Supply Chain Dashboard Analysis for ARV and HIV Lab Shipments

Vijayasree Choudary
Supply Chain Analyst
Date: July 7,2024

Introduction

The dataset provided by **USAID (2023)** offers valuable insights into the global supply chain for HIV/AIDS health commodities. This dataset, available through the USAID Development Data Library, it provides a more comprehensive understanding of pricing and shipment volumes **(U.S. General Services Administration, n.d.; USAID, n.d.)**. It aims to provide valuable insights into the supply chain health commodity shipment and pricing data focusing on ARV, HIV, etc. lab shipments to selected countries.

This report focuses on ARV and HIV lab shipments and addresses key questions regarding target markets, transportation methods, dosage forms, vendor popularity, brand popularity and product growth over years' time. Additionally, it explores the relationship between product weights(kilogram), unit of measure, pack price, and unit price and suggests additional data fields that could enhance the dashboard's effectiveness.

By using this data, stakeholders can effectively and efficiently make data-driven decisions to improve supply chain.

Methods

Data Collection and Sources

To understand further into the dataset, refer **(USAID Development Data Library)** and **(Data.gov.)** The dataset includes detailed information on ARV and HIV lab shipments to selected countries and a detailed description of each field **(USAID. (n.d.)**). The dataset was last updated on November 27, 2023, and provides valuable insights into supply chain health commodity shipment and pricing data.

Tools Used in analyzing the data

Power BI and Excel are the tools used for data extraction, data loading, data cleaning, transforming and visualizing the data.

Data Cleaning and Preparation

The dataset has been thoroughly reviewed using various steps and processes before proceeding further into visualization.

1. Manually reviewed the dataset to check the number of records and identified if there are any duplicate records or missing values using Excel operations. During this analysis identified that the dataset has 10325 records in total and fields such as **Product Sent to vendor Date (record count-3801), Product sent to client date (record count-2681), Shipment mode, dosage, weight(kilograms)** has issues with data

2. Loaded the dataset into Power BI and reviewed the data to cross verify the issues using Power Query (transformation data) click on the fields Column quality, Column distribution, Column profile which shows the empty/black spaces, if there are any errors and below is the Image **Exhibit 1** which shows the details of the Dataset quality view. In **Exhibit 2**, shows the different format data in date fields and weight(kilogram) has numerical and categorical types of data.
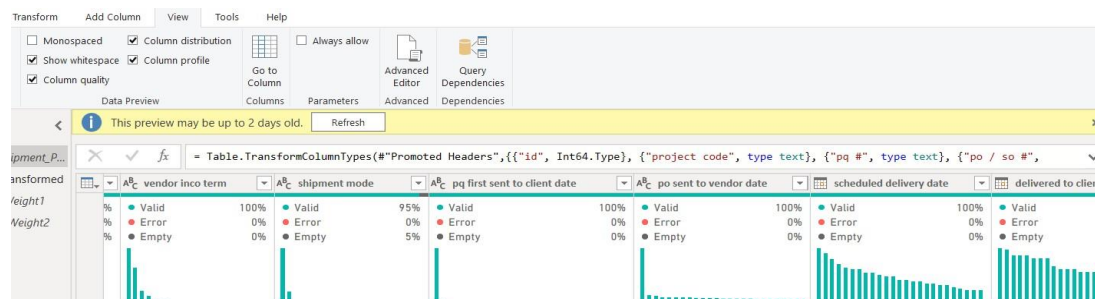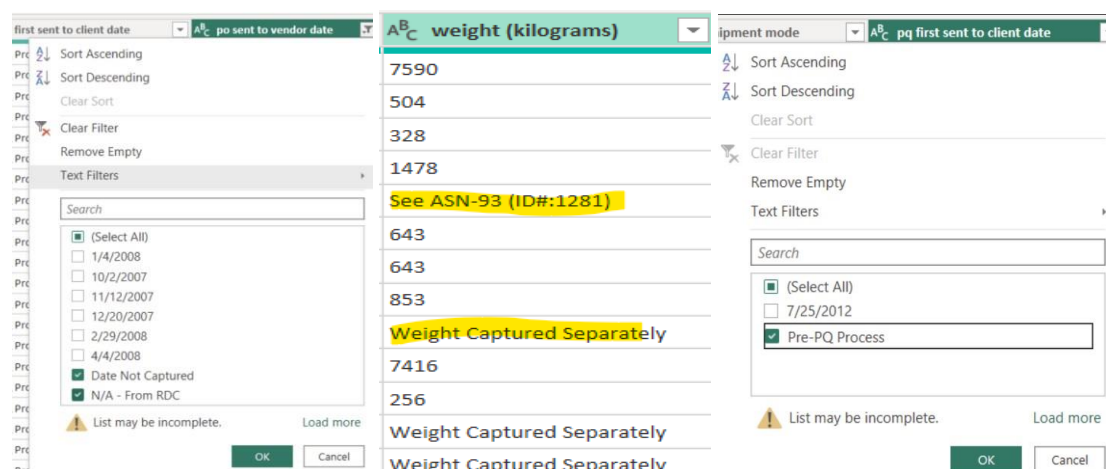
Exhibit -1

Exhibit-2

3. Once the issue is identified, resolved the data using various transformation methods. For column
   Shipment mode – Blank values are replaced keyword by "Missing Value"
   pq sent to vendor date – Included default dates 01/01/9999(Date Not captured) and
   09/09/9999(N/A – from RDA)
   pq first sent to client date – Included default dates 01/01/9999 (Pre-PQ Process)
   dosage - Blank values are replaced keyword by "Missing Value"
   weight(kilograms) – Used a transformation logic and explained below to solve 2 problems
   1. Reference scenario
   2. Categorical column data

   Based on this weight column if it says SEE ASN "ID" I reviewed the ID column from the dataset
   and identified that ID#961 should have value 3509 (highlighted in blue) instead of See ASN-3562
   (ID#:960) and replaced **"Weight Captured Separately"** with -1 value.

Initial Dataset

ID	Weight(kilograms)
44	328
961	See ASN-3562 (ID#:960)
46	See ASN-93 (ID#:1281)
62	Weight Captured Separately
1281	479
960	3509

Transformed Dataset

ID	Weight(kilograms)
44	328
961	3509
46	479
62	-1
1281	479
960	3509

Using Power Query made data transformation changes

- Using **Split delimiter function "**(ID#:") - extracted the data from actual weight data column and created a new table with just ID and weight for only the rows which has **"See ASN-#"** and post that mapped that with id and got values of categorical columns
- Using **Left join** combined the new column with the original table and updated the columns using Power Query Transformation which is in Step 2
- Using DAX formula combined the two tables and created a new column as weight_transformed. Also, new column weight_flag having Weight Captured Separately is updated to -1 and to avoid these values on graphs used weight_flag (**Exhibit 3**)

<div align="center">

Exhibit 3

</div>


```
1  weight_transformed = if(Supply_Chain_Transformed[weight (kilograms)-new] =BLANK(),Supply_Chain_Transformed[weight (kilograms)],if(Supply_Chain_Transformed[weight
   (kilograms)-new]= "Weight Captured Separately","-1",Supply_Chain_Transformed[weight (kilograms)-new]))
```

signation	weight (kilograms)	freight cost (usd)	line item insurance (usd)	weight (kilograms)-new	weight_transformed	weight_flag	Late Delivery	Late Deli
True	55	913.15			55	True	ON TIME	
True	4436	18375.7			4436	True	ON TIME	

```
weight_flag = if(Supply_Chain_Transformed[weight_transformed] < 0,FALSE(),TRUE())
```

Below is the explanation of the theory in excel (**Exhibit 4**)

Exhibit 4

Table 1
Sample data - Actual Table

ID	weight
1	100
2	300
3	abc-2
4	xya-1

Table 2

ID	weight new		Transformation	Power Query
3	2	2	300	
4	1	1	100	

Table 3
replaced the Value of 2 with 300 and 1 with 100 based using power query

ID	weight new
1	
2	
3	300
4	100

Final Table value using DAX formula using measure

	org	Weights new	weight - Transformed
1	100		100
2	300		300
3	abc-2	300	300
4	xya-1	100	100

4. There are few unwanted columns, instead of deleting them I have just ignored them for visualizing the data.
5. Created two new columns Late delivery and Late delivery days to verify the late delivery of the product this is again created using DAX formal as represented below **(Exhibit 5)**

Exhibit 5

```
Late Delivery =
var ld = DATEDIFF(Supply_Chain_Transformed[delivered to client date],Supply_Chain_Transformed[scheduled delivery date],DAY)
return if(ld >= 0,"ON TIME","LATE")
```

| gnation | weight (kilograms) | freight cost (usd) | line item insurance (usd) | weight (kilograms)-new | weight_transformed | weight_flag | Late Delivery | Late Delivery Days |

```
1 Late Delivery Days =
2 var ldd = DATEDIFF(Supply_Chain_Transformed[delivered to client date],Supply_Chain_Transformed[scheduled delivery date],DAY)
3 return ABS(ldd)
```

| ignation | weight (kilograms) | freight cost (usd) | line item insurance (usd) | weight (kilograms)-new | weight_transformed | weight_flag | Late Delivery | Late Delivery Days |

Post transforming the data, now data is ready to be visually represented to address key questions regarding target markets, transportation methods, dosage forms, vendor popularity, brand popularity and product growth over years' time.

Visualization of the data and understanding the results of the graph

Created dashboards to address each key questions, to maintain a consistence across dashboards and to follow a standard view, have add Title and created filters for Date, Year, Country and Product Group on each dashboard so client can easily understand & categorize and visualize the data in an effective and efficient way.

Let us go further on one each graph and understand the result/outcome of each dashboard.

Representation of Target markets by Product Group in the dataset
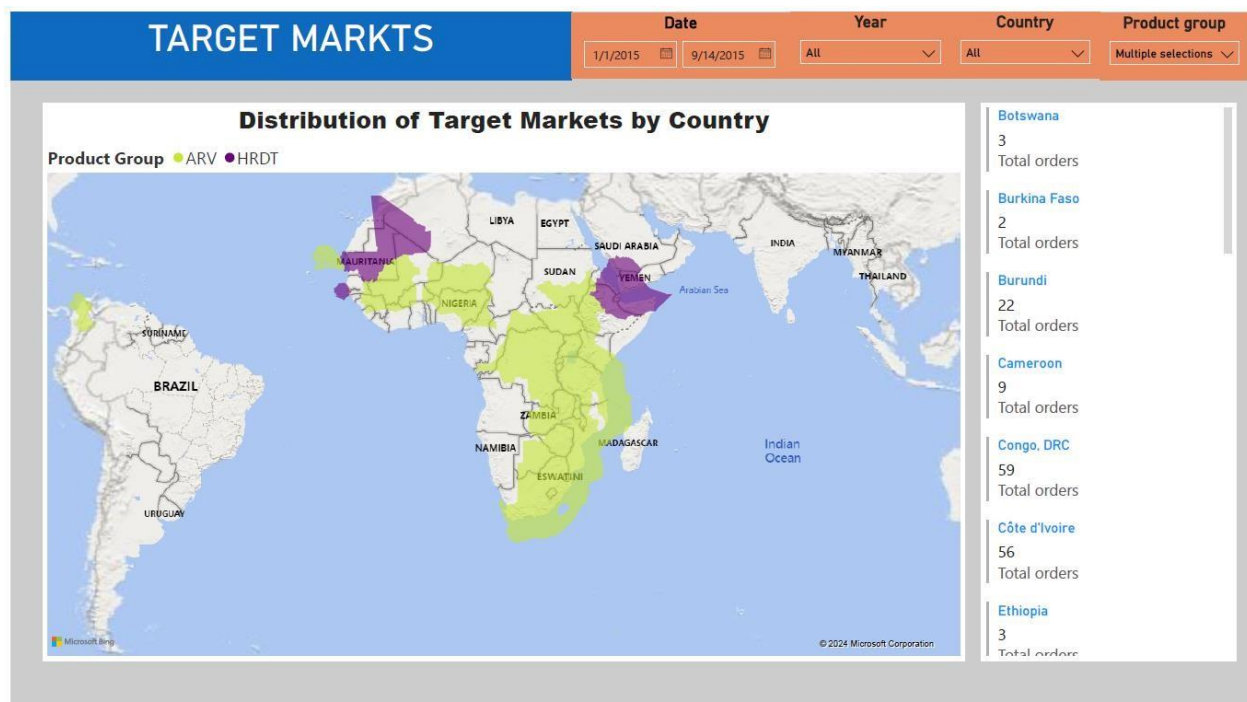
The Map clearly highlights and differentiates the countries based on ARV and HIV lab shipments that were transported across the countries. We can filter product groups across each country and year; and can identify the number of orders and the regions distributed across countries.

Result: Based on this graph, it clearly shows the data is available from May,2006 till September,2015 and key target markets are South Africa, Nigeria, Côte d'Ivoire and others which can viewed in the graph.
Takeaway for the Client: Understanding the Target market across geography plays an important role in shipments. Based on this, resources(logistics) can be effectively utilized and planned to ensure on time delivery.

**Representation of Time Period in the dataset**
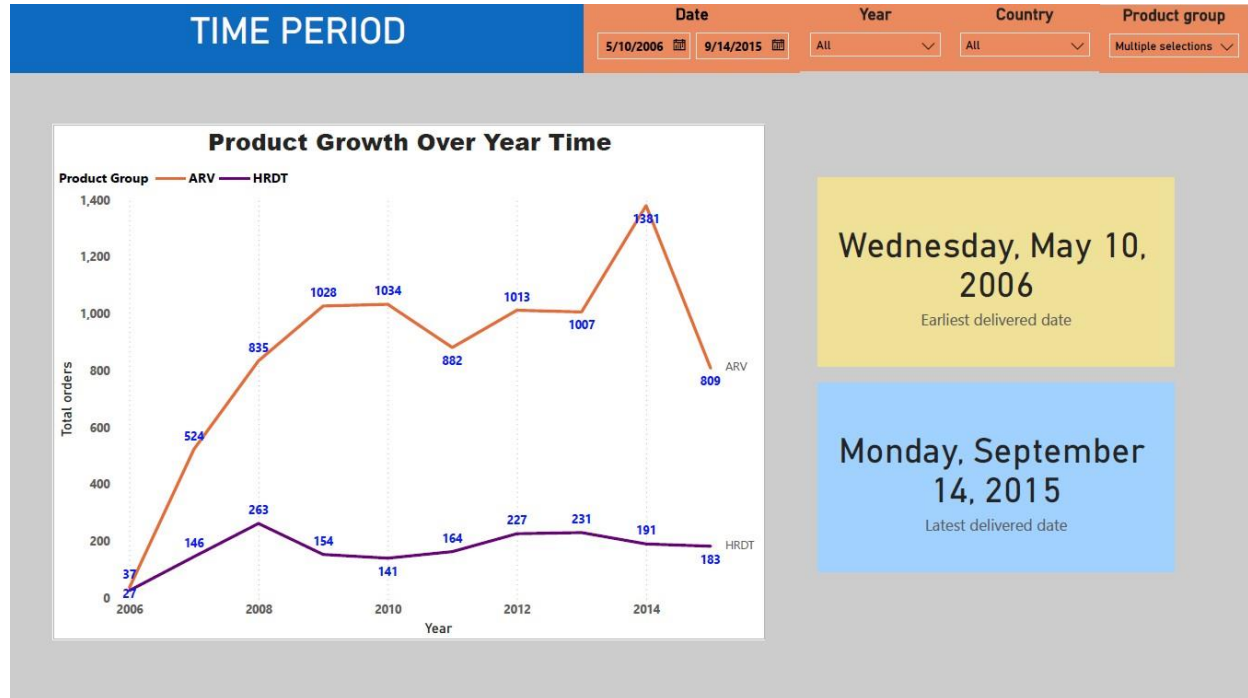
Timeline graph is very straight forward, it clearly shows the total orders over the years from May 10, 2006 (Earliest delivered date) to August 28, 2015 (Latest delivered date)

Result: The Product for ARV grew in 2010 with total orders as 1034 and there is drop again in 2011 to 882, similarly, for HRDT there was a price in 2008 of 263 but again in 2010 it drastically dropped to 141 in 2010.For ARV 2014, was the highest product growth while HRDV's highest was in 2008 after that there is not much rise. This clearly shows that HRDT must improve its growth or take necessary measures.
Takeaway for the Client: This helps in understanding the historical trends of each product and showing shipment performance over the years and take necessary measures to improve the product service or to completely stop the product based on necessary reviews.

Representation of common Transport Shipment Method in the dataset
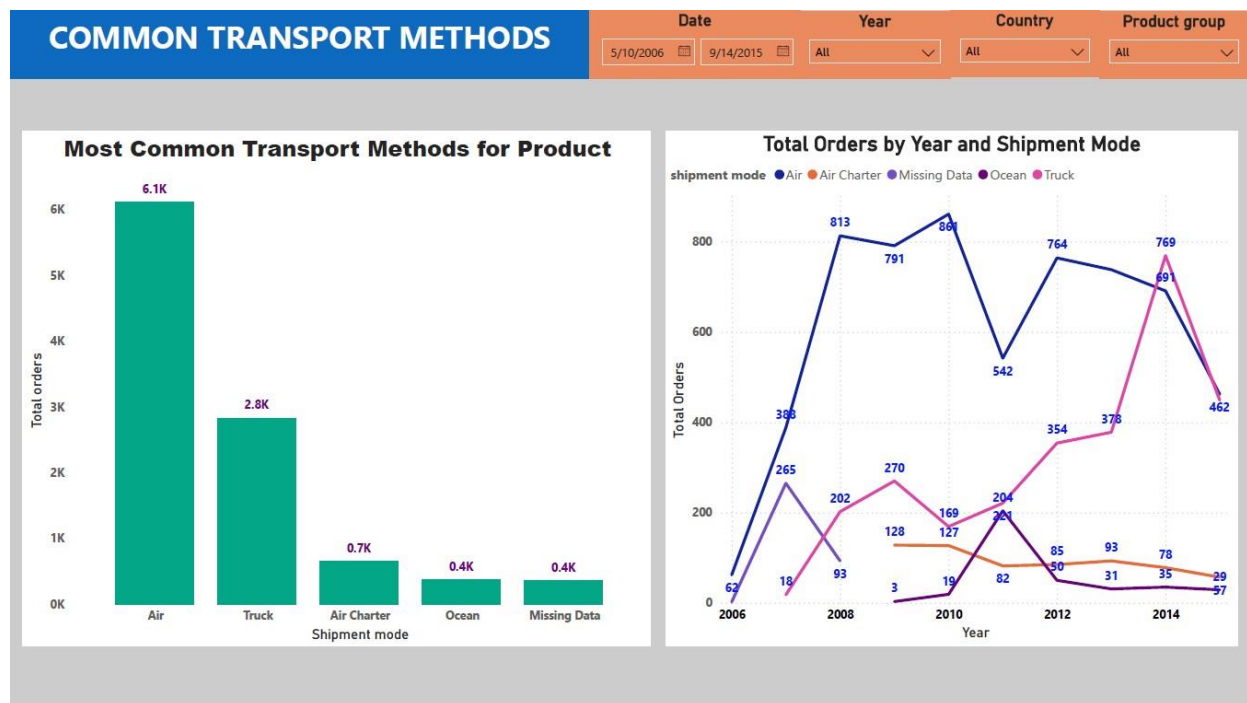

The bar chart shows the distribution of the transportation of product and also shows a line chart which shows the transportation of product over a period of time.

Result: It clearly shows that products transported by Air are more followed by Truck i.e. 6100 and 2800 respectively. While we do have some missing data which might change the stats number by 0.3%
Takeaway for the Client: Have a clear understanding of transportation method will help in utilizing that mode of transport for timely delivery and can reduce other ways of transportation logistics, which will help in reducing the unnecessary expenditure.

Representation of Common Dosage Forms of the product in the dataset
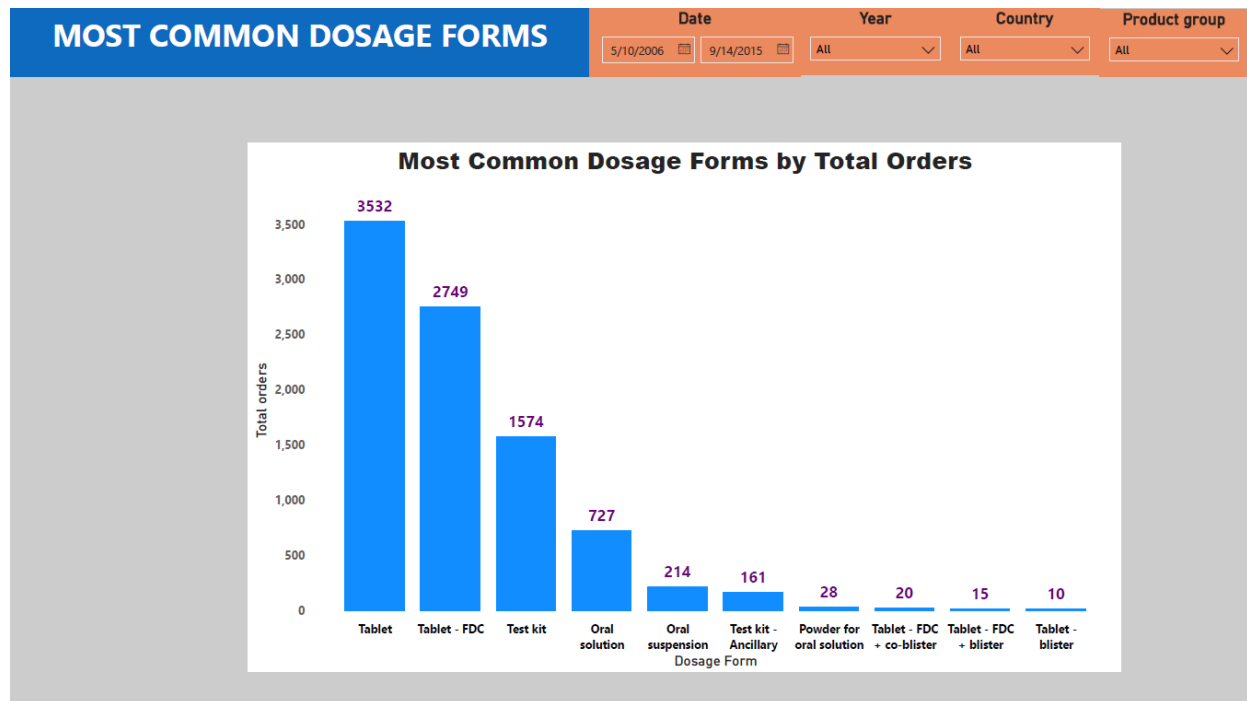

The bar chart shows the top 10 dosage forms of the products ordered. We can again review this data based on country filter and product Group filters.

Result: Tablet dosage stands out first and is highly adopted with 3532 orders overall while tablet-FDC and Test Kit fall with 2749 and 1574 orders keeping at second and third position respectively.
Takeaway for the Client: Dosage form is a very crucial part of analysis as this helps in managing the inventory and to ensure an optimized supply chain is implemented. In nutshell understanding the demand and supply accordingly to save the cost and effort.

Representation of Average Weights (in Kilograms) of the product in the dataset
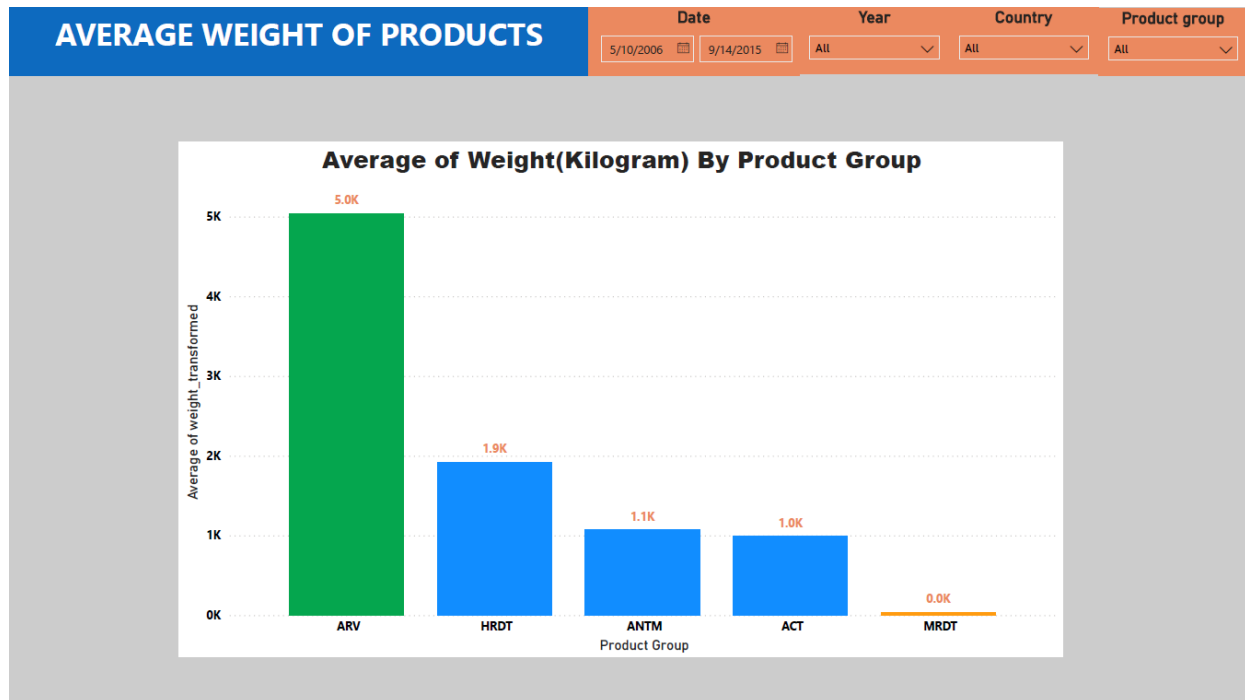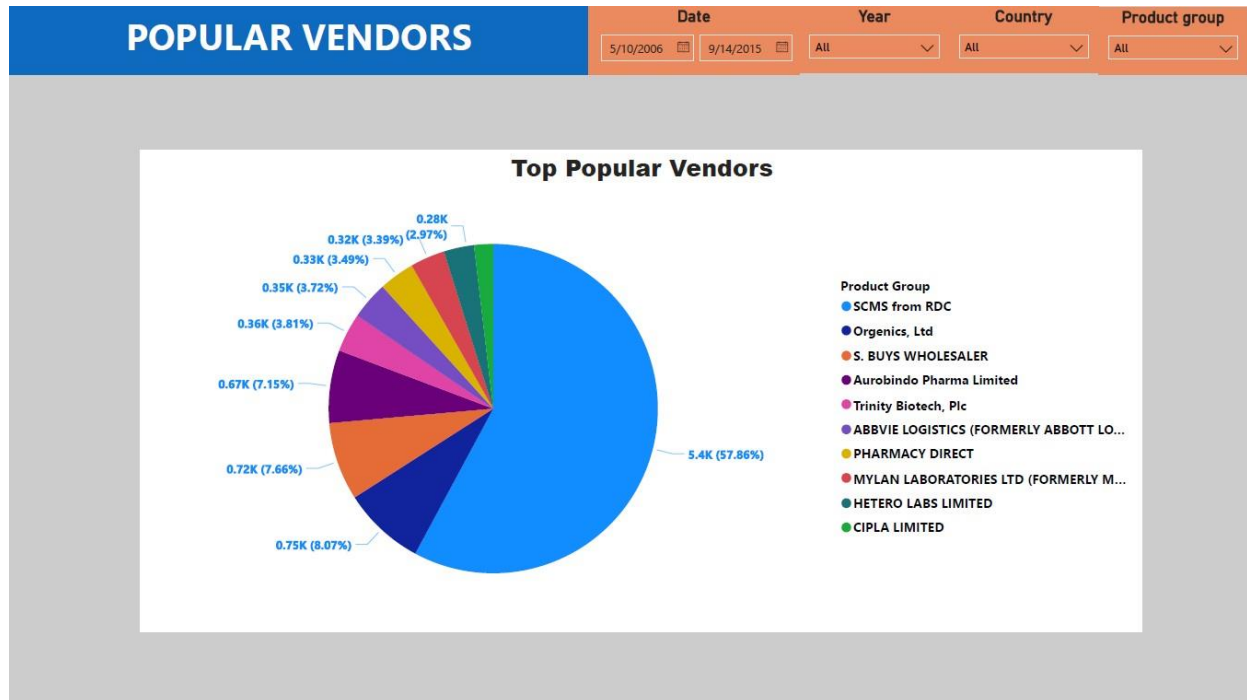

The bar chart shows the average weight of the products. Based on each country and year the average weight of the product can be filtered and reviewed. Also, based on EXHIBIT 3,4,5 transformations in data the above graph is represented.

Result: ARV products are the highest amongst the 5 products with an average weight of 5 Kilograms, HRDT bragging the second position with 1.9 Kilograms, ANTM at 1.1 kilograms, ACT at 1kilogram and MRDT is the least clearly shows it was not sold.
Takeaway for the Client: Average weight of the product is another key factor in planning the logistics, packing costs, increasing or decreasing the product supply and consideration of bulk transportation based on the location.

**Representation of Popular vendors of the product in the dataset**

The pie chart shows the top 10 vendor distribution of the product. Based on the each country and year the average weight of the product can be filtered and reviewed.

Result: SCM from RDC are the top vendors having 5400 orders having 57.86% of the market products. CIPLA LIMITED is the least amongst other vendors which has only 175 orders i.e. it holds only 1.87% of the market share.

Takeaway for the Client: Vendors help in strategic partnerships and consistent supply chain in the market. This is directly dependent on product demand and sales.

**Representation of Brands of the product in the dataset**
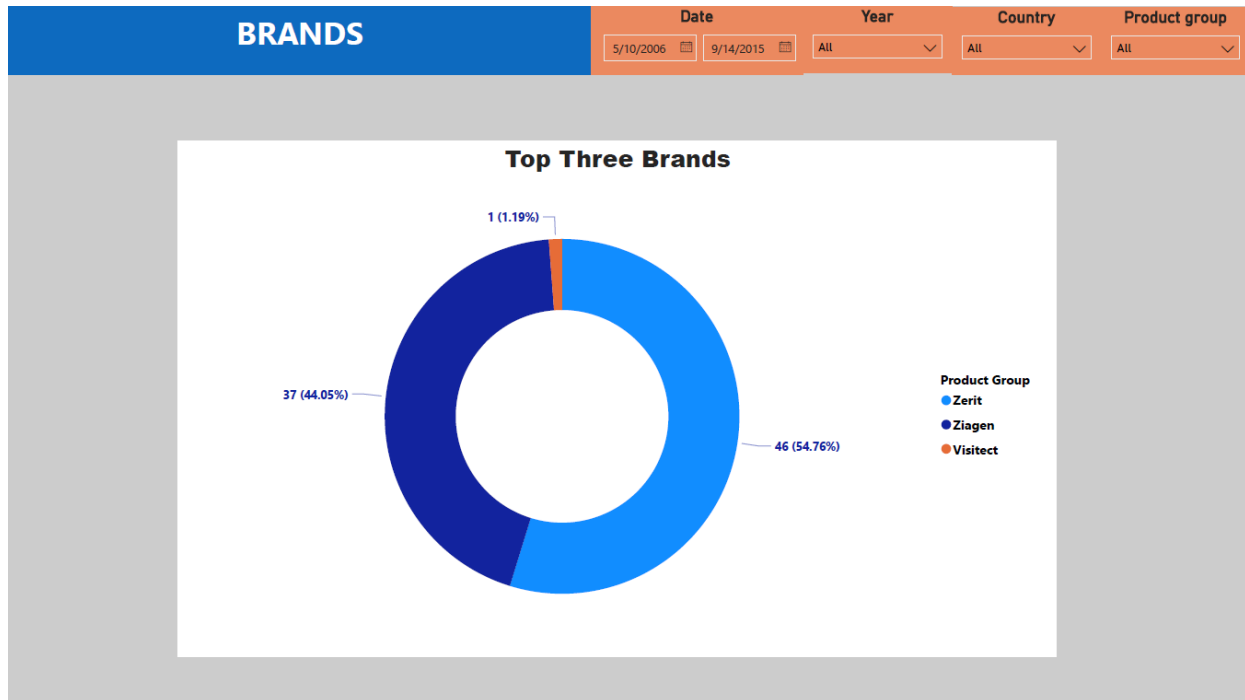
The donut chart shows the top 3 most popular brands of the products based on the number of orders from customers across regions.

Result: Zerit is the top brand while visitect is the least having 56.76% and 1.19% of the market share respectively.

Takeaway for the Client: Brand is a very important criterion of supply, as customer check the high demand brand and they will buy accordingly. Here, this helps in promotional supply of the product

**Representation of Late Delivered Products in the dataset**
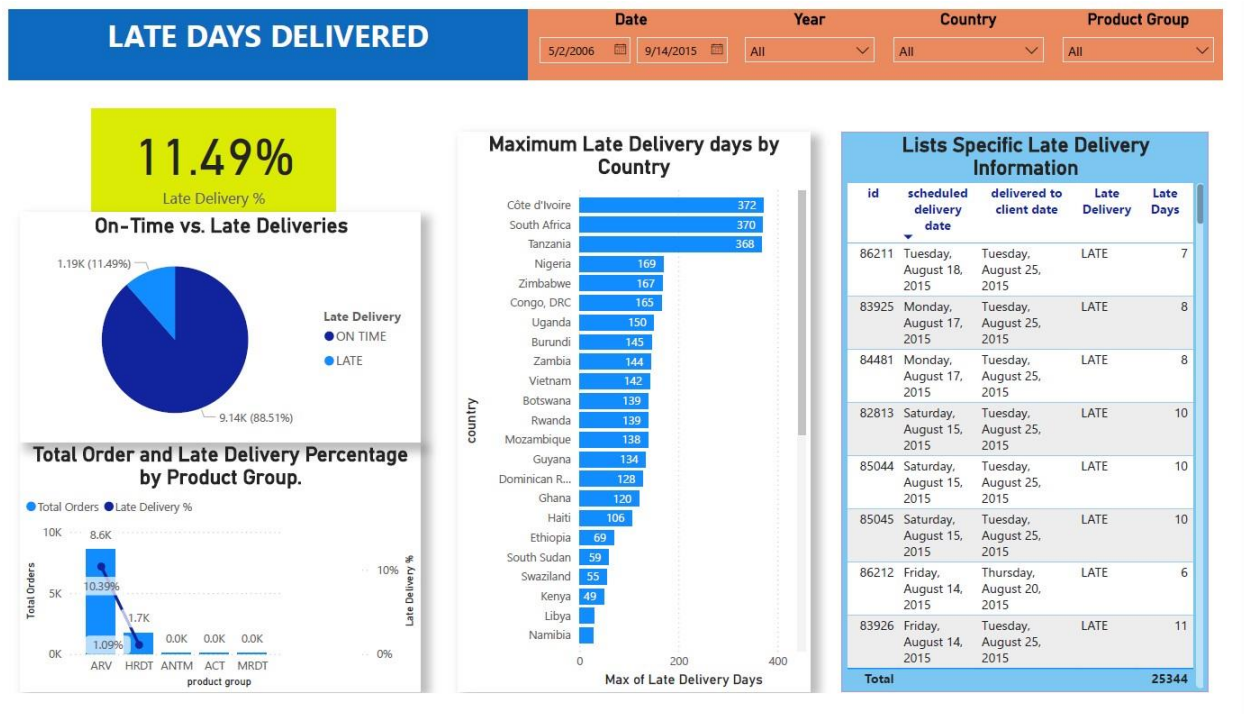
The dashboard has multiple graphs which show the delay delivery details days, percentage. It provides a list of specific delay dates.

Result: 11.49% of the products are delivered late to the destination while 88.51% were delivered on time. ARV products are mostly delivered late to the destination, i.e. out of 8.6K orders 10.39% of the products were delivered late to the destination while HRDT out of 1.7K 1.09% were the delays.
Takeaway for the Client: Delay in delivery insights will help in increasing the transportation service, understanding the client concerns, this will also help in understanding the demand for the product and reducing the delay for better customer service.

Representation of a relationship between Unit Measure, Unit Pack and Unit Price in the dataset
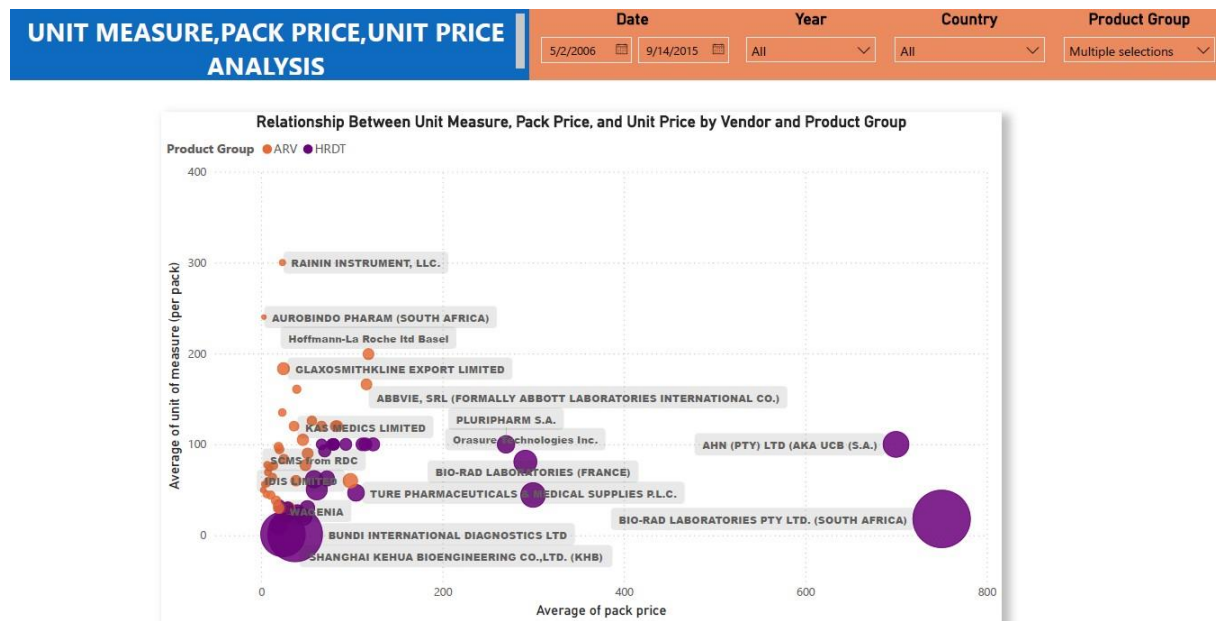
Scatter plot is the visual representation of the relationship between pack price and unit price. Based on the data it was clear that unit price values are really low. Ideally, based on the dataset unit price is calculated based on Pack price by unit of measure. While pack price can be calculated based online item value by line item quantity.

unit of measure (per pack)	line item quantity	line item value	pack price	unit price
30	19	551	29	0.97
240	1000	6200	6.2	0.03
100	500	40000	80	0.8

<div align="center">Unit Price = Pack Price/Unit of measure Ex. 29/30 = 0.97</div>

<div align="center">pack Price = line item value/line item quantity Ex. 551/19 = 29</div>

Unit of Measure: Number of individual items in one pack

Pack Price: Total cost of one pack, calculated by dividing the total cost by the number of packs

Unit Price: Cost of one item within a pack, calculated by dividing the pack price by the number of items in the pack It helps in determining the cost structure in the dataset and how the cost of individual items are related to the total pack cost and the number of items per pack.

Result: Larger points depicts higher unit prices. Vendors like BIO-RAD LABORATORIES PTY LTD. (SOUTH AFRICA) and AHN (PTY) LTD (AKA UCB (S.A.) have higher pack prices and unit measures

Takeaway for the Client: Understanding the price will help in taking a decision to increase or decrease the purchase price, quantity and pack size.

Additional Data and Information for supply chain dashboard

Customer Feedback on delivery, Delay reason, Vendor supply, Product tracking information, returned products, supplier delay information, this information will be helpful in understanding the data in terms of delay, number of purchases for one particular vendor and improving the quality of delivery and product based on customer inputs.

Conclusion

Based on the Data and Visualization of the data for ARV and HIV lab shipments across regions. ARV has been on top in terms of Products, transportation, dosage forms, vendors brands and even in terms of late delivery. It clearly shows the areas of improvements that need to be made for ARV and HIV. Using additional features mentioned above can help in putting the supply chain in a better state as it will give a deeper understanding of the product metrics. Since it's a healthcare industry it's very important to maintain even the minute details like review comments, product storage operations which will help in improving the product supply and take corrective decisions.

References

USAID Dataset: Doby. (2023). Supply Chain Shipment Pricing Dataset. Dataset. USAID Development Data Library. Retrieved from https://data.usaid.gov/d/a3rc-nmf6

Data.gov Dataset: U.S. General Services Administration. (n.d.). Supply Chain Shipment Pricing Data. Data.gov. Retrieved June 4, 2024, from https://catalog.data.gov/dataset/supply-chain-shipment-pricing-data-07d29

USAID. (n.d.). Supply Chain Shipment Pricing Dataset. Retrieved June 4, 2024, from https://data.usaid.gov/HIV-AIDS/Supply-Chain-Shipment-Pricing-Dataset/a3rc-nmf6/about_data