UNIT III Basics of Data Visualization

Introduction to data visualization, challenges of data visualization, Definition of Dashboard, Their type, Evolution of dashboard, dashboard design and principles, display media for dashboard. Types of Data visualization: Basic charts scatter plots, Histogram, advanced visualization Techniques like streamline and statistical measures, Plots, Graphs, Networks, Hierarchies, Reports. Data Science with MS-Excel, Data Science with Google Data Studio.

INTRODUCTION TO DATA VISUALIZATION

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions

Our eyes are drawn to colors and patterns. We can quickly identify red from blue, square from circle. Our culture is visual, including everything from art and advertisements to TV and movies. Data visualization is another form of visual art that grabs our interest and keeps our eyes on the message. When we see a chart, we quickly see trends and outliers. If we can see something, we internalize it quickly. It's storytelling with a purpose. If you've ever stared at a massive spreadsheet of data and couldn't see a trend, you know how much more effective a visualization can be

As the "age of Big Data" , visualization is an increasingly key tool to make sense of the trillions of rows of data generated every day. Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information

Benefits of data visualization:

- Easily, graspable information – Data is increasing day-by-day, and it is not wise for anyone to scram through such quantity of data to understand it. Data visualization comes handy then.
- Establish relationships – Charts and graphs do not only show the data but also established co-relations between different data types and information.
- Share – Data visualization is also easy to share with others.
- Interactive visualization – today, when technological inventions are making waves in every market segment, regardless of big or small, you could also leverage interactive visualization to dig deeper and segment the different portions of charts and graphs to obtain a more detailed analysis of the information being presented.
- Intuitive, personalized, updatable – Data visualization is interactive.

Example of data visualization:

- Government Budget – Government budgets are always tough to understand as they number and more numbers.
- World population –World population along with their density is represented visually
- Profit and loss – Business companies often resort to pie charts or bar graphs showing their annual profit or loss margin.

● Films and dialogues – Out of many characters in the film who will have how many dialogues? Data visualization is the answer here.

CHALLENGES OF DATA VISUALIZATION

Visualization of big data with diversity and heterogeneity (structured, semi-structured, and unstructured) is a big problem.
Speed is the desired factor for big data analysis.
Designing a new visualization tool with efficient indexing is not easy in big data.
Cloud computing and advanced graphical user interface can be merged with the big data for the better management of big data scalability
There are also following problems for big data visualization
• Visual *noise:* Most of the objects in the dataset are too relative to each other. Users cannot divide them as separate objects on the screen.
• Information *loss:* Reduction of visible data sets can be used, but leads to information loss.
• Large *image perception:* Data visualization methods are not only limited by aspect ratio and resolution of device, but also by physical perception limits.
• High *rate of image change: Users* observe data and cannot react to the number of data changes or its intensity on display.
• High *performance requirements:* It can be hardly noticed in static visualization because of lower visualization speed requirements--high performance requirements.
Visualizing every data point can lead to over-plotting and may overwhelm users' perceptual and cognitive capacities

1. Data Inputs and Algorithms Are Prone to Human Error
2. Reliance dependance on Visualization Is unavoidable

DEFINITION OF DASHBOARD, THEIR TYPE

A dashboard is an information management tool used to track KPIs (Key Performance Indicators), metrics (Method of Measuring something), and key data points that are relevant to your business, department, or a specific process.
Dashboards aggregate and visualize data from multiple sources, such as databases, locally hosted files, and web services.
Dashboards allow you to monitor your business performance by displaying historical trends, actionable data, and real-time information.
Dashboards actually take their name from automobile dashboards and they are used in much the same way. For the sake of an analogy, let's look at a car. There may be hundreds of processes that impact the performance of your vehicle if you look under the hood. Your car's dashboard summarizes this using visualizations so you can focus on what matters most: safely driving your vehicle.
For businesses, there are hundreds of processes that impact your performance if you look 'under the hood', so to speak. And with a wealth of data made available these days, managing and extracting value from it can be difficult. Simplifying data analysis and distribution through tools like dashboards is a way to help businesses *rev(increase running speed) their engines* and make smarter, better, faster data-driven decisions.
And a well designed dashboard levels up your approach to information management.

Everyone in the business, regardless of role, has questions about your company performance, whether it be campaign performance, new wins, or churn (Churn is a measurement of the percentage of accounts that cancel or choose not to renew their subscriptions) rate. Dashboards bring everyone (and your metrics) together in one place to answer these questions.

A dashboard is a type of graphical user interface which often provides at-a-glance views of key performance indicators (KPIs) relevant to a particular objective or business process. In other usage, "dashboard" is another name for "progress report" or "report" and considered a form of data visualization.

The “dashboard” is often accessible by a web browser and is usually linked to regularly updating data sources.

Well known dashboards include Google Analytics dashboards, used on 55% of all websites,[1] which show activity on a website; such as visits, entry pages, bounce rate and traffic sources.

The COVID-19 pandemic of 2020 brought other dashboards to the fore, with the Johns Hopkins coronavirus tracker[2] and the UK government coronavirus tracker[3] being good examples.

The term dashboard originates from the automobile dashboard where drivers monitor the major functions at a glance via the instrument panel.

Benefits

Digital dashboards allow managers to monitor the contribution of the various departments in their organization.

Benefits of using digital dashboards include:

- Visual presentation of performance measures
- Ability to identify and correct negative trends
- Measure efficiencies/inefficiencies
- Ability to generate detailed reports showing new trends
- Ability to make more informed decisions based on collected business intelligence
- Align strategies and organizational goals
- Saves time compared to running multiple reports
- Gain total visibility of all systems instantly
- Quick identification of data outliers and correlations
- Consolidated reporting into one location
- Available on mobile devices to quickly access metrics

What's the difference between a dashboard and a report?

Let’s look at similarities first. Dashboards and reports both:

- Provide historical data
- Bring multiple metrics together

Where the differences lie is:

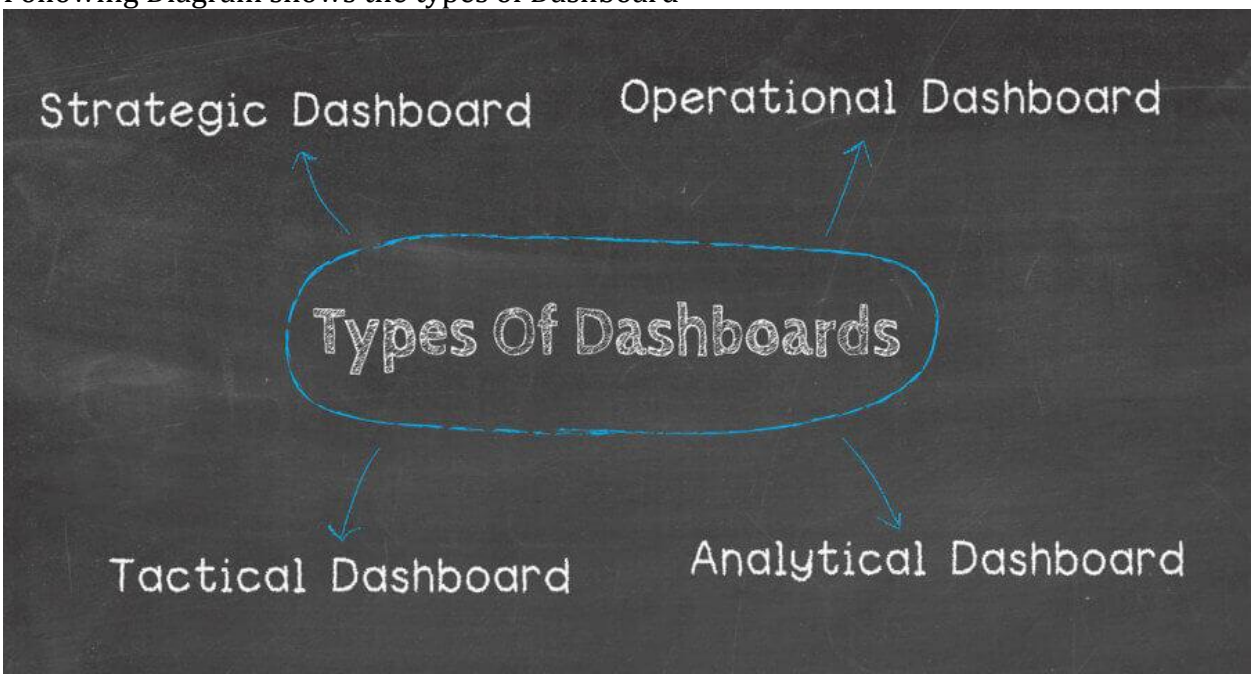
- Reports are static, dashboards are interactive with live, dynamically updated data
- Reports share information on known areas of interest or goals, Dashboard monitor known areas of interest or goals

Reason for using a business dashboard

When used effectively, dashboards can significantly impact business performance. Here are three common use cases for business dashboards.
1. Get more value out of your data
What does this mean exactly? Data is one of the most valuable assets owned by your business.
2. Consolidate(combine number of thing) and automate multiple data points
3. Align teams and departments

DASHBOARD TYPES

There is another important factor to a dashboard's success, It is as simple as choosing the right type of dashboard.
As mentioned, the purpose of a dashboard is to drive action. In this data-driven world, many dashboard types are changing the way a successful business intelligence strategy is conducted.
Following Diagram shows the types of Dashboard



. There are 4 general subtypes of dashboards:

1. Strategic - focused on long-term strategies and high-level metrics
2. Operational - shows shorter time frames and operational processes.
   Operational dashboards look at current performance related to your KPIs. They help organizations understand, in real-time, if their performance is on target. They are often used across various levels of an organization.
3. Analytical - contains vast amounts of data created by analysts.
   Analytical dashboards help organizations establish targets based on insights into historical data. They are often complex: utilizing complex models and what-if statements. Analytical dashboard ownership usually falls on business analysts/experts.
4. Tactical - used by mid-management to track performance.

Strategic Dashboard

A strategic dashboard is a reporting tool for monitoring the long-term company strategy with the help of critical success factors. They're usually complex in their creation, provide an enterprise-wide impact to a business, and are mainly used by senior-level management. Strategic dashboards are commonly used in a wide range of business types while aligning a company's strategic goals. They track performance metrics against enterprise-wide strategic goals. As a result, these dashboards tend to summarize performance over set time frames: past month, quarter, or year. When the strategic dashboard is properly developed, designed, and implemented, it can effectively reduce the amount of time needed to accomplish a specific business key performance indicator, while reducing operational costs.

There are 5 strategic dashboard examples.

a) Management strategic dashboard

This dashboard answers the following: What is my customer base and revenue compared to this time last year?

b) CMO strategic dashboard

Chief Marketing Officers (CMOs) often don't have time to check numbers such as traffic of certain campaigns.

c) SaaS (Software as as Service) management dashboard

d) CFO (Chief financial officers) dashboard for strategic planning

e) Sales KPI dashboard

Operational Dashboard

An operational dashboard is one of the types of dashboards used for monitoring and managing operations that have a shorter time horizon. Since they focus on tracking operational processes, they're usually administered by junior levels of management.

a) Marketing operational dashboard

b) Manufacturing Production dashboard

c) Pick and pack operational dashboard for logistics

d) LinkedIn operations dashboard example

c) Customer service operational metrics dashboard

Analytical Dashboard

An analytical dashboard is a type of dashboard that contains a vast amount of data created and used by analysts to provide support to executives. They supply a business with a comprehensive overview of data, with middle management being a crucial part of its usage.

a) Financial performance dashboard

b) Procurement cost dashboard

c) Healthcare analytical dashboard for patients

d) Analytical retail KPI dashboard

e) KPI Analytics dashboard for FMCG industry

Tactical Dashboard?

A tactical dashboard is utilized in the analysis and monitoring of processes conducted by mid-level management, emphasizing(give special importance or prominence to (something) in speaking or writing) the analysis. Then an organization effectively tracks

the performance of a company's goal and delivers analytic recommendations for future strategies.
a) IT project management dashboard
b) Energy Management tactical dashboard
c) Human Resources talent management dashboard
d) Social media dashboard
e) Supply chain management tactical dashboard

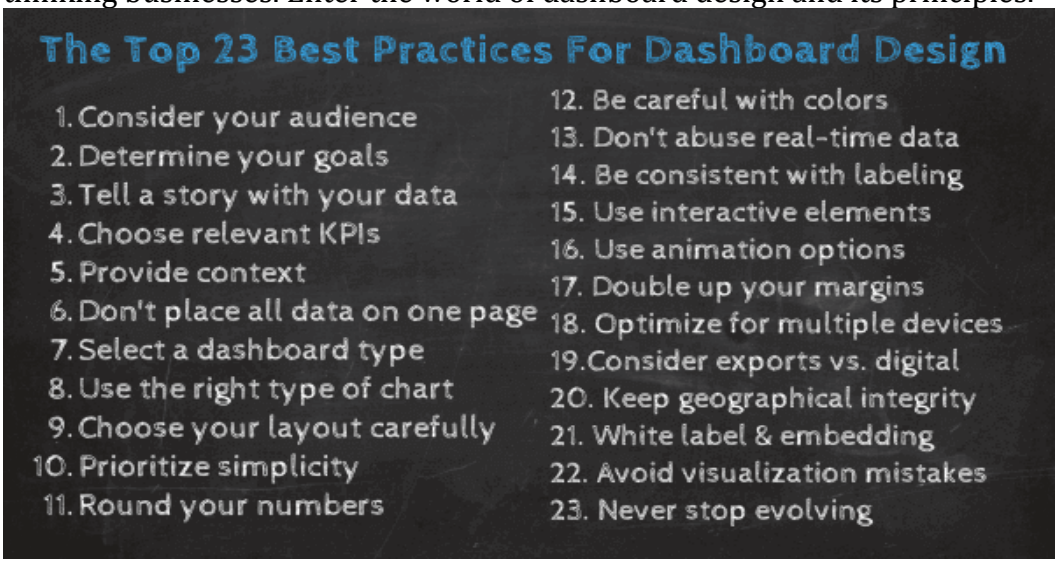
EVOLUTION OF DASHBOARD

For the first three or four hundred years, all a business had to do was make money. Things began to change in the dot-com era as we learned about non-financial business drivers – things like the number of visitors to your site and your market's potential.

Business dashboards emerged into existence four decades back.
In the early 1970s businesses were using decision support system (DSS) to perform their business tasks and this was comparable to the automobile industry where dashboards provided information on the status of critical functions like gasoline and oil levels, speed, engine temperature etc to the driver. Technology revolutionized this display and digital dashboards came into existence. Air crafts went on to 'Glass Cockpits' that displayed all information vital for flight. Dashboards continued to be developed and used in business reviews and were adopted by analytics professionals and managers of global organizations [1].

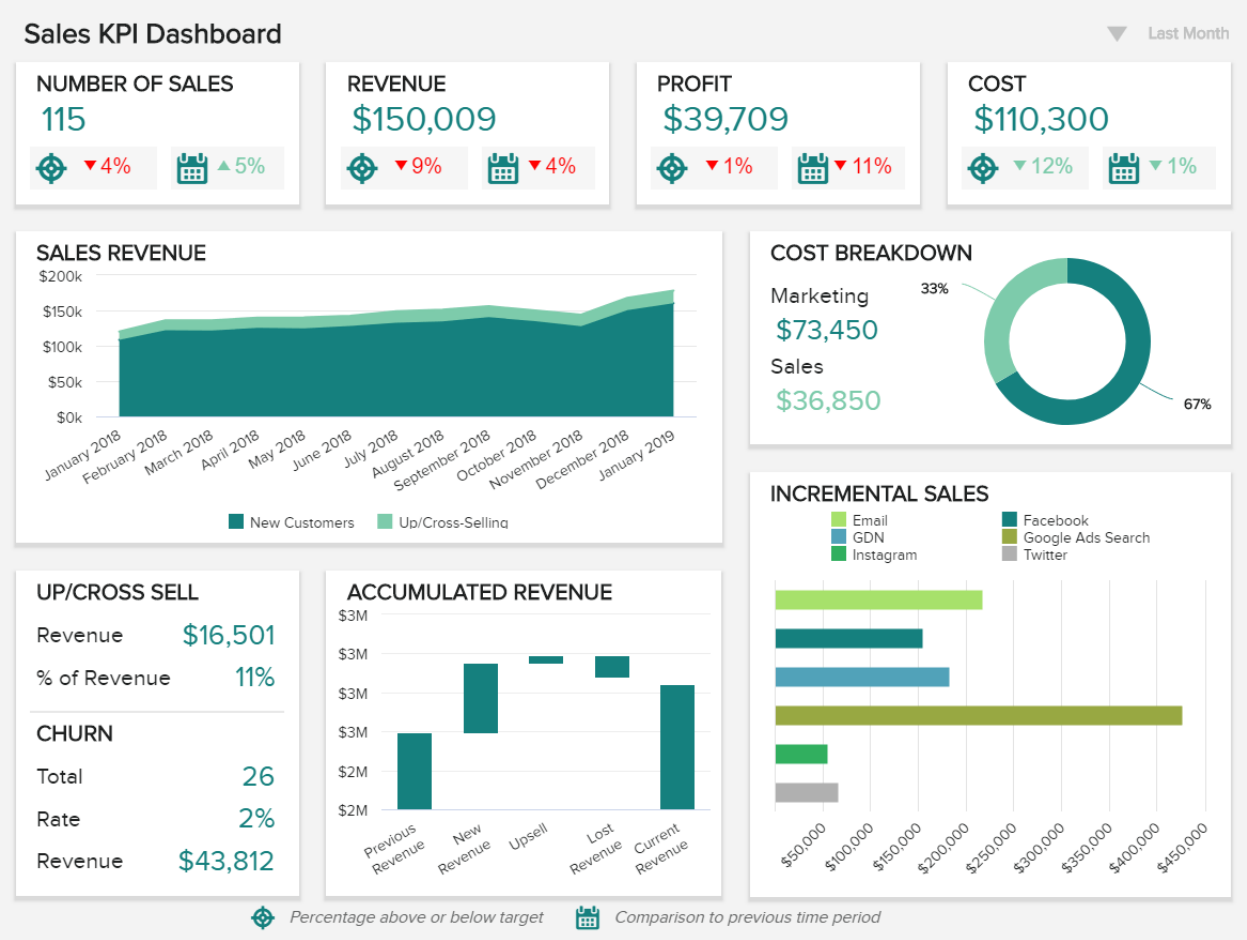
DASHBOARD DESIGN AND PRINCIPLES

The rise of innovative, interactive, data-driven dashboard tools has made creating effective dashboards – like the one featured above – swift, simple, and accessible to today's forward-thinking businesses. Enter the world of dashboard design and its principles.



1. Consider your audience

Concerning dashboard best practices in design, your audience is one of the most important factors you have to take into account. You need to know who's going to use the dashboard and for what purpose they will use it in order to create the best analytical tool for them.

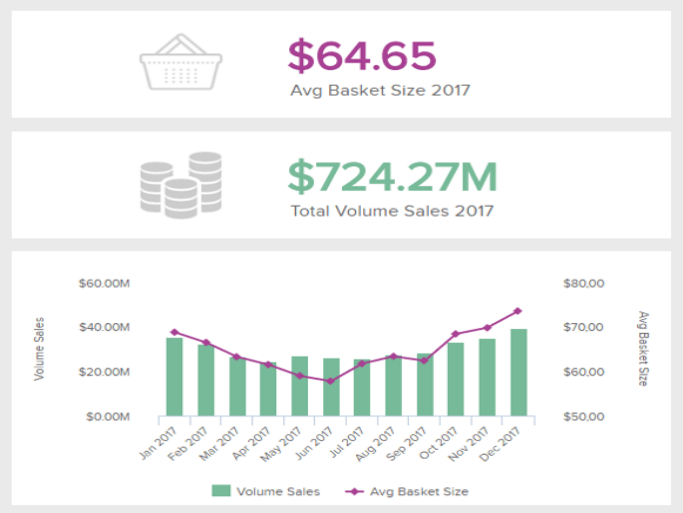


2. Determine your goals

The next dashboard UI design principle has a direct relationship between the user's needs and the purpose of the dashboard, which is to establish your ultimate goals. Rather you are creating a client dashboard or an internal report, each dashboard that you create will serve a purpose and answer key questions through the data. Here, it is important to consider that not all the data available will be useful for the analysis process and that getting this part of the process wrong can render your further efforts meaningless.

3. Choose relevant KPIs

For a truly effective KPI dashboard design, selecting the right key performance indicators (KPIs) for your business needs is a must. Once you've determined your ultimate goals and considered your target audience, you will be able to select the best KPIs to feature in your dashboard.



4. Tell a story with your data

To put it simply, dashboard storytelling is the process of presenting data in a visual manner that will depict the whole narrative of the data analysis process in order to efficiently understand business strategies and goals. In other words, efficient storytelling will help you communicate your message in the clearest way possible.

5. Provide context (Conditions/factors)

Without providing context, how will you know whether those numbers are good or bad, or if they are typical or unusual? Without comparison values, numbers on a dashboard are meaningless for the users. And more importantly, they won't know whether any action is required. For example, a management dashboard design will focus on high-level metrics that are easy to compare and, subsequently, offer a visual story.

6. Don't try to place all the information on the same page

The next in our rundown of dashboard design tips is a question of information. This most golden of dashboard design principles refers to both precision and the right audience targeting.

7. Select the right type of dashboard

Another best practice to consider is to be aware of the type of dashboard that you want to build based on its analytical purpose. As mentioned in previous points, each dashboard should be designed for a particular user group with the specific aim of assisting recipients in the business decision-making process.

For reference, here are the 5 primary types of dashboards for each main branch business-based activity:

- Strategic: A dashboard focused on monitoring long-term company strategies by analyzing and benchmarking a wide range of critical trend-based information.
- Operational: A business intelligence tool that exists to monitor, measure, and manage processes or operations with a shorter or more immediate time scale.
- Analytical: These particular dashboards contain large streams of comprehensive data that allow analysts to drill down and extract insights to help the company to progress at an executive level.
- Platform-specific: As its name suggests, this type of dashboard is used for platform-specific analytics. For example, if you want to track your social media performance

you can use specific metrics and generate a LinkedIn dashboard focusing only on that channel.

- Tactical: These information-rich dashboards are best suited to mid-management and help in formulating growth strategies based on trends, strengths, and weaknesses across departments, such as in the example below:

8. Use the right type of chart

Line charts are great when it comes to displaying patterns of change across a continuum. They are compact, clear, and precise. Line charts format is common and familiar to most people so they can easily be analyzed at a glance.

Choose bar charts if you want to quickly compare items in the same category, for example, page views by country. Again such charts are easy to understand, clear, and compact.

Pie charts aren't the perfect choice. They rank low in precision because users find it difficult to accurately compare the sizes of the pie slices. Although such charts can be instantly scanned and users will notice the biggest slice immediately, there can be a problem in terms of scale resulting in the smallest slices being so small that they even Sparklines usually don't have a scale which means that users will not be able to notice individual values. However, they work well when you have a lot of metrics and you want to show only the trends. They are rapidly scannable and very compact.

It's also not that easy to decipher scatterplots as they are an advanced type of visualization for more knowledgeable users. They aim to find the correlation between two variables. When the data is distributed on the chart, the results show the correlation to be positive, negative, or nonexistent.

Gauge charts are valuable visualizations to provide context. The advantage of these charts lays in the fact that they are easy to interpret as they use various colors to represent different values of the same metric. They are usually used in situations where the expected value is already known, this way the different stakeholders that use the dashboard can understand where they stand just by looking at the gauge chart. For example, to monitor the sales target or sales growth.

Most experts agree that bubble charts are not fit for dashboards. They require too much mental effort from their users even when it comes to reading simple information in a context. Due to their lack of precision and clarity, they are not very common and users are not familiar with them.

9. Choose your layout carefully

Dashboard best practices in design concern more than just good metrics and well-thought-out charts. The next step is the placement of charts on a dashboard. If your dashboard is visually organized, users will easily find the information they need. Poor layout forces users to think more before they grasp the point, and nobody likes to look for data in a jungle of charts and numbers. The general rule is that the key information should be displayed first – at the top of the screen, upper left-hand corner.marketing data, and then again to sales data. This analytics dashboard best practice will enable you to present your data in the most meaningful way and clear to the end-user.

10. Prioritize simplicity

One of the best practices for dashboard design focuses on simplicity. Nowadays, we can play with a lot of options in chart creation and it's tempting to use them all at once. However, try to use those frills sparingly. Frames, backgrounds, effects, gridlines… Yes,

these options might be useful sometimes, but only when there is a reason for applying them.

11. Round your numbers

Continuing on simplicity, rounding the numbers on your dashboard design should be also one of the priorities since you don't want your audience to be flooded with numerous decimal places. Yes, you want to present details but, sometimes, too many details give the wrong impression. If you want to present your conversion rate with 5 more decimal places, it would make sense to round the number and avoid too many number-specific factors. Or, if you want to present your revenue, you don't need to do so by going into cents. 850K looks simpler and more visually effective than $850 010, 25. Especially if you want to implement executive dashboard best practices, where strategic information doesn't need to represent every operational detail of a certain number.

12. Be careful with colors - choose a few and stick to them

Without a shadow of a doubt, this is one of the most important of all dashboard design best practices.

13. Don’t go over the top with real-time data

Next on our list of good dashboard design tips refers to insight: *don’t overuse real-time data*.

14. Be consistent with labeling and data formatting

15. Use interactive elements

16. Additionally, use animation options

Animation options can be one of the dashboard elements that give an additional neat visual impression where you select the appearance of the specific element on the dashboard and assign an animation option. The result is a simple, yet effective automated movement based on the desired speed (slow, medium, or fast,e.g.) and types such as linear, swing, ease-in, or ease-out.

17. Double up your margins

One of the most subtle yet essential dashboard guidelines, this principle boils down to balance. White space – also referred to as negative space – is the area of blankness between elements featured on a dashboard design.

18. Optimize for multiple devices

Optimization for mobile or tablet is another critical point in the dashboard development process. By offering remote access to your most important insights, you can answer critical business questions on the go, without the need for a special office meeting. Benefits such as swift decision-making and instant access ensure everyone can look at the data on the fly.

19. Consider the use in terms of exports vs. digital

In the process of dashboard designing, you also need to think about exports. You can use the dashboard itself and share it, but if you plan on regularly using exports, you might want to consider optimizing towards printing bounds, fewer colors, and different types of line styles to make sure everything is readable even on a black-and-white printout. Hence, when you plan your data dashboard design, you also need to look into the future uses and how to optimize towards different exporting options or simply share the dashboard itself with all its features and options.

Additionally, by assigning viewer roles to users, you can specify the number of features you openly allow, including the number of filters, and all the bits and details of specific

permissions. That way, you have full control over your digital presentation and the amount of analysis you want to share. In this digital case, you don't need to take into account print, but it would certainly help if you ever want to create one.
20. Keep graphical integrity
It might seem like an obvious point, but it is worth mentioning as it is one of the most important dashboard design trends. Graphical integrity basically refers to keeping the truth about the data. This means being objective about the values and not making them 21. White label and embed if you need to
22. Avoid common data visualization mistakes
Data visualization has evolved from simple static presentations to modern interactive software that takes visual perception to the next level. It also enabled average business users and advanced analysts to create stunning visuals that tell a clear data story to any potential audience profile, from beginners in a field to seasoned analysts and strategists.

- Incorrect calculations: The numbers should add up to a total (100%). For example, if you conduct a survey and people have the option to choose more than one answer, you will probably need some other form of visuals than a pie chart since the numbers won't add up, and the viewers might get confused.
- The wrong choice of visualizations: We have mentioned how important it is to choose the right type of chart and dashboard, so if you want to present a relationship between the data, a scatter plot might be the best solution.
- Too much data: Another point you need to keep in mind, and we have discussed in detail, don't put too much data on a single chart because the viewer will not recognize the point.

23. Never stop evolving
Last but certainly not least in our collection of principles of effective dashboards – the ability to tweak and evolve your designs in response to the changes around you will

DISPLAY MEDIA FOR THE DASHBOARD.

Display Media is the most fundamental step in dashboard design. Once you've chosen between text, graphics, or some combination of the two, you must then determine how to organize the text and/or what kinds of graphics to use. Quantitative graphs and several other types of charts that are commonly used in business reporting (for example, process flow and organization charts) work well on dashboards, provided their design is kept clear and simple.
dashboard display media that are used to present actual data. Other display media, such as command buttons, are sometimes needed, It must be the best means to display a particular type of information that is commonly found on dashboards.
The library is divided into six categories:

- Graphs
- Images
- Icons
- Drawing objects
- Text
- Organizers

Graphs

Most dashboard display media fall into the graph category, to represent the quantitative data on most dashboards, this isn't surprising.Different Graphs are:

- Bullet graphs
- Bar graphs (horizontal and vertical)
- Stacked bar graphs (horizontal and vertical)
- Combination bar and line graphs
- Line graphs
- Sparklines
- Box plots
- Scatter plots
- Treemaps

TYPES OF DATA VISUALIZATION

Graph Scatter Plot

A scatter plot is a diagram where each value in the data set is represented by a dot.
A scatter plot is a chart type that is normally used to observe and visually display the relationship between variables.
The values of the variables are represented by dots.
The positioning of the dots on the vertical and horizontal axis will inform the value of the respective data point; hence, scatter plots make use of Cartesian coordinates to display the values of the variables in a data set. Scatter plots are also known as scattergrams, scatter graphs, or scatter charts.

- A scatter plot is a chart type that is normally used to observe and visually display the relationship between variables. It is also known as a scattergram, scatter graph, or scatter chart.
- The data points or dots, which appear on a scatter plot, represent the individual values of each of the data points and also allow pattern identification when looking at the data holistically.
- The most common use of the scatter plot is to display the relationship between two variables and observe the nature of such a relationship. The relationships observed can either be positive or negative, non-linear or linear, and/or, strong or weak.

Scatter Plot Applications and Uses

1. Demonstration of the relationship between two variables
The most common use of the scatter plot is to display the relationship between two variables and observe the nature of the relationship. The relationships observed can either be positive or negative, non-linear or linear, and/or, strong or weak.
The data points or dots, which appear on a scatter plot, represent the individual values of each of those data points and also allow pattern identification when looking at the data holistically.

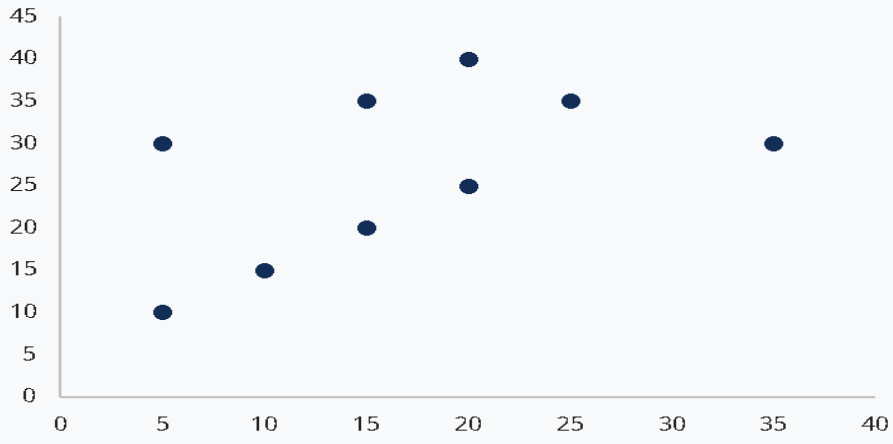
2. Identification of correlational relationships
Another common use of scatter plots is that they enable the identification of correlational(connection between two or more things) relationships. Scatter plots tend to have independent variables on the horizontal axis and dependent variables on the vertical axis.
3. Identification of data patterns
Data pattern identification is also possible with scatter plots. Data points can be grouped together based on how close their values are, and this also makes it easy to identify any outlier points when there are data gaps.

Creating a Scatter Plot Diagram

Raw Data

Month	Series 1	Series 2
Sep	30	5
Aug	35	15
Jul	40	20
Jun	35	25
May	30	35
Apr	25	20
Mar	20	15
Feb	15	10
Jan	10	5

The scatter plot diagram for the data above is seen below:



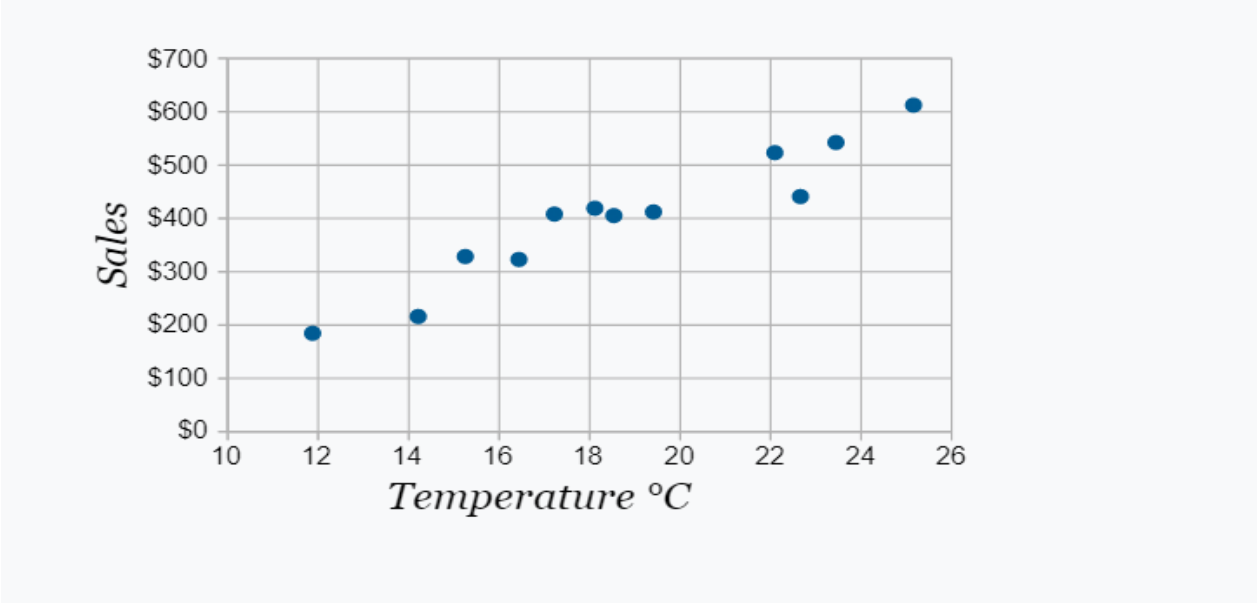
A scatter plot is a chart type that is normally used to observe and visually display the relationship between variables. The values of the variables are represented by dots. The positioning of the dots on the vertical and horizontal axis will inform the value of the respective data point; hence, scatter plots make use of Cartesian coordinates to display the values of the variables in a data set. Scatter plots are also known as scattergrams, scatter graphs, or scatter charts.

The local ice cream shop keeps track of how much ice cream they sell versus the noon temperature on that day. Here are their figures for the last 12 days:

Ice Cream Sales vs Temperature

Temperature °C	Ice Cream Sales
14.2°	$215
16.4°	$325
11.9°	$185
15.2°	$332
18.5°	$406
22.1°	$522
19.4°	$412
25.1°	$614
23.4°	$544
18.1°	$421
22.6°	$445
17.2°	$408

And here is the same data as a Scatter Plot:



HISTOGRAM

A histogram is a graphical representation that organizes a group of data points into user-specified ranges.

Similar in appearance to a bar graph, the histogram represents a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges

A histogram is the graphical representation of data where data is grouped into continuous number ranges and each range corresponds to a vertical bar.

- The horizontal axis displays the number range.
- The vertical axis (frequency) represents the amount of data that is present in each range.

The number ranges depend upon the data that is being used.

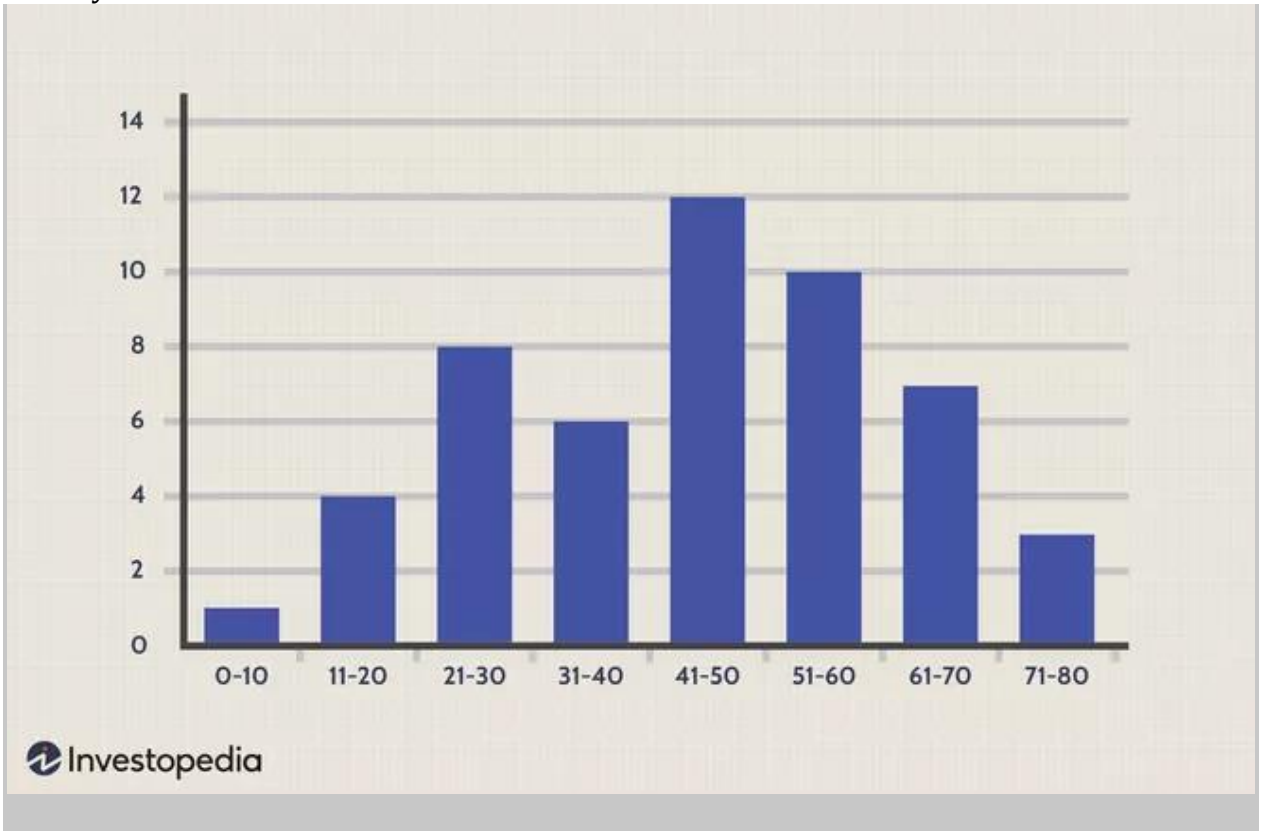
- A histogram is a bar graph-like representation of data that buckets a range of outcomes into columns along the x-axis.
- The y-axis represents the number count or percentage of occurrences in the data for each column and can be used to visualize data distributions.
- In trading, the MACD histogram is used by technical analysts to indicate changes in momentum.

How Histograms Work

Histograms are commonly used in statistics to demonstrate how many of a certain type of variable occurs within a specific range. For example, a census focused on the demography of a country may use a histogram to show how many people are between the ages of 0 - 10, 11 - 20, 21 - 30, 31 - 40, 41 - 50, etc. This histogram would look similar to the example below.

Histograms can be customized in several ways by the analyst. The first is to change the interval between buckets. In the above example, there are 5 buckets with an interval of ten. This could be changed, for example, to 10 buckets with an interval of 5 instead.

The other consideration is how to define the y-axis. The most basic label is to use the frequency of occurrences observed in the data, but one could also use percentage of total or density instead.



Histogram Graph

A histogram graph is a bar graph representation of data. It is a representation of a range of outcomes into columns formation along the x-axis. in the same histogram, the number count or multiple occurrences in the data for each column is represented by the y-axis. It is the easiest manner that can be used to visualize data distributions. Let us understand the histogram graph by plotting one for the given below example.

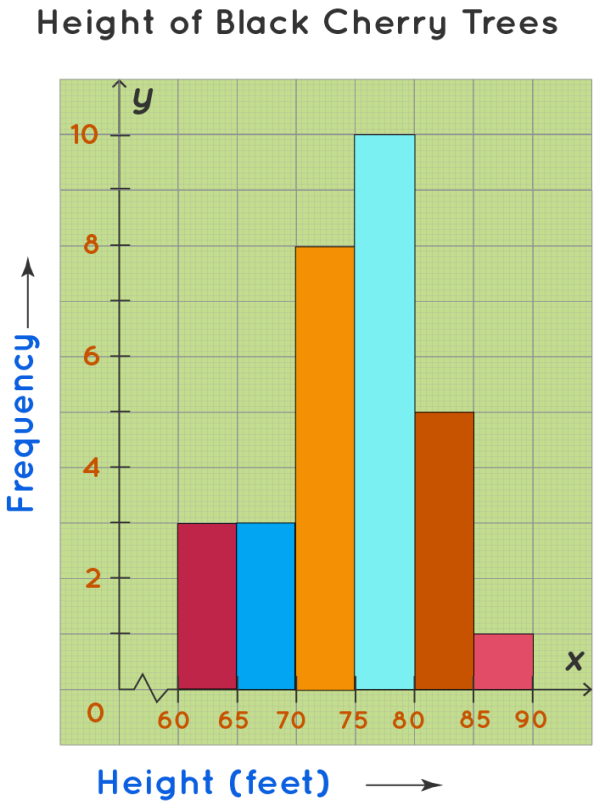
Uncle Bruno owns a garden with 30 black cherry trees. Each tree is of a different height. The height of the trees (in inches): 61, 63, 64, 66, 68, 69, 71, 71.5, 72, 72.5, 73, 73.5, 74, 74.5, 76, 76.2, 76.5, 77, 77.5, 78, 78.5, 79, 79.2, 80, 81, 82, 83, 84, 85, 87. We can group the data as follows in a frequency distribution table by setting a range:

Height Range (ft)	Number of Trees (Frequency)
60 - 65	3
66 - 70	3

71 - 75	8
76 - 80	10
81 - 85	5
86 - 90	1

This data can be now shown using a histogram. We need to make sure that while plotting a histogram, there shouldn't be any gaps between the bars.

Histogram



How to Make a Histogram?

The process of making a histogram using the given data is described below:

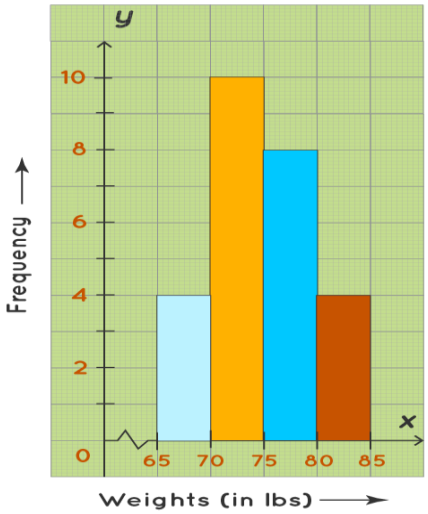
- Step 1: Choose a suitable scale to represent weights on the horizontal axis.
- Step 2: Choose a suitable scale to represent the frequencies on the vertical axis.
- Step 3: Then draw the bars corresponding to each of the given weights using their frequencies.

Example: Construct a histogram for the following frequency distribution table that describes the frequencies of weights of 25 students in a class.

Weights (in lbs)	Frequency (Number of students)
65 - 70	4
70 - 75	10
75 - 80	8
80 - 85	4

Steps to draw a histogram:

- Step 1: On the horizontal axis, we can choose the scale to be 1 unit = 11 lb. Since the weights in the table start from 65, not from 0, we give a break/kink on the X-axis.
- Step 2: On the vertical axis, the frequencies are varying from 4 to 10. Thus, we choose the scale to be 1 unit = 2.
- Step 3: Then draw the bars corresponding to each of the given weights using their frequencies.

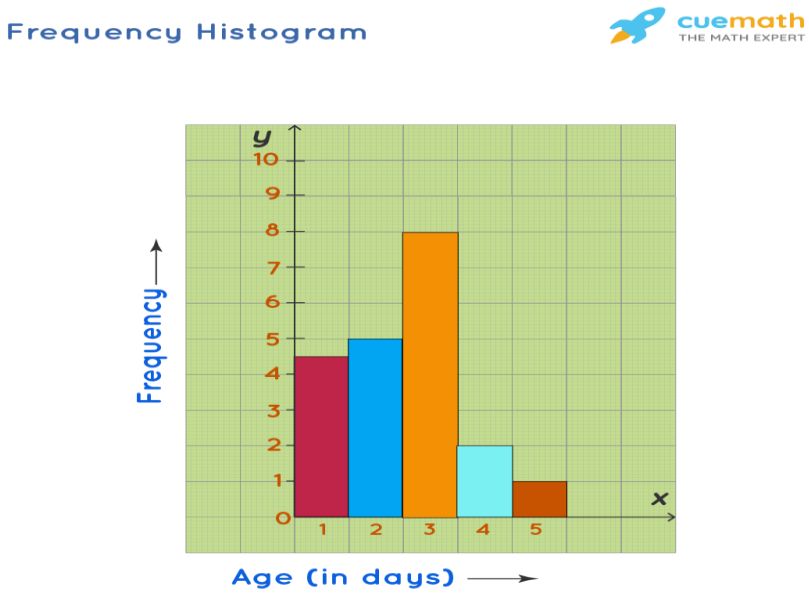


Frequency Histogram

A frequency histogram is a histogram that shows the frequencies (the number of occurrences) of the given data items. For example, in a hospital, there are 20 newborn babies whose ages in increasing order are as follows: 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 5. This information can be shown in a frequency distribution table as follows:

Age (in days)	Frequency
1	4
2	5
3	8
4	2
5	1

This data can be now shown using a frequency histogram.



Histogram Shapes

The histogram can be classified into different types based on the frequency distribution of the data. There are different types of distributions, such as normal distribution, skewed distribution, bimodal distribution, multimodal distribution, comb distribution, edge peak distribution, dog food distribution, heart cut distribution, and so on. The histogram can be used to represent these different types of distributions. We have mainly 5 types of histogram shapes. They are listed below:

1. Bell Shaped Histogram
2. Bimodal Histogram
3. Skewed Right Histogram
4. Skewed Left Histogram
5. Uniform Histogram

PLOTS

A plot is a graphical technique for representing a data set, usually as a graph showing the relationship between two or more variables. The plot can be drawn by hand or by a computer. In the past, sometimes mechanical or electronic plotters were used. Graphs are a visual representation of the relationship between variables, which are very useful for humans who can then quickly derive an understanding which may not have come from lists of values

Plots play an important role in statistics and data analysis

There are also many statistical tools generally referred to as graphical techniques. These include:[1]

- scatter plots
- spectrum plots
- histograms
- probability plots
- residual plots
- box plots, and
- block plots

GRAPHS

Graph Data Science is a science-driven approach to gain knowledge from the relationships and structures in data, typically to power predictions. It describes a toolbox of techniques that help data scientists answer questions and explain outcomes using graph data.

Applications of Graph Data Science

Graph Data Science techniques can be used as part of a variety of different applications and use cases.

- Graph queries support domain experts by answering common questions.
- Graph algorithms help make sense of the global structure of a graph, and the results used for standalone analysis or as features in a machine learning model.
- Graph embeddings are a core component of similarity graphs that power recommendation systems.
- Natural Language Processing techniques support content based filtering recommendations and knowledge graph completion.

NETWORKS

network analysis is a technique that uses graph theory to study complex real-world problems, such as computational biology, engineering, finance, marketing, neuroscience, political science, and public health

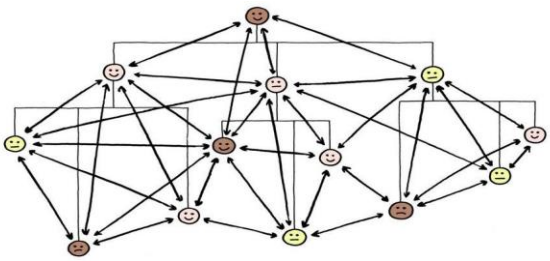
A network refers to an object composed of elements and relationships or connections between those elements.

Network analysis is a collection of techniques for examining the relationships between entities, and depicting the structure of those relationships. Network analysis spans a

number of domains, including social networks, bibliometrics, epidemiology, bioinformatics, complex systems, and text analysis.
Graph theory provides the formal basis for network analysis, across domains, and provides a common language for describing the structure of networks.
Network visualization involves the visualization of the relationships (edges or links) between data elements (nodes).

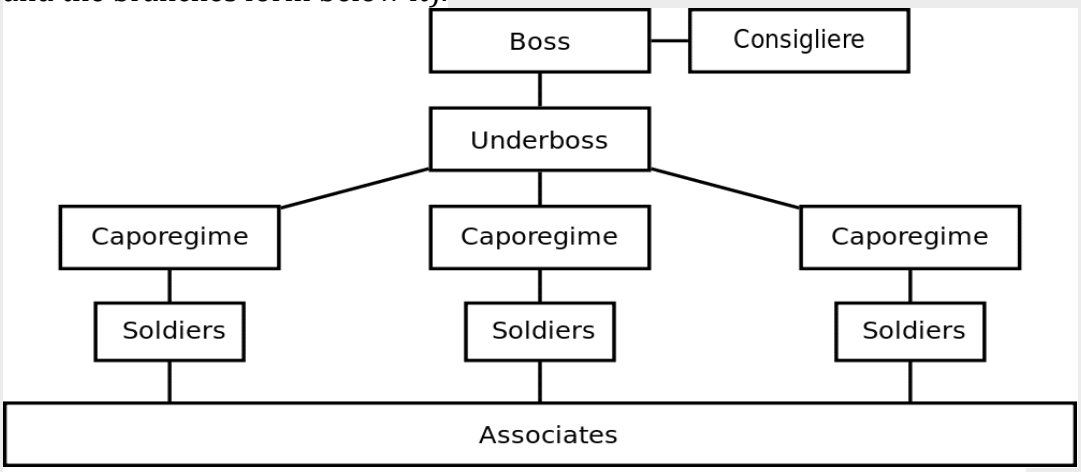
Programming Language	Network Visualization and Analysis Library
Python	NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.
	python-igraph is a Python connector to the igraph collection of network analysis tools.
R	igraph is an R connector to the igraph collection of network analysis tools.
	visNetwork is an R package for interactive network visualization, built on the vis.js Javascript library.
Julia	JuliaGraphs is an aggregation of all the common packages in Julia for using and studying graphs, including visualization using the GraphPlot and NetworkLayout packages.

HIERARCHICAL
Hierarchical data, a very special type of network data, is represented by hierarchical data visualization.
Typically, the principle of connection represents network data. However, hierarchical data relies on the principle of containment!
A classic example of hierarchical data visualization is the file and folder system found on your computer. You have a folder that contains more folders.
Other common types of hierarchical data visualization forms are tree diagrams, cone tree diagrams, botanical tree diagrams and treemap diagrams.



Hierarchical data is essentially a specialized form of network data – in that while entities within the dataset do not have dependent relationships; they are all related to each other

by the principle of containment. They, unlike standard data networks, do not use the principle of connection.
A hierarchy begins with a root entity. This might be the CEO of a company, the name of a book, the title of a folder, etc. and then the root entity has at least one "child node" and every further child node has zero or more children.
An entity which comes below another is a child node to the entity above. Similarly, an entity which comes above another is a parent node to the node below.
Hierarchical data is shown in tree graphs; so called because of their similarity to a tree's structure (though a tree which has been turned upside down so that the root is at the top and the branches form below it).



Author/Copyright holder: MaEr. Copyright terms and licence: Public Domain.
Above we see a simple tree diagram for the structure of a mafia family. The root entity is the boss of the family and the underboss is the first child entity. This is a very basic hierarchical relationship and it is possible to map much more complex hierarchies using information visualization techniques.

REPORTS
A data analysis report is somewhat different from other types of professional writing that you may have done or seen, or will learn about in the future. It is related to but not the same as:
• A typical psych/social science paper orgainzed around "intro/methods/analysis/results/discussion" sections.
• A research article in an academic journal.
• An essay.
• A lab report in a science class.
The overall structure of a data analysis report is simple:
1. Introduction
2. Body
3. Conclusion(s)/Discussion
4. Appendix/Appendices The data analysis report is written for several different audiences at the same time

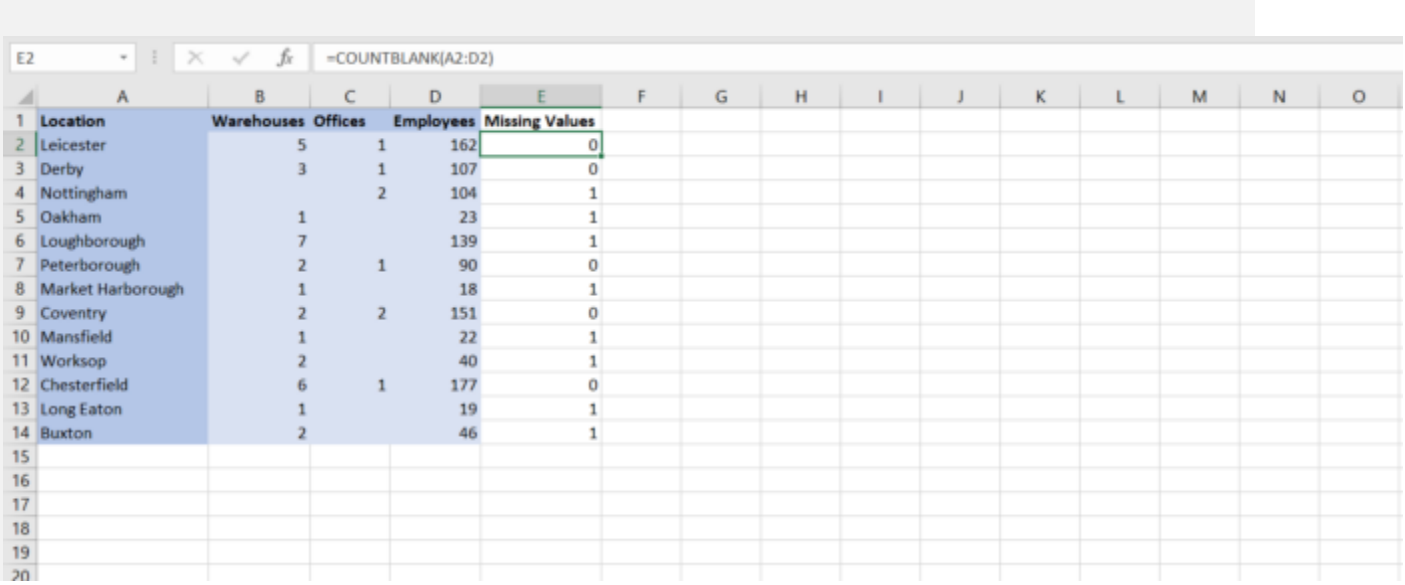
DATA SCIENCE WITH MS-EXCEL

Microsoft Excel provides several means and ways to analyze and interpret data. The data can be from various sources. The data can be converted and formatted in several ways. It can be analyzed with the relevant Excel commands, functions and tools - encompassing Conditional Formatting, Ranges, Tables, Text functions, Date functions, Time functions, Financial functions, Subtotals, Quick Analysis, Formula Auditing, Inquire Tool, What-if Analysis, Solvers, Data Model, PowerPivot, PowerView, PowerMap, etc

Following are some of the important function used in Excel for Data Science

COUNTBLANK

Raw data can be messy, and sometimes it can include missing values, especially when this data is collected in the real world. In Python, we can use a method such as isna() to detect missing values. However, this is also possible with Excel using the COUNTBLANK function. COUNTBLANK will count the number of blank cells in a given range.



	Location	Warehouses	Offices	Employees	Missing Values
2	Leicester	5	1	162	0
3	Derby	3	1	107	0
4	Nottingham		2	104	1
5	Oakham	1		23	1
6	Loughborough	7		139	1
7	Peterborough	2	1	90	0
8	Market Harborough	1		18	1
9	Coventry	2	2	151	0
10	Mansfield	1		22	1
11	Worksop	2		40	1
12	Chesterfield	6	1	177	0
13	Long Eaton	1		19	1
14	Buxton	2		46	1

Using the COUNTBLANK function to check for missing values. Image by author.

The formula for this Excel function is:

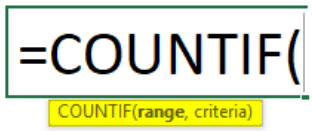
=COUNTBLANK(range)

COUNTIF

COUNTIF Formula in excel is an inbuilt or pre-built integrated function which is categorized under the statistical group of formulae.

Excel COUNTIF Formula counts the number of cells within a specified array or range based on a specific criterion or applied condition.

Below is the Syntax of the COUNTIF Formula in Excel :



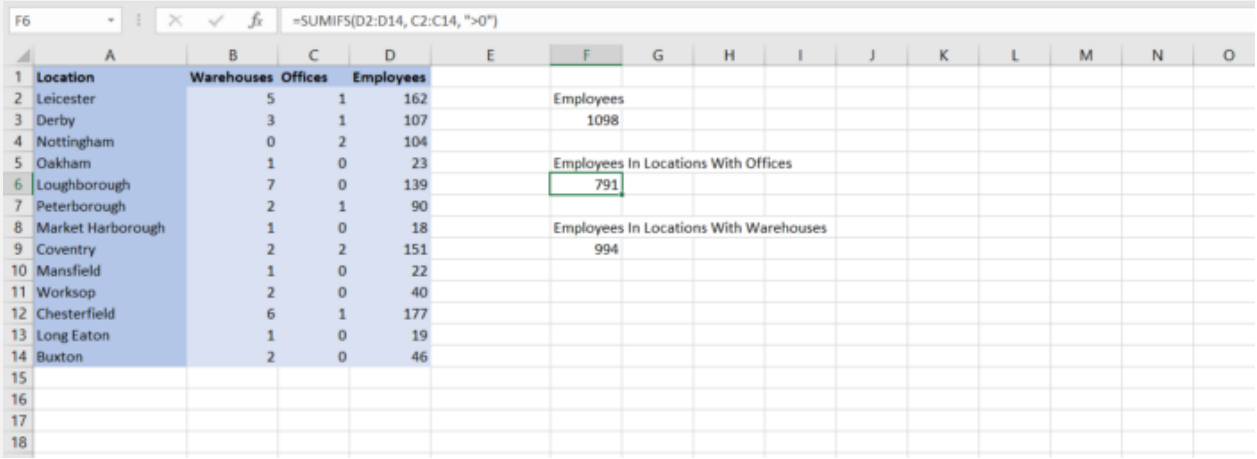
The COUNTIF Formula in Excel has two arguments, i.e. range, criteria.

1. **Range**: (Required & Compulsory argument): The range or array of cells on which the criteria should be applied, here array or range should be mentioned as, e.g. C1:C12.
2. **Criteria:** (Required & Compulsory argument)) It is a specific condition which will be applied to the cell values presented by the range of cells. It indicates the Count IF formula "what cells need to be counted" here. There is a limitation of the criteria argument in the COUNTIF Formula, i.e. If the text string in the criteria argument contains greater than 255 characters in length or more than 255 characters. Then it returns **#VALUE! error.**

SUMIFS

One of the best-known and most basic Excel functions is the SUM function. Sometimes we may want to use SUM, but also exclude cells from our summation using a criterion. This is where SUMIFS comes in.

SUMIFS allows us to sum values in a given range, but only sums the values that match given criteria. As many criteria can be given to this function as desired.



Using the SUMIFS function to count employees in different locations. Image by author.

The formula for this Excel function is:

=SUMIFS(sum_range, criteria_range1, criteria1, ...)

RANK

The RANK function can be used to return the rank of a numerical value when compared to a list of other numerical values. Ranking can be very important in data analysis to know where a particular value would fall in an ordered array. There also exists a similar function known as PERCENTRANK which returns the ranking as a percentage of the dataset's range. It is important to note that if a value cannot be found in the given list or array, Excel will return a #N/A error. This is not true for PERCENTRANK.

E2 =RANK(D2, D$2:D$14)

	A	B	C	D	E
1	Location	Warehouses	Offices	Employees	Employees Rank
2	Leicester	5	1	162	2
3	Derby	3	1	107	5
4	Nottingham	0	2	104	6
5	Oakham	1	0	23	10
6	Loughborough	7	0	139	4
7	Peterborough	2	1	90	7
8	Market Harborough	1	0	18	13
9	Coventry	2	2	151	3
10	Mansfield	1	0	22	11
11	Worksop	2	0	40	9
12	Chesterfield	6	1	177	1
13	Long Eaton	1	0	19	12
14	Buxton	2	0	46	8

Using the RANK function to rank locations by the number of employees. Image by author.
The formula for this Excel function is:
=RANK(number, list)

VLOOKUP

VLOOKUP is one of the most important functions for any data analyst to know. It can be used to retrieve, or look up, data in a table that is organised vertically. This is incredibly useful as it can be used to automatically find data in another spreadsheet, as long as each row has an ID.

The ID column for the table must be the first column. The third parameter in the VLOOKUP function can then be used to refer to the column containing the data that you want to retrieve. It is important to note that this parameter is 1-based in Excel. This means that a value of 2 will get the 2nd column, and a value of 3 will get the third column, and so on. This is different to list indexing in Python, which is 0-based.

The other two parameters in the VLOOKUP function refer to the lookup value (or the ID of the row you are retrieving) and the range of the table itself.

H5 =VLOOKUP(H2, A2:D14, 4, 0)

	A	B	C	D	G	H
1	Location	Warehouses	Offices	Employees		
2	Leicester	5	1	162	Location	Peterborough
3	Derby	3	1	107	Warehouses	TRUE
4	Nottingham	0	2	104	Offices	TRUE
5	Oakham	1	0	23	Employees	90
6	Loughborough	7	0	139		
7	Peterborough	2	1	90		
8	Market Harborough	1	0	18		
9	Coventry	2	2	151		
10	Mansfield	1	0	22		
11	Worksop	2	0	40		
12	Chesterfield	6	1	177		
13	Long Eaton	1	0	19		
14	Buxton	2	0	46		

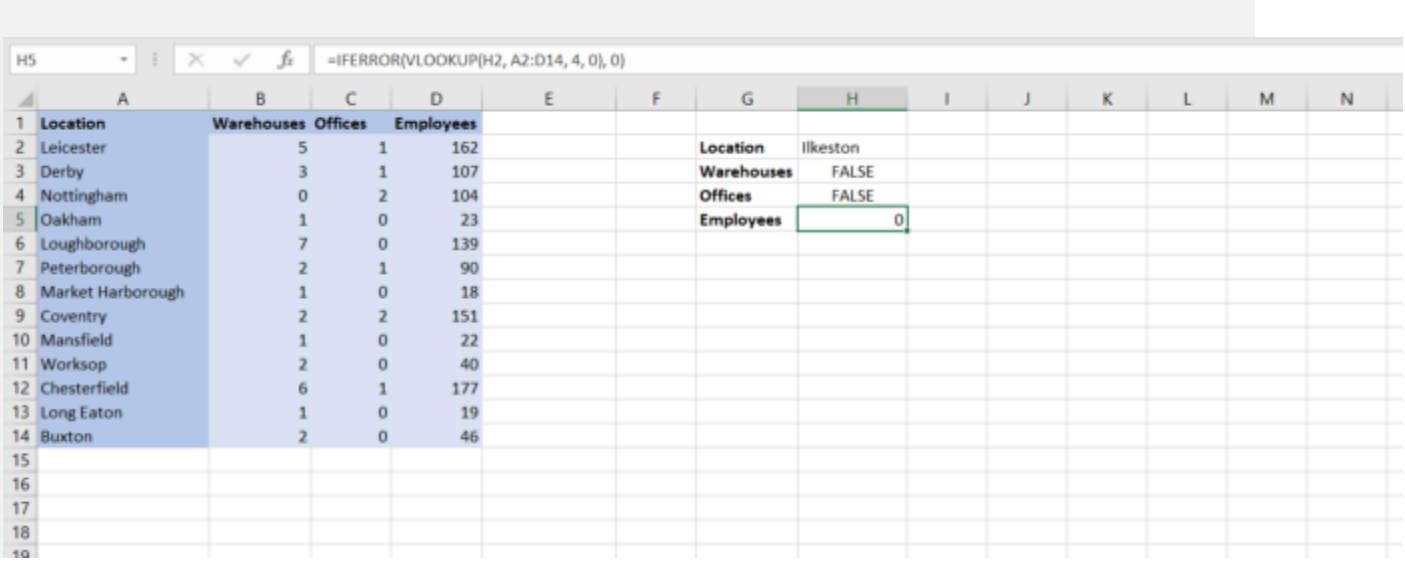
Using the VLOOKUP function to find data on different locations. Image by author.
The formula for this Excel function is:

=VLOOKUP(lookup_value, table_array, col_index_number)

IFERROR

If you are presenting your Excel spreadsheets to other people in your business, then it can be useful to have default values in cells when errors occur in your functions. This can be particularly useful when other users do not understand the meaning of errors, or the value of the cell would be known if an error occurred in calculations (such as defaulting to 0).

The IFERROR function can do exactly that. This function simply takes two parameters, the first is the function, and the second is the default value if the function throws an error. If the function does not throw an error, then the value will be outputted as normal.



	A	B	C	D
1	**Location**	**Warehouses**	**Offices**	**Employees**
2	Leicester	5	1	162
3	Derby	3	1	107
4	Nottingham	0	2	104
5	Oakham	1	0	23
6	Loughborough	7	0	139
7	Peterborough	2	1	90
8	Market Harborough	1	0	18
9	Coventry	2	2	151
10	Mansfield	1	0	22
11	Worksop	2	0	40
12	Chesterfield	6	1	177
13	Long Eaton	1	0	19
14	Buxton	2	0	46

	G	H
1		
2	**Location**	Ilkeston
3	**Warehouses**	FALSE
4	**Offices**	FALSE
5	**Employees**	0

Using the IFERROR function to handle Excel errors. Image by author.

The formula for this Excel function is:

=IFERROR(value, value_if_error)

DAYS

If you have ever worked with data involving time, you will know that there are several key calculations that will keep reoccurring in your work. One of these is calculating the number of days between two dates. In Excel, we can do this with the DAYS functions.

The DAYS function takes two dates as parameters and returns the number of days between them as an integer.

One scenario where this function would be useful is calculating the time between when a product is ordered, and when that product is delivered. This is an important metric for businesses that sell products, and so it is great that Excel provides an easy way to calculate it.

I6 =DAYS(H6,G6)

Order ID	Product ID	Product Description	Quantity	Quantity Price	Overall Price	Order Date	Delivery Date	Days To Deliver
128	662-134	Wooden Table	1	£12.50	£12.50	2021-01-04	2021-01-05	1
129	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-04	2021-01-08	4
130	132-737	Mesh Bin	3	£3.25	£9.75	2021-01-05	2021-01-07	2
131	662-134	Wooden Table	1	£12.50	£12.50	2021-01-05	2021-01-07	2
132	662-134	Wooden Table	2	£12.50	£25.00	2021-01-05	2021-01-11	6
133	356-441	Leather Chair	1	£24.00	£24.00	2021-01-05	2021-01-09	4
134	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-10	4
135	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-06	2021-01-10	4
136	721-180	Wooden Crate	12	£6.00	£72.00	2021-01-06	2021-01-08	2
137	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-08	2
138	356-441	Leather Chair	3	£24.00	£72.00	2021-01-06	2021-01-13	7
139	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-07	2021-01-18	11
140	662-095	Metal Table	3	£18.00	£54.00	2021-01-07	2021-01-12	5
141	662-134	Wooden Table	1	£12.50	£12.50	2021-01-07	2021-01-16	9
142	662-095	Metal Table	1	£18.00	£18.00	2021-01-08	2021-01-13	5
143	662-134	Wooden Table	2	£12.50	£25.00	2021-01-08	2021-01-16	8
144	356-441	Leather Chair	4	£24.00	£96.00	2021-01-08	2021-01-16	8
145	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-08	2021-01-16	8
146	721-180	Wooden Crate	8	£6.00	£48.00	2021-01-08	2021-01-12	4
147	662-134	Wooden Table	1	£12.50	£12.50	2021-01-09	2021-01-10	1
148	829-001	Two Door Wardrobe	2	£62.00	£124.00	2021-01-09	2021-01-13	4
149	662-095	Metal Table	2	£18.00	£36.00	2021-01-10	2021-01-14	4
150	662-134	Wooden Table	1	£12.50	£12.50	2021-01-10	2021-01-16	6
151	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-10	2021-01-11	1

Using the DAYS function to calculate days to deliver a product. Image by author.

The formula for this Excel function is:

=DAYS(end_date, start_date)

MAXIFS

Finding the maximum value that a variable can take is extremely important in business. From knowing the busiest day of the year to the maximum amount of profit a business made in a day — there are many reasons to calculate the maximum.

Sometimes, you might only want to calculate the maximum on a subset of your data. In that case, you can use the MAXIFS function to put constraints on the data that you want to take the maximum of.

In the example below, we use MAXIFS to find the maximum days to deliver each product. The constraint is placed on the product in each row of our dataset.

M3 =MAXIFS($I$2:$I$25, $C$2:$C$25, L3)

Order ID	Product ID	Product Description	Quantity	Quantity Price	Overall Price	Order Date	Delivery Date	Days To Deliver
128	662-134	Wooden Table	1	£12.50	£12.50	2021-01-04	2021-01-05	1
129	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-04	2021-01-08	4
130	132-737	Mesh Bin	3	£3.25	£9.75	2021-01-05	2021-01-07	2
131	662-134	Wooden Table	1	£12.50	£12.50	2021-01-05	2021-01-07	2
132	662-134	Wooden Table	2	£12.50	£25.00	2021-01-05	2021-01-11	6
133	356-441	Leather Chair	1	£24.00	£24.00	2021-01-05	2021-01-09	4
134	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-10	4
135	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-06	2021-01-10	4
136	721-180	Wooden Crate	12	£6.00	£72.00	2021-01-06	2021-01-08	2
137	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-08	2
138	356-441	Leather Chair	3	£24.00	£72.00	2021-01-06	2021-01-13	7
139	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-07	2021-01-18	11
140	662-095	Metal Table	3	£18.00	£54.00	2021-01-07	2021-01-12	5
141	662-134	Wooden Table	1	£12.50	£12.50	2021-01-07	2021-01-16	9
142	662-095	Metal Table	1	£18.00	£18.00	2021-01-08	2021-01-13	5
143	662-134	Wooden Table	2	£12.50	£25.00	2021-01-08	2021-01-16	8
144	356-441	Leather Chair	4	£24.00	£96.00	2021-01-08	2021-01-16	8
145	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-08	2021-01-16	8
146	721-180	Wooden Crate	8	£6.00	£48.00	2021-01-08	2021-01-12	4
147	662-134	Wooden Table	1	£12.50	£12.50	2021-01-09	2021-01-10	1
148	829-001	Two Door Wardrobe	2	£62.00	£124.00	2021-01-09	2021-01-13	4
149	662-095	Metal Table	2	£18.00	£36.00	2021-01-10	2021-01-14	4
150	662-134	Wooden Table	1	£12.50	£12.50	2021-01-10	2021-01-16	6
151	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-10	2021-01-11	1

Max Days To Deliver	
Wooden Table	9
Two Door Wardrobe	11
Mesh Bin	2
Leather Chair	8
Wooden Crate	4
Metal Table	5

Using the MAXIFS function to get maximum delivery times for each product. Image by author.

The formula for this Excel function is:

=MAXIFS(max_range, criteria_range_1, criteria_1, …)

AVERAGEIFS

Earlier we explored the use of SUMIFS to get filtered sums of our data. We have also looked at MAXIFS to do the same but with the maximum values. By now you may have worked out that there are many different IF functions that can be used in Excel.

Some other useful IF functions include COUNTIFS, MINIFS, and the plain-old IF function. One function that does appear often is the AVERAGEIFS function, which is used to calculate averages.

In the example below, we use the AVERAGEIFS function to calculate the average delivery time for each product.

M3 =AVERAGEIFS($I$2:$I$25, $C$2:$C$25, L3)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Order ID	Product ID	Product Description	Quantity	Quantity Price	Overall Price	Order Date	Delivery Date	Days To Deliver				
2	128	662-134	Wooden Table	1	£12.50	£12.50	2021-01-04	2021-01-05	1			Average Days To Deliver	
3	129	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-04	2021-01-08	4			Wooden Table	4.33
4	130	132-737	Mesh Bin	3	£3.25	£9.75	2021-01-05	2021-01-07	2			Two Door Wardrobe	5.33
5	131	662-134	Wooden Table	1	£12.50	£12.50	2021-01-05	2021-01-07	2			Mesh Bin	2.00
6	132	662-134	Wooden Table	2	£12.50	£25.00	2021-01-05	2021-01-11	6			Leather Chair	6.33
7	133	356-441	Leather Chair	1	£24.00	£24.00	2021-01-05	2021-01-09	4			Wooden Crate	3.00
8	134	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-10	4			Metal Table	4.67
9	135	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-06	2021-01-10	4				
10	136	721-180	Wooden Crate	12	£6.00	£72.00	2021-01-06	2021-01-08	2				
11	137	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-08	2				
12	138	356-441	Leather Chair	3	£24.00	£72.00	2021-01-06	2021-01-13	7				
13	139	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-07	2021-01-18	11				
14	140	662-095	Metal Table	3	£18.00	£54.00	2021-01-07	2021-01-12	5				
15	141	662-134	Wooden Table	1	£12.50	£12.50	2021-01-07	2021-01-16	9				
16	142	662-095	Metal Table	1	£18.00	£18.00	2021-01-08	2021-01-13	5				
17	143	662-134	Wooden Table	2	£12.50	£25.00	2021-01-08	2021-01-16	8				
18	144	356-441	Leather Chair	4	£24.00	£96.00	2021-01-08	2021-01-16	8				
19	145	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-08	2021-01-16	8				
20	146	721-180	Wooden Crate	8	£6.00	£48.00	2021-01-08	2021-01-12	4				
21	147	662-134	Wooden Table	1	£12.50	£12.50	2021-01-09	2021-01-10	1				
22	148	829-001	Two Door Wardrobe	2	£62.00	£124.00	2021-01-09	2021-01-13	4				
23	149	662-095	Metal Table	2	£18.00	£36.00	2021-01-10	2021-01-14	4				
24	150	662-134	Wooden Table	1	£12.50	£12.50	2021-01-10	2021-01-16	6				
25	151	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-10	2021-01-11	1				
26													
27													

Using the AVERAGEIFS function to get average delivery times for each product. Image by author.

The formula for this Excel function is:

=AVERAGEIFS(average_range, criteria_range_1, criteria_1, …)

MATCH

Sometimes it can be useful to know the specific column or row that a data value occurs in. We will look at one use-case later in the article, but first, we will introduce the function that allows us to do it.

The MATCH function is used to determine the position of a value in a given array. It is important to remember that this result will be relative to the start of the array, so a result of 4 does not necessarily mean the value you were looking for is in the 4th column or row of the spreadsheet; it would be in the 4th position of the given range.

L3 =MATCH(11, I2:I25)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Order ID	Product ID	Product Description	Quantity	Quantity Price	Overall Price	Order Date	Delivery Date	Days To Deliver					
2	128	662-134	Wooden Table	1	£12.50	£12.50	2021-01-04	2021-01-05	1					
3	129	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-04	2021-01-08	4			12		
4	130	132-737	Mesh Bin	3	£3.25	£9.75	2021-01-05	2021-01-07	2			Two Door Wardrobe		
5	131	662-134	Wooden Table	1	£12.50	£12.50	2021-01-05	2021-01-07	2					
6	132	662-134	Wooden Table	2	£12.50	£25.00	2021-01-05	2021-01-11	6					
7	133	356-441	Leather Chair	1	£24.00	£24.00	2021-01-05	2021-01-09	4					
8	134	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-10	4					
9	135	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-06	2021-01-10	4					
10	136	721-180	Wooden Crate	12	£6.00	£72.00	2021-01-06	2021-01-08	2					
11	137	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-08	2					
12	138	356-441	Leather Chair	3	£24.00	£72.00	2021-01-06	2021-01-13	7					
13	139	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-07	2021-01-18	11					
14	140	662-095	Metal Table	3	£18.00	£54.00	2021-01-07	2021-01-12	5					
15	141	662-134	Wooden Table	1	£12.50	£12.50	2021-01-07	2021-01-16	9					
16	142	662-095	Metal Table	1	£18.00	£18.00	2021-01-08	2021-01-13	5					
17	143	662-134	Wooden Table	2	£12.50	£25.00	2021-01-08	2021-01-16	8					
18	144	356-441	Leather Chair	4	£24.00	£96.00	2021-01-08	2021-01-16	8					
19	145	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-08	2021-01-16	8					
20	146	721-180	Wooden Crate	8	£6.00	£48.00	2021-01-08	2021-01-12	4					
21	147	662-134	Wooden Table	1	£12.50	£12.50	2021-01-09	2021-01-10	1					
22	148	829-001	Two Door Wardrobe	2	£62.00	£124.00	2021-01-09	2021-01-13	4					
23	149	662-095	Metal Table	2	£18.00	£36.00	2021-01-10	2021-01-14	4					
24	150	662-134	Wooden Table	1	£12.50	£12.50	2021-01-10	2021-01-16	6					
25	151	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-10	2021-01-11	1					
26														
27														

Using the MATCH function to get a specific row in the dataset. Image by author.

The formula for this Excel function is:

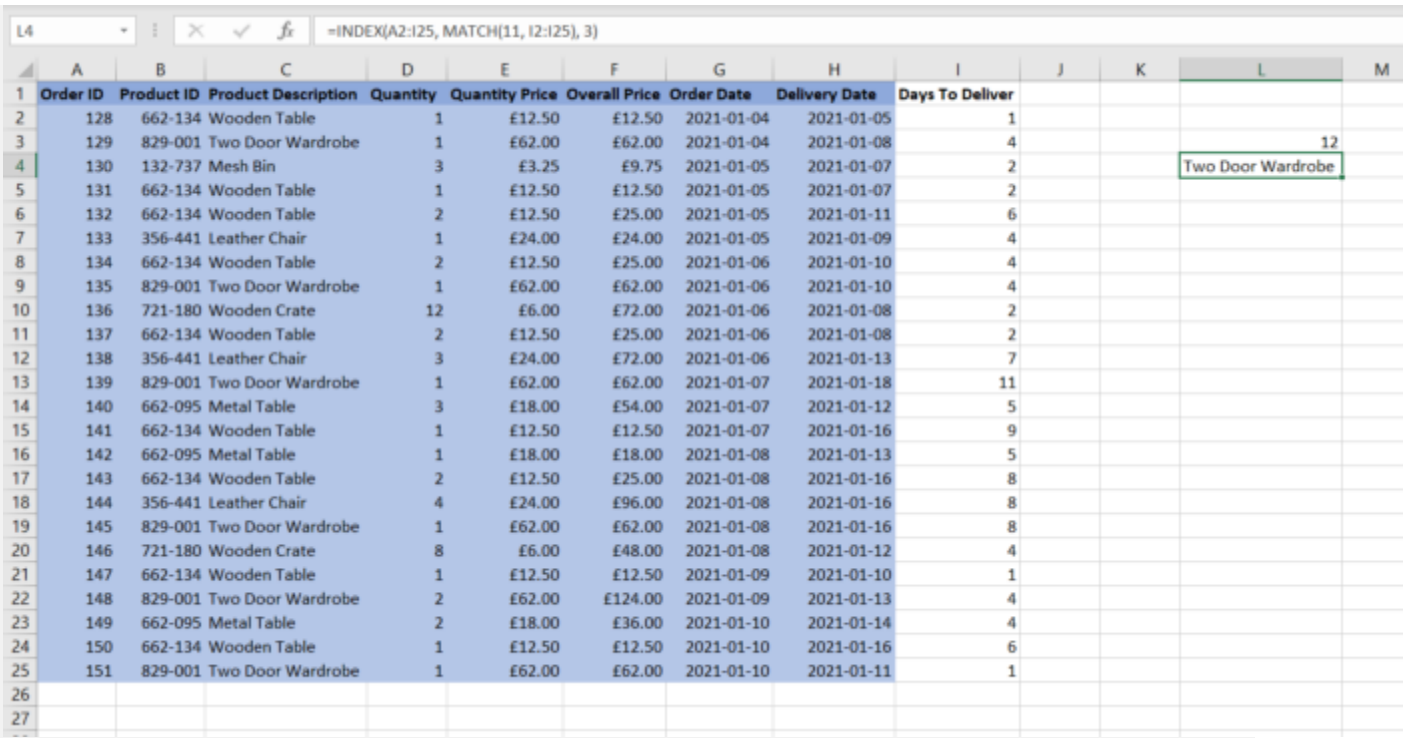
=MATCH(value, array)

INDEX

INDEX is a function that is commonly used with the previous MATCH function. Together, they can be used to get a value in a lookup table based upon given criteria, similar to the VLOOKUP function. However, these two functions can be used in some cases where the singular lookup functions fail, for example when the lookup value is not in the first row/column of the range.

On its own, the INDEX function is used to return the value of a cell at a given row number and index number.

In the example below, we can see that the MATCH function is used to get the row where the product is delivered in 11 days, and then the INDEX function is used to get the product description of that row.



L4 =INDEX(A2:I25, MATCH(11, I2:I25), 3)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Order ID	Product ID	Product Description	Quantity	Quantity Price	Overall Price	Order Date	Delivery Date	Days To Deliver				
2	128	662-134	Wooden Table	1	£12.50	£12.50	2021-01-04	2021-01-05	1				
3	129	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-04	2021-01-08	4			12	
4	130	132-737	Mesh Bin	3	£3.25	£9.75	2021-01-05	2021-01-07	2			Two Door Wardrobe	
5	131	662-134	Wooden Table	1	£12.50	£12.50	2021-01-05	2021-01-07	2				
6	132	662-134	Wooden Table	2	£12.50	£25.00	2021-01-05	2021-01-11	6				
7	133	356-441	Leather Chair	1	£24.00	£24.00	2021-01-05	2021-01-09	4				
8	134	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-10	4				
9	135	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-06	2021-01-10	4				
10	136	721-180	Wooden Crate	12	£6.00	£72.00	2021-01-06	2021-01-08	2				
11	137	662-134	Wooden Table	2	£12.50	£25.00	2021-01-06	2021-01-08	2				
12	138	356-441	Leather Chair	3	£24.00	£72.00	2021-01-06	2021-01-13	7				
13	139	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-07	2021-01-18	11				
14	140	662-095	Metal Table	3	£18.00	£54.00	2021-01-07	2021-01-12	5				
15	141	662-134	Wooden Table	1	£12.50	£12.50	2021-01-07	2021-01-16	9				
16	142	662-095	Metal Table	1	£18.00	£18.00	2021-01-08	2021-01-13	5				
17	143	662-134	Wooden Table	2	£12.50	£25.00	2021-01-08	2021-01-16	8				
18	144	356-441	Leather Chair	4	£24.00	£96.00	2021-01-08	2021-01-16	8				
19	145	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-08	2021-01-16	8				
20	146	721-180	Wooden Crate	8	£6.00	£48.00	2021-01-08	2021-01-12	4				
21	147	662-134	Wooden Table	1	£12.50	£12.50	2021-01-09	2021-01-10	1				
22	148	829-001	Two Door Wardrobe	2	£62.00	£124.00	2021-01-09	2021-01-13	4				
23	149	662-095	Metal Table	2	£18.00	£36.00	2021-01-10	2021-01-14	4				
24	150	662-134	Wooden Table	1	£12.50	£12.50	2021-01-10	2021-01-16	6				
25	151	829-001	Two Door Wardrobe	1	£62.00	£62.00	2021-01-10	2021-01-11	1				
26													
27													

Combining the MATCH and INDEX functions for a lookup. Image by author.
The formula for this Excel function is:
=INDEX(array, row_number, column_number)

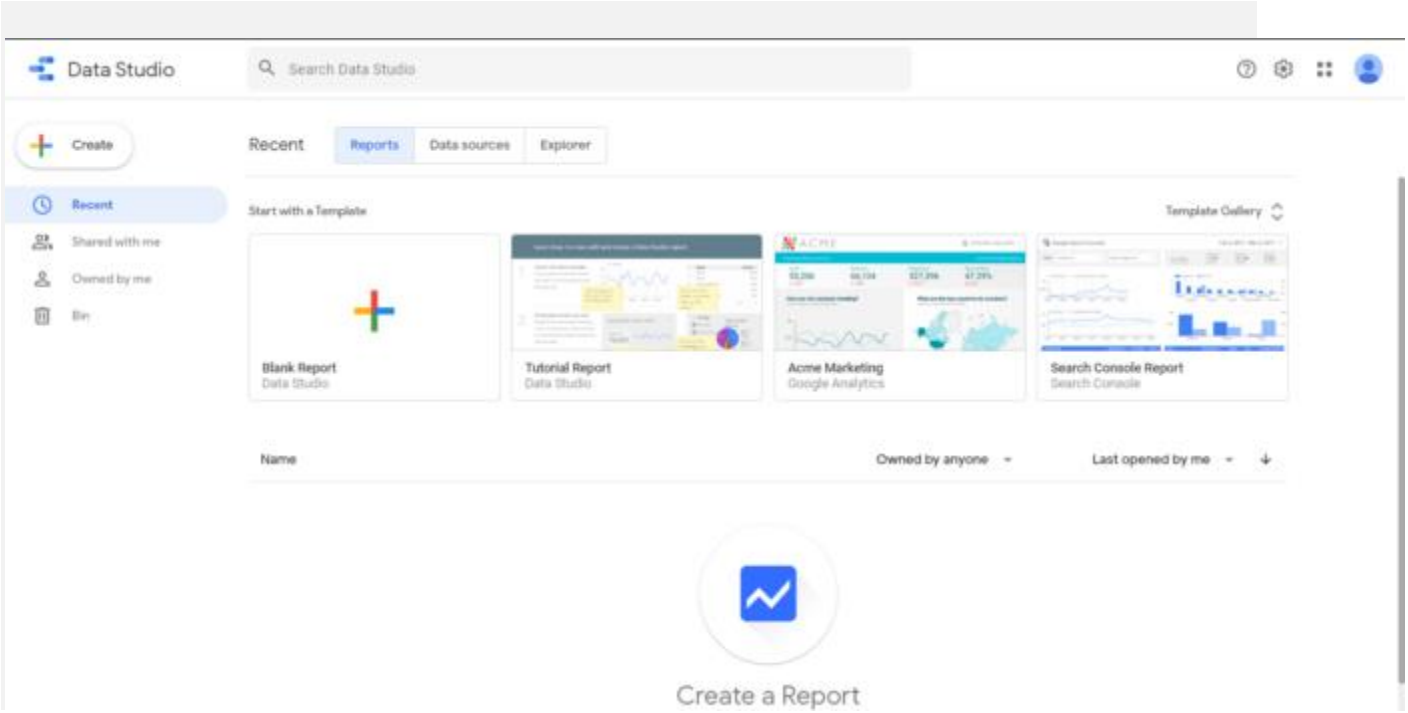
DATA SCIENCE WITH GOOGLE DATA STUDIO.

Google Data Studio **is One of the best data visualization and reporting tools is entirely free to use.**
When it comes to visualizing data, there are tons of platforms available. But the best ones are mostly paid, and the free ones are not up to the mark. Luckily, there is one platform that lets you overcome both these issues — Google Data Studio, which is entirely FREE and accessible to all
Apart from being a free-to-use tool, there are also many other benefits of using Data Studio, not limited to the following.

- It's an entirely web-based tool, so you don't have to install any bulky software on your computer. And again, everything is stored on the cloud, and you can access it from anywhere.
- Sharing reports is also really simple, using the same method that is followed to share documents from google drive.
- Connect to data from 150+ sources
- You can even display reports on your website using the embed code.

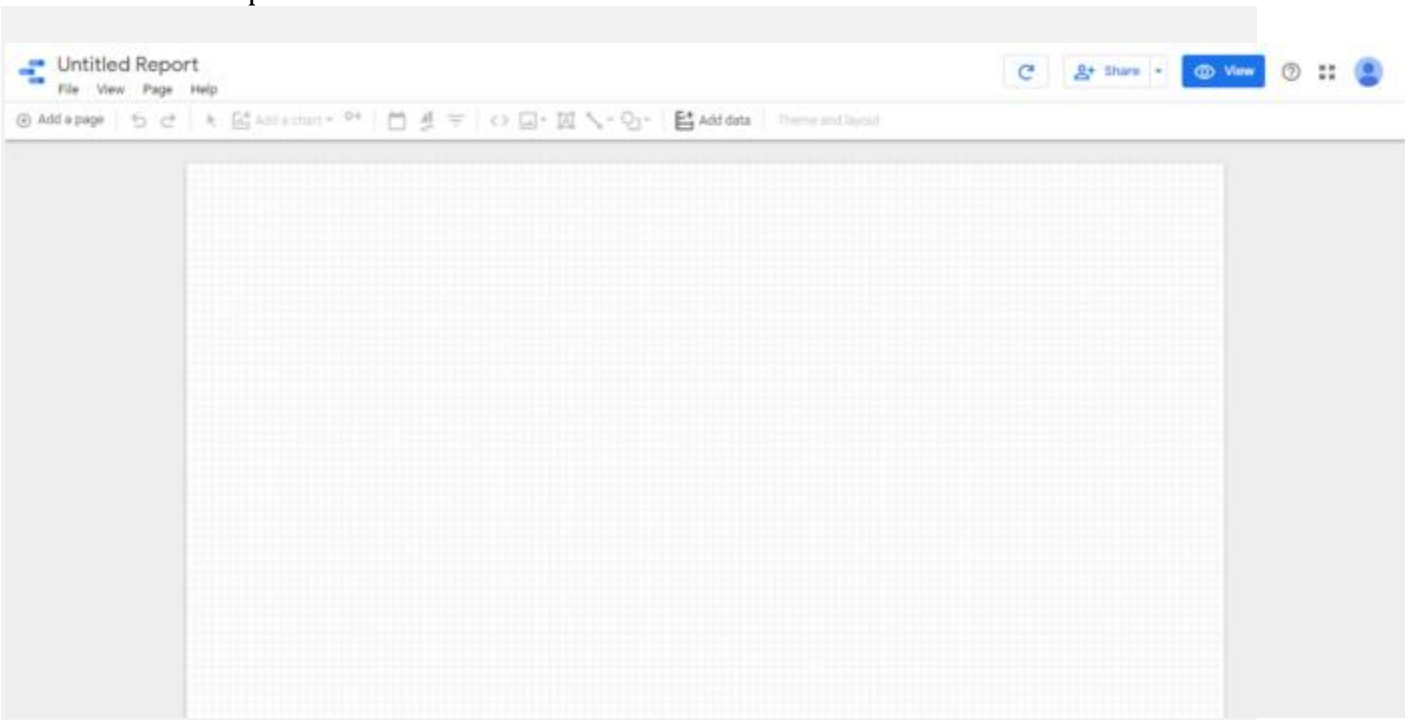
Log into Data Studio

To log in, you'll need a Google account, and once you log in, your home page will have the Data Studio dashboard, which looks like this.



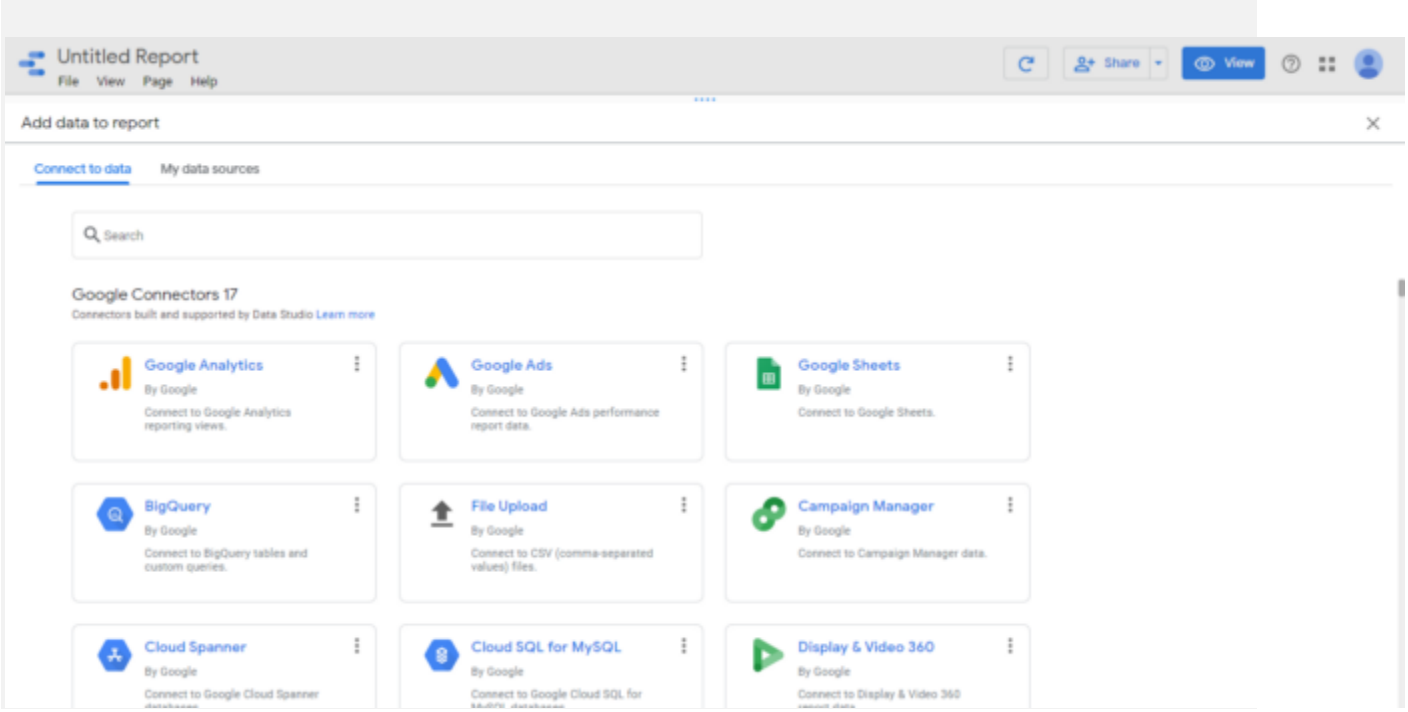
Create a Report

Select the "blank report" option. This is the place where you'll be arranging all your visuals and text for the report.



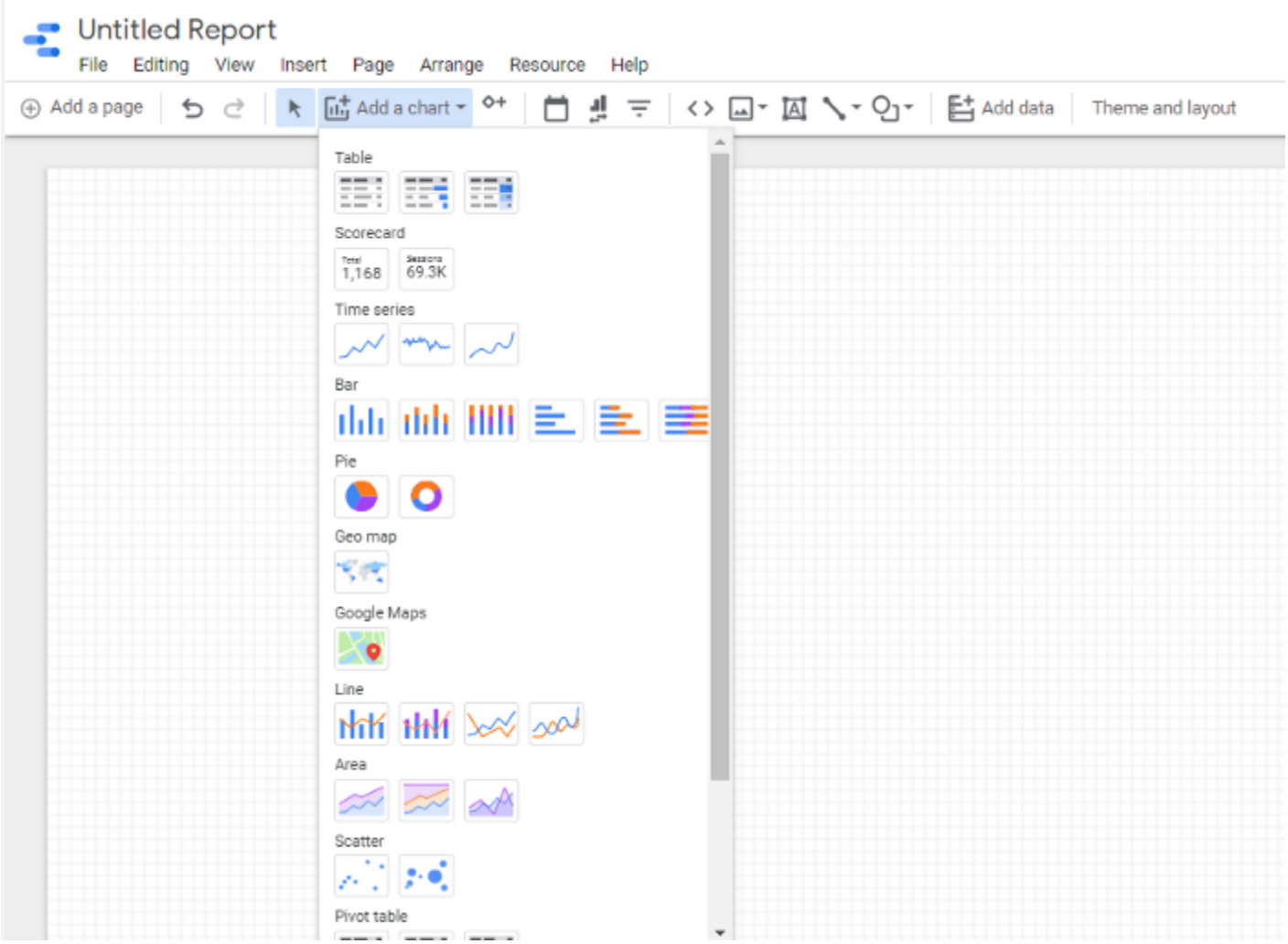
Connect with the data source

Select the "add data" option from the toolbar and connect with the sample Google Analytics data provided.

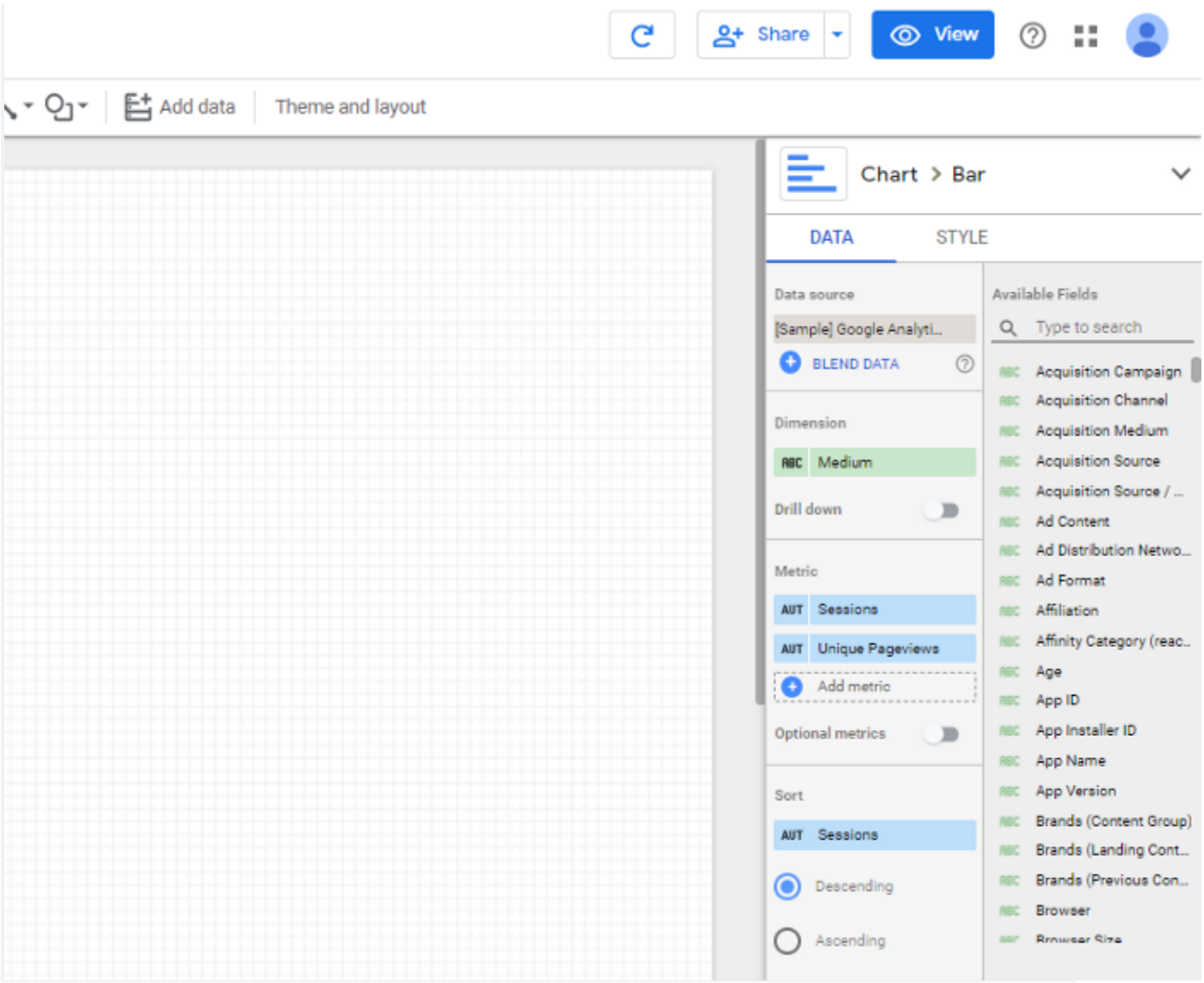


Add a visual

Click “add a chart” in the toolbar. A dropdown with all the chart types will appear. Select the chart that you want.

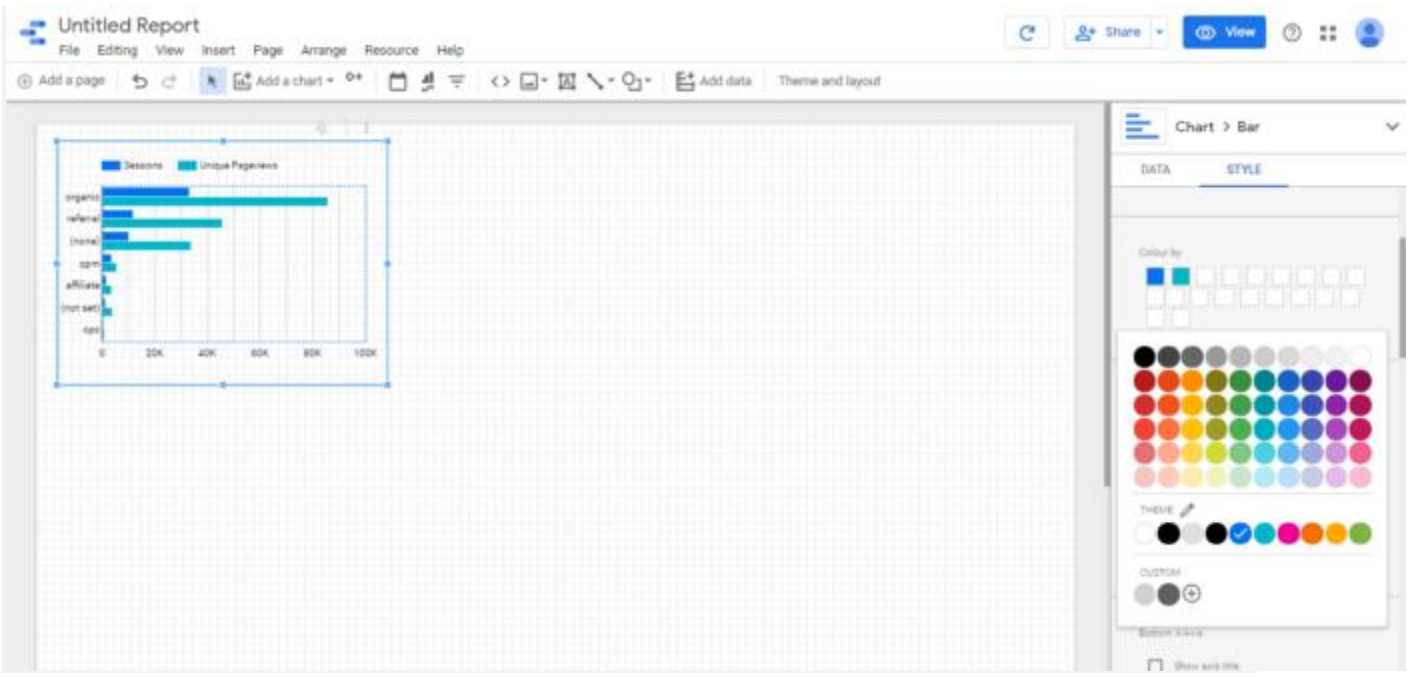


Once you select the chart type, the right-hand pane will appear, and here is what you will see. You can simply drag and drop the dimensions that you want from the "Available Fields," and your visual will be ready!



Customize your visual

You can customize the visual as per your requirement by selecting the "Style" option.



Adding text

Text and headings can be easily added using the "Textbox" option from the toolbar, and again you can customize and format the textbox as you need.

Downloading and sharing

Once your report is ready, you can share it using the "Share" option shown, or you can also download the report as PDF from the file menu.