

Foundation of Data Analytics (CDA-105)

LECTURE-1-3



What is Data Analytics (Introduction to Python – GeeksforGeeks)

Data Analytics is the science of analyzing raw data to make conclusions about that information . It encompasses a wide range of analysis techniques, including math, statistics, and computer science, to draw insights from data sets . Data analytics helps businesses optimize their performance, perform more efficiently, maximize profit, or make more strategically-guided decisions . The techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.



Various Approaches to Data Analytics

- 1 **Descriptive Analytics:** This approach involves looking at what happened in the past and summarizing the data to gain insights into historical trends and patterns.
- 2 **Diagnostic Analytics:** This approach focuses on understanding why something happened by analyzing the data and identifying the root causes of specific outcomes
- 3 **Predictive Analytics:** This approach involves using historical data to make predictions about future events or outcomes
- 4 **Prescriptive Analytics:** This approach aims to determine what should be done next based on the insights gained from descriptive, diagnostic, and predictive analytics



Softwares: (Python – GeeksforGeeks)

Data analytics relies on a variety of software tools, including spreadsheets, data visualization, reporting tools, data mining programs, and open-source languages for the greatest data manipulation . It helps businesses optimize their performance, perform more efficiently, maximize profit, or make more strategically-guided decisions . Data analytics can be applied to any type of information to reveal trends and metrics that would otherwise be lost in the mass of data. This information can then be used to optimize processes and increase the overall efficiency of a business or system

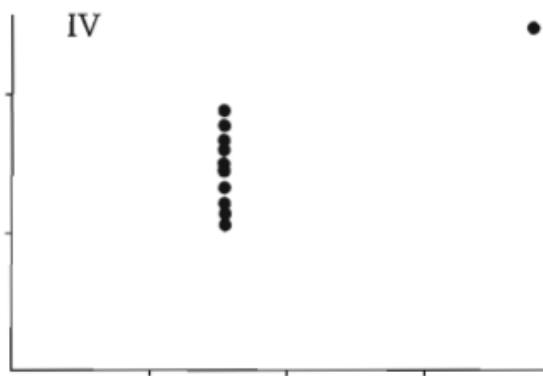
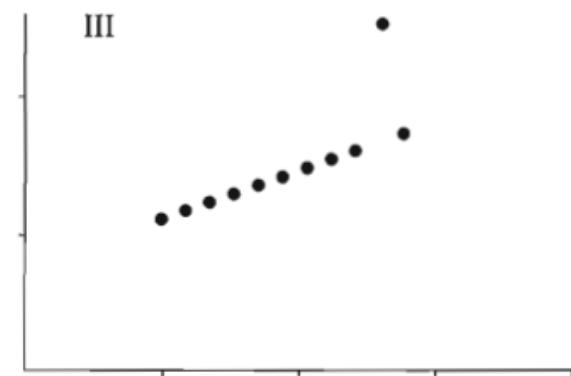
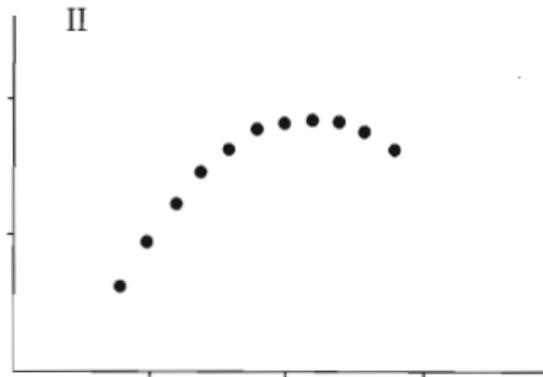
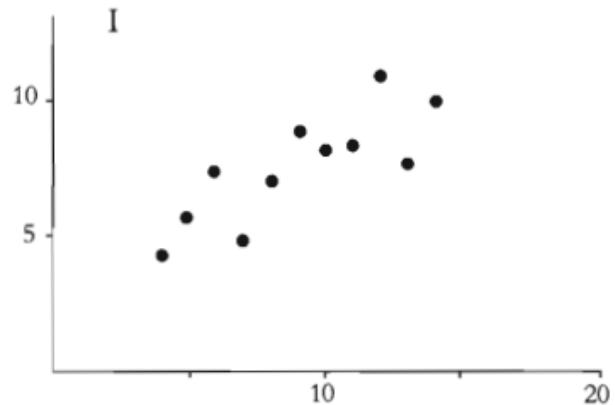


Examples:

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

N = 11
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope = 0.118
 $t = 4.24$
sum of squares $X - \bar{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
 $r^2 = .67$

How to Visualize the Data



Methods of Collection of Data

Data can be collected through the Primary and Secondary Sources:

Primary Data: Primary Data refers to data that is collected first hand by a researcher or a team of researchers for a specific research project or purpose. It is original information that has not been previously published or analyzed, and it is gathered directly from the source or through the use of data collection methods such as surveys, interviews, observations, and experiments .

Some common formats for primary data collection include:

- Textual data: This includes written responses to surveys or interviews, as well as written notes from observations.
- Numeric data: Numeric data includes data collected through structured surveys or experiments, such as ratings, rankings, or test scores.
- Audio data: Audio data includes recordings of interviews, focus groups, or other discussions.
- Visual data: Visual data includes photographs or videos of events, behaviors, or phenomena being studied.

Methods of Collection of Data

- Sensor data: Sensor data includes data collected through electronic sensors, such as temperature readings, GPS data, or motion data.
- Biological data: Biological data includes data collected through biological samples, such as blood, urine, or tissue samples.

Primary data is collected through research methods such as surveys, interviews, experiments, and observations. The purpose of primary data is to gather information directly from the source, without relying on secondary sources or pre-existing data

Secondary Data:

Secondary Data refers to data that is collected by someone other than the primary user. It is information that has been collected, processed, and published by someone else, rather than the researcher gathering the data firsthand.

Secondary Data

There are two types of secondary data based on the data source:

- **Internal sources of data:** This refers to information gathered within the researcher's company or organization, such as a database .
- **External sources of data:** This refers to data collected outside the organization, such as government statistics or mass media channels .

Secondary data sources are extremely useful as they allow researchers and data analysts to build large, high-quality databases that help solve business/ Research problems . Some popular examples of secondary data include tax records, census data, electoral statistics, health records, books, journals, social media monitoring, and more .

If you're interested in collecting secondary data, you can explore external sources such as government publications, academic journals, market research reports, and other existing datasets

Census Method

Census Method is a method of collecting data in which an investigator gathers information related to the problem under investigation by covering every item of the population or universe . It involves a complete enumeration of the population, where data is collected from each and every item . For example, if an investigator wants to investigate the color composition of a cars (Say Tata) in India, they would collect data on the color of each Tata car sold in India . The census of the population is the most essential method of statistical inquiry, and it is conducted every ten years in India . The last census was held in February 2011 . The census method is suitable when the size of the population is small, there are widely diverse items in the population, intensive examination of different items is required, and a high degree of reliability and accuracy is needed

Survey Method

Survey Method is a process, tool, or technique that researchers use to gather information in research by asking questions to a predefined group of people¹. It is a flexible and exciting process that allows you to collect relevant information from research participants or the people who have access to the required data¹. There are different survey methods that allow you to collect information, such as interviews, surveys, and observations¹. Typically, your research context, the type of systematic investigation, and many other factors should determine the survey method you adopt.

There are different types of survey methods, including:

Interviews: An interview is a survey research method where the researcher facilitates some sort of conversation with the research participant to gather useful information about the research subject¹. This conversation can happen physically as a face-to-face interview or virtually as a telephone interview or via video and audio-conferencing platforms¹. During an interview, the researcher has the opportunity to connect personally with the research subject and establish some sort of relationship

Survey Method

- **Surveys:** A survey is a data collection tool that lists a set of structured questions to which respondents provide answers. Surveys can be conducted in various formats, such as online surveys, paper surveys, and telephone surveys. Surveys can be qualitative or quantitative depending on the type of research and the type of data you want to gather in the end .
- **Observations:** Observations involve systematically recording the behavior or activities of individuals or groups in a natural or controlled setting. This type of data collection is often used in fields such as anthropology, sociology, and psychology.
- **Case Studies:** Case studies involve in-depth analysis of a particular individual, group, or organization . They typically involve collecting a variety of data, including interviews, observations, and documents.
- **Action Research:** Action research involves collecting data to improve a specific practice or process within an organization or community . It often involves collaboration between researchers and practitioners .

The choice of survey method depends on the research question, the type of data needed, and the resources available

Advantage of Surveys

Surveys offer several advantages for data collection. Here are some of the key benefits:

Inexpensive: Surveys are one of the most cost-effective methods of gathering quantitative data. They can be self-administered, avoiding the need for in-person interviews. You can distribute surveys through various channels, such as websites, emails, or social media profiles. This flexibility allows you to collect a large amount of information from a diverse demographic in a relatively short time¹.

Practical: Surveys provide a practical way to gather information about specific topics. You have control over the questions asked and the format used, such as polls, questionnaires, quizzes, open-ended questions, or multiple-choice options. The real-time nature of surveys allows for immediate feedback and useful insights¹.

Fast Results: With today's mobile and online tools, surveys can generate results quickly¹. Depending on the scale and reach of your questions, you can receive responses in as little as one day. This speed enables you to make decisions promptly and take necessary actions.

Scalability: Surveys can be scaled to reach a large number of participants. Online surveys, in particular, offer a faster response time compared to traditional methods. You can transfer and use the collected data in various applications to answer important questions.

These advantages make surveys an attractive option for researchers and organizations looking to collect data efficiently and effectively.

Observation Method for Collection of Data:

Observation Method is a process that involves human or mechanical observation to observe and describe the behavior of a subject . It is a way of collecting relevant information and data by observing people's behavior . The observational research method is also referred to as a participatory study because the researcher has to establish a link with the respondent and for this has to immerse himself in the same setting as theirs . Only then can he use the observational research method to record and take notes .

There are different types of observation methods, including:

- **Controlled Observations:** This method is carried out in a closed space. It is the researcher who has the control over the environment and the variables .
- **Naturalistic Observations:** Social scientists and psychologists generally use the naturalistic observation method in their research. This method involves observing people in their natural environment without any interference from the researcher .
- **Participant Observations:** This method involves the researcher becoming a part of the group being studied . The researcher immerses himself in the group and observes their behavior .

Advantages

Observation method has several advantages and disadvantages. Here are some of the key benefits:

Advantages:

- **Easiest Method:** The simplest method of data collection is the method of observation. Very minimal technical knowledge is required, and even though scientifically controlled observations require some technical skills, it is still more accessible and more straightforward than other methods².
- **Natural Surroundings:** The observation method of data collection describes the observed phenomenon precisely and does not introduce any artificiality like other methods. They describe the phenomenon precisely as it occurs in the natural research environment .
- **High Accuracy:** In interview methods and questionnaire methods, the respondents' information provides us the information with which the researchers have to work. These are all indirect methods, and there is no means to investigate the accuracy. But in the observation method, the information accuracy can be checked by various testing. So, the data collected by observation is much more reliable .

Dis-Advantages

Disadvantages:

- **Time-consuming:** Observation takes a lot of time, and it is not always possible to observe everything that is happening .
- **Observer Bias:** The observer's bias can affect the results of the observation method. The observer's bias can be due to the observer's personal beliefs, values, and attitudes .

Limited Generalizability: The observation method is limited in its generalizability. The results of the observation method cannot be generalized to other populations or situations

Experimental Collection of Data

Experimental Method is a systematic process of collecting data that involves manipulating the samples by applying some form of treatment prior to data collection . It refers to manipulating one variable to determine its changes on another variable ¹. The sample subjected to treatment is known as “experimental units” .

Experimental research is primarily a quantitative method . It allows researchers to have a high level of control over the variables being studied, making it possible to determine if a potential outcome is viable . Experimental research can be used in a wide variety of situations and industries . For example, teachers might use experimental research to determine if a new method of teaching or a new curriculum is better than an older system, while pharmaceutical companies use experimental research to determine the viability of a new product

Advantages

The advantages of experimental research include:

Control: Researchers have a high level of control over the variables being studied, allowing them to determine if a potential outcome is viable .

Versatility: Experimental research is not limited to a specific industry or type of idea. It can be used in a wide variety of situations .

Specific Conclusions: Experimental research provides conclusions that are specific and relevant with consistency .

Replicability: The results of experimental research can be duplicated when the same variables are controlled by others, promoting the validity of a concept .

Speed: Research conducted within a laboratory environment allows for the replication of natural settings with faster speeds, enabling researchers to have greater control over variables .

Disadvantages

However, experimental research also has some disadvantages, including:

Time-consuming: Experimental research can be time-consuming, and it may not be possible to observe everything that is happening .

Observer Bias: The observer's bias can affect the results of the experimental method .

Limited Generalizability: The results of experimental research may not be generalizable to other populations or situations.