentation. These was divided into 3 main tasks, one which is Mandatory/Current method and the other two were Optional methods:

**Mandatory/Current Method**

This method consists of mainly of simple filtering with lot of requirements. As these requirements are applied to the filtering, the number of potential consignments that can be consolidated drastically drops. Below is the list of all the requirements that this method demands:

* **Status Released:** Consignments must have their status set to 'Released'.
* **Consignment Starts with Q:** The consignment identifier must start with 'Q'.
* **Ship By Date:** The consignment's ship-by date must be more than four hours from the current time.
* **Is VAS N:** The consignment must not be marked as Value-Added Service (VAS).
* **Carrier ID Not Restricted:** The carrier ID must not be in the list of restricted carriers ('SDS', 'DHL', 'CCO', 'MRT').
* **Full Pallets < 45:** The sum of full pallets per consignment must be less than 45.
* **Customer ID Slim:** The customer ID (slim version) must be the same for both load opportunities.
* **Total Full Pallets:** The sum of full pallets on both load opportunities must be less than or equal to 52.

**Optional Methods**

The optional methods offer additional flexibility by modifying the criteria to allow for more potential consolidation opportunities. These methods relax the requirement that consignments must go to the same customer and instead use geographical proximity as a criterion.

* **Optional Method 1: Within 30 Miles**  
  This method allows for the consolidation of consignments that are within a 30-mile radius of each other.
* **Optional Method 2: Within 1 Hour Travel Distance**  
  This method permits consolidation for consignments that are within a 1-hour travel distance from each other.

1. Findings

Current Methods

In the current exploration of the we followed 3 methods of consolidation.

1. Basic Logic Based
2. Distance Based Filtering
3. Time based Filtering

Let’s explore these avenues one by one and provide you some explanation along the way.

## Basic logic based.

The following pyramid represents the funnel like structure that was followed in this approach to filter out the consignments that were fit for approval based on the current resource standard and the capacity.

### Explanation of the logic.

1. In the first stage of filtration, we remove the consignments which were not released (not ready for deployment).
2. The consignments which starts with `Q`. (as given)
3. Where Is\_Vas is `N`.
4. Then we filter by the carriers, removing the other 3rd party carriers.
5. We remove the trucks which are over the limit of 45 full pallets so as not to overload the truck!
6. We then filter if the trucks are going to the same customer.
7. For the consignments which are going to the same customer we check the size of the combinations and remove anyone which has over 52 full pallets.
8. At last, we combine all the datasets to get a holistic view of the opportunities.

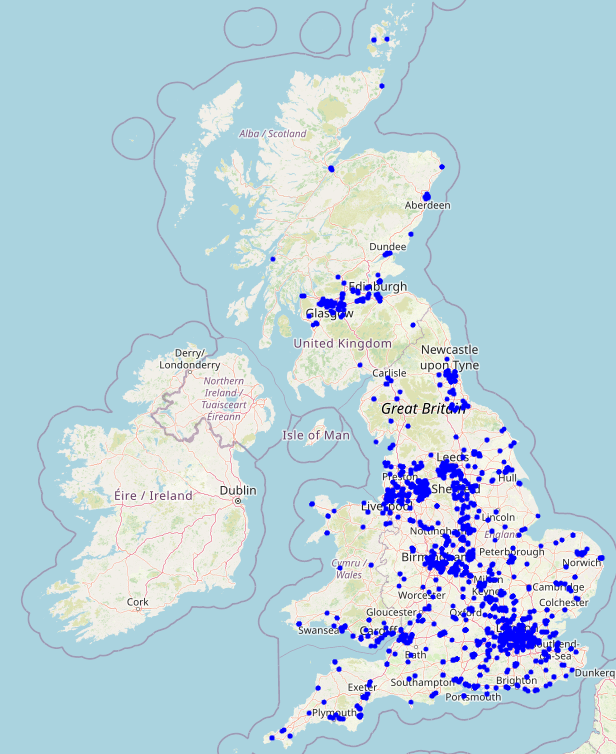
It is important to understand the individual values and the impact of filtering methods at each stage of filtration.

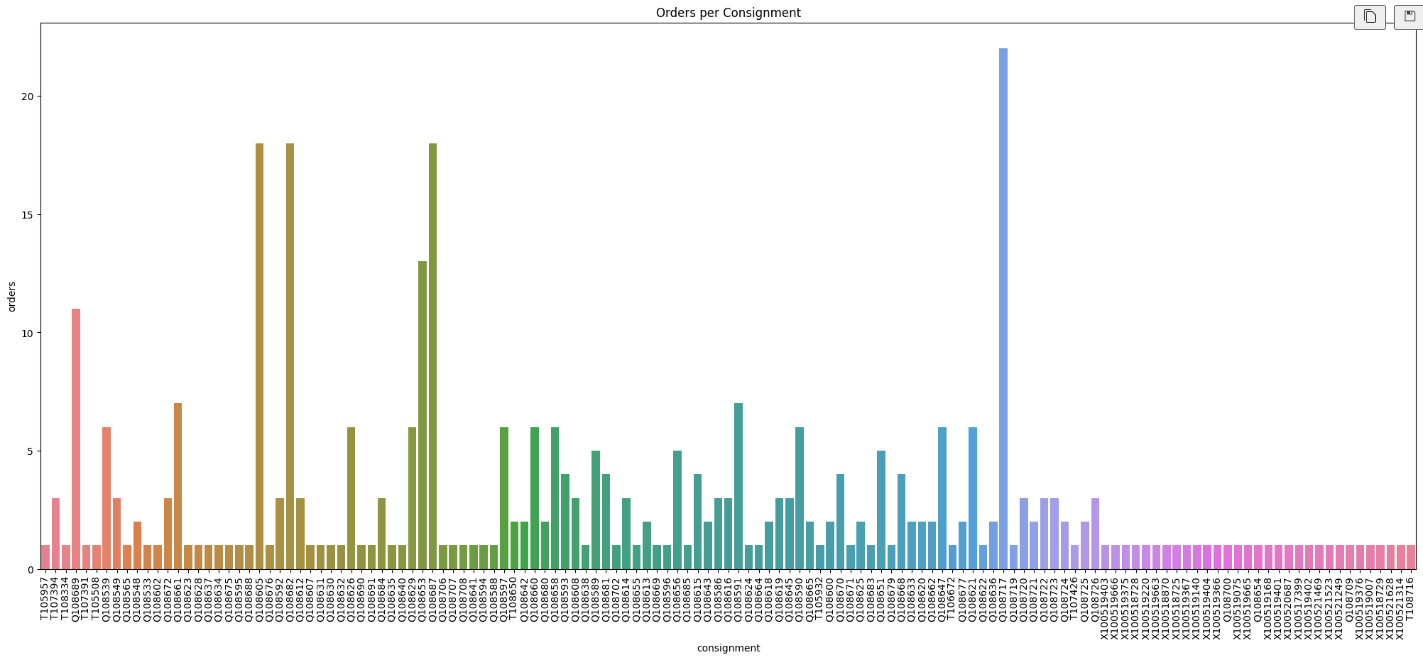
A blue and white rectangular object with text

Description automatically generated

As we can see there is a big drop of numbers in the initial filter of release, which although logical is indicative that the timeframe of different deliveries can make a lot of difference, so there should be a very soft boundary keeping in mind the order fulfilment requirements and it being scheduled at a time where it can be easily consolidated into other containers.

Let us also analyse all these categories one by one to better understand the structure as well as the distribution of the columns in each set of data.

The following is the distribution of customers across the UK. 

One other thing that we should take into condition is order per consignment as this should be able to give us a pretty solid understanding of what an average consignment looks like.  
  


## Distance based consolidations.

In the distance-based consolidation method, we have used the opensource OSRM API.

First, we convert the easting and northing into longitude and latitude.

Out of the above based parameters, we preserve some of the above columns: -

1. For the consignments which have released status.
2. Where consignment ID starts with `Q`.
3. Is Vas is `N`.
4. Filter the 3rd party carriers.

After these basic filtration steps, we follow some steps specific to this approach.

1. Making pairs of orders that are in `30-mile radius` (by route)
   1. We can get this information by using the OSRM API, this API calculates the distance between 2 coordinates based on route rather than straight line distance which leads to better and more qualified consolidation opportunities.
   2. The function for this approach was built with keeping the flexibility in mind, so the parameter for scanning of distance by the algorithm could be changed to give the company opportunity to explore further.
2. The consolidation opportunities are then filtered according to the 52 max pallet condition.

## Time based consolidations.

There is the time-based consolidation we have used OSRM API.

First, we convert the easting and northing into longitude and latitude.

Out of the above based parameters, we preserve some of the above columns: -

1. For the consignments which have released status.
2. Where consignment ID starts with `Q`.
3. Is Vas is `N`.
4. Filter the 3rd party carriers.

After these basic filtration steps, we follow some steps specific to this approach.

1. Making pairs of orders that are in `1 Hour radius` (by time)
   1. We can get this information by using the OSRM API, this API calculates the time based on the time it takes to cover a particular route for example the distance that can be travelled via a Highway will be higher than the distance covered in the city road.
   2. The function for this approach was built with keeping the flexibility in mind, so the parameter for hour radius to be calculated by the algorithm could be changed to give the company opportunity to explore further.
2. The consolidation opportunities are then filtered according to the 52 max pallet condition.

Consolidation Opportunities

### First Method (Logic Based)

The overall findings of this methods unfortunately yielded 0 consolidation opportunities. This is mainly due to the fact we have:

1. Strict consolidation boundaries, with 0 relaxation or situation independent which is not good in long run.
2. We are only looking for consolidation opportunities where the order is going to same customers, hence reducing our search space ridiculously.

### Second Method (Distance Based)

The overall opportunity that was yielded by this model were 0. But we noticed some consolidation opportunities as we relaxed the model parameters.

### Third Method (Time Based)

The overall opportunity that was yielded by this model were unfortunately 0. But similar to the distance based approach, this approach also had some really interesting consolidation opportunities as we further relaxed the model parameters.

Challenges Encountered

The main challenges that were faced in the analysis were as follows: -

1. A somewhat unclear description of the columns so we have to spend a lot of our time to find out the meaning of the columns, in future it will be better to maintain a glossary of columns that is listed in the dataset.
2. The data had a lot of integrity test fails, which means that the data that the data is was a random slice. Hence can lead to a lot of problems in trusting the data and testing rigorously.
3. Due to the data being a limited random slice, there were no consolidation opportunities that were found. This means that we did not have a lot of test cases to properly showcase the power of our modes and we had to improvise by using a relaxed version of parameters that were given to us.
4. One of the main challenges we faced during the analysis was the rigidity of the boundaries and filtering parameters. The strict cutoff of 30 miles or 1 hour for finding consolidation opportunities led to many missed opportunities. It would be beneficial to define parameter standards based on the business season. For example, during busier seasons, the consolidation opportunity radius should be smaller due to the higher frequency of trucks. Conversely, during slower periods, the radius should be extended to capture more opportunities.
5. Evaluation

Successes

The project was an overall success, as we were not only able to comprehensively analyse the requirements of the company, but we also managed to develop a well-defined, modular approach to solve the problem efficiently. Initially, the number of consolidations identified was zero. However, after introducing some relaxation in the metrics, we discovered numerous consolidation opportunities that can significantly streamline operations and enhance productivity.

Our thorough analysis revealed that by adjusting certain parameters, we could unlock potential areas for consolidation that were previously overlooked. These newly identified opportunities have the potential to optimise resource allocation, reduce redundancies, and improve overall operational efficiency. The insights gained from this project will provide the company with actionable strategies to leverage these consolidation opportunities, ultimately driving cost savings and operational improvements.

Furthermore, the modular approach we implemented ensures that the solutions are scalable and adaptable to future changes. This flexibility is crucial for the company as it continues to evolve and grow. The success of this project underscores our ability to not only meet but exceed expectations by delivering innovative and practical solutions tailored to the company's specific needs.

1. Conclusion

Summary of Findings

1. No opportunities were found and no flexibility in the vanilla logic.
2. Although no consolidation opportunities were found in the other logics as well, but we saw interesting patterns after relaxing the bounding parameters especially with the total pallet size, this suggests that either our size of vehicles are not enough for the size of customers that we are dealing with or we will have to adjust the parameters (via a different ML algorithm) to better match the consolidation demand.
3. On the datasets provided when the basic logic is applied the number of consolidations drops to around 10%, the funnel chart shown in previous section better helps to visualise the significant drop once each of the requirements are applied.
4. Two tests were conducted individually on relaxing the boundaries of the requirements which were: Increasing time from 4 hours to 8 hours and increasing full pallets from 52 to 75. The first one did not change anything, on the other hand changing the full pallets size to 75 did result in some opportunities being found which lead to the conclusion that if the size of the trucks can be increased that will certainly help find more opportunities.

Next Steps

A second report has been made which highlights more strategies that have been explored. This report provides an overview of the logistics industry, explaining the importance of effective supply chain management, particularly in order consolidation and logistics routing. It introduces GXO's current methods and specifies which parts of the supply chain these methods are applied to.

The report continues by acknowledging the analysis based on GXO's earlier dataset from this report and reviewing its effectiveness. It includes a literature review of over 50 articles, giving a comprehensive understanding of the industry. The report identifies the best algorithms used by other companies, including traditional methods like heuristics and metaheuristics, big data analytics, as well as AI-based methods such as optimization algorithms, machine learning, and predictive analytics.

Common techniques are defined, their applications and limitations in the logistics industry are reviewed, and various case studies showing the methods in use are examined. The report compares AI-based methods to traditional methods to find the most effective ones, and reviews case studies of some companies to identify the results achieved with the tested algorithms.

It discusses insights from the comparative analysis, identifies the benefits and challenges of AI integration in logistics routing, and outlines the practical implications for GXO. The conclusion summarizes key findings, provides recommendations for GXO based on algorithm performance, and discusses future trends in AI-driven logistics routing and their potential impact on GXO.

1. Appendix