Date
09/09/23.

Ch-8

BDA Introduction

Date __/__/____ Lecture-1.

Saathi

Concept of Data, Information & Knowledge:

Data : (i) Data is a source of information.

(ii) collection of meaningless, text, no., symbols which do not convey any contextual relevance/ information is called as Raw Data.

Eg :- 1, 2, 3, 4, Amul, Nestle, ITC etc.

Information :- (i) Data when processed to obtain contextual relevance is called as information.

Information = Data + Context.

(ii) most commonly we take the helps of computers and software packages for computing of data.

Eg : Amul, Nestle, ITC, (Three FMCG companies listed in NSE)

Knowledge :- when information is used for solving a problem, we called it as knowledge

Knowledge = Information + Application.

Eg : Amul, Nestle, ITC, (Three FMCG Co. listed in NSE) (These three co's financial performance should be analysed to understand the Indian FMCG sector).

Art. 1 = 1.5 = Sales / Avg. Stock

= Sales / 1.5

Saathi

Date __/__/____

Nature of Data

Numerical Data

Any data expressed as a number is called as Numerical Data

eg :- Stock price of HUL i.e, 2500.

Descriptive Data

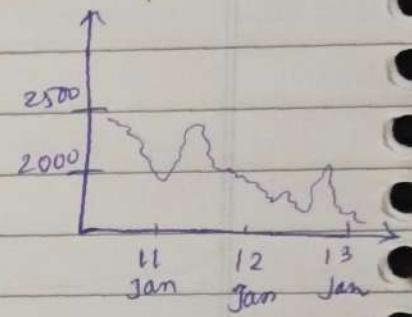
when info. is deciphered in the form of qualitative info. it is called as Descriptive Data.

eg :- HUL's AR may provide a descriptive data abt. its care for stake holder, env, sustainable growth. The data can be used for inv. decision.

Graphical Data

Pictorial representation of data is called Graphical Data.

eg :- HUL stock price may be presented in Graph.



Types of Data in finance and Costing.

Quantitative

Data expressed in numbers.

Eg :- stock price data, Data in Fin. St.

Qualitative

Some data in the financial studies may appear in qualitative format.

Eg :- Text, video, audio.

This info. is useful for understanding the performance of the business.

Eg :- Mgt. discussion & analysis Presented as a part of annual report in the form of text.

Page No.

Date __/__/____

Saathi

Other types of Data (Another way for classifying the types of data).

Nominal Scale

Used for categorisation of data based on certain characteristics. The category may contain numbers but may have no numeric value.

Eg:- Classifying Eqp. into large, medium, small, cap.
- large → RIL
- small → ECL
- cap. → Zomato

Ordinal Scale

It is used for classifying data and put it into order. It do not specify how much better or worse a stock is compared with others

Eg:- Top 10 stock by PE Ratio

Interval scale

It is used for categorising & Ranking using an equal interval scale.

Eg:- Temperature scale. The issue is 0°C does not indicate absence of temp.

Ratio Scale

It possesses characteristic of Nominal, Ordinal, Interval scale. A Ratio scale has a true 0 which means 0 has a significant value.

Eg:- Length, time, money, age.

Ratio scale is used to measure sales, pricing, market shares

Saathi

Date __/__/____

Q. Differentiation between four scale type is based on 3 characteristics. what are they ?

Ans (a) whether the sequence of answers matter or not.
(b) whether the gap between observations is significant or interpretable.
(c) The existence or presence of a genuine zero.

Date 18/09/23

Lecture-02.

Digitization of Data & information.

Q(1) What is digitization ?

Ans (1) It is the process of converting data & information from analouge to digital format.

(2) Data in the original form may be stored as an object, document, image, which has to be converted in the form of binary nos that facilitate processing using computers.

(2) Objective of Digitization

(1) Wide access of data to large group of users simultaneously.

(2) Data preservation for longer period.

Eg :- Aadhar Card / UID - govt. digitization

Page No.

Saathi

Date __/__/____

(3) Advantages of Digitization

(1) Improves classification & indexing for documents which helps in easy retrival of records.

(2) Records can be used by many users simultaneously

(3) Data reuse becomes easier. eg :- large map, micro films.

(4) Helps in work processing.

(5) Easy to back-up the files & data retrival during any unexpected disaster.

(6) Requires less storage space.

(7) Increase organisations productivity.

(8) Can be accissed from multiple location.

(9) Higher integration with busines information systems

(4) Methods of Digitization.

(A) Justification of the proposed digitization project.
cost benefit analysis.
Risk assessment.

(B) Assessment
All records cannot be digitized. Identify the format, files, tables, & documents etc., hardware, software requirement. Risk Assessment (natural disaster, cyber attack)

Page No.

Date __/__/____

Saathi

(c) Planning
- selection of the digitization approach.
- Project documentation, Resource mgt., technical specification & Risk management
- decision of inhouse or outsourcing.

(d) Digitization Activities.
Planning, Capture, Primary quality control, Editing, Secondary quality control, & storage and management.

(e) Processes in the case of records.
Defining the access right, Intellectual control, upkeeping & maintenance of data.
(CIA of the data) (Confidentiality, Integrity & Data)

(f) Evaluation.
To identify changes for any improvement.

Date __/__/____

Saathi

Big Data.

Big e-commerce companies like Amazon, flipkart & conglomerate companies like Reliance have huge database which has to be managed by them, so innovative technology driven software like ERP, OLAP, etc. has to be used.

Top level

Decision making & Risk management by CFO.

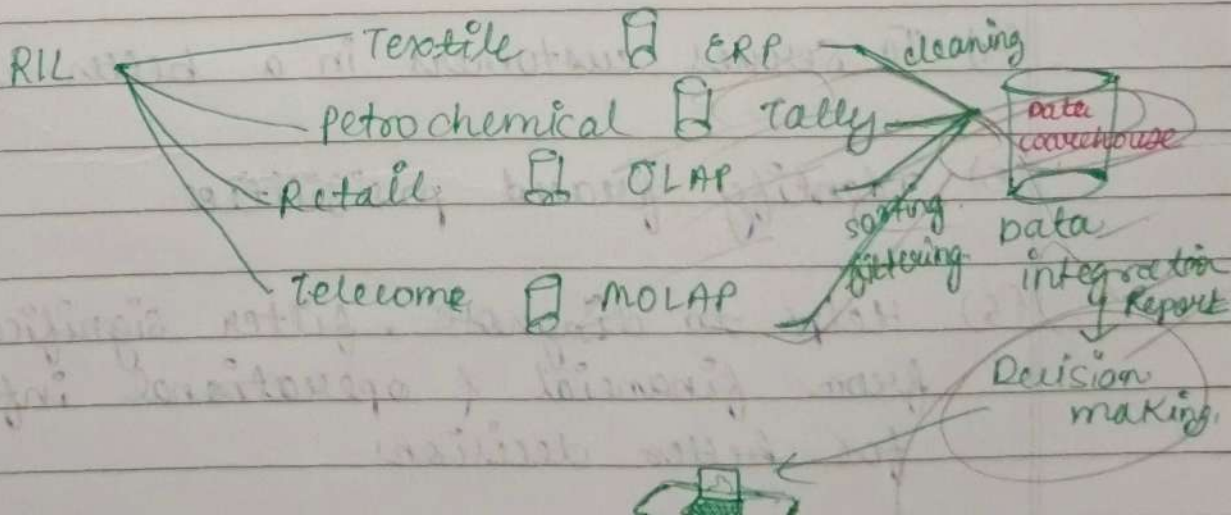
middle level.

Financial reporting & controlling
eg :- CFs, BIs.

Foundation level

Transactions processing and bookkeeping using Tally / ERP software.

Steps of Data transformation



Saathi

Date __/__/____

① collection of data
② Organising of data
③ Data processing.
④ Integration of Data
⑤ Data Reporting
⑥ Data utilization.

Date 08/10/23

Lecture-03.

Q What are the advantages to the firm ~~for~~ for transformation of information into decision. ?

(1) Helps to predict outcomes more effectively than the traditional forecasting techniques

(2) Real time spotting of emerging opportunities

(3) Making strategies for responding to uncertain events & black swan events

(4) To serves customers in a better way.

(5) Identify fraud possibilities.

(6) Helps to diagnose, filter significant values from financial & operational information for better decisions

(7) Building useful dashboards to measure success leading to effective strategy.

(8) Helps in corporate planing & budgeting and forecasting.

Page No.

Saathi

Date __/__/____

Write a short note on Professional Scepticism.

① Professional Scepticism is an important focus area for practitioners, researchers, regulators and standard setters, auditors & manager.

② It may result into additional cost and strained client relationship.

③ A deeper understanding of hidden insights need to be explored so managers should never take an important analysis at its face value.

④ Major concern about the use of DA is the likelihood of false positive which has to be eliminated, therefore nowadays new DA tools (OLAP, MOLAP) are being used in the financial environment to get unique insights in to the data.

Page No.

Ethical Use of Data & Information.

What are the 5 Principles of Data Ethics that a business organisation should follow?

Ans: (1) Regarding ownership :- It is unlawful & unethical to collect someone's personal data without their consent, and the consent may be obtained through digital privacy policies or signed agreements.

(2) Regarding transparency :- The objective of the company is that for the purpose for which user's data is being collected should be known to the users. Eg :- many companies are using "cookies" to track the online behaviour of the users.

(3) Regarding privacy :- The user's data (PII) Personal Identifiable Information, cannot be made publicly available. ~~Many~~ companies have to publish annual report which are more public, so to protect the privacy of the data there should be a data security process like 'file encryption' & "dual authentication" password, etc.

(4) Regarding intention :- The intention of the DA should never be making profits out of other weaknesses. this is unethical & should be avoided.

(5) Regarding outcomes :- Even if the intention are good the result of data analysis may sometimes hurt the clients & this is called disparate impact, which is unethical and should be avoided.

Lecture-04.

Ch-3.

Data Processing, Organisation, cleaning and validation.

Data processing (DP) & its types.

Data processing (DP)

(1) It means organising, categorizing, manipulating data in order to extract meaning info. from it.
(2) Now a days data processing is done by modern tools & technology.
(3) Info. extracted by DP is also heavily reliant on the quality of the data. The data quality
(4) may get affected by missing data, duplication.

Saathi

Date __/__/____

Types / History of DP.

- Manual DP
  - (i) It involves DP without machines assistance.
  - (ii) Suitable for small-scale DP involving manual efforts eg:- Retriving data from ~~out dated text~~ or documents
- Mechanical DP
  - (i) DP done through mechanical tools (but not modern computers).
  - (ii) It began in 1890 when Punch card machines were developed by US Bureau.
  - (iii) It will be quicker than the manual Process.
  - Eg. Calculator, Scanner.
- Electronic DP
  - (i) It involves DP done electronically using sofisticated, modern computers and electronics.
  - (ii) It has replaced the other 2 DPs methods resulting in low mistakes and highly productive.
  - (iii) It is highly / widely used in research, industries, institution, e-comm companies, Bank, etc.

Saathi

Date __/__/____

How data processing and data science is relevant for finance ?

Ans (A) Risk analytics.

(i) Business involves many risks and most of them are data related risk which is unstructured. Once a danger has been recognised precautions may be taken to avoid it in future. ∴ machine learning algorithm can be used to increase the trustworthiness & real time risk assessment.

(B) Real time Analytics.

Generally data analysis is done on historical data however with technological advancement real time analytics are available for eg :- machine learning, Business Intelligence, Data Science, Data engineering to provide optimal experience.

(C) Customers Data management (Internet of Things)

Big data can easily be processed by using IOT devices. we can easily identify data patterns, transaction history to eliminate sub-zero consumers, enhance cross-sales and calculate ~~contu~~ costumer's life time worth.

(D) Customer segmentation.

Automated machine learning algorithm can be used to categorize their clients based on specified attributes (like age wise, geographical area) with this info. organisation may eliminate clients who little / no value.

Saathi

Date __/__/____

(E) Personalised services

By delivering the product that met their needs, cross selling, all these are possible beoz of data processing. Now a days NLP (Natural Language processing) are used to enhance customer's service experience.

(F) Advanced customer Service

(Cross selling, up selling, adressing client's enquiry)

(G) Predictive Analysis

It helps the organisation in the financial sector to anticipate what may occur in the future. for this machine learning techniques are utilized for eg :- in stock market pricing ML is used the future to forcast the future stockmarket trends based on the historical info.

(H) Fraud detection

Eg :- AI & ML algorithms are used to detect credit card and fraud detections, flaging of suspicious transaction, etc.

(I) Anomaly detection.

It means detecting abnormalities in customers bank a/c activities. Eg. Activation of some deormant a/c, unlawful insider trading. The method for anomaly identification are Recurrent Neural Networks & long short-memory models.

Saathi

Date __/__/____

(J) Algorithmic trading.

Algorithmic trading happens when an unsupervised computer utilising the intelligence supplied by an algorithm trade suggestion on the stock market. It is devised based on stringent rules on the basis of which trading is done in a specific market, this method is also known as Reinforcement learning, in which model is taught using penalties & rewards associated with the rules.

Advantages

(1) frequency of deal increases
(2) human indecision can be avoided.
(3) Also, machine will only trade when it precises percives the opportunity.

(K) Customer's Analytics.

same as C

Saathi

Date __/__/____

Steps involved in Data Processing.

(1) Data validation.

(a) Data validation could be operationally defined as a process which ensures the correspondance of the final data with a number of quality data (Acc. to Simon)

(b) Data validation leads to acceptance or rejection of the data.

(c) The objective of data validation is to assure a particular degree of data quality (Relevance, correctness, clarity, coherence, comparibility)

(2) Sorting

(a) It refers to arganising the data into a meaningfull order to make it simple to understand, analyse and visualise.

(b) Sorting can be performed on raw data, aggregated data, summarised data.

(c) Data can be sorted in ascending & descending order with the help of various software

(d) Sorting is frequently used to rank the records.

for eg :- Customers ranking is done on the basis of highest volume accounts or heavy usage customers

(e) It can be done to various visualising techniques like tables, charts, etc.

Page No.

Saathi

Date __/__/____

(f) DBMS (Data Base mgt System) is also used for data sorting

(3) Aggregation

(a) Data aggregation refers to data collection and data summarising.

(b) Groups when data is aggregated individual data rows are compiled together and replaced with summaries.

(c) A data warehouse often contains aggregate data and are used for analytical enquires.

(d) Software tools like data aggregator can be used for aggregating the data

(4) Data Analysis.

(a) Data analysis is a process of cleaning, converting and modelling the data to obtain actionable business intelligence.

(b) Data analysis means future anticipation based on past studies.

(c) Data analysis is done to expand busines & finance.

(d) modern tools, like SQL, JAVA, PYTHON, SAP, & software ERP, SAS, etc can be used for data analysis.

Page No.

Date 16/10/23

Lecture -05

⑤ Reporting.

(a) Data reporting is the act of gathering & structuring raw data & turning into consumable format in order to evaluate organisation's continous performance.

(b) We can use excel documents or simple data visualisation tools.
Eg. of reports are - Fin. data contained in P/L, B/S, CF st., etc.

(c) A data report indicates ~~the~~, where we should devote the most time and money.

(d) There is no space of error in data reporting, therefore it has to be prepared ~~in~~ with high attention.

(e) Although static reporting can be helpful but has certain limitations which is the absence of real time insights.

(f) If data reporting system is effetient it facilitate the formation of judicious judgements that might provides additional income streams.

⑥ classification

(a) It is a process of classifying data according to important categories so that it may be utilized and safeguarded more effectively.

(b) classifying data entails labelling it to make it searchable & trackable. Also it avoids duplication of data, which minimizes storage and backup expenses

(c) Data classification is done on the basis of sensitivity of data in order to ensure ~~e~~ CIA of the data is been maintained.
(Confidentiality Integrity Availability)

Q. What are the three primary methods of data classification as recognised by Industry standards?

Ans: classification can be done based on.

① Content, Sensitivity of the data.

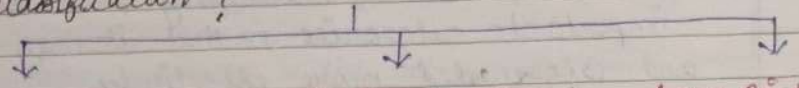
② Context based classification. (characteristics, application, location & creator)

③ User based classification (It relies on human selection of each documents by the end user)

Saathi

Date __/__/____

Q) What are the 3 categories of risk for data classification ?



High Risk

Any data which is very sensitive and critical to the operations of the co is called as High Risk data. Once such data if lost is difficult to retrieve.

Eg :- Secret data, Sensitive, Essential data.

Moderate Risk

Data which are internal i.e, pertains to business or its partners and are non-public is called Moderate Risk data.

Eg :- Cost of products, Op. process, certain corporate paper works

Low Risk.

Public Data whose recovery is simple & there data collection & usage possesses smaller risk than the other.

Saathi

Date __/__/____

Data classification matrix.

Basis	Confidential data (High Risk)	Sensitive data (moderate Risk)	Public data (Low Risk)
① Description	Highly Sensitive & personal data whose CIA cannot be compromised. Access to confidential data must be controlled from creation to distruction & should be given to those individual who are required to perform their job ~~are~~ permitted by Law. Access to such data may be individually requested and then it is authorised by security module Representative.	Internal / Non-public data which are moderately sensitive such data are required ~~for~~ decisions therefore it should remain timely & accurate Access to this data is restricted to the persons based on ~~RBAC~~ RBAC rights. (Role Based Access Control)	Not sensitive & may be granted to any requester ~~who~~ however its integrity has to reserved & representative must authorise replication or copying of the data.
② Access.	Individuals, designated with approved access.	Employees & non-ees based on RBAC	Public & need to know basis & EMU affiliates
③ Impact	The the impact if the data is incorrect on the organisation is very high	moderate	low

Example of data classification
↳ Restricted, Private, public

Steps involved in data classification process

(1) Understand the current setup
(2) creation of data classification policy. (Data Protection standards are laid down)
(3) Priorize & organiz data.

Data Cleaning

Data cleaning is a process of correcting or deleting inaccurate, corrupted, duplicate, insufficient data from a data set. Incorrect data renders outcomes & algorithms untrustworthy, so it is important to apply data cleaning which removes irrelevant data from data set.
It is different from data transformation. Data tranformation involves changing data from one format to another format. (Data wrangling / data munging)

Steps for Data cleaning

(1) Removal of duplicate and irrelevant information
(2) Fix the structural error (typo error, wrong capitalisation, unusual naming standards etc)
(3) filter unwanted outliers.
Occasionally there may be some observation that at first look to fit inside the data that you are evaluating. If we have a valid cause to eliminate an outlier then it will improve the performance.

(4) Handling the missing data - Missing values can be droped or may be entered based on the observation
(5) Validation & QA (Quality Assurance)
(a) We may check that does the data make sense.
(b) Is there any issue with the quality
(c) Does the data adheres to the regulations.

** Main characteristics of the Quality data
(VACC)
(i) Validity (ii) Accuracy (iii) Completeness (iv) Consistency

Benefits of Data cleaning

(1) Error correction
(2) fewer mistakes results in happy customers & less irritated workers
(3) Monitoring mistakes & helps in improving reporting.
(4) Effective corporate procedure & speedier decision-making.

Date 18/10/23

Lecture-06.

Date __/__/____

Data Validation

① It is a process of collecting information, evaluating data and preparing them to deliver data to the stakeholders.

② Data validation is very essential as it leads to precision, clarity & specificity of data is essential for mitigating project failure.

③ Using this validation criteria, it helps to mitigate in garbage in, garbage out problems & ensuring data integrity.

Types of Data Validation.

① Data type check - It verifies that the entered data has the appropriate data type, for eg:- we have specified a field check which can take only numerical values this means that the system will reject data containing letter, charatcrt showing "invalid" data entries.

② Code check - It verifies that the field value is picked from a legitimate set of options.

③ Range check - It verifies that the input data should fall within a specified range.
eg:- latitude & will under -90° to 90° & longitude -180 & 180° any outside range will show invalid data entries.

Date __/__/____

④ Format check - It ensures numerical data should comply with certain format.
Eg:- DD-MM-YYYY any data outside forwat will show invalid entries

⑤ Consistency check - It verifies that the data has been input in a consistent manner.
Eg:- checking package delivery ~~data~~ date is later than the order date or not.

⑥ Uniqueness check - It checks whether unique items are added into the database or not.
Eg:- PAN, email Id, etc.

Also a drop down menu or auto completion form can be used as code checks for data validation process.

Saathi

Date __/__/____

Data Organisation & Distribution

1. It refers to the process of arranging unstructured data into distinct group in a meaning full manner
2. Eg. of Data organisation techniques are
   - frequency dist. table, graphical representation, image repr., histogram, pie-chart, bar chart etc.
3. It is done for understanding and manipulation of the data
4. Structured data consist of tabular info. that can be readily imported into a database & utilised by analytic software.

Data Distribution

1. It is a function which identifies and quantifies all potential values for a variable as well as their relative frequency.
2. The primary benefit of data distn. is the estimation of prob. of any certain observation within a sample space.

Saathi

Date __/__/____

Types of Data Distribution

Discrete

It results from countable data and has a finite no. of potential values. They can be displayed in table.
Eg:- rolling dice, toss of a coin.

Types

1. Binomial dist.
   Tossing of coin and is applicable for mutually exclusive & exhaustive classes
2. Poisson dist.
   It applies to attributes that can take huge values but in reality takes only tiny values.
   Eg:- no. of accident, mistakes, etc.
3. Hyper-geometric dist
   It is similar to Binomial dist. but unlike the binomial dist. here the chance of success is same for all the trials
4. Geometric dist.
   Eg:- A marketic representative from an advertising firm choose hockey players only (Negative Binomial dist.)

Continuous

It has unlimited no. of data points that may be represented on a continuous measuring scale. It is described using probability density function. Clustered frequency dist. is seen.
Eg:- weights, Height.

Types

1. Normal dist. (Gaussian dist.)
   It is a bell shaped curve. For symmetric dist mean = mode = median
2. lognormal dist.
3. F dist. table
   (It is a asymmetric dist with no maximum value or min. value of 0)
4. chi square dist.
5. Exponential dist.
   (It is closely connected with the poisson dist.)
6. T dist. method.
   Student

ate 9/10/23

Ch-10.

Lecture-07

Saathi

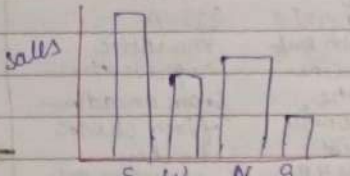
Date __/__/____

Data Visualisation.

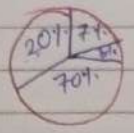
(1) A picture speaks a 1000 words.
(2) Recent study reveal that top-performing finance directors are more likely than their peers to emphasise data visualisation abilities.
(3) It has the capacity to explain complicated ideas, identify info. linkages, providing captivating narrative resulting from data not only a elevates finance position in strategic decision making but also democratises data throughout the business.

Tools / Techniques of Data Visualisation.

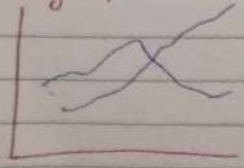
(a) Bar Chart.



(b) Pie Chart



(c) Time series graph.



standard to actual comparision

Page No.

Saathi

Date __/__/____

Q. Why Data Visualisation is important?

Ans (1) Scott Berinato (Specialist of Harvard Business Review) writes in a recent post that DV which was once a talent of design manager have become indispensable for all the managers who wish to analysis and communicate the significance of data.

(2) According to a research 65% of the individuals are visual learners.
DV is used in finance to communicate its findings to the entire business.

(3) It aids finance in identifying move patter and gaining deeper insights.

Q. How can we do Data Visualisation in the right way?

Ans (1) Know the objective - Before the development of great images, one must first grasp the objectives. HBR's Berinato suggests, first establishment of the information if its conceptual or data-driven (i.e. does it rely on qualitative or quantitative data) is required.

(2) Always keep the audience in mind - who views the data visualizations will determine the degree of detail required.
eg:- C-Suite managers like CFO, CEO, if any visualisation charts are made for them then it should be in very detailed for decision making.

Page No.

Saathi

Date __/__/____

(3) Invest in best technology - There are a multitude of technological tools that make it simple to produce engaging visualizations in the current digital age.
Eg :- ERP software can be used which maintains centralised database & avoids data deplication.

(4) Improve the team's ability to visualize data - Most of the agile finance directors rank their team data visualization abilities as good, compared to only twenty four percent of their counterpart, according to an AICPA survey

Objectives / Functions of Data Presentation.

(1) Making a better data analysis.
(2) Faster decision making
(3) Analysing complicated data.

However DV assists specialist remaining informed and composed despite the vol. of data. Further the objectives can also be summarized in the following :-
(i) Effective communication for audience.
(ii) Interprets large data volume more quickly
(iii) Deeper insights available to the decision maker to facilitates decisions. analysists.

Saathi

Date __/__/____

Q. According to an article published by HBR, what are the most common errors made by analysts that makes DV unsuccessful?

Ans Common error are as follows :-

(i) Understanding the audience.
(a) who will read and understand the material & how will they do so? (i.e, a specialist audience or a broader public)
(b) what are the expectations of audience?
(c) what is the functional role of visualization? A visualisation is exploratory should leave viewers with questions to investigate, but if visualisation are confirmatory should not.

(ii) setting up a clear framework.
(a) The designer must guarantee that all viewers have the same understanding of what the visualisation represents. They must use semantics which are simple to understand.

(iii) Telling a story.
Visualisation must give a viewers a story, so it should be presented into a narrative medium specially when the tale is comprises of large no. of data.

Saathi

Date __/__/____

Data Presentation Architecture (DPA)

(I) DPA is a set of skills that aims to identify, find, modify, format the data in a manner that conveys the meaningful insight.

According to

(II) Kelly Lautt DPA is a rarely applied skill set critical for the success and value of intelligence.

Objectives of DPA

(i) Utilise the data in the most efficient, feasible method which will convey timely & comprehensive data to audience. Utilize data to deliver data effectively & feasibly (minimize noise, complexity).

Scope of DPA

(i) Defining significant meaning required by each audience.
(ii) Obtaining the proper data.
(iii) Determining the needed frequency of data
(iv) Determining the optimal presentation.
(v) Using suitable analysis, categorising style
(vi) Developing appropriate delivery techniques based on job, duties, location.

Lecture-08.

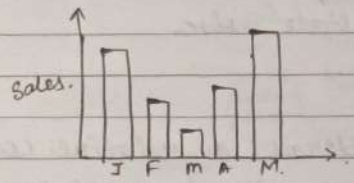
Saathi

Date __/__/____

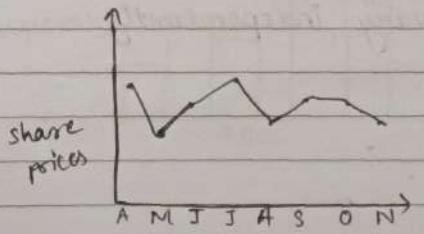
Q. What is a data Visualisation Dashboard ?
It is an interactive dashboard that enables to manage important matrix across numerous financial channels, visualise the data points and generate reports for customers that summarizes the results.

Q. What are the different types of DV techniques?

(1) Bar chart
It is the most commonly used types of DV. Used for comparing data across categories, illustrates historical highs & lows.



(2) Line chart/graph.
It joins various data points displaying them as a continous progression. It helps to observe trends in data over time. It is used for showing price fluctuation in the stock market or sharemarket.



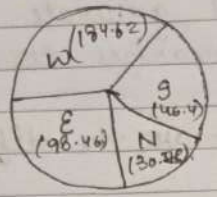
Saathi

Date __/__/____

(3) Pie chart / circle chart

Is a circular graphical representation of the statistical data which is segmented to demonstrate numerical proportion.

Direction	Sales (in Lac)
N	100
S	150
E	320
W	600



(4) Map

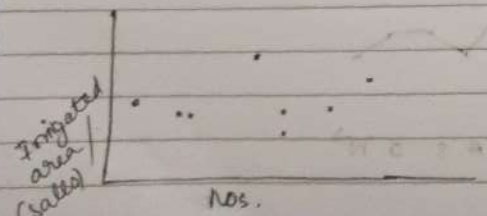
For displaying any types of locational data maps are used. Eg. postal codes, state abbreviations. A map can be used to show the average of GDP state wise.

(5) Density Map

It indicates patterns or relative concentrations. It allows to deal with larger data sets in a limited geographical area.

(6) Scatter plots

Scatter plots help to examine the connection between many variables revealing whether one variable is a good predictor of another or they tend to vary independently.



Saathi

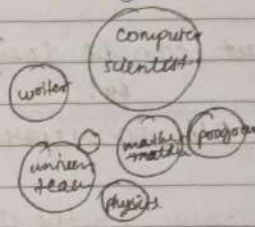
Date __/__/____

(7) Gantt chart

It represents a project's timeline, ie, time taken by each activity to accomplish. Also it states which activities must be completed in advance so that other activities can start (predecessor & successor). Also state allocation of resources.

(8) Bubble chart

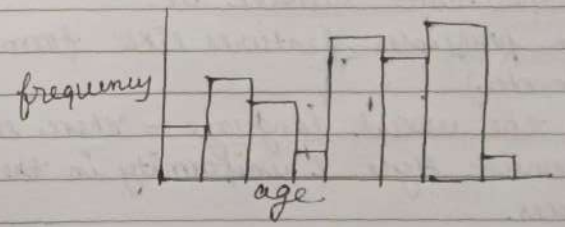
By using different size & colour of the circles. Enormous data can be displayed in an aesthetically engaging manner.



proportions of professions of people who generate Programming languages.

(9) Histogram

It illustrates distribution of data among various groups. It divides the data into discrete categories called as bins.



(10) Tables.

Tables also known as "crosstabs" or "matrices".
Table arranges the data along X-axis & Y-axis
Here, X (Row), Y (columns).
Tables may arrange qualitative data & show their linkages, it is one of the most prevalent method of showing the data.

	Name	Rollno	class.
1)			
2)			
3)			

Q How a report can be designed using a DV?
OR.
How to use 'DV' in report design?

Ans (A) Find a story in the data.
(B) Create a narrative - This means to compose an excellent story one must have a
- catchy title & sub-heading.
- Incoporate the content
- Hilights & significant discoveries.
(C) Choose the most suitable DV.
(Design principles, features like form, size color, etc.)
(D) Follow the visual language - there should be a consist style & uniformity in the DV Principles.
(E) Publicize the report.

Tools & Techniques of Visualisation & Graphical Presentation.

Tableau	Microsoft Power BI	Microsoft Excel	Qlik. view
It is a DV software for creating interactive graphs, charts, map from the source data & within few minutes it is capable of producing static visualisation which can be publish in more than one pages & can't produce interactive map. Eg :- Tableau Public	It is a DV application software for Business Intelligence Data. Reporting, Self serving analytics & predective analytics are supported.	It provides various options for viewing the data such as pie chart, bar chart, histogram, etc. It helps to illustrate the relationship between 2 or more Data sets i.e, wished to compare. It helps in examining, statistical, scientific data for market research.	It is a data discovery platform that unables the users to make quicker & more informed choices by speeding up the analytics. It includes various types like color-coded tables, bar graph, line chart, etc. It has been designed using a "drag & drop" interface.

Date 20/10/23

Lecture-09. Ch-11.

Data Analysis & Modelling

Date __/__/____

① Data Analysis is the science of evaluating unprocessed Data sets to draw conclusion about the info they contain

② Applications like machine learning (ML), AI, BI, Simulation, etc may be used for DA.

Process of Data Analytics.

(1) Criteria for Grouping data.
Data can be classified on the basis of various parameters like, age, income, sex, etc.

(2) Collecting the data
(3) Organizing the data.
(4) Cleaning the data
(5) Adopt the right type of data analytics process -
(i) Descriptive (patient telling the symptoms).
(ii) Diagnostics (why it has occured)
(iii) Predictive (what can happen in futures)
(iv) Prescriptive (what can be done).

Date __/__/____

Benefits of DA.

(1) Improves decision making process.
Used DA gained info. resulting in enhanced outcomes. Using DA reduces the amount of guesswork involved in preparing marketing plans, deciding what materials to produce, and many more.

(2) Increase in efficiency of operations
minimizes the wastage of resources, increasing profitability, better understanding of their audience's demands.

(3) Improved service to Stakeholders.
(customers, Employees & other stakeholders)

Data Mining / Kdd (Knowledge Discovery in Data)

It is the extraction of patterns & other useful info. from massive data sets. In other words it means identification of unknown trends patterns between the variable for enhanced corporate decision making we can use modern technologies like Data warehousing technology, Big Data, AI, BI,.

Saathi

Date __/__/____

What are the 2 primary process on which Data mining techniques can be categories?

Ans (1) Describing the target data sets or predicting resulting using ML algorithms, it helps to detect fraud, users habit, bottle neck.

(2) It can also be paired with a DA & visualisation technologies like, Apache spark.

Process of DM.

(a) Setting the business objective
(b) Preparation of Data.
(c) Models building & patter mining.
(d) Result evaluation & implementation of knowledge

Techniques of DM.

- Association Rule
- Neural Network
- Decision tree
- K-nearest Neighbour. (KNN)

Association Rule: Discovering Association between variable in a given data set. Eg:- Analyse the client's consumption pattern for developing cross-selling tactics.

Neural Network: used for deep learning. It replicates the interconnection of human brain. Every nodes has inputs, weights, threshold, output. If the output exceeds threshold then the node fires & passes data to subsequent network for analysis.

Decision tree: It classifies or predict outcomes based on collection of decision, it employs a tree-like representation

K-nearest Neighbour: It classifies data points depending on there closeness & correlation with other accessable data. It attempts to measure the dist. between the data points

Saathi

Date __/__/____

Implicati Implementation of DM in Finance & Management -

(1) Detecting money laundering & other financial crimes.

Money laundering is the illegal conversion of black money to white money. Most of the banks are unknowingly involved in it. So proper data mining methodology will help the banks to detect money laundering activities.

(2) Prediction of loan repayment & customer credit policy analysis.

The loan prediction system automatically computes the size by examining data pertaining to the size. In other words it helps to access whether the loan should be sanction for amount of loan, etc.

(3) Target marketing.

It helps to target certain market By knowing about your customer's market, etc.

(4) Design & construction of data warehouse -

If all the data is stored in a data warehouse, it becomes easier for Data analysis.

Lecture - 10.

Saathi

Date __/__/____

Descriptive Analysis

(1) It focuses exclusively ~~focuses on ex~~ on what has occured in an organisation unlike other type of analysis.
(2) It is the basic starting point to prepare data for subsequent analysis.
(3) It is the simplest ~~king~~ of DA as it employs simple mathematical & statistical methods.

How does descriptive analytics work?

According to Dan Vesset, there are 5 steps.

step (1) : Decide the business metrics :
(Identification of KPI's)

step (2) : Identification of Data requirement :

Step (3) : Preparation & collection of data

Step (4) : Analysis of the data.

Step (5) : Presentation of the data.

Saathi

Date __/__/____

Advantages & disadvantages

Advantages

(1) As it is based on historical data it is easily applicable to day to day operations.
(2) Firm may report on performance very quickly and changes can be made.

Examples of descriptive analytics :-

(1) Past events, such as sales & operational data or marketing campaigns are summarised.
(2) Reporting general trends
(3) Compiling survey data.
(4) Social media usage such as instagram, fb.

Diagnostic Analytics

(1) It highlights the tools which are employed to question the data "why did this occur?"
(2) It involves a deeper & thorough examination of the data to discover important insights.
(3) It includes data discovery, drill down, data mining & correlations.
Data mining is the extraction of the unstructured data.

Advantages

(1) It helps to make most of the data by turning it into visuals & insights that can be ~~from~~ utilized by everyone.
(2) Derive value from the data by indepth analysis.

Example :-

(1) HR dept. can use it to examine the performance of their employees, utilising Python, R language for indepth examination
(2) cyber security

(2) cyber security is another example. The cyber security team may determine the relationship between security rating & the no. of incidents & also average time for the resolution time, etc.

Predictive Analysis

(1) It focuses on forecasting & understanding what might occur in the future.
(2) By analysing the past trends / historical info. it is possible to predict what can occur in future.

How does it work ?

(a) The foundation of predictive analysis is probability.
(b) Using techniques like DM, statistical modelling, ML algorithms, predictive analysis to predict the future outcomes.
(c) Deep learning which is the most recent subfield of machine learning, which is trying to imitate the building of human brain network.
(d) This method enable the executive & managers to take a more pro-active approach to corporate planning.

Advantages :

(a) Effectiveness increases (eg :- inventory forecasting)
(b) Better customer services by gaining a deeper knowledge of who its client are what they want.
(c) Detection & prevention of fraud.
(d) Risk mitigation in the financial sector

Examples

(1) Ecommerce, using the search history & previous purchase info. client's behaviour can be predicted
(2) Sales -
(3) Human resource - identifying ~~ee~~ who have to retained & eliminate.
(4) IT security (5) Health care industry.

Prescriptive analytics

(1) Descriptive analytics talks about "what has occured".
(2) Diagnostic analytics talks about "why it occured".
(3) Predictive talks about "what could occur".
(4) & Prescriptive talks about "what best could be done"
(5) It is the fourth, final, most sophisticated step of business analysis process.
(6) It urges the executive, managers, personnel in making the most informed decision based on the available data.

Saathi

Date __/__/____

How does it work?
It uses different tools like statistics, ML algorithms to access internal data & external data.
Predictive analysis is different from ML.
Predictive analysis uses historical data & statistical techniques to predict about future.
whereas ML is a subset of AI which uses an enormous data and is self learning to analyse data and exhibit intelligence human.

Advantages
It gives important insights for making most optimal data driven decision.

Disadvantage
(1) This technique requires enormous volumes of data which is not always possible.
(2) It cannot account for all the external variables.

Examples
(i) Oil & manufacturing for monitoring price fluctuations.
(ii) Mfg industries for enhancing equipment maintenance, prodn, storage.
(iii) Health care industries enhancing patient care, evaluating customers risk in terms of price & premium info.
(iv) pharmaceutical research

Page No.

Lecture-11.

Saathi

Date __/__/____

Cloud Computing

Use of any computing services via network typically by internet is called as cloud computing.

Eg:- Google is giving us bundled services.

Email, G-meet, G-form, G-map, GPay, gdrive

Above examples can be categorised as -
(i) Infrastructure as a service (IAAS)
(ii) Platform as a service (PAAS)
(iii) Software as a service (SAAS).
(iv) Communication as a service
(v) Security " " "
(vi) APIS.

Now a days most of the businesses are relying on cloud computing for reducing ~~their~~ cost, maintenance staff, physical IT resources.

Page No.

Saathi

Date __/__/____

Types of Cloud Computing

Public cloud

(i) It stores & manages access to the data & application through internet

(ii) Easy to access.
- These resources are available through web. it unables enterprises to grow & expand with ease.

(iii) Payment on need to use basis

(iv) Dynamic, elastic & scalable

(v) Since no. of users are large, cost per user becomes very low (economical)

(vi) Strict Stranguent SLA.

(vii) Additionally the cloud service provider use rigorous security measures to prevent the data confidentiality.
Eg: Google.

Private cloud

(i) It is a cloud where usuage is exclusively reserved for a single corporate organisation.

(ii) Private cloud offers high level of security but is very expensive

(iii) Two types.
- In house pvt cloud/ On the premise
- Out sourced pvt. cloud

Eg: AWS (Amazon web service)

(viii) Resiliance
This means that public cloud users are not at all bothered about what is happening backend office of the public cloud.

Hybrid cloud

(i) It is the combination of public & pvt. cloud.

(ii) In this cloud the sensitive & confidential data are kept on premise pvt. cloud & the non-sensitive data on the public cloud.

(iii) It unables the enterpise to exploit the benefit of shared resources while leveraging there IT infra - for mission/ critical security need.

Eg: Govt. Org. uses this.

Community cloud.

Page No.

Saathi

Date __/__/____

Business Intelligence

BI tools includes DA, DM, BA, Data Visualisation, AI, BI, ML to assist business in making best-choices.

What are the different methods of BI.?

(1) Data Mining - Identification of unknown trends patterns between the variable for enhanced corporate decision making.

(2) Data Visualisation - Transforming the data into the visual format like pie, charts, table, Gantt charts, etc.

(3) Reports - It consists of Conclusion to be presented before Stakeholders & for decision making.

(4) Performance Matrix & Benchmarking -

(5) Descriptive Analysis

(6) Data Querring -

(7) Statistical Analysis

(8) Visual Analysis

(9) Data preparation

Page No.

Saathi

Date __/__/____

XBRL (Extensible Business Reporting Language)

(I) It's a data description language that is facilitates the interchange of standard, comprehensible corporate data

(II) It is based on XML (Extensible markup language).

(III) It unables automated interchange & trustworthy extraction of financial data across all software types & advanced technology, including internet.

(IV) XBRL allows organisations to arrange data using tags.

Eg:- When a piece of data is labelled as "revenue", for instance, XBRL enabled applic... know that it pertains to revenues

Benefits

(1) All reports are automatically created from a single source of information which reduces the chance of erroneous data entry and hence increases data reliability.

(2) Reduces expenses by automating the report production for various client.

(3) Accelerates the decision-making of financial entities such as bank & rating services

(4) Access, comparison, and analytic capabilities for information are unparalleled

Page No.

Saathi

Date __/__/____

to who are the participants?

① Companies
(a) auditor
(b) creditor partner
(c) MA

② Financial Publishers
(a) Regulars
(b) SEBI

③ Investors. Software vendors

XBRL process.

(1) Business operations
(2) Internal fin-reporting
(3) External fin-reporting.
(4) Investment & landing analysis.

Short note of XBRL (XML (Extensible markup language)

(1) It is a file format & markup language for storing, transferring & recreating arbitrary data.
(2) It specifies set of standards for encoding texts in a format that is understandable both by machines & human.
(3) It is defined by the 1998 XML 1.0 - specification of the world wide web Consortium & numerous other related specifications, which are all free open standards.
(4) Several Schema systems exists to help in the design of XML-based languages, and numerous APIs have been developed by programmers to facilitate the processing of XML data.
(5) Eg: IETF RFC 7303, RFC 3470, commonly known as IETF BCP 70.

Page No.

Saathi

Date __/__/____

Application of XML

(1) It is most widely used for the exchange of data via Internet.

(2) There have been 100s of document formats created using XML Syntax, including office Open XML, Open doc., XHTML, etc.

(3) XML is also the foundational language for communication protocols like SOAP & XMPP.

(4) It is also the message interchange format for the programming approach Asynchronous JavaScript and XML (AJAX)

Date 21/10/23

Lecture - 12.

AI
Artifical Intelligence

(1) John McCarthy defined AI ~ Science & Engineering of making intelligent machines / intelligent computer programs.

(2) Alan Turing - "Father of Computer Science".
Posed the question - Can machines think?

(3) Stuart Russell published "AI - A modern approach".

Saathi

Date __/__/____

AI has two approaches

Human approach	Ideal approach.
(i) Systems that thinks like human.	(i) Systems that think rationally
(ii) Systems that act like human.	(ii) Systems that act rationally.

(4) AI includes 2 sub fields.
(a) ML (machine learning)
(b) DL (Deep learning).

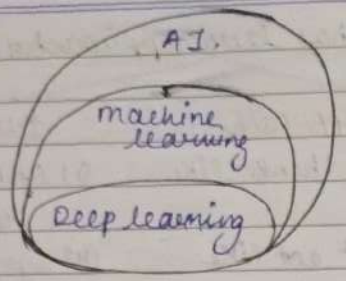
Application of AI in the area of Financial services

(A) Investment Services - (i) Algorithmic trading
(ii) Robo Advisory.
(iii) Insurance claim processing
(iv) Pricing of insurance products

(B) Lending - (i) Retail & Commercial lending op.
(ii) Retail & commercial lending scores
(iii) Detecting possibility of default

(C) Audit & compliance - (i) Fraud detection
(ii) Regulatory compliance
(iii) Travel & expense mgt

(D) Customer Service - (i) Detecting new sell opportunities
(ii) Know your customer (KYC)
(iii) Prediction of customer churning



Types of AI.

Weak AI.

- called as narrow AI/ ANI.
- It's a kind of AI which is trained to do a particular task.
- It is feeble & today weak AI mostly arounds us.
  eg :- Apple Siri, Amazon's Alexa, Tesla's autodriven car.

Strong AI.

- called as ~~or~~ ASI, AGI.
- when a machine possesses human level intellect, and self aware consciousness, has the ability to solve problems, learn, plan for future.
  eg :- Sofia Robot

Sub-field of AI.

Machine learning

① It is a subset of AI which involves manual human intervention to interpret the data.

② ML ~~doesn't~~ require more structure data to learn.

Deep learning

① It is a subset of MI which automates a significant portion of the features extraction step reducing the need of manual human intervention.

② They can work on bigger data sets which are unstructured.

③ They can also utilise labled data sets known as supervised learning.

Robotic Process Automation (RPA)

① RPA are capable of learning, simulating & executing rules based business process.

② RPA software bots can communicate with any application or system in the same manner humans can.

③ RPA bots can function continuously around the clock with 100% accuracy.

④ RPA bots possess a digital skill that exceeds that of humans.

⑤ RPA bots are simple to configure or utilise. (We) will be able to configure RPA bots, if you know how to record & do. It has pressing record, play, pause, drop, stop, drag & drop button.

MDP → Markov decision process in ML

Saathi

Date __/__/____

Benefits of RPA.

(i) Higher productivity.
(ii) Higher accuracy
(iii) Saving of cost.
(iv) Integration across platforms.
(v) Better customer experience
(vi) Harnessing AI.
(vii) Scalability.

Machine Learning

(1) ML is a branch of study devoted for understanding and developing systems that learns & ways that use data to improve performance on a set of tasks.

(2) Its application includes fields of medicine, email filtering, speech recognition, computer vision.

Approaches towards ML.

Supervised learning

(1) It constructs a mathematical model of data sets which includes both input & expected outcomes. This data acts as training data

Unsupervised

It utilises a data set comprising of just the inputs i.e the algorithms are taught using unlabelled, unclassified, uncategorised test data

Semi-supervised

It is the intermediate between supervised & unsupervised. This means when unlabeled data combined with little amt. of labeled data there is a significant gain in learning accuracy.

Reinforcement

It is a subfield of ML, which is determining how software agents should operate in a given environment Eg: game theory, control theory, info theory, etc.

Dimensionality reduction

It is a process of acquiring a set of major variables in order reduce the no. of random variables. It includes lowering the size of features or no. of features, etc.

Page No.

Saathi

Date __/__/____

Two Schools of thoughts for AI.

Data driven AI

It focuses on enhancing the data quality & data governance in order to enhance the performance of a particular problem Statement.

Model Driven AI.

It attempts to increase the performance by developing new models and algorithmic manipulation

** During Covid-19 AI's contribution was remarkable.
X-ray, CTScan, MRI
However it should be noted that AI largely depends on input data & poor input data will never lead to superior performance