Orders Dataset Documentation

* Initially I connected the following dataset to the tableau:
  + South Dataset – from year 2015 to 2018
  + Orders East
  + Orders West
  + Orders Central
  + Orders Returns
* South Dataset:
  + Initially I added the first dataset for year 2015 and selected the option to join more dataset and it tableau automatically selected all south dataset in the folder as all have the similar data structure/format.
  + After this I have changed the datatypes of various fields like date columns as “Date Type” and zipcode of “Pincode Types”.
  + I also removed the files path column added automatically after join as it is not necessary.
  + At last I excluded all the rows which have a null value in any field.
* Central Dataset:
  + I have changed the datatypes of various fields like date columns as “Date Type” and zipcode of “Pincode Types” and also added a new field named “Region” with the value “Central” as the other data previously cleaned where having it.
  + I have also merged the day , month and year of both the orders and ship in there respective field names- “Ship Date” and “Order Date”.
  + Renamed product field to “Product Name”
  + At last I excluded all the rows which have a null value in any field.
* West Dataset:
  + I have changed the datatypes of various fields like date columns as “Date Type” and zipcode of “Pincode Types” and also added a new field named “Region” with the value “Central” as the other data previously cleaned where having it.
  + I have also merged the day , month and year of both the orders and ship in there respective field names- “Ship Date” and “Order Date”.
  + And also merged the column with same name to avoid irrelevant field duplicacy.
  + At last I excluded all the rows which have a null value in any field.
* East Dataset:
  + I have changed the datatypes of various fields like date columns as “Date Type” and zipcode of “Pincode Types” and also added a new field named “Region” with the value “Central” as the other data previously cleaned where having it.
  + At last I excluded all the rows which have a null value in any field.
* Returns Dataset:
  + I have changed the datatypes of various fields like date columns as “Date Type” and zipcode of “Pincode Types” and also added a new field named “Region” with the value “Central” as the other data previously cleaned where having it.
  + And also I trimmed the “Customer Name” and “Reasons” column for the extra white spaces.
  + At last I excluded all the rows which have a null value in any field.
* Union Orders Dataset:
  + I now Union all the orders data of all regions and created a **mega dataset**.
  + Now from this dataset I removed all the rows which have null values in any column and also deleted a field name Table names created after union which is not necessary.
  + Also changed the ”date” field datatype from datetime to date.
* Join the Mega Dataset with the Returns order dataset
  + Here I joined the data using the full outer join and on the field names – “Order ID” and “Product ID”.
  + **Which column(s) are the potential key columns for the above JOIN:**
    - Order Id
    - Product ID
  + **What type of JOIN would ensure that maximum records are extracted for analysis?**
    - I have used **Full Outer Join**  because we want all the data from both the tables.
  + **How would you validate the number of records extracted out of the JOIN?**
    - After the join I will remove the column Like “RowID”, “Order ID”,”Product ID”,”Sub-Category”,”Product Name”, and “Order Date”
    - Changed the date field from datetime to date
    - Removed the rows where the join result null in the field columns.