A diagram of a program

Description automatically generated

A diagram of data

Description automatically generated A white background with black text

Description automatically generated A black text on a white background

Description automatically generated

---------------------------------------------------------------------------------------------------------------------------------------------------------------------

**BI : Introduction :** 2030 More than 180 Zettabytes of data. How to harness the power of data? Make informed decisions.

|  |  |
| --- | --- |
| A diagram of a person typing on a keyboard  Description automatically generated | A diagram of data ecosystem  Description automatically generated |
| A diagram of a company's landscape  Description automatically generated | A diagram of a diagram  Description automatically generated |
| A close-up of a chart  Description automatically generated | A diagram of a triangle with points and a flag  Description automatically generated |

BI : The ability to apprehend the interrelationships of preented facts in such a was a to guide action toward a desired goal.

BI Also can be defined of **combination of various technologies, tools and methodologies**:

* **Gather, analyzes and transform data into meaningful information**.
* Included processes like data Mining, analysis, benchmarking, Descriptive analysis, Visualization and reporting.
* Presents information into understandable and actionable items.
* BI analysis based on current data and historical data to uncover patterns, trends and relationships within data, and also to explores problems and root causes.

Overview of BI Components, Steps in the BI Process :

1. Identify sources of dat (from which to gather and extract data).
2. Determine the data warehouses or repositories where the data from different sources is stored.
3. Analyzing and making the data meaningful , such as Extract, Transform and load process (ETL process).
4. Various reporting and presentation tools help to manipulate and present complex data into visual and understandable reports that help make business decisions.

**Key Performance Indicators (KPIs) and Metrics**

KPIs are not goals, rather **measurable outcomes** that reveal the organization’s performance against its established goals. First are to define your KPIs. For examples of various types of KPIs :

1. Financial KPIs include gross and net profit margin, inventory turnover and liquidity ratio.
2. Marketing KPIs are customer acquisition cost and conversion rate.
3. Project management KPIs include productivity and return on investment.
4. Customer service KPIs are net promoter score and customer effort score.
5. Human resources KPI include turnover rate, vacancy rate, employee satisfaction level and costs per hire.

Metrics are another measurement to track the performance of specific business processes**. Metrics provide context to performance of key business goals**, but are not critical like KPIs. For example :

1. In IT industry, are total number of bugs reported and total monthly support tickets.

In Customer service industry example :

1. Average ticket resolution time are KPI because it is measured against a goal
2. Total number of support ticket received are Metrics because tracks the performance of service/software is offered.

**All KPIs are metrics, but not all metrics are KPIs.** Both are quantitative measurements, for different purposes.

KPI are typically strategic, whereas Metrics are tactical

KPI how successful the business is in achieving the goal whereas Metrics show how successful the tasks are at achieving the KPIs

KPIs have a high-level perspective (HAVE Targets) whereas Metrics are considered lower level indicators (measure performance)

KPIs aligned with the business plan and impact businesss goals whereas Metrics provide additional information to support the results.

KPIs drive actions and used more by upper management whereas Metrics measures the success of everyday business activities.

**KPIs is an indicator of business goals or failure whereas Metrrics track the performance of specific process.**

SMART KPIs : Specific, Measurable, Attainable, Realistic, and time-bounded. (Set realistic target and goals).Monitor KPI using visual dashboards.

HOW?? Identify the critical aspects of your business, and set realistic goals,

Data analytics can be categorized into four key types :

1. Descriptive analytic : Helps to understand what happened.
2. Diagnostic analytics : Answer why did this happen.
3. Predictive analytics : Answer the question or what might happen in the future.
4. Prescriptive analytics : What should you do next.

**DESCRIPTIVE ANALYTICS :**

* Initial phase of data processing.
* Type of analytics that uses historical (ie. Compare current moths sales performance with previous months) and current data to spot any trends in relationships between that data (count year on year growth).
* Doesn’t help for the future predictions, but it draws insights considering what happened.
* Helps to create business metrics and comprehensive reports
* Define KPIs that enable businesses to monitor their performance. (for example Number of followers, Engagement rates and also revenue generated through social media platforms).

HOW descriptive analytics works :

CREATE METRICS aligning with business goals => Identify the right data to support these metrics => Collect, clean and prepare the relevant data for analysis => To analyze the clean data using BI software to dive Deeper to uncover valuable insights and performance patterns.

Descriptive analytics uses :

* measures of distribution like frequeny or count,
* measures of central tendency like mean, median and mode.
* measures of variability such as Variance and standard deviation.

**Descriptive analytics showcase** : PIVOT tables, Reports, Visual components (histograms, pie chart, line graphs)

Ie. Netflix use Descriptive analytics to find customer preferences, behavior trends, and forecast demand.

**Descriptive analytics can do for a business : Evaluate, compare, spot anomalies, and then find relative strengths and weaknesses.**

|  |  |
| --- | --- |
| PROS Descriptive analytics | CONS Descriptive analytics |
| 1. Allows for simple analysis without having a lot experience 2. It helps present data insights in a graphical format 3. Answer business performance questions 4. Summarize the central characteristics of your data 5. Help business to compare their performance with competitros or other product lines. | 1. Only limited to answering what and why 2. Doesn’t offer future predicitons 3. Data collection is time consuming and expensive 4. Cann’t generalize findings to a bigger population. |

**DIAGNOSTIC ANALYTICS :**

* Can help identify the factors or conditions that lead to a specific results such as decline in sales.
* Determine the reason an event occurred.
* Discover hidden correlations and connections between variables
* Determine causal relationships (one event influence another event) ie. Decrease in followup appointment to doctor, causes increase in readmission patient in doctor,
* Detect anomalies (ie.customer complaints increase from credit card customers,a bank deteted an anomaly in credit card spending. There was a significant deviation from cardholders typical spending behavior. )
* Isolate patterns (examined energy consumption data, helped isolate a recurring pattern of increase energy usage during specific hours. Further investigation revealed that a specific piece of equipment that ran during these hours had missed its maintenance schedule, causing higher consumption)

HOW diagnostic analytics works :

Use past data to provide answers to queries => employ several techniques (ie. Probability theory, filtering, time series analysis, regression).

1. Probability theory helps to find the likelihood of specific outcomes based on historical data. Example : Probability of a customer making a purchase based on past purchased history.
2. Time series analysis helps uncover pattern, trends, and anomalies in time-based data. Example . Detect seasonal fluctuations in sales.
3. Filtering helps you to isolate relevant data points or patterns from irrelevant data, so we can focus on the most critical information for analytics. Example Filter our new customers when you want to analyze customer churn in your long term customer base.
4. Regression analysis help you to find relationships between variables,

|  |  |
| --- | --- |
| PROS Diagnostic analytics | CONS Diagnostic analytics |
| 1. Provides more granular insights 2. Helpful in exploring anomalies and outliers. | 1. Possibility of mistaking correlation for causation (might result in a wrong conclusion). 2. Requires additional sources to supplement analysis(include third party historical and real time data) 3. Requires more time and 4. Requires higher level skills than descriptive analytics. |

**PREDICTIVE ANALYTICS :**

* Predicts future outcomes
* Relies on historical data (ie Weather forecast rely on temperature and humidity level)
* Employs various techniques such as statistical modeling, data mining and machine learning for predictions.

SCENARIO Coffee Shop :

* Using Descriptive analytics to know your sales data and inventory level of coffee beans.
* Using Diagnostic analytics to understand sales patterns and find reason variance from expected sales.
* Can we predict the quantity of coffee bean bags required at different times to ensure optimal stock levels??
* Descriptive and diagnostic analytics focus on what happened, predictive analytics focus on what will happen in the future.

|  |  |
| --- | --- |
| PROS Predictive analytics | CONS Predictive analytics |
| 1. Save time by eliminating the need for manual research and analysis. 2. It brings more accuracy by using advanced algorithms and machine learning techniques to analyze data 3. It can analyze massive data sets by processing large volumes of data, resulting in more accurate predictions. 4. ML can continuously refine predictive analytics to ensure that they stay relevant. 5. Parameters adjusted more accurate prediction based on data patterns, and can be setup update themselves automatically when new data is available. | 1. Requires time 2. Requires large enough data set to produce meaningful results. 3. Need to put significant effort into gathering and preparing data. 4. Might require high upfront costs for several reasons, such as need experts and specialized tools and platforms. |

**PRESCRIPTIVE ANALYTICS :**

* Recommend the optimal course of action to achieve a specific goal.
* It draws from inputs from descriptive, diagnostic and predictive analytics processes.
* Descriptive analytics Reveals data trends, whereas diagnostic analytics uncover why trends occurred . Predictive analytics forecast future trends based on current data. Prescriptive analytics uses all these inputs to recommend what should be done next.

Prescriptive analytics harnesses technology like Machine learning and AI to implement techniques like data optimization, simulation and decision analysis methods. All algorithms apply rules to interpret data to generate recommendations.

|  |  |
| --- | --- |
| PROS Prescreptive analytics | CONS Prescriptive analytics |
| 1. Making unbiased data driven decisions, rather than relying solely on instincts. 2. Provide insights that human mind may not decipher. 3. Helps automaking the decision making process, thus increasing operational efficiency (adjust Uber fares based on demand) 4. Effective forecast tool 5. Making complex decisions considering numerous data points. | 1. The results are only as good as the inputs, so make sure the inputs are valid. 2. Building a prescriptive analytics models are complex (requires training, testing and adjusting parameters to obtain desired result) 3. Contonuous monitoring is crucial 4. Certain decisions, like venture capital investments could need human judgement (tools to support decision making) |

A screenshot of a computer

Description automatically generated

The differences between descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics.

* Descriptive analytics summarizes raw data to reveal trends. (most basic step in BI, involves historical and current data to spot trends, Descriptive analytics also helps spot the business’s strengths and weaknesses) **WHAT HAPPENED – *MEASURES WHAT YOUR ORGANIZATION is DOING***
* Diagnostic analytics correlates the data with other data sets to find the reason for the trends.(Why?? Using some techniques)

**WHY EVENT OCCURRED-**

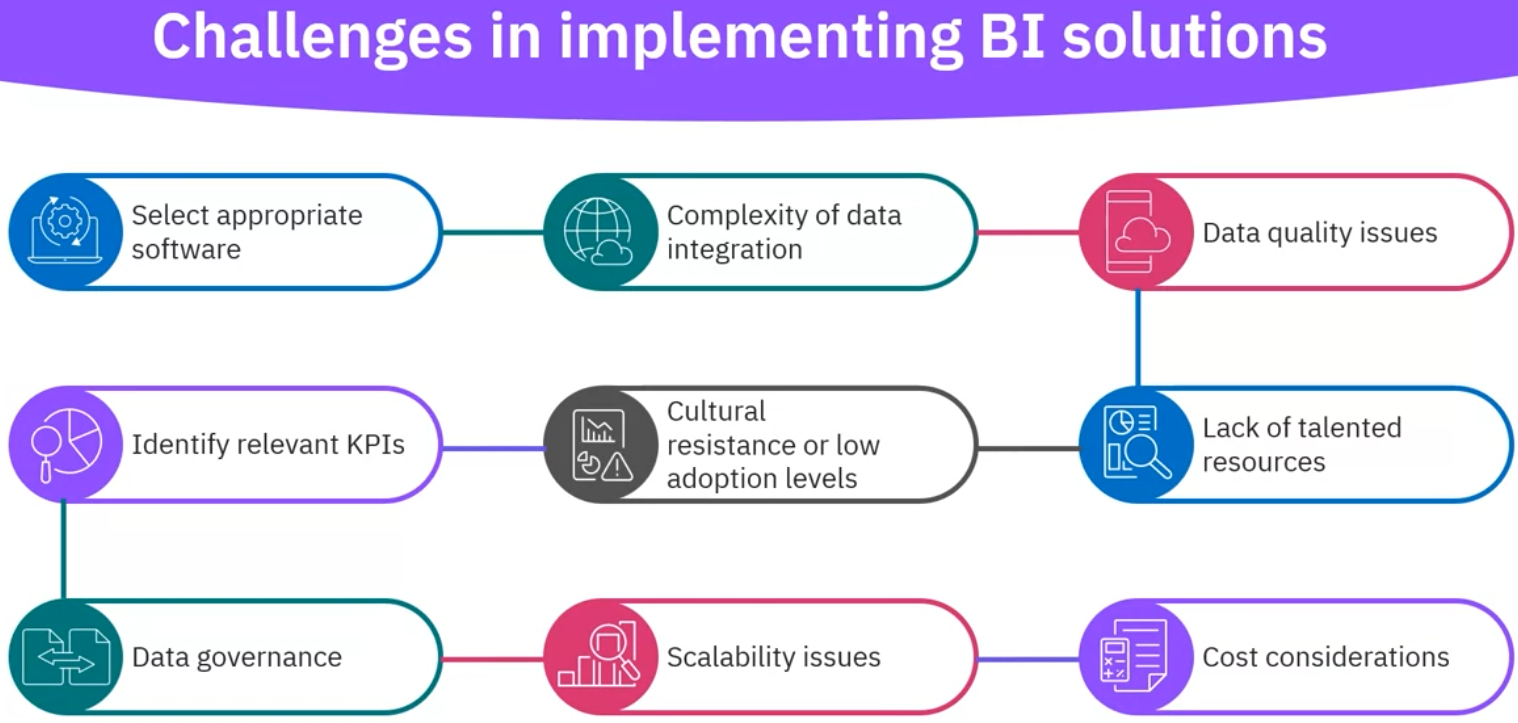
* Predictive analytics uses historical data *to forecast/predict trends*. (using statistical model or Machine Learning)

**WHAT WILL HAPPEN – *WHAT THE COMPANY WILL BE DOING***

* Prescriptive data goes beyond descriptive, diagnostic, and predictive analytics to recommend the next course of action based on various data points. **WHAT SHOULD BE DONE- *SHOW WHAT THE COMPANY SHOULD BE DOING***

METRIC SELECTION depends on activity measurement :

1. Transactional activity : Events like orders, sales, support tickets.
2. Periodic snapshots : Bank Balances.
3. Accumulating snapshots : Measures elapsed time between an order.
4. Factless measurement : Activities outside your company (not happening in your company).
5. Customer metrics (evaluate customer satisfaction and behavior)
6. Operational metrics (monitor the efficiency and effectiveness)
7. Sales and marketing metrics (assess the effectiveness of sales and marketing efforts)
8. Financial data (time series data)
9. Production data (enhancing production efficiency, refresh data frequently).
10. HR data(data governance and data security is extremely important).



A screenshot of a medical report

Description automatically generated A screenshot of a computer

Description automatically generated

A screenshot of a questionnaire

Description automatically generated A screenshot of a website

Description automatically generated

A diagram of a business intelligence

Description automatically generated A diagram of data reporting

Description automatically generated

A diagram of data analysis

Description automatically generated

Data Analysis involves interpreting historical data to identify trends, patterns and anomalies.

Data Mining extract valuable insights from large datasets using techniques examples Machinelearning.

Key components of a business intelligence (BI) system, including :

* data sources and integration,
* data warehousing,
* data analysis and mining,
* reporting and visualization systems,
* software technologies, and
* advanced analytics

BI ecosystem or BI environment comprises four key elements:

* Data, Data is the most important element of the ecosystem, as it is the raw material for analytics and reporting.
* People, People involved in BI include data analysts responsible for sourcing and processing data and users who query the data to generate the required reports to make informed decisions.
* Processes, must be designed to fulfill the requirements of the business and produce accurate results. The BI architecture typically consists of data collection, data integration and management, data analysis, and data reporting and visualization processes.
* Technologies. must be carefully chosen to align with the purpose of the BI system and always be kept up to date.

A diagram of software applications

Description automatically generated A black text on a white background

Description automatically generated

BI Tool categories :

1. Unified Modeling Language (UML)

* Visual modeling language used to design, visualize and document software systems.
* UML can model and represents the components, relationships and processes of data analysis.
* Example draw.io for drawing UML diagrams, lucidchart best UML tools

1. ETL Tools (Extract, Transform, Load)

* Extracting data from various sources
* Transforming it into a consistent format
* Loading it into a target system or data warehouse
* Requires building complex data pipelines that move data between locations.
* Example IBM Infosphere Datastage, Talend, AWS Glue, Oracle Data Integrator, Azure Data Factory

1. Data Warehousing tools

* Collect, organize and store large amounts of data from various sources in a single central location.
* This enable easy organization and processing of data for the analysis phase.
* Support online analytical processing or OLAP tools that enable multi dimensional analysis of large volumes of data.
* Choose cloud or on-premises data warehouse solution based on the use cases.
* Example IBM DB2 Warehouse, AWS Redhsift, AWS S3 , AWS RDS, Google BigQuery, Google DataFlow, Google Data Studio, MS Azure SQL Database and Azure Synapse Analytics, Snowflake, SAS Cloud, SAP Data Warehouse cloud.

1. Data preparation and cleaning tools

* Standardizing the data
* Eliminating unwanted outliers
* Fixing cross data errors
* Dealing with missing data
* Validating the data
* Example : MS Excel, IBM Infosphere Quality Stage, Power BI, Talend, Tableau

1. Visualization and dashboarding tools

* Enable users to represent visual data through interactive charts, graphs and maps to gain desired insights.
* Provide a user friendly interface with sorting, filtering, drag and drop functionalities
* Allowing users to interact with the interface to get real time insights without help from IT Team.
* Example IBM Cognos Analytics, Tableau, Power BI; Sisense, Looker, Domo, QlikSense.

1. Reporting and querying tools

* Allow users to generate reports
* Perform ad-hoc queries to analyze and interpret large amounts of data.
* Example IBM Cognos Analytics, SAP Business Object, Oracle BI, MicroStrategy, Zoho Analytics.

1. Mobil BI Software tools

* Provide BI On mobile devices such as smartphones and tablets.
* Example IBM Cognos Analytics, Sisense, Power BI, Google Analytics, Amazon Quicksight

1. Cloud based SAAS BI software

* Can be accessed by users through a web browser.
* Subscription based pricing model.
* Example IBM Cognos Analytics, Alteryx, Domo, Looker, PowerBI

1. Spreadsheet

* Store data in a tabular format
* Making it easy to query and format.
* Uses a data exploration, cleansing and reporting.
* They support performance tracking , accounting and customer intelligence.
* Example MS Excel

A white text on a white background

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **IBM Cognos Analytics Features:**   * Self-service and geospatial capabilities * Seamless access on a desktop or mobile device * Advanced analytics, customization, scalable distribution, and scheduling abilities to meet business goals | **Tableau Features:**   * User-friendly interface that allows users to create intuitive visualizations and interactive dashboards * Drag-and-drop functionality that makes it easy for users to explore data and gain insights * Advanced analytics capabilities that support complex calculations, statistical analysis, and forecasting * Built-in functions and integration with R and Python for advanced analytics * Seamless integration capabilities with various data sources, including databases, spreadsheets, cloud services, and web connectors | **Power BI Features:**   * Seamless integration with Microsoft products such as Excel, SharePoint, Teams, Power Automate, and Azure * Ability to leverage existing data sources and collaborate within the Microsoft ecosystem * Question-and-answer feature enabling interaction using natural language to get instant answers through visualizations * Access to reports and dashboards on the go through Power BI Mobile apps for iOS and Android devices * Collection of pre-trained machine learning models enhancing your data preparation efforts | **Looker Features:**   * Cloud-based tool * Powerful data modeling layer allowing users to define relationships between different datasets and create reusable metrics and dimensions * Data visualization integrations into applications or websites * Granular access controls, auditing capabilities, and integration with single sign-on providers, thus securing sensitive data and making it accessible only to authorized users |
| **Pros:**  • Preferred by cloud users • High scalability and customization capabilities • Built-in AI features to accelerate and improve blending data | **Pros:**   * Extensive library of pre-built visualizations and interactive elements * A vast repository of resources * Community support | **Pros:**   * Popular with Microsoft Office users * Robust sharing and collaboration capabilities with access control policies * Cheap licensing costs attracting small and medium businesses | **Pros:**   * Favored by companies that have invested in cloud infrastructure * Model customization and complex calculations through LookML, or Looker Modeling language * Capability to extend functionality as needed through comprehensive APIs * Detailed user-interaction history through granular access control and auditing capabilities |
| **Cons:**   * Lacks comprehensive dashboard features * Requires a steep learning curve to learn about all the features * Comes at a high price * Has additional paid features | **Cons:**   * Limited data preparation capabilities and thus relies on other tools * Needs a strong IT team due to its dependency on other tools * Expensive licensing options making it less accessible for small businesses * Steep learning curve for navigating the advanced features and complex data models | **Cons:**   * Limited customization and data modeling capabilities * Premium subscriptions for some advanced features | **Cons:**   * Need for SQL and database expertise * Long render times for large data sets * No budget-friendly pricing plan |

Reqirement Gathering Steps :

1. Identify stakeholders (internal/external) to understand requirements and identify resources.
2. Identify project objectives, KPIs and build a roadmap for the BI Project, list specific business objectives and define the BI’s scope.
3. Meet with stakeholders to understand the requirements and gather feedback.
4. Categorized them (from the stakeholders) into functional (data integration, ETL Steps, Data modelling, Database design) and non-functional requirements (qualitative attributes such as scalability, performance, security, reability, maintenance.
5. List assumptions, risks, and deliverables for the BI Project.
6. Obtain approval or sign off to meet stakeholder requirements, to avoid scope creep in the future.

**Primary FOCUS : DATA., data regulations regarding how data should be managed.**

* Ensuring data privacy, security, and compliance with regulations is essential to avoid such issues to legal / financial consequences.
* Data privacy is the practice of protecting personal information from unauthorized access, use and disclosure.
* Data security is the practice of protecting digital information, including personal information, from unauthorized access, tampering, theft throughout its entire life cycle.

Regulatory standards :

1. General Data Protection Regulation (GDPR)
2. Payment Card Industry Data Security Standard (PCI-DSS)
3. Health Insurance Portability and Accountability Act (HIPAA)
4. Family educational rights and Privay act (FERPA)

Preventing data privacy and security issues :

1. Identify and classify data (data sensitivity)
2. Identify the source of the data and where it goes (map the data sources),data will be used.
3. Understand who uses the data, so we can define roles and permissions for users.
4. Implement data privacy and security techniques ie. Encryption, authentication, authorization, anonymization(modifying personal data).

Setting a minimum access policy (User management), who has to approve access requests (contributor:edit access), GDPR compliance (needs to be corresponds is GDPR),

* Key components of BI include data sources and integration, warehousing, analytics tools, reporting and visual tools, OLAP, CPM, and real-time BI.
* The BI ecosystem or BI environment comprises four key elements: Data, people, processes, and technologies. While creating a business intelligence strategy, one should consider all four elements.
* BI tools can be categorized based on their capabilities into various categories:
  + unified modeling language (UML);
  + extract, transform, load (ETL) tools;
  + data warehousing tools;
  + data preparation and cleaning tools;
  + visualization and dashboarding tools;
  + reporting and querying tools;
  + mobile BI software tools;
  + cloud-based-SaaS BI software;
  + Spreadsheets.

1. Identifying & classifying data involves understanding the data sensitivity and the associated regulatory compliance (public data and company data)
2. Data users need skills to query the BI system to generate the required reports. Therefore, they need to be trained to use the BI system.
3. Offering a clear insight into the historical performance enables you to analyze past data and understand how the new POS system impacts sales performance in a quarter.
4. After analysis, you can use various BI tools for reporting and visualization tools to manipulate and present complex data into visual and understandable reports that help make critical business decisions. It involves creating dashboards, reports, and visualizations providing a comprehensive view of key performance indicators (KPIs) and trends.
5. Identifying relevant data sources and gathering them involves locating and understanding the various data repositories that will be used to extract information for analysis and reporting.

|  |  |
| --- | --- |
| **Data analysis** | The process that involves cleaning, transforming, and modeling data to uncover useful information to aid business decision-making. |
| **Data analytics** | The process that focuses on extracting valuable information from data using various tools, techniques, processes, and algorithms. It includes data analysis and the interpretation of the results, keeping in mind specific business objectives. |
| **Data encryption** | The process of translating data from a readable to an unreadable format, also known as ciphertext, to protect sensitive information during transmission or storage. |
| **Data governance** | The management of data in enterprise systems that involves ensuring availability, usability, integrity, and security, all of which are controlled by internal data standards and policies. |
| **Data integration** | A combination of technical and business processes that combine data from disparate sources into meaningful and valuable information. |
| **Data marts** | Data warehouses have data marts or subsets, which are like a partition to manage one specific business function, department, or subject area. Data marts make specific data available to a defined group of users, which allows those users to quickly access critical insights without wasting time searching through an entire data warehouse. |
| **Data mining** | The process of uncovering patterns and other valuable information from large data sets. |
| **Data modeling** | A visual representation of either a whole information system or parts of it to communicate connections between data points and structures. |
| **Data privacy** | Protection of personal data from those who should not have access to it and the ability of individuals to determine who can access their personal information. |
| **Data repositories** | Data sets identified to be mined for reporting and analysis. It is also known as a data archive or library. |
| **Data security** | The practice of protecting digital information, including personal information, from unauthorized access, tampering, or theft throughout its entire lifecycle. |
| **Data storytelling** | Skills that involve analyzing data to communicate insights and influence decisions. |
| **Data visualization** | Graphical representation of information and data. It helps data visualization to understand trends, outliers, and patterns in data. |
| **Data warehouse** | A storage architecture that pulls data from many sources into a single data repository for sophisticated analytics and decision support. |
| **Descriptive analysis** | The process of utilizing statistical techniques to explain or summarize a specific set of data. Descriptive analysis is also called descriptive statistics. |
| **Descriptive analytics** | A type of analytics that uses historical and current data to spot any trends and relationships between that data. |
| **Diagnostic analytics** | A type of analytics that helps identify the reason an event occurred. It lets you discover hidden correlations and connections between variables, determine causal relationships, detect anomalies, and isolate patterns. |
| **Predictive analytics** | A type of analytics used to predict future outcomes. It relies on historical data and employs various techniques such as statistical modeling, data mining, and machine learning for predictions. |
| **Prescriptive analytics** | A type of data analytics that recommends the optimal course of action to achieve a specific goal, drawing from inputs from descriptive, diagnostic, and predictive analytics processes. |

**Business Intelligence Process**

A diagram of steps and steps

Description automatically generated

* Data gathering involves identifying and collecting data from both internal and external sources.
* Data preparation helps to convert complex data into accurate and reliable data sets suitable for further processing and analysis.
* Data wrangling refers to a series of processes to explore, transform, and verify raw data sets from their unstructured and complex forms into high-quality data.
* Data integration merges data from various sources and organizes it within the data warehouse, database or data lakes, and data marts. Data may require cleaning and transformation using the extract, transform, and load (ETL) process.
* Typically, data gets stored in a centralized repository, such as a data warehouse or a data mart, to optimize for efficient querying and analysis.
* A data analyst uses four analytical techniques: Descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics.
* BI leverages various reporting and visualization tools to transform complex data sets and patterns into visually engaging and easily understandable reports for stakeholders. This process helps to get efficient data analysis, enabling stakeholders to make informed decisions.