**Trends**

In technical analysis, trends are identified by trend lines or price action that highlight when the price is making higher swing highs and higher swing lows for an uptrend, or lower swing lows and lower swing highs for a downtrend. The three basic types of trends are up, down, and sideways.

An uptrend is marked by an overall increase in price. Nothing moves straight up for long, so there will always be oscillations, but the overall direction needs to be higher.

A downtrend occurs when the price of an asset moves lower over a period of time. While the price may move intermittently higher or lower, downtrends are characterized by lower peaks and lower troughs over time.

Trends may be discovered in the short, medium, and long term. Generally, investors take positions in assets that will be profitable as long as the current trend continues. Taking positions that profit only if a trend reverses is riskier. Analysts use trendlines and channels, which are essentially boundaries for price fluctuations, in an attempt to spot and define trends. Upward trends are characterized by an asset price hitting a series of higher highs and higher lows, while downward trends are marked by lower highs and lower lows. Most traders trade in the direction of the trend. Traders who go opposite the trend are called contrarian investors.

**Alerts**

A Alert is a series of data that repeats in a recognizable way. It can be identified in the history of the asset being evaluated or other assets with similar characteristics. Alerts often include the study of sale volume, as well as price. Alerts can occur within a downward or upward trend, or they can mark the beginning of a new trend.

Alerts are the distinctive formations created by the movements of security prices on a chart. A pattern is identified by a line that connects common price points, such as closing prices or highs or lows, during a specific period of time. Chartists seek to identify Alerts as a way to anticipate the future direction of a security’s price.

There are bottoming, topping, and continuation Alerts. A "follow-through day" pattern is an example of a pattern used by some analysts to identify market bottoms. The "head-and-shoulders" topping pattern is popular among day and swing traders, while continuation Alerts include the "cup-and-handle," "flat base," and "three weeks tight."

"The trend is your friend" is a common catchphrase among technical analysts. A trend can often be found by establishing a line chart. A trend line is the line formed between a high and a low. If that line is going up, the trend is up. If the trend line is sloping downward, the trend is down. Trend lines are the foundation for most chart Alerts.

**Self Service Reporting**.

The relevance of data with the success of an enterprise goes hand-in-hand. What do we know about Self Service Reporting?

Well, the data is compelling. However, it is self-service reporting that offers all the access we require to the power. Comprehensive analysis is the prerequisite for IT leaders who face challenges with data accessibility that comes with the changing infrastructure management. With Self-Service Business Intelligence, you can profess about the future of your organization and gain access to thorough analysis, reports, and dashboards.

Business Intelligence (BI) solutions require individual users to be able to access reports and dashboards. The need to hire a data-scientist for every statistical move to be taken within an organization is increasingly becoming redundant. Providing users access to data to help them stay informed with the changing trends is what self service reporting is all about. It is irrelevant as to which domain or analytical background the user belongs.

Every company requires a self service reporting tool for the kind of benefits it offers. Not only to the end user, but for IT and the company itself.

### Benefits of Self Service Reporting tool

- It improves flexibility and agility within the business departments by providing user independence.
- IT departments are not overloaded in data requests or drowned in information. Their primary and straightforward tasks are removed or reduced, which reduces their overall workload.
- IT department can focus more on value-added tasks and help the organization with a higher degree of expertise from their department.
- Tackle business problems at the right time with self service BI's flexibility. The flexibility it offers to the users with the information is precious. With, ad hoc reporting, reports, and dashboards, once can take decisions time specifically.
- Efficiency is the key and the most promising attribute of self-service analytics tools; with it, there is minimal dependence on resources (external). It eliminates the rapid translation process and replaces it with business needs.
- Make quicker decisions based on data and compete with your competition in the forefront.
- IT is relieved of the tedious changing requests that come from various departments.
- Quick insights and data discovery are the primary mottoes of self-service business intelligence. It brings all the power to the users. Well, in the past enterprise space, traditional BI was rigid, slow, time-consuming and a burden for IT teams. That said, let's weigh in the pros and cons that are making self-service reporting emerge and unleash the power of data and analytics.

| Cons of Traditional BI | Pros of Self-Service Reporting |
|---|---|

| | |
|---|---|
| The technology is way past its prime; it is around 20 years old. Traditional BI is a design with architectural limitations, i.e., memory, hardware, CPU, etc. The age-old use of filters, disk space access, aggregation, etc. requires a lot of technologies to support the architecture. | Advanced technology for any operations, any tasks are performed on the go, user independence. Reports and dashboards replace any raw data existing and give practical and time-saving information. |
| It has only a pre-defined view on data, before the users view the data, it gets aggregated. | Decreases longer turnaround times and improves productivity with predictive analysis. |

With the dreadful technology dying with age, business intelligence (BI) solutions are paving the way to transitioning and replacing the traditional BI. Users require advanced analytics and search-based analytics. The liberation from conventional systems and access to easy-to-use tools helps the users to gain access to data and flexibility in performing operations.

### Can they two co-exist?

Many organizations use both the business intelligence tools, traditional and self-service BI. That is to say; it is a common practice for businesses to perform their daily business operations, for functional reporting with conventional BI.

Traditional BI is still great for a few organizations for dash boarding and compliance reporting. Organizations that have been using traditional BI for a long time still use it to seek answers about what happened back then.

With the changing dimensions, what users require is individual power. That being said, traditional BI is not designed to provide the users what it wants most. Which is the reason for self-service to be the preferred one?

Characteristics of  a good Report:

the reports have some objective and purpose behind it. That objective and purpose can only be achieved if a report has the following qualities and characteristics:

1. It should be factual: Every report should be based on facts, verified information and valid proofs.
2. Clear and Easily understandable: Explained below
3. Free from errors and duplication
4. Should facilitate the decision makers in making the right decision:
5. Result focused and result oriented
6. Well organized and structured
7. Ethical reporting style

**Reader-Friendly**

Readers are various stakeholders who receive reports generated by M&E. If reports are reader-friendly, they are likely to be read, remembered and acted upon. Following decisions need to be made by CSOs to make their reports reader-friendly:

- What do they need to know?
- When do they need to know?
- How do they like to know?

**Easy, Simple Language**

M&E reports are meant to inform not impress. Using easy, simple language, be it Urdu or English makes the report friendly on reader. To do this, here are some useful tips:

- Write only what is necessary
- Avoid repetition and redundancy
- Give interesting and relevant information
- Avoid preaching or lecturing
- Compose short and correct sentences

**Purposeful Presentation**

Each report has some objective(s) to meet. The “objective” comes from analyzing the needs of the reader. A CSO is working for a project that has several donors, and is channeled through an agency that needs to be informed about some specific things going on in the field. CSOs reports are the main pathways or channels of information to the people who decide to fund this and other such projects. Similarly, field reports are the admin vehicles for the management of the CSOs to make decision regarding the project itself. A good report presents facts and arguments in a manner that supports the purpose of the report.

**Organized and Well-Structured**

Each CSO comes up with a format of internal reporting to suit its requirements. Reporting to donors is done on their prescribed formats. The M&E system should be able to generate information that can be organized using different formats. In the annex, this manual provides some useful formats that can be customized by a CSO.

**Result-Focused**

In general, all readers are interested in the RESULTS. Therefore, one over-riding principle that CSOs should aim for in all report writing is to report on the results of their activities. This requires some analysis on their part that goes beyond a mere description of their activities. Result-focused means that description of activities is liked with the project objectives. This aspect must be addressed especially in the project progress reports. According to Phil Bartle, “A good progress report is not merely a descriptive activity report, but must analyze the results of those reported activities. The analysis should answer the question, "How far have the project objectives been reached?"

**Timely Prepared and Dispatched**

M&E generate “Information Products”, a customized set of information according to needs to a defined group of users. M&E’s information products are time-bound for both internal and external stakeholders. Reports, in suitable formats, need to be timely produced and made available to the readers. It is useful to develop an Information Product Matrix (IPM) like the one described below:

**Straightforward**

A good report is straight forward, honest description. It contains no lies, no deception, no fluff. It is neat, readable and to-the-point. It is well spaced, has titles and subtitles and is free of language errors.

**Features of Power BI Layout**

### Layouts Are Custom

Each layout is a totally custom setup. Each visual has been modified in a way to make the presentation of the data look the best. This means in some cases the visuals fonts, labels, label precision, x or y axis has been turned off or modified in some way.

Since the layout has been highly stylized, you may not recognize the visual or feel like something is missing, for example the y-axis. This was intentional and done to tell the data story.

You can modify all the properties by clicking on the visual and selecting the **Paint Roller** to modify the properties.

### Duplicating a Page

Once you have found a layout that you like, you might want to reuse that page layout multiple times. To reuse a layout, simply **Right Click** on the report layout tab that you like and select **Duplicate Page**.

Duplicate a Page

### Delete Unwanted Page

Most layout files will contain multiple report pages. Some layout documents have multiple layouts per PBIT file. Thus, there will be situations where you want to delete un-needed pages. This is a simple **Right Click** on the report tab that you wish to remove and selecting **Delete Page**.

The Layout files are created with many different report tabs, so you can have incredible flexibility. For example, you may want a main page with two (2) visuals, and slicers, but later in the report you need a three (3) column layout. By including many different designs within one PBIT Layout file you can choose what works best with your data.

### Visuals Layout in Selection Pane

In the selection window there will be multiple items which represent each item on the report page. Each item in the **Selection** pane will attempt to be laid out with applicable names and in order of left to right typically. There will be some exceptions to this rule, but items will be laid out as intuitive as possible.
*Note: If you don't see the selection window you can turn this on by clicking on the **View** ribbon, and then checking the box titled **Selection Pane**.*

### Report Pages Background

Each Layout Report page will come with a background image. If it is desired this background image can be removed. To remove the background image, make sure no visuals are selected. This can be achieved by pressing the **ESC** key. Then in the **Visualizations** pane expand the **Page Background**. To remove the background image, click the little **X** next to the image name, found under the Transparency setting. By default, the **Image Fit** is set to **Fit**.
*Caution: If a background image is removed, it can't be replaced. So, once it is gone, you won't be able to re-apply a background image.*

### Changing Visuals

Now, not all layouts will be able to meet your report needs. Maybe you want to add a custom visual, or maybe change a visual from a map to a bar chart. Even though this is a predefined layout you can easily change the visual. To change the visual first select the visual you want to change, and then on the **Visualizations** pane click a new visual type. It is that easy.

### Moving Visuals

All the visuals on each report page are locked by default and therefore you cannot move them with the mouse. This is a setting that has been turned on. The setting is called **Lock Objects** and it can be changed by navigating to the **View Ribbon** and then **Unchecking** the **Lock Objects** item. By unchecking this feature, you will be able to move the visuals around the page.

### Adding Custom Report Themes

he layouts have been made using the default colors for Power BI Desktop. However, once you have connected your data you can easily add your own color theme. If you need help making color themes using JSON you can use this **free Theme Generator** to help add custom colors to your report. For the full specification on the Reports Themes visit the **official Microsoft documentation page**..

ou may have seen various people use the terms “reporting” and “analysis” as though they were interchangeable terms or almost synonyms. While both of these areas of web analytics draw upon the same collected web data, reporting and analysis are very different in terms of their **purpose**, **tasks**, **outputs**, **delivery**, and **value**. Without a clear distinction of the differences, an organization may sell itself short in one area (typically analysis) and not achieve the full benefits of its web analytics investment. Although I’m primarily focusing on web analytics, companies can run into the same challenge with other analytics tools (e.g., ad serving, email, search, social, etc.).

Most companies have analytics solutions in place to derive greater value for their organizations. In other words, the ultimate goal for reporting and analysis is to increase sales and reduce costs (i.e., add value). Both reporting and analysis play roles in influencing and driving the actions which lead to greater value in organizations.

For the purposes of this blog post, I’m not going delve deeply into what happens before or after the reporting and analysis stages, but I do recognize that both areas are critical and challenging steps in the overall data-driven decision-making process. It’s important to remember that reporting and analysis only have the opportunity of being valuable **if they are acted upon**.

**Purpose**

Before covering the differing roles of reporting and analysis, let’s start with some high-level definitions of these two key areas of analytics.****

***Reporting:*** *The process of organizing data into informational summaries in order to monitor how different areas of a business are performing.*****

**Analysis:** *The process of exploring data and reports in order to extract meaningful insights, which can be used to better understand and improve business performance.*

__Reporting translates raw data into **information**. Analysis transforms data and information into **insights**. Reporting helps companies to monitor their online business and be alerted to when data falls outside of expected ranges. Good reporting should **raise questions** about the business from its end users. The goal of analysis is to **answer questions** by interpreting the data at a deeper level and providing actionable recommendations. Through the process of performing analysis you may raise additional questions, but the goal is to identify answers, or at least potential answers that can be tested. In summary, reporting shows you ***what is happening*** while analysis focuses on explaining ***why it is happening*** and ***what you can do about it.***

**Tasks**

Companies can sometimes confuse "analytics" with "analysis". For example, a firm may be focused on the general area of analytics (strategy, implementation, reporting, etc.) but not necessarily on the specific aspect of analysis. It's almost like some organizations run out of gas after the initial set-up-related activities and don't make it to the analysis stage. In addition, sometimes the lines between reporting and analysis can blur – what feels like analysis is really just another flavor of reporting.

One way to distinguish whether your organization is emphasizing reporting or analysis is by identifying the primary tasks that are being performed by your analytics team. If most of the team's time is spent on activities such as *building, configuring, consolidating, organizing, formatting,* and *summarizing* – that's reporting. Analysis focuses on different tasks such as *questioning, examining, interpreting, comparing*, and *confirming* (I've left out testing as I view optimization efforts as part of the action stage). Reporting and analysis tasks can be intertwined, but your analytics team should still evaluate where it is spending the majority of its time. In most cases, I've seen analytics teams spending most of their time on reporting tasks.

**Outputs**

When you look at reporting and analysis deliverables, on the surface they may look similar – lots of charts, graphs, trend lines, tables, stats, etc. However, there are some subtle differences. One of the main differences between reporting and analysis is the overall approach. Reporting follows a **push approach**, where reports are pushed to users who are then expected to extract meaningful insights and take appropriate actions for themselves (i.e., self-serve). I've identified three main types of reporting: *canned reports*, *dashboards*, and *alerts*.

1. **Canned reports:** These are the out-of-the-box and custom reports that you can access within the analytics tool or which can also be delivered on a recurring basis to a group of end users. Canned reports are fairly static with fixed metrics and dimensions. In general, some canned reports are more valuable than others, and a report's value may depend on how relevant it is to an individual's role (e.g., SEO specialist vs. web producer).
2. **Dashboards:** These custom-made reports combine different KPIs and reports to provide a comprehensive, high-level view of business performance for specific audiences. Dashboards may include data from various data sources and are also usually fairly static.
3. **Alerts:** These conditional reports are triggered when data falls outside of expected ranges or some other pre-defined criteria is met. Once people are notified of what happened, they can take appropriate action as necessary.

In contrast, analysis follows a **pull approach**, where particular data is pulled by an analyst in order to answer specific business questions. A basic, informal analysis can occur whenever someone simply performs some kind of mental assessment of a report and makes a decision to act or not act based on the data. In the case of analysis with actual deliverables, there are two main types: *ad hoc responses* and *analysis presentations*.

1. **Ad hoc responses:** Analysts receive requests to answer a variety of business questions, which may be spurred by questions raised by the reporting. Typically, these urgent requests are time sensitive and demand a quick turnaround. The analytics team may have to juggle multiple requests at the same time. As a result, the analyses cannot go as deep or wide as the analysts may like, and the deliverable is a short and concise report, which may or may not include any specific recommendations.
2. **Analysis presentations:** Some business questions are more complex in nature and require more time to perform a comprehensive, deep-dive analysis. These analysis projects result in a more formal deliverable, which includes two key sections: *key findings* and *recommendations*. The key findings section highlights the most meaningful and actionable insights gleaned from the analyses performed. The recommendations section provides guidance on what actions to take based on the analysis findings.

When you compare the two sets of reporting and analysis deliverables, the different purposes (information vs. insights) reveal the true colors of the outputs. Reporting pushes information to the organization, and analysis pulls insights from the reports and data. There may be other hybrid outputs such as annotated dashboards (analysis sprinkles on a reporting donut), which may appear to span the two areas. You should be able to determine whether a deliverable is primarily focused on reporting or analysis by its purpose (information/insights) and approach (push/pull).

Another key difference between reporting and analysis is **context**. Reporting provides **no or limited context** about what's happening in the data. In some cases, the end users already possess the necessary context to understand and interpret the data correctly. However, in other situations, the audience may not have the required background knowledge. Context is critical to good analysis. In order to tell a meaningful story with the data to drive specific actions, context becomes an essential component of the storyline.

Although they both leverage various forms of data visualization in their deliverables, analysis is different from reporting because it **emphasizes data points that are significant, unique, or special – and explain why they are important to the business**. Reporting may sometimes automatically highlight key changes in the data, but it's not going explain why these changes are (or aren't) important. Reporting isn't going to answer the "so what?" question on its own.

If you've ever had the pleasure of being a new parent, I would compare canned reporting, dashboards, and alerts to a six-month-old infant. It cries – often loudly – when something is wrong, but it can't tell you what is exactly wrong. The parent has to scramble to figure out what's going on (hungry, dirty diaper, no pacifier, teething, tired, ear infection, new Baby Einstein DVD, etc.). Continuing the parenting metaphor, reporting is also not going to tell you how to stop the crying.

The **recommendations** component is a key differentiator between analysis and reporting as it provides specific guidance on what actions to take based on the key insights found in the data. Even analysis outputs such as ad hoc responses may not drive action if they fail to include recommendations. Once a recommendation has been made, **follow-up** is another potent outcome of analysis because recommendations demand decisions to be made (go/no go/explore further). Decisions precede action. Action precedes value.

**Delivery**

As mentioned, reporting is more of a push model, where people can access reports through an analytics tool, Excel spreadsheet, widget, or have them scheduled for delivery into their mailbox, mobile device, FTP site, etc. Because of the demands of having to provide periodic reports (daily, weekly, monthly, etc.) to multiple individuals and groups, **automation** becomes a key focus in building and delivering reports. In other words, once the report is built, how can it be automated for regular delivery? Most of the analysts who I've talked to don't like manually building and refreshing reports on a regular basis. It's a job for robots or computers, not human beings who are still paying off their student loans for 4-6 years of higher education.

On the other hand, analysis is all about human beings using their superior reasoning and analytical skills to extract key insights from the data and form actionable recommendations for their organizations. Although analysis can be "submitted" to decision makers, it is more effectively presented **person-to-person**. In their book "Competing on Analytics", Thomas Davenport and Jeanne Harris emphasize the importance of **trust and credibility between the analyst and decision maker**. Decision makers typically don't have the time or ability to perform analyses themselves. With a "close, trusting relationship" in place, the executives will frame their needs correctly, the analysts will ask the right questions, and the executives will be more likely to take action on analysis they trust.

**Value**

When it comes to comparing the different roles of reporting and analysis, it's important to understand the relationship between reporting and analysis in driving value. I like to think of the data-driven stages (data > reporting > analysis > decision > action > value) as a series of dominoes. If you remove a domino, it can be more difficult or impossible to achieve the desired value.

In the "Path to Value" diagram above, it all starts with having the **right data that is complete and accurate**. It doesn't matter how advanced your reporting or analysis is if you don't have good, reliable data. If we skip the "reporting" domino, some seasoned analysts might argue that they don't need reports to do analysis (i.e., just give me the raw files and a database). On an individual basis that might be true for some people, but it doesn't work at the organizational level if you're striving to democratize your data.

Most companies have abundant reporting but may be missing the "analysis" domino. Reporting will rarely initiate action on its own as analysis is required to **help bridge the gap between data and action**. Having analysis doesn't guarantee that good decisions will be made, that people will actually act on the recommendations, that the business will take the right actions, or that teams will be able to execute effectively on those right actions. However, it is a necessary step closer to action and the potential value that can be realized through successful web analytics.

**Core Tools for Reporting and Dashboard layout Design:**

the most commonly cited among the websites that ranked the best BI tools. These five BI tools are not ranked in any particular order.

### 1. Microsoft Power BI

One of the most popular BI tools is Power BI, offered by leading software giant Microsoft. This tool is downloadable software, so you can choose to run analytics either on the cloud or in a reporting server. Syncing with sources such as Facebook, Oracle, and more, generate reports and dashboards in minutes with this interactive tool. It comes with built-in AI capabilities, Excel integration, and data connectors, and offers end-to-end data encryption and real-time access monitoring.

**Learn Power BI with Coursera:** In just two hours, you can learn the basics of Power BI Desktop with this guided project. You'll load and transform data to create interactive reports and dashboards.

### 2. Tableau

Tableau is known for its user-friendly data visualization capabilities, but it can do more than make pretty charts. Their offering includes live visual analytics, an interface that allows users to drag and drop buttons to spot trends in data quickly. The tool supports data sources such as Microsoft Excel, Box, PDF files, Google Analytics, and more. Its versatility extends to being able to connect with most databases.

**Learn Tableau with Coursera:** There are several options for learning how to use Tableau.

- Data Visualization with Tableau specialization from the University of California Davis
- Use Tableau for Your Data Science Workflow specialization from the University of California Irvine
- Data Visualization and Communication class with Tableau from Duke University
- Guided Project on Visualizing Citibike Trips with Tableau

### 3. QlikSense

QlikSense is a BI tool that emphasizes a self-service approach, meaning that it supports a wide range of analytics use cases, from guided apps and dashboards to custom and embedded analytics. It offers a user-friendly interface optimized for touchscreens, sophisticated AI, and high-performance cloud platforms. Its associative exploration capability, Search & Conversational Analytics, allows users to ask questions and uncover actionable insights, which helps increase data literacy for those new to using BI tools.

### 4. Dundas BI

Dundas BI is a browser-based BI tool that's been around for 25 years. Like Tableau, Dundas BI features a drag-and-drop function that allows users to analyze data on their own, without involving their IT team. The tool is known for its simplicity and flexibility through interactive dashboards, reports, and visual analytics. Since its inception as a data visualization tool in 1992, it has evolved into an end-to-end analytics platform that is able to compete with the new BI tools available today.

### 5. Sisense

Sisense is a user-friendly BI tool that focuses on being simplified and streamlined. With this tool, you can export data from sources like Google Analytics, Salesforce, and more. Its in-chip technology allows for faster data processing compared to other tools. Key features include the ability to embed white-label analytics, meaning a company can fully customize the services to its needs. Like others, it has a drag-and-drop feature. Sisense allows you to share reports and dashboards with your team members as well as externally.

**Other popular BI tools include:** Zoho Analytics, Oracle BI, SAS Visual Analytics, Domo, Datapine, Yellowfin BI, Looker, SAP Business Objects, Clear Analytics, Board, MicroStrategy, IBM Cognos Analytics, Tibco Spotfire, BIRT, Intercom, Google Data Studio, and HubSpot.

**Trends**

In technical analysis, trends are identified by trend lines or price action that highlight when the price is making higher swing highs and higher swing lows for an uptrend, or lower swing lows and lower swing highs for a downtrend. The three basic types of trends are up, down, and sideways.

An uptrend is marked by an overall increase in price. Nothing moves straight up for long, so there will always be oscillations, but the overall direction needs to be higher.

A downtrend occurs when the price of an asset moves lower over a period of time. While the price may move intermittently higher or lower, downtrends are characterized by lower peaks and lower troughs over time.

Trends may be discovered in the short, medium, and long term. Generally, investors take positions in assets that will be profitable as long as the current trend continues. Taking positions that profit only if a trend reverses is riskier. Analysts use trendlines and channels, which are essentially boundaries for price fluctuations, in an attempt to spot and define trends. Upward trends are characterized by an asset price hitting a series of higher highs and higher lows, while downward trends are marked by lower highs and lower lows. Most traders trade in the direction of the trend. Traders who go opposite the trend are called contrarian investors.

**Alerts**

A Alert is a series of data that repeats in a recognizable way. It can be identified in the history of the asset being evaluated or other assets with similar characteristics. Alerts often include the study of sale volume, as well as price. Alerts can occur within a downward or upward trend, or they can mark the beginning of a new trend.

Alerts are the distinctive formations created by the movements of security prices on a chart. A pattern is identified by a line that connects common price points, such as closing prices or highs or lows, during a specific period of time. Chartists seek to identify Alerts as a way to anticipate the future direction of a security's price.

There are bottoming, topping, and continuation Alerts. A "follow-through day" pattern is an example of a pattern used by some analysts to identify market bottoms. The "head-and-shoulders" topping pattern is popular among day and swing traders, while continuation Alerts include the "cup-and-handle," "flat base," and "three weeks tight."

"The trend is your friend" is a common catchphrase among technical analysts. A trend can often be found by establishing a line chart. A trend line is the line formed between a high and a low. If that line is going up, the trend is up. If the trend line is sloping downward, the trend is down. Trend lines are the foundation for most chart Alerts.

**Self Service Reporting**.

The relevance of data with the success of an enterprise goes hand-in-hand. What do we know about Self Service Reporting?

Well, the data is compelling. However, it is self-service reporting that offers all the access we require to the power. Comprehensive analysis is the prerequisite for IT leaders who face challenges with data accessibility that comes with the changing infrastructure management. With Self-Service Business Intelligence, you can profess about the future of your organization and gain access to thorough analysis, reports, and dashboards.

Business Intelligence (BI) solutions require individual users to be able to access reports and dashboards. The need to hire a data-scientist for every statistical move to be taken within an organization is increasingly becoming redundant. Providing users access to data to help them stay informed with the changing trends is what self service reporting is all about. It is irrelevant as to which domain or analytical background the user belongs.

Every company requires a self service reporting tool for the kind of benefits it offers. Not only to the end user, but for IT and the company itself.

### Benefits of Self Service Reporting tool

- It improves flexibility and agility within the business departments by providing user independence.
- IT departments are not overloaded in data requests or drowned in information. Their primary and straightforward tasks are removed or reduced, which reduces their overall workload.
- IT department can focus more on value-added tasks and help the organization with a higher degree of expertise from their department.
- Tackle business problems at the right time with self service BI's flexibility. The flexibility it offers to the users with the information is precious. With, ad hoc reporting, reports, and dashboards, once can take decisions time specifically.

- Efficiency is the key and the most promising attribute of self-service analytics tools; with it, there is minimal dependence on resources (external). It eliminates the rapid translation process and replaces it with business needs.
- Make quicker decisions based on data and compete with your competition in the forefront.
- IT is relieved of the tedious changing requests that come from various departments.
- Quick insights and data discovery are the primary mottoes of self-service business intelligence. It brings all the power to the users. Well, in the past enterprise space, traditional BI was rigid, slow, time-consuming and a burden for IT teams. That said, let's weigh in the pros and cons that are making self-service reporting emerge and unleash the power of data and analytics.

| Cons of Traditional BI | Pros of Self-Service Reporting |
|---|---|
| The technology is way past its prime; it is around 20 years old. Traditional BI is a design with architectural limitations, i.e., memory, hardware, CPU, etc. The age-old use of filters, disk space access, aggregation, etc. requires a lot of technologies to support the architecture. | Advanced technology for any operations, any tasks are performed on the go, user independence. Reports and dashboards replace any raw data existing and give practical and time-saving information. |
| It has only a pre-defined view on data, before the users view the data, it gets aggregated. | Decreases longer turnaround times and improves productivity with predictive analysis. |

With the dreadful technology dying with age, business intelligence (BI) solutions are paving the way to transitioning and replacing the traditional BI. Users require advanced analytics and search-based analytics. The liberation from conventional systems and access to easy-to-use tools helps the users to gain access to data and flexibility in performing operations.

### Can they two co-exist?

Many organizations use both the business intelligence tools, traditional and self-service BI. That is to say; it is a common practice for businesses to perform their daily business operations, for functional reporting with conventional BI.

Traditional BI is still great for a few organizations for dash boarding and compliance reporting. Organizations that have been using traditional BI for a long time still use it to seek answers about what happened back then.

With the changing dimensions, what users require is individual power. That being said, traditional BI is not designed to provide the users what it wants most. Which is the reason for self-service to be the preferred one?

Characteristics of a good Report:

the reports have some objective and purpose behind it. That objective and purpose can only be achieved if a report has the following qualities and characteristics:

1. It should be factual: Every report should be based on facts, verified information and valid proofs.
2. Clear and Easily understandable: Explained below
3. Free from errors and duplication
4. Should facilitate the decision makers in making the right decision:
5. Result focused and result oriented
6. Well organized and structured
7. Ethical reporting style

**Reader-Friendly**

Readers are various stakeholders who receive reports generated by M&E. If reports are reader-friendly, they are likely to be read, remembered and acted upon. Following decisions need to be made by CSOs to make their reports reader-friendly:

- What do they need to know?
- When do they need to know?
- How do they like to know?

**Easy, Simple Language**

M&E reports are meant to inform not impress. Using easy, simple language, be it Urdu or English makes the report friendly on reader. To do this, here are some useful tips:

- Write only what is necessary
- Avoid repetition and redundancy
- Give interesting and relevant information
- Avoid preaching or lecturing
- Compose short and correct sentences

**Purposeful Presentation**

Each report has some objective(s) to meet. The “objective” comes from analyzing the needs of the reader. A CSO is working for a project that has several donors, and is channeled through an agency that needs to be informed about some specific things going on in the field. CSOs reports are the main pathways or channels of information to the people who decide to fund this and other such projects. Similarly, field reports are the admin vehicles for the management of the CSOs to make decision regarding the project itself. A good report presents facts and arguments in a manner that supports the purpose of the report.

**Organized and Well-Structured**

Each CSO comes up with a format of internal reporting to suit its requirements. Reporting to donors is done on their prescribed formats. The M&E system should be able to generate information that can be organized using different formats. In the annex, this manual provides some useful formats that can be customized by a CSO.

**Result-Focused**

In general, all readers are interested in the RESULTS. Therefore, one over-riding principle that CSOs should aim for in all report writing is to report on the results of their activities. This requires some analysis on their part that goes beyond a mere description of their activities. Result-focused means that description of activities is liked with the project objectives. This aspect must be addressed especially in the project progress reports. According to Phil Bartle, “A good progress report is not merely a descriptive activity report, but must analyze the results of those reported activities. The analysis should answer the question, "How far have the project objectives been reached?"

**Timely Prepared and Dispatched**

M&E generate “Information Products”, a customized set of information according to needs to a defined group of users. M&E’s information products are time-bound for both internal and external stakeholders. Reports, in suitable formats, need to be timely produced and made available to the readers. It is useful to develop an Information Product Matrix (IPM) like the one described below:

**Straightforward**

A good report is straight forward, honest description. It contains no lies, no deception, no fluff. It is neat, readable and to-the-point. It is well spaced, has titles and subtitles and is free of language errors.

**Features of Power BI Layout**

### Layouts Are Custom

Each layout is a totally custom setup. Each visual has been modified in a way to make the presentation of the data look the best. This means in some cases the visuals fonts, labels, label precision, x or y axis has been turned off or modified in some way.

Since the layout has been highly stylized, you may not recognize the visual or feel like something is missing, for example the y-axis. This was intentional and done to tell the data story.

You can modify all the properties by clicking on the visual and selecting the **Paint Roller** to modify the properties.

### Duplicating a Page

Once you have found a layout that you like, you might want to reuse that page layout multiple times. To reuse a layout, simply **Right Click** on the report layout tab that you like and select **Duplicate Page**.

Duplicate a Page

### Delete Unwanted Page

Most layout files will contain multiple report pages. Some layout documents have multiple layouts per PBIT file. Thus, there will be situations where you want to delete un-needed pages. This is a simple **Right Click** on the report tab that you wish to remove and selecting **Delete Page**.

The Layout files are created with many different report tabs, so you can have incredible flexibility. For example, you may want a main page with two (2) visuals, and slicers, but later in the report you need a three (3) column layout. By including many different designs within one PBIT Layout file you can choose what works best with your data.

## Visuals Layout in Selection Pane

In the selection window there will be multiple items which represent each item on the report page.  Each item in the **Selection** pane will attempt to be laid out with applicable names and in order of left to right typically.  There will be some exceptions to this rule, but items will be laid out as intuitive as possible.

*Note: If you don't see the selection window you can turn this on by clicking on the **View** ribbon, and then checking the box titled **Selection Pane**.*

## Report Pages Background

Each Layout Report page will come with a background image.  If it is desired this background image can be removed.  To remove the background image, make sure no visuals are selected.  This can be achieved by pressing the **ESC** key.  Then in the **Visualizations** pane expand the **Page Background**.  To remove the background image, click the little **X** next to the image name, found under the Transparency setting.  By default, the **Image Fit** is set to **Fit**.

*Caution: If a background image is removed, it can't be replaced.  So, once it is gone, you won't be able to re-apply a background image.*

## Changing Visuals

Now, not all layouts will be able to meet your report needs.  Maybe you want to add a custom visual, or maybe change a visual from a map to a bar chart.  Even though this is a predefined layout you can easily change the visual.  To change the visual first select the visual you want to change, and then on the **Visualizations** pane click a new visual type.  It is that easy.

## Moving Visuals

All the visuals on each report page are locked by default and therefore you cannot move them with the mouse.  This is a setting that has been turned on.  The setting is called **Lock Objects** and it can be changed by navigating to the **View Ribbon** and then **Unchecking** the **Lock Objects** item.  By unchecking this feature, you will be able to move the visuals around the page.

## Adding Custom Report Themes

he layouts have been made using the default colors for Power BI Desktop.  However, once you have connected your data you can easily add your own color theme.  If you need help making color themes using JSON you can use this **free Theme Generator** to help add custom colors to your report.  For the full specification on the Reports Themes visit the **official Microsoft documentation page**..

ou may have seen various people use the terms "reporting" and "analysis" as though they were interchangeable terms or almost synonyms. While both of these areas of web analytics draw upon the same collected web data, reporting and analysis are very different in terms of their **purpose**, **tasks**, **outputs**, **delivery**, and **value**. Without a clear distinction of the differences, an organization may sell itself short in one area (typically analysis) and not achieve the full benefits of its web analytics investment. Although I'm primarily focusing on web analytics, companies can run into the same challenge with other analytics tools (e.g., ad serving, email, search, social, etc.).

Most companies have analytics solutions in place to derive greater value for their organizations. In other words, the ultimate goal for reporting and analysis is to increase sales and reduce costs (i.e., add value). Both reporting and analysis play roles in influencing and driving the actions which lead to greater value in organizations.

For the purposes of this blog post, I'm not going delve deeply into what happens before or after the reporting and analysis stages, but I do recognize that both areas are critical and challenging steps in the overall data-driven decision-making process. It's important to remember that reporting and analysis only have the opportunity of being valuable **if they are acted upon**.

**Purpose**

Before covering the differing roles of reporting and analysis, let's start with some high-level definitions of these two key areas of analytics.****

***Reporting:*** *The process of organizing data into informational summaries in order to monitor how different areas of a business are performing.*****

**Analysis:** *The process of exploring data and reports in order to extract meaningful insights, which can be used to better understand and improve business performance.*

__Reporting translates raw data into **information**. Analysis transforms data and information into **insights**. Reporting helps companies to monitor their online business and be alerted to when data falls outside of expected ranges. Good reporting should **raise questions** about the business from its end users. The goal of analysis is to **answer questions** by interpreting the data at a deeper level and providing actionable recommendations. Through the process of performing analysis you may raise additional questions, but the goal is to identify answers, or at least potential answers that can be tested. In summary, reporting shows you ***what is happening*** while analysis focuses on explaining ***why it is happening*** and ***what you can do about it.***

**Tasks**

Companies can sometimes confuse "analytics" with "analysis". For example, a firm may be focused on the general area of analytics (strategy, implementation, reporting, etc.) but not necessarily on the specific aspect of analysis. It's almost like some organizations run out of gas after the initial set-up-related activities and don't make it to the analysis stage. In addition, sometimes the lines between reporting and analysis can blur – what feels like analysis is really just another flavor of reporting.

One way to distinguish whether your organization is emphasizing reporting or analysis is by identifying the primary tasks that are being performed by your analytics team. If most of the team's time is spent on activities such as *building, configuring, consolidating, organizing, formatting,* and *summarizing* – that's reporting. Analysis focuses on different tasks such as *questioning, examining, interpreting, comparing*, and *confirming* (I've left out testing as I view optimization efforts as part of the action stage). Reporting and analysis tasks can be intertwined, but your analytics team should still evaluate where it is spending the majority of its time. In most cases, I've seen analytics teams spending most of their time on reporting tasks.

**Outputs**

When you look at reporting and analysis deliverables, on the surface they may look similar – lots of charts, graphs, trend lines, tables, stats, etc. However, there are some subtle differences. One of the main differences between reporting and analysis is the overall approach. Reporting follows a **push approach**, where reports are pushed to users who are then expected to extract meaningful insights and take appropriate actions for themselves (i.e., self-serve). I've identified three main types of reporting: *canned reports*, *dashboards*, and *alerts*.

1. **Canned reports:** These are the out-of-the-box and custom reports that you can access within the analytics tool or which can also be delivered on a recurring basis to a group of end users. Canned reports are fairly static with fixed metrics and dimensions. In general, some canned reports are more valuable than others, and a report's value may depend on how relevant it is to an individual's role (e.g., SEO specialist vs. web producer).
2. **Dashboards:** These custom-made reports combine different KPIs and reports to provide a comprehensive, high-level view of business performance for specific audiences. Dashboards may include data from various data sources and are also usually fairly static.
3. **Alerts:** These conditional reports are triggered when data falls outside of expected ranges or some other pre-defined criteria is met. Once people are notified of what happened, they can take appropriate action as necessary.

In contrast, analysis follows a **pull approach**, where particular data is pulled by an analyst in order to answer specific business questions. A basic, informal analysis can occur whenever someone simply performs some kind of mental assessment of a report and makes a decision to act or not act based on the data. In the case of analysis with actual deliverables, there are two main types: *ad hoc responses* and *analysis presentations*.

1. **Ad hoc responses:** Analysts receive requests to answer a variety of business questions, which may be spurred by questions raised by the reporting. Typically, these urgent requests are time sensitive and demand a quick turnaround. The analytics team may have to juggle multiple requests at the same time. As a result, the analyses cannot go as deep or wide as the analysts may like, and the deliverable is a short and concise report, which may or may not include any specific recommendations.
2. **Analysis presentations:** Some business questions are more complex in nature and require more time to perform a comprehensive, deep-dive analysis. These analysis projects result in a more formal deliverable, which includes two key sections: *key findings* and *recommendations*. The key findings section highlights the most meaningful and actionable insights gleaned from the analyses performed. The recommendations section provides guidance on what actions to take based on the analysis findings.

When you compare the two sets of reporting and analysis deliverables, the different purposes (information vs. insights) reveal the true colors of the outputs. Reporting pushes information to the organization, and analysis pulls insights from the reports and data. There may be other hybrid outputs such as annotated dashboards (analysis sprinkles on a reporting donut), which may appear to span the two areas. You should be able to determine whether a deliverable is primarily focused on reporting or analysis by its purpose (information/insights) and approach (push/pull).

Another key difference between reporting and analysis is **context**. Reporting provides **no or limited context** about what's happening in the data. In some cases, the end users already possess the necessary context to understand and interpret the data correctly. However, in other situations, the audience may not have the required background knowledge. Context is critical to good analysis. In order to tell a meaningful story with the data to drive specific actions, context becomes an essential component of the storyline.

Although they both leverage various forms of data visualization in their deliverables, analysis is different from reporting because it **emphasizes data points that are significant, unique, or special – and explain why they are important to the business**. Reporting may sometimes automatically highlight key changes in the data, but it's not going explain why these changes are (or aren't) important. Reporting isn't going to answer the "so what?" question on its own.

If you've ever had the pleasure of being a new parent, I would compare canned reporting, dashboards, and alerts to a six-month-old infant. It cries – often loudly – when something is wrong, but it can't tell you what is exactly wrong. The parent has to scramble to figure out what's going on (hungry, dirty diaper, no pacifier, teething, tired, ear infection, new Baby Einstein DVD, etc.). Continuing the parenting metaphor, reporting is also not going to tell you how to stop the crying.

The **recommendations** component is a key differentiator between analysis and reporting as it provides specific guidance on what actions to take based on the key insights found in the data. Even analysis outputs such as ad hoc responses may not drive action if they fail to include recommendations. Once a recommendation has been made, **follow-up** is another potent outcome of analysis because recommendations demand decisions to be made (go/no go/explore further). Decisions precede action. Action precedes value.

**Delivery**

As mentioned, reporting is more of a push model, where people can access reports through an analytics tool, Excel spreadsheet, widget, or have them scheduled for delivery into their mailbox, mobile device, FTP site, etc. Because of the demands of having to provide periodic reports (daily, weekly, monthly, etc.) to multiple individuals and groups, **automation** becomes a key focus in building and delivering reports. In other words, once the report is built, how can it be automated for regular delivery? Most of the analysts who I've talked to don't like manually building and refreshing reports on a regular basis. It's a job for robots or computers, not human beings who are still paying off their student loans for 4-6 years of higher education.

On the other hand, analysis is all about human beings using their superior reasoning and analytical skills to extract key insights from the data and form actionable recommendations for their organizations. Although analysis can be "submitted" to decision makers, it is more effectively presented **person-to-person**. In their book "Competing on Analytics", Thomas Davenport and Jeanne Harris emphasize the importance of **trust and credibility between the analyst and decision maker**. Decision makers typically don't have the time or ability to perform analyses themselves. With a "close, trusting relationship" in place, the executives will frame their needs correctly, the analysts will ask the right questions, and the executives will be more likely to take action on analysis they trust.

**Value**

When it comes to comparing the different roles of reporting and analysis, it's important to understand the relationship between reporting and analysis in driving value. I like to think of the data-driven stages (data > reporting > analysis > decision > action > value) as a series of dominoes. If you remove a domino, it can be more difficult or impossible to achieve the desired value.

In the "Path to Value" diagram above, it all starts with having the **right data that is complete and accurate**. It doesn't matter how advanced your reporting or analysis is if you don't have good, reliable data. If we skip the "reporting" domino, some seasoned analysts might argue that they don't need reports to do analysis (i.e., just give me the raw files and a database). On an individual basis that might be true for some people, but it doesn't work at the organizational level if you're striving to democratize your data.

Most companies have abundant reporting but may be missing the "analysis" domino. Reporting will rarely initiate action on its own as analysis is required to **help bridge the gap between data and action**. Having analysis doesn't guarantee that good decisions will be made, that people will actually act on the recommendations, that the business will take the right actions, or that teams will be able to execute effectively on those right actions. However, it is a necessary step closer to action and the potential value that can be realized through successful web analytics.

**Core Tools for Reporting and Dashboard layout Design:**

the most commonly cited among the websites that ranked the best BI tools. These five BI tools are not ranked in any particular order.

### 1. Microsoft Power BI

One of the most popular BI tools is Power BI, offered by leading software giant Microsoft. This tool is downloadable software, so you can choose to run analytics either on the cloud or in a reporting server. Syncing with sources such as Facebook, Oracle, and more, generate reports and dashboards in minutes with this interactive tool. It comes with built-in AI capabilities, Excel integration, and data connectors, and offers end-to-end data encryption and real-time access monitoring.

**Learn Power BI with Coursera:** In just two hours, you can learn the basics of Power BI Desktop with this guided project. You'll load and transform data to create interactive reports and dashboards.

### 2. Tableau

Tableau is known for its user-friendly data visualization capabilities, but it can do more than make pretty charts. Their offering includes live visual analytics, an interface that allows users to drag and drop buttons to spot trends in data quickly. The tool supports data sources such as Microsoft Excel, Box, PDF files, Google Analytics, and more. Its versatility extends to being able to connect with most databases.

**Learn Tableau with Coursera:** There are several options for learning how to use Tableau.

- Data Visualization with Tableau specialization from the University of California Davis
- Use Tableau for Your Data Science Workflow specialization from the University of California Irvine
- Data Visualization and Communication class with Tableau from Duke University

- Guided Project on Visualizing Citibike Trips with Tableau

### 3. QlikSense

QlikSense is a BI tool that emphasizes a self-service approach, meaning that it supports a wide range of analytics use cases, from guided apps and dashboards to custom and embedded analytics. It offers a user-friendly interface optimized for touchscreens, sophisticated AI, and high-performance cloud platforms. Its associative exploration capability, Search & Conversational Analytics, allows users to ask questions and uncover actionable insights, which helps increase data literacy for those new to using BI tools.

### 4. Dundas BI

Dundas BI is a browser-based BI tool that's been around for 25 years. Like Tableau, Dundas BI features a drag-and-drop function that allows users to analyze data on their own, without involving their IT team. The tool is known for its simplicity and flexibility through interactive dashboards, reports, and visual analytics. Since its inception as a data visualization tool in 1992, it has evolved into an end-to-end analytics platform that is able to compete with the new BI tools available today.

### 5. Sisense

Sisense is a user-friendly BI tool that focuses on being simplified and streamlined. With this tool, you can export data from sources like Google Analytics, Salesforce, and more. Its in-chip technology allows for faster data processing compared to other tools. Key features include the ability to embed white-label analytics, meaning a company can fully customize the services to its needs. Like others, it has a drag-and-drop feature. Sisense allows you to share reports and dashboards with your team members as well as externally.

**Other popular BI tools include:** Zoho Analytics, Oracle BI, SAS Visual Analytics, Domo, Datapine, Yellowfin BI, Looker, SAP Business Objects, Clear Analytics, Board, MicroStrategy, IBM Cognos Analytics, Tibco Spotfire, BIRT, Intercom, Google Data Studio, and HubSpot.

## Trends

In technical analysis, trends are identified by trend lines or price action that highlight when the price is making higher swing highs and higher swing lows for an uptrend, or lower swing lows and lower swing highs for a downtrend. The three basic types of trends are up, down, and sideways.

An uptrend is marked by an overall increase in price. Nothing moves straight up for long, so there will always be oscillations, but the overall direction needs to be higher.

A downtrend occurs when the price of an asset moves lower over a period of time. While the price may move intermittently higher or lower, downtrends are characterized by lower peaks and lower troughs over time.

Trends may be discovered in the short, medium, and long term. Generally, investors take positions in assets that will be profitable as long as the current trend continues. Taking positions that profit only if a trend reverses is riskier. Analysts use trendlines and channels, which are essentially boundaries for price fluctuations, in an attempt to spot and define trends. Upward trends are characterized by an asset price hitting a series of higher highs and higher lows, while downward trends are marked by lower highs and lower lows. Most traders trade in the direction of the trend. Traders who go opposite the trend are called contrarian investors.

**Alerts**

A Alert is a series of data that repeats in a recognizable way. It can be identified in the history of the asset being evaluated or other assets with similar characteristics. Alerts often include the study of sale volume, as well as price. Alerts can occur within a downward or upward trend, or they can mark the beginning of a new trend.

Alerts are the distinctive formations created by the movements of security prices on a chart. A pattern is identified by a line that connects common price points, such as closing prices or highs or lows, during a specific period of time. Chartists seek to identify Alerts as a way to anticipate the future direction of a security’s price.

There are bottoming, topping, and continuation Alerts. A "follow-through day" pattern is an example of a pattern used by some analysts to identify market bottoms. The "head-and-shoulders" topping pattern is popular among day and swing traders, while continuation Alerts include the "cup-and-handle," "flat base," and "three weeks tight."

"The trend is your friend" is a common catchphrase among technical analysts. A trend can often be found by establishing a line chart. A trend line is the line formed between a high and a low. If that line is going up, the trend is up. If the trend line is sloping downward, the trend is down. Trend lines are the foundation for most chart Alerts.

**Self Service Reporting**.

The relevance of data with the success of an enterprise goes hand-in-hand. What do we know about Self Service Reporting?

Well, the data is compelling. However, it is self-service reporting that offers all the access we require to the power. Comprehensive analysis is the prerequisite for IT leaders who face challenges with data accessibility that comes with the changing infrastructure management. With Self-Service Business Intelligence, you can profess about the future of your organization and gain access to thorough analysis, reports, and dashboards.

Business Intelligence (BI) solutions require individual users to be able to access reports and dashboards. The need to hire a data-scientist for every statistical move to be taken within an organization is increasingly becoming redundant. Providing users access to data to help them stay informed with the changing trends is what self service reporting is all about. It is irrelevant as to which domain or analytical background the user belongs.

Every company requires a self service reporting tool for the kind of benefits it offers. Not only to the end user, but for IT and the company itself.

### Benefits of Self Service Reporting tool

- It improves flexibility and agility within the business departments by providing user independence.
- IT departments are not overloaded in data requests or drowned in information. Their primary and straightforward tasks are removed or reduced, which reduces their overall workload.
- IT department can focus more on value-added tasks and help the organization with a higher degree of expertise from their department.
- Tackle business problems at the right time with self service BI's flexibility. The flexibility it offers to the users with the information is precious. With, ad hoc reporting, reports, and dashboards, once can take decisions time specifically.
- Efficiency is the key and the most promising attribute of self-service analytics tools; with it, there is minimal dependence on resources (external). It eliminates the rapid translation process and replaces it with business needs.
- Make quicker decisions based on data and compete with your competition in the forefront.
- IT is relieved of the tedious changing requests that come from various departments.
- Quick insights and data discovery are the primary mottoes of self-service business intelligence. It brings all the power to the users. Well, in the past enterprise space, traditional BI was rigid, slow, time-consuming and a burden for IT teams. That said, let's weigh in the pros and cons that are making self-service reporting emerge and unleash the power of data and analytics.

| **Cons of Traditional BI** | **Pros of Self-Service Reporting** |
|---|---|
| The technology is way past its prime; it is around 20 years old. Traditional BI is a design with architectural limitations, i.e., memory, hardware, CPU, etc. The age-old use of filters, disk space | Advanced technology for any operations, any tasks are performed on the go, user independence. Reports and dashboards replace any raw data existing and give |

| | |
|---|---|
| access, aggregation, etc. requires a lot of technologies to support the architecture. | practical and time-saving information. |
| It has only a pre-defined view on data, before the users view the data, it gets aggregated. | Decreases longer turnaround times and improves productivity with predictive analysis. |

With the dreadful technology dying with age, business intelligence (BI) solutions are paving the way to transitioning and replacing the traditional BI. Users require advanced analytics and search-based analytics. The liberation from conventional systems and access to easy-to-use tools helps the users to gain access to data and flexibility in performing operations.

### Can they two co-exist?

Many organizations use both the business intelligence tools, traditional and self-service BI. That is to say; it is a common practice for businesses to perform their daily business operations, for functional reporting with conventional BI.

Traditional BI is still great for a few organizations for dash boarding and compliance reporting. Organizations that have been using traditional BI for a long time still use it to seek answers about what happened back then.

With the changing dimensions, what users require is individual power. That being said, traditional BI is not designed to provide the users what it wants most. Which is the reason for self-service to be the preferred one?

Characteristics of a good Report:

the reports have some objective and purpose behind it. That objective and purpose can only be achieved if a report has the following qualities and characteristics:

1. It should be factual: Every report should be based on facts, verified information and valid proofs.
2. Clear and Easily understandable: Explained below
3. Free from errors and duplication
4. Should facilitate the decision makers in making the right decision:
5. Result focused and result oriented
6. Well organized and structured

7. Ethical reporting style

**Reader-Friendly**

Readers are various stakeholders who receive reports generated by M&E. If reports are reader-friendly, they are likely to be read, remembered and acted upon. Following decisions need to be made by CSOs to make their reports reader-friendly:

- What do they need to know?
- When do they need to know?
- How do they like to know?

**Easy, Simple Language**

M&E reports are meant to inform not impress. Using easy, simple language, be it Urdu or English makes the report friendly on reader. To do this, here are some useful tips:

- Write only what is necessary
- Avoid repetition and redundancy
- Give interesting and relevant information
- Avoid preaching or lecturing
- Compose short and correct sentences

**Purposeful Presentation**

Each report has some objective(s) to meet. The "objective" comes from analyzing the needs of the reader. A CSO is working for a project that has several donors, and is channeled through an agency that needs to be informed about some specific things going on in the field. CSOs reports are the main pathways or channels of information to the people who decide to fund this and other such projects. Similarly, field reports are the admin vehicles for the management of the CSOs to make decision regarding the project itself. A good report presents facts and arguments in a manner that supports the purpose of the report.

**Organized and Well-Structured**

Each CSO comes up with a format of internal reporting to suit its requirements. Reporting to donors is done on their prescribed formats. The M&E system should be able to generate information that can be organized using different formats. In the annex, this manual provides some useful formats that can be customized by a CSO.

**Result-Focused**

In general, all readers are interested in the RESULTS. Therefore, one over-riding principle that CSOs should aim for in all report writing is to report on the results of their activities. This requires some analysis on their part that goes beyond a mere description of their activities. Result-focused means that description of activities is liked with the project objectives. This aspect must be addressed especially in the project progress reports. According to Phil Bartle, “A good progress report is not merely a descriptive activity report, but must analyze the results of those reported activities. The analysis should answer the question, "How far have the project objectives been reached?"

**Timely Prepared and Dispatched**

M&E generate “Information Products”, a customized set of information according to needs to a defined group of users. M&E’s information products are time-bound for both internal and external stakeholders. Reports, in suitable formats, need to be timely produced and made available to the readers. It is useful to develop an Information Product Matrix (IPM) like the one described below:

**Straightforward**

A good report is straight forward, honest description. It contains no lies, no deception, no fluff. It is neat, readable and to-the-point. It is well spaced, has titles and subtitles and is free of language errors.

**Features of Power BI Layout**

### Layouts Are Custom

Each layout is a totally custom setup. Each visual has been modified in a way to make the presentation of the data look the best. This means in some cases the visuals fonts, labels, label precision, x or y axis has been turned off or modified in some way.

Since the layout has been highly stylized, you may not recognize the visual or feel like something is missing, for example the y-axis. This was intentional and done to tell the data story.

You can modify all the properties by clicking on the visual and selecting the **Paint Roller** to modify the properties.

### Duplicating a Page

Once you have found a layout that you like, you might want to reuse that page layout multiple times. To reuse a layout, simply **Right Click** on the report layout tab that you like and select **Duplicate Page**.

Duplicate a Page

### Delete Unwanted Page

Most layout files will contain multiple report pages. Some layout documents have multiple layouts per PBIT file. Thus, there will be situations where you want to delete un-needed pages. This is a simple **Right Click** on the report tab that you wish to remove and selecting **Delete Page**.

The Layout files are created with many different report tabs, so you can have incredible flexibility. For example, you may want a main page with two (2) visuals, and slicers, but later in the report you need a three (3) column layout. By including many different designs within one PBIT Layout file you can choose what works best with your data.

### Visuals Layout in Selection Pane

In the selection window there will be multiple items which represent each item on the report page. Each item in the **Selection** pane will attempt to be laid out with applicable names and in order of left to right typically. There will be some exceptions to this rule, but items will be laid out as intuitive as possible.
*Note: If you don't see the selection window you can turn this on by clicking on the **View** ribbon, and then checking the box titled **Selection Pane**.*

### Report Pages Background

Each Layout Report page will come with a background image. If it is desired this background image can be removed. To remove the background image, make sure no visuals are selected. This can be achieved by pressing the **ESC** key. Then in the **Visualizations** pane expand the **Page Background**. To remove the background image, click the little **X** next to the image name, found under the Transparency setting. By default, the **Image Fit** is set to **Fit**.
*Caution: If a background image is removed, it can't be replaced. So, once it is gone, you won't be able to re-apply a background image.*

### Changing Visuals

Now, not all layouts will be able to meet your report needs. Maybe you want to add a custom visual, or maybe change a visual from a map to a bar chart. Even though this is a predefined layout you can easily change the visual. To change the visual first select the visual you want to change, and then on the **Visualizations** pane click a new visual type. It is that easy.

### Moving Visuals

All the visuals on each report page are locked by default and therefore you cannot move them with the mouse. This is a setting that has been turned on. The setting is called **Lock Objects** and it can be changed by navigating to the **View Ribbon** and then **Unchecking** the **Lock Objects** item. By unchecking this feature, you will be able to move the visuals around the page.

### Adding Custom Report Themes

he layouts have been made using the default colors for Power BI Desktop. However, once you have connected your data you can easily add your own color theme. If you need help making color themes using JSON you can use this **free Theme Generator** to help add custom colors to your report. For the full specification on the Reports Themes visit the **official Microsoft documentation page**..

ou may have seen various people use the terms "reporting" and "analysis" as though they were interchangeable terms or almost synonyms. While both of these areas of web analytics draw upon the same collected web data, reporting and analysis are very different in terms of their **purpose**, **tasks**, **outputs**, **delivery**, and **value**. Without a clear distinction of the differences, an organization may sell itself short in one area (typically analysis) and not achieve the full benefits of its web analytics investment. Although I'm primarily focusing on web analytics, companies can run into the same challenge with other analytics tools (e.g., ad serving, email, search, social, etc.).

Most companies have analytics solutions in place to derive greater value for their organizations. In other words, the ultimate goal for reporting and analysis is to increase sales and reduce costs (i.e., add value). Both reporting and analysis play roles in influencing and driving the actions which lead to greater value in organizations.

For the purposes of this blog post, I'm not going delve deeply into what happens before or after the reporting and analysis stages, but I do recognize that both areas are critical and challenging steps in the overall data-driven decision-making process. It's important to remember that reporting and analysis only have the opportunity of being valuable **if they are acted upon**.

**Purpose**

Before covering the differing roles of reporting and analysis, let's start with some high-level definitions of these two key areas of analytics.****

***Reporting:*** *The process of organizing data into informational summaries in order to monitor how different areas of a business are performing.*****

**Analysis:** *The process of exploring data and reports in order to extract meaningful insights, which can be used to better understand and improve business performance.*

__Reporting translates raw data into **information**. Analysis transforms data and information into **insights**. Reporting helps companies to monitor their online business and be alerted to when data falls outside of expected ranges. Good reporting should **raise questions** about the business from its end users. The goal of analysis is to **answer questions** by interpreting the data at a deeper level and providing actionable recommendations. Through the process of performing analysis you may raise additional questions, but the goal is to identify answers, or at least potential answers that can be tested. In summary, reporting shows you ***what is happening*** while analysis focuses on explaining ***why it is happening*** and ***what you can do about it.***

**Tasks**

Companies can sometimes confuse "analytics" with "analysis". For example, a firm may be focused on the general area of analytics (strategy, implementation, reporting, etc.) but not necessarily on the specific aspect of analysis. It's almost like some organizations run out of gas after the initial set-up-related activities and don't make it to the analysis stage. In addition, sometimes the lines between reporting and analysis can blur – what feels like analysis is really just another flavor of reporting.

One way to distinguish whether your organization is emphasizing reporting or analysis is by identifying the primary tasks that are being performed by your analytics team. If most of the team's time is spent on activities such as *building, configuring, consolidating, organizing, formatting,* and *summarizing* – that's reporting. Analysis focuses on different tasks such as *questioning, examining, interpreting, comparing*, and *confirming* (I've left out testing as I view optimization efforts as part of the action stage). Reporting and analysis tasks can be intertwined, but your analytics team should still evaluate where it is spending the majority of its time. In most cases, I've seen analytics teams spending most of their time on reporting tasks.

**Outputs**

When you look at reporting and analysis deliverables, on the surface they may look similar – lots of charts, graphs, trend lines, tables, stats, etc. However, there are some subtle differences. One of the main differences between reporting and analysis is the overall approach. Reporting follows a **push approach**, where reports are pushed to users who are then expected to extract meaningful insights and take appropriate actions for themselves (i.e., self-serve). I've identified three main types of reporting: *canned reports*, *dashboards*, and *alerts*.

1. **Canned reports:** These are the out-of-the-box and custom reports that you can access within the analytics tool or which can also be delivered on a recurring basis to a group of end users. Canned reports are fairly static with fixed metrics and dimensions. In general, some canned reports are more valuable than others, and a report's value may depend on how relevant it is to an individual's role (e.g., SEO specialist vs. web producer).
2. **Dashboards:** These custom-made reports combine different KPIs and reports to provide a comprehensive, high-level view of business performance for specific audiences. Dashboards may include data from various data sources and are also usually fairly static.

3. **Alerts:** These conditional reports are triggered when data falls outside of expected ranges or some other pre-defined criteria is met. Once people are notified of what happened, they can take appropriate action as necessary.

In contrast, analysis follows a **pull approach**, where particular data is pulled by an analyst in order to answer specific business questions. A basic, informal analysis can occur whenever someone simply performs some kind of mental assessment of a report and makes a decision to act or not act based on the data. In the case of analysis with actual deliverables, there are two main types: *ad hoc responses* and *analysis presentations*.

1. **Ad hoc responses:** Analysts receive requests to answer a variety of business questions, which may be spurred by questions raised by the reporting. Typically, these urgent requests are time sensitive and demand a quick turnaround. The analytics team may have to juggle multiple requests at the same time. As a result, the analyses cannot go as deep or wide as the analysts may like, and the deliverable is a short and concise report, which may or may not include any specific recommendations.
2. **Analysis presentations:** Some business questions are more complex in nature and require more time to perform a comprehensive, deep-dive analysis. These analysis projects result in a more formal deliverable, which includes two key sections: *key findings* and *recommendations*. The key findings section highlights the most meaningful and actionable insights gleaned from the analyses performed. The recommendations section provides guidance on what actions to take based on the analysis findings.

When you compare the two sets of reporting and analysis deliverables, the different purposes (information vs. insights) reveal the true colors of the outputs. Reporting pushes information to the organization, and analysis pulls insights from the reports and data. There may be other hybrid outputs such as annotated dashboards (analysis sprinkles on a reporting donut), which may appear to span the two areas. You should be able to determine whether a deliverable is primarily focused on reporting or analysis by its purpose (information/insights) and approach (push/pull).

Another key difference between reporting and analysis is **context**. Reporting provides **no or limited context** about what's happening in the data. In some cases, the end users already possess the necessary context to understand and interpret the data correctly. However, in other situations, the audience may not have the required background knowledge. Context is critical to good analysis. In order to tell a meaningful story with the data to drive specific actions, context becomes an essential component of the storyline.

Although they both leverage various forms of data visualization in their deliverables, analysis is different from reporting because it **emphasizes data points that are significant, unique, or special – and explain why they are important to the business**. Reporting may sometimes automatically highlight key changes in the data, but it's not going explain why these changes are (or aren't) important. Reporting isn't going to answer the "so what?" question on its own.

If you've ever had the pleasure of being a new parent, I would compare canned reporting, dashboards, and alerts to a six-month-old infant. It cries – often loudly – when something is wrong, but it can't tell you what is exactly wrong. The parent has to scramble to figure out what's going on (hungry, dirty diaper, no pacifier, teething, tired, ear infection, new Baby Einstein DVD, etc.). Continuing the parenting metaphor, reporting is also not going to tell you how to stop the crying.

The **recommendations** component is a key differentiator between analysis and reporting as it provides specific guidance on what actions to take based on the key insights found in the data. Even analysis outputs such as ad hoc responses may not drive action if they fail to include recommendations. Once a recommendation has been made, **follow-up** is another potent outcome of analysis because recommendations demand decisions to be made (go/no go/explore further). Decisions precede action. Action precedes value.

**Delivery**

As mentioned, reporting is more of a push model, where people can access reports through an analytics tool, Excel spreadsheet, widget, or have them scheduled for delivery into their mailbox, mobile device, FTP site, etc. Because of the demands of having to provide periodic reports (daily, weekly, monthly, etc.) to multiple individuals and groups, **automation** becomes a key focus in building and delivering reports. In other words, once the report is built, how can it be automated for regular delivery? Most of the analysts who I've talked to don't like manually building and refreshing reports on a regular basis. It's a job for robots or computers, not human beings who are still paying off their student loans for 4-6 years of higher education.

On the other hand, analysis is all about human beings using their superior reasoning and analytical skills to extract key insights from the data and form actionable recommendations for their organizations. Although analysis can be "submitted" to decision makers, it is more effectively presented **person-to-person**. In their book "Competing on Analytics", Thomas Davenport and Jeanne Harris emphasize the importance of **trust and credibility between the analyst and decision maker**. Decision makers typically don't have the time or ability to perform analyses themselves. With a "close, trusting relationship" in place, the executives will frame their needs correctly, the analysts will ask the right questions, and the executives will be more likely to take action on analysis they trust.

**Value**

When it comes to comparing the different roles of reporting and analysis, it's important to understand the relationship between reporting and analysis in driving value. I like to think of the data-driven stages (data > reporting > analysis > decision > action > value) as a series of dominoes. If you remove a domino, it can be more difficult or impossible to achieve the desired value.

In the "Path to Value" diagram above, it all starts with having the **right data that is complete and accurate**. It doesn't matter how advanced your reporting or analysis is if you don't have good, reliable data. If we skip the "reporting" domino, some seasoned analysts might argue that they don't need reports to do analysis (i.e., just give me the raw files and a database). On an individual basis that might be true for some people, but it doesn't work at the organizational level if you're striving to democratize your data.

Most companies have abundant reporting but may be missing the "analysis" domino. Reporting will rarely initiate action on its own as analysis is required to **help bridge the gap between data and action**. Having analysis doesn't guarantee that good decisions will be made, that people will actually act on the recommendations, that the business will take the right actions, or that teams will be able to execute effectively on those right actions. However, it is a necessary step closer to action and the potential value that can be realized through successful web analytics.

**Core Tools for Reporting and Dashboard layout Design:**

the most commonly cited among the websites that ranked the best BI tools. These five BI tools are not ranked in any particular order.

### 1. Microsoft Power BI

One of the most popular BI tools is Power BI, offered by leading software giant Microsoft. This tool is downloadable software, so you can choose to run analytics either on the cloud or in a reporting server. Syncing with sources such as Facebook, Oracle, and more, generate reports and dashboards in minutes with this interactive tool. It comes with built-in AI capabilities, Excel integration, and data connectors, and offers end-to-end data encryption and real-time access monitoring.

**Learn Power BI with Coursera:** In just two hours, you can learn the basics of Power BI Desktop with this guided project. You'll load and transform data to create interactive reports and dashboards.

### 2. Tableau

Tableau is known for its user-friendly data visualization capabilities, but it can do more than make pretty charts. Their offering includes live visual analytics, an interface that allows users to drag and drop buttons to spot trends in data quickly. The tool supports data sources such as Microsoft Excel, Box, PDF files, Google Analytics, and more. Its versatility extends to being able to connect with most databases.

**Learn Tableau with Coursera:** There are several options for learning how to use Tableau.

- Data Visualization with Tableau specialization from the University of California Davis
- Use Tableau for Your Data Science Workflow specialization from the University of California Irvine
- Data Visualization and Communication class with Tableau from Duke University
- Guided Project on Visualizing Citibike Trips with Tableau

### 3. QlikSense

QlikSense is a BI tool that emphasizes a self-service approach, meaning that it supports a wide range of analytics use cases, from guided apps and dashboards to custom and embedded analytics. It offers a user-friendly interface optimized for touchscreens, sophisticated AI, and high-performance cloud platforms. Its associative exploration capability, Search & Conversational Analytics, allows users to ask questions and uncover actionable insights, which helps increase data literacy for those new to using BI tools.

### 4. Dundas BI

Dundas BI is a browser-based BI tool that's been around for 25 years. Like Tableau, Dundas BI features a drag-and-drop function that allows users to analyze data on their own, without involving their IT team. The tool is known for its simplicity and flexibility through interactive dashboards, reports, and visual analytics. Since its inception as a data visualization tool in 1992, it has evolved into an end-to-end analytics platform that is able to compete with the new BI tools available today.

### 5. Sisense

Sisense is a user-friendly BI tool that focuses on being simplified and streamlined. With this tool, you can export data from sources like Google Analytics, Salesforce, and more. Its in-chip technology allows for faster data processing compared to other tools. Key features include the ability to embed white-label analytics, meaning a company can fully customize the services to its needs. Like others, it has a drag-and-drop feature. Sisense allows you to share reports and dashboards with your team members as well as externally.

**Other popular BI tools include:** Zoho Analytics, Oracle BI, SAS Visual Analytics, Domo, Datapine, Yellowfin BI, Looker, SAP Business Objects, Clear Analytics, Board, MicroStrategy, IBM Cognos Analytics, Tibco Spotfire, BIRT, Intercom, Google Data Studio, and HubSpot.

UNIT III: PROBABILITY FOR DATA ANALYTICS : Basic Probability: Uses of probability - Differentiate between sample space, event, independent and dependent - Calculate probability - Probability and Ven Diagramming: Analyze “this” OR “that” diagram - Analyze “this” AND “that” diagram - Analyze exclusive diagram - Joint probability - Conditional probability - Calculating Probability: Calculate P using a contingency table - Calculate P from trees - Calculate Bayes’ theorem - Calculate the mean in terms of probabilities - Calculate the variance and standard deviation in terms of probabilities - Calculate conditional probability

Probability Definition

*Probability is the measure of the likelihood that an event will occur. It is quantified as a number between 0 and 1.*

**Terms in Probability**

Here are some basic concepts or terminologies used in probability:

| Term | Definition |
| --- | --- |
| Sample Space | The set of all possible outcomes in a probability experiment. For instance, in a coin toss, it’s “head” and “tail”. |
| Sample Point | One of the possible results in an experiment. For example, in rolling a fair six-sided dice, sample points are 1 to 6. |
| Experiment | A process or trial with uncertain results. Examples include coin tossing, card selection, or rolling a die. |

| | |
|---|---|
| Event | A subset of the sample space representing certain outcomes.<br>Example: getting "1" when rolling a die. |
| Favorable Outcome | An outcome that produces the desired or expected consequence |

Probability Formula

***P(A) = Number of favorable outcomes / Total number of outcomes***

=n(A)/n(S)

Types of Probability

These are 3 major types of probability:

1. Theoretical Probability
2. Experimental Probability
3. Axiomatic Probability

**Theoretical Probability**

It is focused on the likelihood of anything occurring. The reasoning behind probability is the foundation of **scientific probability**.

If a coin is flipped, the statistical chance of having a head is 1/2.

**Experimental Probability**

It is founded on the results of an experiment. The experimental chance can be determined by dividing the total number of trials by the number of potential outcomes.

**For example**, if a coin is flipped ten times and heads are reported six times, the experimental chance of heads is 6/10 or 3/5.

**Axiomatic Probability**

A collection of laws or axioms that apply to all forms is established in axiomatic probability. **Kolmogorov** developed these axioms, which are known as **Kolmogorov's three axioms**.

**Uses of probability**: Probability is used in many practical aspects of everyday life.

- it is used in weather forecasting to predict the likelihood of rain or sunshine.
- used in insurance to assess risk and set premiums.
- sports to predict the outcome of games
- medical diagnosis to assess the likelihood of certain conditions.
- In finance, probability is used to assess investment risks and make informed decisions.

Overall, probability helps us make sense of uncertainty and make informed choices in various real-life situations.

**Differentiate b/w independent and dependent event**

**What are Dependent Events?**

For events to be considered dependent, one must have an influence over how probable another is. In other words, a dependent event can only occur if another event occurs first.

For example, say you'd like to go on vacation at the end of next month, but that depends on having enough money to cover the trip.

Consider the following examples:

- Getting into a traffic accident is dependent upon driving or riding in a vehicle.
- If you park your vehicle illegally, you're more likely to get a parking ticket.

**What are Independent Events?**

An event is deemed independent when it isn't connected to another event, or Independent events don't influence one another or have any effect on how probable another event is.

Other examples of pairs of independent events include:

- Taking an Uber ride and getting a free meal at your favorite restaurant
- Winning a card game and running out of bread
- Finding a dollar on the street and buying a lottery ticket; finding a dollar isn't dictated by buying a lottery ticket, nor does buying the ticket increase your chances of finding a dollar.

**Some example problems of probability:**

**Example 1:** A coin is thrown 3 times .what is the probability that atleast one head is obtained?

**Sol:** Sample space = [HHH, HHT, HTH, THH, TTH, THT, HTT, TTT]

Total number of ways = 2 × 2 × 2 = 8.  Fav. Cases = 7

P (A) = 7/8

OR

P (of getting at least one head) = 1 – P (no head)⇒1 – (1/8) = 7/8

**Example 2:** Find the probability of getting a numbered card when a card is drawn from the pack of 52 cards.

**Sol:** Total Cards = 52. Numbered Cards = (2, 3, 4, 5, 6, 7, 8, 9, 10) 9 from each suit 4 × 9 = 36

P (E) = 36/52 = 9/13

**Example 3:** There are 5 green 7 red balls. Two balls are selected one by one without replacement. Find the probability that first is green and second is red.

**Sol:** P (G) × P (R) = (5/12) x (7/11) = 35/132

**Example 4:** What is the probability of getting a sum of 7 when two dice are thrown?

**Sol:**  Probability math - Total number of ways = 6 × 6 = 36 ways. Favorable cases = (1, 6) (6, 1) (2, 5) (5, 2) (3, 4) (4, 3) --- 6 ways. P (A) = 6/36 = 1/6

**Example 5:** 1 card is drawn at random from the pack of 52 cards.

(i) Find the Probability that it is an honor card.

(ii) It is a face card.

**Sol:** (i) honor cards = (A, J, Q, K) 4 cards from each suits = 4 × 4 = 16

P (honor card) = 16/52 = 4/13

(ii) face cards = (J,Q,K) 3 cards from each suit = 3 × 4 = 12 Cards.

P (face Card) = 12/52 = 3/13

**Example 6:** Two cards are drawn from the pack of 52 cards. Find the probability that both are diamonds or both are kings.

**Sol:** Total no. of ways = ${}^{52}C_2$

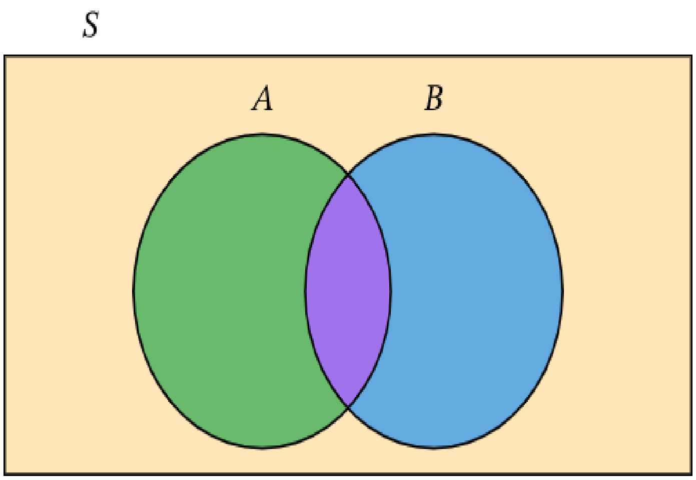
**Case I:** Both are diamonds = ${}^{13}C_2$

**Case II:** Both are kings = $^{4}C_{2}$

P (both are diamonds or both are kings) = ($^{13}C_{2}$ + $^{4}C_{2}$ ) / $^{52}C_{2}$

In probability, a Venn diagram is a figure with one or more circles inside a rectangle that describes logical relations between events. The rectangle in a Venn diagram represents the sample space or the universal set, that is, the set of all possible outcomes. A circle inside the rectangle represents an event, that is, a subset of the sample space.

We consider the following Venn diagram involving two events, ▯ and ▯.




In the diagram above, we have two events ▯ and ▯ within the sample space (or universal set) ▯.Sometimes, the sample space is denoted by ▯ or ▯ instead of ▯.Colored regions in this Venn diagram represent the following events: GreenandpurpleregionsBlueandpurpleregionsPurpleregionGreen,purple,andblueregionsYellowregionalternatively,: ▯,: ▯,: ▯∩▯,: ▯∪▯,: ▯∪▯,(▯∪▯)′.

**Definition: Two-Event Venn Diagrams**

Let ▯ and ▯ be events described in a Venn diagram. Then,

- the circles do not overlap if ▯ and ▯ are mutually exclusive events, that is, ▯∩▯=∅;
- the circles overlap if ▯∩▯≠∅,in which case the intersection ▯∩▯ is represented by the overlapping region;
- the region outside both circles but within the rectangle represents the complement of the union of both events, that is, ▯∪▯ or, alternatively, (▯∪▯)′.

Within each divided region of a Venn diagram, we can add data in one of the following ways:

- the outcomes of the event,
- the number of outcomes in the event,
- the probability of the event.

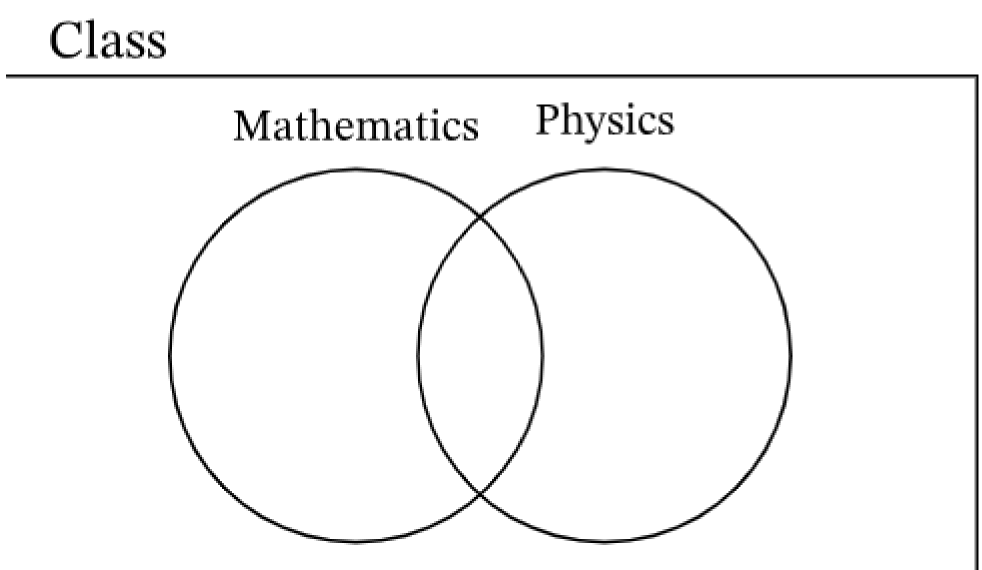
In out first example, we will use a Venn diagram to organize our data and use it to compute the probability of an event.

**Example 1: Organizing Data using Venn Diagrams to Find Probabilities**

A class contains 100 students; 70 of them like mathematics, 60 like physics, and 40 like both. If a student is chosen at random, using a Venn diagram, find the probability that they like mathematics but not physics.

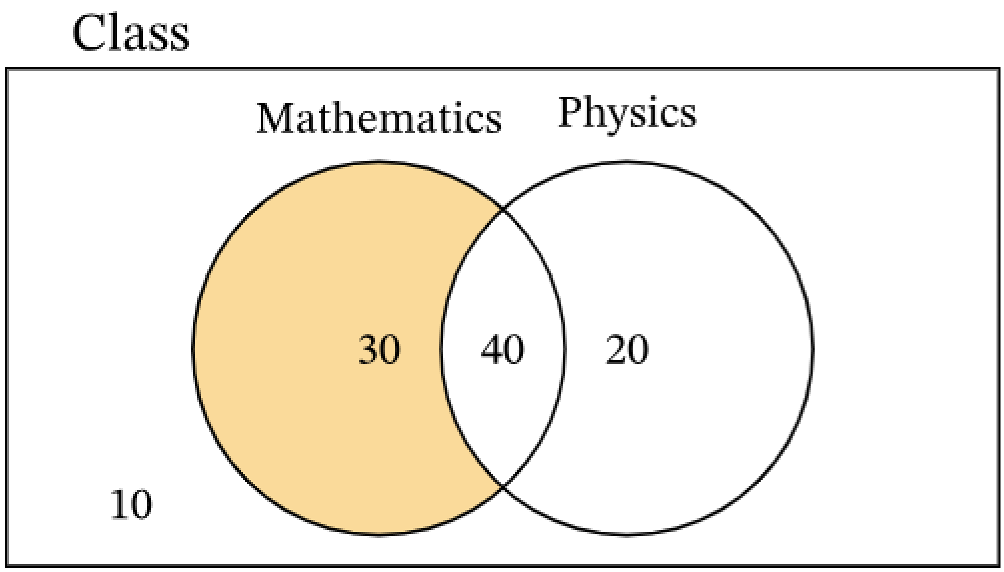
**Answer**

We begin by drawing an empty Venn diagram to represent this example.




We know that the overlapping region in a Venn diagram represents the intersection of the events. We are given that 40 students belong to this intersection, since they like both mathematics and physics.

70 students like mathematics, and 40 of them also like physics. This tells us that the number of students who like mathematics only is $70-40=30$.Similarly, we can deduce that $60-40=20$ students like physics only. This leads to the following Venn diagram.




We note that the number 10 outside is to ensure that the sum of all values within the Venn diagram is equal to 100, since the class contains 100 students total. We are looking for the probability that a randomly chosen student likes mathematics but not physics. The region of the Venn diagram representing this event is highlighted below.

From the Venn diagram, we note that 30 students like mathematics but not physics. Since a student is randomly selected, we can obtain the probability of this event when dividing this number by the total number of students, which is 100.

Hence, the probability that a randomly selected student likes mathematics but not physics is 30100=0.3.

**What is Joint Probability?**

Joint probability is the statistical metric that quantifies the chances of multiple events happening at the same time. Mathematically, the joint probability of two events, A and B is represented as P(A or simply P.

**1.1. Joint Probability Formula**

If A and B are two events, then the joint probability that both A and B occur is given by:

- $P(A \cap B) = P(A) * P(B)$

## 3. Why is Joint Probability Important?

Joint probability is foundational across various sectors, from finance to artificial intelligence. It asists experts in risk assessment, predictive analysis, and decoding complex scenarios with intertwined events.

## 2. Calculating Joint Probability

The method to calculate joint probability hinges on whether our events are independent (one doesn't influence the outcome of the other) or dependent (one has an effect on the other).

### 2.1. Example 1: Independent Events (Rolling Dice)

Let's consider rolling two dice:
– Event A: Rolling a 3 on the first die.
– Event B: Rolling a 4 on the second die.

The outcome of one dice roll doesn't impact the other. Therefore, the joint probability is just the product of their individual chances:

- $P(A\cap B)=P(A)\cdot P(B)=16\cdot 16=136$ $PA\cap B=PA\cdot PB=16\cdot 16=136$

So, the probability of rolling a 3 on the first die and a 4 on the second simultaneously is 1 in 36.

### 2.2. Example 2: Dependent Events (Drawing Cards)

Imagine drawing two cards from a standard 52-card deck:
– Event A: Drawing an Ace first.
– Event B: Drawing a King second, without putting the Ace back.

The first draw affects the probabilities of the second draw. The joint probability calculation, therefore, is:

- $P(A\cap B)=P(A)\times P(BA)=152\times 151$ $PA\cap B=PA\times PBA=152\times 151$

This equation factors in the reduced number of cards once the Ace is drawn.

**What Is Conditional Probability?**

**Conditional probability** is one of the important concepts in probability and statistics. The "probability of A given B" (or) the "probability of A with respect to the condition B" is denoted by the conditional probability P(A | B) (or) P (A / B) (or) PB�A). Thus, P(A | B) represents the probability of A which happens after event B has happened already. the probability of an event may alter if there is a condition given.

**Definition of Conditional Probability**

If A and B are two events associated with the same sample space of a random experiment, **the conditional probability of event A given that B has occurred is given by P(A/B) = P( A ∩ B)/ P (B)**, provided P(B) ≠ 0.

Let us assume the two events A and B as follows:

- A = the event of getting at least two tails
- B = the event of getting a head on the first toss

Then, A = {HTT, THT, TTH, TTT} and B = {HHH, HHT, HTH, HTT}.

Then P(A) = 4/8 = 1/2 and P(B) = 4/8 = 1/2.

We have to find the probability of getting at least two tails given that it is a head on the first toss. It means, out of all elements of B, we have to choose only the ones with two tails. We can see that among the elements of B, there is only one element (which is HTT) with two tails. Thus, the required probability is P(A | B) = 1/4 (only 1 outcome of B is favorable to A out of 4 outcomes of B).

**Conditional Probability Formula**

In the above example, we have got P(A | B) = 1/4, here 1 represents the element HTT which is present both in "A and B" and 4 represents the total number of elements in B. Using this, we can derive the formula of conditional probability as follows.

**P(A | B) = P(A ∩ B) / P(B)** (Note that P(B) ≠ 0 here)

Similarly, we can define P(B | A) as follows:

**P(B | A) = P(A ∩ B) / P(A)** (Note that P(A) ≠ 0 here)

These formulas are also known as the "Kolmogorov definition" of conditional probability.

Here:

- P(A | B) = The probability of A given B (or) the probability of A which happens after B
- P(B | A) = The probability of B given A (or) the probability of B which happens after A
- P(A ∩ B) = The probability of happening of both A and B
- P(A) = The probability of A
- P(B) = The probability of B

**Properties of Conditional Probability**

**Property 1**

Let S be the sample space of an experiment and A be any event. Then P(S | A) = P(A | A) = 1.

**Proof:**

By the formula of conditional probability,

P(S | A) = P(S ∩ A) / P(A) = P(A) / P(A) = 1

P(A | A) = P(A ∩ A) / P(A) = P(A) / P(A) = 1

Hence property 1 is proved.

**Property 2**

Let S be the sample space of an experiment and A and B be any two events. Let E be any other event such that P(E) ≠ 0. Then P((A ∪B) | E) = P(A | E) + P(B | E) - P((A ∩ B) | E).

**Proof:**

By the formula of conditional probability,

P((A ∪B) | E) = [P((A ∪B) ∩ E)] / P(E)

= [ P(A ∩ E) ∪P(B ∩ E) ] / P(E) (using a property of sets)

= [P(A ∩ E) + P(B ∩ E) - P(A ∩ B ∩ E)] / P(E) (using addition theorem of probability)

= P(A ∩ E) / P(E) + P(B ∩ E) / P(E) - P(A ∩ B ∩ E) / P(E)

= P(A | E) + P(B | E) - P((A ∩ B) | E) (By conditional probability formula)

Hence property 2 is proved.

**Property 3**

P(A' | B) = 1 - P(A | B), where A' is the complement of the set A.

**Proof:**

By Property 1, we have P(S | B) = 1.

We know that S = A ∪A'. Thus by the above property,

P( A ∪A' | B) = 1

Since A and A' are disjoint events,

P(A | B) + P(A' | B) = 1

P(A' | B) = 1 - P(A | B)

Hence property 3 is proved.

**Dependent and Independent Events**

The definition of independent and dependent events is connected to conditional probability. Let us see the definitions of independent and dependent events along with their formulas.

**Dependent Events**

Dependent events, as the name suggests, are any two events in which the happening of one event depends on the happening of the other event.

- If A depends on B, then the probability of A is P(A | B).
- If B depends on A, then the probability of B is P(B | A).

By the conditional probability formulas,

P(A | B) = P(A ∩ B) / P(B) ⇒P(A ∩ B) = P(A | B) · P(B)

P(B | A) = P(A ∩ B) / P(A) ⇒P(A ∩ B) = P(B | A) · P(A)

Thus, two event A and B are said to be dependent events if one of the conditions is satisfied.

- $P(A \cap B) = P(A \mid B) \cdot P(B)$ (or)
- $P(A \cap B) = P(B \mid A) \cdot P(A)$

**Independent Events**

Independent events, as the name suggests, are any two events in which the happening of one event does not depend on the happening of the other event. i.e., if A and B are independent then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$. Thus, to get the formula of independent events, we just need to replace $P(A \mid B)$ with $P(A)$ (or $P(B \mid A)$ with $P(B)$) in one of the above (dependent events) formulas. Hence, two events are said to be independent if

$P(A \cap B) = P(A) \cdot P(B)$

This is also called as multiplication rule of probability.

**Important Notes:**

- The probability of A given B is called the conditional probability and it is calculated using the formula $P(A \mid B) = P(A \cap B) / P(B)$.
- The events that are part of conditional probability are dependent events. For example, if we have $P(A \mid B)$ anywhere in the problem, then it means that A and B are dependent.
- If two events A and B are independent, then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.
- For any two events A and B, $P(A \cap B) = P(A) \cdot P(B)$. This is called the multiplication theorem of probability.

**Examples of Conditional Probability**

**Example 1: The table below shows the occurrence of diabetes in 100 people. Let D and N be the events where a randomly selected person "has diabetes" and "not overweight". Then find P(D | N).**

| | Diabetes (D) | No Diabetes (D') |
|---|---|---|
| Not overweight (N) | 5 | 45 |

| | | |
|---|---|---|
| Overweight (N') | 17 | 33 |

**Solution:**

From the given table, P(N) = (5+45) / 100 = 50/100.

P(D ∩ N) = 5/100.

By the conditional probability formula,

P(D | N) = P(D ∩ N) / P(N)

= (5/100) / (50/100)

= 5/50

= 1/10

**Answer: P(D | N) = 1/10.**

- **Example 2: The probability that it will be sunny on Friday is 4/5. The probability that an ice cream shop will sell ice creams on a sunny Friday is 2/3 and the probability that the ice cream shop sells ice creams on a non-sunny Friday is 1/3. Then find the probability that it will be sunny and the ice cream shop sells the ice creams on Friday.**

  **Solution:**

  Let us assume that the probabilities for a Friday to be sunny and for the ice cream shop to sell ice creams be S and I respectively. Then,

  P(S) = 4/5.

  P(I | S) = 2/3.

  P(I | S') = 1/3.

  We have to find P(S ∩ I).

  We can see that S and I are dependent events. By using the dependent events' formula of conditional probability,

  P(S ∩ I) = P(I | S) · P(S) = (2/3) · (4/5) = 8/15.

  **Answer: The required probability = 8/15.**

**USING TREE DIAGRAMS TO CALCULATE PROBABILITIES**

Tree diagrams can be helpful in organizing information in probability problems; they help provide a structure for understanding probability. We assign the appropriate probabilities to the events shown on the branches of the tree.
By multiplying probabilities along a path through the tree, we can find probabilities for "and" events, which are intersections of events with an example.

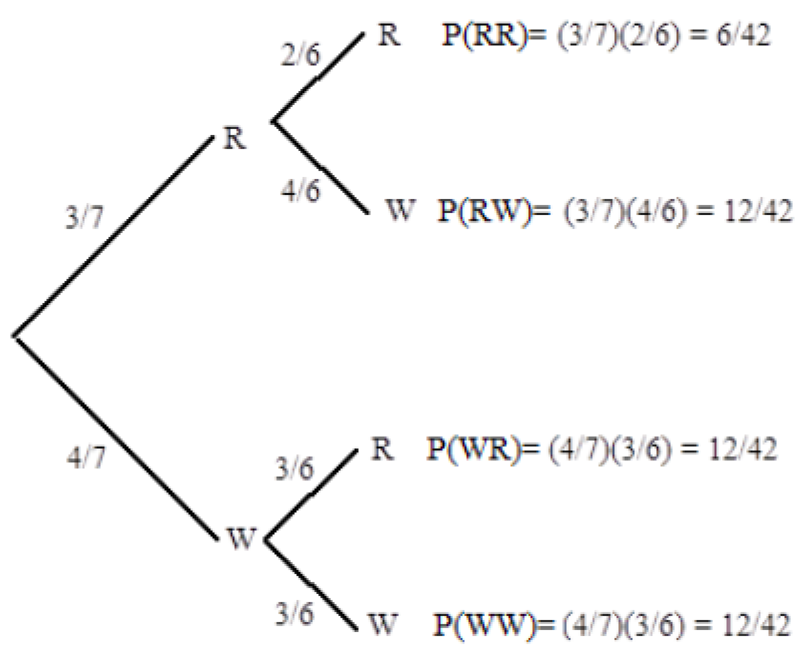
Suppose a jar contains 3 red and 4 white marbles. If two marbles are drawn without replacement, find the following probabilities using a tree diagram.

a. The probability that both marbles are red.
b. The probability that the first marble is red and the second white.
c. The probability that one marble is red and the other white.

**Solution**

Let R*R* be the event that the marble drawn is red, and let W be the event that the marble drawn is white.

We draw the following tree diagram.




d. The probability that both marbles are red is P(RR)=6/42*P(RR)=6/42*

e. The probability that the first marble is red and the second is white is P(RW)=12/42*P(RW)=12/42*
f. For the probability that one marble is red and the other is white, we observe that this can be satisfied if the first is red and the second is white, **or** if the first is white and the second is red. The "or" tells us we'll be using the Addition Rule from Section 7.2.

Furthermore events RW and WR are mutually exclusive events, so we use the form of the Addition Rule that applies to mutually exclusive events.

Therefore

P(one marble is red and the other marble is white)

=P(RW or WR)=P(RW)+P(WR)=12/42+12/42=24/42

**Bayes Theorem Basics and Derivation:**

- In probability theory, **Bayes theorem** describes probability of an event based on prior knowledge of conditions that might be related to the event
- **Derivation:** Bayes theorem is derived through conditional probability equation by equating P(A and B) of below mentioned equation 1 and equation 2

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{-- equation 1}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \text{-- equation 2}$$

From equation 1 and 2 on equating for expression of P(A∩B)

$$P(A/B) * P(B) = P(B/A) * P(A)$$

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)} \quad \text{--- Bayes Theorem}$$

- P(A) also known as prior probability or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- P(A|B) is the conditional probability of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- P(B|A) is the conditional probability of B given A. Also known as likelihood.
- P(B) is the prior or marginal probability of B, and acts as a normalizing constant.

Calculate the mean in terms of probabilities:

probability distribution tells us the probability that a random variable takes on certain values.

For example, the following probability distribution tells us the probability that a certain soccer team scores a certain number of goals in a given game:

| Goals (X) | Probability P(X) |
|---|---|
| 0 | 0.18 |
| 1 | 0.34 |
| 2 | 0.35 |
| 3 | 0.11 |
| 4 | 0.02 |

***Note:*** *The probabilities in a valid probability distribution will always add up to 1. We can confirm that this probability distribution is valid: 0.18 + 0.34 + 0.35 + 0.11 + 0.02 = 1.*

To find the **mean** (sometimes called the "expected value") of any probability distribution, we can use the following formula:

**Mean (Or "Expected Value") of a Probability Distribution:**

**$\mu = \Sigma x * P(x)$**

**where:**

- **x: Data value**
- **P(x): Probability of value**

**Example : Mean Number of Vehicle Failures**

The following probability distribution tells us the probability that a given vehicle experiences a certain number of battery failures during a 10-year span:

| Failures (X) | Probability P(X) |
|---|---|
| 0 | 0.24 |
| 1 | 0.57 |
| 2 | 0.16 |
| 3 | 0.03 |

**Question:** What is the mean number of expected failures for this vehicle?

**Solution:** The mean number of expected failures is calculated as:

$\mu = 0*0.24 + 1*0.57 + 2*0.16 + 3*0.03 = 0.98$ failures.

UNIT IV: DISTRIBUTIONS: Distributions: Analyze distributions - Discrete distributions - Binomial distributions - Poisson distributions - Continuous Distributions: Identify continuous distributions - Calculate continuous distributions - Identify cumulative distributions - Identify normal distributions - Calculate normal distributions - Compare quartiles and normal distributions - Identify skew

Distibution: Distribution can be defined as a mathematical function or a set of probabilities that assigns a probability to each possible outcome of a random variable. This function or set of probabilities can be represented graphically using various types of plots, such as histograms, density plots, or cumulative distribution functions.

Understanding the distribution of a dataset is essential for making statistical inferences, predicting future outcomes, and understanding the behavior of random variables in various situations.

Types of Distributions:

Types of Probability Distributions

There are two types of distributions based on the type of data generated by the experiments.

1. Discrete Probability Distributions

These distributions model the probabilities of random variables that can have discrete values as outcomes. For example, the possible values for the random variable X that represents the number of heads that can occur when a coin is tossed twice are the set {0, 1, 2} and not any value from 0 to 2 like 0.1 or 1.6.

Examples: Binomial Distribution , Poisson Distribution

2. Continuous Probability Distributions

These distributions model the probabilities of random variables that can have any possible outcome. For example, the possible values for the random variable X that represents weights of citizens in a town which can have any value like 34.5, 47.7, etc.,

Discrete vs Continuous Distribution

| **Discrete Distribution** | **Continuous Distribution** |
|---|---|
| Random variable can only take on a finite or countable number of values | Random variable can take on any value within a certain range or interval |
| Probability function assigns probabilities to each possible outcome | Probability function assigns probabilities to each possible value within a range or interval |
| Examples: binomial distribution, Poisson distribution, geometric distribution | Examples: normal distribution, exponential distribution, beta distribution |
| Used to model events with discrete outcomes, such as number of successes in a fixed number of trials | Used to model events with continuous outcomes, such as the height of individuals in a population |
| Probability of any single outcome is non-zero | Probability of any single value is zero |
| Cumulative distribution function is stepwise | Cumulative distribution function is continuous |

Types of Discrete Probability Distributions

Here are the list of types of Discrete probability distributions explained with examples.

**Binomial Distribution**

This is generated for random variables with only two possible outcomes. Let p denote the probability of an event is a success which implies 1 – p is the probability of the event being a failure. Performing the experiment repeatedly and plotting the probability each time gives us the Binomial distribution.

The most common example given for Binomial distribution is that of flipping a coin n number of times and calculating the probabilities of getting a particular number of heads. More real-world examples include the number of successful sales calls for a company or whether a drug works for a disease or not.

The PMF is given as,

$$n_{C_x} p^x (1-p)^x$$

where p is the probability of success, n is the number of trials and x is the number of times we obtain a success.

**Poisson Distribution**

This distribution describes the events that occur in a fixed interval of time or space. An example might make this clear. Consider the case of the number of calls received by a customer care center per hour. We can estimate the average number of calls per hour but we cannot determine the exact number and the exact time at which there is a call. Each occurrence of an event is independent of the other occurrences.

The PMF is given as,

$$P(X = x) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

where λ is the average number of times the event has occurred in a certain period of time, x is the desired outcome and e is the Euler's number.

**Why do normal distributions matter?**

All kinds of variables in natural and social sciences are normally or approximately normally distributed. Height, birth weight, reading ability, job satisfaction, or SAT scores are just a few examples of such variables.

Because normally distributed variables are so common, many statistical tests are designed for normally distributed populations.

Understanding the properties of normal distributions means you can use inferential statistics to compare different groups and make estimates about populations using samples.

**What are the properties of normal distributions?**

Normal distributions have key characteristics that are easy to spot in graphs:

- The mean, median and mode are exactly the same.
- The distribution is symmetric about the mean—half the values fall below the mean and half above the mean.

- The distribution can be described by two values: the mean and the standard deviation.

The mean is the location parameter while the standard deviation is the scale parameter.

The mean determines where the peak of the curve is centered. Increasing the mean moves the curve right, while decreasing it moves the curve left.

The standard deviation stretches or squeezes the curve. A small standard deviation results in a narrow curve, while a large standard deviation leads to a wide curve.

### Empirical rule

The **empirical rule**, or the 68-95-99.7 rule, tells you where most of your values lie in a normal distribution:

- Around 68% of values are within 1 standard deviation from the mean.
- Around 95% of values are within 2 standard deviations from the mean.
- Around 99.7% of values are within 3 standard deviations from the mean.

**Example: Using the empirical rule in a normal distribution**You collect SAT scores from students in a new test preparation course. The data follows a normal distribution with a mean score (*M*) of 1150 and a standard deviation (*SD*) of 150.

Following the empirical rule:

- Around 68% of scores are between 1,000 and 1,300, 1 standard deviation above and below the mean.
- Around 95% of scores are between 850 and 1,450, 2 standard deviations above and below the mean.
- Around 99.7% of scores are between 700 and 1,600, 3 standard deviations above and below the mean.

The empirical rule is a quick way to get an overview of your data and check for any outliers or extreme values that don't follow this pattern.

If data from small samples do not closely follow this pattern, then other distributions like the t-distribution may be more appropriate. Once you identify the distribution of your variable, you can apply appropriate statistical tests.

### Central limit theorem

The central limit theorem is the basis for how normal distributions work in statistics.

In research, to get a good idea of a population mean, ideally you'd collect data from multiple random samples within the population. A **sampling distribution of the mean** is the distribution of the means of these different samples.

The central limit theorem shows the following:

- Law of Large Numbers: As you increase sample size (or the number of samples), then the sample mean will approach the population mean.
- With multiple large samples, the sampling distribution of the mean is normally distributed, even if your original variable is not normally distributed.

Parametric statistical tests typically assume that samples come from normally distributed populations, but the central limit theorem means that this assumption isn't necessary to meet when you have a large enough sample.

You can use parametric tests for large samples from populations with any kind of distribution as long as other important assumptions are met. A sample size of 30 or more is generally considered large.

For small samples, the assumption of normality is important because the sampling distribution of the mean isn't known. For accurate results, you have to be sure that the population is normally distributed before you can use parametric tests with small samples.

## Quartiles of the Standard Normal Distribution:

A random variable is said to follow the standard normal distribution if it follows the normal distribution with mean $\mu$=0 and variance $\sigma^2$=0. The values of the quartiles for the standard normal distribution can be found by looking at the Z table.

First Quartile = -0.675 (One-fourth of the area under the standard normal curve lies below this value)

Second Quartile = 0 (Half of the area under the standard normal curve lies below this value)

Third Quartile = +0.675 (Three-fourths of the area under the standard normal curve lies below this value)

## Quartiles of Normal Distribution:

Let X be a random variable that follows the normal distribution with mean $\mu$ and variance $\sigma^2$. We first convert the random variable into a Z score by the formula,

$$Z = X\text{-}\mu/\sigma.$$

We then set the value of Z equal to the value of the quartile of the standard normal curve given above and solve for X.

$$X = \mu + Z\sigma.$$

This gives us the value of the quartile for the given normal distribution.

The first, second, and third quartiles of a normal distribution are given by the formulae,

$$\text{First Quartile} = \mu - 0.675\sigma.$$

$$\text{Second Quartile} = \mu.$$

$$\text{Third Quartile} = \mu + 0.675\sigma.$$

### Example:

Let X be a normal random variable with a mean of 110 and a variance of 100. Then the quartiles of the normal distribution are given as,

$$\text{First Quartile} = \mu - 0.675\sigma = 110 - 0.675*10 = 103.25.$$

$$\text{Second Quartile} = \mu = 110.$$

$$\text{Third Quartile} = \mu + 0.675\sigma = 116.75.$$

UNIT III: PROBABILITY FOR DATA ANALYTICS : Basic Probability: Uses of probability - Differentiate between sample space, event, independent and dependent - Calculate probability - Probability and Ven Diagramming: Analyze “this” OR “that” diagram - Analyze “this” AND “that” diagram - Analyze exclusive diagram - Joint probability - Conditional probability - Calculating Probability: Calculate P using a contingency table - Calculate P from trees - Calculate Bayes’ theorem - Calculate the mean in terms of probabilities - Calculate the variance and standard deviation in terms of probabilities - Calculate conditional probability

Probability Definition

*Probability is the measure of the likelihood that an event will occur. It is quantified as a number between 0 and 1.*

**Terms in Probability**

Here are some basic concepts or terminologies used in probability:

| Term | Definition |
|---|---|
| Sample Space | The set of all possible outcomes in a probability experiment. For instance, in a coin toss, it’s “head” and “tail”. |
| Sample Point | One of the possible results in an experiment. For example, in rolling a fair six-sided dice, sample points are 1 to 6. |
| Experiment | A process or trial with uncertain results. Examples include coin tossing, card selection, or rolling a die. |

| | |
|---|---|
| Event | A subset of the sample space representing certain outcomes.<br>Example: getting “1” when rolling a die. |
| Favorable Outcome | An outcome that produces the desired or expected consequence |

Probability Formula

***P(A) = Number of favorable outcomes / Total number of outcomes***

=n(A)/n(S)

Types of Probability

These are 3 major types of probability:

1. Theoretical Probability
2. Experimental Probability
3. Axiomatic Probability

**Theoretical Probability**

It is focused on the likelihood of anything occurring. The reasoning behind probability is the foundation of **scientific probability**.

If a coin is flipped, the statistical chance of having a head is 1/2.

**Experimental Probability**

It is founded on the results of an experiment. The experimental chance can be determined by dividing the total number of trials by the number of potential outcomes.

**For example**, if a coin is flipped ten times and heads are reported six times, the experimental chance of heads is 6/10 or 3/5.

**Axiomatic Probability**

A collection of laws or axioms that apply to all forms is established in axiomatic probability. **Kolmogorov** developed these axioms, which are known as **Kolmogorov’s three axioms**.

**Uses of probability**: Probability is used in many practical aspects of everyday life.

- it is used in weather forecasting to predict the likelihood of rain or sunshine.
- used in insurance to assess risk and set premiums.
- sports to predict the outcome of games
- medical diagnosis to assess the likelihood of certain conditions.
- In finance, probability is used to assess investment risks and make informed decisions.

Overall, probability helps us make sense of uncertainty and make informed choices in various real-life situations.

**Differentiate b/w independent and dependent event**

**What are Dependent Events?**

For events to be considered dependent, one must have an influence over how probable another is. In other words, a dependent event can only occur if another event occurs first.

For example, say you'd like to go on vacation at the end of next month, but that depends on having enough money to cover the trip.

Consider the following examples:

- Getting into a traffic accident is dependent upon driving or riding in a vehicle.
- If you park your vehicle illegally, you're more likely to get a parking ticket.

**What are Independent Events?**

An event is deemed independent when it isn't connected to another event, or Independent events don't influence one another or have any effect on how probable another event is.

Other examples of pairs of independent events include:

- Taking an Uber ride and getting a free meal at your favorite restaurant
- Winning a card game and running out of bread
- Finding a dollar on the street and buying a lottery ticket; finding a dollar isn't dictated by buying a lottery ticket, nor does buying the ticket increase your chances of finding a dollar**.**

**Some example problems of probability:**

**Example 1:** A coin is thrown 3 times .what is the probability that atleast one head is obtained?

**Sol:** Sample space = [HHH, HHT, HTH, THH, TTH, THT, HTT, TTT]

Total number of ways = 2 × 2 × 2 = 8. Fav. Cases = 7

P (A) = 7/8

OR

P (of getting at least one head) = 1 – P (no head)⇒1 – (1/8) = 7/8

**Example 2:** Find the probability of getting a numbered card when a card is drawn from the pack of 52 cards.

**Sol:** Total Cards = 52. Numbered Cards = (2, 3, 4, 5, 6, 7, 8, 9, 10) 9 from each suit 4 × 9 = 36

P (E) = 36/52 = 9/13

**Example 3:** There are 5 green 7 red balls. Two balls are selected one by one without replacement. Find the probability that first is green and second is red.

**Sol:** P (G) × P (R) = (5/12) x (7/11) = 35/132

**Example 4:** What is the probability of getting a sum of 7 when two dice are thrown?

**Sol:** Probability math - Total number of ways = 6 × 6 = 36 ways. Favorable cases = (1, 6) (6, 1) (2, 5) (5, 2) (3, 4) (4, 3) --- 6 ways. P (A) = 6/36 = 1/6

**Example 5:** 1 card is drawn at random from the pack of 52 cards.

(i) Find the Probability that it is an honor card.

(ii) It is a face card.

**Sol:** (i) honor cards = (A, J, Q, K) 4 cards from each suits = 4 × 4 = 16

P (honor card) = 16/52 = 4/13

(ii) face cards = (J,Q,K) 3 cards from each suit = 3 × 4 = 12 Cards.

P (face Card) = 12/52 = 3/13

**Example 6:** Two cards are drawn from the pack of 52 cards. Find the probability that both are diamonds or both are kings.

**Sol:** Total no. of ways = ${}^{52}C_2$

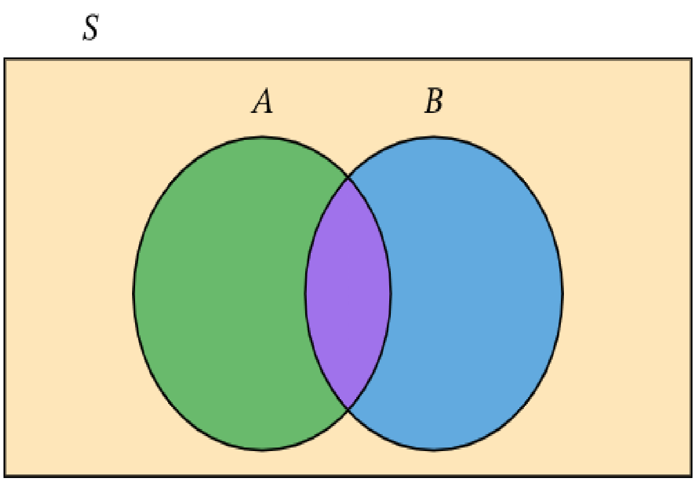
**Case I:** Both are diamonds = ${}^{13}C_2$

**Case II:** Both are kings = ${}^{4}C_{2}$

P (both are diamonds or both are kings) = $({}^{13}C_{2} + {}^{4}C_{2}) / {}^{52}C_{2}$

In probability, a Venn diagram is a figure with one or more circles inside a rectangle that describes logical relations between events. The rectangle in a Venn diagram represents the sample space or the universal set, that is, the set of all possible outcomes. A circle inside the rectangle represents an event, that is, a subset of the sample space.

We consider the following Venn diagram involving two events, □ and □.




In the diagram above, we have two events □ and □ within the sample space (or universal set) □.Sometimes, the sample space is denoted by □ or □ instead of □.Colored regions in this Venn diagram represent the following events: GreenandpurpleregionsBlueandpurpleregionsPurpleregionGreen,purple,andblueregionsYellowregionalternatively,: □,: □,: □∩□,: □∪□,: □∪□,(□∪□)′.

**Definition: Two-Event Venn Diagrams**

Let □ and □ be events described in a Venn diagram. Then,

- the circles do not overlap if □ and □ are mutually exclusive events, that is, □∩□=∅;
- the circles overlap if □∩□≠∅,in which case the intersection □∩□ is represented by the overlapping region;
- the region outside both circles but within the rectangle represents the complement of the union of both events, that is, □∪□ or, alternatively, (□∪□)′.

Within each divided region of a Venn diagram, we can add data in one of the following ways:

- the outcomes of the event,
- the number of outcomes in the event,
- the probability of the event.

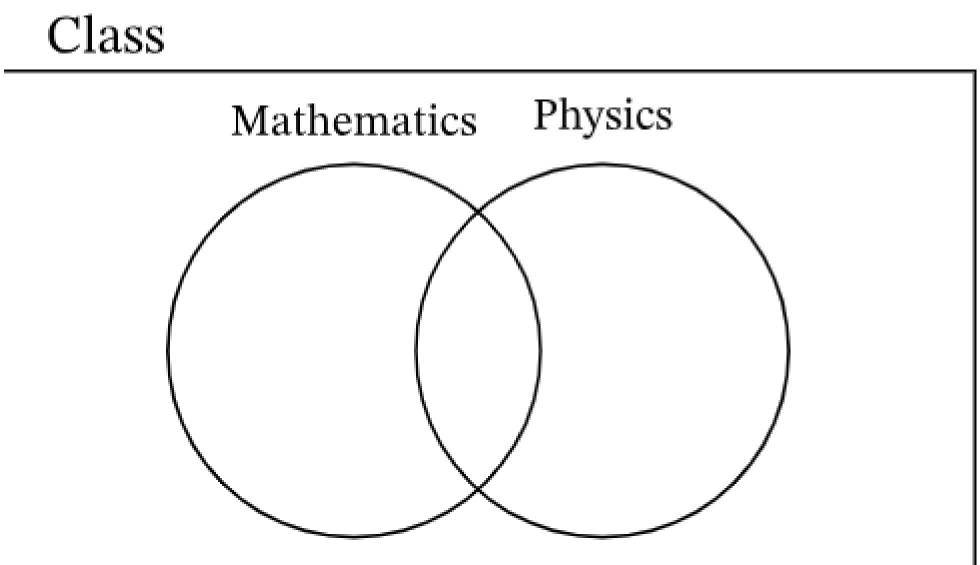
In out first example, we will use a Venn diagram to organize our data and use it to compute the probability of an event.

**Example 1: Organizing Data using Venn Diagrams to Find Probabilities**

A class contains 100 students; 70 of them like mathematics, 60 like physics, and 40 like both. If a student is chosen at random, using a Venn diagram, find the probability that they like mathematics but not physics.

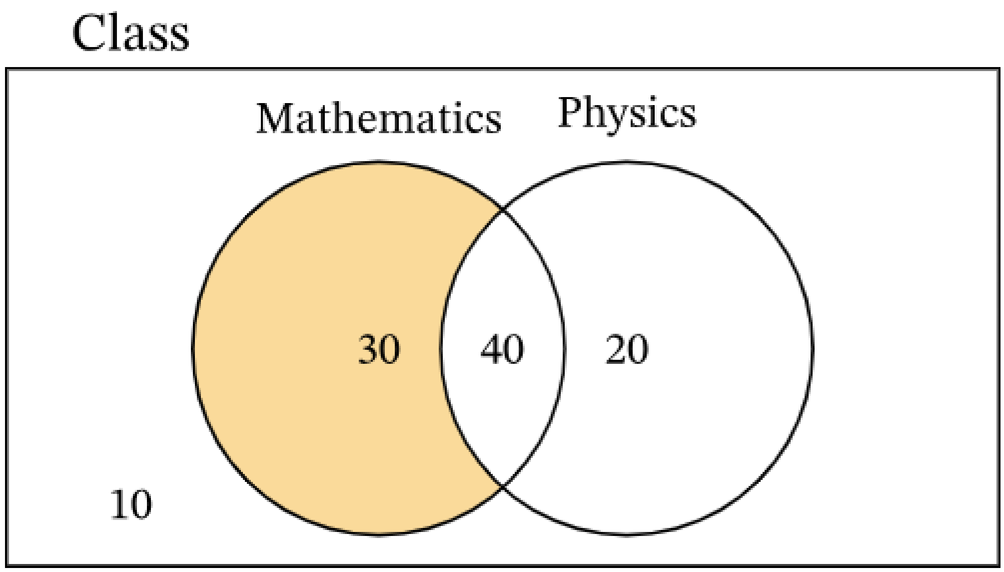
**Answer**

We begin by drawing an empty Venn diagram to represent this example.




We know that the overlapping region in a Venn diagram represents the intersection of the events. We are given that 40 students belong to this intersection, since they like both mathematics and physics.

70 students like mathematics, and 40 of them also like physics. This tells us that the number of students who like mathematics only is $70-40=30$.Similarly, we can deduce that $60-40=20$ students like physics only. This leads to the following Venn diagram.




We note that the number 10 outside is to ensure that the sum of all values within the Venn diagram is equal to 100, since the class contains 100 students total. We are looking for the probability that a randomly chosen student likes mathematics but not physics. The region of the Venn diagram representing this event is highlighted below.

From the Venn diagram, we note that 30 students like mathematics but not physics. Since a student is randomly selected, we can obtain the probability of this event when dividing this number by the total number of students, which is 100.

Hence, the probability that a randomly selected student likes mathematics but not physics is 30100=0.3.

**What is Joint Probability?**

Joint probability is the statistical metric that quantifies the chances of multiple events happening at the same time. Mathematically, the joint probability of two events, A and B is represented as P(A or simply P.

**1.1. Joint Probability Formula**

If A and B are two events, then the joint probability that both A and B occur is given by:

- $P(A \cap B) = P(A) * P(B)$

### 3. Why is Joint Probability Important?

Joint probability is foundational across various sectors, from finance to artificial intelligence. It asists experts in risk assessment, predictive analysis, and decoding complex scenarios with intertwined events.

### 2. Calculating Joint Probability

The method to calculate joint probability hinges on whether our events are independent (one doesn't influence the outcome of the other) or dependent (one has an effect on the other).

#### 2.1. Example 1: Independent Events (Rolling Dice)

Let's consider rolling two dice:
– Event A: Rolling a 3 on the first die.
– Event B: Rolling a 4 on the second die.

The outcome of one dice roll doesn't impact the other. Therefore, the joint probability is just the product of their individual chances:

- P(A∩B)=P(A)· P(B)=16· 16=136 $PA\cap B=PA\cdot PB=16\cdot 16=136$

So, the probability of rolling a 3 on the first die and a 4 on the second simultaneously is 1 in 36.

#### 2.2. Example 2: Dependent Events (Drawing Cards)

Imagine drawing two cards from a standard 52-card deck:
– Event A: Drawing an Ace first.
– Event B: Drawing a King second, without putting the Ace back.

The first draw affects the probabilities of the second draw. The joint probability calculation, therefore, is:

- P(A∩B)=P(A)×P(BA)=152×151 $PA\cap B=PA\times PBA=152\times 151$

This equation factors in the reduced number of cards once the Ace is drawn.

**What Is Conditional Probability?**

**Conditional probability** is one of the important concepts in probability and statistics. The "probability of A given B" (or) the "probability of A with respect to the condition B" is denoted by the conditional probability P(A | B) (or) P (A / B) (or) PB�A). Thus, P(A | B) represents the probability of A which happens after event B has happened already. the probability of an event may alter if there is a condition given.

**Definition of Conditional Probability**

If A and B are two events associated with the same sample space of a random experiment, **the conditional probability of event A given that B has occurred is given by P(A/B) = P( A ∩ B)/ P (B)**, provided P(B) ≠ 0.

Let us assume the two events A and B as follows:

- A = the event of getting at least two tails
- B = the event of getting a head on the first toss

Then, A = {HTT, THT, TTH, TTT} and B = {HHH, HHT, HTH, HTT}.

Then P(A) = 4/8 = 1/2 and P(B) = 4/8 = 1/2.

We have to find the probability of getting at least two tails given that it is a head on the first toss. It means, out of all elements of B, we have to choose only the ones with two tails. We can see that among the elements of B, there is only one element (which is HTT) with two tails. Thus, the required probability is P(A | B) = 1/4 (only 1 outcome of B is favorable to A out of 4 outcomes of B).

**Conditional Probability Formula**

In the above example, we have got P(A | B) = 1/4, here 1 represents the element HTT which is present both in "A and B" and 4 represents the total number of elements in B. Using this, we can derive the formula of conditional probability as follows.

**P(A | B) = P(A ∩ B) / P(B)** (Note that P(B) ≠ 0 here)

Similarly, we can define P(B | A) as follows:

**$P(B | A) = P(A \cap B) / P(A)$** (Note that $P(A) \neq 0$ here)

These formulas are also known as the "Kolmogorov definition" of conditional probability.

Here:

- $P(A | B)$ = The probability of A given B (or) the probability of A which happens after B
- $P(B | A)$ = The probability of B given A (or) the probability of B which happens after A
- $P(A \cap B)$ = The probability of happening of both A and B
- $P(A)$ = The probability of A
- $P(B)$ = The probability of B

**Properties of Conditional Probability**

**Property 1**

Let S be the sample space of an experiment and A be any event. Then $P(S | A) = P(A | A) = 1$.

**Proof:**

By the formula of conditional probability,

$P(S | A) = P(S \cap A) / P(A) = P(A) / P(A) = 1$

$P(A | A) = P(A \cap A) / P(A) = P(A) / P(A) = 1$

Hence property 1 is proved.

**Property 2**

Let S be the sample space of an experiment and A and B be any two events. Let E be any other event such that $P(E) \neq 0$. Then $P((A \cup B) | E) = P(A | E) + P(B | E) - P((A \cap B) | E)$.

**Proof:**

By the formula of conditional probability,

$P((A \cup B) | E) = [P((A \cup B) \cap E)] / P(E)$

$= [ P(A \cap E) \cup P(B \cap E) ] / P(E)$ (using a property of sets)

$= [P(A \cap E) + P(B \cap E) - P(A \cap B \cap E)] / P(E)$ (using addition theorem of probability)

$= P(A \cap E) / P(E) + P(B \cap E) / P(E) - P(A \cap B \cap E) / P(E)$

$= P(A | E) + P(B | E) - P((A \cap B) | E)$ (By conditional probability formula)

Hence property 2 is proved.

**Property 3**

$P(A' | B) = 1 - P(A | B)$, where A' is the complement of the set A.

**Proof:**

By Property 1, we have $P(S | B) = 1$.

We know that $S = A \cup A'$. Thus by the above property,

$P(A \cup A' | B) = 1$

Since A and A' are disjoint events,

$P(A | B) + P(A' | B) = 1$

$P(A' | B) = 1 - P(A | B)$

Hence property 3 is proved.

**Dependent and Independent Events**

The definition of independent and dependent events is connected to conditional probability. Let us see the definitions of independent and dependent events along with their formulas.

**Dependent Events**

Dependent events, as the name suggests, are any two events in which the happening of one event depends on the happening of the other event.

- If A depends on B, then the probability of A is $P(A | B)$.
- If B depends on A, then the probability of B is $P(B | A)$.

By the conditional probability formulas,

$P(A | B) = P(A \cap B) / P(B) \Rightarrow P(A \cap B) = P(A | B) \cdot P(B)$

$P(B | A) = P(A \cap B) / P(A) \Rightarrow P(A \cap B) = P(B | A) \cdot P(A)$

Thus, two event A and B are said to be dependent events if one of the conditions is satisfied.

- $P(A \cap B) = P(A \mid B) \cdot P(B)$ (or)
- $P(A \cap B) = P(B \mid A) \cdot P(A)$

**Independent Events**

Independent events, as the name suggests, are any two events in which the happening of one event does not depend on the happening of the other event. i.e., if A and B are independent then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$. Thus, to get the formula of independent events, we just need to replace $P(A \mid B)$ with $P(A)$ (or $P(B \mid A)$ with $P(B)$) in one of the above (dependent events) formulas. Hence, two events are said to be independent if

$P(A \cap B) = P(A) \cdot P(B)$

This is also called as multiplication rule of probability.

**Important Notes:**

- The probability of A given B is called the conditional probability and it is calculated using the formula $P(A \mid B) = P(A \cap B) / P(B)$.
- The events that are part of conditional probability are dependent events. For example, if we have $P(A \mid B)$ anywhere in the problem, then it means that A and B are dependent.
- If two events A and B are independent, then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.
- For any two events A and B, $P(A \cap B) = P(A) \cdot P(B)$. This is called the multiplication theorem of probability.

**Examples of Conditional Probability**

**Example 1: The table below shows the occurrence of diabetes in 100 people. Let D and N be the events where a randomly selected person "has diabetes" and "not overweight". Then find P(D | N).**

| | Diabetes (D) | No Diabetes (D') |
|---|---|---|
| Not overweight (N) | 5 | 45 |

| | | |
|---|---|---|
| Overweight (N') | 17 | 33 |

**Solution:**

From the given table, P(N) = (5+45) / 100 = 50/100.

$P(D \cap N) = 5/100$.

By the conditional probability formula,

$P(D \mid N) = P(D \cap N) / P(N)$

= (5/100) / (50/100)

= 5/50

= 1/10

**Answer: $P(D \mid N) = 1/10$.**

- **Example 2: The probability that it will be sunny on Friday is 4/5. The probability that an ice cream shop will sell ice creams on a sunny Friday is 2/3 and the probability that the ice cream shop sells ice creams on a non-sunny Friday is 1/3. Then find the probability that it will be sunny and the ice cream shop sells the ice creams on Friday.**

  **Solution:**

  Let us assume that the probabilities for a Friday to be sunny and for the ice cream shop to sell ice creams be S and I respectively. Then,

  P(S) = 4/5.

  $P(I \mid S) = 2/3$.

  $P(I \mid S') = 1/3$.

  We have to find $P(S \cap I)$.

  We can see that S and I are dependent events. By using the dependent events' formula of conditional probability,

  $P(S \cap I) = P(I \mid S) \cdot P(S) = (2/3) \cdot (4/5) = 8/15$.

  **Answer: The required probability = 8/15.**

**USING TREE DIAGRAMS TO CALCULATE PROBABILITIES**

Tree diagrams can be helpful in organizing information in probability problems; they help provide a structure for understanding probability. We assign the appropriate probabilities to the events shown on the branches of the tree.
By multiplying probabilities along a path through the tree, we can find probabilities for "and" events, which are intersections of events with an example.

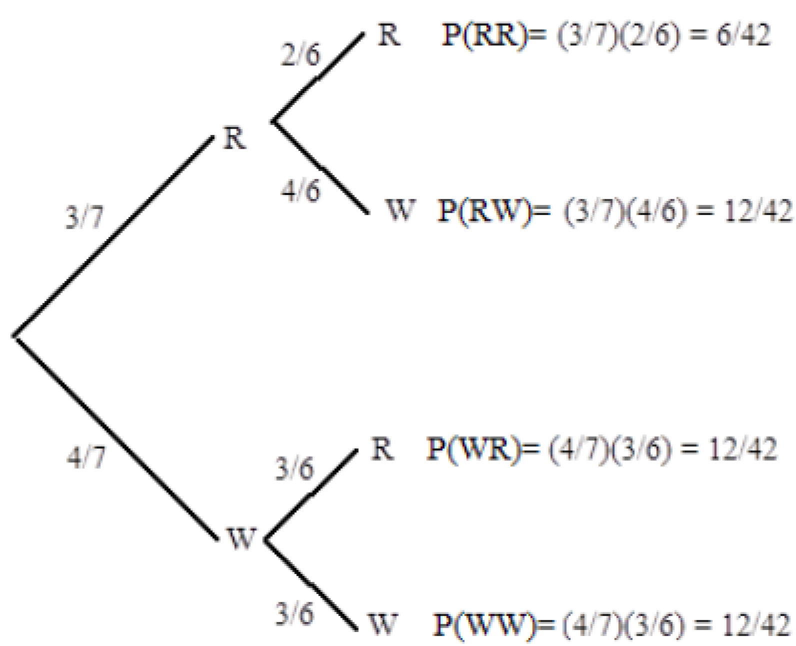
Suppose a jar contains 3 red and 4 white marbles. If two marbles are drawn without replacement, find the following probabilities using a tree diagram.

a. The probability that both marbles are red.
b. The probability that the first marble is red and the second white.
c. The probability that one marble is red and the other white.

**Solution**

Let R*R* be the event that the marble drawn is red, and let W be the event that the marble drawn is white.

We draw the following tree diagram.




d. The probability that both marbles are red is P(RR)=6/42*P(RR)=6/42*

e. The probability that the first marble is red and the second is white is P(RW)=12/42*P(RW)=12/42*
f. For the probability that one marble is red and the other is white, we observe that this can be satisfied if the first is red and the second is white, **or** if the first is white and the second is red. The "or" tells us we'll be using the Addition Rule from Section 7.2.

Furthermore events RW and WR are mutually exclusive events, so we use the form of the Addition Rule that applies to mutually exclusive events.

Therefore

P(one marble is red and the other marble is white)

=P(RW or WR)=P(RW)+P(WR)=12/42+12/42=24/42

**Bayes Theorem Basics and Derivation:**

- In probability theory, **Bayes theorem** describes probability of an event based on prior knowledge of conditions that might be related to the event
- **Derivation:** Bayes theorem is derived through conditional probability equation by equating P(A and B) of below mentioned equation 1 and equation 2

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad \text{-- equation 1}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \qquad \text{-- equation 2}$$

From equation 1 and 2 on equating for expression of P(A∩B)

$$P(A/B) * P(B) = P(B/A) * P(A)$$

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)} \qquad \text{--- Bayes Theorem}$$

- P(A) also known as prior probability or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- P(A|B) is the conditional probability of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- P(B|A) is the conditional probability of B given A. Also known as likelihood.
- P(B) is the prior or marginal probability of B, and acts as a normalizing constant.

Calculate the mean in terms of probabilities:

probability distribution tells us the probability that a random variable takes on certain values.

For example, the following probability distribution tells us the probability that a certain soccer team scores a certain number of goals in a given game:

| Goals (X) | Probability P(X) |
|---|---|
| 0 | 0.18 |
| 1 | 0.34 |
| 2 | 0.35 |
| 3 | 0.11 |
| 4 | 0.02 |

***Note:*** *The probabilities in a valid probability distribution will always add up to 1. We can confirm that this probability distribution is valid: 0.18 + 0.34 + 0.35 + 0.11 + 0.02 = 1.*

To find the **mean** (sometimes called the "expected value") of any probability distribution, we can use the following formula:

**Mean (Or "Expected Value") of a Probability Distribution:**

**$\mu = \Sigma x * P(x)$**

**where:**

- **x: Data value**
- **P(x): Probability of value**

**Example : Mean Number of Vehicle Failures**

The following probability distribution tells us the probability that a given vehicle experiences a certain number of battery failures during a 10-year span:

| Failures (X) | Probability P(X) |
|---|---|
| 0 | 0.24 |
| 1 | 0.57 |
| 2 | 0.16 |
| 3 | 0.03 |

**Question:** What is the mean number of expected failures for this vehicle?

**Solution:** The mean number of expected failures is calculated as:

$\mu = 0*0.24 + 1*0.57 + 2*0.16 + 3*0.03 = 0.98$ failures.