SETS IN TABLEAU:

Tableau important points:

1. sets can be constant or dynamic , by using marks we create constant sets.
2. For sets to be dynamic select the required dimension on which you want to create a set.eg say order Id.
3. We can create combined dynamic sets , but two sets can only be combined if they are made on the same dimensions.
4. Can be thought of similarly to filter results.
5. Can be created by marks , calculations , combining and can be edited
6. If we want some specific members of the filter results, we can create their individual set and proceed
7. Calculated fields can also be created on different sets.

FILTERS

1. interactive filters can let users interact with them , created by show filter.
2. One way to drag the dimension/measure to the filter shelf

For filtering a dimension:

1. Wildcard option in the edit filter shelf let’s us choose very specifically , like say if we do not want a mail to be addressed to the employes of our company we can simply choose exclude ends with @tableau.com
2. For filtering a measure:
3. level of aggregation can be specified.
4. Range can be specified , lower and upper limit
5. At most and atleast can also be used
6. We can even filter out null values using these quantative filter
7. dates can also be filtered , like relative dates , where we can slect like for the past 2 years or so and we can choose a static date from where we want to start or end.
8. Interactive filters can be used by right clicking on the dimension or measure to be filtered and then using show filter
9. Basic way of filtering involves dragging and dropping the field to be filtered by on the filter shelf

5 filtering can be of many types , aggregate level filtering in which we can filter out the sub-categories which are not profitable , in record level filtering we filter the records of the database which are not profitable , still we have all our subcategories intact instead of the aggregate level filtering.

Data source filtering is used when we only the data based on a particular dimension for eg suppose we only want the data of the European market so we can aplly data source filtering to gather only the data of that market , but we have to be carefull since all the worksheets will be influenced by this as we are filtering the primary data source all together.

Context filters can be used when we want to add more filters to a particular filter where one filter becomes the primary filter and the rest become secondary which are applied on that.

PARAMETERS

1. can be used to add control to our worksheet.
2. Can be used in filters,
3. For eg if we want to find out our top 10 customers by sale then one way is to simply filter them , other way is to create a parameter in this way we can have the total control of the view.

ANNOTATIONS OR TOOLTIPS

They are a very powerful analytical tool can be used to add extra information to our visualization, like we can use annotations to highlight or mark specific areas or points in our visualization, tooltips can give information about the way our visualization is and information related to it.

1. Formatting done in the pane can only be copied BUT the formatting done in the view cannot be copied.

TREND LINES

R square and p value are one way to determine wheather the trend line is accurate or not but , this can be misleading sometimes , the points (error points), distance between the predicted value and the actual they should be normally distributed along the 0 line when plotted against the explanatory variable on the X axis , if this is true then our model is good else NOT. This is called a residual plot.

DASHBOARD AND STORIES

When we create a dashboard , and we do any modifications in the sheets in the dashboard , those changes are reflected back in the original sheets , BUT when a story is created this does not happen and when something is changed it is not reflected back in the sheets , but if a modification is made in the story that persisits whether it is updated in the sheets or not.

If we want to filter a dashboard with a filter which is not available in the original

View then we can simply choose the appropriate sheet and use it to select the filter but this will only filter the view of the sheet from where it is selected.

However if we want the filter to be applied to all the sheets we simply can select the option of apply to all the relevant sheets using the data source , we can also use a particular view as a filter by simply selecting the filter option on the corresponding view and then we can filter the entire dashboard according to that view.

In tableau we have an option of designing a particular dashboard for different devices like a tablet , phone and desktop etc. we have a default view or a master view which will be seen when no particular device is selected , and a custom view , which can be used to modify the default view and can be viewed on a tablet or mobile. We can adjust the size , width and the way it can be viewed by using the options like fit all , fit width etc. This whole exercise can be done for specific tablets.

There are various actions which can be performed in a dashboard like , highlighting , filter and URL. URL can use any field as a dynamic data input , whereas filter and highlighting are source and destination driven. We can use a field in the url for eg we can use Wikipedia to provide information on all the cities in our data source , by simply copying the link and adding the field as city.

There are two types of views in tableau a tile view and a float view , in the float view we can change the dimensions of the legends etc , while in tile view it by default occupies all the space , but in float view we can change it according to us.

Stories can be used to show the audience discoveries made as we are analyzing the data , multiple dashboards can be used in a story , and various story points are created . If any changes are made to the worksheets individually which are there in our story , they will be reflected back in the story as well. Various dashboards and sheets can be adjusted according to the story layout , by using options like fit to story etc. New points can be created and if we want to preserve a point in a particular sheet (in which a filter has been used) we can simply use the update option, we can even save the modification we made to the visualization as a new point , which will by default use the filter which we made on the original visualization.

MAPPING IN TABLEAU

Lattitude indicates how far up or down from the equator , whereas longitude indicates how far east or west from the prime meridian. If we don’t have the latitude or longitude for any place and we only have the name of the city or the province then tableau will automatically generate the latitude and the longitude (generated) for that place. The places on the map can be represented by a simple dot or simply if we hover over the area or by polynomial shapes on the map which shows the area.

Tableau automatically recongnises the geographical data and puts a globe symbol next to it , and in case it does not we can simply use the options to make sure it does and choose the most relevant data type. We can also use filled maps for visualisation in which the sizes of the states are according to the geographical data and we can use measures like profit to be displayed in our map. In tableau we can use a map to search for a particular city or district , maps in tableau are fully user interactive and we have options like zoom , pan and we also have the option to disable them if we want to.We can add layers in maps like coastlines , view highways , streets etc.

When we add city to the latitude and the longitude there can be a lot of ambiguity in our visualisation since there can be a lot of cities with the same name in different states of different countries , to resolve this we can drag state to the view and tableau will then automatically map the cities to their respective countries and states.If tableau does not recongnise a particular state or city in our map then we can use edit locations options to correct that , one error can be that we have spelled the name of the city wrong , hence we can simply correct it , other can be the name of that city might have changed , and one thing can be that the particular city which we have in our data source is not there in the tableau database and we can simply enter the latitude and longitude for that city and that error will be removed.

Tableau interprets spatial information as a new field called geometry and it has globe icon next to it meaning that it can visualized on a map.

Spatial files contain information about the location and latitude etc of the place , and two different spatial files can also be joined with each other.we can expand the mapping capabilities of tableau even further , if we have the coordinates for a particular place then we can run geocoder on it , say we have the address , then we can simply map that data according to our own needs. Importing data directly and using custom geocoding there is a difference that we can use custom polygons , in data in which there already is polygon information it is easier.If we don’t want to use the latitude and longitude everytime then we can create a csv file which contains geocoding data and then import that file to tableau (custom geocoding), it will import all the files and after that we will be able to use this data everytime WITHOUT using latitude or longitude.

A Csv file must have consistent spelling capitalization and column names.

Custom geocoding can be of 3 types one of them being extending an existing role in which new towns are added which tableau does not recongnise,adding new roles creates new levels with in an existing hierarchy, and finally creating a new hierarchy itself , so basically when we create a new hierarchy in tableau itself we need to create different csv files , for each level of the hierarchy and then after each level of the hierarchy is added the previous level is also to be included in the other file we have made.Custom geocoding comes in as very handy when we know we have to do a visualisation in which there will be some very small towns and tableau wont be able to recongnise them , so we can use this in which we obviously also enter latitude and longitude thorough which it recongnises what we need.

The various fields which are required to create a polygon map are-

1. latitude and longitude are the coordinates of the various points to be connected
2. point order tells tableau which points are to be connected and in what order
3. polygon id breaks each sets of points into individual polygons
4. sub-polygon id does the same work
5. finally there has to be a unique identifier for all the polygons which are being made , for eg in case of national parks if we are trying to make a polygon map , then the name for each national park can be the unique identifier.

To add a mapbox to tableau we need to add the url , and integrate it further we need to have the required url or api token to add the required mapbox map , and we have to go to:

1. maps menu
2. then to background maps
3. add map services
4. choose between classic and other option
5. name the map according to us and then continue

background images can be imported in tableau and data can be plotted on it just like a map.

There can be dimensions in our excel file which contains information and then there need to be X and Y coordinates which need to be measures and there has to be a scale assigned to them , in some cases that can simply be the size of the image

1. go to the map menu
2. choose the option of background images
3. add the url of the image by simply using the location of where the image is stored
4. if the image is a map then we can simply use the latitude and the longitude as the X and Y coordinate
5. in our excel file we need to have the x and y coordinates of all the places that we need to mark on our background image , say it is the image of the metro line of hongkong , then we can plot the sum(x) and sum(y) of the points and bring the station to detail and hence we can simply get all the information of all the points that is all the stations in our background image.

Calculations in Tableau

Regular calculations are the queries which tableau ask of the datasource and they are computed by the data source only and the result set is returned to tableau.

Table calculations are performed on top of the returned calculations and they are performed by tableau itself.

Level of detail expressions or LOD expressions allow us to compute the aggregations that are not at the level of detail of the view.eg if we want to find avg profit per country we noticed that multiple orders were placed from a single order id and the profit was for line items , now if we wish to find profit for all the order id’s in our data set , what will we do? Now basically for these sort of problems we use , lod expressions and through LOD expressions we are able to find profit on all the orders of all the order id’s including the line orders and this is made possible by the level of detail expressions and if we want to find the aveage , then we can simply change the aggregation to average.

1. now in lod expressions we can use three keywords { keyword dimension:aggregation}.
2. Fixed , in fixed we only include the dimension which we have specified and we DO NOT include the other dimensions in the view
3. INCLUDE , it includes the dimension which we have specified along with all the other dimensions in the view
4. EXCLUDE , it excludes the dimension specified in the view and only considers the dimensions in the view.

Now in our data set that we had we wanted to find the average of the profit per order id in different countries , now we have to keep one thing in our mind that the order id’s which we have in our data set , they are unique in a particular country but not unique for all countries. This detail makes a big difference as when we use fix LOD expressions if by chance in two countries the order id comes out to be the same then tableau will compute the profit as the sum of both of them , as in fixed LOD expressions it only computes on the basis of order id only , but on the other hand if we use INCLUDE then tableau will also consider the country and now here in this case the profit will be calculated differently for each country. However if the order id’s were unique all together in different countries and in the country itself then both FIXED and INCLUDE would have given the same result.

Now in tableau even the table calculations can be modified , they can either be partitioned i.e table vise or pane vise , or they can be addressing like across or down across then down , so basically when we are modifying the table calculations the dimensions which are checked are the addressing type , i.e they specify the direction in which the table calculation will happen , and if we uncheck them they become partitioning i.e they can be pane vise or table vise , there is a lot of difference when they are partitioned or if they addressing , for eg if subcategory ( phones , printer) of technology category then tableau will tell us the percentage of sub-category sold with reference to the technology category as 100% if both of them are unchecked then tableau will tell us the sale of any individual sub-category with reference to the sale of all the categories and sub-categories as 100%, hence there is a lot of difference between the two.

Attribute or ATTR aggregation checks wheather the function returns only one value or not , if yes then okay , else an asterickt is returned.Make date function in tableau is a very useful function which helps us to combine various fields in the right order to make it a date that tableau recongnises.

If we want the data to be dynamic , i.e if we want the data to be updated as we use our filter , then we can use parameters to do that , they are dynamic data pieces and can be changed like for eg if we want to use different company names at the same time we can do that using parameters.the zn function can help us compare two values in which one of the value contains null values , and hence the zn function converts all the null values to 0 and hence we can compare the two measures or fields we want to compare.

Any aggregate expressions cannot contain attr or table calculations. In a LOD expression the aggregation can be as simple as sum(dimension) or it can be a complex statement , within an aggregation which includes iff and may include 1 or more dimension or measure in the aggregation itself!

LOD expressions which use include and exclude as keywords will always result in a measure whereas the lod expressions which use FIXED can result either in a dimension or a measure.If the aggregate expression yields a dimension then the overall expression will be a dimension , and this will be the same for measure. Since fixed LOD expressions are independent of the other dimensions of the view , the result can be more aggregated or granulated than expected. No aggregation is needed when the level of detail of the expression is courser than of the view as it igets ignored.usually tableau will define the exclude LOD expression in the view as ATTR as there is no aggregation specified in the view.In tableau we can also use nested LOD expressions to solve our problems and the inner expression is usually determined by the outer expression and NOT with respect to the sheet! , it is also called the parent expression. If the inner calculation is more aggregated than the outer calculation then the result is granular , while if the inner calculation is granulated than the outer than the result is aggregated.

When we are using table calculations the fields which we are using should be aggregated.Usually dimensions are discrete and measures are continuous but this is not true always , and also a measure can be converted into a discrete value and then we can see that the sign of the pill changes to blue. Discrete pills are blue and continuous pills are green , dimensions come out in the view themselves whereas measures come out in the view as aggregates.Even the color pane works differently for dimensions and measures , for a measure when it is dropped into the color , a color gradient is created ( light to dark shades) , BUT on the other hand for a dimension a whole new color pallete is created to show the difference.For a continuous date there is a single and progressional trend line , whereas for a dimensional date different categories are created and there is a unique trend line for every category.

While sorting our data in tableau say for eg , we want to sort customer id’s by profit and the customer id’s were to be first determined by the ship mode and we see that the order is incorrect then we can easily infer that , the customer whose id seems to be wrong may have placed multiple orders via multiple ship modes hence obviously the profit for a particular ship mode may be less or more , hence there is this kind of inconsistency.Granularity means we want to increase the no of points in the view or further drill down and this can be done by adding more dimensions in the view and the other thing which tableau does is aggregation . the way tableau calculates or proceeds with a calculation depends on the granularity of the view on which further the aggregation depends.Usually what tableau does while calculating the sum/profit is it individually calculates the the ratio for every entry at the view’s granularity , but this is not we want , we want the ratios of the aggregations that is first sum all the profit and sum all the sales and then takes the ratio that is what we want.when we want to calculate the top 10 products by sales and in a particular region sometimes there will be a condition in which in a particular region we find only 7 products , this is because it may be possible that a top selling product is not available in a particular market , and we need to proceed in a different way , that is this can either be done by using a context filter , or filtering out by the region first.The correct way to do this is by using the rank function and then in the filter column dropping the rank function to select only the countries from 1 to 10 , and in this way we can get all the products in a country which are top selling.

FINDING THE % OF SECOND PURCHASE DATE OF CUSTOMERS WITHIN 6 MONTHS.

So basically here we want to answer the primary question that how many of our customers are willing to buy again from us within 6 months , so basically here what we have to do is find the first order date of the customers , i.e when did they place their first order , then after this we want to find the date of their second purchase , now we have to create a calculated field again using LOD expressions , and then obviously this date has to be minimum as there can be multiple orders so we have to find the date of the order after our first purchase , then the next question is are these orders within 6 months? For this again we have to create a calculated field and then finally we have to tell the % of people who have made their second purchase after 6 months of their first purchase.

A parameter can be very useful if we want give user the power to choose between two fields and we want to do that in a single view itself , first the parameter is to be created and then we have to choose the fields which we want to use in the parameter and then create a calculated field which includes the parameter and use the case function to select the functionality of the field in the parameter and then finally use the show parameter function to access it.

A control chart in tableau is used establish whether the variation in a measure is between acceptable limits or not usually determined by the standard deviation. Window function is used to calculate a particular aggregation in the underlying view and not in the data source itself , any aggregation can be used with this.

Bollinger bands are usually used in tableau for financial analysis to see that a particular stock is oversold or undersold as compared to a moving average. Moving average is calculated over a certain period of time , and we also have to consider the standard deviation during that time period to calculate our moving average. To create a view we need the moving avg and the standard deviation at the same time period as the moving avg , also the upper bound and the lower bound which means moving avg + standard deviation and moving avg- standard deviation.

A bump chart is a line chart which shows in rank over some time , it can convey changes in rank at the start and the end periods and also over time.

Funnel chart are basically used for analysis of prospects for eg in our sales or how candidates move through an interview process.

Pareto charts are used to see what percentage of a dimension is contributing to what percentage of a measure. For eg what percentage of products is contributing to what percentage of sales, it works in conjunction with the 80/20 rule , 80% of sales is coming by 20 % of customers etc. we use the percent of total for the running total of the measure and the count distinct of the measure.

Step or jump lines can be used to show the rank of a book for eg for a particular duration of time for eg by weeks continuously for for 4 months and from this we can see that at which week the rank of our book was 1 , and for how long , we can use jump and step lines to further enhance our view , step lines are obviously like a staircase , whereas jump lines are NOT continuous.

A waterfall chart shows the running total of a measure against a dimension to show the contribution of its members and it shows both positive as well as negative values. It shows how each member of the DIMENSION is contributing to the overall total either positively or negatively. First we select the running total of the measure change the mark type to GANT and then create a field for negative profit and drag it to size and then move it to color.

Fixing incorrect sorts in tableau is an important issue , this can be done by creating a calculated field called “Rank” and then we can simply specify the field on the basis of which we want to sort and then the order in which we want to sort i.e ascending or descending order and the IMPORTANT step here is to convert the measure to DISCRETE.