Module-1 What is Data Analytics

In this module, you will learn about the different types of data analysis and the key steps in a data analysis process. You will gain an understanding of the different components of a modern data ecosystem, and the role Data Engineers, Data Analysts, Data Scientists, Business Analysts, and Business Intelligence Analysts play in this ecosystem. You will also learn about the role, responsibilities, and skillsets required to be a Data Analyst, and what a typical day in the life of a Data Analyst looks like.

Learning Objectives

- Explain the different components of a modern data ecosystem.
- Describe and differentiate between the role different data professionals play in a modern data ecosystem.
- Explain what data analysis is, the different types of data analysis, and the key steps in the data analysis process.
- Describe the responsibilities and skillsets of a Data Analyst.
- Summarize a typical day in the life of a Data Analyst.

Welcome

Course Introduction

Introduction to Data Analytics Course

- This is the **first course** in a series aimed at preparing learners for a career as a **Junior Data Analyst**.
- The course sets the foundation for:
 - Understanding data analytics.
 - Exploring career paths within the data domain.
 - Acquiring practical skills in analytics tools and methodologies.

Importance of Data in Business

- Quoting a **Forrester consulting report**:
 - Businesses now recognize the **untapped value in data**.
 - Data analytics is seen as crucial for competitiveness.

- Companies are:
 - **Hiring** new talent.
 - **Upskilling** current employees.
 - Creating **centers of excellence** to foster robust data practices.

Job Market & Career Opportunities

- There's a significant **supply-demand mismatch** for skilled data analysts.
- Data analytics is:
 - Highly sought-after
 - Well-paid
- Potential career growth into:
 - Data Science
 - Data Engineering
 - Business Analytics
 - Business Intelligence Analytics

6 Who Is This Course For?

- Fresh graduates from any field.
- **Working professionals** exploring a career transition.
- Data-driven decision makers in business roles.
- Anyone in **analytics-enabled roles** seeking to upgrade skills.

📤 What You Will Learn

- Core Concepts:
 - The data ecosystem
 - o Basics of data analysis
- Fundamental Processes:

- Data gathering
- Data wrangling
- Data mining
- Data analysis
- o Data visualization

Real-World Perspective:

- A glimpse into "A day in the life" of a data analyst.
- o **Insights** from practicing data analysts:
 - How they entered the field.
 - Career advice and learning paths.
 - Employer expectations.
 - Best practices for analysis.

🚀 Looking Ahead

- The field of data analytics is **rapidly evolving** and **full of opportunities**.
- Learners are encouraged and congratulated for embarking on this exciting journey.

* Summary

This introductory course on Data Analytics provides foundational knowledge and insight into becoming a junior data analyst. It highlights the growing importance of data in driving business decisions and the rising demand for analytics professionals. The course is suitable for fresh graduates, career changers, and professionals in analytics-enabled roles. Learners will gain exposure to essential analytics concepts, tools, and career insights from industry experts, making it a stepping stone into various high-growth data professions.

Key Takeaways:

- Data analytics is crucial for business success.
- There's a talent gap, leading to high demand and salaries.
- The course offers a strong foundation for several data-related career paths.
- Real-world advice from experienced analysts enhances learning.

Table: What We Learnt in the Video

Topic/Section	What We Learnt	
Introduction to the Course	First in a series, aims to build foundational skills for junior data analysts.	
Business Perspective	Data is seen as a key competitive asset; businesses are investing in analytics.	
Job Market Trends	High demand, low supply of data analysts = great career potential.	
Target Audience	Ideal for fresh grads, mid-career professionals, and decision-makers.	
Course Coverage	Covers data collection, wrangling, mining, analysis, and visualization.	
Career Insights	Includes real-life examples, best practices, and employer expectations.	
Future Outlook	Promising and expanding field with varied roles and learning paths.	

Before you start

Practically Speaking: From Messy Data to Meaningful Insights

♦ Introduction to Practically Speaking

• Purpose of the Series:

- o Connects learning to real-world applications
- Focuses on how data analytics is used across industries

Key Idea:

 Much of the value in data analytics comes from behind-the-scenes work before any insights or charts are produced

✓ Data Preparation: The Unsung Hero

• Time Spent Cleaning Data:

 $\circ~$ Data analysts spend 60--80% of their time on data cleaning and preparation

• Why It Matters:

- Essential for transforming raw data into **trustworthy insights**
- Foundational step that enables meaningful analysis

The IBM Course and Its Relevance

- Practical Learning Focus:
 - Not just theory helps learners understand the **modern data ecosystem**
- Skill Development:
 - o Equips learners to handle and interpret massive volumes of data
 - o Builds competence for real-world decision-making
- Real-World Applications of Data Analytics
- Finance
 - Use of Alternative Data:
 - o Data from sources like social media sentiment and satellite imagery
 - Supports complex investment decisions
- Healthcare
 - Predictive Analytics:
 - Data used to predict patient outcomes
 - Enables more personalized treatment and better healthcare results

Retail

• Scenario: Stagnant Sales

- o Data analysts examine sales by region and product
- May discover regional demand-supply mismatches

Outcome:

- Create visualizations (charts showing demand vs stock)
- Business adjusts inventory → **Sales improve**

Becoming a Data Detective

Mindset Shift:

From passive student to active data detective

• Core Role of a Data Analyst:

- o Uncover hidden patterns and tell stories with data
- Communicate insights that drive decisions

🧸 Data as Ingredients, Analyst as the Chef

• Analogy:

- Raw data = ingredients (flour, sugar, eggs)
- Analyst = chef using tools and techniques to turn it into something useful (like a delicious meal)

Practical Tips for Beginners

1. Clarity is Key

• Focus on the **main message** of your analysis

		0	Keep it simple and impactful
2.	Thi	ink	About Action
		0	Always ask: "So what?"
		0	Move beyond description to actionable insights
3.	Em	bra	ice the Mess
		0	Imperfect data is the norm
		0	Cleaning, fixing, exploring data is a valuable skill
Aı	oply	ing	Learning from Module One
⊚ Si	igge	stec	d Exercise:
•	Thi	nk (of a real-life decision you care about
•	Ask	k: W	hat data could help you make that decision better?
😉 Ed	cosys	sten	n Understanding:
•	Ref	lect	on roles and technologies introduced in the module
•	Ima	agin	e how a data analyst interacts with:
		0	Teams (engineers, business leaders, etc.)
		0	Tools (software, databases)
		0	Challenges (communication, clarity, messy data)

★ Summary

In this introductory episode of *Practically Speaking*, learners are welcomed to the world of data analytics with a focus on real-world relevance. The episode emphasizes how much of a data analyst's job happens behind the scenes — especially cleaning and preparing data — which is essential for building trust in insights. It introduces how data analytics is used in finance, healthcare, and retail, and encourages a mindset shift toward becoming a "data detective." Key beginner tips include focusing on clarity, actionable insights, and embracing messy data. Learners are invited to start thinking practically about the decisions data can help improve and how data analysts function within the modern data ecosystem.

Table: What We Learnt in the Video

Topic/Section	What We Learnt	
Data Preparation	Takes 60-80% of analysts' time; essential for reliable insights	
IBM Course Value	Teaches practical data skills, not just theory	
Finance Example	Uses social media and satellite data for investment decisions	
Healthcare Example	Predicts patient outcomes to personalize care	
Retail Example	Data reveals regional inventory issues → boosts sales	
Analyst's Role	Like a detective or chef — combining messy data into useful outcomes	
Beginner Tips	Focus on clarity, actionable insights, and embrace imperfect data	
Learning Application	Start with a real decision you care about and imagine how data could inform it	
Ecosystem Reflection	Consider how analysts interact with people, roles, and tech in organizations	

Modern Data Ecosystem and the Role of Data Analytics

Modern Data Ecosystem

The Modern Data Ecosystem

• Data Explosion:

- Driven by constant increases in data processing speed and bandwidth.
- Emergence of new tools for creating, sharing, and consuming data.
- Ever-growing community of data creators and consumers.
- o "Data begets more data": A virtuous cycle of continuous growth.

• Interconnected Entities:

- The data ecosystem is a network of independent yet interconnected components.
- o Continually evolving due to technological advancements and user needs.

• Ecosystem Components:

- Data Sources
- Data Integration
- o Data Analysis and Insight Generation
- Stakeholder Collaboration
- Storage and Infrastructure

🖢 Data Sources

• Types of Data:

 Structured and unstructured data formats. 			
o Examples:			
■ Text, images, videos			
■ Clickstreams, user conversations			
■ Social media, IoT devices			
■ Real-time event data streams			
 Legacy databases 			
 Data from professional providers/agencies 			
Key Characteristics:			
Diversity and dynamic nature are at an all-time high.			
 Data Acquisition Initial Step: Pull a copy of data from original sources into a data repository. Tasks Involved: Handling various formats, sources, and interfaces. Ensuring reliability, security, and data integrity. 			
✓ Data Organization & Governance● Data Processing Tasks:			

- Organizing, cleaning, and optimizing data.
 Ensuring easy and secure access for end users.
- Compliance and Standards:
 - o Conform to legal and organizational guidelines:
 - Personal data regulations (e.g., health, biometric, household data via IoT)
 - Standardization using master data tables
- Challenges:
 - Data management
 - o Ensuring high availability, flexibility, accessibility, and security of data repositories

🦍 Data Access and Utilization

- Stakeholders:
 - o Business users, analysts, data scientists, developers
- Data Needs:
 - Analysts → raw data
 - Business users → dashboards, reports
 - \circ Applications \rightarrow custom APIs
- Challenges:
 - o Interface design, API development, meeting diverse user needs

🚀 Emerging Technologies in Data Ecosystem

- Key Technologies:
 - Cloud Computing:
 - Offers virtually limitless storage and compute power
 - Enables access to latest tools, open-source tech, and ML frameworks
 - Machine Learning:
 - Used to create predictive models
 - Requires training algorithms on historical (big) data
 - o Big Data:
 - Refers to vast and varied datasets
 - Demands new tools and analytical techniques beyond traditional methods

Summary

The video explores the ever-expanding modern data ecosystem shaped by rapid technological growth and diverse data sources. It highlights the continuous cycle of data generation and consumption, the need for robust integration and management, and the importance of compliance and stakeholder-centric delivery. It further emphasizes the transformational impact of cloud computing, machine learning, and big data in generating insights and shaping business decisions.

Key Takeaways:

- Data is growing faster than ever due to tech advancements and user activity.
- Ø Modern ecosystems are complex and interconnected.
- II Data must be acquired, cleaned, governed, and tailored for specific users.

• Cloud and AI/ML tools are revolutionizing how we store and analyze data.

Table: What We Learnt in the Video

Topic/Section	What We Learnt	
Modern Data Ecosystem	Composed of interconnected, evolving components with increasing data complexity	
Data Sources	Come from diverse, dynamic origins (e.g., IoT, social media, legacy systems)	
Data Acquisition	Initial step of pulling data into repositories; focuses on reliability and integrity	
Data Organization & Governance	Involves cleaning, optimizing, and ensuring compliance with standards and policies	
Data Utilization	Different users need data in different formats (raw, reports, APIs)	
Emerging Technologies	Cloud, machine learning, and big data are transforming data storage and analysis	

Key Players in the Data Ecosystem

Importance of Data in Business Today

- Organizations leveraging data to uncover opportunities are leading the future.
- Data is used to:
 - Detect fraud via pattern recognition in financial transactions.
 - o Drive conversions with recommendation engines.
 - o Analyze social media for customer sentiment.
 - Personalize offerings through behavioural analysis.
- Competitive advantage is driven by data-driven insights.
- Achieving this requires diverse roles and skill sets in data.

🧘 Role of a Data Engineer

• Core Responsibilities:

- Develop and maintain data architectures.
- Extract, integrate, and organize data from multiple sources.
- Clean, transform, and prepare data.
- Design and manage data repositories.
- Ensure data is accessible and usable by analysts and scientists.

• Skills Required:

- o Programming knowledge.
- Understanding of technology architectures.
- Deep knowledge of relational and non-relational databases.

Role of a Data Analyst

• Core Responsibilities:

- Translate complex data into plain language.
- Clean and inspect data.
- Identify correlations and patterns.
- Apply statistical methods for analysis.
- Visualize and present data insights (dashboards, charts).

• Typical Questions Answered:

- Are search experiences good on our site?
- What is public sentiment on our rebranding?
- Is there a link between sales of different products?

• Skills Required:

Spreadsheet tools and query writing.

- Statistical tools for visualizations.
- Programming basics.
- Strong analytical and storytelling abilities.

Role of a Data Scientist

• Core Responsibilities:

- Analyze data for insights and build predictive models.
- Use machine learning and deep learning.
- Train models on historical data to predict future trends.

• Typical Questions Answered:

- How many social media followers will we gain?
- What is the customer churn rate forecast?
- Is this financial transaction anomalous?

• Skills Required:

- Math, statistics, and domain expertise.
- Proficiency in programming and database handling.
- Experience in model building.

Role of Business Analysts & BI Analysts

Business Analysts:

- Use insights from data analysts/scientists.
- Focus on implications and decision-making for business growth.

• BI (Business Intelligence) Analysts:

- Analyze market and external forces.
- Organize and monitor functional data.

• Provide insights that improve business performance.

• Focus Areas:

- Internal vs. external data monitoring.
- Actionable decision-making for strategy and performance.

Career Flexibility in Data Roles

- Professionals often start in one data role and transition to others.
- This is possible by **upskilling and learning new tools and techniques**.

* Summary

This video explores how modern businesses rely on data for strategic decisions and outlines the key roles that turn raw data into valuable insights. It introduces the essential responsibilities and skills of **data engineers**, **data analysts**, **data scientists**, **business analysts**, and **BI analysts**, emphasizing their interdependent roles within the data ecosystem. The takeaway is that data professionals transform information into actionable intelligence, fueling business growth and innovation.

Key Takeaways:

- Data is central to business success.
- Each data role has distinct functions but overlaps in skills.
- Career transitions are common with continuous learning.
- Businesses benefit when insights lead to action.

Table: What We Learnt in the Video

Topic/Section	What We Learnt	
Importance of Data	Data enables fraud detection, personalization, and customer insight.	
Data Engineer	Builds the infrastructure and makes data usable.	
Data Analyst	Finds patterns and insights; translates numbers into business decisions.	
Data Scientist	Predicts future trends using ML and data modeling.	

Business Analyst	Applies insights to business strategy and decision-making.	
BI Analyst	Focuses on market trends and external data for business improvement.	
Skill Requirements	Vary by role but include tech, statistics, communication, and domain expertise.	
Career Mobility	Data professionals can shift roles by acquiring new skills.	

Defining Data Analysis

Q What is Data Analysis?

• Definition:

• The process of **gathering, cleaning, analyzing**, and **mining data**, interpreting results, and reporting findings.

Purpose:

- Identifies **patterns and correlations** within data.
- Helps generate **insights** and draw **conclusions**.
- Supports business decisions, enhances performance, and validates actions to save time and resources.

Four Primary Types of Data Analysis

1. Descriptive Analytics

• **Goal**: Answers "What happened?"

• Function:

- Summarizes past data.
- Presents insights to stakeholders.

• Examples:

Tracking KPIs.
Analyzing cash flow.
2. Diagnostic Analytics
• Goal: Answers "Why did it happen?"
• Function:
 Digs deeper into data from descriptive analytics to find root causes.
a Evamples

- o Investigating sudden traffic spikes.
- o Analyzing unexplained sales increase.

3. Predictive Analytics

Goal: Answers "What will happen?"

Function:

- Uses historical data to forecast future outcomes.
- o All predictions are **probabilistic**, not certain.

Examples:

- Risk assessment.
- o Sales forecasting.

4. Prescriptive Analytics

Goal: Answers "What should be done?"

Function:

0	Analyzes decisions and events to estimate outcomes and suggest actions.			
• Examp	• Examples:			
0	 Self-driving cars making real-time decisions. 			
0	Airlines adjusting ticket prices based on demand, weather, traffic.			
Key Steps	in the Data Analysis Process			
1. Understan	ding the Problem			
• Clearly	define:			
0	The problem.			
0	The desired outcome.			
0	Current vs. target state.			
2. Setting a C	lear Metric			
• Decide	what to measure and how.			
• Examp	les:			
0	Metric: Number of product X sold.			
0	Time frame: A quarter, festive season.			
3. Gathering	Data			
• Identify	y:			
0	Required data.			
0	Data sources.			

Appropriate tools.		
4. Cleaning Data		
Why it's critical:		
o Ensures analysis accuracy.		
• Tasks involved:		
Fix missing/incomplete values.		
 Identify and handle outliers (e.g., age = 150). 		
Standardize data from various sources.		
5. Analyzing and Mining Data		
Manipulate and explore data to:		
 Understand trends. 		
o Identify correlations.		
Detect patterns and anomalies.		
6. Interpreting Results		
Review insights critically.		
• Check if:		
Analysis is defendable.		

 $\circ\quad$ There are limitations or exceptions.

7. Presenting Findings

- Use clear, impactful visuals and formats:
 - o Reports
 - Dashboards
 - Charts and graphs
 - Maps 🗺
 - Case studies

→ Summary

This guide explains the **process of data analysis**—a crucial discipline for making informed business decisions. It describes four primary types of data analytics (Descriptive, Diagnostic, Predictive, Prescriptive), each answering a specific question about data. The content also walks through the **step-by-step methodology** used in data analysis, from problem definition to communicating findings. This structured approach helps organizations extract insights, improve strategies, and forecast outcomes, ensuring **data-driven decision-making**.

Key Takeaways:

- Data analysis turns raw data into actionable insights.
- There are **four core types** of analytics, each serving a unique purpose.
- Every analysis follows a **structured process**: understanding the problem, setting metrics, gathering, cleaning, analyzing, interpreting, and presenting.

Table: What We Learnt in the Video

Topic/Section	What We Learnt		
What is Data Analysis?	Process of extracting insights through data collection, cleaning, and interpretation.		
Descriptive Analytics	Summarizes past data to explain "what happened."		
Diagnostic Analytics	Dives deeper into data to explain "why it happened."		

Predictive Analytics Forecasts possible future outcomes based on trends and past

data.

Prescriptive Analytics Recommends actions by evaluating outcomes of past

decisions.

Understanding the Clearly define the problem and goals.

Problem

Setting a Clear Metric Choose what to measure and define timeframes.

Gathering Data Identify data sources and tools.

Cleaning Data Ensure data accuracy by removing errors and outliers.

Analyzing & Mining Discover patterns, trends, and correlations.

Data

Interpreting Results Evaluate findings, address limitations.

Table: Four Primary Types of Data Analysis

Type of Analysis	Question Answered	Purpose	Examples
Descriptive Analytics	What happened?	Summarize historical data to understand past events.	- Tracking key performance indicators (KPIs) - Cash flow analysis
Diagnostic Analytics	Why did it happen?	Investigate root causes of outcomes identified in descriptive analysis.	- Analyzing sudden traffic changes to a website - Investigating sales spikes
Predictive Analytics	What will happen?	Use historical data and trends to forecast future outcomes.	- Risk assessment - Sales forecasting

Prescriptive	What should be	Recommend actions by analyzing	- Self-driving cars deciding
Analytics	done about it?	past decisions and estimating likely results.	speed/routes - Airlines adjusting prices
		mery resures.	prices

Viewpoints: What is Data Analytics?

Q What is Data Analytics?

• General Definitions:

- Process of collecting, cleaning, analyzing, and presenting data to confirm hypotheses.
- A method for making decisions based on factual data.
- Described as "storytelling with data" using insights to communicate the state of the world clearly.

• Components of Data Analytics:

- **Collection** Gathering data from various sources.
- **Cleansing** Removing or correcting inaccurate records or irrelevant data.
- **Analysis** Drawing meaningful insights by applying statistical or logical techniques.
- **Presentation** Communicating findings to stakeholders.
- **Sharing** Ensuring analysis reaches decision-makers to influence strategy or action.

Real-World Analogy

• Everyday Decision-Making:

- Watching a weather report to decide daily attire or activities is a form of data analysis.
- Emphasizes that data analytics is not abstract but embedded in daily life just with a technical term now.

Problem-Solving Approach

• Step-by-Step Process:

- Define the problem.
- o Formulate a hypothesis.
- Collect relevant data.
- Clean the data.
- Analyze findings.
- Present to stakeholders.

• Purpose:

- Use data to test ideas or assumptions (hypotheses).
- o Make informed, evidence-based decisions.

Professional Perspectives

- CPA (Certified Public Accountant):
 - Uses data analytics to:
 - Understand financial health.
 - Predict trends: past, present, and future positioning of businesses.
- Digital Marketing and Content Strategy:
 - Data is gathered from audiences or customers.
 - Broken into subsets to:
 - Guide product/service offerings.
 - Optimize content for engagement and relevance to target audiences.

Summary

In the video, several professionals define and describe the meaning and importance of data analytics. They outline data analytics as a systematic process involving collecting, cleansing, analyzing, and communicating data to support decision-making. Real-world examples, such as checking weather forecasts and financial forecasting, help demystify the concept. The speakers emphasize that data analytics is not just technical — it's an intuitive, everyday activity that now plays a critical role in businesses, especially in finance and digital content. The takeaway is that data analytics empowers clearer understanding and better decisions across various industries.

Key Takeaways:

- Data analytics is both technical and natural.
- It involves hypothesis testing, not just data processing.
- Its power lies in transforming raw information into actionable insights.
- It is applicable across fields from accounting to digital marketing.

Table: What We Learnt in the Video

Topic/Section	What We Learnt
Definition of Data Analytics	A process of collecting, analyzing, and presenting data to test hypotheses.
Real-Life Analogy	Similar to using weather reports to decide daily plans — simple, practical usage.
Problem-Solving in Analytics	Starts with a problem \rightarrow hypothesis \rightarrow data collection \rightarrow analysis \rightarrow presentation.
CPA's Use of Data	Used to assess financial health and predict business trends.
Storytelling with Data	Data used to communicate reality clearly and persuasively.
Digital Application of Analytics	Helps refine products, services, and content based on audience data.
Decision-Making Support	Data analytics guides informed, evidence-backed decisions.

Data Analytics vs. Data Analysis

The terms Data Analysis and Data Analytics are often used interchangeably, including in this course.

However it is important to note that there is a subtle difference between the terms and meaning of the words *Analysis* and *Analytics*. In fact some people go far as saying that these terms mean different things and should not be used interchangeably. Yes, there is a technical difference...

The dictionary meanings are:

Analysis - detailed examination of the elements or structure of something

Analytics - the systematic computational analysis of data or statistics

Analysis can be done without numbers or data, such as business analysis psycho analysis, etc. Whereas *Analytics*, even when used without the prefix "Data", almost invariably implies use of data for perfoming numerical manipulation and inference.

Some experts even say that *Data Analysis* is based on inferences based on historical data whereas *Data Analytics* is for predicting future performance. The design team of this course does not subscribe to this view, and you will see why later in the course as you become familiar with the terms like *predictive analytics, prescriptive analytics, etc.*

So in this course we take a more liberal view, and use the terms Data Analysis and Data Analytics to mean the same thing. For example, an earlier video is titled *Defining Data Analysis*, whereas the preceding video with the viewpoints of several data professionals is titled *What is Data Analytics*. The difference in these titles is not intentional.

1. Introduction to Terms

- Data Analysis and Data Analytics are often used interchangeably.
- Some experts argue these terms have different meanings and should not be used interchangeably.
- The course takes a liberal view and treats both as synonymous.

2. Dictionary Meanings

• Analysis:

- Defined as a detailed examination of the elements or structure of something.
- Can be applied broadly and is not limited to numerical data.
- Examples include business analysis, psychoanalysis, etc.

• Analytics:

- Defined as the systematic computational analysis of data or statistics.
- Implies the use of numerical data for manipulation and inference.
- Even without the word "Data," analytics almost always involves data and numbers.

3. Technical Differences Discussed by Experts

- Some experts differentiate based on focus:
 - Data Analysis: Mainly inference based on historical data.
 - **Data Analytics**: More about **predicting future performance**.
- The course design team disagrees with this strict differentiation.

• Terms like **predictive analytics** and **prescriptive analytics** show overlap and complicate strict definitions.

4. Course Approach

- The course uses **Data Analysis** and **Data Analytics** interchangeably.
- Examples from the course include:
 - Video titled "Defining Data Analysis"
 - Video titled "What is Data Analytics"
- The difference in video titles is **not intentional**, reflecting the flexible usage of terms.

Summary

The terms **Data Analysis** and **Data Analytics** are often used interchangeably, although there are subtle technical differences. Analysis broadly refers to examining components or structure, not necessarily involving numerical data, while analytics involves systematic computational methods applied to data. Some experts distinguish data analysis as working mainly with historical data and data analytics as focused on future predictions, but this course takes a more liberal view, treating both as synonyms. Understanding this flexible usage helps when encountering terms like predictive and prescriptive analytics.

Table: What We Learnt in the Video

Topic/Section	What We Learnt
Difference in terms	Analysis = detailed examination (broader, not always numerical); Analytics = computational data analysis
Usage of Analysis	Can be non-numerical (business, psychoanalysis)
Usage of Analytics	Almost always involves numerical data manipulation
Expert views	Some say Data Analysis = historical data inference; Data Analytics = future prediction
Course stance	Terms used interchangeably; no strict differentiation
Examples from course	Video titles show interchangeable use of terms

Table between Analysis and Analytics based on the content you provided:

Aspect	Analysis	Analytics
Definition	Detailed examination of elements or structure	Systematic computational analysis of data/statistics
Nature	Can be qualitative or quantitative	Primarily quantitative and computational
Use of Data	May or may not involve numerical data	Almost always involves numerical data
Examples	Business analysis, psychoanalysis	Data analytics, predictive analytics, prescriptive analytics
Focus	Understanding past or present	Often focused on inference, manipulation, prediction
Scope	Broader, applies to many fields	Specific to data and statistics
Historical vs Future	Sometimes seen as based on historical data	Sometimes associated with predicting future trends
Course View	Used interchangeably with Analytics	Used interchangeably with Analysis

Summary and Highlights

In this lesson, you have learned the following information:

A modern data ecosystem includes a network of interconnected and continually evolving entities that include:

- Data that is available in a host of different formats, structure, and sources.
- Enterprise Data Environment in which raw data is staged so it can be organized, cleaned, and optimized for use by end-users.
- End-users such as business stakeholders, analysts, and programmers who consume data for various purposes.

Emerging technologies such as Cloud Computing, Machine Learning, and Big Data, are continually reshaping the data ecosystem and the possibilities it offers. Data Engineers, Data Analysts, Data Scientists, Business Analysts, and Business Intelligence Analysts, all play a vital role in the ecosystem for deriving insights and business results from data.

Based on the goals and outcomes that need to be achieved, there are four primary types of Data Analysis:

- Descriptive Analytics, that helps decode "What happened."
- Diagnostic Analytics, that helps us understand "Why it happened."
- Predictive Analytics, that analyzes historical data and trends to suggest "What will happen next."
- Prescriptive Analytics, that prescribes "What should be done next."

The Data Analysis process involves:

- Developing an understanding of the problem and the desired outcome.
- Setting a clear metric for evaluating outcomes.

- Gathering, cleaning, analyzing, and mining data to interpret results.
- Communicating the findings in ways that impact decision-making.

The Data Analyst Role

Responsibilities of a Data Analyst

- **K** Responsibilities of a Data Analyst
 - **b** Data Acquisition
 - Collect data from **primary** and **secondary sources**.
 - **Q** Data Extraction
 - Create **queries** to pull data from databases and systems.
 - / Data Preparation
 - o Filter, clean, standardize, and reorganize data for analysis.
 - **II** Data Interpretation
 - Use statistical tools to analyze data sets.
 - Identify patterns and correlations using statistical techniques.
 - **Pattern Analysis**
 - o Analyze complex data to interpret trends.
 - Properting
 - Prepare **reports** and **charts** to communicate trends clearly.
 - **Documentation**
 - o Document each step of the data analysis process.

Technical Skills Required

• | Spreadsheets

Microsoft Excel, Google Sheets expertise.

• Statistical & Visualization Tools

o IBM Cognos, IBM SPSS, Oracle Visual Analyzer, Microsoft Power BI, SAS, Tableau.

• **A** Programming Languages

• At least one: **R**, **Python** (sometimes C++, Java, MATLAB).

• 📔 Database Knowledge

- Strong SQL skills.
- Work with relational & NoSQL databases.

• Data Repositories

o Data marts, warehouses, lakes, pipelines.

• 🔅 Big Data Tools

o Hadoop, Hive, Spark familiarity.

Functional Skills Required

• Statistics

• Validate analysis and spot errors.

• Analytical Skills

o Research, interpret data, theorize, forecast.

• * Problem-Solving

• Use analysis to solve real-world problems.

• 🚡 Probing Skills

o Understand problems from varied stakeholder views.

• P Data Visualization

• Pick tools & techniques to suit audience and data.

• III Project Management

• Manage processes, teams, timelines, and dependencies.

Soft Skills for a Data Analyst

• Collaboration

• Work effectively with business and cross-functional teams.

• Present findings clearly and tell compelling data stories.

• **a** Curiosity

• Explore new questions from data patterns and anomalies.

• • Intuition

• Sense future trends via pattern recognition and experience.

🍃 Summary

This video covers the **core responsibilities and skills** a Data Analyst needs in today's organizations. It emphasizes a mix of **technical skills** (like programming, databases, and tools), **functional expertise** (statistics, problem-solving, visualization), and crucial **soft skills** (communication, curiosity, intuition). Data analysis is described as both a science and an art, where success depends not only on tools but also on creativity, collaboration, and curiosity to discover meaningful insights.

📊 Table: What We Learnt in the Video

Topic/Section	What We Learnt
* Responsibilities	Acquire, extract, clean, analyze data, report, document
Technical Skills	Excel, statistical tools, programming (R/Python), SQL, Big Data tools
Functional Skills	Statistics, analytical thinking, problem-solving, probing, visualization, project management
Soft Skills	Collaboration, communication, curiosity, storytelling, intuition
Intuition Importance	Use past experience & pattern recognition to anticipate future trends

Viewpoints: Qualities and Skills to be a Data Analyst

1. Qualities of a Data Analyst

• Curiosity

- Naturally curious individuals seek answers even when no questions are posed.
- Willingness to research unconventional areas.

• Attention to Detail

- Noticing patterns and small details (e.g., counting people in a room or observing setup).
- Critical to detect anomalies or inconsistencies in data.

• Enjoyment of Working with Computers

• Technology evolves rapidly; analysts must adapt and learn new tools continuously.

2. Skills Required for a Data Analyst

A. Technical Skills

Programming Languages

• Python, R (useful but not mandatory initially).

• Query Languages

- SQL is essential and widely used for extracting data from databases.
- Fundamental skill to master before moving on to more complex tools.

• Data Visualization Tools

• Tableau, Power BI (important to be proficient in at least one).

• Data Manipulation & Analysis

- Ability to clean, wrangle, and join multiple data tables.
- Mining data to extract actionable insights.

B. Soft Skills (Interpersonal and Cognitive)

• Business Acumen

• Understanding which data to use and how to apply it effectively.

Presentation Skills

• Ability to present data clearly and persuasively to stakeholders.

• Curiosity and Critical Thinking

• Asking relevant questions and listening carefully to users and coworkers.

• Detail-Oriented Mindset

• Not taking data at face value, verifying against previous data to detect fraud or errors.

• Continuous Learning

• Staying updated with fast-evolving analytics tools and techniques.

3. The Evolving Role of Data Analysts

- Analysts need to:
 - Understand the business problem they are solving.
 - Extract and prepare data appropriately from complex data lakes.

- Join and manipulate data from various tables.
- o Derive insights and present them effectively using visual storytelling.

Summary

This video highlights the essential qualities and skills required to become a proficient data analyst. Key qualities include natural curiosity, attention to detail, and a passion for working with evolving technology. Critical technical skills emphasize mastery of SQL, familiarity with programming languages like Python or R, and proficiency in data visualization tools such as Tableau and Power BI. Equally important are soft skills like business acumen, communication, critical thinking, and continuous learning to adapt to a fast-changing field. The modern data analyst's role extends beyond just data extraction to interpreting data, solving business problems, and effectively communicating insights through compelling storytelling.

Table: What We Learnt in the Video

Topic/Section	What We Learnt
Qualities of a Data Analyst	Curiosity, attention to detail, enjoyment of computers, adaptability to rapid technology changes.
Technical Skills	Mastery of SQL, knowledge of Python/R, expertise in Tableau/Power BI, data cleaning and manipulation.
Soft Skills	Business acumen, presentation skills, curiosity, critical thinking, listening skills, continuous learning.
Role of Data Analysts Today	Solve business problems, extract and join data, clean and analyze data, visualize and communicate insights clearly.

Generative AI: An essential Skill for today's Data Analysts

Introduction

As a beginner in data analytics, you're stepping into a field that's rapidly evolving. Generative AI is becoming an essential tool for data analysts, allowing them to create new content and gain deeper insights. Let's explore what generative AI is and how it can enhance your skills.

What is generative AI?

Generative AI refers to a class of artificial intelligence models that create new content such as text, images,

music, and more by learning patterns from existing data.

Generative AI can respond naturally to human conversation and serve as a tool for customer service and personalization of customer workflows. For example, you can use AI-powered chatbots, voice bots, and virtual assistants that respond more accurately to customers for first-contact resolution.

How does generative AI work?

Generative AI starts with a prompt that could be in the form of a text, an image, a video, a design, musical notes, or any input that the AI system can process. Various AI algorithms then return new content in response to the prompt. Content can include essays, solutions to problems, or <u>realistic fakes</u> created from pictures or audio of a person.

Early versions of generative AI required submitting data via an API or an otherwise complicated process. Developers had to familiarize themselves with special tools and write applications using languages such as Python.

Now, pioneers in generative AI are developing better user experiences that let you describe a request in plain language. After an initial response, you can also customize the results with feedback about the style, tone, and other elements you want the generated content to reflect.

Key techniques in generative AI:

Generative adversarial networks (GANs): GANs consist of two neural networks: the generator and the discriminator. The generator creates new data, whereas the discriminator evaluates it. Over time, the generator improves to produce realistic data.

Variational autoencoders (VAEs): VAEs encode input data into a compressed format and then decode it back, generating new data points similar to the input data.

Transformers: Used primarily in natural language processing (NLP), transformers generate human-like text by predicting the next word in a sequence. Generative Pre-trained Transformer 3 (GPT-3) is a notable example.

Generative AI models

Generative AI models combine various AI algorithms to represent and process content. For example, to generate text, various NLP techniques transform raw characters (e.g., letters, punctuation, and words) into sentences, parts of speech, entities, and actions, which are represented as vectors using multiple encoding

techniques. Similarly, images are transformed into various visual elements, also expressed as vectors. One caution is that these techniques can also encode the biases, racism, deception, and puffery contained in the training data.

Once developers settle on a way to represent the world, they apply a particular neural network to generate new content in response to a query or prompt. Techniques such as GANs and VAEs—neural networks with a decoder and encoder—are suitable for generating realistic human faces, synthetic data for AI training, or even facsimiles of particular humans.

Recent progress in transformers, such as Google's Bidirectional Encoder Representations from Transformers (BERT), OpenAI's GPT, and Google AlphaFold, have also resulted in neural networks that can not only encode language, images, and proteins but also generate new content.

What are the use cases for generative AI?

Generative AI can be applied in various use cases to generate virtually any kind of content. The technology is becoming more accessible to users of all kinds thanks to cutting-edge breakthroughs like GPT that can be tuned for different applications.

Some of the use cases for generative AI include the following:

- Implementing chatbots for customer service and technical support.
- Deploying deepfakes for mimicking people or even specific individuals.
- Improving dubbing for movies and educational content in different languages.
- Writing email responses, dating profiles, resumes, and term papers.
- Creating photorealistic art in a particular style.
- Improving product demonstration videos.
- Suggesting new drug compounds to test.
- Designing physical products and buildings.
- Optimizing new chip designs.
- Writing music in a specific style or tone.

What are the benefits of generative AI?

Generative AI can be applied extensively across many areas of the business. It can make it easier to interpret and understand existing content and automatically create new content. Developers are exploring ways that generative AI can improve existing workflows, with an eye to adapting workflows entirely to take advantage of

the technology. Some of the potential benefits of implementing generative AI include the following:

- Automating the manual process of writing content.
- Reducing the effort of responding to emails.
- Improving the response to specific technical queries.
- Creating realistic representations of people.
- Summarizing complex information into a coherent narrative.
- Simplifying the process of creating content in a particular style

What are the limitations of generative AI?

Early implementations of generative AI vividly illustrate its many limitations. Some of the challenges generative AI presents result from the specific approaches used to implement particular use cases. For example, a summary of a complex topic is easier to read than an explanation that includes various sources supporting key points. The readability of the summary, however, comes at the expense of a user being able to vet where the information comes from.

Here are some of the limitations to consider when implementing or using a generative AI app:

- It does not always identify the source of content.
- It can be challenging to assess the bias of original sources.
- Realistic-sounding content makes it harder to identify inaccurate information.
- It can be difficult to understand how to tune in to new circumstances.
- Results can gloss over bias, prejudice, and hatred.

What are the concerns surrounding generative AI?

•The rise of generative AI is also fueling various concerns. These relate to the quality of results, the potential for misuse and abuse, and the potential to disrupt existing business models. Here are some of the specific types of problematic issues posed by the current state of generative AI:

- It can provide inaccurate and misleading information.
- It is more difficult to trust without knowing the source and provenance of information.
- It can promote new kinds of plagiarism that ignore the rights of content creators and artists of original content.
- It might disrupt existing business models built around search engine optimization and advertising.
- It makes it easier to generate fake news.

- It makes it easier to claim that real photographic evidence of wrongdoing was just an AI-generated fake.
- It could impersonate people for more effective social engineering cyberattacks.
- Given the newness of GenAI tools and their rapid adoption, enterprises should prepare for the inevitable "trough of disillusionment" that's part and parcel of emerging technology by adopting sound AI engineering practices and making responsible AI a cornerstone of their GenAI efforts, ensuring transparency, ethical considerations, and long-term sustainability in their AI implementations.

What are some examples of generative AI tools?

Generative AI tools exist for various modalities, such as text, imagery, music, code, and voices. Some popular AI content generators to explore include the following:

- Text generation tools include GPT, Jasper, AI-Writer, and Lex.
- Image generation tools include Dall-E 2, Midjourney, and Stable Diffusion.
- Music generation tools include Amper, Dadabots, and MuseNet.
- Code generation tools include codeStarter, Codex, GitHub Copilot, and Tabnine.
- Voice synthesis tools include Descript, Listnr, and Podcast.ai.
- AI chip design tool companies include Synopsys, Cadence, Google, and NVIDIA.

Applications of generative AI in data analytics

Generative AI has many applications that can enhance your data analytics work:

- Data augmentation: Create synthetic data to augment existing data sets, which is especially useful when data is scarce or imbalanced. This can improve predictive model performance.
- Anomaly Detection: Identify anomalies or outliers by understanding the distribution of normal data.
 This is valuable in fraud detection, network security, and quality control.
- Text and image generation: Generate realistic text and images for marketing, content creation, and customer engagement, such as automatic product descriptions and marketing visuals.
- Simulation and forecasting: Simulate scenarios and forecast future events by generating potential outcomes from historical data. This is crucial in financial planning, supply chain management, and strategic decision-making.

Conclusion

Generative AI is a transformative technology that can significantly enhance your capabilities as a data analyst. By mastering generative AI techniques, you can unlock new possibilities in data augmentation, anomaly detection, content creation, and forecasting. As you embark on this journey, remember to balance innovation with ethical responsibility, ensuring that AI is used positively.

A Day in the Life of a Data Analyst

Introduction: Role of a Data Analyst

- Data Analysts perform a variety of tasks:
 - Acquiring data from multiple sources
 - Writing queries to extract data from repositories
 - Examining large datasets to find insights
 - Creating reports and dashboards
 - Interacting with stakeholders to gather requirements and present findings
 - Cleaning and preparing data to ensure credible analysis (a significant portion of the role)
- The speaker, Sivaram Jaladi, finds the investigative part of analyzing data for insights especially fascinating.

About the Speaker and Company

- Sivaram Jaladi, Data Analyst at Fluentgrid
- Fluentgrid is a smart grid technology solutions company in Vishakhapatnam, India
- Fluentgrid is an IBM partner and recipient of IBM Beacon awards
- Fluentgrid provides integrated operations center solutions for power utilities and smart cities
- Uses a platform called Fluentgrid Actilligence for actionable intelligence

Project Context

- Client: Power utility company in South India
- Issue: Increase in complaints about overbilling
- Pattern suggests the issue is more than random incidents
- Task: Analyze complaints and billing data to identify causes

Initial Steps

- Data sources to analyze:
 - 1. Complaint data
 - 2. Subscriber information
 - 3. Billing data
- Develop initial hypotheses/questions to guide analysis:
 - 1. Usage patterns of complainants is overbilling related to certain consumption ranges?
 - 2. Geographic concentration are complaints clustered in specific localities?
 - 3. Repeat complaints are same subscribers repeatedly reporting overbilling? If yes, what is the frequency?
 - 4. Consistency if a subscriber is overbilled once, does it continue every month or occur sporadically?

Data Analysis Process

- Extract average billing amounts (annual, quarterly, monthly) of complainants to find complaint concentration in usage ranges
- Analyze location data (zip codes) for spatial concentration of complaints
 - Found complaints concentrated in specific areas
- Investigated subscriber tenure using connection date
 - Over 95% of complainants were subscribers for more than 7 years
 - Not all long-term subscribers had complaints
- Examined meter details (make and serial number)
 - Found that affected meters belonged to the same batch from a single supplier
 - These meters were installed in the complaint-heavy areas

Findings and Conclusion

- Overbilling complaints clustered geographically in areas with meters from a specific batch
- Majority of complainants were long-term subscribers, indicating the problem wasn't random
- Findings were clear enough to present to stakeholders

- Shared data sources and methodology to build credibility
- Project may conclude or lead to further investigations with new complaint patterns

Summary

Sivaram Jaladi, a Data Analyst at Fluentgrid, describes a typical investigative day focused on analyzing overbilling complaints for a South Indian power utility. The process involved gathering and cleaning data from complaints, subscriber info, and billing records. He formed hypotheses about usage patterns, geography, repeat complaints, and consistency. By examining billing amounts, location, subscriber tenure, and meter batch details, he discovered a concentration of complaints linked to a specific meter batch installed in certain areas. This data-backed insight was then presented to stakeholders, illustrating the crucial role of data cleaning, analysis, and communication in solving real-world utility issues.

Key takeaways:

- Data analysts handle varied tasks from data extraction to stakeholder communication
- Forming hypotheses early helps direct analysis efficiently
- Data cleaning and validation are vital for credible insights
- Geographical and equipment-based patterns can reveal root causes of issues
- Transparent sharing of methods boosts stakeholder confidence

Table: What We Learnt in the Video

Topic/Section	What We Learnt
Role of Data Analyst	Multi-faceted role including data acquisition, cleaning, analysis, reporting, and communication
Company & Context	Fluentgrid provides smart grid solutions; project deals with overbilling complaints for power utility
Initial Hypotheses	Usage patterns, geographic concentration, repeat complaints, consistency of overbilling
Data Sources	Complaint data, subscriber information, billing data
Analytical Approach	Extract billing averages, analyze zip codes, check subscriber tenure, review meter batch info

Findings	Overbilling linked to specific meter batch installed in complaint-heavy areas
Importance of Process Sharing	Sharing data sources and analysis methods builds credibility
Project Outcome	Possible resolution or further investigation depending on new complaint patterns

Viewpoints: Applications of Data Analytics

1. Applications of Data Analytics in Everyday Life

- Data analytics is ubiquitous and integrated into daily life.
- Examples in everyday contexts:
 - Commercials: Data analysis helps identify what information to highlight (e.g., "4 out of 10 dentists recommend").
 - Health monitoring: Diabetes patients track sugar levels using data analysis.
 - Consumer information such as calorie counts or reactions to products.
- The process of analysis is continuous and universal, embedded in many small daily decisions.

2. Broad Applicability Across Industries and Functions

- Analytics benefits every industry, vertical, and organizational function.
- Examples of applications:
 - Sales pipeline analysis.
 - Financial reporting and monthly financial analysis.
 - Headcount planning and review.
- Industries mentioned:
 - Airlines
 - Pharmaceuticals
 - Banking
- Analytics improves decision-making across diverse business functions.

3. Role of Data Analytics During the Pandemic

• The pandemic caused shifts in consumer buying habits.

- Companies are using data analytics to:
 - Monitor and understand these changed behaviors.
 - Pivot quickly to meet new demands.
 - o Cater effectively to evolving customer needs.
- Data analytics has become more critical in adapting to rapid market changes.

4. Specific Applications in Finance

- Increasing use of alternative data in financial analysis.
- Examples include:
 - Sentiment analysis of tweets and news to supplement traditional finance metrics and guide investment decisions.
 - Satellite imagery to track industrial activity and growth.
 - Geolocation data to monitor store traffic and predict sales volumes.
- These innovative data sources enhance financial insights beyond conventional data.

Summary

The video highlights the pervasive and vital role of data analytics across everyday life, multiple industries, and especially during dynamic conditions such as the pandemic. It illustrates how analytics is not just a specialized skill but a fundamental part of how information is processed and decisions are made, from simple consumer information to complex financial strategies. Key takeaways include the universal presence of data analytics, its broad cross-industry applicability, its crucial role in adapting to pandemic-driven market shifts, and the innovative use of alternative data sources in finance.

Table: What We Learnt in the Video

Topic/Section	What We Learnt
Everyday Life Applications	Data analytics is embedded in daily activities, commercials, health monitoring, and consumer info.
Industry and Functional Uses	All industries and functions (sales, finance, HR) benefit from analytics, enhancing decision-making.
Pandemic Impact	Analytics helps businesses track changing consumer behavior and adapt quickly to maintain demand.

Finance	Sector
Applica	ations

Use of alternative data like sentiment analysis, satellite imagery, and geolocation enriches financial insights.

Summary and Highlights

In this lesson, you have learned the following information:

The role of a Data Analyst spans across:

- Acquiring data that best serves the use case.
- Preparing and analyzing data to understand what it represents.
- Interpreting and effectively communicating the message to stakeholders who need to act on the findings.
- Ensuring that the process is documented for future reference and repeatability.

In order to play this role successfully, Data Analysts need a mix of technical, functional, and soft skills.

- Technical Skills include varying levels of proficiency in using spreadsheets, statistical tools, visualization tools, programming and querying languages, and the ability to work with different types of data repositories and big data platforms.
- An understanding of Statistics, Analytical techniques, problem-solving, the ability to probe a situation from multiple perspectives, data visualization, and project management skills – all of which come under Functional Skills a Data Analyst needs in order to play an effective role.
- Soft Skills include the ability to work collaboratively, communicate effectively, tell a compelling story with data, and garner support and buy-in from stakeholders. Curiosity to explore different pathways and intuition that helps to give a sense of the future based on past experiences are also essential skills for being a good Data Analyst.