

Google Data Studio

In **Google Data Studio**, one can connect to data from different tools (sheets, excel, google analytics etc) and then visualise that data into beautiful reports and share it with others.

The flow of data is in the following manner

Dataset → Data connector → Data source → Data report.

Building blocks:

Data Level:

Data source:

It is an instance of a connection when we connect the data from different resources to Data Studio. It's the place where we can configure the fields and options according to our need. It provides a secure way to share our data for the report viewers who don't have access to the particular data. It automatically get updated when we have new data.

It can be used in 2 different ways namely Embedded data source and Reusable data source.

Embedded data source is the one which allows all users to edit the data source and connection. It is the one which is created while editing a report. It is the most commonly used type.

Reusable data source is the one which does not allow the users to edit the data source and connection except the one with edit access.

Data connector:

It is the one which connects Data studio to our data which in result creates the data source. Google itself provide some connectors like Google Sheets, Google Analytics, Google Ads, Google Search Console, YouTube, BigQuery, MySQL, CSV, Google Surveys.

Dimensions :

Dimensions are the ones you want to measure, or that serve as ways to categorise your data.

Metrics:

Metrics measure the things contained in dimensions and provide value for the charts.

Calculated Fields (Data source specific):

Calculated field is used to create new metric or dimension by performing some operations like arithmetic, conditional, geographic on already present fields in the data source.

- Used in charts, controls, and other calculated fields, like you use the normal fields.
- Filter can be used on this like in normal fields.
- This would be available for all report which use that data source.

Limits of data source calculated fields:

- Can't be used with blended data.
- Edit rights to the data source is must to create or edit calculated fields.

Filter:

Creating a filter by adding it to a component, data source is the one in use by the component. The filter created using the filter manager can be used to in any of the data sources currently added to the report. But the selected data source should contain the field used in filtering.

Version control:

Version control saves the previous versions of the report and data source and allows one to view them and if needed one can restore the report or data source to the previous version at any time.

Data access controls:

- Owner's credentials let viewers see the data from our data source without requiring them to have their own access to that data.
- Viewer's credentials require views to have the data source for accessing the report data.

View/ Edit:

- Edit mode allows one to make change, add, or remove data sources.
- View mode allows one to see all the data which are authorised to see.

Data Visualisation level:**Formatting:**

Conditional formatting is used to colour some particular text or background based on the condition we give. We can have 0 or more conditions combined with and/or. In conditional formatting we can give formula to which we need to represent in different colour.

Header formatting

Footer formatting

Body formatting

Colours

Background and Border

Handle missing data

Metric and dimension styling

Calculated Fields (Chart specific):

As like calculated fields in data source, it is used to create metrics by performing some operations based on math, functions or case specific.

- Adding fields is quicker and easier without needing access to the data source.
- Fields can be created on the basis of blended data.
- Data source calculated fields can be used in chart-specific calculated fields.

Limits of chart-specific calculated fields:

- Exist only in the chart in which you create them.
- A field created in the chart does not create it in the data source of the chart.
- Other chart's chart-specific calculated field can't be referenced in a formula of the chart which one use in present.
- Editor can only create a chart-specific calculated fields.
- In the data source, Field Editing in reports must be enabled.

Prediction display:

We can also use google data studio to visualise the predicted ML models which we write in Google colab and take the predictions into a google sheet and deploy it in Google data studio to visualise.

Filter:

Filter is used to refine or reduce the data which is shown according to the specific need. It can be set only by the editors. Meanwhile, reports can have interactive filters in charts. Viewers can filter the chart according to the need if interactive filter option is enabled.

There are 2 types of filters:

- Include filters retrieve only the records that match the conditions.
- Exclude filters retrieve only the records that DON'T match the conditions.

E.g., Suppose you want to limit a chart's information to just data coming from the United States. Select the chart and add filters

Components:

- The report itself
- Pages
- Charts, such as bar, line, pie, tables, etc.
- Controls, such as the date range control, filter control, and data control
- Text areas
- Shapes
- Images

Rows per page:

Rows per page is used to control the number of table rows to be displayed per table page. **Show pagination** must be enabled before.

Show summary row:

Show summary row is used to create an extra row at the end the table to show the summary of themetric column.

Sort:

Sort enables viewers to sort the data either in ascending or descending order based on the field selected. Secondary sort allows to sort a table based on the combination of field and dimension.

Collaboration:

Generate a link for the report and then share the report(link) through gmail to the people with whom one need to work together. One also share publicly so that anyone with the link can collaborate. Owner of the report can modify the view and edit permission to the users at any time.

View/ Edit:

- Edit mode allows one to edit the report.
- View mode allows one to see the report and interact with the view controls.

Calculated field allows one to create a new field using different types of functions like aggregation, arithmetic, Conditional, Date, Geo, Text which is common for both data level and data visualisation level.

- Some of the **aggregation functions** are AVG, COUNT, MAX, MEDIAN, STDDEV.
- Some of the **arithmetic functions** are COS, CEIL, LOG10, POWER.
Some of the **conditional functions** are CASE, IF, IFNULL.
Some of the **date functions** are CURRENT_DATE, DATE_DIFF, DATETIME.
- Some of the **geo functions** are TOCOUNTRY, TOCONTINENT, TOCITY.
- Some of the **text functions** are CONCAT, LENGTH, LOWER.
- Some other functions are IMAGE, CAST, HYPERLINK.

Demonstration of a simple conditional metric

Formula ?

```
1 CASE
2 WHEN ( Years experience >= 0 AND Years experience < 4) THEN "JUNIOR"
3 WHEN ( Years experience >=4 AND Years experience < 8) THEN "SENIOR"
4 WHEN ( Years experience >=8) THEN "SUPER SENIOR"
5 END
```

	Name	Years experience	Experience	Position	Salary
1.	Hank	15	SUPER SENIOR	Manager	1,500,000
2.	Julian	9	SUPER SENIOR	Manager	1,250,000
3.	Macy	12	SUPER SENIOR	Data scientist	1,240,000
4.	Cadence	10	SUPER SENIOR	Data scientist	1,100,000
5.	Isla	9	SUPER SENIOR	Data scientist	1,020,000
6.	John	6	SENIOR	Developer	1,011,000
7.	Henry	6	SENIOR	Manager	1,000,000
8.	Adam	12	SUPER SENIOR	Designer	950,000
9.	William	6	SENIOR	Data scientist	900,000
10.	Josephine	8	SUPER SENIOR	Designer	900,000
11.	Doug	7	SENIOR	Data scientist	870,000
12.	Rick	4	SENIOR	Designer	800,000

Demonstration of a simple aggregation function

Field Name

Average Salary

Field ID

calc_gyp8tlwzkc

Formula ?

1 AVG(Salary)

CLOSE

UPDATE

Features:

Data studio supports data integration from 200+ connectors.

Drilling down helps the user to get more details in a chart. E.g., Suppose we want a chart with COVID state in all states and districts, there is no need for creating two charts. With drill down option enabled we can drill down to districts from state and drill up from districts to state.

The **Breakdown Dimension** displays the numerical data(metrics) categorised according to the selected dimension. For example, a chart showing annual income data could be categorised by a designation dimension to show salary by designation, or by a qualification dimension to show income by different level of education.

Blend data allows to create charts with multiple data sets. As long as there is a Join key, you can blend two data sources and get a more complete picture.

When data is taken from two different data sources, related data have to be considered separately this is known as **Data mapping**.

Data conversion is changing one type of data to another type for better interpretation, this can be done by 'type' option in the edit page of your google data studio.

'Data Control' option in Google Data Studio lets you give your viewers the ability to select which data they want to see in the report.

In google data studio after collecting data from the data sources if we make changes to the data, it gets reflected automatically after refreshing. This is known as Real time data collection.

In google data studio u can add multiple pages to a single report.

Enterprise reporting includes reports on key performance indicators.

Sharing and distribution of reports can be done by **sharing** with email addresses, **Google** Groups, or public links.

We can add Google Docs, Images, YouTube Videos, Google Sheets, and web pages into our report via the **URL embed feature**.

Reference lines can be added to charts, which can be either horizontal or vertical lines. These will help us in understanding how the data in the chart compares to a reference value.

Metric sliders allow you to adjust the values like a price range when you're shopping for products

We can create our own custom visualizations using community visualizations.

Themes are a set of predefined color and style combinations you can apply to your reports. Applying a theme updates the style settings of your charts, tables, background color, and text, helping to ensure your reports are visually attractive and stylistically coherent.

CHART TYPES:

AREA CHART

An area chart contains shaded areas under the plotted lines on the graph. It also displays the volume of data represented by the lines.

There are 2 other area charts named stacked area chart and 100% stacked area chart. Stacked area chart is used to represent the evaluation of data of different group in same graph where as 100% stacked area chart represents the percentage of data.

WHEN TO USE

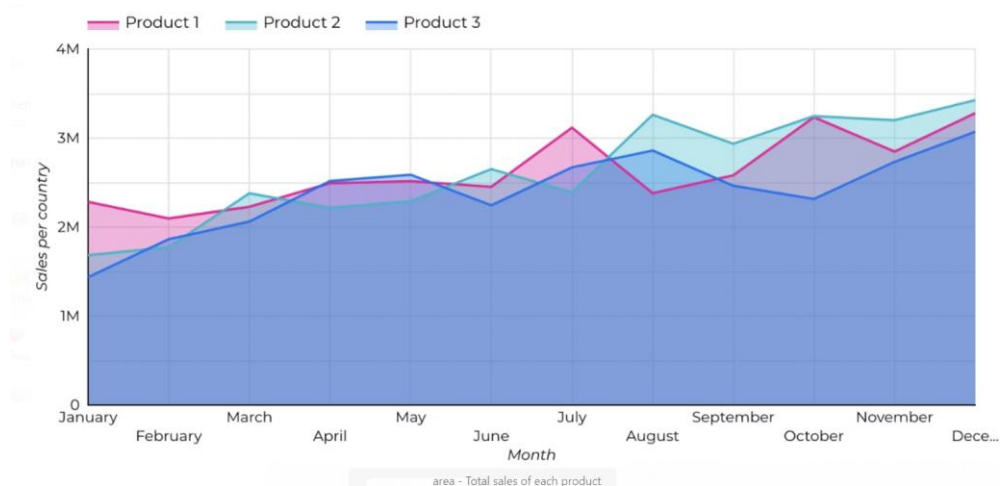
An area chart is used to represent trends over time. It is perfect to use when we communicate the overall trend, with respect to the individual values. Stacked area chart can be used for multiple data series with part-to-whole relationships.

Area chart can visualise 1 dimension which can be breakdown dimension and 1 metric

E.g., Showing increase in sales over the course of a year.

Comparing the contribution of different teams made to overall sales. - Stacked area chart.

- The below area chart shows the sales for 3 products across the 12 months.



BAR CHART

A bar chart uses rectangular bars to compare different categories of data. The longer the bar, the greater the value it represents.

WHEN TO USE

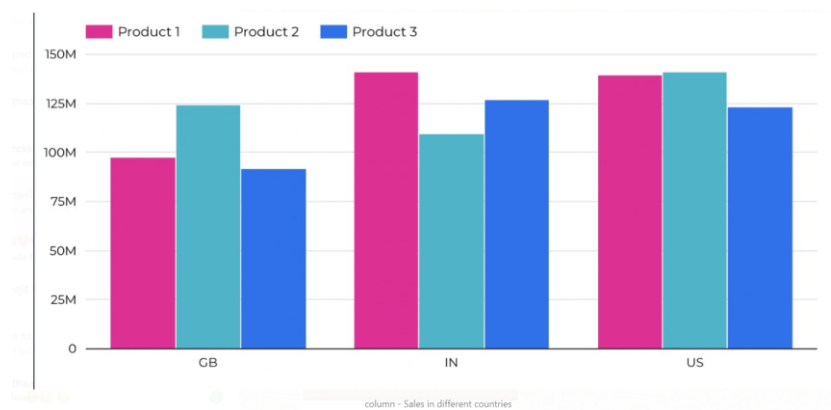
Bar charts are used to show discrete data (that's data that is based on counts and can only be certain values). They are most effectively to show change over time (e.g., the net monthly earnings of Tesla Motors in a year), Compare values of different categories (e.g., a year's fishing yields for different species of fish), Compare parts of a whole (e.g., the percent distribution of Netflix rentals across genres)

Bar chart should have at least 1 metric. It can visualise 20 metrics when there is a single dimension and 1 metric when there are 2 dimensions.

E.g., A company showing the sales activity of each month for a year to its stakeholders.

There are different bar charts like column chart, stacked column chart, 100% stacked column chart, bar chart, stacked bar chart, 100% stacked bar chart.

- The below column chart shows the 3 product's sales across different countries.

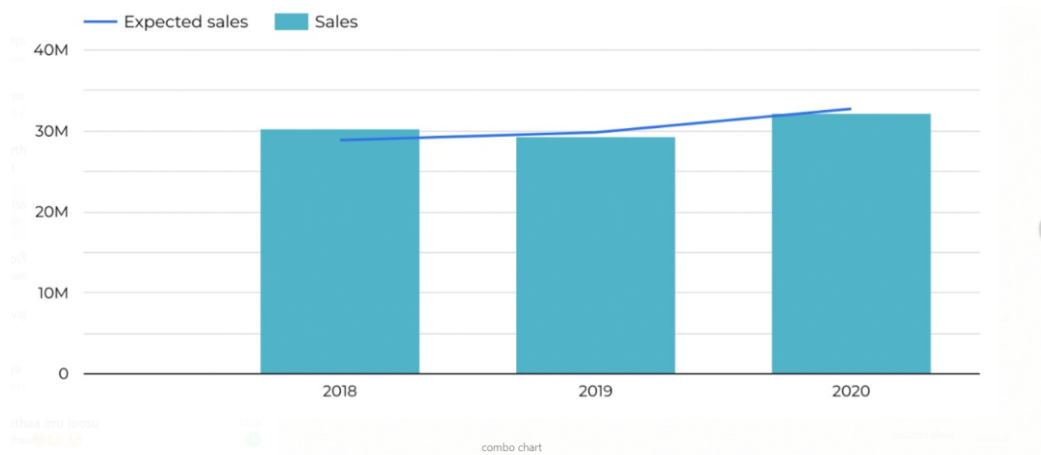


A grouped bar graph is a used to show information on sub-groups of the main categories.

A **stacked bar** chart also shows sub-groups, but the sub-groups are stacked on the same bar.

The stacked bar chart is used for **grouping or categorizing the parts of a whole**.

- The below combo chart shows the sales in different years comparing it with the expected sales.



BULLET CHART

Bullet charts give you a way to quickly see how well is the performance against target performances. It has a maximum of 3 ranges. Bullet charts having 3 components:

- A centre bar showing the actual value of the metric you are graphing
- A vertical line showing a target value
- Coloured bands that represent threshold ranges, such as poor, average, and good

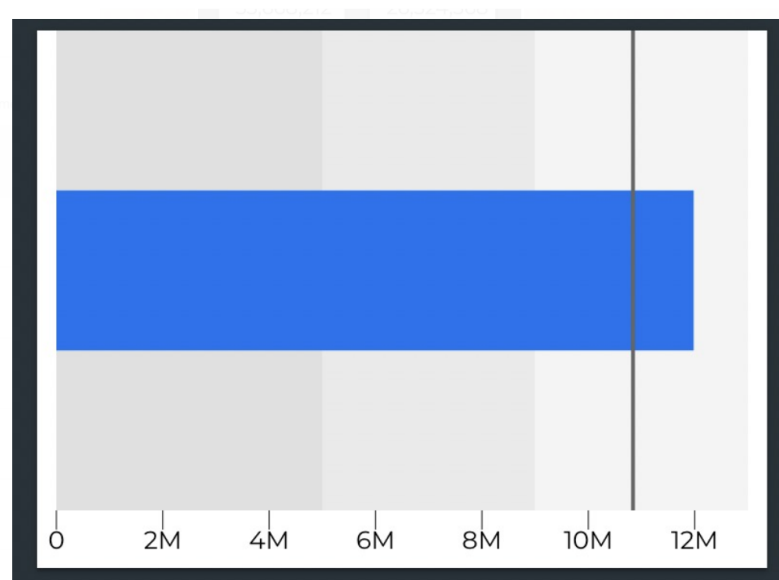
WHEN TO USE

Bullet charts are often used to monitor KPI (key performance indicators). A bullet graph is useful for comparing the performance to what was predicted or expected. From bullet graphs we can also understand poor, average and good performances with the presence of different shades in the graph.

Bullet chart can visualise only 1 metric.

E.g., Expenses, revenue (actual and predicted values).

- The chart below shows total mobile subscriptions in India from 2004 to 2017 and the target value which is the 3/4th of the total population.



Product 1's sales in 2020 compared with expected sales - bullet

GEO CHART

A Geo chart provides an easy way to visualise how a measurement varies across a geographic area. E.g., the chart below shows the population in each country.

WHEN TO USE

Use a geo chart to visualise the variation of data across a country, continent, or region.

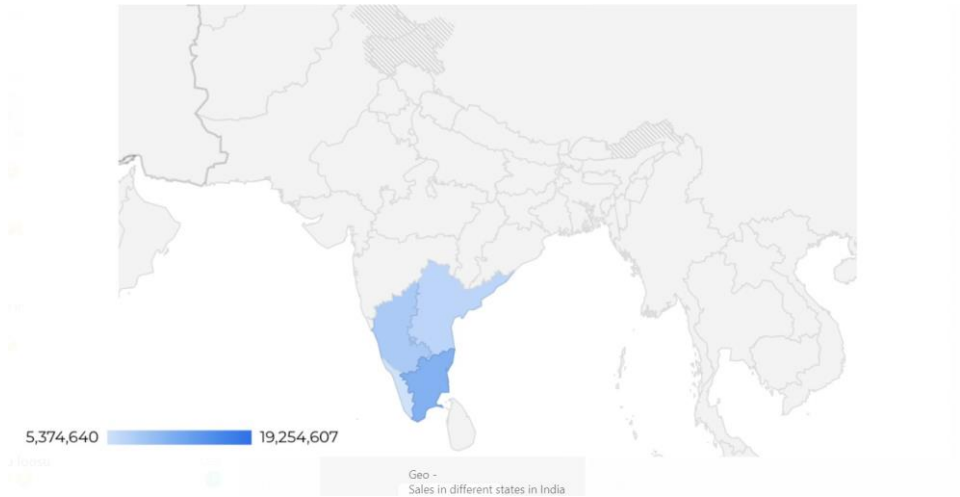
Any single available metric is required.

Drill down options for a geo chart				
World	Continent	Subcontinent	Country	Region
Continent Continent Code Sub Continent Sub Continent Code Country Country Code Latitude Longitude City City Code	Sub Continent Sub Continent Code Country Country Code Latitude Longitude City City Code	Country Country Code Latitude Longitude City City Code	City City Code Metro Metro Code Region Region Code Latitude Longitude	City Metro Code (US only)

E.g., a map that shows the birth/death rate across countries in Asia

If you'd like to create a geo chart that displays sales data across U.S. states. Your preexisting data would likely include values corresponding to state names (for example, California) or state codes (for example, CA)

- The below chart represents the sales for different states in India.



LINE CHART

Line charts are useful when you want to see how your data changes with respect to another parameter. E.g., the number of mobile subscriptions with respect to years.

WHEN TO USE

Line chart is used to represent a continuous data. Used to identify a trend in the data. Easy way to show time series relationships and analyse increase and decrease. Good for tracking multiple data sets on the same chart to see any correlation in trends.

Line charts can plot a single dimension with up to 5 metrics, or 2 dimensions with a single metric.

There are different line charts like combo chart, stacked combo chart, lone chart, smoothed line chart.

E.g. A line graph can be used in comparing the amount of money held by each branch office with a separate line for each office.

A finance department may plot the change in the amount of cash the company has had over time.

To see the no of users a particular app over a time period.

- Showing the change in average salary with increase in the years of experience.



PIE CHART

Pie charts display your data in sections (slices of the pie). The sizes of the slices are proportional to the value of the metric you are plotting. E.g., percentage of total population of different countries.

WHEN TO USE

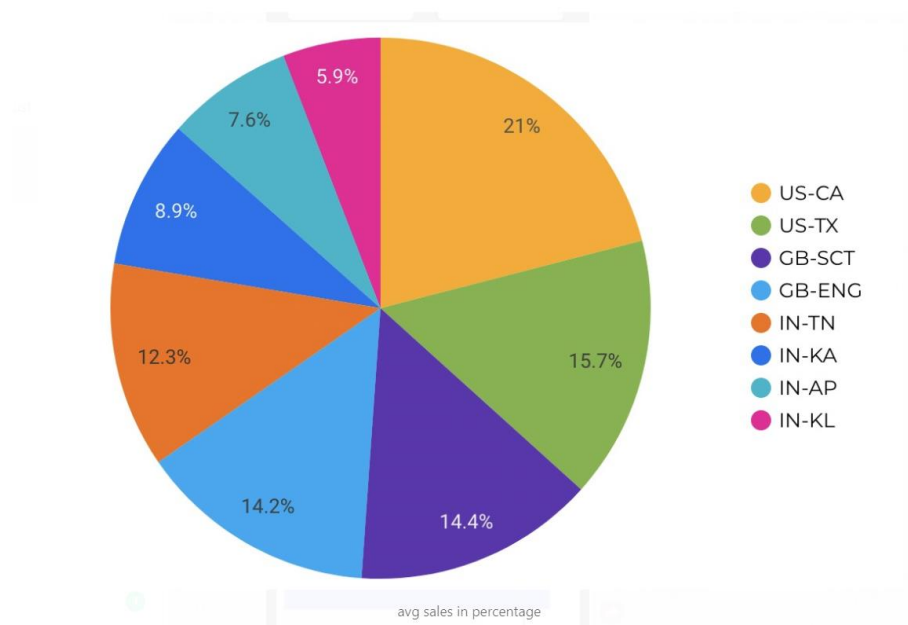
A pie chart is used to visualise the how much is the contribution of a category as a part of whole from which the maximum and minimum contributors can be inferred easily. Used to express a whole data divided into portions than the specific numbers. Mostly used for visualising smaller data sets.

Pie chart can visualise the relative values for a single dimension/metric combination. Dimension is used to determine the number and names of the slices. Metric is used to determine the proportion of each slice.

It can support a maximum of 10 slices. Pie chart can have 1 dimension and 1 metric.

E.g, Comparing the proportion of turnover, profit and exposure of a company.

- The below pie chart shows the average sales across different states in 3 countries in percentage.



SCORECARD

Scorecards display a summary of a **single metric**. Scorecards are commonly used to visualize **key performance indicators** (variable that measures the performance of your business or area of activity).

- For example, a scorecard can summaries total sales, sales across different countries and average sales



TIME SERIES

A time series chart shows how your data changes over a period of time. For example, you can display the increase in population with respect to years.

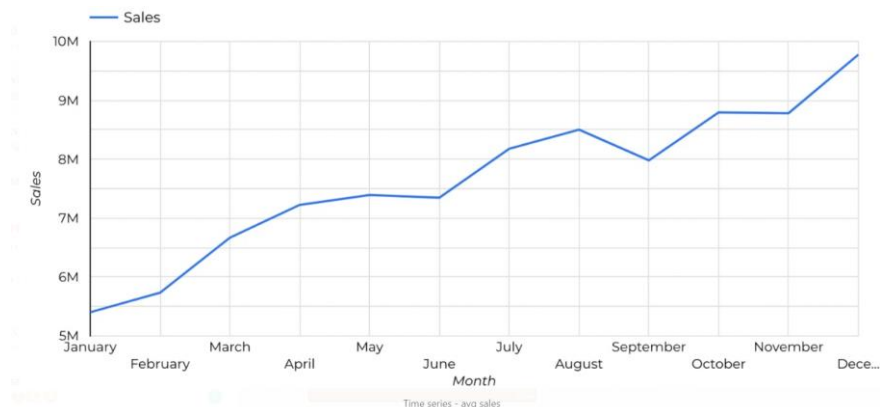
WHEN TO USE

Time series graphs is used to visualise trends in counts or numerical values varying over time. As data and time are continuous, we can plot the points and connect them to get a continuous line. The X-axis of time series chart always represents data or time.

Time series chart can visualise upto 5 metrics in same chart.

E.g., An agricultural analyst is studying trend in growth of crops to determine if initiatives worked effective. The analyst uses a time series graph and compares the growth of crops over time.

- The below chart shows the average sales across all the months in a year.



TREEMAP

A tree map shows your data organized into dimension hierarchies. For example, you can use a tree map to show the average annual sales of each item in a product category > subcategory > product hierarchy. The tree map below displays the number of items sold by an online pet store.

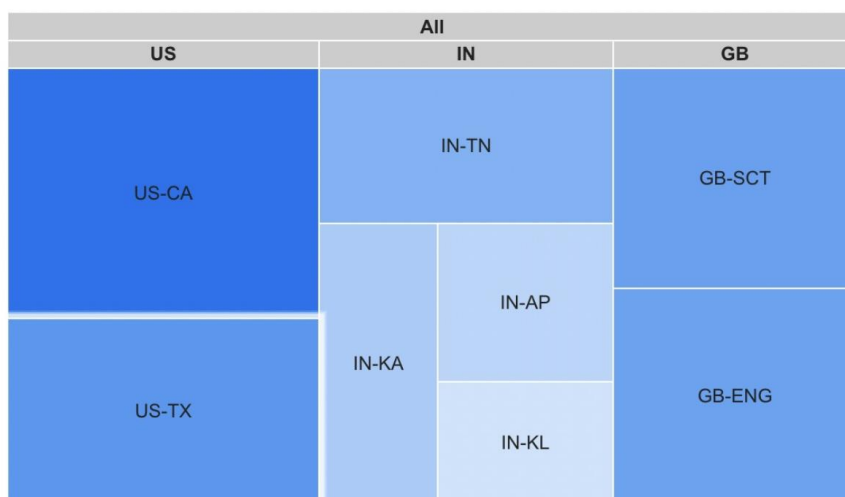
WHEN TO USE

Tree-maps are used to display data that is grouped and nested in structure. It is very helpful while representing data which is a part of a sub category of a topic.

Tree-maps can display from 5 to 5000 rows of data.

Tree maps are often used for sales data, as they capture relative sizes of data categories, allowing for quick perception of the items that are large contributors to each category.

- The below chart represents the amount of sales in each country and states



Sales in each country and its states

SCATTER CHART

Scatter charts are used to look for relationships between variables. These charts show your data as points or circles on a graph using X (left to right) and Y (top to bottom) axes. Scatter charts can include a trendline that shows how the variables in the chart are related.

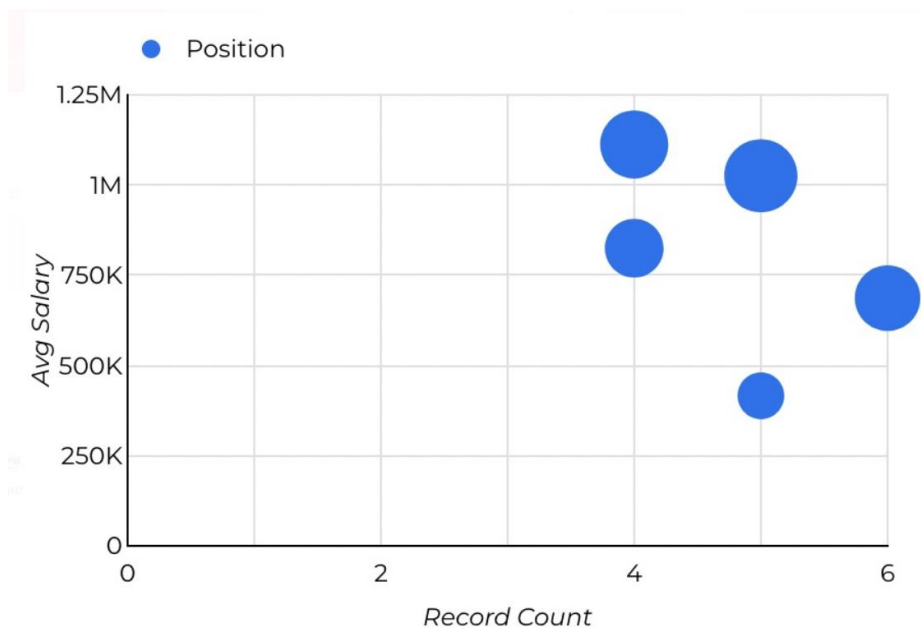
WHEN TO USE

Scatter chart can be used when you have paired numerical data Or your dependent variable may have multiple values for each value of your independent variable or while trying to determine whether the two variables are related.

Scatter charts in Data Studio let you see the relationship between 2 metrics for up to 3 dimensions.

E.g. Marks of different students in a class.

- No of employees in a specific position and their avg salary



PIVOT TABLE

Pivot tables in Data Studio take the rows in a standard table and pivot them so they become columns. This lets you group and summarize the data in ways a standard table can't provide. Pivot tables let you narrow down a large data set or analyze relationships between data points.

WHEN TO USE

A PivotTable is an interactive way to quickly summarize large amounts of data. You can use a PivotTable to analyze numerical data in detail, and answer unanticipated questions about your data. A PivotTable is especially designed for: Querying large amounts of data in many user-friendly ways.

Pivot tables reorganize your dimensions and metrics to help you quickly summarize your data and see relationships that might otherwise be hard to spot.

E.g. Summary of salary for various employees.

- Sales in different countries for different products across 3 years

		Country / Sales per country		
<input type="checkbox"/> Product	Year	US	IN	GB
Product 1	2020	4,029,000	5,402,902	2,547,730
	2019	3,610,000	5,100,500	1,927,334
	2018	3,982,733	1,256,844	3,648,782
	Total	11,621,733	11,760,246	8,123,846
Product 2	2020	2,691,020	3,028,950	4,602,000
	2019	4,063,000	2,102,400	3,509,600
	2018	5,012,321	4,012,342	2,435,096
	Total	11,766,341	9,143,692	10,546,696
Grand total		33,668,212	31,792,547	26,324,568

The dataset used and the charts link in google data studio are attached below.

Salary dataset:

<https://datastudio.google.com/u/0/reporting/7c686ee1-6b98-4c83-8028-03d006e6c241/page/da9SC>

Sales dataset:

https://docs.google.com/spreadsheets/d/1xwFxBnbjBFaBvsjCmFauZdTPttKS0hXWw0RaZv5A-_M/edit#gid=684740176

Charts:

<https://datastudio.google.com/u/0/reporting/7c686ee1-6b98-4c83-8028-03d006e6c241/page/da9SC>