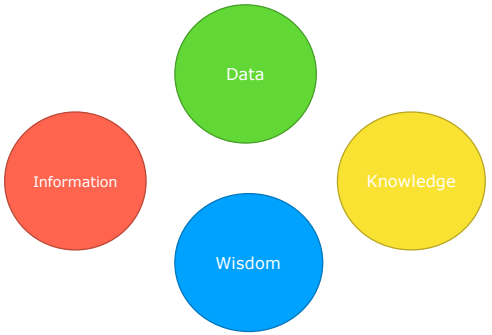
Data Science and Data Visualization

DR. SHAILESH SIVAN
Assistant Professor
Department of Computer Science
Cochin University of Science and Technology

DATA

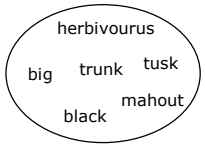
- Data is discrete, objective facts or observations, which are unorganized and unprocessed and therefore have no meaning or value because of lack of context and interpretation.
- Data is recorded (captured or stored) symbols, including , words (text and/or verbal), numbers, diagrams, and images (still &/or video), which are the building blocks of communication"



JUST START WITH THIS !

DATA, INFORMATION & KNOWLEDGE

Elephant

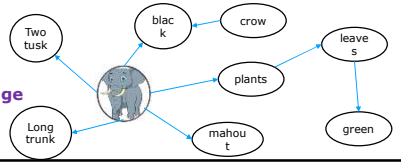


Data

- Biggest animal in the land
- Has two tusk
- Black in Colour
- Has a long trunk
- Eats plants
- Mahout is the keeper

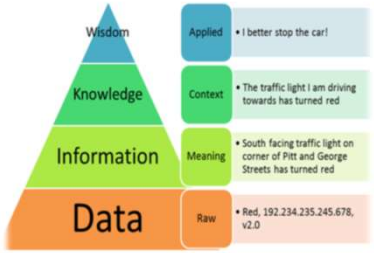
Information

Knowledge



AT GLANCE

- Data – facts and statistics collected for reference or analysis.
- Information – Facts provided or learned about somethingor someone.
- Knowledge – Facts, information, and skills acquired by a person through experience or education.
- Wisdom – the quality of having experience, knowledge, and good judgment – or my favorite…knowledge rightly applied.

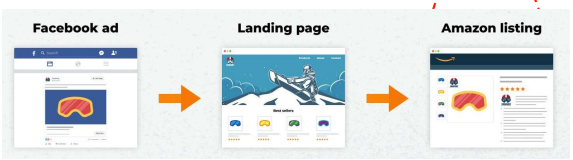


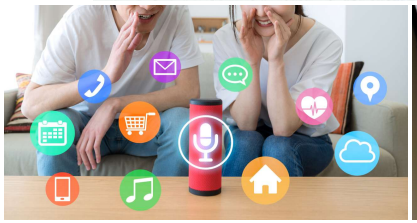
WHY DATA IS IMPORTANT ?

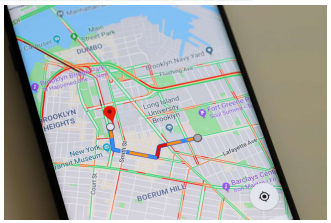
- Make Informed Decisions
- Get The Results You Want
- Find Solutions To Problems
- Back Up Your Arguments
- Be Strategic In Your Approaches
- Keep Track Of It All
- Know What You Are Doing Well

DATA IS NEW OIL

IS DATA IMPORTANT?





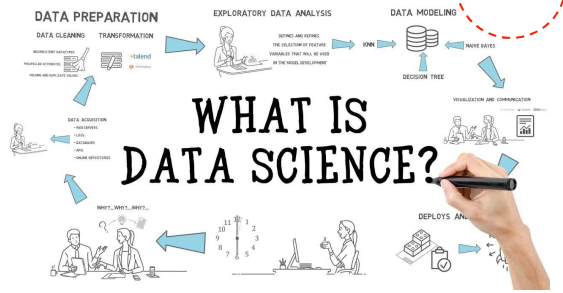


REALITY ABOUT DATA

We are now **drowing** in data ! but **starving** for insights ☹

SOLUTION ?

DATA SCIENCE



NEED AND TYPE OF DATA ANALYTICS

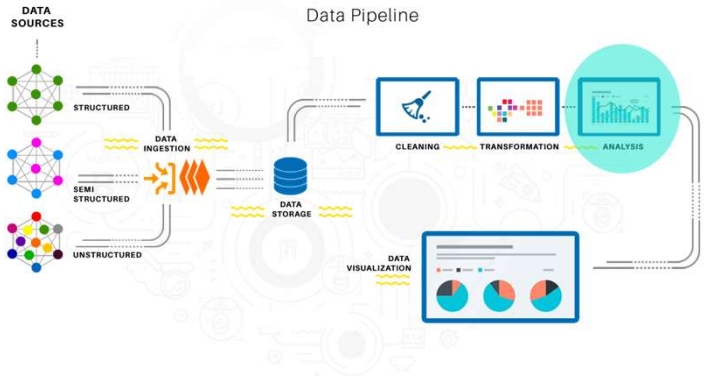
NEED

- Describe past results
- Diagnose why past results occurred
- Predict what might happen in the future
- Recommend actions for the organization's next steps

TYPES

- Descriptive Analytics.
- Diagnostic Analytics.
- Predictive Analytics.
- Prescriptive Analytics

DATA SCIENCE PIPELINE



DESCRIPTIVE ANALYTICS

***What* Happened ?**

- This type of analytics is rooted in facts
- Descriptive analytics can serve as that starting point or base for building your analytics strategy.
- With some professionals estimating that 80% of analytics involve descriptions based on aggregations of past performance.
- Techniques like data aggregation, data mining, clustering and/or summary statistics all serve to provide analytics that describe a past state—descriptive analytics

- A report provided by your sales team may include
  - Information on new clients, win rates and goals.
  - To customize the data view to look at only specific data to answer a specific question, such as
    - "How many customers do we have in Kochi?"
    - "How much volume did we sell last month in the Kochi City?"

DIAGNOSTIC ANALYTICS

***Why* Did it Happen ?**

- Diagnostic analytics also focus on the past.
- These types of analyses look for cause and effect to illustrate why something occurred.
- The objective is to compare past occurrences to determine causes.
- This is ofcourse a black-and-white answer, but instead provided in the context of probability, likelihood or a distributed outcome.

Diagnostic analytics can provide guidance by helping to:

- **Identify outliers**. For example, a sudden drop in sales or an explosion in website traffic that can't be explained may indicate a need for additional examination.
- **Isolate patterns.** Analysts may need to look outside the existing dataset to identify the source of the pattern. For example, a sudden drop in sales may have stemmed from the launch of a disruptor.
- **Uncover relationships.** Using more complex analytics, analysts may employ probability theory, regression analysis, or time series to isolate cause and effect relationships.

PRESPECTIVE ANALYTICS

How to Make it Happen ?

- **Prescriptive analytics** builds on predictive analytics by helping determine recommended (prescribed) actions based on desired potential (predicted) outcomes, helping organizations achieve their business objectives.
- Prescriptive analytics models are constantly "learning" through feedback mechanisms to continuously analyze action and event relationships and recommend the optimal solution.
- By simulating the solution, prescriptive analytics can examine all the key performance criteria to ensure the outcome would achieve the correct metric goals before anything is implemented.

- **Artificial intelligence**, **machine learning** and neural network algorithms are often employed to support prescriptive analytics
- To make specific suggestions based on nuanced patterns and perceptions of organizational goals, limitations and influencing factors.

PREDICTIVE ANALYTICS

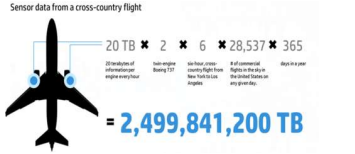
What Happens *If* ?

- Predictive analytics is the branch of the advanced analytics which is used to make predictions about unknown future events.
- Predictive analytics uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future.

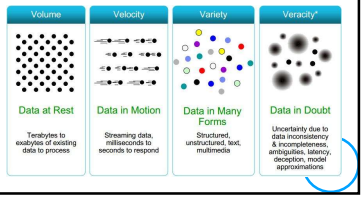
- Organizations are turning to predictive analytics to help solve difficult problems and uncover new opportunities. Common uses include:
- **Detecting fraud.** Combining multiple analytics methods can improve pattern detection and prevent criminal behavior.
- **Optimizing marketing campaigns.** Predictive analytics are used to determine customer responses or purchases, as well as promote cross-sell opportunities.

BIG DATA AND ANALYTICS

Is Big Data = Large Data ?



Properties of Bigdata



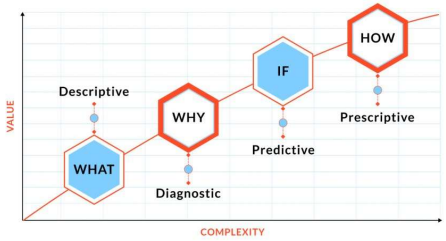
TOOLS FOR DATA ANALYTICS

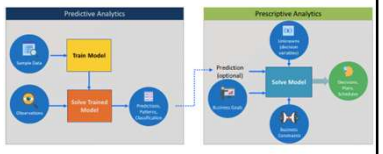
INDUCING INTELLIGENCE IN ANALYTICS

- Big data and artificial intelligence have a synergistic relationship
- AI requires a massive scale of data to learn
- Big data analytics leverages AI for better data analysis.
- With this convergence, one can more easily leverage advanced analytic.

Take Advantage of **AI/ML for Big Data Analytics**

INDUCING INTELLIGENCE IN ANALYTICS

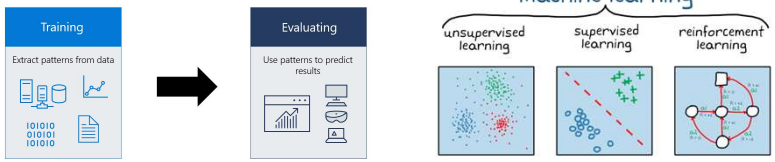




LESS HIGH

Need of AI

MACHINE LEARNING



Problems with Uncertainity

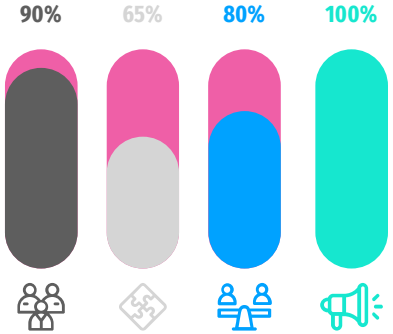
Approximating the patterns to Generalization

Data with hidden patterns

Minimum error to predict unseen data

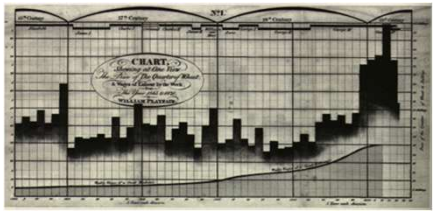
Data Visualization

- Graphical representation of information and data.
- Uses visual elements like charts, graphs, and maps
- Data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
- This makes the data more natural for the human mind to comprehend



History of Data Visualization

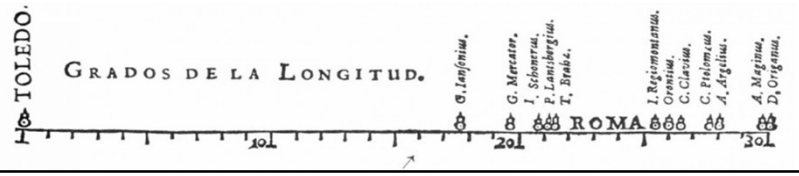
- 18th century saw the beginning of thematic mapping.
  - Attempts at the thematic mapping of geologic, economic, and medical data
- William Playfair
  - widely considered to be the inventor of many of the most popular graphs we use today
  - line, bar, circle, and pie charts



A graph by Playfair (1821), shown below, shows the price of wheat, weekly wages, and reigning monarch over a two hundred fifty year span from 1565 to 1820.

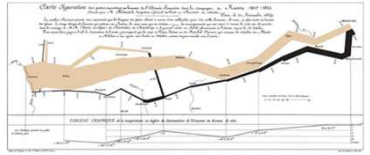
History of Data Visualization

- Prior to the 17th century, data visualization existed mainly in the realm of maps, displaying land markers, cities, roads
- In 1644, Michael Florent Van Langren, a Flemish astronomer, is believed to have provided the first visual representation of statistical data
- Van Langren could have provided this information in a table, it is the use of the graph that really visually displays the wide variations in estimates



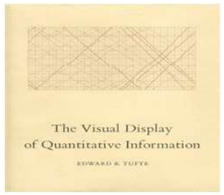
History of Data Visualization

- The latter half of the 19th century is what Friendly calls the Golden Age of statistical graphics.
- Two famous examples of data visualization
  - map of cholera outbreaks in the London epidemic of 1854
  - number of men in Napoleon's 1812 infamous Russian campaign army, with army location indicated by the X-axis, and extreme cold temperatures indicated at points when frostbite took a fatal toll.



History of Data Visualization

- The latter half of the 20th century is what Friendly calls the 'rebirth of data visualization', brought on by the emergence of computer processing
- The early 80s saw the emergence of Edward Tufte, whose seminal work, *The Visual Display of Quantitative Information* is still used today in university courses for data visualization and statistical analysis



Why Data Visualization

A picture is worth a thousand words

Humans just understand data better through pictures rather than by reading numbers in rows and columns.

If the data is presented in a graphical format, people are more able to effectively find correlations and raise important questions.

Data Visualization - A Formal Definition

Data visualization is the process of acquiring, interpreting and comparing data in order to clearly communicate complex ideas, thereby facilitating the identification and analysis of meaningful patterns.

Data visualization helps people understand the significance of data by summarizing and presenting a huge amount of data in a simple and easy-to-understand format in order to communicate the information clearly and effectively.

How visualization help businesses?

01 Converting the business data into interactive graphs for dynamic interpretation to serve the business goals

02 Transforming data into visually appealing, interactive dashboards of various data sources to serve the business with the insights

03 Creating more attractive and informative dashboards of various graphical data representations

04 Making appropriate decisions by drilling into the data and finding the insights

05 Figuring out the patterns, trends, and correlations in the data being analyzed to determine where they must improve their operational processes and thereby grow their business

06 Organizing and presenting massive data intuitively to present important findings from the data and making better, quick, and informed decisions with data visualization

Data Visualization Process

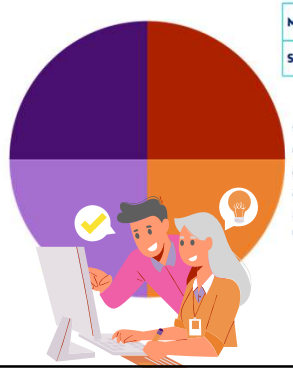
01	Filtering & Processing	Refining and cleaning data to convert it into information through analysis, interpretation, contextualization, comparison and research
02	Translation & Visual Representation	Shaping the visual representation by defining graphic resources, language, context, and the tone of the representation, all of which are adapted for the recipient.
03	Perception & Interpretation	Finally, the visualization becomes effective when it has a perceptive impact on the construction of knowledge

Importance of Visualization

We typically process images 60,000 times faster than a table or a text

13 milliseconds and store the information, provided that it is associated with a concept.

A visual representation that uses images is a much more effective way to communicate information than text or a table; it also takes up much less space.



Month	Jan	Feb	Mar	Apr	May	Jun
Sales	45	56	36	58	75	62