

Session 1: Introduction to Business Analytics

Agenda

- **The Four V's Of Big Data**
- **What Is Business Analytics?**
- **Business Analytics Applications**
- **Importance Of Business Analytics**
- **Types Of Analytics**
- **Example**
- **Data Analytics Vs Statistical Analysis**
- **Data For Business Analytics**
- **Model**
- **Decision Model**
- **Steps In Problem Solving Process**
- **Companies Using R For Analytics**
- **Role Of A Data Scientist**
- **Roadmap To Become A Data Scientist**
- **Quote**

The Four V's Of Big Data

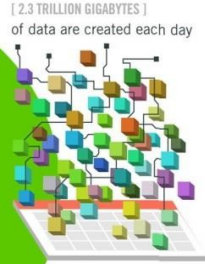
40 ZETTABYTES

[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005



Volume SCALE OF DATA

It's estimated that **2.5 QUINTILLION BYTES** [2.3 TRILLION GIGABYTES] of data are created each day



Most companies in the U.S. have at least **100 TERABYTES** [100,000 GIGABYTES] of data stored

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]



**30 BILLION
PIECES OF CONTENT**
are shared on Facebook every month

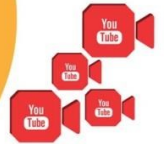


Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be

**420 MILLION
WEARABLE, WIRELESS
HEALTH MONITORS**

**4 BILLION+
HOURS OF VIDEO**
are watched on
YouTube each month



400 MILLION TWEETS
are sent per day by about 200 million monthly active users



The New York Stock Exchange captures

**1 TB OF TRADE
INFORMATION**
during each trading session



Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

**18.9 BILLION
NETWORK
CONNECTIONS**

— almost 2.5 connections per person on earth



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure



Poor data quality costs the US economy around

\$3.1 TRILLION A YEAR



**1 IN 3 BUSINESS
LEADERS**

don't trust the information they use to make decisions

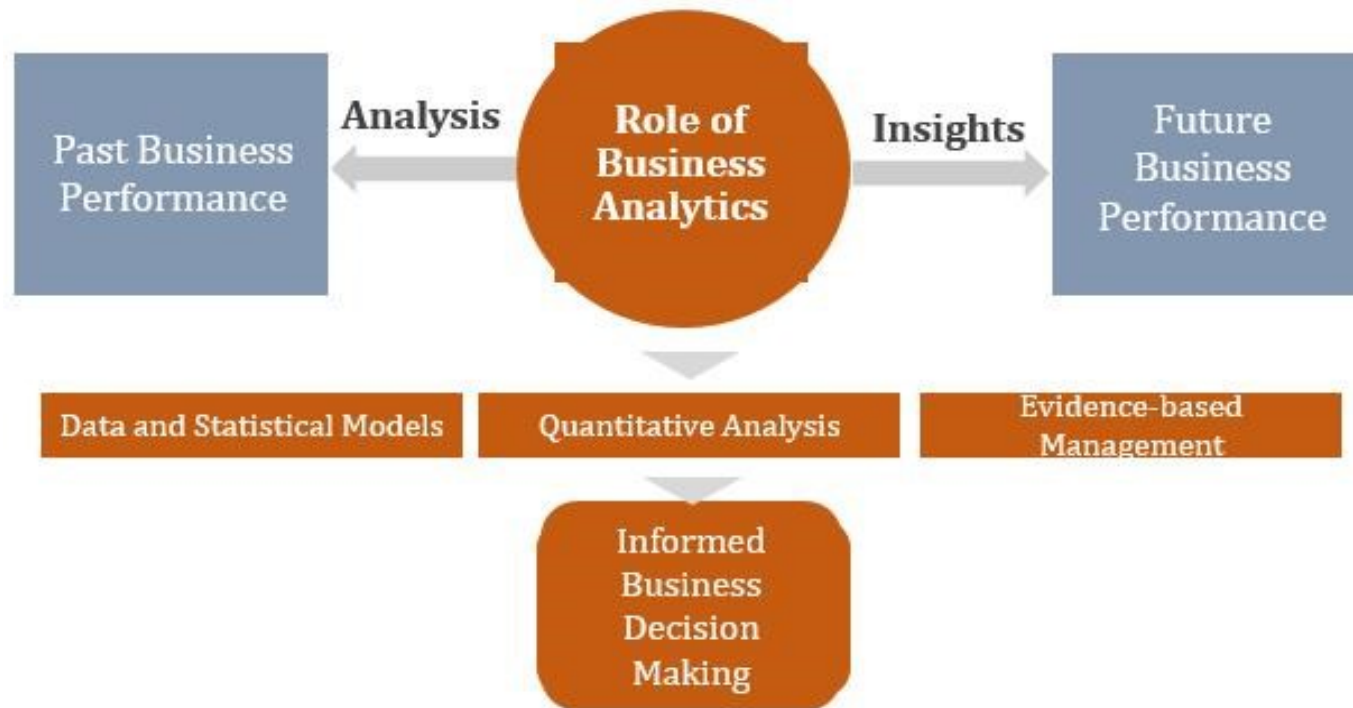


in one survey were unsure of how much of their data was inaccurate

Veracity UNCERTAINTY OF DATA

What Is Business Analytics?

- 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.



Business Analytics Applications

- Email marketing
- Promotions
- Health care
- Insurance
- Management of customer relationships
- Financial and marketing activities
- Supply chain management
- Human resource planning
- Pricing decisions
- Sport team game strategies

Importance Of Business Analytics

- Business Analytics helps in:
 - profitability of businesses
 - revenue of businesses
 - shareholder return
 - enhances understanding of data
 - helps businesses to remain competitive
 - enables creation of informative reports

Types Of Analytics

- **Descriptive analytics** - uses data to understand past and present i.e. describes a set of data.
- **Predictive analytics** - analyze current and historical facts to make predictions about future events. In essence, to use the data on some objects to predict values for another object.
- **Prescriptive analytics** - uses optimization techniques and suggests what should be done to optimize results

Example

- Retail stores usually run clearance sale by reducing the prices to finish their inventory
- **The question is: When to reduce the price and by how much?**
- **Descriptive analytics:** examine historical data for similar products (prices, units sold, advertising etc.)
- **Predictive analytics:** predict sales based on price by analysing historical sales data
- **Prescriptive analytics:** find the best sets of pricing and advertising to maximize sales revenue

Data Analytics Vs Statistical Analysis

Data Analytics

- Utilizes data mining techniques
- Identifies inexplicable or novel relationships/trends
- Seeks to visualize the data to allow the observation of relationships/trends

Statistical Analysis

- Utilizes statistical and/or mathematical techniques
- Used based on theoretical foundation
- Seeks to identify a significant level to address hypotheses or RQs

Data For Business Analytics

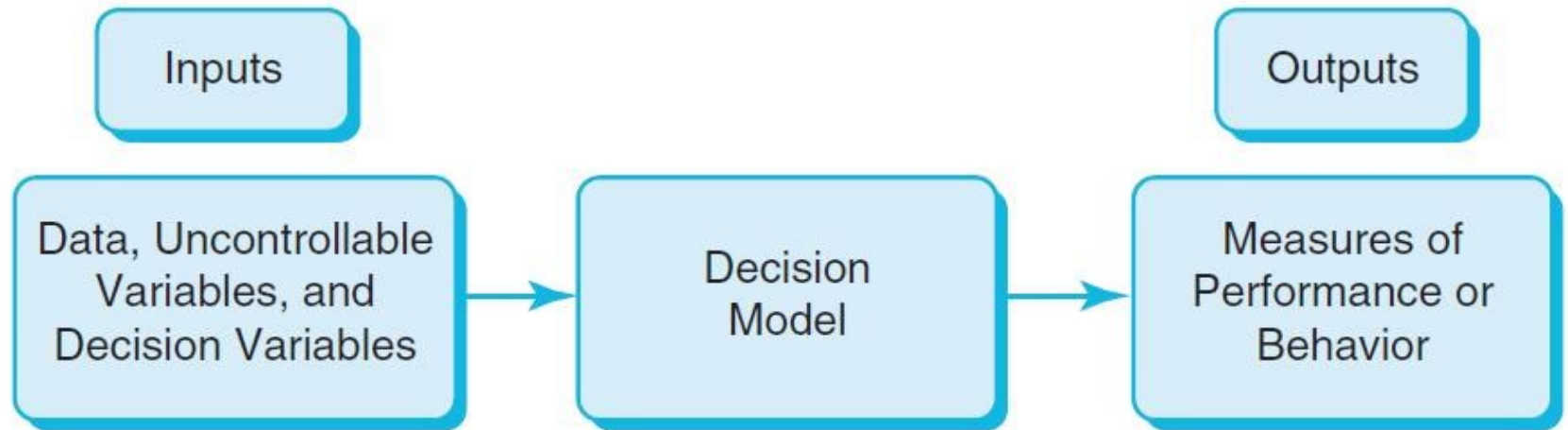
- DATA - collected facts and figures
- DATABASE - collection of computer files containing data
- INFORMATION - comes from analyzing data

Model

Model:

- An abstraction or representation of a real system, idea, or object
- Captures the most important features
- Can be a written or verbal description, a visual display, a mathematical formula, or a spreadsheet representation

Model (Contd.)



Decision Model

- A decision model is a model used to understand, analyze or facilitate decision making.

Types of model input:

- data
- uncontrollable variables
- decision variables (controllable)

Types of model output:

- performance measures
- behavioral measures

Steps In Problem Solving Process

Steps In Problem Solving Process

1. Recognizing the problem
2. Defining the problem
3. Structuring the problem
4. Analyzing the problem
5. Interpreting results and making a decision
6. Implementing the solution

Recognizing The Problem

- Problems exist whenever there is a gap between what is happening and what we think should be happening.
- For example, costs are too high compared with competitors and customers are opting for competitors because of that

Defining The Problem

- Clearly defining the problem is not an easy task.
- Complexity increases when the following occur:
 - several competing objectives
 - external groups are impacted
 - problem owner and problem solver are not the same person
 - time constraints exist

Structuring The Problem

- State goals and objectives
- Characterize the possible decisions
- Identify any constraints or restrictions
- Clearly break down the goals into further stories to be dealt with

Analyzing The Problem

- Identifying and applying appropriate Business Analytics techniques such as experimentation, statistical analysis, or a solution process

Interpreting Results & Making Decision

- Managers interpret the results from the analysis phase.
- Incorporate subjective judgment as needed.
- Understand limitations and model assumptions.
- Make a decision utilizing the above information.

Implementing The Solution

- Translate the results of the model back to the real world.
- Make the solution work in the organization by providing adequate training and resources.

Companies Using R For Analytics

Company Name	Uses
New York Times	Used for interactive and print visualization
Google	Has more than 500 R users
FDA	Uses R for clinical trials of new drugs
John Deere	Uses R to forecast crop yield and optimize tractor manufacturing
National Weather Service	Uses R to predict the extent of flooding events
The Consumer Financial Protection Bureau	Uses R and other open source tools
Twitter	Uses R for data science applications on the Twitter database
FourSquare	Uses R to develop its recommendation engine
Facebook	Uses R to model all sorts of user behaviour

Role Of A Data Scientist

- Develop and plan required analytic projects in response to business needs.
- In conjunction with data owners and department managers, contribute to the development of data models and protocols for mining production databases.
- Develop new analytical methods and/or tools as required.
- Contribute to data mining architectures, modeling standards, reporting, and data analysis methodologies.
- Conduct research and make recommendations on data mining products, services, protocols, and standards in support of procurement and development efforts.
- Work with application developers to extract data relevant for analysis.

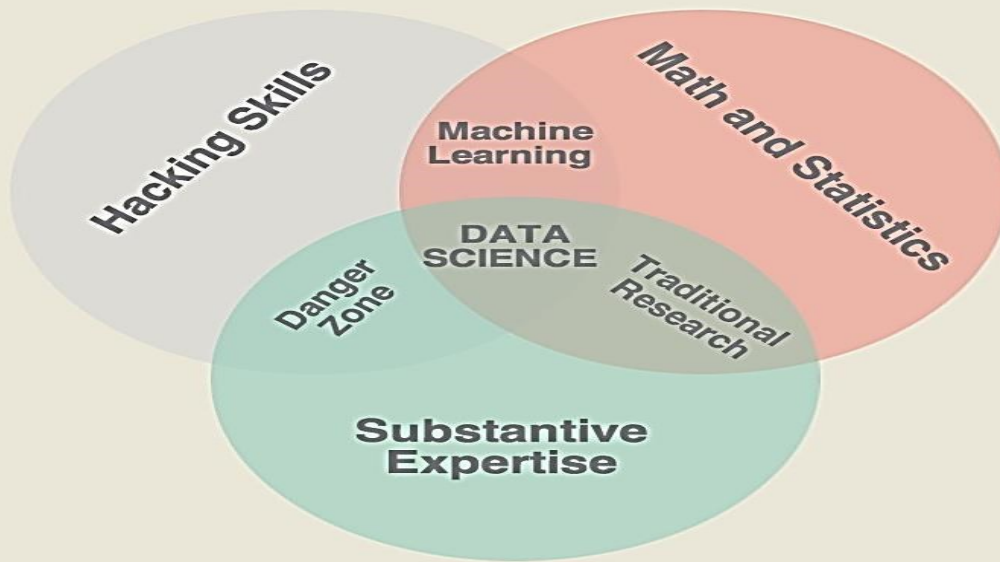
Roadmap To Become A Data Scientist



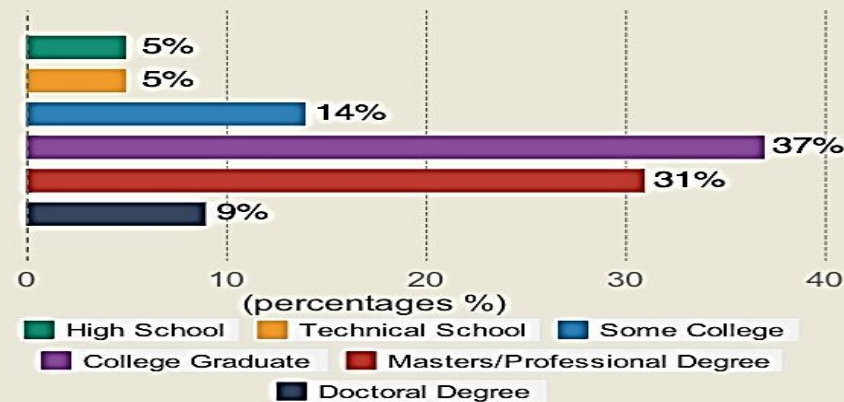
Data Scientist

in 8 easy steps

What's a data scientist?



Typical Background



A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician.