

SUMMARY: INTRODUCTION TO DATA ANALYTICS

SESSION OVERVIEW-

By the end of this session, students will be able to:

1. Understand data analytics and its use in the real world.
2. Understand career scope and opportunities in the data domain.
3. Distinguish between different types of data and data analytics.

KEY TOPICS AND EXAMPLES:

1. Understand data analytics and its use in the real world.

- **What is data analytics?**

Data analytics refers to analyzing, interpreting and deriving insights from raw data to help decision making, solving problems and optimize outcomes.

- Now with the help of real time examples, explanations on data analytics will help the students to have a better understanding of data analytics.

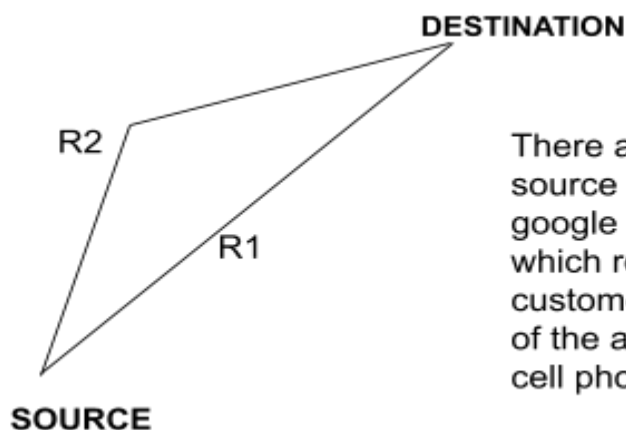
- **Example-**

1. **e-commerce product recommendation.**

We'll look at an example here of how data analytics are used by e-commerce platforms to provide consumers with customized product recommendations. We will go over the fundamental ideas, methods that underpin recommendation systems. Imagine you are a user browsing for a new laptop. Based on the past purchases and present browsing history, the platform recommends different laptops with the same specifications, price range, etc. These recommendations are created with the help of past data fed to an algorithm.

2. **Uber finding the nearest driver.**

3. **Google maps showing the shortest route**, amount of traffic and estimation of time interval between source and destination.



There are two routes from source to destination and google maps is suggesting which route should the customers choose on the basis of the amount of traffic (active cell phones) in both the route.

4. **Risk detection:**

In earlier times, amid financial challenges, organizations turned to data analytics as a tool for addressing their predicaments. With access to extensive customer data, including transaction histories and demographic details, they employed systematic data analysis techniques. By carefully dissecting this data and scrutinizing patterns, such as recent spending behaviors and customer profiles, they could discern indicators suggestive of potential defaults or fraudulent activities. Through this methodical approach, organizations successfully mitigated risks and enhanced their fraud detection capabilities.

5. **How Data analytics helped the world to fight COVID:**

COVID-19 has had a huge impact on the world. But it could have been even more catastrophic if not for the use of big data analytics in the healthcare industry.

If you were in India at the time, you must have downloaded the Aarogya Setu app on your phone to detect any threat of infection. The app uses Bluetooth and data analytics for contact tracing.

2. **Understand career scope and opportunities in the data domain.**

- **Explanation on the vast array of career opportunities** that the data domain offers with the requirement of skills w.r.t each opportunity.
 - **Data analyst**- data collection, data cleaning and analyzing data.
 - **Data science**
 - **Business Intelligence developer**- develop dashboards using power BI, tableau, etc.
- **How is data analytics important in today's world?**
 - **Procter and gamble (FMCG):**
 - Consumer based companies like Procter and Gamble (P&G) heavily rely on data analysis to gain a competitive advantage.
 - Data analysis allows P&G to gather and **analyze vast amounts of data** from various sources, including **sales data, customer feedback, and market trends**.
 - Data analysis helps P&G to find **emerging trends and customer needs** which helps companies to improve their products.
 - Data analysis enables P&G to **optimize its marketing and advertising campaigns** for maximum effectiveness.
 - Through data analysis, P&G can gather **competitive intelligence** by monitoring competitor activities, pricing strategies, and market share trends.
 - **Healthcare:**
 - Data analytics allows healthcare providers to **analyze patient data**, such as electronic health records (EHRs), medical imaging, and genetic information, to identify patterns, trends, and correlations that can improve diagnosis, treatment, and patient care.

- Enables healthcare organizations to **optimize clinical workflows, resource allocation**, and operational efficiency.
- **Manufacturing:**
 - To implement predictive maintenance programs, which involve analyzing equipment sensor data to predict when machines are likely to fail.
 - To optimize their supply chain operations by analyzing data from suppliers, logistics providers, and internal production processes.
 - To optimize production processes, improve efficiency, and reduce waste.
- **How the course will help students build a career in data analytics.**
 - **Foundational knowledge**
 - **Technical skills** - excel, SQL, Power BI, Python
 - **Soft skills**- Improving problem solving skills, solution designing, deriving insights and communication.
 - **Hands-on experience**- Real time projects and case studies.

3.A. Introduction to data and types of data:

- **What is data?**
 - Collection of facts.
 - Types of data based on organization:
 - Unstructured data
 - Semi-structured data
 - Structured data
- **How to convert data to information?**

UNSTRUCTURED DATA- Data which do not have predefined structure or organization.
Example- Audio, text, video, photographs, etc.

EXAMPLE:

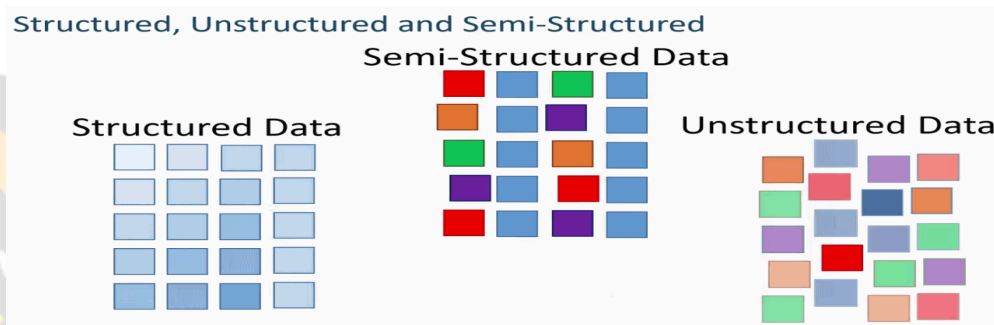
1. **CUSTOMER REVIEWS:**

Online review platforms like Yelp, TripAdvisor, and Amazon host millions of customer reviews containing valuable insights about products, services, and businesses. By applying text mining and sentiment analysis techniques to analyze customer reviews, businesses can identify recurring themes, sentiments, and patterns, enabling them to improve product quality, address customer concerns, and enhance customer satisfaction.

2. **SOCIAL MEDIA POSTS:**

Social media platforms generate vast amounts of unstructured data in the form of text, images, and videos. Data analytics techniques such as sentiment analysis can be applied to analyze social media posts and understand public sentiment towards products, brands, or events. Natural language processing (NLP) algorithms can extract topics, trends, and opinions from unstructured text data, enabling

organizations to make data-driven decisions based on customer feedback and market trends.

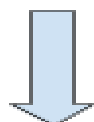


SEMI-STRUCTURED DATA
PID= 1
LONGITUDE= 133
LATITUDE= -33
SOURCE = 2
DESTINATION= 3



STRUCTURED DATA:

PID	LONGITUDE	LATITUDE	SOURCE	DESTINATION
1	133	-33	2	3



INFORMATION:

PID	LONGITUDE	LATITUDE	SOURCE	DESTINATION
1	133	-33	A1	A3

SEMI-STRUCTURED DATA:

- Semi-structured data does not reside in relational database (but structured data does)
- Has organizational properties making it easier to analyze
- Example - emails (which have structured headers and metadata but unstructured body content).

STRUCTURED DATA:

- When we receive **semi-structured** data, then we convert it into **structured data** and after data processing, we finally end up getting **information** from the data.
- In the above hypothetical example of google maps, we have data of certain areas, let's say A1, A2 and A3. Variables involved in the data are personal ID, longitude, latitude, source and

destination. When we convert the raw data into structured data, basically we have converted the data into **tabular form** which is considered to be highly structured and can be stored in a relational database.

- After converting the data into structured format which is meaningful and understandable, now we can derive insights. Here, each source and destination suggests a particular area and if replaced with the area number then it will be more organized.

- **Importance of understanding type of data**

<u>STRUCTURED DATA</u>	<u>SEMI-STRUCTURED DATA</u>	<u>UNSTRUCTURED DATA</u>
Highly organized data	Partially organized data	Unorganized data
Easily machine-readable	Moderately machine readable	Challenging for machines
Guarantees highest performances with structure	Only allows anonymous queries	Only textual query is executable
Relational database, spreadsheets	HTML, JSON files, XML	Images, audios, videos

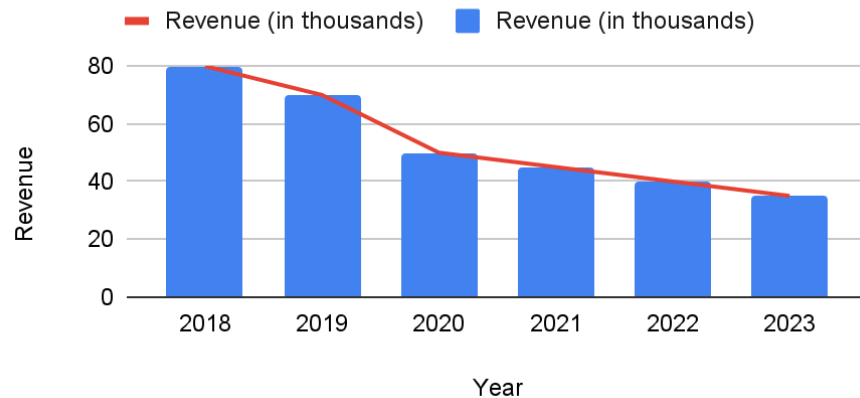
3.B. Introduction to types of data analytics:

- **Types of data analytics:**
 - **Descriptive (WHAT):**
 - >Analysis of the past performances to analyze the trend.
 - >Helps in understanding any underlying issues of the company.
 - >Utilized by almost every company to create dashboards and reports to figure out existing trends from historical data. E.g.Sales trends, etc.

1. **EXAMPLE- (Hypothetical)**

A company has been following a downward trend for the past 6 years which has become a little concerning for the company as well as the stakeholders of the company. The stakeholders are worried and want to understand the reason behind the downward trend and want to discuss what can be done to increase the revenue.

Revenue vs Year



2. EXAMPLE-

○ Diagnostic (WHY): (Root cause analysis)

—> It tells us the reason why a particular trend is visible.

- Now we know that the company is facing a downward trend for a couple of years.
- The next step is to understand the reasons why the company is facing a downward trend.
 - Better products in the market (New competitor)
 - Less differentiation.

Other examples:

Starbucks case:

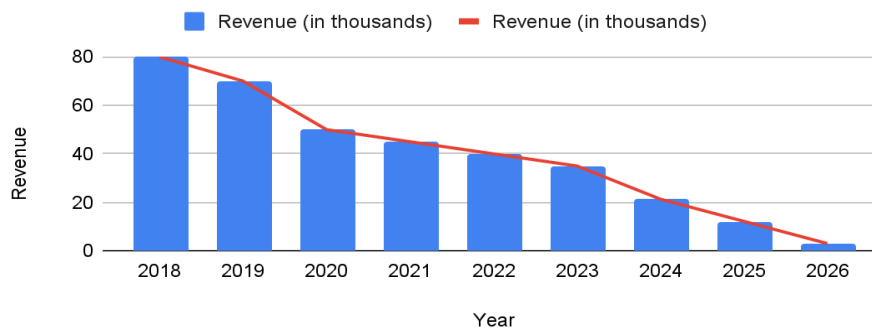
- Starbucks was facing customer retention and store performance issues
- It employed diagnostic analysis by using
 1. customer feedback analysis (semi structured data)
 2. sales data analysis at store level (structured data)
 3. operations analysis (staff schedule, inventory, and order fulfillment efficiency)
- Based on above analysis, it-
 - offered new products (based on customer survey & sales)
 - improved operations and redesigned store layout
 - personalized marketing (based on customer level sales analysis)

○ Predictive (Forecasting future predictions):

—> Tell us what is likely to happen in the future.

—> It is based on historical data and helps us make assumptions about the future trend.

Revenue vs Year



- If we forecast the revenue of the company for the upcoming 3 years, then we notice that there is a downward trend in future revenue.

Example 2- How **amazon uses predictive analysis** to increase revenue.

- Personalized Recommendations
- Enhanced Customer Experience
- Optimized Inventory Management

Example 3- Finance and banking

- Assess credit risk
- Detect fraudulent transactions
- Personalize customer experiences.
- customer transaction data.
- Credit scores.
- Market trends to make informed decisions on lending

○ Prescriptive:

- >Based on the current data, it helps in future decision making.
- >Providing suggestions based on the analysis performed.
 - Improvement in the product.
 - Marketing of the product.