MSIS 2629- Dashboards

Final Project
(President's Malaria Initiative)

Submitted By:

Bhagyashri Badgujar
Kashish Parikh
Priyansh Bendre
Srivatsav Vuddamarry
Vishwadeep Dutta

Data Description:

Malaria_death_rates_by_age:

This data gives us the exact number of deaths caused due to malaria every year from 1990 to 2018 across all the countries in the world. This data was used to get the required information to check if these deaths had any correlation with the PMI program.

Code: Country code.
Entity1: Country.
Year 1: year in which the data was collected.
Medical_Deaths: Death caused due to malaria
Percentage_change: Percent change in deaths.
15-49 Years Old(Per 100,000): Total number of deaths caused in this age group
5-14 Years Old(Per 100,000): Total number of deaths caused in this age group
50-69 Years Old(Per 100,000): Total number of deaths caused in this age group
70+ Years Old(Per 100,000): Total number of deaths caused in this age group
Age-standardized(Per 100,000): Total number of deaths caused in this age group
All ages(Per 100,000): Total number of deaths caused in this age group
Under-5s(Per 100,000): Total number of deaths caused in this age group
Medical Deaths: Total number of deaths.

annual_number_of_deaths_by_cause:
This data describes the total number of deaths occurring due to all the diseases across all the countries. It was used in this analysis to get the total number of deaths happening all around the world and compare it with the deaths caused due to malaria.
Code: Country code.
Entity: Country.
Year: The year in which the data was collected.

Cleaning:

Tableau Prep 1: PMI_Success_Data

1) **Step 1:** Load the annual - number of deaths -by cause data set. (Link for that dataset)
2) **Step 2:** Load the malaria death-rate-by-age data set. (Link for that dataset)
3) Step 3: Load the Share-of-death-by-cause dataset. (link)
4) **Step 4:** Inner Join all three datasets on Year and Code.
5) **Step 5:** Create a calculated field “Medical_Deaths” that included a number of deaths by any Medical Condition.
6) **Step 6:** Removed all unwanted fields and outliers.

Tableau Prep 2: PMI_Delivery_Death

1) **Step 1:** Load the PMI_DELIVER_dataset.
2) **Step 2:** Load Malaria_death_rates dataset.
3) **Step 3:** Clean 1: the dataset by performing the following steps:

- Changed data type for multiple fields.
- Filtered null values.
- Transformed UOM to numeric only and multiplied it with Quantity to get Total_Quantity.
- Created calculated field for all dates.

Name	**Calculated field**
R_Arrival_Date	if YEAR([Arrival Date])>2012 and YEAR([Arrival Date])<=2018 then [Arrival Date] END
R_Shipment_Date	if YEAR([Arrival Date])>2012 and YEAR([Arrival Date])<=2018 then [Arrival Date] END
Total_Quantity	IF ISNULL([Uom]) THEN [Quantity] ELSE ([Uom]*[Quantity]) END

4) **Step 3:** Clean 2: Change data type of year for join.
5) **Step 4:** Use Inner join on R_Arrival_Year and Year.

ASSUMPTIONS FOR THE PROJECT:

Tableau Desktop

Goal 1 : Is the PMI Successful or not?	
Stakeholder	Government
Questions	What is Meant by Success? Long term goal of PMI is to eradicate Malaria from the focus countries in the 20-25 year period.
	How do you measure Success? Significant reduction in the number of deaths in the focus countries would be considered as successful.
Metrics:	Change in Death rates from 2005 (PMI founding year) to 2016 (most recent official complete dataset)

Descriptive Analysis:

Our approach towards doing the descriptive analysis was first to find out how the trend of the deaths have been across years, using various databases before and comparing it with the deaths when PMI was initiated focussing on 3 countries initially in 2005.

Focus Country Deaths:

- The aim of this graph is to represent whether the PMI is successful or not.
- The x-axis contains deaths occurred in all ages(per 100,000) and the y-axis contains the years from 1996 to 2010.
- The graph is divided into 3 parts:

1. 90's Malaria
2. Post_PMI
3. Pre_PMI.

- There has been a significant decrease in deaths from 2001-2005 by 14.38% but after the introduction of PMI program we can see that the deaths due to malaria decreased by 66% in the years 2006 - 2010.
- This graph can therefore help us to conclude that PMI was a success.

- Calculated Fields used:

Name	Calculation
percentage_change	IF [Year]>1995 and [Year]<=2000 THEN "90's Malaria" ELSEIF [Year]>2000 and [Year]<=2005 THEN "Pre_PMI" ELSEIF [Year]>=2005 AND [Year]<=2010 THEN "Post_PMI" END

- Sets:90's Malaria(1996-2000), Post_PMI(2006-2010), Pre_PMI(2001-2005).

Predictive Analysis:

- From the sheet “Focus country death”, we can say that as the program was successful in its past years, if it is continued in the same way then the deaths caused by malaria could be reduced to a very great extent.

Stakeholder: Supplier

Goal: How can the Supply Chain and logistics can be made efficient for the better results of the PMI?	
Stakeholder	Supplier
Questions	What is meant by better? According to the graphs and the average date diff if the supply is sent through ocean and is reaching in the wrong quarter then that can be made better.
	What all come under supply chain and logistics? Mode of transport, what kind of shipments come through what mode, and what is the average date difference between the products for each country.

Metrics:	When is the best quarter the products can be shipped and through what mode of transport.

Descriptive Analysis:

The goal of this analysis is to find how to make the Supply chain better, by focussing on the shipment mode for each product - subcategory. Analysing the date difference between the Shipment date and the Arrival date for each quarter across each country and for each mode.

Z-Score:

The Z- score is calculated to see the distribution of date difference in accordance with the number of records. From the Z-score we can say that most of the date difference lies on 0 - 1.65.

- **Calculated Fields used:**

Name	**Formula**
z_score(bin)	Size of bin: 0.55
Z_Score_Country (bin)	Size of bin: 0.55
Cleaned_Date_Diff	IF [Date_Diff] >0 then [Date_Diff] ELSE 0 END
country_avg_date_diff	{FIXED [Country]:AVG([Valid Date Difference])}
Country_Avg_Diff	{ FIXED [Country]:AVG([Date_Diff])}
Country_Std_Deviation	{ FIXED [Country]:STDEV([Date_Diff])}
Date_Diff	DATEDIFF('day',[R Shipment date],[R Arrival Date])
Valid Date Difference	if [Date_Diff] > 0 then [Date_Diff] END
valid_date_diff	if [Date_Diff]>0 THEN [Date_Diff] END
Z_Score_Country	([Date_Diff] - [Country_Avg_Diff])/[Country_Std_Dev

	iation]
z_score	([Cleaned_Date_Diff]-[Country_Avg_Diff]) / [Country_Std_Deviation]

Shipment Arrival vs Total_Quantity:

- This graph was designed to compare the average total quantity across quarters from 2012 to 2016.
- The x-axis consists of Average total quantity and the y-axis contains R Arrival Date.
- The graph is separated into 4 parts each containing the average of quantities across the four quarters for the years 2012 - 2016.
- We can see that the supply of medicine has been increased gradually in Q2 and Q3 over the years. Generally, in sub-Saharan regions are the months December - April with the highest occurrences of malaria as known from the source on WHO data. Inference from the graph: It is efficient to make sure that the products arrive in the Q2 and Q3.

Shipment Arrival vs Death Graph:

- The aim of this graph is to show the stats for how many orders arrived per quarter yearly.
- The x-axis contains the no. of yearly deaths per 100k. each quarter from 2013-2016.
- The y-axis denotes the Number of deaths (per 100k). This death count covers all the age groups(0-70+ years)
- The graph is significant enough to show that the death counts is lowered after 2015. Reason behind that is the supply in 2013 and 2014 for 1st quarter was low compared to the other quarters.
- As there was surplus supply in quarter one of 2015 and 2016,we can see that the death rate gradually decreased

Inference from the Graph:

This can be understood from the Arrival vs Total Quantity that in year 2013 and 2014 there wasn't enough supply of Malaria medicines in Q1 of 2013 and 2014. From the data it is observed that the prevailing season for Malaria is in extreme cold i.e from December to April.

Country Average date diff:

- The goal of this sheet is to describe in which country, how much is the delay in transportation of supplies via different mode of transport.
- A continuous data is made on valid_date_diff.
- The y-axis consists of distinct count of valid_date_diff and the x-axis consists of different countries present in the database.
- Color filter is applied on Mode.

The graph shows the average of date difference for each country present in the data set and we are providing the average date difference for the various modes of transport.

Inference from the graph:

This graph helps the supplier to understand the average date difference for each country along with the mode of transport so that he can start the shipment on time based on the earlier analysis.

Prescriptive:

- It can be deduced that the number of deaths is decreased when there is surplus supply in Q2 and Q3 each year. The malaria spread is highest in the months of December - April (Extreme cold or extreme hot climatic condition).
- Thus, the medicines should be sent in Q2 and Q3 each year for the best results.

Sheet 17

- The Graph is used to show the country with the count of death in all age groups (0-71+ years) with the average date difference from Shipment date to arrival date.
- The x axis shows the number of deaths in all age groups.
- The y-axis shows the country for which the user need the details of death count and average date difference.
- Here we have applied for a filter of Action: Country and Avg(Valid Date Difference) as a label.

Sheet 19

- This graph shows the country wise analysis of the Product which is distributed the most in that country and who is the manufacturer.

- The x-axis shows the product count.
- The y-axis show countries in the dataset.
- Here the dataset is having the fields
- Max(Product_max) : Product_max ({ FIXED [Manufacturer (long)],[Product]:MAX([Total Quantity])}) and taking the top rows from the calculated field.
- Country : The countries in the dataset.
- As a filters we have used Supply Type, MAX(Product), MAX(Manufacturer(long).

Qty vs Mode Graph:

This graph shows the total number of products shipped based on the mode of transport. And also we are categorising the Product Subcategories into
The product subcategories are classified into Essential and Non-Essential supplies.

Assumption: After analysing the data we understood that products such as mosquito nets, medical equipment etc are usually shipped through ocean mostly and the main medicines are shipped through air.

Essential supplies :
1. ALu Generic
2. AS/AQ FDC
3. Coatem
4. Essential Medicines malaria
5. Essential Medicines PH
6. Malaria Pharmaceuticals.
7. Rapid Diagnostic Test Kit

Non-Essential Supplies :
1. Long Lasting Insecticides Treated net.
2. Malaria Misc. Commodities.
3. Medical Equipment

World Map graph:
This graph shows that the countries in the dataset with total deaths in all age groups.

PMI Analysis Overview Dashboard:

- This is the dashboard which shows the death rate reduction after initiating the PMI.
- The dashboard also shows the Qty vs Mode graph and deaths across the countries as per the date difference.

Conclusion:

In conclusion, we analyzed the success of the President's Malaria Initiative by correlating the official PMI dataset with the datasets available with other organizations on deaths and making appropriate assumptions.

We started with finding out whether to continue the PMI initiative or not. From the data that we got, we found out that the PMI is highly successful because the number of deaths in the focus countries has sharply reduced over the period when the PMI started its operations which has helped save thousands of lives.

We then moved forward to find out any issues that the PMI may have been facing or would have gone through over the last 15 years of its operations. We saw that there were some irregularities in the data, a handful of outliers and multiple records of delay in the shipment. Therefore, we decided to analyze from the suppliers point of view what has been happening since the last 5 years.

We saw that the most shipments reached the destination in the Q2 and Q3 and supporting this fact we also saw that the death for all ages, decreased from Q3 and Q4 of 2015 and 2016 in contrast to 2013 and 2014. Which led us to conclude that the shipment arrival delay was directly proportional to the deaths in focus countries. We also saw that most of the non-essential product (supplies) were sent via ocean which also led to a delay in the delivery and resulted in demands not being met.

We also saw that on an average it takes 8 days by air and approximately 48 days by ocean for the medicines to reach a particular destination.

We prescribe that if the shipment has to be sent via Air then it has to be sent before 8 days and 48 days via ocean for efficient supply chain and increase in the success rate of the PMI. We also suggest the supplier that if all the medicines were supplied by Q1 and Q2 instead of Q2 and Q3 could help decrease the death rate.

Also if the manufacturers were able to manufacture, stock and supply the average number of quantity before hand (Q1 and Q2). This can also help in possible decrease in the death rates.

Dashboard Link:
https://public.tableau.com/views/PMIAnalysisOverview/PMIAnalysisOverview?:embed=y&:display_count=yes&publish=yes&:origin=viz_share_link

References:
https://data.usaid.gov/Malaria-/PMI_GHSC_dataset/ax4e-gv2e
https://data.usaid.gov/Malaria-/PMI_DELIVER_dataset/k8ym-fihd
https://data.usaid.gov/Malaria-/PMI_PSM_dataset/h7qj-2w82
https://ourworldindata.org/malaria (Datasets)
https://www.cdc.gov/malaria/malaria_worldwide/impact.html