



IIT ROORKEE

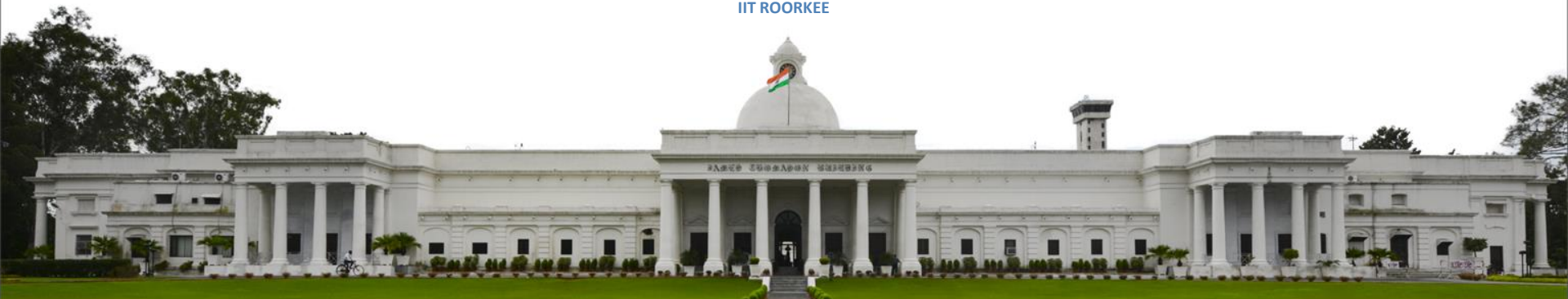


NPTEL ONLINE
CERTIFICATION COURSE

Data Analytics with Python

Lecture 1: Introduction to data analytics

Dr. A. Ramesh
DEPARTMENT OF MANAGEMENT
IIT ROORKEE

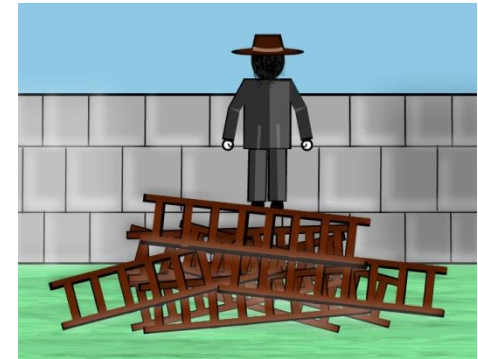


Objective of the course

- The principle focus of this course is to introduce conceptual understanding using simple and practical examples rather than repetitive and point click mentality
- This course should make you comfortable using analytics in your career and your life
- You will know how to work with real data, and might have learned many different methodologies but choosing the right methodology is important

Objective of the course Contd...

- The danger in using quantitative method does not generally lie in the inability to perform the calculation
- The real threat is lack of fundamental understanding of:
 - Why to use a particular technique or procedure
 - How to use it correctly and,
 - How to correctly interpret the result



Learning objectives

1. Define data and its importance
2. Define data analytics and its types
3. Explain why analytics is important in today's business environment
4. Explain how statistics, analytics and data science are interrelated
5. Why python?
6. Explain the four different levels of Data:
 - Nominal
 - Ordinal
 - Interval and
 - Ratio

1. Define Data and its importance

- Variable, Measurement and Data
- What is generating so much data?
- How data add value to the business?
- Why data is important?

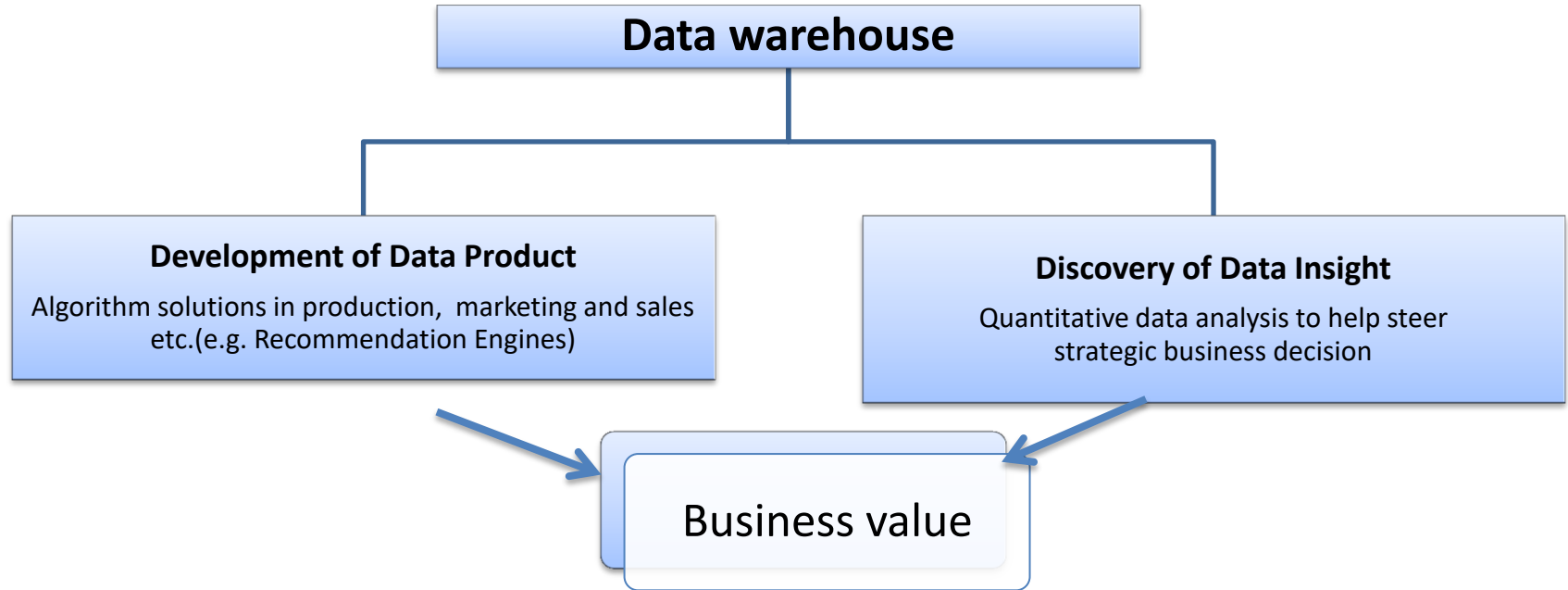
1.1 Variable, Measurement and Data

- Variables – is a characteristic of any entity being studied that is capable of taking on different values
- Measurements – is when a standard process is used to assign numbers to particular attributes or characteristic of a variable
- Data – data are recorded measurements

1.2 What is generating so much data?

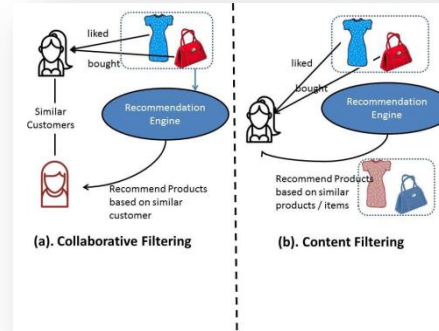
- Data can be generated by
 - Humans,
 - Machines or
 - Humans-machines combines
- It can be generated anywhere where any information is generated and stored in structured or unstructured formats

1.3 How data add value to business?



Source: <https://datajobs.com/>

Data Products



1.4 Why Data is important?

- Data helps in make better decisions
- Data helps in solve problems by finding the reason for underperformance
- Data helps one to evaluate the performance.
- Data helps one improve processes
- Data helps one understand consumers and the market

2. Define data analytic and its types

- Define data analytics
- Why analytics is important?
- Data analysis
- Data analytics vs. Data analysis
- Types of Data analytics

2.1. Define data analytics

- Analytics is defined as “the scientific process of transforming data into insights for making better decisions”
- Analytics, is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions – James Evans
- Analysis = Analytics ?

2.2 Why analytics is important?

- Opportunity abounds for the use of analytics and big data such as:



1. Determining credit risk
2. Developing new medicines
3. Finding more efficient ways to deliver products and services
4. Preventing fraud
5. Uncovering cyber threats
6. Retaining the most valuable customers



2.3 Data analysis

- Data analysis is the process of examining, transforming, and arranging raw data in a specific way to generate useful information from it
- Data analysis allows for the evaluation of data through analytical and logical reasoning to lead to some sort of outcome or conclusion in some context
- Data analysis is a multi-faceted process that involves a number of steps, approaches, and diverse techniques

Analysis

2.4 Data analytics vs. Data analysis

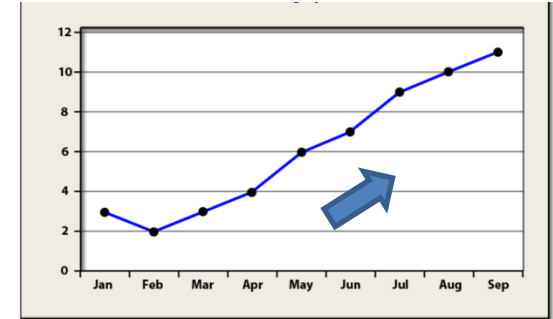


Past

Explain

How?

Why?



2.4 Data analytics vs. Data analysis

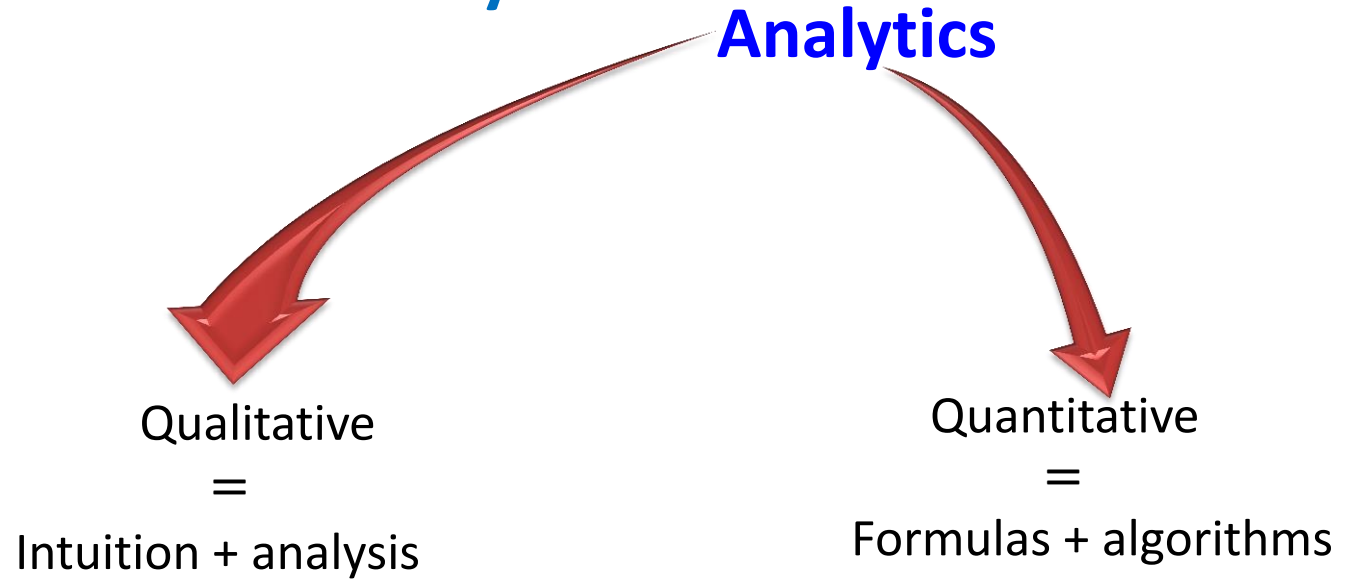
Analytics



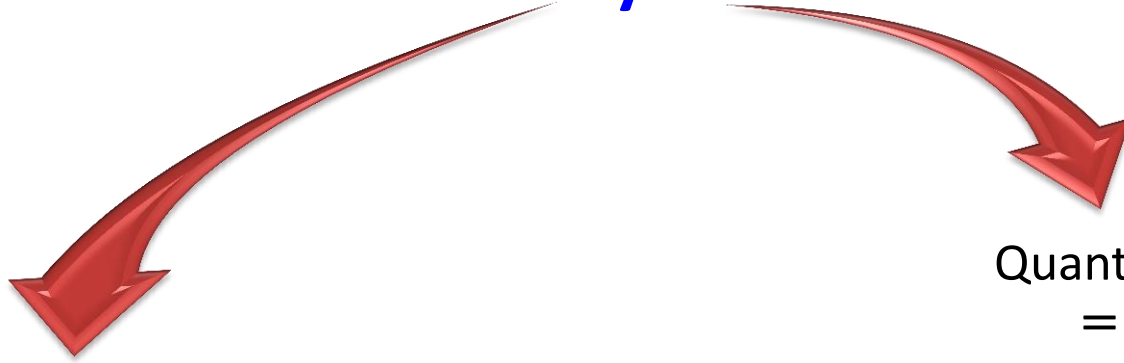
Future

Explore potential future events

2.4 Data analytics vs. Data analysis



Analysis



Qualitative

=

Explains **How** And **Why** Story ends the way it did ?

Quantitative

=

Data + **how** the sale **decreased** last summer

Analysis \neq Analytics

Data Analysis \neq Data analytics

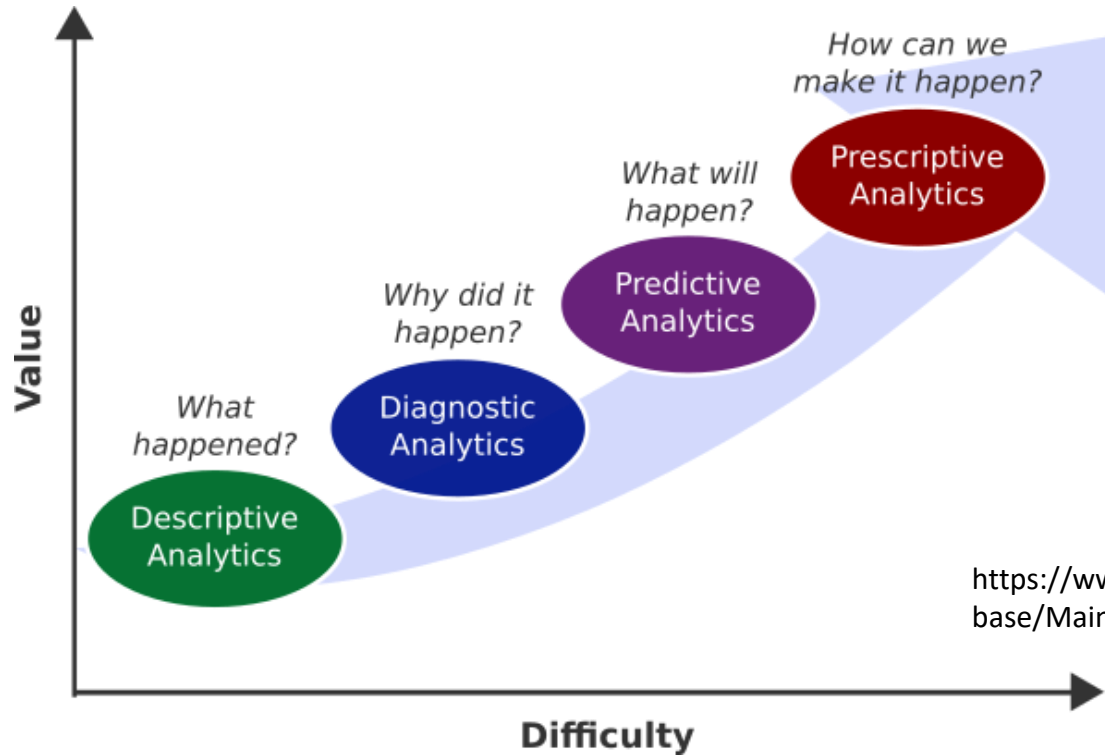
Business Analysis \neq Business analytics

2.5 Classification of Data analytics

Based on the phase of workflow and the kind of analysis required, there are four major types of data analytics.

- Descriptive analytics
- Diagnostic analytics
- Predictive analytics
- Prescriptive analytics

Classification of Data analytics



https://www.governanceanalytics.org/knowledge-base/Main_Tools/Data_classification_and_analysis

Descriptive Analytics

- Descriptive Analytics, is the conventional form of Business Intelligence and data analysis
- It seeks to provide a depiction or “summary view” of facts and figures in an understandable format
- This either inform or prepare data for further analysis
- Descriptive analysis or statistics can summarize raw data and convert it into a form that can be easily understood by humans
- They can describe in detail about an event that has occurred in the past

Example

A common example of Descriptive Analytics are company reports that simply provide a historic review like:

- Data Queries
- Reports
- Descriptive Statistics
- Data Visualization
- Data dashboard



Source: <https://www.linkedin.com/learning/478e9692-d13d-338f-907e-d76f0724d773>

Diagnostic analytics

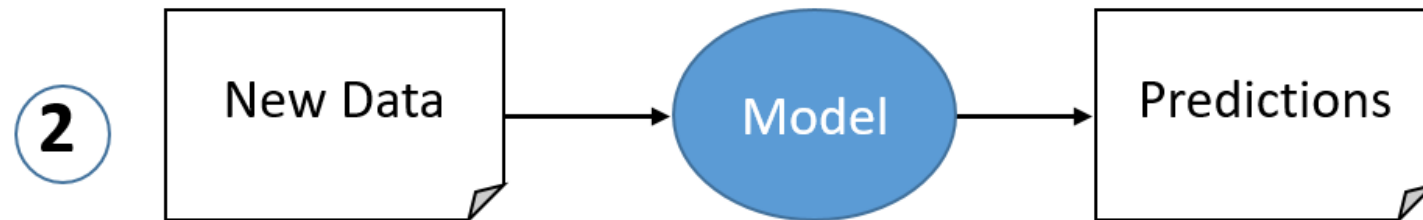
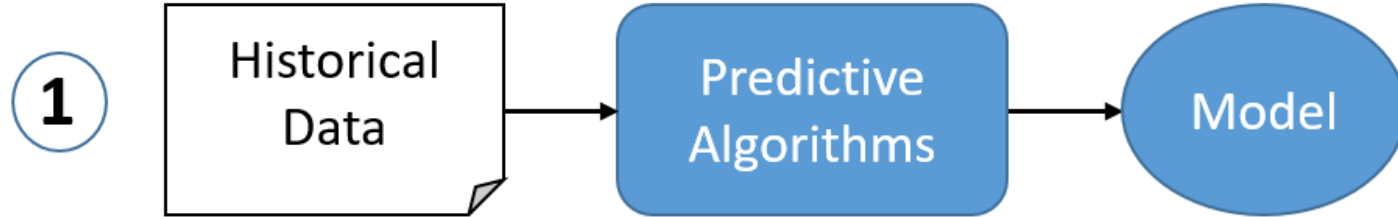
- Diagnostic Analytics is a form of advanced analytics which examines data or content to answer the question “Why did it happen?”
- Diagnostic analytical tools aid an analyst to dig deeper into an issue so that they can arrive at the source of a problem
- In a structured business environment, tools for both descriptive and diagnostic analytics go parallel

Example

- It uses techniques such as:
 1. Data Discovery
 2. Data Mining
 3. Correlations

Predictive analytics

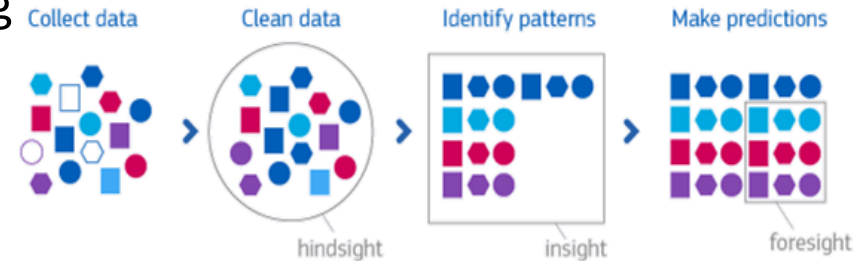
- Predictive analytics helps to forecast trends based on the current events
- Predicting the probability of an event happening in future or estimating the accurate time it will happen can all be determined with the help of predictive analytical models
- Many different but co-dependent variables are analysed to predict a trend in this type of analysis



Source: <https://www.logianalytics.com/wp-content/uploads/2017/11/predictive-1.png>

Example

- Set of techniques that use model constructed from past data to predict the future or ascertain impact of one variable on another:
 - Linear regression
 - Time series analysis and forecasting
 - Data mining



Source: <https://bigdata-madesimple.com/5-examples-predictive-analytics-travel-industry/>

Prescriptive analytics

- Set of techniques to indicate the best course of action
- It tells what decision to make to optimize the outcome
- The goal of prescriptive analytics is to enable:
 1. Quality improvements
 2. Service enhancements
 3. Cost reductions and
 4. Increasing productivity

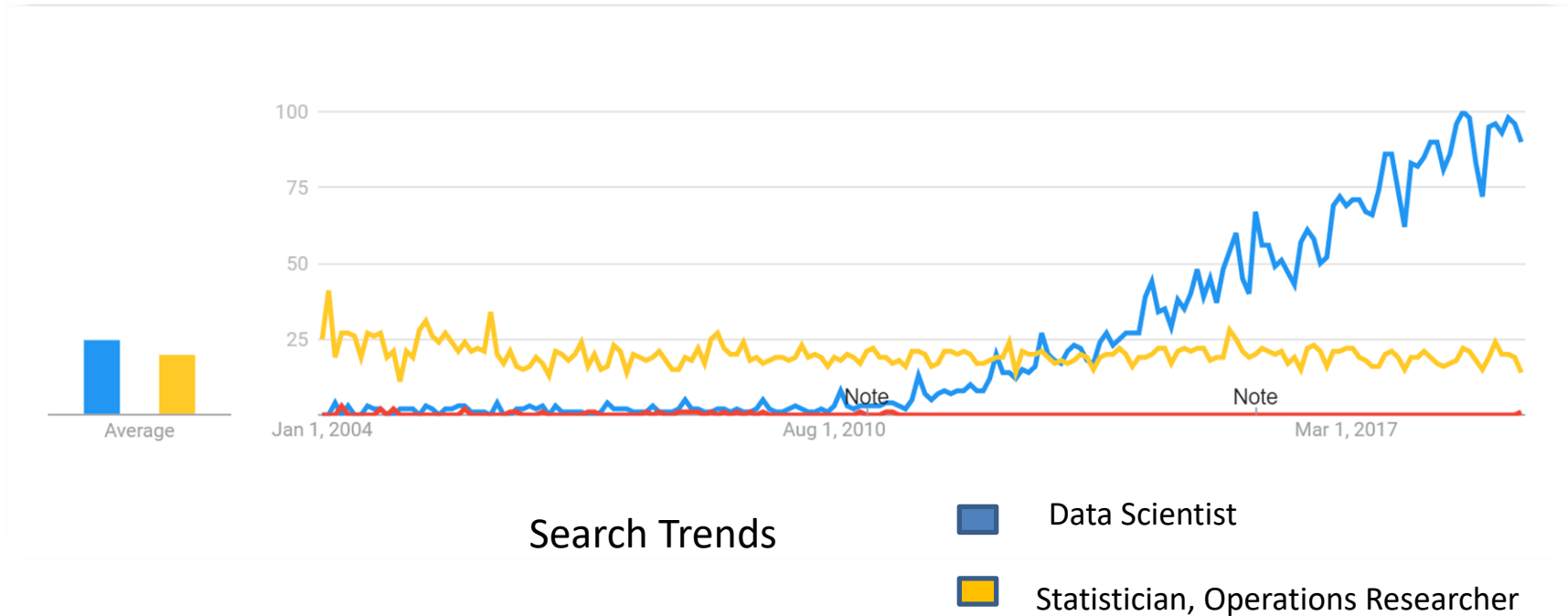
Prescriptive analytics: Example

- Optimization Model
- Simulation
- Decision Analysis

3. Explain why analytics is important

- Demand for Data Analytics
- Element of data Analytics

3. Explain why analytics is important





THIS STORY IS FROM MAY 8, 2016

Data scientists earning more than CAs, engineers

Namrata Singh | TNN | Updated: May 8, 2016, 14:50 IST



A-

A+

<https://timesofindia.indiatimes.com/india/Data-scientists-earning-more-than-CAs-engineers/articleshow/52171064.cms>

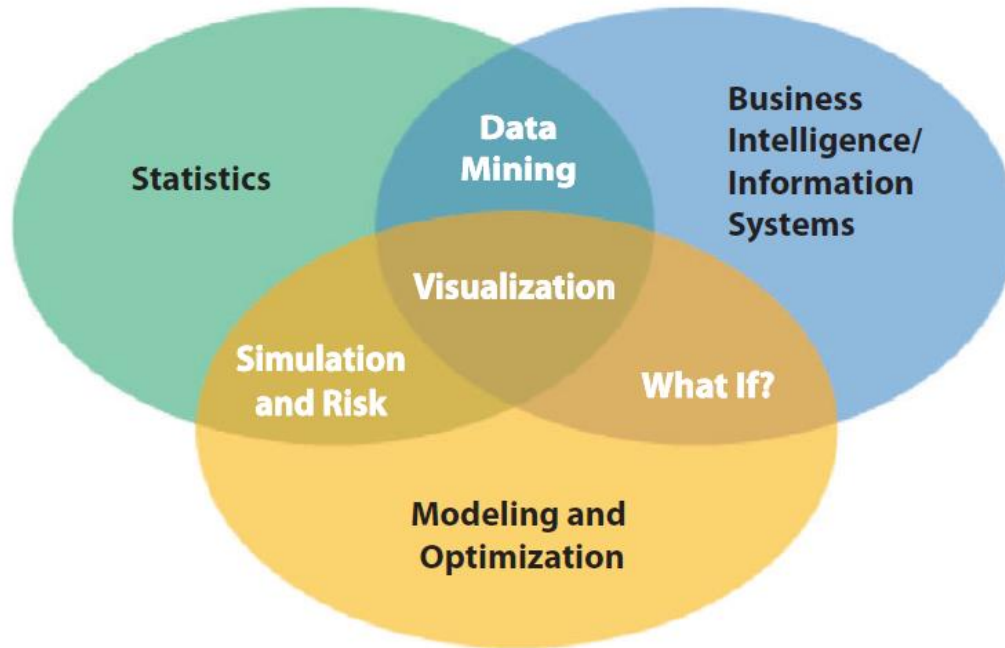
3.1 Demand for Data Analytics

With companies across industries striving to bring their research and analysis (R&A) departments up to speed, the demand for qualified data scientists is rising.

"India will face a demand-supply gap of 2,00,000 analytics professionals over the next three years. Even in the US, only 40 out of 100 positions for analytics professionals can be filled," said Rituparna Chakraborty , co-founder & senior VP of TeamLease Services.

http://timesofindia.indiatimes.com/articleshow/52171064.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

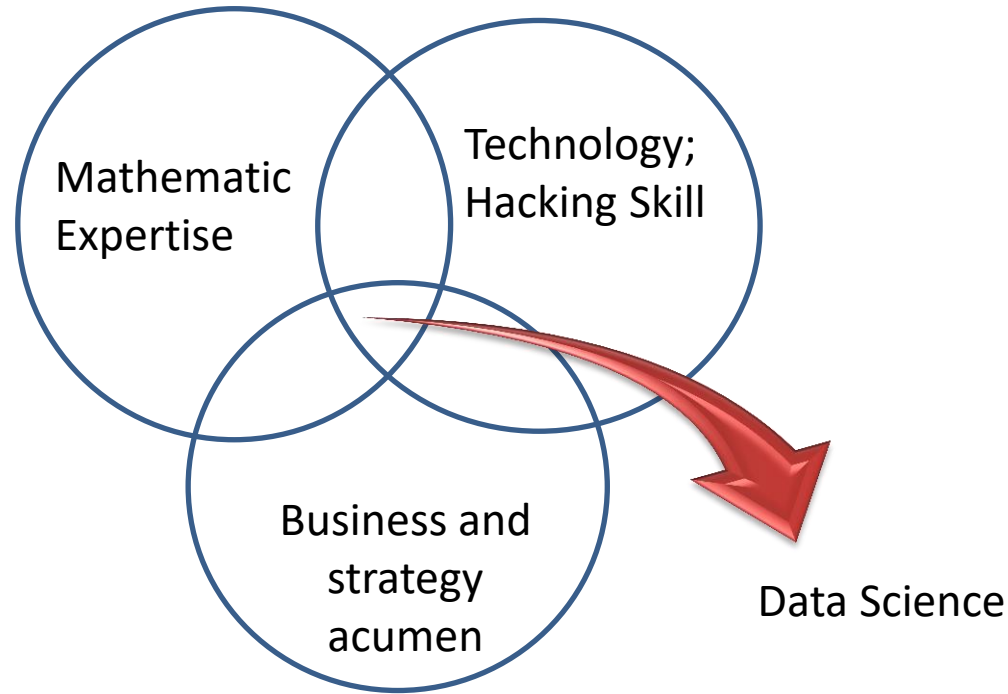
3.2 Element of data Analytics



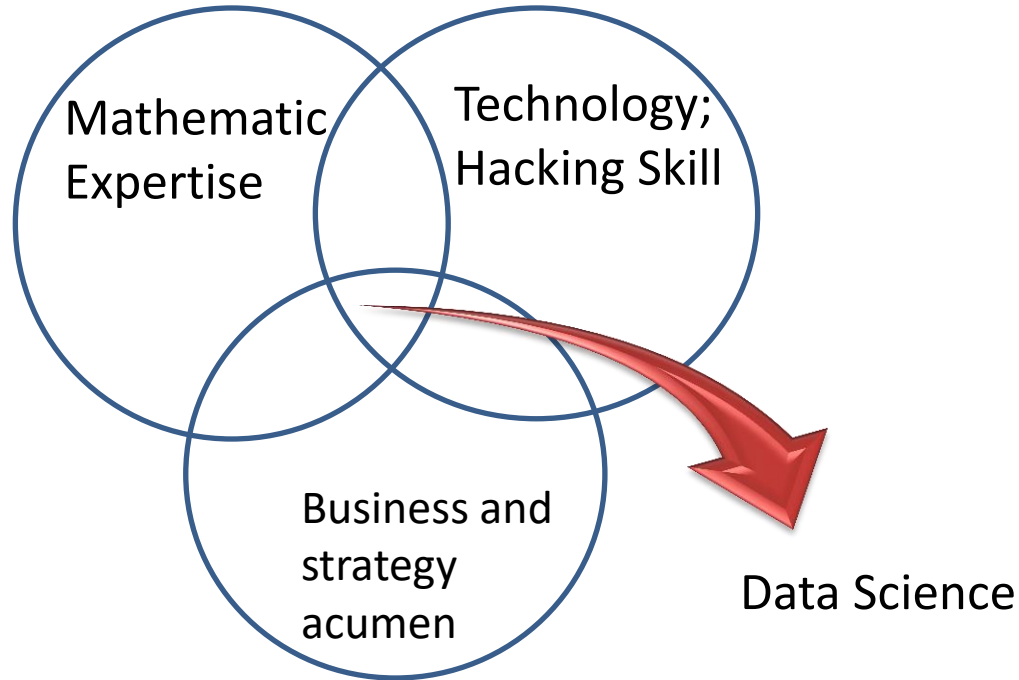
4. Data analyst and Data scientist

- The requisite skill set
- Difference between Data analyst and Data Scientist

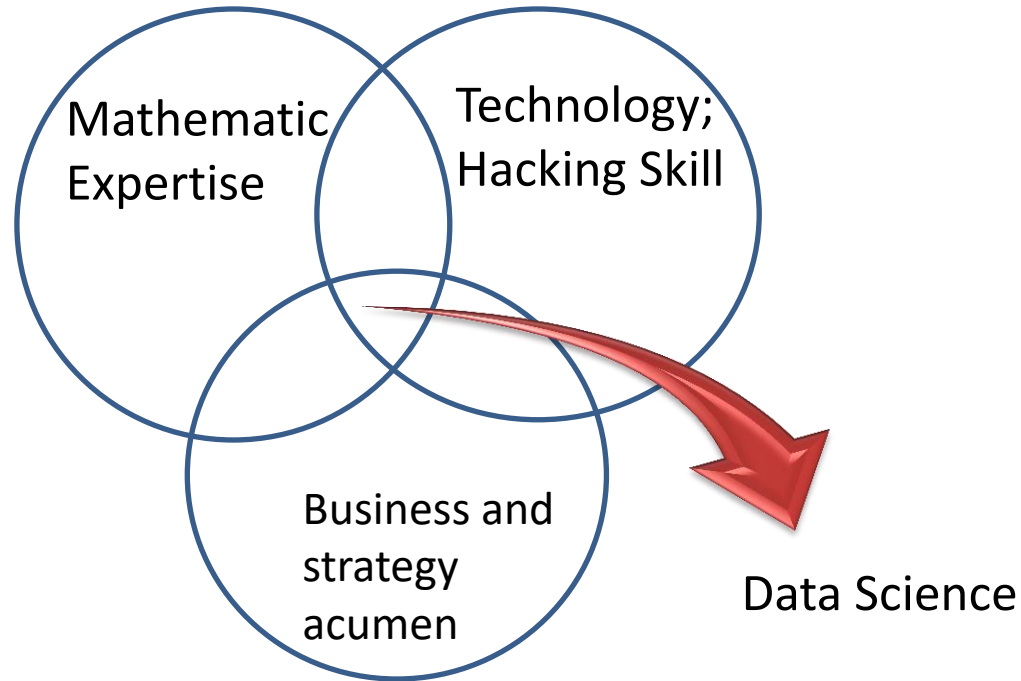
4.1 The requisite skill set



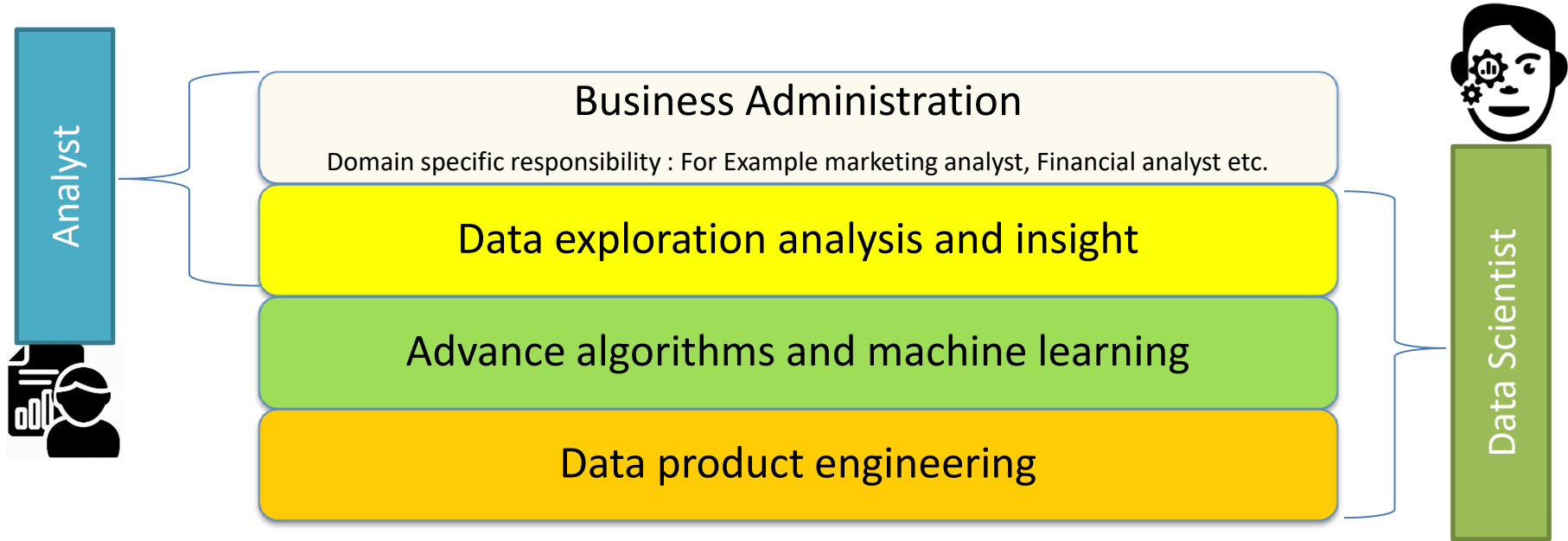
4.1 The requisite skill set



4.1 The requisite skill set



4.2 Difference between Data analyst and Data Scientist



Source: <https://datajobs.com/>

5. Why python?



Features

- Simple and easy to learn
- Freeware and Open source
- Interpreted
- Dynamically Typed
- Extensible
- Embedded
- Extensive library

5. Why python?



Usability

- Desktop and web applications
- Database applications
- Networking applications
- Data analysis (Data Science)
- Machine learning
- IoT and AI applications
- Games

Companies using Python



Why Jupyter Notebook?



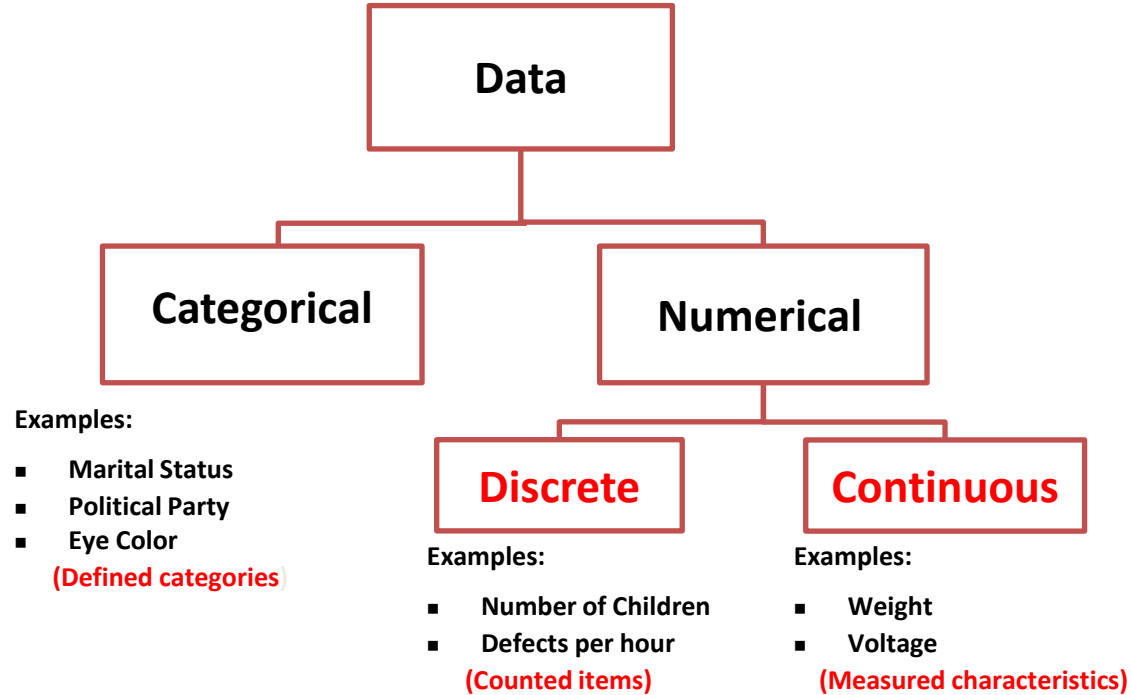
Why?

- Client – Server Application
- Edit code on web browser
- Easy in documentation
- Easy in demonstration
- User- friendly Interface

6. Explain the four different levels of Data

- Types of Variables
- Levels of Data Measurement
- Compare the four different levels of Data:
Nominal
Ordinal
Interval and
Ratio
- Usage Potential of Various Levels of Data
- Data Level, Operations, and Statistical Methods

6.1 Types of Variables



6.2 Levels of Data Measurement

- Nominal — Lowest level of measurement
- Ordinal
- Interval
- Ratio — Highest level of measurement

6.3.1 Nominal

- A **nominal scale** classifies data into distinct categories in which no ranking is implied
- Example : Gender, Marital Status

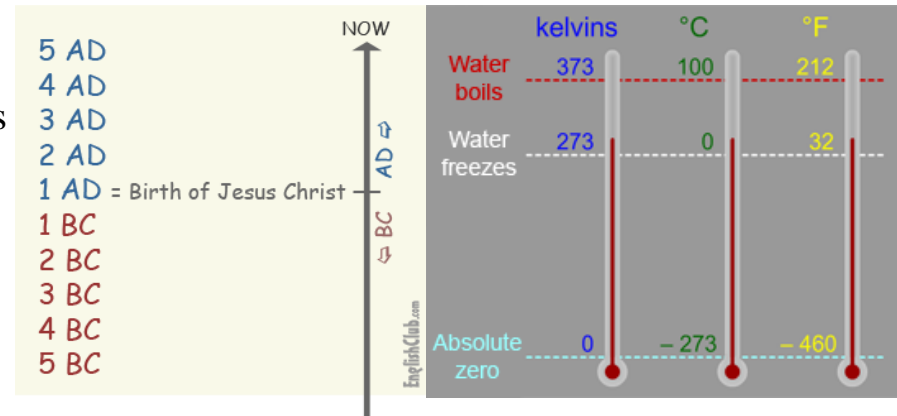


6.3.2 Ordinal scale

- An **ordinal scale** classifies data into distinct categories in which ranking is implied
- Example:
 - Product satisfaction → Satisfied, Neutral, Unsatisfied
 - Faculty rank → Professor, Associate Professor, Assistant Professor
 - Student Grades → A, B, C, D, F

6.3.3. Interval scale

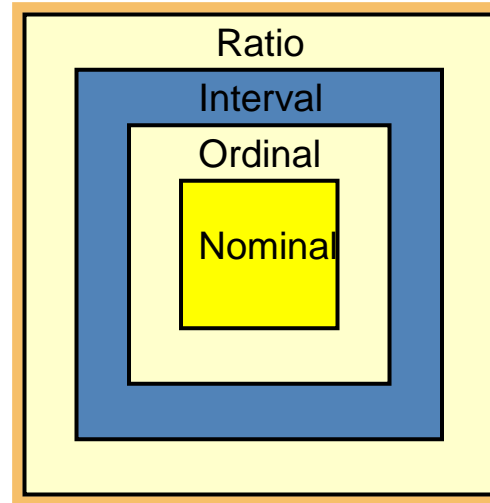
- An **interval scale** is an ordered scale in which the difference between measurements is a meaningful quantity but the measurements do not have a true zero point.
- Example
 - Temperature in Fahrenheit and Celsius
 - Year



6.3.4 Ratio scale

- A **ratio scale** is an ordered scale in which the difference between the measurements is a meaningful quantity and the measurements have a true zero point.
- Example
 - Weight
 - Age
 - Salary

6.4 Usage Potential of Various Levels of Data



6.5 Impact of choice of measurement scale

Data Level	Meaningful Operations	Statistical Methods
Nominal	Classifying and Counting	Nonparametric
Ordinal	All of the above plus Ranking	Nonparametric
Interval	All of the above plus Addition, Subtraction	Parametric
Ratio	All of the above plus multiplication and division	Parametric

Thank You

