SMU
SINGAPORE MANAGEMENT UNIVERSITY

ANLY482 Analytics Practicum Project Proposal

Optimizing operations productivity and developing an interactive dashboard for local Supply Chain Company

Team 03

Submitted by OPTEAMIZATION

1. Tay Jing Ying (jytay.2014@sis.smu.edu.sg)

2. Russell Yap Song Chen (russell.yap.2014@sis.smu.edu.sg)

3. Chua Weilun (weilun.chua.2014@sis.smu.edu.sg)

20 February 2018

Contents

Project Sponsor

Our sponsor is a home-grown supply chain solutions company and leading regional supply chain management partner to a number of the world's leading brands such as Dell, Motorola, Samsung, LG, Exxon Mobil, Unilever and Heinemann. The company has operations all over Asia Pacific, such as China, the whole of ASEAN, India and Korea and they service various industry segments: Chemicals & Healthcare, Consumers & Retail, Electronics & Technology, E-Commerce and Cold Chain.

A strong advocate for technological and operational innovation, the company advocates productivity and sustainability improvements through interactive solutions so that they can provide their diverse consumer-base from various industry clusters with world-class logistics services.

Motivation

The company has stayed ahead of their competition by embracing technology in the early days and now, they are starting to dabble into analytics in order to help make use of the large amount of data that they have generated throughout the course of their business. Currently, these information is being stored in their proprietary Warehouse Management System (WMS) which captures transaction information whenever goods are moving in, out or within the warehouse.

The amount of information being managed by the system is massive and it is very complicated, thus managers have to spend time to use pivot tables in excel to better visualize the information being generated in order to get a better view of how the current operations is being run. This uses up valuable time which could have been used for other more important operational matters.

One important information that is very useful to a warehouse manager is an ABC analysis which is an inventory categorization technique. The purpose of this analysis is to classify the inventory into 3 categories. One possible way of doing categorization is shown below.

- 'A' items – 20% of the items accounts for 70% of the annual consumption value of the items.
- 'B' items - 30% of the items accounts for 25% of the annual consumption value of the items.
- 'C' items - 50% of the items accounts for 5% of the annual consumption value of the items.

Not only that, in order to provide even more better insight for their customers, they would also like to be able to have a better view of the **seasonality** patterns of the products, as well as an **affinity analysis** of the products to further improve the way the products can be organized in the warehouse to increase the efficiency of the warehouse operations.

Objectives

The main objectives of the project would be to do the following:

1. Create an Operations Dashboard to visualize the following information
    - Inbound / Outbound Volume Time Series Chart
        - The purpose of this chart is to visualize the operations performance at various time slices so as to allow operations to have a clear view of the inventory movement patterns for the various SKUs
        - X-axis: Time unit (day, week, month, quarter, year)
        - Y-axis: No. of units of product
        - Lines: Each line would symbolize the trend for a different SKU.
    - Product Ranking Chart
        - The purpose of this chart is to find out what the best-selling items are and what are the least popular items.
        - Tree map view of the products filtered by month / year
    - Product Seasonality Chart
        - The purpose of this chart is to see the seasonality factors of the different products.
        - X-axis: Month of the year
        - Y-axis: Number of units of product
        - Lines: Each line would symbolize the trend for a different SKU.

2. Conduct **Associative Rule learning** to attempt to see if there are affinities within the products within the warehouse so that inventory can be better organized within the warehouse.

Data

The data provided by our sponsor are obtained from their proprietary Warehouse Management System (WMS) which is developed in-house. The Warehouse Management System is used to keep track of the movement of the goods coming in, moving out and within the warehouse itself. The process of goods coming into the warehouse is known as the Inbound process while the process of goods moving out of the warehouse is known as the Outbound process.

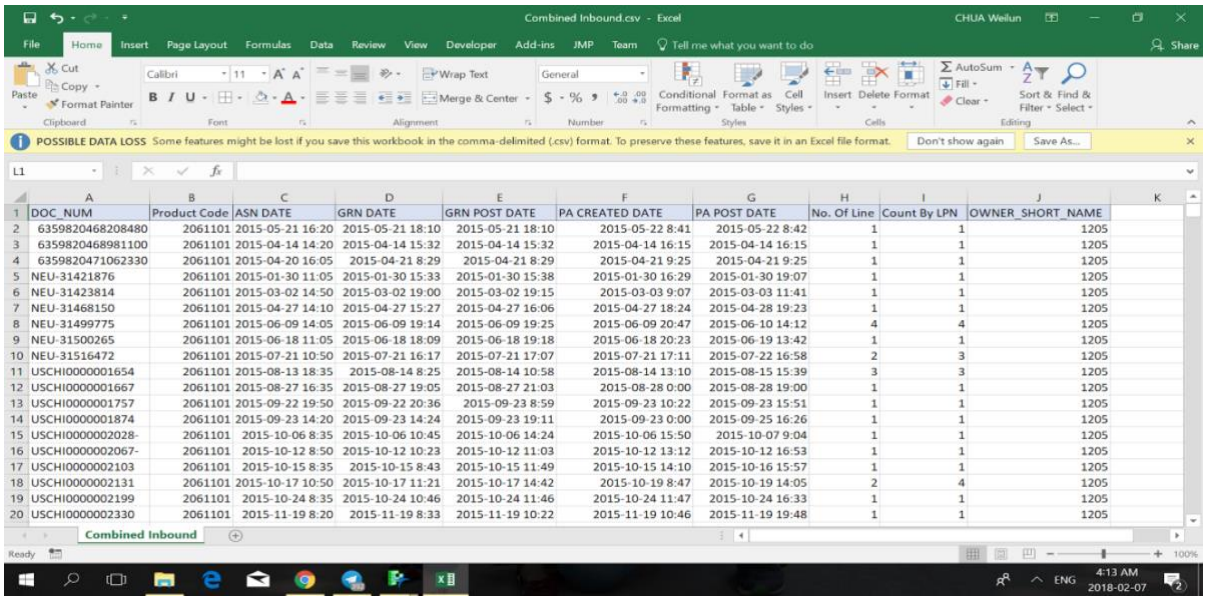
For this project, our sponsor has provided us with data for one of their client, which is a company selling cables and connectors. We will be provided with 3 years worth of data from year 2015 to 2017. There are 2 Excel workbooks provided, one for the Inbound goods and the other for the Outbound goods. A more in-depth explanation of the Excel workbooks will be explained below.

Inbound Report

There are a total of 308,730 rows of data spanning across January 2015 to December 2017 for the Inbound orders for this client. In 2015, there were 99,586 rows of data and this increased in 2016 to 102,379. In 2017, the number of rows of data was 106,765.

Metadata Dictionary

Terminology	Description
Doc_Num	Unique identifier for an inbound order.
Product_Code	Unique identifier of a product.
ASN Date	Advanced Shipping Notice. ASN date represents the date when the ASN is downloaded.
GRN Date	Goods Received Note. The date when the GRN job is created
GRN Post Date	The date when the GRN job is completed
PA Created Date	The date when the putaway job is created.
PA Post Date	The date when the putaway job is completed.
No. Of Line	The number of lines in the storing list.
Count by LPN	Number of cartons.
Owner_Short_Name	The short name of the client.



DOC_NUM	Product Code	ASN DATE	GRN DATE	GRN POST DATE	PA CREATED DATE	PA POST DATE	No. Of Line	Count By LPN	OWNER_SHORT_NAME
6359820468208480	2061101	2015-05-21 16:20	2015-05-21 18:10	2015-05-21 18:10	2015-05-22 8:41	2015-05-22 8:42	1	1	1205
6359820468981100	2061101	2015-04-14 14:20	2015-04-14 15:32	2015-04-14 15:32	2015-04-14 16:15	2015-04-14 16:15	1	1	1205
6359820471062330	2061101	2015-04-20 16:05	2015-04-21 8:29	2015-04-21 8:29	2015-04-21 9:25	2015-04-21 9:25	1	1	1205
NEU-31421876	2061101	2015-01-30 11:05	2015-01-30 15:33	2015-01-30 15:38	2015-01-30 16:29	2015-01-30 19:07	1	1	1205
NEU-31423814	2061101	2015-03-02 14:50	2015-03-02 19:00	2015-03-02 19:15	2015-03-03 9:07	2015-03-03 11:41	1	1	1205
NEU-31468150	2061101	2015-04-27 14:10	2015-04-27 15:27	2015-04-27 16:06	2015-04-27 18:24	2015-04-28 19:23	1	1	1205
NEU-31499775	2061101	2015-06-09 14:05	2015-06-09 19:14	2015-06-09 19:25	2015-06-09 20:47	2015-06-10 14:12	4	4	1205
NEU-31500265	2061101	2015-06-18 11:05	2015-06-18 18:09	2015-06-18 19:18	2015-06-18 20:23	2015-06-19 13:42	1	1	1205
NEU-31516472	2061101	2015-07-21 10:50	2015-07-21 16:17	2015-07-21 17:07	2015-07-21 17:11	2015-07-22 16:58	2	3	1205
USCHI0000001654	2061101	2015-08-13 18:35	2015-08-14 8:25	2015-08-14 10:58	2015-08-14 13:10	2015-08-15 15:39	3	3	1205
USCHI0000001667	2061101	2015-08-27 16:35	2015-08-27 19:05	2015-08-27 21:03	2015-08-28 0:00	2015-08-28 19:00	1	1	1205
USCHI0000001757	2061101	2015-09-22 19:50	2015-09-22 20:36	2015-09-23 8:59	2015-09-23 10:22	2015-09-23 15:51	1	1	1205
USCHI0000001874	2061101	2015-09-23 14:20	2015-09-23 14:24	2015-09-23 19:11	2015-09-23 0:00	2015-09-25 16:26	1	1	1205
USCHI0000002028-	2061101	2015-10-06 8:35	2015-10-06 10:45	2015-10-06 14:24	2015-10-06 15:50	2015-10-07 9:04	1	1	1205
USCHI0000002067-	2061101	2015-10-12 8:50	2015-10-12 10:23	2015-10-12 11:03	2015-10-12 13:12	2015-10-12 16:53	1	1	1205
USCHI0000002103	2061101	2015-10-15 8:35	2015-10-15 8:43	2015-10-15 11:49	2015-10-15 14:10	2015-10-16 15:57	1	1	1205
USCHI0000002131	2061101	2015-10-17 10:50	2015-10-17 11:21	2015-10-17 14:42	2015-10-19 8:47	2015-10-19 14:05	2	4	1205
USCHI0000002199	2061101	2015-10-24 8:35	2015-10-24 10:46	2015-10-24 11:46	2015-10-24 11:47	2015-10-24 16:33	1	1	1205
USCHI0000002330	2061101	2015-11-19 8:20	2015-11-19 8:33	2015-11-19 10:22	2015-11-19 10:46	2015-11-19 19:48	1	1	1205

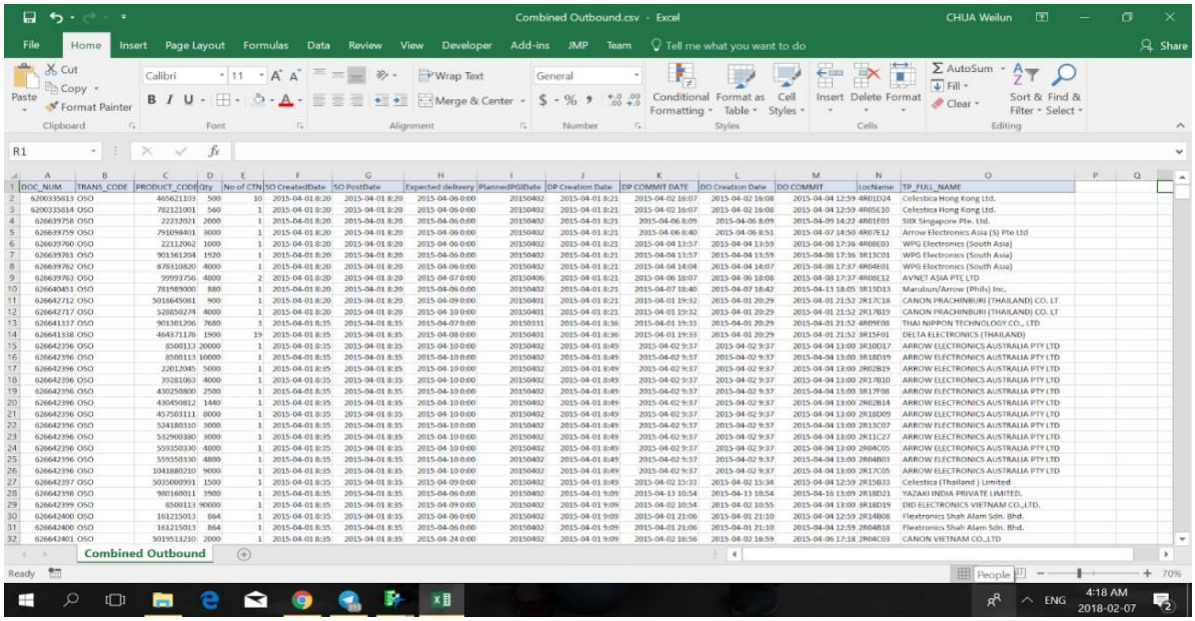
The report shows the order level where each row represents an order by the client.

Outbound Report

The number of Outbound orders for this client from January 2015 to December 2017 is approximately 657,578 rows of data. In 2015, there were 220,387 rows of data and this increased in 2016 to 221,498. In 2017, the number of rows of data was 215,667.

Metadata Dictionary

Terminology	Description
Doc_Num	Unique identifier for an outbound order.
Trans_Code	There are several trans code, but we will only focus on: • OSA (outbound stock adjustment – might want to consider this also as usually when there is shortage or excess physically in the warehouse, we need to perform this in the system to ensure that the quantity tallies between system and physical) • OSO (For shipment outbound) • OOT (Outbound Ownership Transfer - When there is a change of plant code (example: from plant code 1201 to plant code 1402) • OTA (Outbound Transaction Adjustment - When there is a data entry error and this transaction code is used to reverse the wrong transaction out) • OVT (Outbound Vendor Transfer - When there is short shipment)
Product_Code	Unique identifier of a product.
Qty	The quantity of products needed for the order.
No of CTN	Number of cartons for the order.
SO Created Date	The date when the sales order is downloaded.
Expected Delivery Date	The date which the goods is expected to be shipped out.
DP Creation Date	The date when the picking list is printed for the picker to pick the goods in warehouse.
DP Commit Date	The date when the picker has finished picking.
DO Created Date	The date which the delivery order is created (the date indicates that we are ready for shipment and waiting for transporter to load our goods)
DO Commit	The date when the delivery order is committed (this happens after the goods left the warehouse)
LocName	The name of the location in our warehouse.
Ship_To/TP_FULL_NAME	Outbound customer name
PlannedPGIDate	Planned DP post date



The report shows the order level where each row represents an order by the client.

Methodology

In this section, we will explain the methodology which our team plan to implement to perform analysis on the data provided by our sponsor.

We will be using Python, Tableau Software and Microsoft Excel for Exploratory Data Analysis (EDA) to better understand the dataset given and its characteristics. As part of data preprocessing, our team will be performing the following steps to obtain a clean dataset. These steps will eventually be converted into a script which will be used to clean the data file that is uploaded into the dashboard which we will develop for our sponsor.

Data Preprocessing

With every new dataset, we first must clean the data to remove irrelevant data that should not be included in our analysis. For data cleaning, the steps include:

1. Handling missing values. If there are missing values in a row of record, the entire row will be excluded because it will be inaccurate to include it.
2. Handling duplicate data. Duplicate data could occur when the employees double scan the barcode upon inbound of goods. Similarly, in the event of duplicate data, we will remove the entire row as well.
3. Resolving redundancies caused by data integration.

With the clean dataset, we will proceed to further explore the data and find out potential visualizations and analysis that can be done with the dataset to provide a more in-depth analysis and dashboard that will be useful for our sponsor.

Visualizations

The final product of our project is to create an Operations Dashboard to visualize the following KPIs:

1. Product Volume Chart
2. Product Inbound / Outbound Volume Time Series Chart
3. Product Ranking Chart
4. Product Seasonality Chart
5. Warehouse Heat map

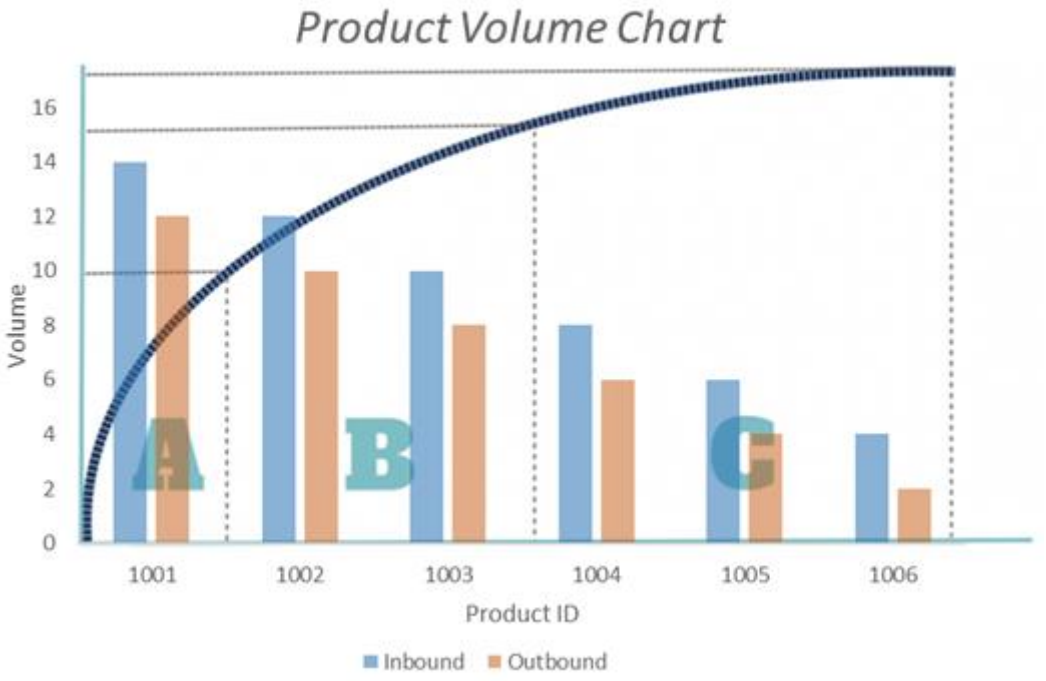
The users will have the ability to upload the Inbound and Outbound reports as CSV files on the dashboard. After the data cleaning script is run, the information will be stored into the database. The data in the database will be used for data visualizations on the dashboard, which will be explained in greater depth later.

The visualizations to be used on our dashboard include Bar Chart, Time Series Line Chart and Tree map. The users will have the ability to select the period to view as well.

Bar Chart

Product Volume Chart

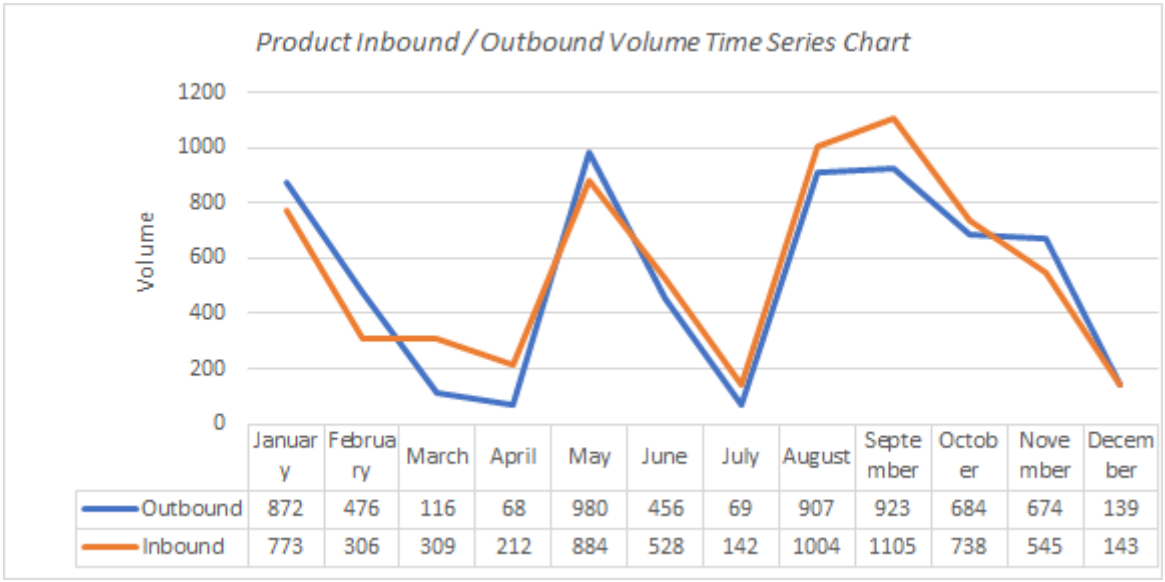
For the Product Volume Chart, the x-axis will be the products while the y-axis will be the volume. For each product, a dual bar chart will be used, one for Inbound volume and another for Outbound volume. A Pareto chart will also be used to show the inventory which are in top 20% based on Outbound volume, which gives us the fast-moving products that are in category A.



Time Series Line Chart

Product Inbound / Outbound Volume Time Series Chart

This chart aims to show the volume of Inbound or Outbound inventory over the period through a Time Series Line Chart. It can be accessed by clicking the bar for the product in the Products Volume Chart. It allows the operations to have a clear view of the inventory movement patterns for the various SKUs to understand which items have the highest or lowest volume across the period. This helps our sponsor to understand the trend of an inventory to see if there are any special period when there is a peak in movement.



	January	February	March	April	May	June	July	August	September	October	November	December
Outbound	872	476	116	68	980	456	69	907	923	684	674	139
Inbound	773	306	309	212	884	528	142	1004	1105	738	545	143

Example of a time series line chart which shows the inflow and outflow quantities over time along the same axis.

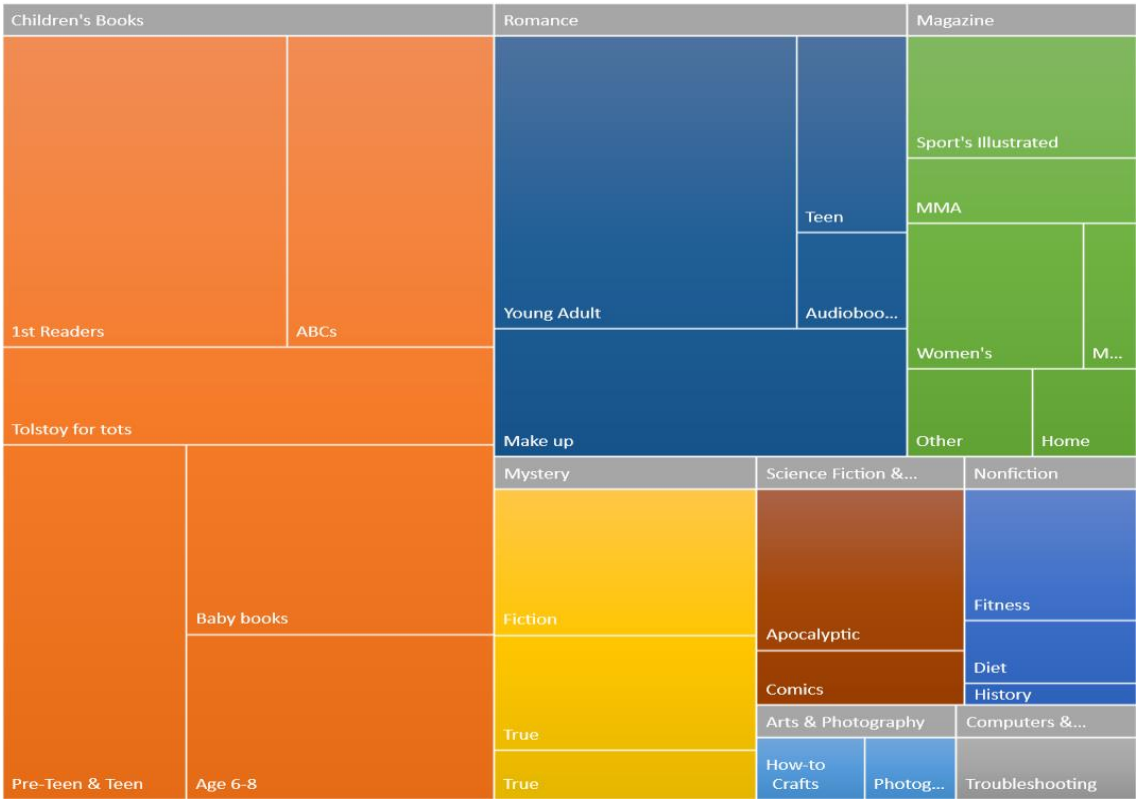
Product Seasonality Chart

We can drill down into the individual product seasonality performance through a Time Series Line Chart, where we can observe the Inbound and Outbound volume of the product over a period. This allows us find insights on which products are more popular during which period and that can help us to determine the period where each product belongs to category A, B or C.

Treemap

Product Ranking Chart

A Treemap can be used to identify which product is the best-selling item and which is the least selling item. They are being ranked according to the size and proportion of their sales as compared to one another, it will be based on the Outbound volume. As seen from the example below, we can identify the best-selling product by the proportion of its sales. The Treemap view of the products can be filtered by period, either month or year.



Example of a Treemap that is used to understand which is the best selling categories of books.

Heatmap

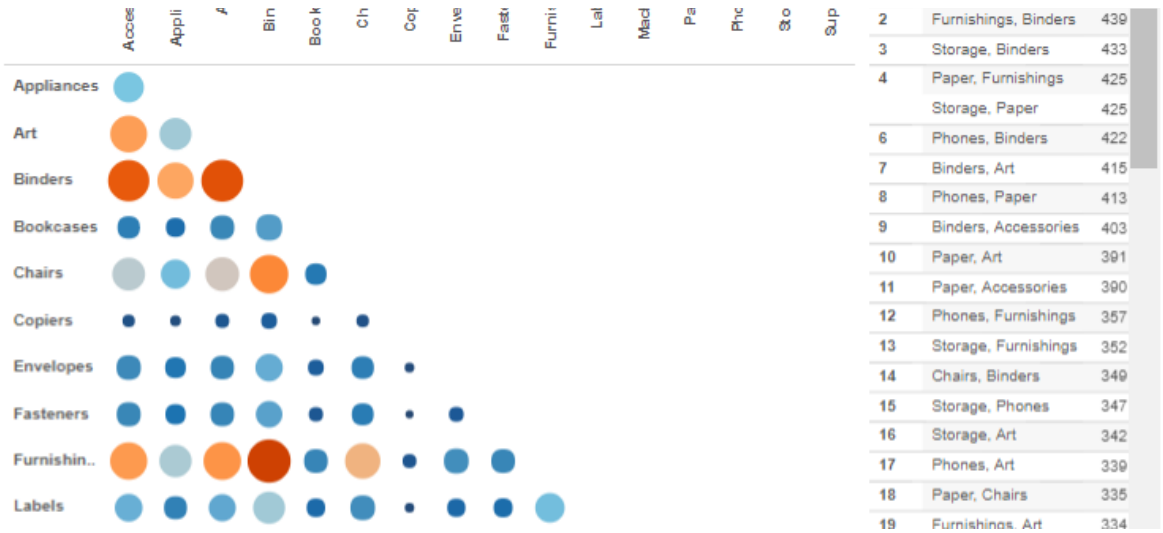
Warehouse Heatmap

A heatmap will be used to showcase the locations where the different categories (A, B and C) goods are usually taken from for Outbound shipment. This can tell us where the category A goods are usually placed and it gives a good indication of whether the location in the warehouse is properly utilized. This is because if a good is fast-moving, it should be placed near the staging area so that it can be quickly packed by the workers.

Example of a warehouse heatmap visualization

Market Basket Analysis

Our team also aims to perform Market Basket Analysis in attempt to find out affinities between products that are within the warehouse. For instance, in a particular month, when product X is moved out of the warehouse, product Y is usually moved out as well. This can help us to better organize the inventory in the warehouse as product X may belong to category A while product Y may belong to category B. However, because of this affinity they have, it might be wiser to put product Y beside product X during that particular month.



Example of a market basket analysis chart that is used to understand which products tends to be requested together

Technology

We will be utilizing several tools to aid us in the development of our dashboard for our sponsor. This includes:

Highcharts and JavaScript	Highcharts is a simple-to-use charting library written in JavaScript which allows us to add interactive charts to our web application for the dashboard. It will be used to develop our dashboard for the various visualizations mentioned above.
Python and Jupyter	Python is a programming language which we will use to create a script meant for cleaning the dataset when it is uploaded and initial Exploratory Data Analysis. It will be done using Jupyter notebook and relevant python packages such as pandas and numpy.
PostgreSQL	PostgreSQL is an object-relational database management system which will be used to store the data from the CSV files.
Tableau Software	Tableau will be utilized for Exploratory Data Analysis.
Microsoft Excel	Microsoft Excel will be utilized for Exploratory Data Analysis.

Scope of Work

The following describes the scope of work which the team will be working on across the span of the next few months till the final submission.

Data Gathering and Scoping

Our sponsor has given us various excel spreadsheets, both Inbound and Outbound, with data from 2015 to 2017, for a client selling cables and connectors.

Our project scope includes analysing the products ranking, products seasonality and products inbound and outbound volumes. Additionally, we will also look at the affinity between products to see if there are trends of any products frequently being shipped together. With the data, we will perform Exploratory Data Analysis using Tableau Software, Microsoft Excel and Python.

Data Cleaning, Wrangling and Restructuring

Our team aims to combine the 3 years' worth of data for Inbound products into 1 excel spreadsheet. This will be the same for the Outbound report. We will also try to combine the Inbound and Outbound reports by using their unique identifier to link the data together. Data cleaning will be done using Python to handle missing values and duplicate data, after integrating the data.

The team will also use Python to write a script which will combine the various excel spreadsheets and extract the relevant fields before consolidating them into a single spreadsheet that will be used as an input for the analysis and the Operations Dashboard. All the data that is uploaded will be stored into a PostgreSQL database, which serves as the database of information for the dashboard.

Data Visualization and Reporting

Once the data is cleaned and combined, we will extract the data and create a dashboard with various visualizations as mentioned earlier. This will be done using Highcharts and JavaScript. The interactive dashboard will allow managers to have easy access to visualizations showing the performance of the current operations as compared to past operations.

On top of the visualizations, we will also do Market Basket Analysis to find out the affinity between products. Also, we will report all our findings to our sponsor and propose how the goods should be arranged in the warehouse based on their categories (A, B or C).

Work Plan

The following Gantt chart shows the task allocations to each member for each iteration, as well as the important project milestones which we will prepare for.

	Tasks	Task Allocation	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16
Iteration 1: Project Proposal Preparation		**20th Dec 2017 - 31st Jan 2018**																	
1	Meeting Client to understand project specifications	JY & Russ	✓																
2	Internal Meeting: Discuss methodology and segregation of work	JY & Russ	✓																
3	Internal Meeting: Analyze dataset and work on wiki	JY & Russ	✓																
4	Internal Meeting: Finalize proposal (Sponsor 1)	ALL	✓																
Milestone 1: Project Proposal Submission		**14th Jan**																	
5	Meeting Client to understand project specifications	ALL	✓																
6	Project Re-scheduling	ALL	✓																
Change of Sponsor																			
Iteration 2: Project Re-scheduling		**1st Feb - 11th Feb**																	
7	Internal Meeting: Finalize proposal (Finalized Sponsor)	ALL						✓											
8	Client Meeting To Understand Data and Objectives	ALL						✓											
9	Update Wiki	ALL						✓											
Iteration 3: Analyze and Clean Data		**12th Feb - 25th Feb**																	
10	Data Cleaning and Scripting	Russ & Weilun																	
11	Analysis Research	JY & Russ																	
12	Supervisor Meeting (Data Cleaning and Preparation)	ALL																	
13	EDA	ALL																	
14	Update Wiki	Weilun																	
15	Interim Meeting with Sponsor	ALL																	
16	Interim Meeting With Supervisor	ALL																	
Milestone 2.1: Interim Report Submission		**25th Feb**																	
Iteration 4: Implementing Analysis Method, Interim Report Presentation, and Dashboard Preparation		**26th Feb - 4th March**																	
17	Automation script (Data Cleaning)	Russ & Weilun																	
18	Analysis Implementation (ABC)	ALL																	
19	Preparing Interim Report	ALL																	
20	Presentation slides + Rehearsal	JY & Russ																	
21	Update Wiki	ALL																	
Milestone 2.2: Interim Report Presentation		**28th Feb**																	
22	Dashboard Prototyping 1	JY																	
Iteration 5: Implementing Analysis Methods, Scripting and Dashboard Preparation		**5th March to 18th March**																	
23	Analysis Implementation (Seasonal + MBA)	ALL																	
24	Suprvisor Meeting for Analysis Reporting	ALL																	
25	Automation script (Analysis)	Russ & Weilun																	
26	Dashboard Prototyping 2	JY & Russ																	
27	Sponsor Meeting for Analysis Reporting	ALL																	
28	Update Wiki	ALL																	
Iteration 6: Dashboard Implementation		**19th March - 1st April**																	
29	Architecture Diagram	Russ																	
30	Dashboard implementation	JY & Russ																	
31	Supervisor Meeting (Dashboard Feedback)	ALL																	
32	Sponsor Meeting for Dashboard Implementation	ALL																	
33	Update Wiki	Weilun																	
Milestone 3: Abstract Submission		**1st April**																	
Iteration 7: Abstract preparation + Post Processing		**2nd - 15th April**																	
34	Prepare abstract	ALL																	
35	Update Wiki	ALL																	
36	Data post-processing	ALL																	
37	Supervisor Meeting for Final Product	ALL																	
38	Sponsor Meeting for Final Product	ALL																	
Milestone 4: Full paper Submission		**8th April**																	
Milestone 5: Conference Day		**14th - 15th April**																	
Iteration 8: Final Submission Preparation		**16th April - 29th April**																	
39	Finalize poster (14th - 15th)	ALL																	
40	Conference paper	ALL																	
41	Finalize wiki	ALL																	
42	Presentation preparation	ALL																	
43	Final submission	ALL																	
Milestone 6: Final Submission		**22nd April**																	
44	Sponsor Meeting for Project Handing-Over	ALL																	

Limitations

As with any other real life projects, there are limitations which we face especially with regards to the availability of data.

S/N	Limitations	Conclusion
1	There is no way to accurately link an Outbound shipment to an Inbound shipment. We can only link the Outbound product and Inbound product but not the exact shipment where the product is being brought into the warehouse.	We will assume a First In First Out (FIFO) order of shipment out of the warehouse as that is usually how logistics company operates. This means that the first Outbound in 2015 will be linked to the first Inbound in 2015, unless there are excess goods

		brought over from 2014. If so, calculations will be done based on the quantity to determine the number of each products left after Outbound shipment.
2	The current data provided only consists of information from 1 company, which cables and connectors, that our sponsor company is handling.	We assume that the same analysis can be replicated for the logistic warehousing of other companies' products which are similar in nature.
3	The data provided from WMS might not contain all the transactions that happened due to human errors such as missed barcode scanning.	With a large enough dataset, it can be assumed that the results found by our team can be representative and considered accurate.