**Global Trade Analysis Project**

**Domain – International Trade**

**By Sridhar M M**

***Project Over View***

To create interactive Tableau dashboard for An Indian manufacturing company who wants to launch a new business unit focusing on global trade and logistics.

The new business unit is targeted for major countries such as USA, Canada and Australia.

Analyze to find out and focus on potential commodities for trade.

Show countries that has most profit-based commodities for trade, products that are most traded on Quantity in Billions

Yearly based trade profits (USD) in Billions.

***Methodology:***

Raw **data is imported Using R** required Dataset has been created.

Used **Structure Function str () in R** to see if the data is compactly provided and if the relevant information of the Global Trade dataset provided.

Used **Summary ()** function of R to get more detailed **statistical information** of each column.

I have cleaned the data set by **treating the Null and Zero Values records**, deleting them and verifying if the remaining observations are good sample to work on.

At last the data set has been **imported to tableau for better Visualization**.

Observations are noted for better business decision making that would be presented for the business panel.

***Dataset***

We have an Excel file (Dataset\_Int\_business.xlsx) of import and export data of 12 major commodities including the commodity name, year (1988 to 2017), Trade amount in USD, quantity

Install the required Packages in R for our Analysis

install.packages("readxl") #to read Excel file#

install.packages("xlsx")

install.packages("writexl") #to copy file from R to Excel#

library(xlsx)

library(readxl)

library(writexl)

Using R data is imported and I have created a Dataset “GTA\_Data” for further Analysis…

[1] GTA\_Data<-read\_excel("Dataset\_Int\_business.xlsx") # Import file into R from Excel

|  |
| --- |
| ***Explore & cleaned data.frame Global Trade.***  ***Step 1:*** Let us find the [DIM]ension(number of rows and columns in a matrix) of the dataset  > dim(GTA\_Data)  [1] 59090 10 |
| The “Global Trade” Dataset consists of 59090 observations of 10 attributes (Fields).  ***Step 2:*** Let us see what are the 10 attributes (Fields) names  objects(GTA\_Data)  [1] "category" "comm\_code" "commodity" "country\_or\_area" "flow" "quantity"  [7] "quantity\_name" "trade\_usd" "weight\_kg" "year" |

***Step 3:*** I have used **Structure Function str()** to see the compactly provided & relevant information of the Global Trade dataset provided

> str(GTA\_Data)

Classes ‘tbl\_df’, ‘tbl’ and 'data.frame': 59090 obs. of 10 variables:

$ country\_or\_area: chr "Australia" "Australia" "Australia" "Australia" ...

$ year : num 2016 2016 2016 2016 2016 ...

$ comm\_code : num 10111 10119 10119 10119 10120 ...

$ commodity : chr "Horses, live pure-bred breeding" "Horses, live except pure-bred breeding" "Horses, live except pure-bred breeding" "Horses, live except pure-bred breeding" ...

$ flow : chr "Export" "Re-Import" "Export" "Import" ...

$ trade\_usd : num 1.29e+08 4.93e+06 1.18e+07 9.04e+07 5.85e+04 ...

$ weight\_kg : num 900450 47240 153587 1082493 5805 ...

$ quantity\_name : chr "Number of items" "Number of items" "Number of items" "Number of items" ...

$ quantity : num 1882 104 276 2073 416 ...

$ category : chr "01\_live\_animals" "01\_live\_animals" "01\_live\_animals" "01\_live\_animals" ...

**Structure Function str() shows the** ., “number of observations”, “number of variables”, “names of  
each column”, the “class of each column”, and “sample values from each column”.

***Step 4:*** I have used **Summary ()** function to get more detailed statistical information of each column.

summary(GTA\_Data)

country\_or\_area year comm\_code commodity flow trade\_usd

Length:59090 Min. :1988 Min. : 10111 Length:59090 Length:59090 Min. :1.000e+00

Class :character 1st Qu.:1996 1st Qu.: 30559 Class :character Class :character 1st Qu.:6.785e+04

Mode :character Median :2003 Median : 70690 Mode :character Mode :character Median :8.304e+05

Mean :2003 Mean : 64722 Mean :3.107e+07

3rd Qu.:2010 3rd Qu.: 90620 3rd Qu.:7.308e+06

Max. :2016 Max. :121490 Max. :1.373e+10

weight\_kg quantity\_name quantity category

Min. :0.000e+00 Length:59090 Min. :0.000e+00 Length:59090

1st Qu.:1.829e+04 Class :character 1st Qu.:1.901e+04 Class :character

Median :2.747e+05 Mode :character Median :2.854e+05 Mode :character

Mean :8.739e+07 Mean :8.806e+07

3rd Qu.:3.174e+06 3rd Qu.:3.317e+06

Max. :6.144e+11 Max. :6.144e+11

NA's :715 NA's :705

It displays summary statistics like **“minimum”, “maximum”, “median”, “mean”,** which can be observed for relevant columns in our dataset. Here we would be concentrating more on **Weight/Quantity**.

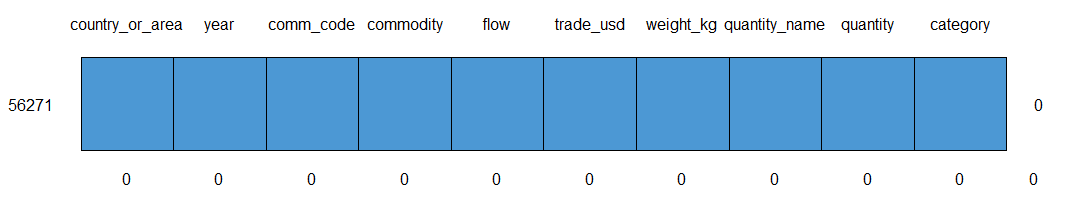
Summary also provides information about the missing values, if any is present. *We see some missing values in our variables Viz., Weight and Quantity which would influence our data to an extent. This is not acceptable as per the data collection that has to be cross verified*

***Step 5:***

*I have cleaned the data set by treating the Null and Zero Values records. After deleting Null and Zero values records remaining observations are good sample to work on.*

GTA\_Data\_NoZ<- GTA\_Data\_NoZ[!is.na(GTA\_Data\_NoZ$weight\_kg) & !is.na(GTA\_Data\_NoZ$quantity),]

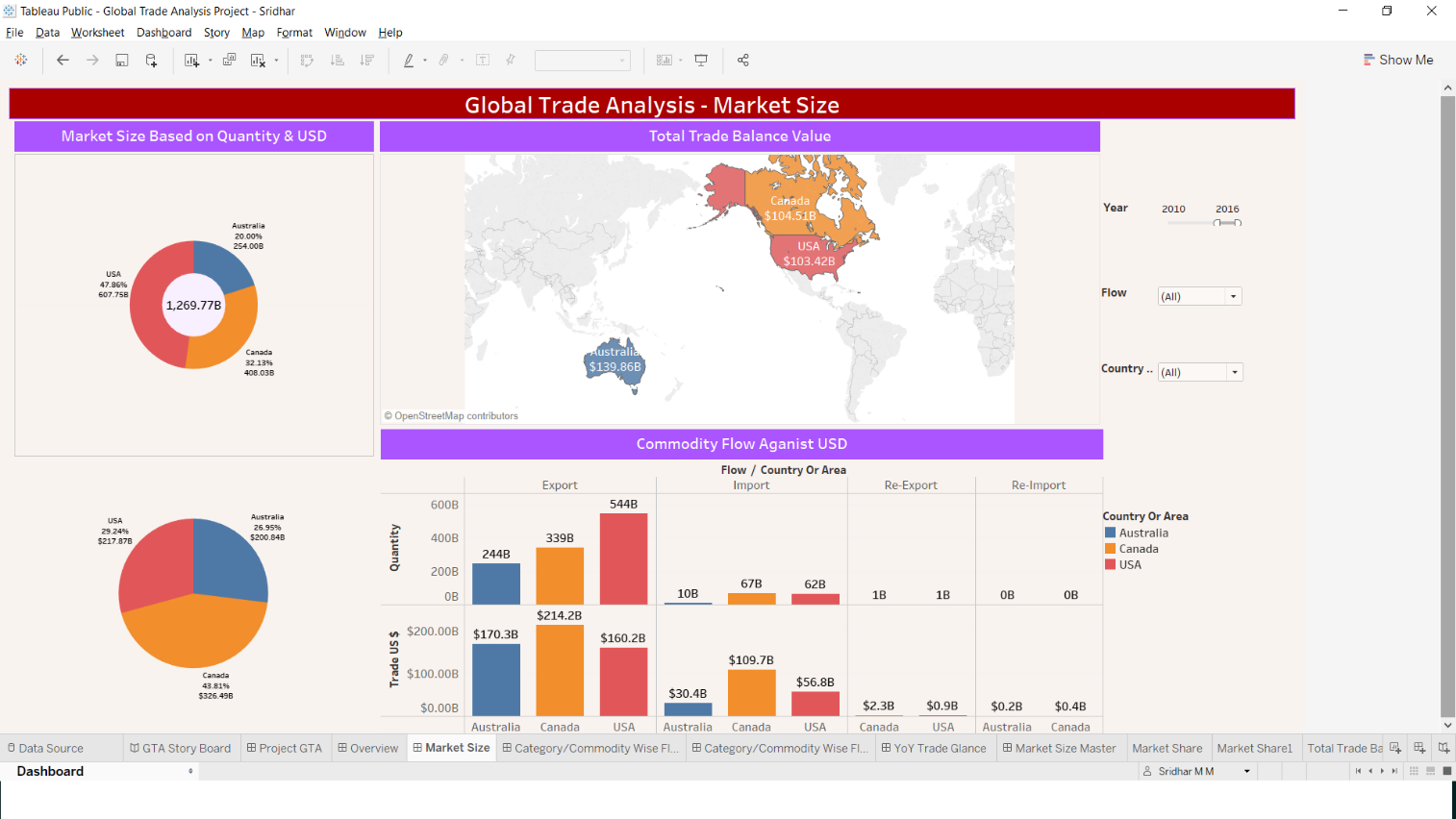
GTA\_Data\_NoZ<-GTA\_Data\_NoZ[-row(GTA\_Data\_NoZ)[GTA\_Data\_NoZ==0],]

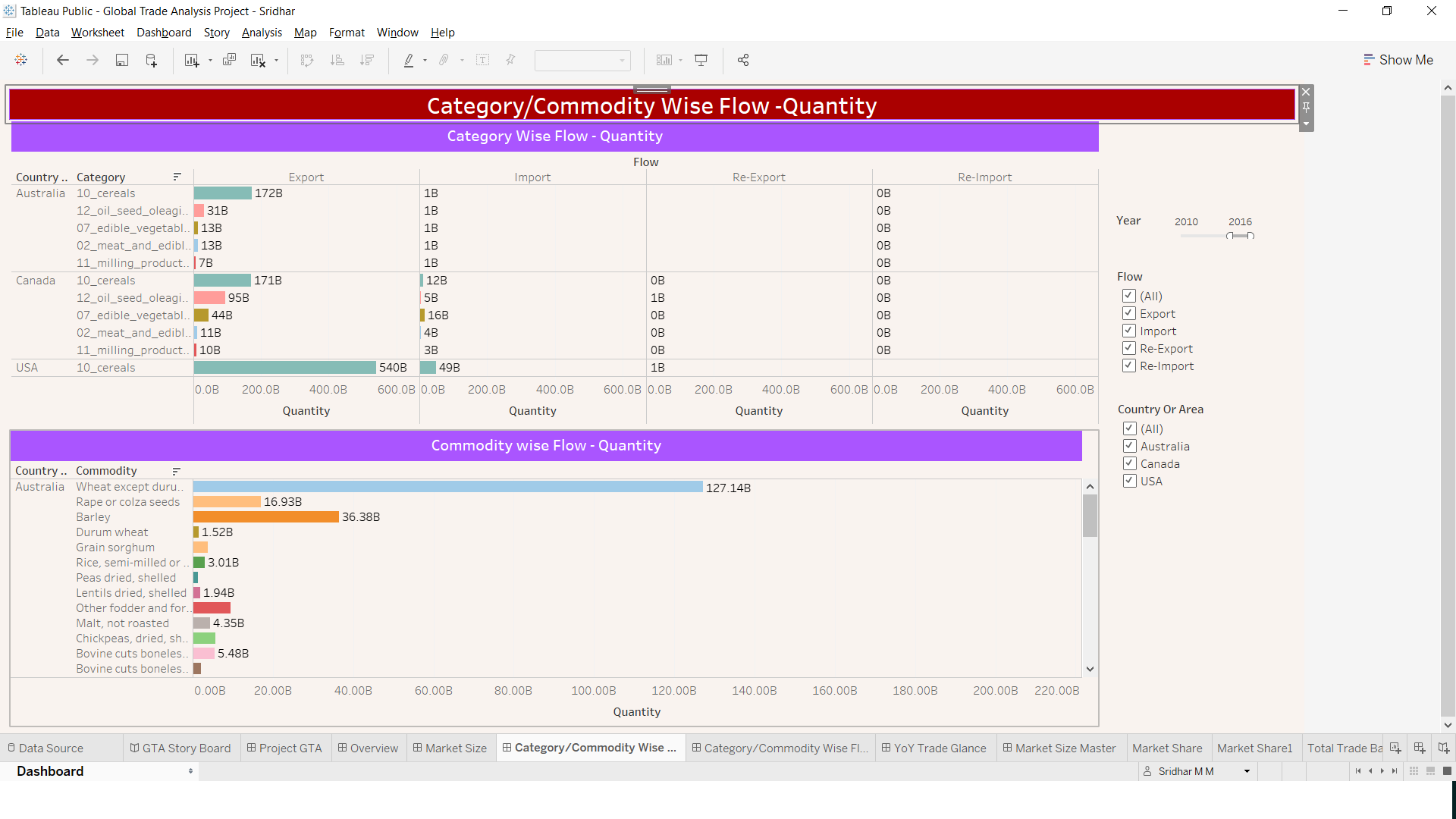


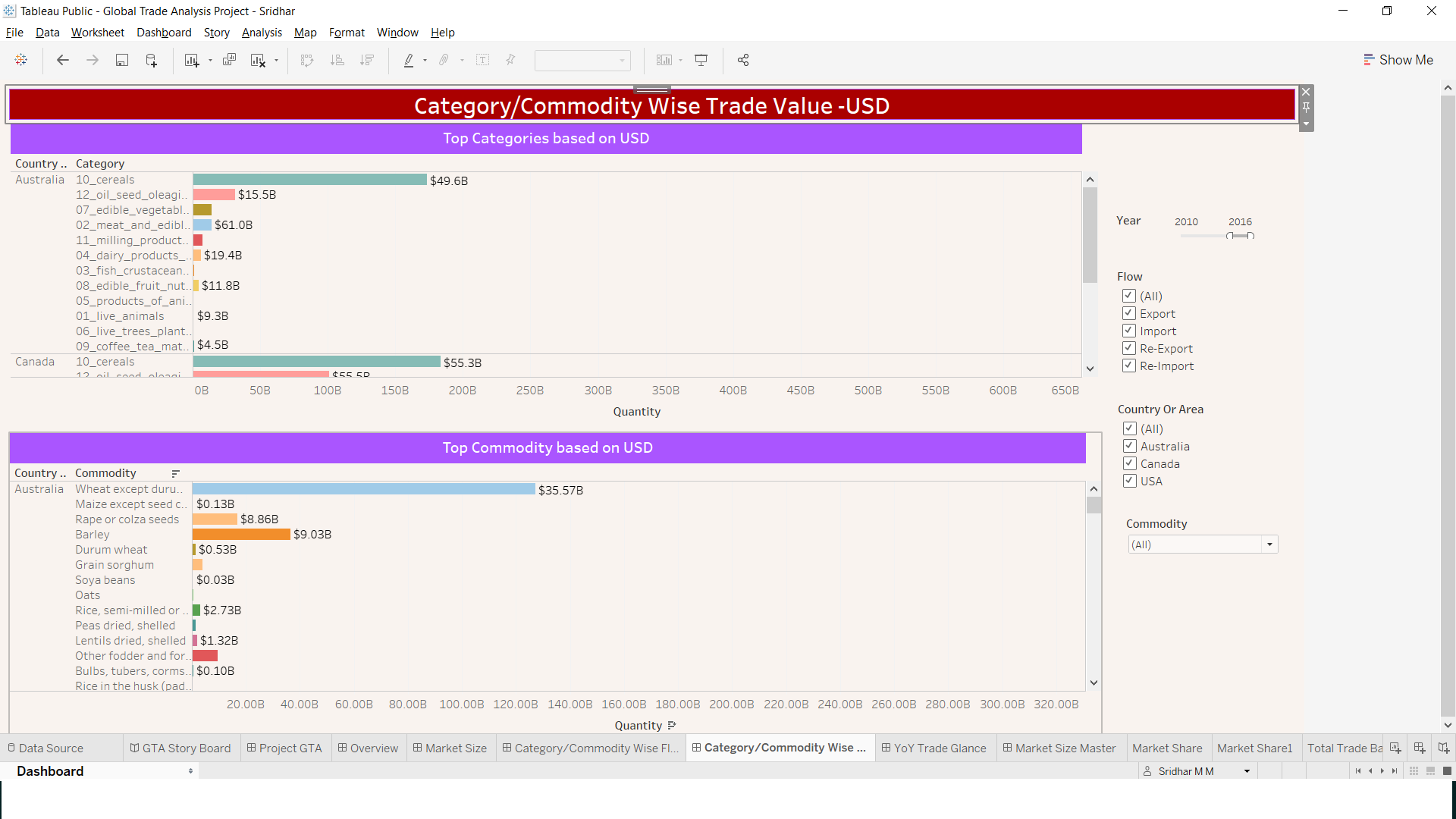
***Step 6:***

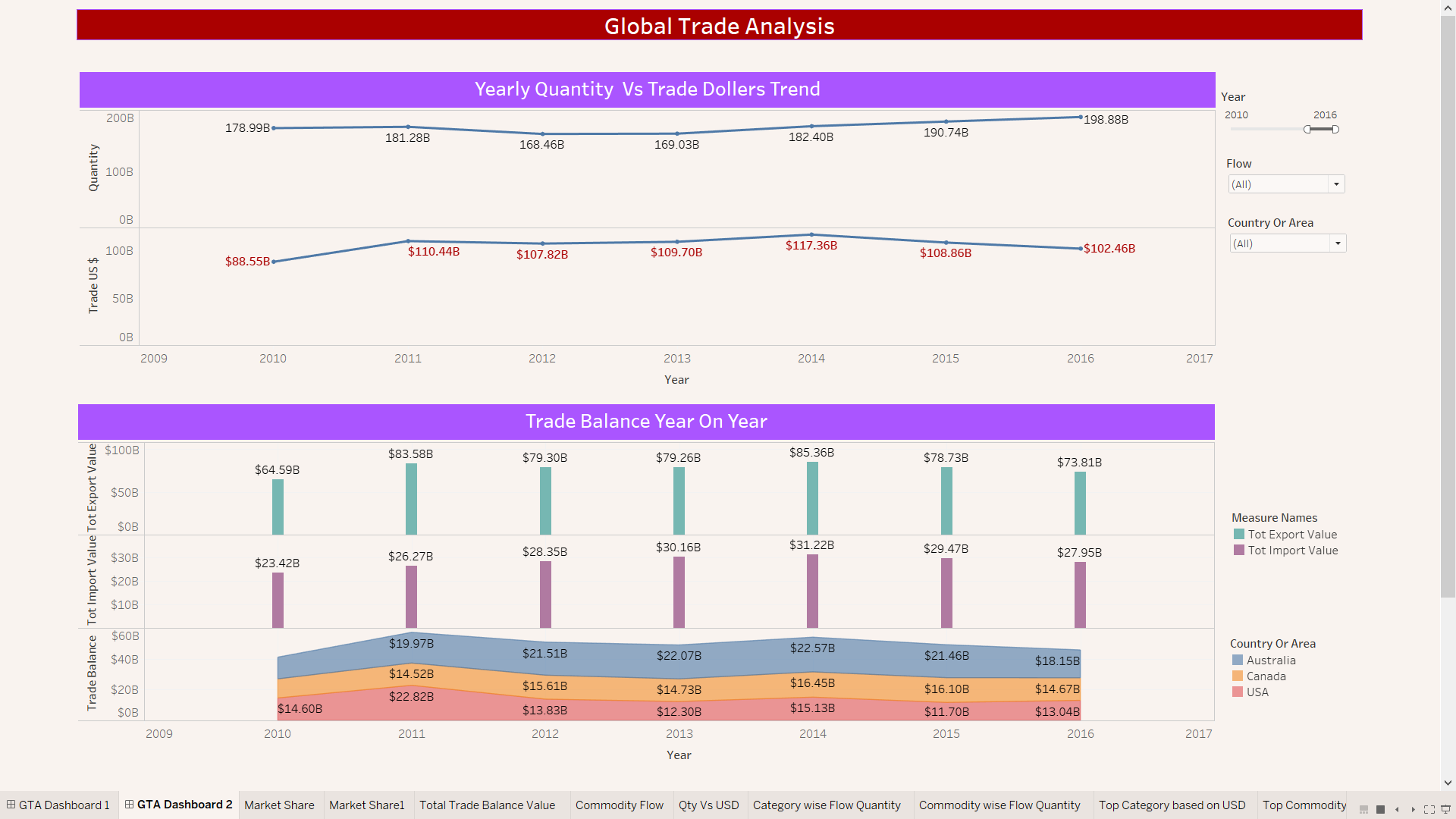
*I have imported the data set to tableau for Visualization as below*

*The Dashboard shows % share held with respect to Quantity & Dollar Value by each country for all the products they have been Exported, Imported, Re-Exported and Re-Imported apart from the details of Trade Balance, Drilldown of Category and Commodity wise trade done for the period 1988 to 2016*









*Observations:*

* *USA has been the major player in the Market for goods trading with 60.54% of share in the overall trading for the sum of all products, followed by Canada at 24.14% and Australia at 15.33%*
* *Further it is observed that Canada happens to be major trader in terms of Dollar value for total sum of products at 42.37%($742.58B) of the total trade followed by US at 32.23%($564.99B) and Australia at 25.04%($445.17B).*
* *When carefully looked for the Trade Balance value it was clear that Canada had More export balance value than the rest of the two Countries.*
* *All 3 countries happened to be major Exporters of Cereals while US did Import the same category most of the time.*
* *Other major export categories are Oil Seeds\_Fruits\_Grains, Edible\_Vegitables and Meat.*
* *The Major Commodities that are exported as below* 
  + - *Wheat except durum wheat, and meslin*
    - *Rape or colza seeds*
    - *Barley*
    - *Durum Wheat*
    - *Soya beans*
    - *Oats*
    - *Other fodder and forage products, roots, etc.*
* *When checked for last 2 years it is observed Wheat, Maize, Rape or colza seeds, Grain sorghum, Durum Wheat, Barley, Peas dried Shelled and Lentils dried Shelled are the products that are in demand.*

#### *Conclusion*

*After cleaning the data in R and visualizing the cleaned data in Tableau for Global trade data set. It was found that there is a lot of demand for*

***Wheat except durum wheat.***

***meslin.***

***Rape or colza seeds.***

***Barley.***

***Durum Wheat.***

***Soya beans and***

***Oats****.*

*I would recommend the Indian Manufacturing company to focus on how they can source the above as they are highly consumed and in demand products across. They need to work and draw export strategies keeping in mind the demand before launching themselves in the market.*

https://public.tableau.com/views/GlobalTradeAnalysisProject-Sridhar/GTAStoryBoard?:embed=y&:display\_count=yes&publish=yes&:origin=viz\_share\_link