



**ANLY482 - Analytics Practicum**

Project Proposal by Hiryuu

Group 1

Chua Wan Theng

Jouta Lim ZiYu

Lin Qianpin

Table of Contents

[**Project Background** 1](#_Toc471063767)

[**Motivation** 2](#_Toc471063768)

[**Data** 2](#_Toc471063769)

[**Methodology** 4](#_Toc471063770)

[**Introduction** 4](#_Toc471063771)

[**Tools Used** 4](#_Toc471063772)

[**Analysis** 4](#_Toc471063773)

[**1.** **Exploratory Analysis** 5](#_Toc471063774)

[**2.** **Geospatial Analysis** 5](#_Toc471063775)

[**3.** **Clustering** 5](#_Toc471063776)

[**4.** **Time Series Data Analysis** 5](#_Toc471063777)

[**Work Plan** 5](#_Toc471063778)

[**Application** 6](#_Toc471063779)

[**Limitations and Risk** 7](#_Toc471063780)

# **Project Background**

The company been managing the distribution network for one of its major clients, handing both reverse logistics and forward logistics, resulting in both inbound and outbound shipments. Thus far, the lack of a dedicated dashboard to show summary tables regarding the status of the shipments have made it very difficult for the program managers and the sales team to monitor and follow up with requests from the client regarding specific shipments. Furthermore, the shift of the company’s direction to be proactive instead of reactive has prompted for the development of systems that would aid in monitoring and also provide analysis for users to not just follow up on cases, but to also seek avenues for further improvements. Hence, we will be working on this project to help formulate an effective solution to their needs whilst providing deeper analysis into the data that will help identify potential problems before they arise.

# **Motivation**

As final year students graduating soon from university, we seek to find a project that would suit our future directions in career path. What we believe that will be rising in the future is e-Commerce and Logistics. Being familiar with logistics and distribution would aid us greatly in managing e-Commerce solutions and systems that we will work on in the future. Following this, we engaged a logistics company in the market that has partnered with various firms handling the logistics solutions and e-Commerce platforms for them, where we will learn from hands on experience in this project.

# **Data**

Total timespan of available data is 3 months worth. There are 2 main datasheets by 2 tracker systems, App1 and App2, owned by the company. Generally they display shipment and customer data for all the items that the company delivers. The dataset is utilized by the sales team to track and monitor shipments.

The App1 TrackHeader and the App2 ShipmentInfo datasets store information pertaining to the shipment itself, such as the sender and receiver data. The App1 Track Detail and the App2 StatusInfo datasets store information pertaining to the tracking data of the shipment, with each row representing a status update for each tracking number. Such data is split into these 2 types following the database normalization processes as per standard data management practices.

All dates and timings in the dataset are stored in String format, so we will classify such data types as Text/String format for easier reference.

Legend:

# Numeric/ID/int or double

^ Text/String format (Eg. words/letters)

\* Fixed/Constant

App1

|  |  |  |  |
| --- | --- | --- | --- |
| TrackHeader |  |  | Track Detail |
| \* Application | ^Destination Country | ^Sender Country | \*Application |
| #Tracking No | ^Pickup Contact Name | ^Receiver Name | ^Tracking No |
| ^Shipment Date | #Pickup Contact No | #Receiver Contact No | ^Stage Code |
| ^Reference No | ^Pickup Address | ^Receiver Address1 | ^Stage Completed Date |
| #Master Con Note | ^Pickup City | ^Receiver Address2 | ^Reason Code |
| #Quantity | ^Pickup State | ^Receiver Address3 | ^Remarks |
| #Initial Weight | #Pickup ZipCode | ^Receiver Address4 | ^Updated City |
| #Volume Metric Weight | ^Pickup Country | ^Receiver City | ^Updated Country |
| #Total Weight | ^Sender Name | ^Receiver State | ^Last Updated Date |
| ^Special Delivery Instruction | #Sender ContactNo | #Receiver ZipCode | ^Last Updated By |
| ^Service Type | ^Sender Address1 | ^Receiver Country |  |
| Customer Info | ^Sender Address2 | ^Last Updated Date |  |
| ^Customer Name | ^Sender Address3 | ^Agent / Vendor Name |  |
| #Customer Account No | ^Sender Address4 | ^ Vendor\_Reference No |  |
| ^Origin City | ^Sender City | ^ ThirdParty\_Reference No |  |
| ^Origin State | ^Sender State | ^ Network Classification |  |
| ^Origin Country | #Sender ZipCode | ^ PODFileLocation  |  |
| ^Destination City |  |  |  |
| ^Destination State |  |  |  |

App2

|  |  |  |  |
| --- | --- | --- | --- |
| Shipment Info |  | Status Info |  |
| \* Application | ^ Destination City | \* Application | ^ Tracking No |
| # Weight | ^ Order\_Date | # Stage Code | ^ Stage\_Updated Date |
| # Quantity | ^ PickUp\_Date |  | ^ Updated\_Location |
| # Account No | ^ Customer Name |  | ^ Remarks |
| # Master\_Con\_Note(MAWB) | ^ Special Delivery Instruction |  | ^ Last\_Updated\_By |
| # Pickup\_Contact No | ^ Pickup\_Contact\_Name |  | ^ Last\_Updated\_Date |
| # Receiver\_Contact No | ^ Pickup\_Address |  |  |
| # Receiver\_ZipCode | ^ Receiver\_Name |  |  |
| ^ Reference No | ^ Receiver\_Address |  |  |
| ^ Tracking No | ^ Receiver\_City |  |  |
| ^ Origin | ^ Shipment Type ID |  |  |
| ^ Destination | ^ Agent |  |  |

# **Methodology**

## **Introduction**

 The main aim of this practicum is to give our sponsor an insight into the delivery patterns in the different countries managed, focusing on Australia and Japan as these 2 countries have posed the most problems. To do so we will first analyse the trends from 3 months worth of data using 4 main techniques, Exploratory, Clustering, Time Series, and Geospatial.

With these analysis done, we hope to give our sponsors a clearer picture as to the reasons for failed deliveries to aid the company in avoiding similar pitfalls in the future.

## **Tools Used**

 We’ll be manually extracting the data we need from the raw data sheets provided. There is also the need to combine the data from both company’s applications (App1 and App2). After which we will proceed with the analysis using JMPro and Power BI to perform exploratory analysis, clustering, and time series. We agree that JMPro is a more powerful too but the reason for using Power BI is because our sponsors are familiar with the software so we want to get familiarise with its display as well so that we can have a better idea how to construct our final web app. QGIS will be our main application for the Geospatial analysis.

Eventually we will display our findings on a single display (most probably Javascript) as per requested by the sponsor.

## **Analysis**

We will be performing the analysis in four stages.

### **Exploratory Analysis**

An exploratory analysis will be conducted first to analyse the shipping behaviour of different customers in different countries.

* Determine the average turnaround time from the first to the last stage.
* Determine the average turnaround time for the statuses closure
* Identify patterns between destinations and shipment issues.
* Identify types of shipments with frequent shipment issues.

### **Geospatial Analysis**

Shipping patterns and behaviour can be identified using geospatial analysis. The analysis will be narrowed down to the country, state/city and postal code. We will seek to answer the following questions:

* Where different customers lie on the map and hopefully identify the more popular areas and their reasons
* How different locations and proximity to the warehouses can affect shipment time and procedures.
* Identify and flag out destinations with high probability of shipment issues.
* Track different shipping routes from the start to the final to determine the average time required.
* Track different shipment status gap to determine partner’s performance in data provision/updates

### **Clustering**

We plan to cluster our data based on type of customer, shipping history, activity level and any other potential classifications which we may identify in the future. Each customer/vendor will then be assigned a cluster number.

### **Time Series Data Analysis**

As the data could be organised by the date, a time series analysis could be conducted. The time series analysis would be broken down into time periods of weeks and month to analyse and identify patterns and trends in the shipment and customer data.

We will also attempt to determine if there are seasonality trends in shipment patterns across different countries for different shipments.

# **Work Plan**

* Fortnightly meeting with Supervisor
* Monthly meeting with Sponsor (last Wed of every month. Tentatively set for Week 4/8/12/16)

On top of these meetings with our sponsor and instructor, we also have a weekly meeting diary which will be uploaded on Wiki every Sunday night 2359. On top of that we will upload the wiki every week regarding the project’s progress as well.



# **Application**

We aim to build a single display. The sponsors we’re working with wish to have easy to read data as they are focused on the marketing simplicity and efficiency only. The functions we hope to show includes:

1. Showing data records of parcels picked up but not replied
2. Show visual summary of shipments and current status
3. View failed deliveries at a single glance and detailed breakdown at a single click, including track by ref for both inbound and outbound
4. Peak of the failure points when time series analysis
5. Simple to understand bar charts and histograms

# **Limitations and Risk**

Inconsistency in data across countries may make it difficult for comparison. For example, we may be able to do some geospatial analysis for countries such as Australia but the same cannot be easily done for countries like Hong Kong which have a postal code system with non standardized addresses written by the end consumers.