A diagram of a triangle with points and a flag

Description automatically generated

how BI techniques and tools are applied in real-world business scenarios.

* Data mining helps identify patterns and establish correlations between parameters based on which companies can make decisions.
* Powerful dashboarding and visualization tools enable users to generate customized reports without the help of a data professional.
* BI helps performance management by providing insights into where they are doing well and where they need improvement.
* BI contributes to benchmarking by tracking certain metrics that need to be standardized eventually.
* BI processes boost sales by providing predictions and tracking the effect of marketing campaigns.
* Advanced statistical analysis helps in predictive modeling, eventually helping in decision-making, helps in predictive modeling based on multiple factors
* Data warehousing helps collect and store data from various sources in a centralized database..

A diagram of a step of a project

Description automatically generatedSteps to develop a BI Project

1. Planning

* Assessing the current system,
* Identifying the project goals
* Planning resources and timelines accordingly

1. Implementing BI Platform
2. Implementing data governance and monitoring the systems

BI Project PLANNING

1. **PLANNING,** starts with creating a project plan, that serves as a structured framework for documenting all essential aspects of the project.

* Project planning entails,
* assessing the data maturity, to understand your data quality and availability.
* outlining clear objectives and goals, define project objectives. And establish measurable KPIs
* Engaging stakeholders to determine requirements, and understand the purpose of the BI system (outcomes expected).
* Determining the budget ex. Software licences, hardware infrastructure, skilled personal based on the size and scope project.
* Putting a core team together to lead the project.
* Establishing a timeframe, set a realistic timeline based on project objectives and budget.

1. **IMPLEMENTING BI Platform**, it involves selecting, procuring, and setting up BI Software and infrastructure.

* Identifying your data sources, encompassing internal databases, external data sources and third party data providers. Data sources should be trustworthy and aligned with objectives.
* Connecting the sources, initiating data collection. This task involves data wrangling to connect disparate data sources and ETL Process.
* Developing a data model, deciding data storage requirements and database schema (ER model, relationship model, network model and graph model) based on the project’s data characteristics and suits your data requirements.
* Deciding on an analytics model, choice of types of analytics (descriptive, diagnostic, predictive, prescriptive) based on the business requirements.
* Visualizing data using reports and dashboards. Retrieve data through queries and create visual representations required for reporting purposes.Determine format which report or dashboards are required and determine the distribution method. Your dashboard consider the requirements of the various stakeholders.
* Testing and validating the results, use a sample data set outcomes to validate your outcomes before applying them to production data.

1. **GOVERNANCE and MONITORING,** defining and implementing policies around the BI infrastructure.

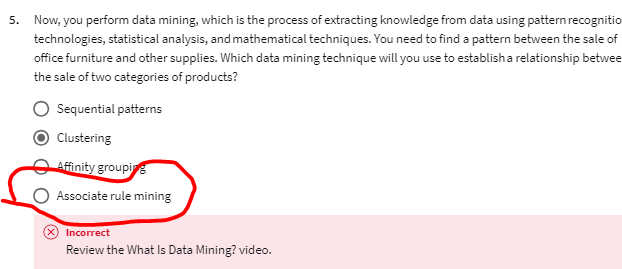
* These policies must be implemented before the project goes live or before users start working with production data.
* Data governance plan is a framework of policies, processes, and strategies that govern the management of data assets including collection, curation, storage and presentation for insights. This governance framework also ensures adherence to regulatory compliance,
* Monitor the performance of the BI platform and the governance system continuously. Address any performance issues that may arise in the BI platform. Automate routine process as needed.
* A company can use BI to set up a robust performance management system to track sales, compare with targets, and use the insights to set up marketing plans to focus on potential customers.
* The first step is assessing the data maturity of the current system to help you understand the data’s quality, availability, and structure. It helps bridge the gap between the current state and the required state.
* Monitoring the performance of the BI platform and the governance system continuously to identify anomalies and then correcting them in time ensures that the user experience remains largely unaffected.

1. APIs are popularly used for extracting data from a variety of data sources. APIs are invoked from applications that require the data and access an end-point containing the data.
2. Descriptive analytics is a type of analytics that uses historical and current data to spot any trends and relationships within that data.
3. The bar chart can effectively represent the trend over time.
4. The results section would contain charts and graphs that would substantiate the results and highlight the more complex or crucial findings.

A screenshot of a computer

Description automatically generatedA screenshot of a questionnaire

Description automatically generated

A screenshot of a computer

Description automatically generated

A close-up of a data

Description automatically generatedA screenshot of a computer

Description automatically generated

| **Term** | **Definition** |
| --- | --- |
| **Affinity grouping** | A technique used to discover co-occurrences in relationships. |
| **Amazon Web Services (AWS)** | A cloud-computing platform that offers economical online app development tools and services. |
| **Analytical strategies** | A long-term decision for an organization to use data to take actions to satisfy organizational vision and mission. It includes research and analysis for internal and external environments, evaluating data, and using strategic analysis tools. |
| **Analytics technologies** | Employs a variety of data-driven methodologies and techniques to address queries, identify patterns, forecast outcomes, and automate decision-making processes. |
| **Anonymization** | The process of modifying personal data so it is no longer possible to identify the individuals associated with it. |
| **Apache Airflow** | An open-source workflow management platform for data engineering pipelines. |
| **Apache Beam** | An open-source, unified programming model for batch and streaming data processing pipelines. |
| **Apache HBase** | A non-relational database that runs on Hadoop, providing real-time access to large data sets. |
| **Apache Kafka** | An open-source software platform used to handle real-time data feeds. |
| **Apache Spark Streaming** | An extension of the core Spark API that allows for fault-tolerant stream processing of live data streams with high throughput and scalability. |
| **Apache Storm** | A framework for distributed stream processing computation primarily written in the Clojure programming language. |
| **Application programming interface (API)** | APIs are predefined rules that allow software applications to communicate with each other, process data transfers, and enable the sharing of application data and functionality between companies and external parties. |
| **Association rule mining** | A technique that helps establish our relationship between two data events. |
| **Atomicity, Consistency, Isolation, and Durability (ACID) Compliance** | A group of characteristics that ensure dependable and uniform processing of transactions in a database system. |
| **Authentication** | The process of confirming the identity of a user or application seeking access to data. |
| **Authorization** | The process of determining whether an authenticated user has the right to perform an operation. |
| **AWS data pipeline** | Web service that helps you reliably process and move data between different AWS compute and storage services, as well as on-premises data sources, at specified intervals. |
| **Azure data factory** | Cloud-based ETL and data integration service that allows you to create data-driven workflows for orchestrating data movement and transforming data at scale. |
| **BeautifulSoup** | A Python library to get data out of HTML, XML, and other markup languages. |
| **Benchmarking** | A process to measure your business's success compared to similar organizations to identify any performance gaps or areas of improvement. |
| **Big data stores** | A larger, more complex data set, especially from new data sources. |
| **Big data** | A dynamic, large, and disparate volume of data being created by people, tools, and machines. |
| **Bokeh** | A Python library that provides interactive charts and plots and is known for delivering high-performance interactivity over large or streaming datasets. Bokeh offers flexibility for applying interaction, layouts, and different styling options to visualization. |
| **Business analyst** | Data professionals who bridge the gap between business objectives and technical solutions. They identify and document business requirements, conduct feasibility studies, and propose improvements to business processes. |
| **Business intelligence** | A combination of various technologies, tools, and methodologies that gather, analyze, and transform data into meaningful information. |
| **Business intelligence analyst** | Professional analyzes, models, and visualizes data to help businesses make informed decisions. |
| **Business intelligence architect** | Professional who designs and implements solutions that help organizations make decisions. |
| **Business intelligence developer** | Engineer who is an in charge of developing, deploying, and maintaining BI interfaces. |
| **Business intelligence initiative** | Organizations can improve their revenue, operational efficiency, and competitive edge with better business decisions. |
| **Business intelligence manager** | Professional who uses data analysis to suggest and execute decisions for a company. |
| **Business intelligence platform** | A cluster of tools and applications that enables organizations to analyze data and get the insights to make decisions. |
| **Business intelligence strategies** | Blueprint to use the data that a company generates. |
| **Business intelligence tools** | Proprietary or open-source application software that is used to collect, process, analyze, sort, filter, and report large data from the systems and transform raw data into useful information. |
| **Business intelligence** | A combination of various technologies, tools, and methodologies that gather, analyze, and transform data into meaningful information. |
| **Causal relationship** | A type of relationship between two events where one event influences the occurrence of another event. |
| **Central tendency** | The statistical measure that represents the central value of the entire distribution or a data set. |
| **Centralized database** | A database that is situated in and maintained from a single location. |
| **Classification** | A technique that classifies attributes into target categories. |
| **Cloudant** | A fully managed, distributed database optimized for heavy workloads and fast-growing web and mobile apps. |
| **Clustering** | A technique of grouping data into clusters so they can be treated as groups. |
| **Comma-separated values (CSV)** | A text-formatted file uses commas to separate the values. |
| **Conceptual data model** | The model, created by business stakeholders and data architects, defines the system's scope, concepts, and rules. |
| **Confidence intervals** | Methodology of inferential statistics that incorporates uncertainty and sample error to create a range of values, the actual population value is likely to fall within. |
| **Corporate performance management (CPM)** | Software that integrates information from various sources to support operational plans and align key performance indicators or KPIs to improve financial planning. |
| **CouchDB** | An open-source NoSQL document database that collects and stores data in JSON-based document formats. Unlike relational databases, CouchDB uses a schema-free data model, which simplifies record management across various computing devices, mobile phones, and web browsers. |
| **Customer relationship management (CRM) software** | Software that helps companies measure and control their lead generation and sales pipelines. |
| **Dash** | A Python framework for creating interactive web-based visualizations. Dash can be used to build highly interactive web applications using Python code. |
| **Dashboards** | Tool to present a bird's-eye view of the complete picture while also allowing you to drill down into the next level of information for each parameter. These are easy to comprehend by an average user, make collaboration easy between teams, and allow you to generate reports on the go. |
| **Data abstraction** | The process of simplifying a set of data to represent the whole. |
| **Data analysis** | Process that involves cleaning, transforming, and modeling data to uncover useful information to aid business decision-making. |
| **Data analyst** | A data professional who first gathers and understands the data, then analyzes and interprets it before visualizing it and, finally, weaving it into a story. |
| **Data analytics** | 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 cleaning** | An essential process of preparing and validating data before performing analysis. It is an integral part of building a data model. |
| **Data collection** | The collection and evaluation of information from multiple sources to find answers, assess outcomes, and forecast trends. |
| **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 engineer** | Responsible for building robust data architecture and easily accessible data warehouses. They also design, construct, and evaluate scalable big data environments to support businesses with the key objective of providing stable and highly optimized data pipelines and systems. |
| **Data fabric** | An architecture that facilitates the end-to-end integration of various data pipelines and cloud environments through intelligent and automated systems. |
| **Data governance analyst** | Data professionals who help a business ensure that its data is accurate, consistent, and compliant with legal and regulatory requirements. They also plan and execute security measures to protect and preserve computer databases. |
| **Data governance policy** | A document that outlines an organization's data management policies, roles, responsibilities, data quality, security, and access. |
| **Data governance** | The process of managing data quality to ensure that data is available, accurate, consistent, usable for the desired purpose, and secure. |
| **Data integration** | Combination of technical and business processes that combine data from disparate sources into meaningful and valuable information. |
| **Data lakes** | A centralized repository designed to store, process, and secure large amounts of structured, semi-structured, and unstructured data. It can store data in its native format and process a variety of data while ignoring size limits. |
| **Data lookup** | A way to fill in information based on rules. |
| **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 maturity** | The level at which an organization uses data in its day-to-day operations. |
| **Data mining** | Process of uncovering patterns and other valuable information from large data sets. |
| **Data modeling** | Process of creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures. |
| **Data modeling tool** | Software applications that help you to design, create, and manage data models. |
| **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 profiling** | The process used to analyze data in a data warehouse for structure, content, relationships, and derivation rules. Profiling identifies anomalies, assesses data quality, and discovers, registers, and assesses metadata. |
| **Data refinery** | A tool that helps to explore data residing in a spectrum of data sources. |
| **Data reporting** | The gathering of raw data from various sources and transforming it into meaningful, easy-to-understand information to gain valuable insights into business performance. |
| **Data repository** | Also termed a data archive or library, it refers to a data set identified to be mined for reporting and analysis. |
| **Data science** | Focuses on understanding the data. This involves data analysis, beginning with data loading, exploring, and cleaning. It creatively explores data, coming up with new solutions and inventions. |
| **Data scientist** | Data professionals who develop algorithms, build predictive models, and uncover patterns and trends in large data sets. They apply statistical analysis, especially inferential statistics, machine learning, and predictive modeling, to extract insights from data and make predictions. |
| **Data security** | The practice of protecting digital information, including personal information, from unauthorized access, tampering, or theft throughout its entire lifecycle. |
| **Data source** | The physical or digital location where the data is held in a data table, object, or other storage format. |
| **Data storage** | The process of saving digital information using computer devices facilitates the efficient completion of various digital tasks. |
| **Data streams** | The process of transmitting continuous data and feeding it into stream processing software to derive valuable insights. |
| **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. |
| **Data warehousing** | The process of pulling together data from many different sources into a single data repository for sophisticated analytics and decision support. |
| **Data wrangling** | An iterative process that involves data exploration, transformation, and validation, making it available for a credible and meaningful analysis. |
| **Database as a service** | A cloud-computing service that allows users to access and use a cloud database system without purchasing and setting up their own hardware, installing their own database software, or managing the database themselves. |
| **Database Management System (DBMS)** | Software to store and retrieve users' data by considering the security of their information. |
| **Decision trees** | A tree structure with multiple branches, each representing a probable occurrence, helps build classification models. This technique helps to build a clear understanding of the relationship between input and output. |
| **Denodo** | A unified virtual data layer that allows enterprise users to access data across formats, protocols, and locations using techniques like search. |
| **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. |
| **Descriptive statistics** | Statistics to summarize information about the sample to present data in a meaningful way, allowing simpler interpretation of the 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. |
| **Dispersion** | The measure of variability in a data set. |
| **DocumentDB** | A NoSQL database service that supports document data structures with some MongoDB 3.6 and 4.0 compatibility. |
| **DynamoDB** | A type of database developed by Amazon Web Services (AWS). |
| **E-commerce** | Platform for buying and selling products, services, transmitting funds and data over the internet. |
| **Enterprise resource planning (ERP) systems** | A type of software system that enables businesses to automate and efficiently manage their key business processes to gain optimal performance. |
| **Entity-relationship model (E-R model)** | A high-level data model is created to define the data elements and their relationships for a specific system. It develops a conceptual design for the database and presents a simple and easy-to-design data view. |
| **Extract, load, and transport (ETL) process** | A process that extracts, loads, and transforms data from multiple sources to a data warehouse or other unified data repository. |
| **Flat files** | Collection of data that is stored specifically in a two-dimensional database. It usually contains a series of records (or lines), where each record is usually a sequence of fields. |
| **Global Positioning Systems (GPS)** | A radio navigation system that accurately determines location, time, and velocity regardless of weather conditions. |
| **Google DataPrep** | An intelligent cloud data service allows visually exploring, cleaning, and preparing both structured and unstructured data for analysis. |
| **Hadoop Distributed File System (HDFS)** | A storage system for big data that runs on multiple commodity hardware devices connected through a network. HDFS provides scalable and reliable big data storage by partitioning files over multiple nodes. |
| **Hadoop** | A collection of tools that provides distributed storage and processing of big data. |
| **Hierarchical model** | A data model in which the data are organized into a tree-like structure. |
| **Hive** | A data warehouse for data query and analysis built on top of Hadoop. |
| **Hypothesis testing** | Methodology of inferential statistics that uses data from a sample to test a hypothesis. |
| **IBM Cognos Analytics** | An end-to-end analytics solution known for its superior visualizations and overlaying data on the physical world using its geospatial capabilities. |
| **IBM InfoSphere DataStage** | ETL tool and part of the IBM Information Platforms Solutions suite and IBM InfoSphere. It uses a graphical notation to construct data integration solutions. |
| **Inferential statistics** | Statistics to make inferences or generalizations about the broader population. |
| **Java** | A programming language known for its platform independence, which allows Java programs to run on different operating systems without modification. |
| **JavaScript** | High-level scripting or programming language that empowers the implementation of complex and sophisticated functionalities on web pages. |
| **JavaScript object notation (JSON)** | An open standard file format that uses readable text to store and transmit data objects consisting of attributes. |
| **Jupyter Notebook** | An open-source web application widely used for data cleaning and transformation, statistical modeling, and data visualization. |
| **Key performance indicators (KPI)** | A performance measure for a specific objective that provides targets, milestones, and insights to help teams and individuals make better decisions. |
| **Linux** | An open-source operating system developed from Unix. |
| **Loatame** | A data management platform. |
| **Logical data model** | Provides detailed descriptions of data elements and is utilized to create visual representations of data entities, attributes, keys, and relationships. |
| **Looker** | A business intelligence app and big data analytics platform that explore, analyze, and share real-time business analytics. |
| **Machine learning engineers** | Typically, programmers who construct the algorithms, systems, models, and frameworks that enable machines to learn and perform functions independently and effectively. |
| **Machine learning** | The branch of artificial intelligence (AI) and computer science focuses on using data and algorithms to imitate how humans learn, gradually improving accuracy. It helps systems learn and improve at forecasting, much like how people learn from experience. |
| **Matplotlib** | A widely used Python data visualization library. It provides different kinds of 2D and 3D plots and the flexibility to create plots in several different ways. |
| **Mean** | A measure of central tendency. The arithmetic average is the sum divided by the number of cases. |
| **Median** | The middle value of a data set arranged in ascending order of values. |
| **Metadata management** | Best-practice processes and technologies that manage the metadata of your data. Metadata management helps provide insights for more effective data management. |
| **Metrics** | Another measure to track the performance of specific business processes. They help provide context to the performance of key business goals but are not critical to its success like key performance indicators (KPIs). |
| **Mode** | The most frequently occurring value. If several values share the greatest occurrence frequency, each is a mode. |
| **MongoDB** | An open source, nonrelational database management system (DBMS) that uses flexible documents instead of tables and rows to process and store various forms of data. |
| **MySQL** | An open-source relational database management system (RDBMS). |
| **Network model** | A database model conceived as a flexible way of representing objects and their relationships. |
| **NoSQL database** | A non-tabular database that stores data with different data storage tables than relational tables. |
| **NumPy** | A tool for mathematical computing and data preparation in Python. It offers a host of built-in functions and capabilities for data mining. |
| **Online analytical processing (OLAP)** | A software that is used to conduct multidimensional analysis on large volumes of data from a data warehouse, data mart, or other centralized data store. |
| **Online Transaction Processing (OLTP)** | A computerized system that allows real-time data processing and immediate response to users' queries. |
| **OpenRefine** | An open-source tool that allows you to import and export data in various formats, such as TSV, CSV, XLS, XML, and JSON. Using OpenRefine, you can clean data, transform it from one format to another, and extend data with web services and external data. |
| **Oracle Cloud** | A cloud platform that offers complete cloud application suites across software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). |
| **Oracle database** | A multi-model database management system generally used for online transaction processing (OLTP), data warehousing, and both workloads. |
| **Outlier detection** | A technique that helps find patterns and data that are not normal or unexpected. |
| **Pandas** | A Python library used to work with data sets. |
| **Pattern recognition** | The discovery of regularities or commonalities in data |
| **Performance management** | The process of tracking and evaluating employee performance against pre-determined objectives. |
| **Physical data model** | A database-specific model that represents relational data objects (for example, tables, columns, primary and foreign keys) and their relationships. |
| **Platform as a service** | A cloud computing model that provides customers a complete cloud platform, hardware, software, and infrastructure, for developing, running, and managing applications without the cost, complexity, and inflexibility that often come with building and maintaining that platform on-premises. |
| **Population** | A discrete group of people or things that can be identified by at least one common characteristic for purposes of data collection and analysis. |
| **PostgreSQL** | An open-source database that has a strong reputation for its reliability, flexibility, and support of open technical standards. |
| **Power BI** | A cloud-based business analytics service from Microsoft that enables you to create reports and dashboards. |
| **PowerShell** | A cross-platform command-line shell and scripting language designed for automating tasks and managing configurations. |
| **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. |
| **Predictive modeling** | A statistical method of using past and current data to predict future trends and outcomes. |
| **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. |
| **Probability theory** | The mathematical framework that can help you find the likelihood of specific outcomes. |
| **Python** | An agile, dynamically typed, expressive, open-source programming language that supports multiple programming philosophies, including procedural, object-oriented, and functional. Python is a popular high-level programming language that is easily extensible through the use of third-party packages and often allows powerful functions to be written with few lines of code. |
| **R Studio** | A popularly used open-source integrated development environment (IDE) for working with the R programming language. |
| **R** | One of the most widely used languages for performing statistical modeling and computations by statisticians and data miners. |
| **Radio Frequency Identification (RFID) tags** | A method for tracking goods through their tags. |
| **Range** | The distance between the smallest and largest values in data sets. |
| **Real-time BI** | Processes and tools that provide up-to-the-minute information and analysis, which enables businesses to monitor KPIs in real-time. |
| **Regression analysis** | Statistical methods can help find relationships between variables, such as how changes in independent variables may affect a dependent variable. |
| **Regression** | A technique that helps identify the nature of the relationship between two variables, which could be causal or correlational. |
| **Relational Database Service (RDS)** | Organizes data into rows and columns, which collectively form a table. Data is typically structured across multiple tables, which can be joined together via a primary key or a foreign key. |
| **Relational database** | Structured data with a well-defined schema. |
| **Relational model** | An approach to managing data using a structure and language consistent with first-order predicate logic. |
| **Return on investment (ROI)** | A performance measure that is used to evaluate the efficiency of an investment. |
| **Sample** | A representative selection drawn from the total population. |
| **SAS Enterprise Miner** | A comprehensive, graphical workbench for data mining. It provides powerful capabilities for interactive data exploration, which enables users to identify relationships within data. |
| **Scala** | A programming language designed for concise, elegant, and type-safe expression of programming patterns. This language seamlessly integrates object-oriented and functional features. |
| **Scrapy** | A free and open-source web-crawling framework written in Python and developed in Cambuslang. |
| **Selenium** | A testing platform for an open-source web user interface. |
| **Self-service business intelligence (SSBI)** | The process, tools, or software that allows business users to ask questions about their data, get insights, and analyze data without relying on IT, BI specialists, or SQL to take business actions. |
| **Semi-structured data** | Data that has some organizational properties but not a rigid schema, such as data from emails, XML, zipped files, binary executables, and TCP/IP protocols. |
| **Sequential patterns** | A technique that traces a series of events that take place in a sequence. |
| **Shiny** | An R package that helps build interactive web apps that you can host as standalone apps on a webpage. |
| **Skewness** | The measure of whether the distribution of values is symmetrical around a central value or skewed left or right. |
| **Snowflake schema** | A multi-dimensional data model, an expansion of a star schema. It breaks down dimension tables into subdimensions. |
| **Spark** | A distributed data analytics framework designed to perform complex data analytics in real-time. |
| **Spreadsheets** | Applications such as Microsoft Excel and Google Sheets that have a host of features and built-in formulae that can help you identify issues and clean and transform data. |
| **SQL database** | A collection of highly structured tables where each row represents a data entity and every column represents a specific information field. |
| **SQL server integration services (SSIS)** | Enterprise data integration, data transformation, and data migration tool built into Microsoft's SQL Server database. It can be used for a variety of integration-related tasks, such as analyzing and cleansing data and running extract, transform, and load (ETL) processes to update data warehouses. |
| **SQL** | Computer language used to interact with a relational database. |
| **Standard deviation** | A measure of how tightly the data is clustered around the mean. |
| **Statistical analysis system (SAS)** | A programming language that provides all the tools necessary to read, write, and create system files, SAS databases, and reports. |
| **Statistical analysis** | The process of collecting large volumes of data and then using statistics and other data analysis techniques to identify trends, patterns, and insights. |
| **Statistical methods** | Methods useful to ensure that data is interpreted correctly and apparent relationships are meaningful and not just happening by chance. |
| **Statistical modeling** | A method that uses mathematical equations to analyze data and identify patterns. |
| **Statistical Package for Social Sciences (SPSS)** | Software used for advanced analytics, text analytics, trend analysis, validation of assumptions, and translation of business problems into data science solutions. |
| **Statistical tool** | A tool that converts, analyzes, interprets, and uses data in different forms and purposes. |
| **Statistics** | Branch of mathematics dealing with the collection, analysis, interpretation, and presentation of numerical or quantitative data. |
| **Structured data** | The data that conforms to a defined structure follows a consistent order and is easily accessible to people or computer programs. |
| **Structured query language (SQL)** | Computer language, which is used to interact with a relational database. |
| **Tableau** | A tool to analyze and report large amounts of data through visual representation. |
| **Tab-separated values (TSV)** | A text-based file format that stores data. |
| **Talend** | ETL tool for Data integration provides software solutions for data preparation, quality, integration, application integration, management, and big data. Talend has a separate product for all these solutions. |
| **Talend Open Studio** | A free, open-source ETL tool for data integration and big data. |
| **Tidyverse** | A collection of R packages that help you really easily load data, aggregate it at different levels, and also quickly and easily visualize it. |
| **Time-series analysis** | A way of analyzing data that can help you uncover patterns, trends, and anomalies in time-based data. |
| **Trend** | The general tendency of a set of data to change over time. |
| **Trifacta wrangler** | An interactive cloud-based service for cleaning and transforming data. It takes messy, real-world data and cleans and rearranges it into data tables, which can then be exported to Excel, Tableau, and R. |
| **Unix** | A group of multitasking, multiuser computer operating system. |
| **Unstructured data** | Typically categorized as qualitative data that cannot be processed and analyzed via conventional data tools and methods. |
| **Variance** | Defines how far away the data points fall from the center, that is, the distribution of values. |
| **Velocity** | A tool to provide insights to the business about how well software delivery works and where to focus new processes, resources, or more automation. |
| **Veracity** | The term "Veracity" was coined by IBM to describe the challenges of managing data from disparate sources, which can be inconsistent and unreliable. |
| **Watson Studio refinery** | Available via IBM Watson Studio, it allows for discovering, cleaning, and transforming data with built-in operations. It transforms large amounts of raw data into consumable, quality information that's ready for analytics. |
| **Web scraping** | A technique used to collect online content and data generally gets saved in a local file so as to manipulate and analyze as needed. |

A screenshot of a computer

Description automatically generatedA screenshot of a survey

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

The Business Intelligence (BI) Essentials course is a part of the IBM BI Analyst Professional Certification (PC). As a next step, you are encouraged to explore other courses in this PC:

* [Excel Basics for Data Analysis](https://www.coursera.org/learn/excel-basics-data-analysis-ibm) : <https://www.coursera.org/learn/excel-basics-data-analysis-ibm>

(This course is part of multiple programs.)

* [Data Visualization and Dashboards with Excel, Cognos, and Looker](https://www.coursera.org/learn/data-visualization-dashboards-excel-cognos) : <https://www.coursera.org/learn/data-visualization-dashboards-excel-cognos> (This course is part of multiple programs.)
* [Statistical Analysis Fundamentals using Excel](https://www.coursera.org/learn/statistics-fundamentals-using-excel) : <https://www.coursera.org/learn/statistics-fundamentals-using-excel>

(This course is part of [**IBM Business Intelligence (BI) Analyst Professional Certificate**](https://www.coursera.org/professional-certificates/bi-analyst))

* [Introduction to Relational Databases (RDBMS)](https://www.coursera.org/learn/introduction-to-relational-databases) : <https://www.coursera.org/learn/introduction-to-relational-databases>

(This course is part of multiple programs.)

* [SQL: A Practical Introduction for Querying Databases](https://www.coursera.org/learn/sql-practical-introduction-for-querying-databases) : <https://www.coursera.org/learn/sql-practical-introduction-for-querying-databases> (This course is part of multiple programs.)
* [Getting Started with Data Warehousing and BI Analytics](https://www.coursera.org/learn/getting-started-with-data-warehousing-and-bi-analytics) : <https://www.coursera.org/learn/data-warehouse-fundamentals>

(This course is part of multiple programs.)

* [Getting started with Tableau](https://www.coursera.org/learn/getting-started-with-tableau) : <https://www.coursera.org/learn/getting-started-with-tableau>

(This course is part of multiple programs.)