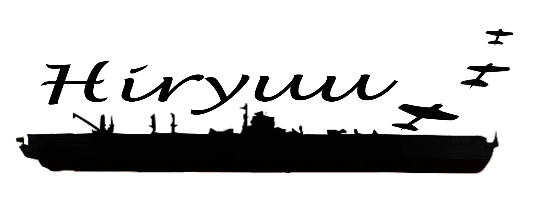


**ANLY482 - Analytics Practicum**

Project Proposal



Jouta Lim Zi Yu | Lin Qianpin | Chua Wan Theng

**1.0 Project Background**

Our sponsors are a Regional Logistics Provider that has done many different projects, handling areas pertaining to warehousing, distribution and e-Commerce. One of the new projects taken on by the company is to manage the distribution network for one of its major clients, handling both reverse logistics and forward logistics, hence having both inbound and outbound shipments. With the start of this new project, the complexities in the processes have resulted in complications that require finding solutions from the data which gives us an understanding of what is happening on ground as we cannot be physically present. Similarly, all questions posed by the client are to be answered accompanied by data evidence to prove the legitimacy of the answers. Hence it becomes crucial to find ways to better read and understand the data to gain deeper insights from it.

**1.1 Business Problem**

Despite the amount of data handled by the team, the lack of a dashboard to show summary tables regarding the status of the shipments has made it difficult for the program managers and the sales team to monitor and follow up with requests from the client regarding specific shipments. The data used by them comes from a number of different sources, which needs to be mapped to their own templates to be used effectively by them.

Furthermore, the shift of the company’s approach to be proactive instead of reactive has prompted the development of systems to aid in monitoring and also provide analysis for users to not just follow up on cases, but to also seek avenues for further improvements.

With this opportunity, we have begun working with the company to take the first step into developing a tool that would provide them with better insights into their data.

**1.2 Motivation**

As final year students graduating soon from university, we sought to find a project that would suit our future directions in career path. All of us in the team are keen to take on a career related to either e-Commerce or logistics. Hence, we believe that the project we have chosen would familiarise us with the workings of logistics and distribution and aid us greatly in managing e-Commerce solutions and systems that we will have the opportunity to work on in the future.

For our choice of project, we engaged a logistics company in the market that has partnered with various firms handling the logistics solutions and e-Commerce platforms for them. From this project, we will have the opportunity to learn from the hands on experience with the data provided to us by the sponsor. This project involves various countries and has data originating from each of the countries, hence the language used for certain data columns such as address and instructions are in the country’s native language (E.g. Japanese, Chinese, Thai). The challenge to handle such non-standardized data fields and to gain insights from it gives us the opportunity to learn beyond the classroom.

From this project, we will gain many opportunities like these to familiarize ourselves with handling real world data unlike what we have used thus far in our classes. Having to handle the unexpected and learning how projects are managed in the real world through hands-on are unlike what we have experienced thus far in our classes. We believe that this project will provide us with invaluable experiences that allow us to put what we learnt into practice, while receiving further guidance on how to better improve our work before we step out into the real world to handle such things ourselves.

**2.0 Data**

Data provided to us are distribution data that span over a period of 3 months. There are 2 main datasheets by 2 tracker systems, App1 and App2, owned by the company. They display the shipment and customer data for all the items the company delivers. The dataset is utilized by the sales team to track and monitor shipments.

The App1 TrackerHeader and the App2 ShipmentInfo datasets store information pertaining to the shipment itself, such as the sender and receiver data. The App1 Tracker Detail and the App2 StatusInfo datasets store information pertaining to the tracking statuses of the shipment, with each row representing a status update for each tracking number.

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

The tracking data lists the shipment data of the following countries:

* Hong Kong
* Japan
* Australia
* Korea
* Taiwan
* Thailand
* Malaysia
* New Zealand
* Singapore

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 |  |  |

**3.0 Methodology**

**3.1 Introduction**

The main aim of this project is to give our sponsor a deeper insight into the delivery patterns in the different countries managed, focusing on Australia and Japan as these 2 countries have posed the most problems.

**3.2 Literary Research**

To better understand the nature of the data in the domain of e-Commerce and logistics, we did some general research to gain a better understanding of the workings of the industry. We believe this would help us make better decisions on how to approach the analysis.

We learnt that big data drives decision making support in several industries including logistics. Big data applications can provide value by reducing costs, decision improvements in products and series and improving productivity and competitiveness. A survey conducted from Accenture shows that big data application in logistics can:

* Improve customer service
* Provide faster and more effective reaction time to supply chain issues
* Increase the efficiency in supply chains
* Enable integration across the supply chain
* Allow better decision making and better customer and supplier relationships.

For example, in distribution and logistics optimisation, solutions such as the GPS-enabled telematics and route optimisation can help manage distribution better. (Ziora, Brzozowska, Sałek, & Wiśniewska-Sałek, 2016) As such, our project will aim to utilise data analytics to help optimise shipment routes using geospatial analytics.

Additionally, part of the data we received from our sponsor pertains to Reverse Logistics. Burnson, P. (2016) explains about the reverse logistics process and its increasing importance today. Reverse logistics, as suggested by its name, refers to the process of returning goods to the manufacturer. Manufacturers may choose to recondition, refurbish, remanufacture, resell, or recycle such returned goods depending on its condition. The importance of having effective reverse logistics would improve customer satisfaction and allow cost savings from reduced spending. One of the key needs of an effective reverse logistics solution is visibility into the inventory at all stages of its lifecycle, highlighting a need to a clear view into the data of the products.

Lastly, as the shipment of products from our sponsor contain sensitive items such as laptop batteries, we have researched on the shipment procedures with regards to shipping dangerous items.

Laptops unlike other average goods, contains lithium ion batteries. According to the UN3481, it is labelled as a “dangerous good” due to the inherent instability of the lithium itself. Any excessive heating that was not factored in would lead to the battery exploding, as demonstrated by the recent news on the Samsung phones. (University, n.d.)

To ship a dangerous good, shipping entities would need to adhere to packaging protocols depending on the scenario.

|  |
| --- |
| Scenario 1 – (1-2 batteries)   * Box itself must be strong as to not break or allow batteries to drop out/expel fumes (usually industrial box) * No markings/labels required on the box |
| Lithium Battery Label for Laptop Shipments.jpgScenario 2 ( >2 batteries)   * Box itself must be strong as to not break or allow batteries to drop out/expel fumes (usually industrial box) * Markings/labels ARE required on the box * Documentation required on the marking/label * Packing contains Lithium ion Batteries * Package must be handled with care and a flammable hazard exists if packaged is damaged * If package is damaged special procedures including inspection and repacking may be necessary   Telephone number to call for additional information |

(Gotz, 2012)

**3.3 Objectives**

There are 5 main objectives we aim to achieve:

1. Understanding the patterns and trends across shipment routes in different countries
2. Identify patterns such as the locations and timing for shipments with frequent issues
3. Identify possible clusters based on types of shipments or customers to easily classify shipments
4. Conducting time series analysis to determine the presence of seasonality in shipments
5. Building a dashboard for single view of all data statistics and to understand KPI easily. The sponsors we’re working with are focused on the marketing simplicity and efficiency only. The functions we hope to show includes the following:

**Design Specification**

* Showing data records of parcels picked up but not replied
* Show visual summary of shipments and current status
* View failed deliveries at a single glance and detailed breakdown at a single click, including track by reference number for both inbound and outbound
* Peak of the failure points when time series analysis
* Simple to understand bar charts and histograms
* Testing Iteration of Application

Multiple iterations of the dashboard will be conducted to increase the usability for our sponsor. We will conduct frequent feedbacks with our supervisor and sponsor to ensure that the dashboard is equipped with the data statistics and KPI most readily useful for the decision making.

**3.4 Analysis**

1. **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.

1. **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

1. **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.

1. **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.

**4.0 Scope of Work**

1. **Data Gathering and Scoping**

Data have been gathered from our sponsor which consists of data across 3 months and spanning across several countries. The data given to us are all in .csv formats.

1. **Research on Software and Proposal Preparation**

Our team has researched and discussed on the software and applications used in analysing and visualising the data with our supervisor and sponsor.

We will mainly utilise JMP Pro for the clustering, exploratory, and seasonality analysis and QGIS for our geospatial analysis. Our sponsor has expressed their preference for the final product to be in Power BI as they are familiar with it. In addition, they have expressed interest for the final product to be dynamic, meaning any input of future data can tap on the methods used to produce the results.

Currently we will focus on analysis on JMP Pro and QGIS and have plans to decide how to compromise the dynamic capabilities within our limitations.

1. **Data Cleaning and restructuring**

After looking and familiarising ourselves with the data, we discovered the complexity of the data. The data is not only large in size, contains a wide variety, and also inconsistent across different countries. This inconsistency is due to the differences in language used for reporting as well as different systems used. As such, certain coding languages could not be used to process the values and other solutions have to be looked into.

Due to the complexity and variety of the data, we will be conducting data cleaning by removing duplicates, ensuring consistency by inserting the header column, merging the data files, filtering the data by countries and checking that the data is appropriate and ready for analysis.

1. **Data Modelling**
2. Exploratory Analysis
3. Geospatial Analysis
4. Clustering
5. Time Series Data Analysis

**5.0 Work Plan**

**5.1 Milestone Deliverables**

1. Proposal Deadline (15 Jan)
2. Interim Report + Presentation (19 Feb)
3. Abstract Paper Submission (2 April)
4. Full Paper Submission (20 April)
5. Final Presentation/Conference Exhibition (22-23 April)

**5.2 Gantt Chart**

* Fortnight meeting with Supervisor
* Monthly meeting with Sponsor (last Wed of every month)



**6.0 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 whom do have a postal code system with non-standardized addresses written by the end consumers.

# **Bibliography**

Burnson, P. (2016). Time to Embrace Reverse Logistics. *Logistics Management (2002),* *55*(3), 36-38.

Gotz, M. (2012, February 27). *Shipping Laptop Computers - Packaging, Labeling and Documentation Requirements*. Retrieved from Quick International Courier: http://quick.aero/quickintl/blog/shipping-laptop-computers-packaging-labeling-documentation-requirements

University, B. (n.d.). *Lithium-ion Safety Concerns*. Retrieved from Battery University: http://batteryuniversity.com/learn/archive/lithium\_ion\_safety\_concerns

Ziora, L., Brzozowska, A., Sałek, R., & Wiśniewska-Sałek, A. (2016). The Possibilities of Big Data Solutions Application in Logistics. *MultiScience - XXX. microCAD International Multidisciplinary Scientific Conference,University of Miskolc, Hungary*.