Toppersnotes ®

Unleash the topper in you

SET

State Eligibility Test

D.I, ICT, People, Development & Environment and Higher Education System

SIERRA INNOVATIONS

SET

Dear Aspirant,
Thank you for making the right decision by choosing ToppersNotes.

To use the QR codes in the book, Please follow the below steps :-

To install the app, scan the QR code with your mobile phone camera or Google Lens

ToppersNotes Exam Prepration app

»
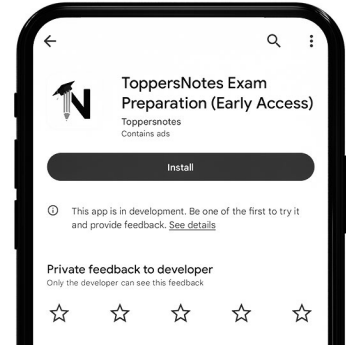
Download the app from Google play store
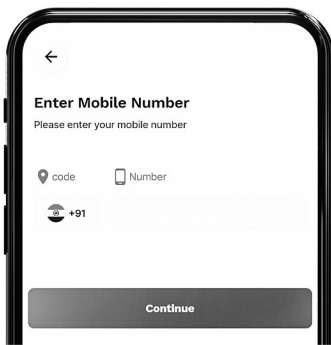
To Login **enter your phone number**

»
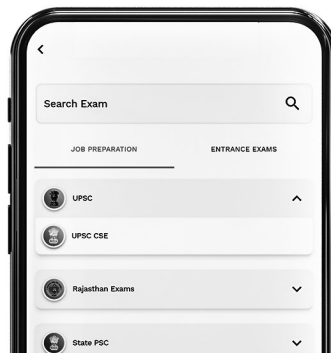
Choose your **exam**

»

Click on **search Button**
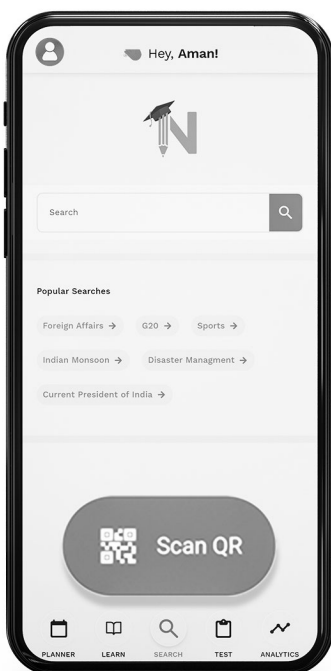
Click on **Scan QR**

»

Choose the **QR from book**

→ • Solution Videos
• Concept Videos
• Doubt Videos

→ • Additional Learning Material

→ • Topic wise practice
• Weakness analysis

→ • Rank Predictor
• Test Practice

For any help,
write us at **hello@toppersnotes.com or whatsapp on** ⓧ **7665641122.**

How to use the Coupon Code?
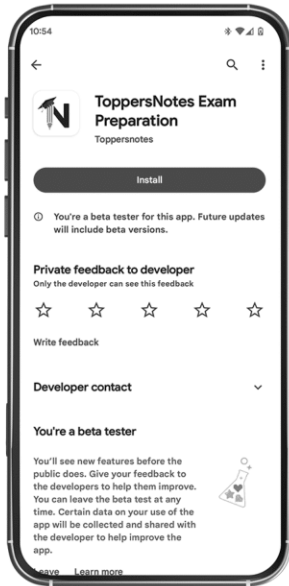
STEP:1

Scan the QR code from the back page and install the Toppersnotes learning app.

→
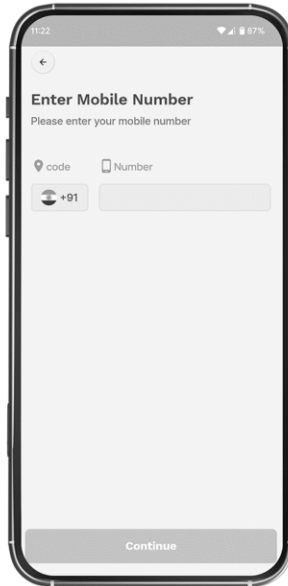
STEP:2

login with your registered phone number and select your exam.

→
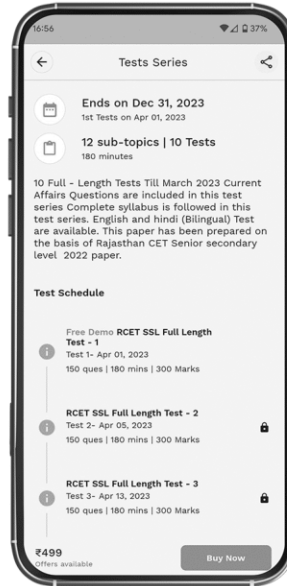
STEP:3

On the test series page you can try demo test or Click on buy now

→
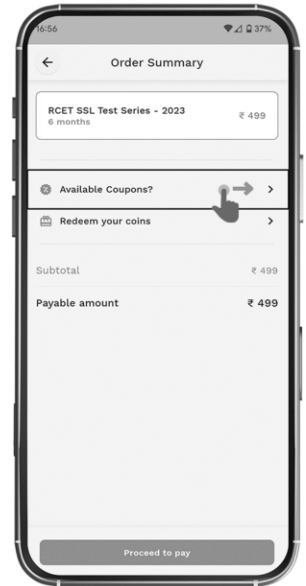
STEP:4

Click on apply coupon
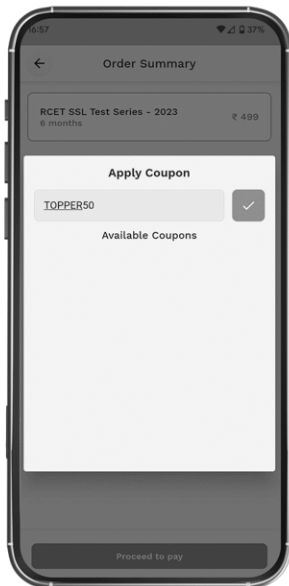
STEP:5

Enter the coupon code.

→
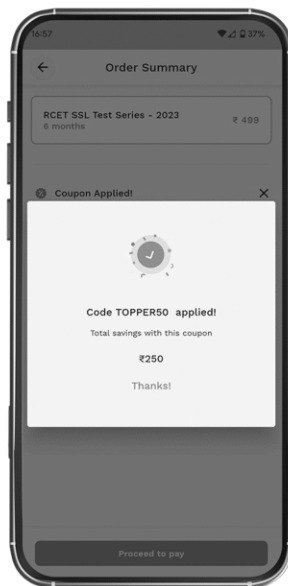
STEP:6

Your code will be applied and then proceed with the payment.

→
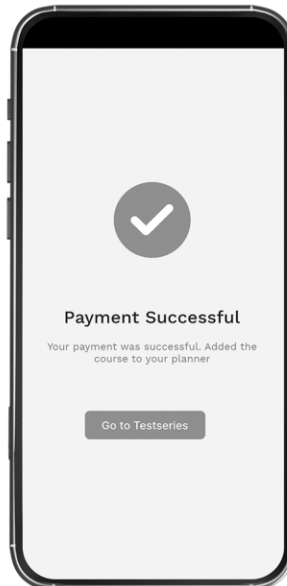
STEP:7

After successful payment click on go to test series

→
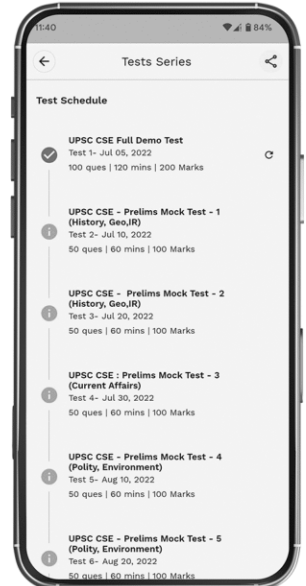
STEP:8

Your test series subscription is active now

For any technical support or queries call
📞 **9614-828-828**

Email
✈ **apps@toppersnotes.com**

Chapter Analysis

UNIT - 7

Data Interpretation

Chapter	1
Questions	5 Questions

- ✓ Basic Calculative & easy to moderate level questions asked from this unit

- ✓ Practice basic math concepts including general calculations, percentage, average & ratio.

- ✓ Practice PYQ as much as possible

Key Points

- ➢ Basics of
- ➢ Mathematics
  - > Average
  - > Percentage
  - > Ratio
- ➢ Table Chart

7 UNIT Data Interpretation

Chapter - 1

Data Sources, Acquisition and Classification of Data

Learning Objectives

● Data acquisition and classification

● Qualitative and Quantitative Data

● Photographic description and data map

● Interpretation of data

● Statistics and Good Governance

Introduction

The material provided is basically a bunch of raw facts, figures and statistics. The process by which meaningful information is obtained - data. For example, the number is often called statistics or the receipt of any information, its calculation, process or processing. The data is compiled, checked and stored after arranging in some sequence. After this it is sent to different person. The process consists of the following steps

1. Calculation - Addition, Subtraction, Multiplication, Division

2. Comparison - Equal, Greater, Smaller, Zero, Positive, Negative

3. Decision making - different stages depending on some condition

4. Reasoning - order of terms to get the required result

Mere counting of numbers is not called a process. Finding errors in documents with the help of computer. Arranging tax etc. is also called process or processing.

Information

The given matter which has been processed is called information. Meaning is a complete fact number or statistical information. In simple words, the meaningful data that is obtained after processing the data is called information. Information is useful material having properties according to different categories. Information is essential and helpful for the following reasons -

1. It presents information in a concise and more meaningful way

2. It helps in taking decisions for present and future.

3. It is helpful in evaluating the future

Properties of Information

We know that information is an essential factor for a system so information should have the following properties

1. Semantic completeness

2. Chastity

3. Accuracy

4. Improving prior knowledge and continuity as much as Possible

5. Conciseness

6. Timeliness

7. Assistant in Editing the Work

Data can be defined as the quantitative or qualitative values of a variable. Data is plural of datum which literally means to give or something given. Data is thought to be the lowest unit of information from which other measurements and analysis can be done.

Data can be numbers, images, words, figures, facts or ideas. Data in itself cannot be understood and to get information from the data one must interpret it into meaningful information. There are various methods of interpreting data.

Data Sources are Broadly Classified into Primary and Secondary Data

Data collection plays a very crucial role in the statistical analysis. In research, there are different methods used to gather information, all of which fall into two categories, i.e. primary and secondary data (Douglas, 2015).

As the name suggests, primary data is one which is collected for the first time by the researcher while secondary data is the data already collected or produced by others.

There are many differences between primary and secondary data, which are discussed in this work. But the most important difference is that—

* Primary data is factual and original whereas secondary data is just the analysis and interpretation of the primary data.

* While primary data is collected with an aim for getting a solution to the problem at hand, secondary data is collected for other purposes.

* The fundamental differences between primary and secondary data are; the term primary data refers to the data originated by the researcher for the first time while secondary data is the already existing data collected by the investigator agencies and organizations earlier.

* Primary data sources include surveys, observations, experiments, questionnaire, personal interview etc. on the other contrary, secondary data collection sources are government publications, websites, books, journal articles, internal records etc.

Primary Data

Primary data means original data that has been collected specially for the purpose in mind. It means someone collected the data from the original source first hand.

* Data collected this way is called primary data. Primary data has not been published yet and is more reliable, authentic and objective. Primary data has not been changed or altered by human beings; therefore, its validity is greater than secondary data.

Secondary Data

Secondary data is the data that has been already collected by and readily available from other sources. When we use Statistical Method with Primary Data from another purpose for our purpose we refer to it as Secondary Data. It means that one purpose Primary Data is another purpose Secondary Data. So that secondary data is data that is being reused. Such data are more quickly obtainable than the primary data.

* These secondary data may be obtained from many sources, including literature, industry surveys, compilations from computerized databases and information systems, and computerized or mathematical models of environmental processes.

Data Acquisition

Data is one of the most important and vital aspects of any research studies. Researchers conducted in different fields of study can be different in methodology but every research is based on data which is analyzed and interpreted to get information. Data is the basic unit in statistical studies. Statistical information like census, population variables, health statistics, and road accidents records are all developed from data.

There are two sources of data collection techniques. Primary and Secondary data collection techniques, Primary data collection uses surveys, experiments or direct observations.

Secondary data collection may be conducted by collecting information from a diverse source of documents or electronically stored information, census and market studies are examples of common sources of secondary data. This is also referred to as "data mining."

Survey

The survey is the most commonly used method in social sciences, management, marketing and psychology to some extent. Surveys can be conducted in different methods.

Questionnaire

The questionnaire is the most commonly used method in the survey. Questionnaires are a list of questions either an open-ended or close-ended for which the respondent give answers. A questionnaire can be conducted via telephone, mail, live in a public area, or in an institute, through electronic mail or through fax and other methods.

Interview

An interview is a face-to-face conversation with the respondent. It is slow, expensive, and they take people away from their regular jobs, but they allow in-depth questioning and follow-up questions.

Observations

Observations can be done while letting the observing person know that he is being observed or without letting him know. Observations can also be made in natural settings as well as in the artificially created environment.

Published Printed Sources

There are varieties of published printed sources. Their credibility depends on many factors. For example, on the writer, publishing company and time and date when published. New sources are preferred and old sources should be avoided as new technology and researches bring new facts into light.

Books

Books are available today on any topic that you want to research. The uses of books start before even you have selected the topic. After selection of topics books provide insight on how much work has already been done on the same topic and you can prepare your literature review. Books are a secondary source but most authentic one in secondary sources.

Journals/Periodicals

Journals and periodicals are becoming more important as far as data collection is concerned. The reason is that journals provide up-to-date information which at times books cannot and secondly, journals can give information on the very specific topic on which you are researching rather talking about more general topics.

Magazines/Newspapers

Magazines are also effective but not very reliable. Newspaper, on the other hand, is more reliable and in some cases, the information can only be obtained from newspapers as in the case of some political studies.

Published Electronic Sources

As internet is becoming more advance, fast and reachable to the masses; it has been seen that much information that is not available in printed form is available on internet. In the past the credibility of internet was questionable but today it is not.

The reasons that in the past journals and books were seldom published on internet but today almost every journal and book is available online. Some are free and for others you have to pay the price.

E-journals: e-journals are more commonly available than printed journals. Latest journals are difficult to retrieve without a subscription but if your university has an e-library you can view any journal, print it and those that are not available you can make an order for them.

General Websites; Generally, websites do not contain very reliable information so their content should be checked for the reliability before quoting from them.

Weblogs: Weblogs are also becoming common. They are actually diaries written by different people. These diaries are as reliable to use as personal written diaries.

Classification of Data

Data classification is the process of organizing data into categories for its most effective and efficient use.

Classification is the way of arranging the data in different classes in order to give a definite form and a coherent structure to the data collected, facilitating their use in the most systematic and effective manner. It is the process of grouping the statistical data under various understandable homogeneous groups for the purpose of convenient interpretation.

There are three different approaches are the industry standard for data classification:

* Content-based classification
* Context-based classification
* User-based classification

Objectives of Classification of Data

* To group heterogeneous data under the homogeneous group of common characteristics;
* To facility similarity of the various group;
* To facilitate effective comparison;
* To present complex, haphazard and scattered dates in a concise, logical, homogeneous, and intelligible form;
* To maintain clarity and simplicity of complex data;
* To identify independent and dependent variables and establish their relationship;
* To establish a cohesive nature for the diverse data for effective and logical analysis;
* To make logical and effective quantification
* A good classification should have the characteristics of clarity, homogeneity, and equality of scale, purposefulness, accuracy, stability, flexibility, and unambiguity.

Classification is of two types, viz., quantitative classification, which is on the basis of variables or quantity; and qualitative classification (classification according to attributes). The former is the way of grouping the variables, say quantifying the variables in cohesive groups, while the latter group the data on the basis of attributes or qualities. Again, it may be multiple classification or dichotomous classification.

The former is the way of making many (more than two) groups on the basis of some quality or attributes, while the latter is the classification into two groups on the basis of the presence or absence of a certain quality.

Data classification, in the context of information security, is the classification of data based on its level of sensitivity and the impact to the University should that data be disclosed, altered or destroyed without authorization. The classification of data helps determine what baseline security controls are appropriate for safeguarding that data. All institutional data should be classified into one of three sensitivity levels, or classifications:

A.	Restricted Data
	Data should be classified as Restricted when the unauthorized disclosure, alteration or destruction of that data could cause a significant level of risk to the University or its affiliates. Examples of Restricted data include data protected by state or federal privacy regulations and data protected by confidentiality agreements. The highest level of security controls should be applied to Restricted data.
B.	Private Data
	Data should be classified as Private when the unauthorized disclosure, alteration or destruction of that data could result in a moderate level of risk to the University or its affiliates. By default, all Institutional Data that is not explicitly classified as Restricted or Public data should be treated as Private data. A reasonable level of security controls should be applied to Private data.
C.	Public Data
	Data should be classified as Public when the unauthorized disclosure, alteration or destruction of that data would results in little or no risk to the University and its affiliates. Examples of Public data include press releases, course information and research publications. While little or no controls are required to protect the confidentiality of Public data, some level of control is required to prevent unauthorized modification or destruction of Public data.

Chapter – 2

Quantitative and Qualitative Data

Qualitative Data

Qualitative data analysis can be summed up in one word - categorical. With qualitative analysis, data is not described through numerical values or patterns, but through the use of descriptive context (i.e., text). Typically, narrative data is gathered by employing a wide variety of person-to-person techniques. These techniques include:

- **Observations:** detailing behavioral patterns that occur within an observation group. These patterns could be the amount of time spent in an activity, the type of activity, and the method of communication employed.
- **Focus groups:** Group people and ask them relevant questions to generate a collaborative discussion about a research topic.
- **Secondary Research:** much like how patterns of behavior can be observed, different types of documentation resources can be coded and divided based on the type of material they contain.
- **Interviews:** one of the best collection methods for narrative data. Inquiry responses can be grouped by theme, topic, or category. The interview approach allows for highly-focused data segmentation.

Quantitative Data

If quantitative data interpretation could be summed up in one word (and it really can't) that word would be "numerical." There are few certainties when it comes to data analysis, but you can be sure that if the research you are engaging in has no numbers involved, it is not quantitative research. Quantitative analysis refers to a set of processes by which numerical data is analyzed. More often than not, it involves the use of statistical modeling such as standard deviation, mean and median. Let's quickly review the most common statistical terms:

- **Mean:** a mean represents a numerical average for a set of responses. When dealing with a data set (or multiple data sets), a mean will represent a central value of a specific set of numbers. It is the sum of the values divided by the number of values within the data set. Other terms that can be used to describe the concept are arithmetic mean, average and mathematical expectation.
- **Standard deviation:** this is another statistical term commonly appearing in quantitative analysis. Standard deviation reveals the distribution of the responses around the mean. It describes the degree of consistency within the responses, together with the mean, it provides insight into data sets.
- **Frequency distribution:** this is a measurement gauging the rate of a response appearance within a data set. When using a survey, for example, frequency distribution has the capability of determining the number of times a specific ordinal scale response appears (ie., agree, strongly agree, disagree, etc.). Frequency distribution is extremely keen in determining the degree of consensus among data points.

The difference between qualitative and quantitative is given below.

Qualitative	Quantitative
1. The qualitative method develops an understanding of the human and social sciences	**1.** Quantitative method is used to generate numerical data from scientific and empirical research method.
2. The qualitative method is holistic in nature.	**2.** Quantitative research is specific.

3. In this method the reasoning used to synthesize the data is inductive	3. The reasoning used in quantitative method is deductive
4. It is Inventor	4. Quantitative is crucial
5. Qualitative method is based on purposive sampling of data where a small group or sample size is selected to get an in-depth understanding of the target concept.	5. Quantitative method relies on random sampling in which large representative sampling method is used for the entire population.
6. Oral data is collected in qualitative method.	6. In quantitative method measurable data is done.
7. Qualitative method remains process oriented	7. Not such a quantitative method
8.	8.

Chapter – 3

Graphical Representation

Data Visualization

Data can be serialized and represented in many ways, for example

1. Table

2. Pictogram

3. Bar Chart

4. Histogram

5. Pie Chart

6. Line Graph

Sometimes the data can be in more than one table, pie chart etc. Their purpose is not only quantitative cell testing. But also to explore comparative and analytical skills.

Main parts of the table

A Table Must Have the Following Main Parts

1. **Title of the Table**

   It is very important to have a proper title of each table. Which should reveal and clarify that what kind of figures are there? At what time? And to which place is it related? Title should be clear, concise and at some places explanatory. The quality of attractiveness is also necessary in the title so that the attention of the reader goes towards it and it does not have to spend much time and effort to understand it.

2. **Table No.**

   Each table should be numbered at the beginning. The table number makes it easy to find a term. When the number of tables is large, these tables should be numbered judiciously. Usually placed above the title. In such a way that it comes to the center of the title.

3. **Subtitle Each**

   A table consists of several columns. The headings in the columns are called subheadings or captions. The subheading should be placed in the middle of the column. There can be many headings under one subheading. When the terms in different columns are measured in different units of measurement, the concerned unit is omitted from the subheading.

4. **Row Title**

   The title of the row is called row title or stn, it is given twice in the table.

5. **Table Cover**

   This is the main and important part of the table. Its size format should be decided in advance on the basis of data. In this section, the data is arranged according to the distribution of subheadings and line headings.

6. **Underline and Spacing**

   Leaving drawing space is also an important part of the table. By leaving proper space and making proper drawing, the table becomes more attractive and effective so that improvements can be made as much as possible and an attractive and clear table can be made.

7. **Foot Note**

   Sometimes comments are considered necessary for the explanation of the figures or words
   Given in the table, then they are given below the table, but as far as possible, the comments
   Should be used as little as possible.

Table No. "Title"

| Line Subheading	Main Column Title			Total
	Sub headline	Sub headline	Sub headline	
Sub-entries		Middle part		
Total				

Bar Chart

A bar chart is a graph that uses bars to show a comparison between categories of data. The bars can be either horizontal or vertical. There are 2 points in the graph. One axis will describe the types of categories being compared and the other will have a numerical value. Which represent the values of the data, it doesn't matter which axis or axes. But it will determine what is shown in the bar graph. If the description is on the horizontal axis, the outside will be oriented vertically. If the values are along the horizontal axis, the bars will be oriented horizontally.

Types of Bar Charts

There are many types of bar charts or bar graphs, they are not always interchangeable. Each type will work best with a different type of fry. What we want to compare helps us determine what type of bar graph to use. First we will discuss some simple bar graph

Vertical or Vertical Bar Chart

A simple vertical bar graph is best. When you do not have to make a comparison between two or more than two independent variables. Each variable will be related to a fixed value and hence can be fixed for horizontal value.
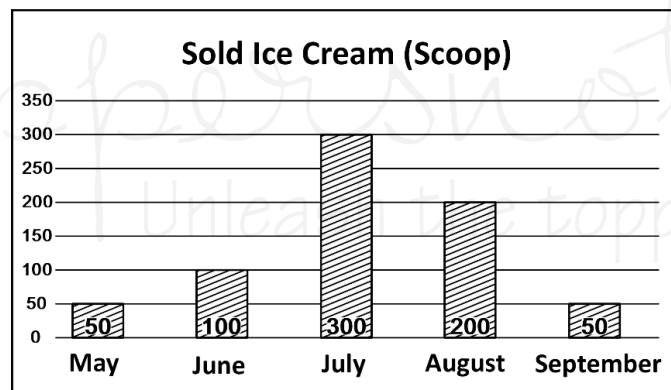
Horizontal Bar Chart

If your data has negative and positive values, but is still a comparison between two or more fixed independent variables, then it is appropriate to have a horizontal graph. The vertical axis can be oriented in the middle of the horizontal axis. Whose negative and positive value can be represented?
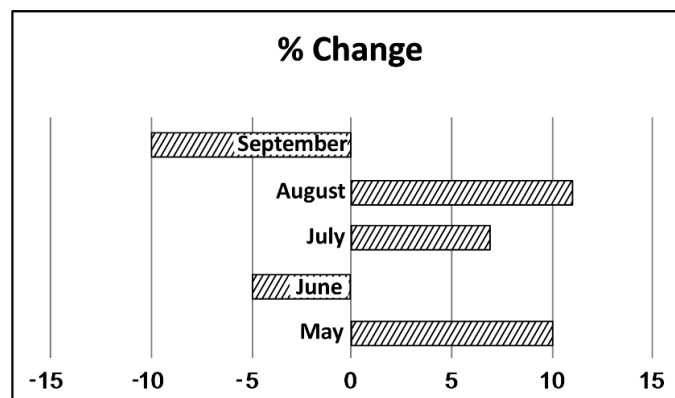
Range Bar Chart

A range bar chart represents a range of data for each independent variable. Temperature ranges or value ranges are common sets of data for range graphs. Unlike the above graph the data does not start from a common listening point. But that particular point starts at a low number for the range of the data. A range bar graph can be either horizontal or vertical.
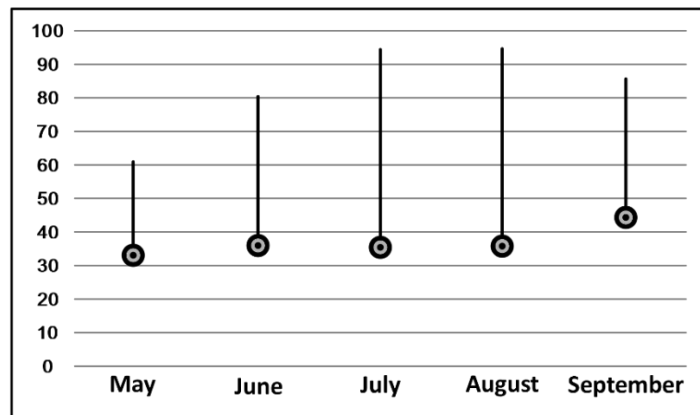
Histogram Data

Statistical information is shown by histogram. The histogram describes the mean between two variables. It is a graphical representation of numerical or histogram data distribution in equal parts. It is a priority distribution of a continuous variable. It was first proposed by Karl Priusen. The histogram shows tabulated frequencies. Which is represented by a rectangle shown on a discrete interval, in which the area of the respective interval is directly proportional to the frequency of observation. The height of the rectangle is also proportional to the frequency density of the corresponding interval, ie the ratio of the shape and the interval width.
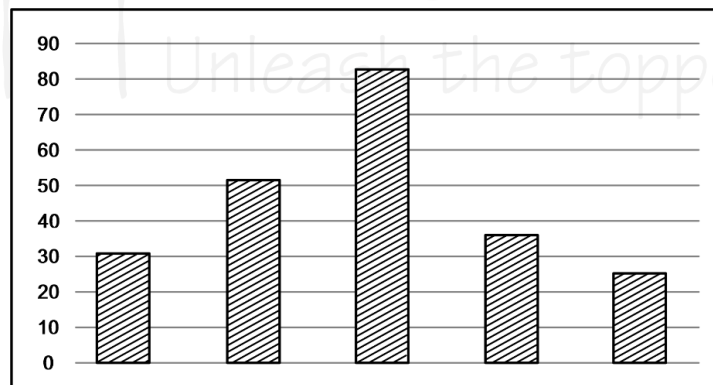
Squar ⟶ Bar Chart
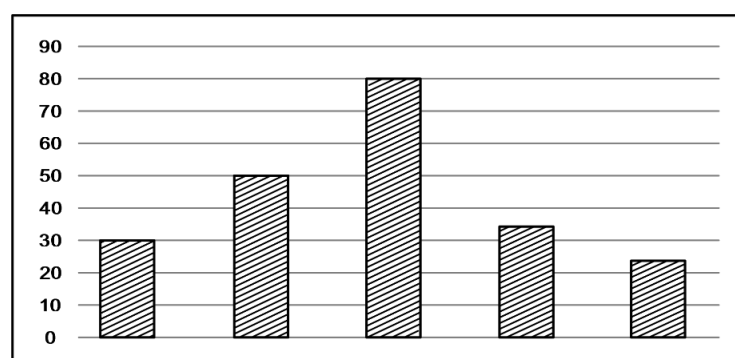
Number ⟶ Histogram

The total area of the map is equal to the total number of figures. A histogram is also displayed by overnormalizing it with the corresponding frequency. Talking about a common form of a histogram, it is shown with the dependent or independent variable on the horizontal line and the dependent or dependent variable along the vertical, these data are shown separately in some color or like a covered moving area.

It is very similar to a bar chart. But there are constant variables in it. The difference between a histogram and a simple bar graph is that each bar in a histogram represents a series of dependent variables, rather than just a single dotted point.

DATA MAPPING

- Data mapping is the process of mapping data fields from a source file to their related target fields. The accessibility to required data can make some organization more successful. Somehow, data is easier to use when it can be visualized as well.

- Visual data help people to understand how different concepts originate and their relation with each other.

- Data mapping helps in all these. For example, "Name, "Email,' and 'Phone' fields from an Excel source are mapped to the relevant fields in a delimited file, which is our destination.

- Data mapping helps by providing organizations with procedure links to show how certain tasks are to be utilized. Forty per cent of our nerve fibres linking to the brain are in the retina only. Data mapping helps us to see what makes different pieces of data useful and helpful

- The customer trends can be traced in the real time The causes of trends and past data numbers can be analysed and other calculations of information and variables can be done. We can also use data mapping software to compare our date with that of competitors. This should make it easier for your business to grow when chosen right.

- They also work by establishing larger maps. Salesforce of any organization has a particularly strong data mapping software program that can be put to use. This helps in real time also. We can get connected to a cloud network to get information in real time.

- Data mapping works for all businesses. For example, if we were in the retail sector, then we can use data mapping to calculate how discount sales can influence the overall sales totals in our business. Similarly, financing, investment type decisions can also be made.

- Data may be internal or external, but it is getting more dispersed and voluminous, then its data leverage is important and actionable insights are developed.

- **In general, data mapping helps with the following activities.**

    o **Data Integration**
       Data mapping tools to cover differences in the schemas of data source and destination, allowing businesses to consolidate information from different data points easily.

    o **Data Migration**
       It is moving data from one database to another. Here, using a code-free data mapping solution that can automate the process is important to migrate data to the destination successfully.

    o **Data Warehousing**
       Data mapping in a data warehouse is the process of creating a connection between the source and target tables or attributes.

- o **Data Transformation**

  It is essential to break information silos and draw insights. Data mapping is the first step in data transformation.

- o **Data Mapping Techniques**

  Although an essential step in any data management process, data mapping can be complex and time consuming. Based on the level of automation, data mapping techniques can be divided into two types and they are as follows.

  1. **Manual data mapping -** Although hand-coded, manual data mapping process offers unlimited flexibility.

  2. **Semi-automated data mapping -** Schema mapping is often classified as a semi-automated data mapping technique. The process involves identifying two data objects that are semantically related and then building mappings between them.

Figure or Data Interpretation

Data interpretation assigns a meaning to the information analyzed and determines its signature and implication. It refers to the implementation of procedures through which data are reviewed for the purpose of arriving at an informed conclusion. A good judgment should be made for its comparative measurement. It includes the following scales.

1. Nominal Scale It consists of non-numerical scale which cannot be ranked or compared quantitatively. Variable is specific and exhaustive

2. Karma Indicative Number or Ordinal Scale. It consists of categories that are other wholes, but with a logical dema. Examples of quality ratings and agreement ratings are Good, Very, Good, Fair, or Agree.

3. Interval scale. Here the data is sorted and grouped into categories on the basis of equal distance. It is always an arbitrary 0 point.

4. Proportion It includes all the above three features while interpreting the data an analysis should try to understand the difference between correlation factor and coincidence etc. apart from some other factors.

DATA AND GOVERNANCE

- Data governance is a requirement in today's fast-moving and highly competitive enterprise environment. Now that organizations have the opportunity to capture massive amounts of diverse internal and external data, they need a discipline to maximize their value, manage risks and reduce cost.

- Data governance is a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals. Data governance defines who can take what action, upon what data, in what situations, using what methods.

- Data governance ensures that roles related to data are clearly defined, and that responsibility and accountability are agreed upon across the enterprise. A well-planned data governance framework covers strategic, tactical, and operational roles and responsibilities.

- While crafting data and governance strategy, we need to be careful.

- **Data Governance is not data management:** Data management refers to the management of the full data lifecycle needs of an organization. Data governance is the core component of data management such as data warehousing.

    1. **Data Governance is not master data management:** Master data management focuses on identifying an organization's key entities and then improving the quality of this data.

    2. **Data Governance is not data stewardship:** Data stewards take care of data assets, making certain that the actual data is consistent with the data governance plan, linked with other data assets and in control in terms of data quality, compliance, or security.

Benefits of Data Governance

An effective data governance strategy provides many benefits to an organization, where it includes the following

- **There is a common understanding of data.**

- **There is improved quality of data, such as data** accuracy, completeness and consistency.

- **Data map is available.**

- **Holistic view:** A 360-degree view of each customer and other business entities basically a single version of the truth'.

- **Consistent compliance:** Data governance provides a platform for meeting the demands of government regulations.

- **Improved data management:** It brings the human dimension into a highly automated and data-driven world.

We can use technology as the enabler for the same.

Open source and cloud are the basic strategies for data governance tools. IPaas is also closely linked with them. These tools also help us achieve the following.

1. Capture and understand our data,

2. Improve the quality of our data.

3. Managing data: With metadata-driven ETL and ELT, and data integration applications.

4. Controlling data.

5. Document our data.

6. Empower the people that know the data best: To contribute to the data stewardship.

7. Protecting sensitive data.

We need to understand that data governance is not optional.

The implementation known as a 'data lake' necessarily requires processes that allow you to keep the data you need in a way that eliminates technical barriers and gives new capabilities to process that data.