



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Data Science

Data Analytics

Sankara Nayaki K

sankaranayaki@wilp.bits-pilani.ac.in

Introduction to Data Science



Data Analytics Module Topics

- Defining Analytics
- Types of data analytics
 - Descriptive, Diagnostic
 - Predictive, Prescriptive
- Data Analytics – methodologies
 - CRISP-DM Methodology
 - SEMMA
 - BIG DATA LIFE CYCLE
 - SMAM
- Analytics Capacity Building
- Challenges in Data-driven decision making





Defining Analytics

Defining Analytics



What is Analytics? – Dictionary meaning

- It can be used as a noun or verb.
- Let's look at the dictionary meaning first
 - Analytics is the systematic computational analysis of data or statistics
 - -- Oxford
 - Analytics is a process in which a computer examines information using mathematical methods in order to find useful patterns
 - -- Cambridge
 - Analytics is the analysis of data, typically large sets of business data, by the use of mathematics, statistics, and computer software
 - -- dictionary.com



What is Analytics? – From websites

- It is the process of discovering, interpreting, and communicating significant patterns in data and using tools to empower your entire organization to ask any question of any data in any environment on any device
 - --Oracle
- Data Analytics refers to the techniques used to analyze data to enhance productivity and business gain
 - -- Edureka
- Data analytics is the pursuit of extracting meaning from raw data using specialized computer systems
 - -- Informatica
- I couldn't find a definition from popular websites such as Google, IBM, Microsoft, Amazon, etc.,

Defining Analytics



What is Analytics?

- Analytics is the process of extracting and creating information from raw data by using techniques such as:
 - filtering, processing, categorizing, condensing and contextualizing the data
- Analytics is a broad term that encompasses the processes, technologies, frameworks and algorithms to extract meaningful insights from data
- This information thus obtained is then used to infer knowledge about the system and/or its users, and its operations to make the systems smarter and more efficient

Defining Analytics



What is Analytics?

- What drives the choice of technologies, algorithms, and frameworks used for analytics?
 - Analytics goals of the analytic task at hand
 - For example
 - To predict something (for example whether a transaction is a fraud or not, whether it will rain on a particular day, or whether a tumor is benign or malignant)
 - To find patterns in the data (for example, finding the top 10 coldest days in the year, finding which pages are visited the most on a particular website, or finding the most searched celebrity in a particular year)
 - To find relationships in the data (for example, finding similar news articles, finding similar patients in an electronic health record system, finding related products on an eCommerce website, finding similar images, or finding correlation between news items and stock prices)

What is Analytics?

- The National Research Council characterized computational tasks for massive data analysis into seven groups (aka "giants")
- These groups include:
 - (1) Basis Statistics
 - (2) Generalized N-Body Problems
 - (3) Linear Algebraic Computations
 - (4) Graph-Theoretic Computations
 - (5) Optimization
 - (6) Integration and
 - (7) Alignment
- This grouping of computational tasks provides a taxonomy of tasks that is useful in analyzing data according to mathematical structure and computational strategy



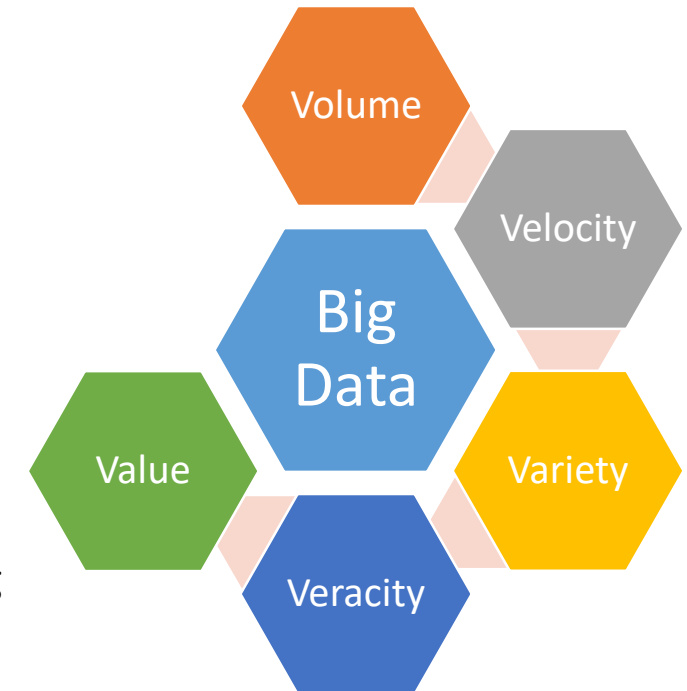
Defining Analytics



Characteristics of Big Data (5 Vs)

- Volume

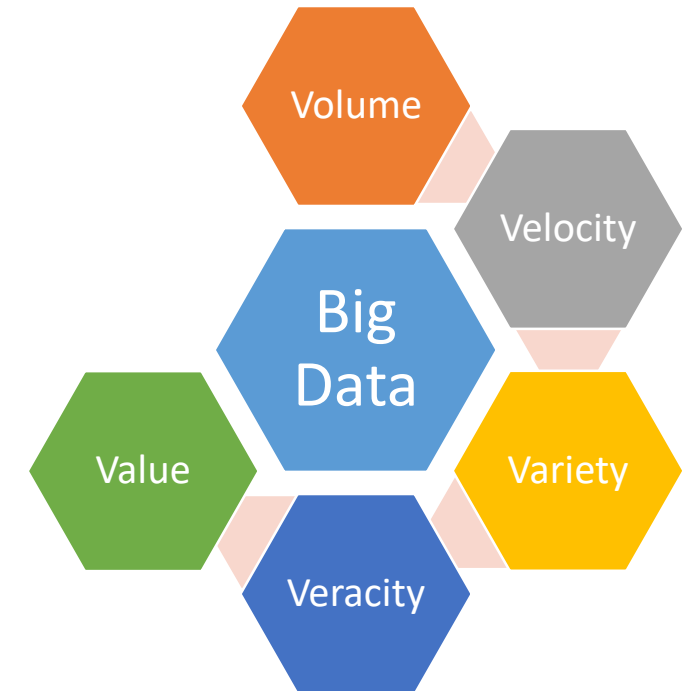
- Big data is so large that it would not fit on a single machine
- Specialized tools and frameworks are required to store, process and analyze such data
- For example:
 - Social media applications process billions of messages everyday, IoT can generate terabytes of sensor data everyday, etc.
- The data generated by modern IT, industrial, healthcare, Internet of Things, and other systems is growing exponentially
- Cost of storage and processing architectures is going down
- The need to extract valuable insights from the data to improve business processes, efficiency and service to consumers
- Though there is no fixed threshold for the volume of data to be considered as big data
- Typically, the term big data is used for massive scale data that is difficult to store, manage and process using traditional databases and data processing architectures



Characteristics of Big Data?

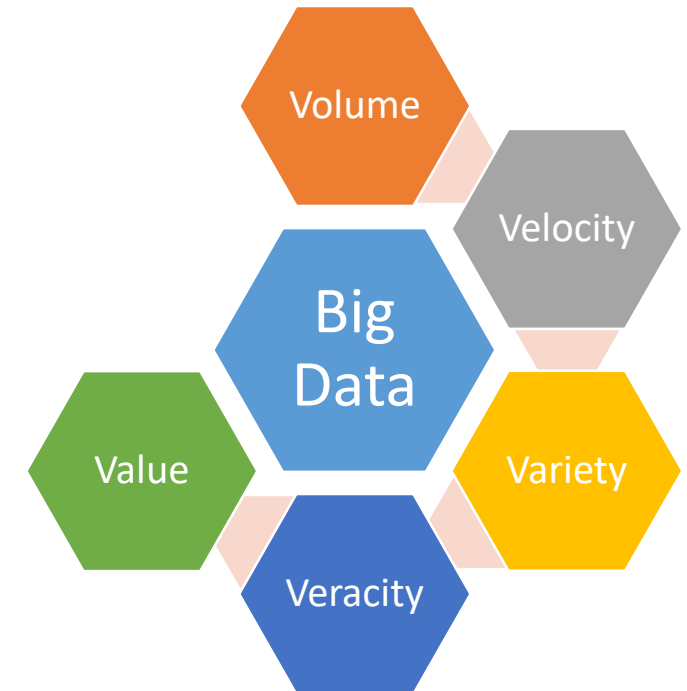
- Velocity

- Velocity of data refers to how fast the data is generated
- Data generated by certain sources can arrive at very high velocities, for example, social media data or sensor data
- Velocity is the primary reason for the exponential growth of data in short span of time
- Some applications can have strict deadlines for data analysis (such as trading or online fraud detection) and the data needs to be analyzed in real-time
- Specialized tools are required to ingest such high velocity data into the big data infrastructure and analyze the data in real-time



Characteristics of Big Data?

- Variety
 - Variety refers to different forms of the data
 - Big data comes in different forms such as structured, unstructured or semi-structured, including text data, image, audio, video and sensor data
 - Big data systems need to be flexible enough to handle such variety of data
- Veracity
 - Veracity refers to how accurate is the data
 - To extract value from the data, the data needs to be cleaned to remove noise
 - Data-driven applications can reap the benefits of big data only when the data is meaningful and accurate
 - Therefore, cleansing of data is important so that incorrect and faulty data can be filtered out



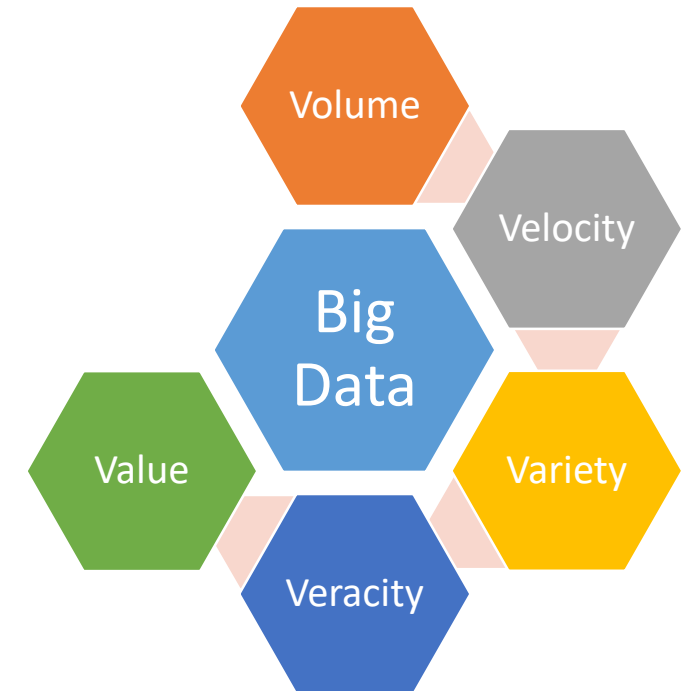
Defining Analytics



Characteristics of Big Data?

- Value

- Value of data refers to the usefulness of data for the intended purpose
- The end goal of any big data analytics system is to extract value from the data
- The value of the data is also related to the veracity or accuracy of the data
- For some applications value also depends on how fast we are able to process the data



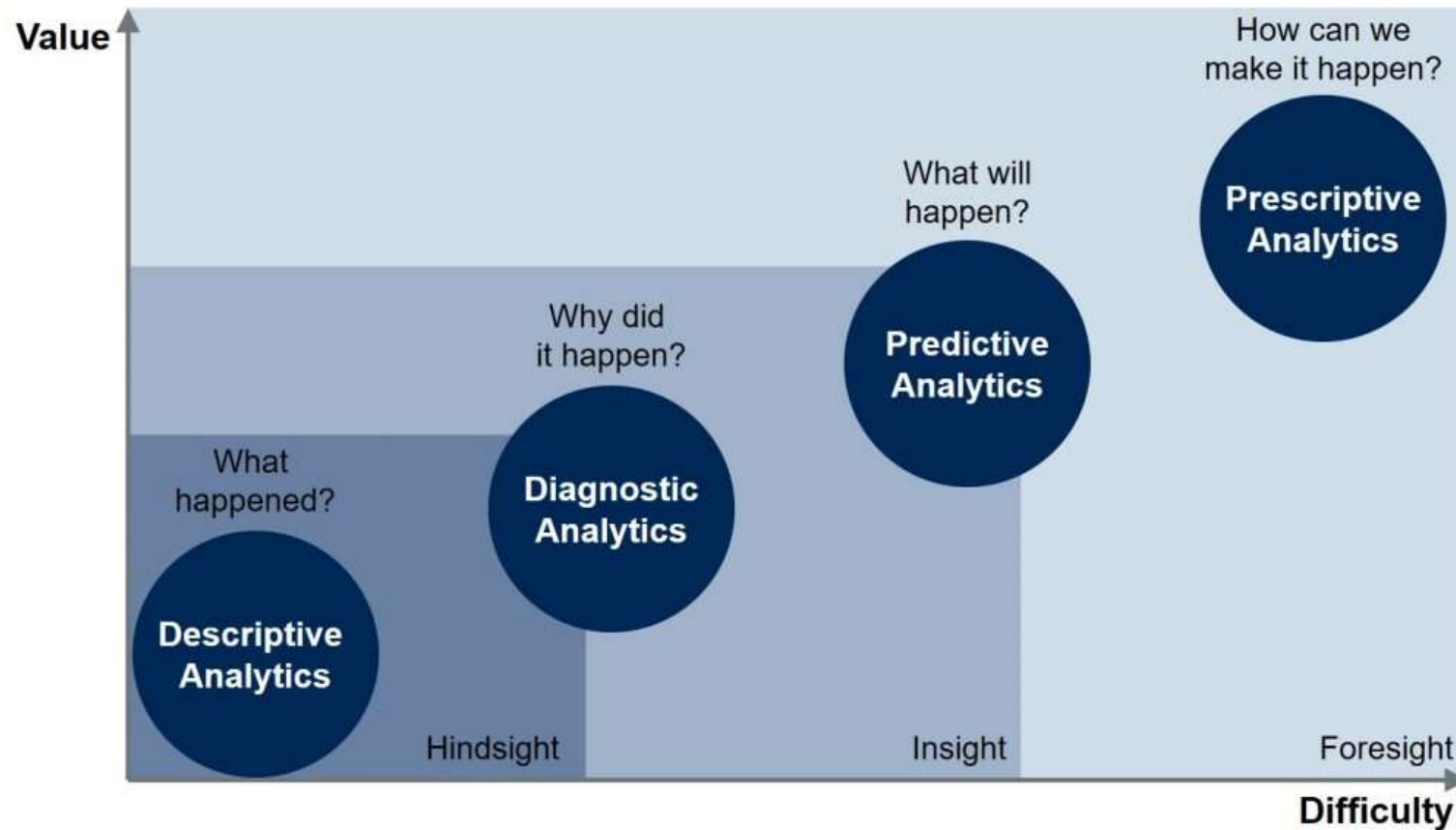


Types of Analytics

Types of Analytics



Types of analytics according to the objective



Types of Analytics



Types of analytics according to the objective

- Descriptive Analytics – What is happening?
 - Involves retrospective analysis of what happened using historical data
 - Data mining methods help uncover patterns that offer insight
 - Lays foundation for diagnosis and further analysis of what might happen in the future
- Diagnostic Analytics – Why did it happen?
 - Looks for the root cause of a problem
 - Used to determine why something happened
 - Attempts to find and understand the causes of events and behaviors
- Predictive Analytics – What is likely to happen?
 - Uses past data and external sources to predict the future
 - Involves forecasting
 - Uses data mining and AI techniques to analyze current data and develop future scenarios
- Prescriptive Analytics – What should be done?
 - Dedicated to finding the right action to be taken.
 - Descriptive analytics provides a historical perspective, and predictive analytics helps forecast what might happen.
 - Prescriptive analytics uses these parameters to find the best solution

Types of Analytics



Next level of Analytics

- Cognitive Computing

- Human cognition is based on the context and reasoning
- Cognitive systems mimic how humans reason and process
- Cognitive systems analyze information and draw inferences using probability



- They continuously learn from data and reprogram themselves
- According to one source
 - "The essential distinction between cognitive platforms and artificial intelligence systems is that you want an AI to do something for you. A cognitive platform is something you turn to for collaboration or for advice"

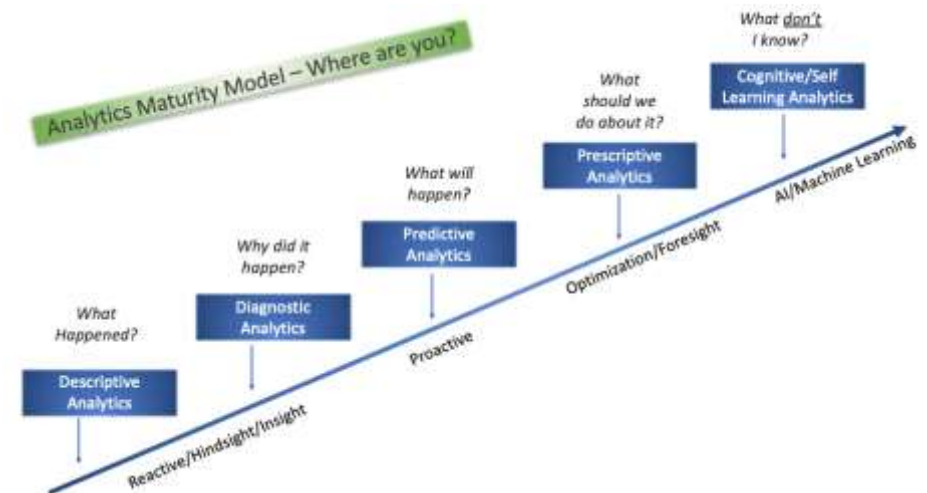
Types of Analytics



Next level of Analytics

- Cognitive Analytics – *What Don't I Know?*

- Involves Semantics, AI, Machine learning, Deep Learning, Natural Language Processing, and Neural Networks
- Simulates human thought process to learn from the data and extract the hidden patterns from data
- Uses all types of data: audio, video, text, images in the analytics process
- Although this is the top tier of analytics maturity, Cognitive Analytics can be used in the prior levels
- It extends the analytics journey to areas that were unreachable with more classical analytics techniques like business intelligence, statistics, and operations research."
 - -- Jean Francois Puget,
- Healthcare, Financial Services, Supply Chains, Customer Relationship Management, Cybersecurity, etc.,.



Types of Analytics



Types of analytics according to the domain

- Marketing Analytics
- Financial Analytics
- Healthcare Analytics
- Sports Analytics
- HR Analytics
- Customer Analytics
- Web Analytics
- Social Analytics
- Cultural Analytics
- Political Analytics



Types of Analytics



Types of analytics according to the type of data

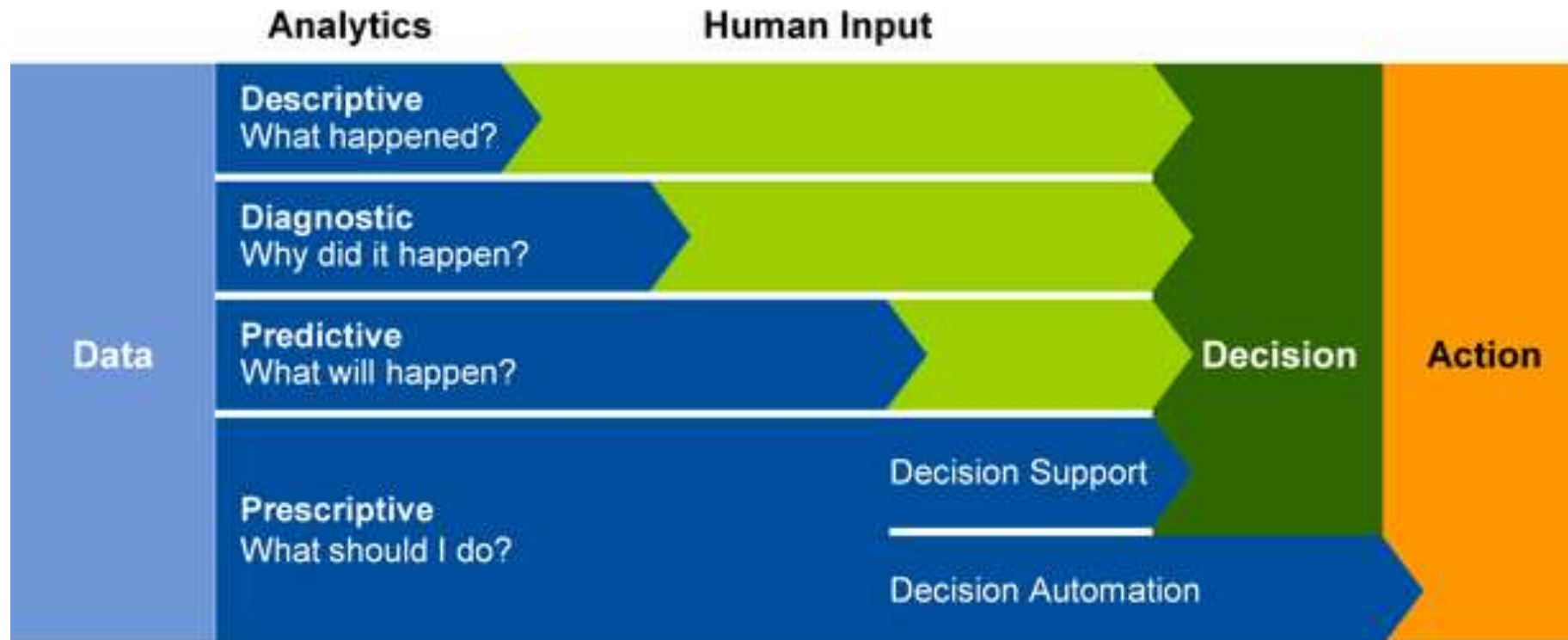
- Text analytics
- Real-time data analytics
- Multimedia analytics
- Geo analytics
- Mobile analytics



Types of Analytics



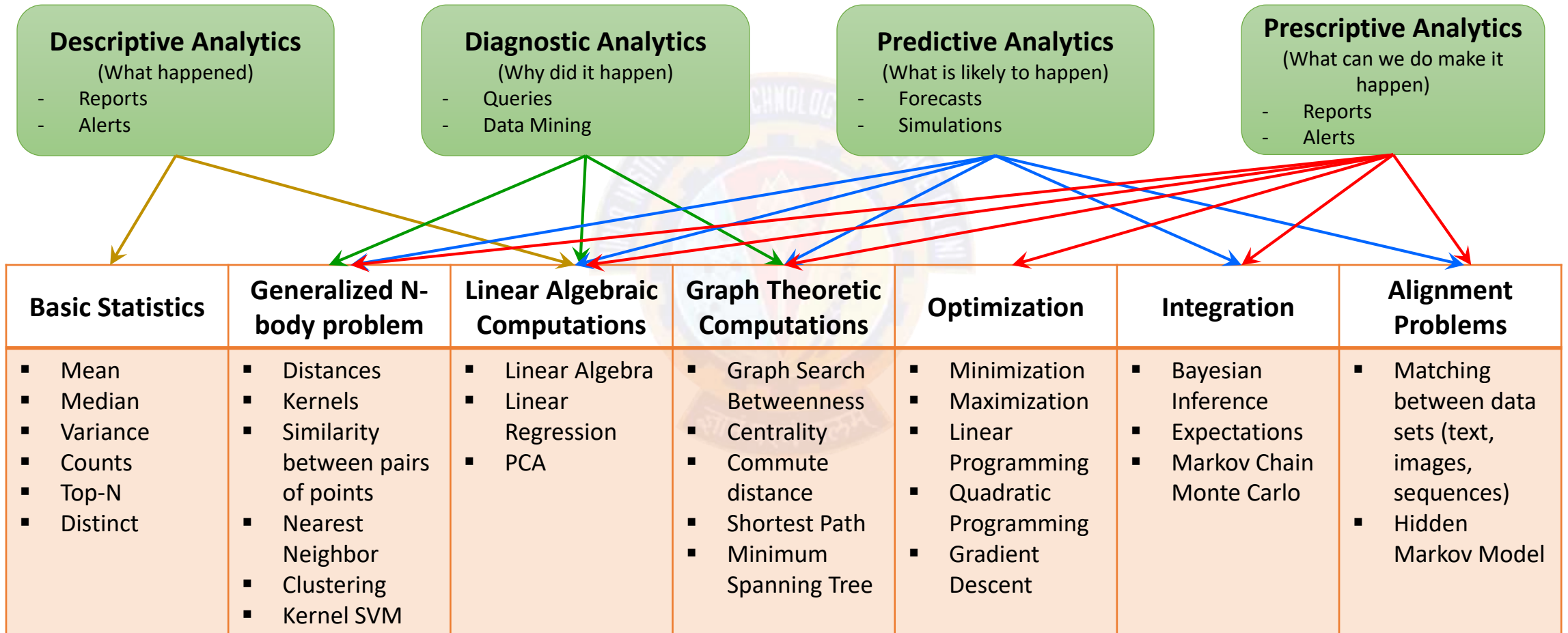
Level of automation in different types of analytics



Types of Analytics



Mapping between types of analytics and computational tasks or 'giants'



Data Analytics

innovate

achieve

lead

THE INFLUENCE OF ANALYTICS ON BUSINESS PERFORMANCE





Case Studies on Data Analytics

Descriptive Analytics



Example#1

Problem Statement : “Market research team at Aqua Analytics Pvt. Ltd is assigned a task to identify profile of a typical customer for a Digital fitness band that is offered by Titanic Corp. The market research team decides to investigate whether there are differences across the usage patterns and product lines with respect to customer characteristics”

- Gender
 - Age (In years)
 - Education(In years)
 - Relationship Status(Single or Partnered)
 - Annual Household income
 - Average number of times customer tracks activity each week
 - Number of miles customer expect to walk each week
 - Self-rated fitness on a scale 1–5 where 1 is poor shape and 5 is excellent
- Models of the product purchased—IQ75, MZ65, DX87

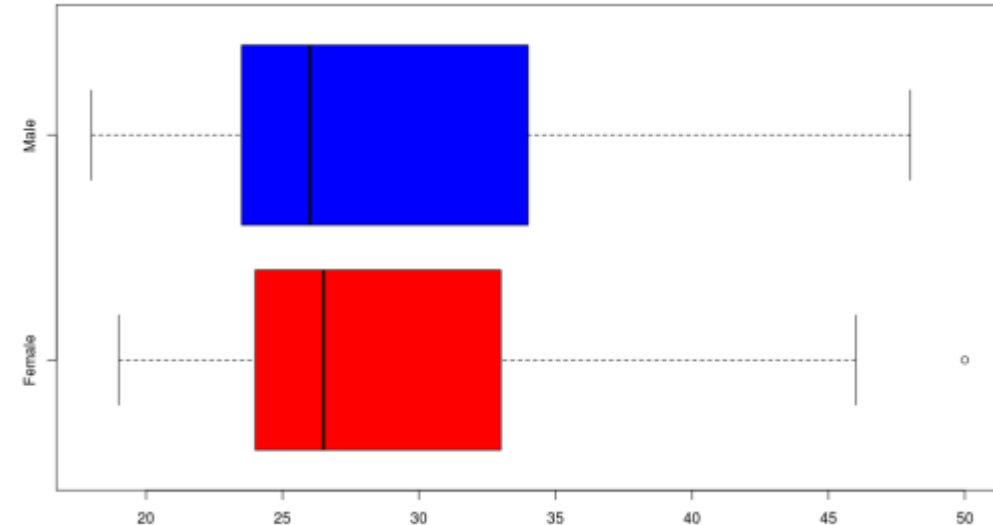
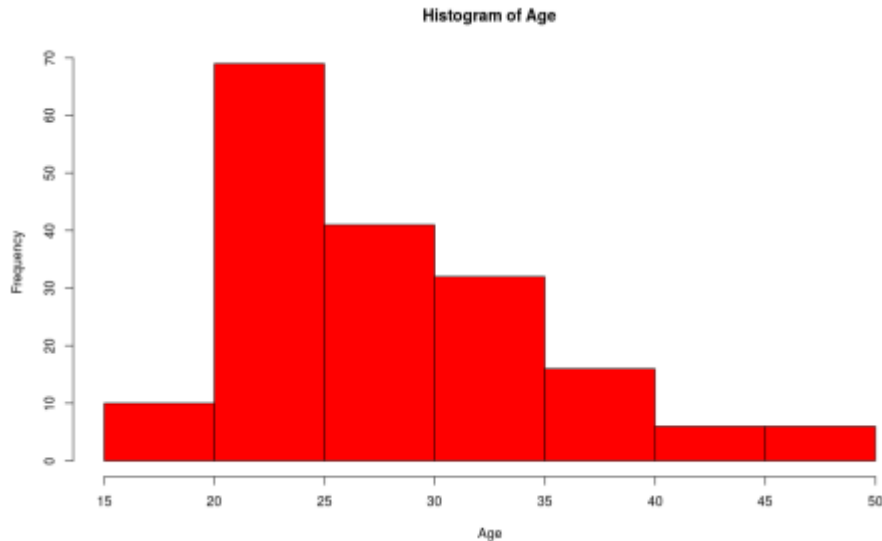
Descriptive Analytics



Results of Descriptive Analytics

Most of the customers use it 3–4 times a week. No female consumer has ever used IQ75 and MZ65 more than 5 time a week.

Product	Usage	2		3		4		5		6		7		Totals
	Gender													
		Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Male		
DX87					1	2	16	3	9	2	5	2	40	
IQ75		13	6	10	18	7	15	1	1				80	
MZ65		7	7	14	17	5	7	3					60	
Totals		20	13	33	36	14	38	7	10	2	5	2	180	

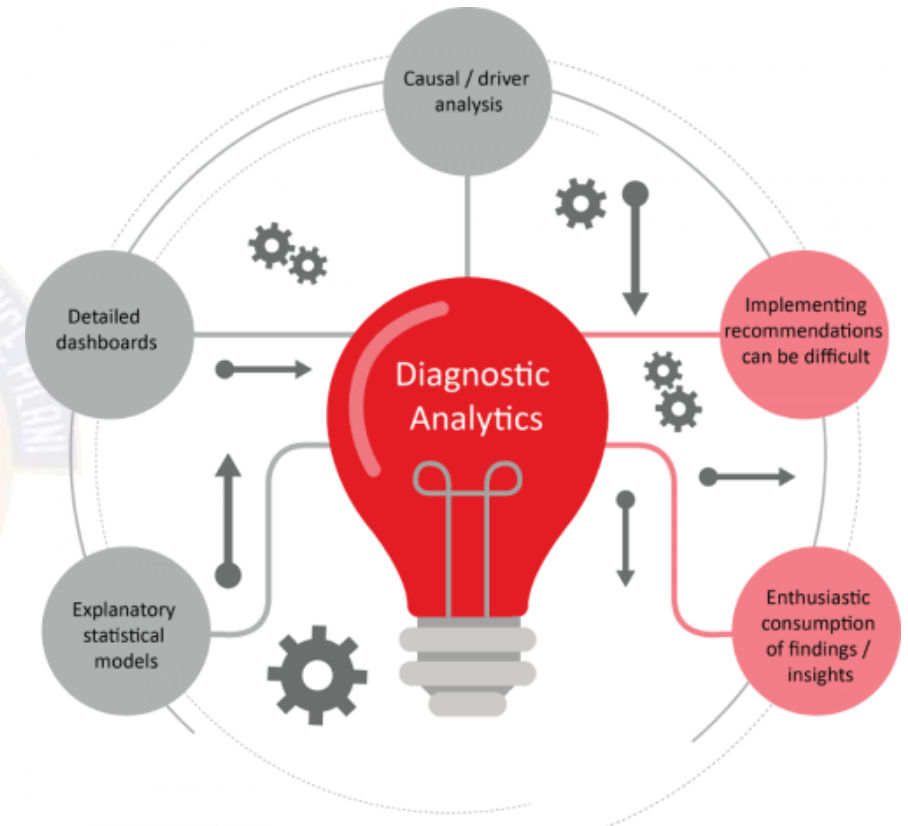


Diagnostic Analytics



Example#1

Problem Statement : During the 1980s General Electric was selling different products to its customers such as light bulbs, jet engines, windmills, and other related products. Also, they separately sell parts and services this means they would sell you a certain product you would use it until it needs repair either because of normal wear and tear or because it's broken. And you would come back to GE and then GE would sell you parts and services to fix it. Model for GE was focusing on how much GE was selling, in sales of operational equipment, and in sales of parts and services. And what does GE need to do to drive up those sales??



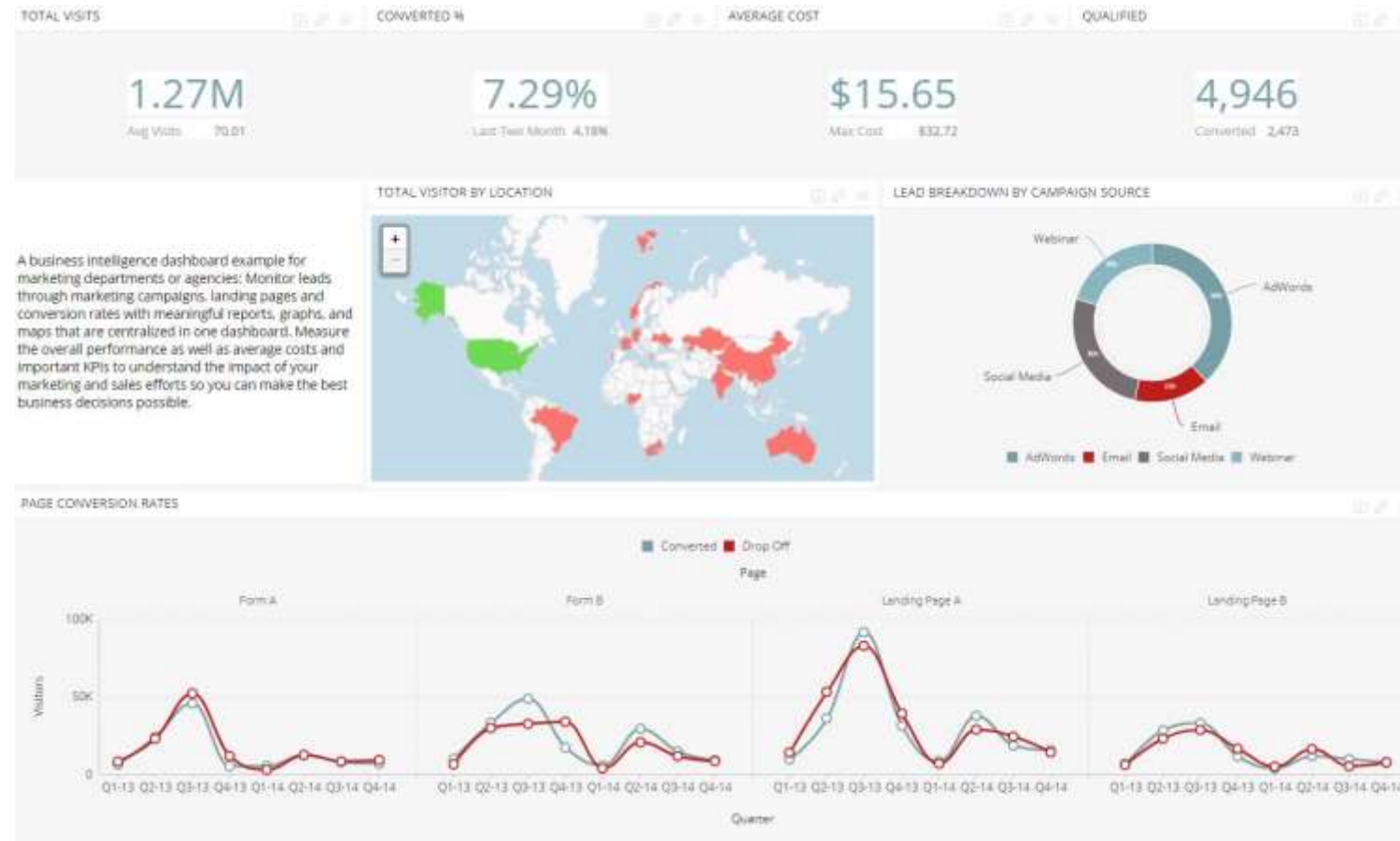
Diagnostic Analytics

innovate

achieve

lead

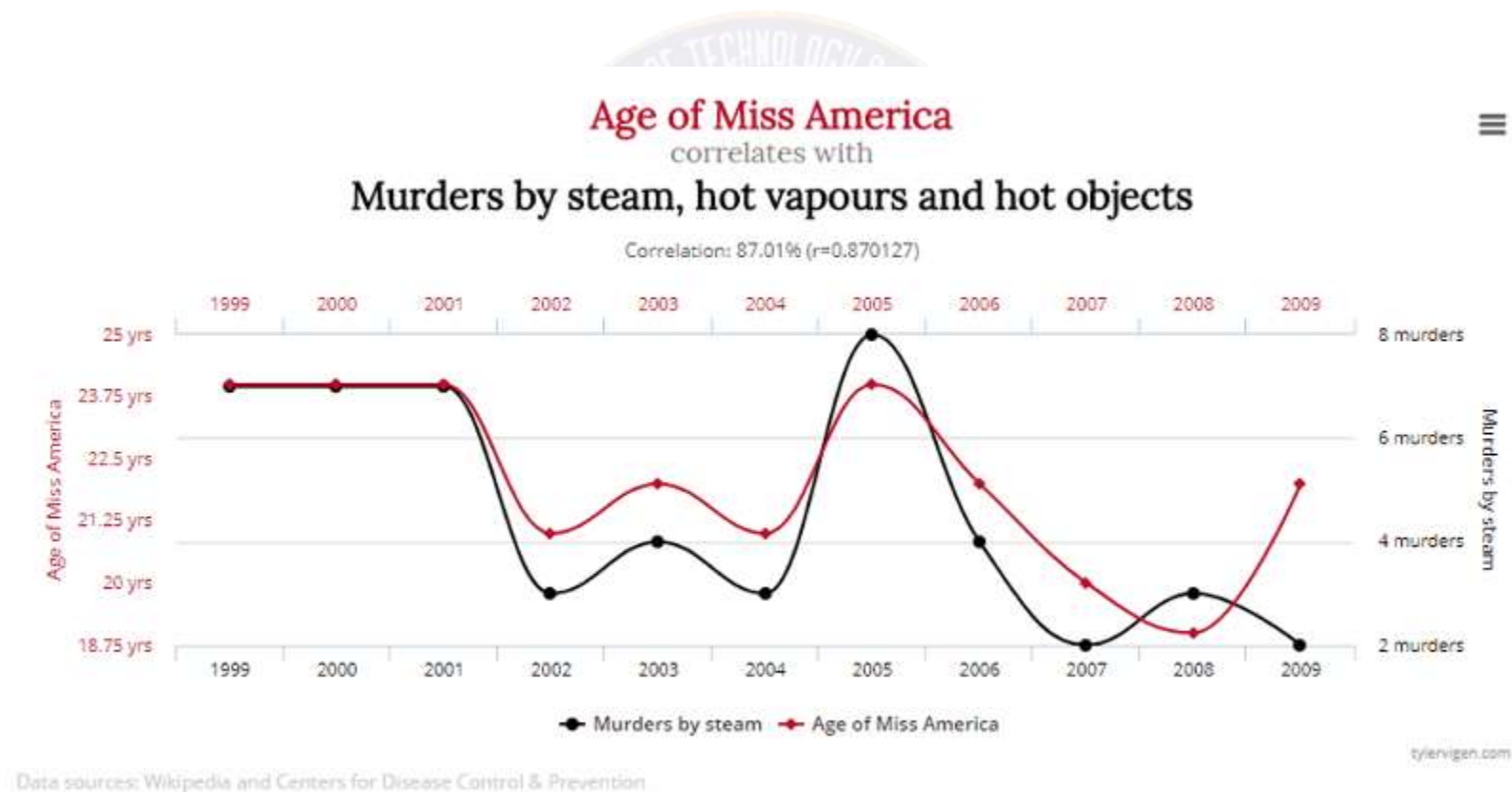
Results of Diagnostic Analytics



Diagnostic Analytics



New Perspective can also be the result of Diagnostic Analytics



Predictive Analytics



Example#1

- Google launched **Google Flu Trends (GFT)**, to collect predictive analytics regarding the outbreaks of flu. It's a great example of seeing big data analytics in action.
- So, did Google manage to predict influenza activity in real-time by aggregating search engine queries with this big data and adopting predictive analytics?
- Even with a wealth of big data analytics on search queries, [GFT overestimated the prevalence of flu](#) by over 50% in 2012-2013 and 2011-2012.

They matched the search engine terms conducted by people in different regions of the world. And, when these queries were compared with traditional flu surveillance systems, *Google found that the predictive analytics of the flu season pointed towards a correlation with higher search engine traffic for certain phrases.*



In 2012 Google announced its ability to predict a flu outbreak in the U.S. five days before it would happen based on big data and search patterns.

In 2015 the center for disease control revealed that Google's prediction was twice the size of the correct number.

Predictive Analytics



Example#2

- Colleen Jones applied predictive analytics to FootSmart (a niche online catalog retailer) on a content marketing product. It was called the [FootSmart Health Resource Center](#) (FHRC) and it consisted of articles, diagrams, quizzes and the like. On analyzing the data around increased [search engine visibility](#), FHRC was found to help FootSmart reach more of the *right kind of target customers*.
- They were receiving more traffic, primarily consisting of people that cared about foot health conditions and their treatments.
- So, FootSmart decided to push more content at FHRC and also improve its merchandising of the product.

The result of such informed data-driven decision-making?

A 36% increase in weekly sales.



Predictive Analytics

innovate

achieve

lead

Results of Predictive Analytics

Is your organization trying to find new ways to generate revenue?

Most organizations apply predictive analytics to core functions that produce revenue.



52%

use predictive analytics to increase profitability

55%

use predictive analytics to create new revenue opportunity

Are you looking for ways to increase customer satisfaction?

45% of organizations currently use predictive analytics for customer services.



What data in your organization can predictive analytics tap to help you discover new trends and opportunities?

The top 5 sources of data tapped for predictive analytics:



Are you using the right product offers and targeting the best customers for your up-sell and cross-sell?

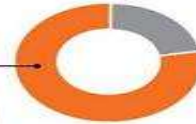
43% of organizations currently use the results of predictive analytics for product recommendations and offers.



What impact can predictive analytics have on your organization?

86% assert that predictive analytics will have a major positive impact on their organization.

with nearly **one-third** indicating it could be transformative in enabling them to do things they couldn't do before.



How can your organization gain a competitive edge and respond in real time?

68% of organizations who use predictive analytics have realized a competitive advantage.

With **real-time predictive analytics**, you can make sure your company doesn't miss its window of opportunity.



What are the 5 things predictive analytics can do for you?

- 1 Instantly predict market trends and customer needs
- 2 Create customized offers for each segment and channel
- 3 Predict how market-price volatility will impact your production plans
- 4 Foresee changes in demand and supply across your entire supply chain
- 5 Proactively manage your workforce by attracting and retaining talent

Where do you want your company analytics to be?

Predictive analysis enables you to extend your analytics capabilities: Moving from the rearview mirror to a **forward-looking view**.



Predictive Policing – Seeing the crime before it happens

- Refers to the usage of mathematical, probabilistic, geospatial, and other analytical techniques in law enforcement to identify potential criminal activity
- This involves two aspects:
 - Applying advanced analytics to various data sets to predict a crime
 - Offer intervention (deploying resources in the areas where there is a potential for crime)
- Predictive policing falls into four categories
 - Predicting crimes
 - Predicting offenders
 - Predicting perpetrators' identities
 - Predicting victims of crime



Predictive Policing – Seeing the crime before it happens

- At the start of each shift, crime data fed into the PredPol system provides officers with 15 different zones for four types of crime:
 - Auto theft, vehicle burglary, burglary and gang-related activity
- Each zone covers an area of 500 square feet

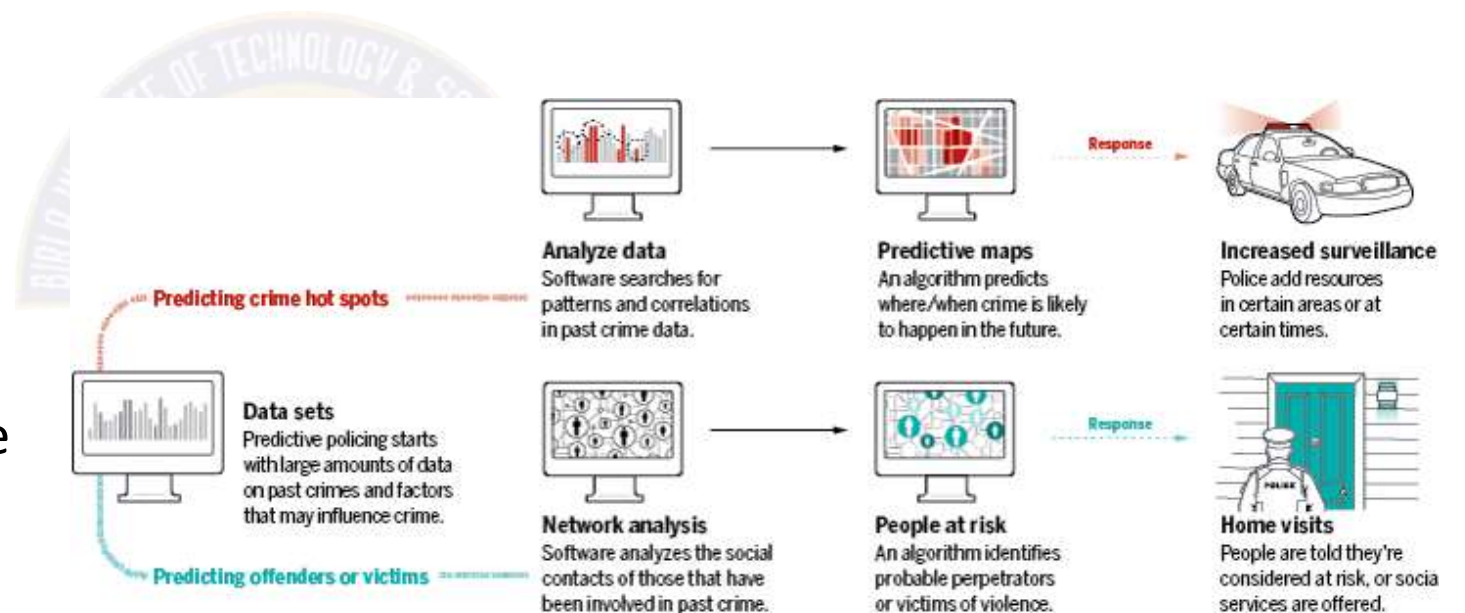


Predictive Analytics



Predictive Policing – Predicting crime hot spots

- Descriptive and Diagnostic Analytics
 - Software searches for patterns and correlations in past crime data
- Predictive Analytics
 - Algorithm predicts when/where crime is likely to happen in the future
- Response
 - Police deploy resources in certain areas at certain times

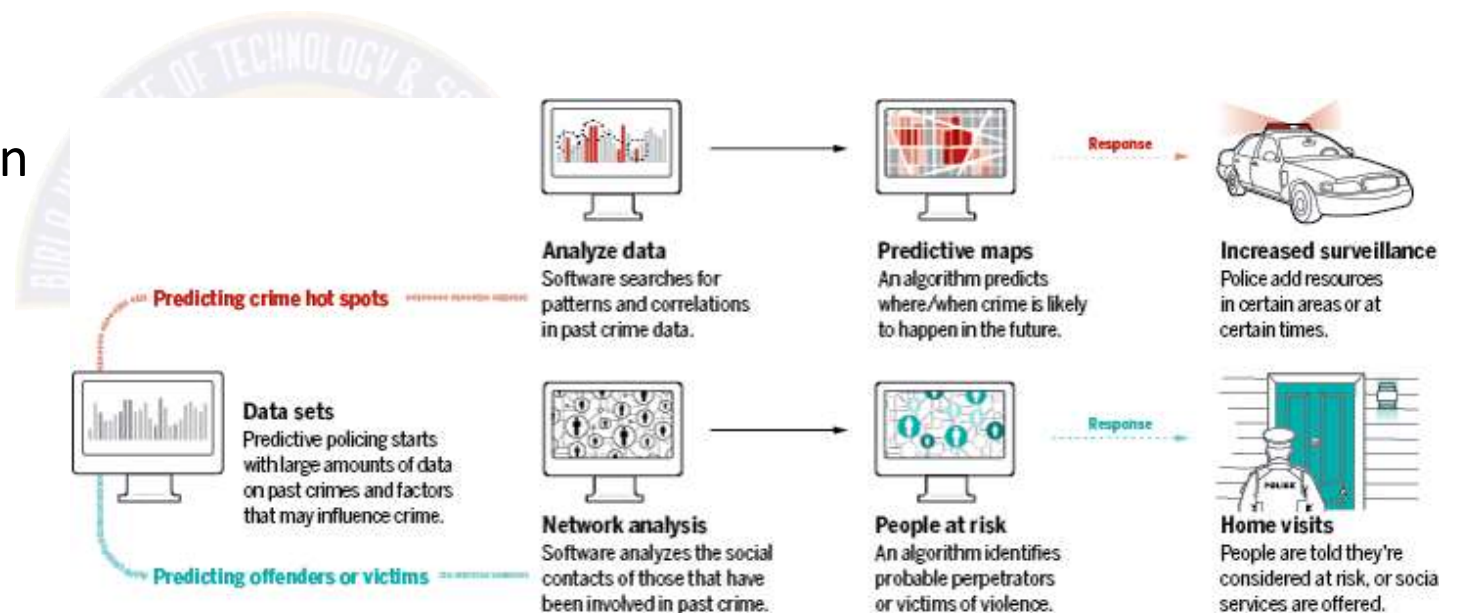


Predictive Analytics



Predictive Policing – Predicting crime hot spots

- Descriptive and Diagnostic Analytics
 - Software analyzes the social contacts of those that have been involved in the past
- Predictive Analytics
 - Algorithm identifies probable perpetrators or victims of violence
- Response
 - People are told they're considered at risk, or social services are offered



Prescriptive Analytics



Example#1

- Sajan Kuttappa, a product marketing manager at IBM, a health insurance company analyzes its data and determines that many of its diabetic patients also suffer from retinopathy.
- With this information, the provider can now use predictive analytics to get an idea of how many more ophthalmology claims it might receive during the next year.
- Then, using prescriptive analytics, the company can look at scenarios where the reimbursement costs for ophthalmology increases, decreases, or holds steady. These scenarios then allow them to make an informed decision about how to proceed in a way that's both cost-effective and beneficial to their customers.
- Analyzing data on patients, treatments, appointments, surgeries, and even radiologic techniques can ensure hospitals are properly staffed, the doctors are devising tests and treatments based on probability rather than gut instinct, and the facility can save costs on everything from medical supplies to transport fees to food budgets.



Example#2

- Whenever you go to Amazon, the site recommends dozens and dozens of products to you. These are based not only on your previous shopping history (reactive), but also based on what you've searched for online, what other people who've shopped for the same things have purchased, and about a million other factors (proactive).
- Amazon and other large retailers are taking deductive, diagnostic, and predictive data and then running it through a prescriptive analytics system to find products that you have a higher chance of buying.
- Every bit of data is broken down and examined with the end goal of helping the company suggest products you may not have even known you wanted.

YouTube's algorithm factors in billions of data points in order to create a customized viewing experience unique to you every time you visit the site's home page. The more data you give the algorithm (by selecting videos, liking and disliking, subscribing, leaving comments, and watch time), the better it gets at surfacing videos that are likely to be of interest to you.

Back over in retail, prescriptive analytics can also help with scheduling, shipping logistics, inventory control, and countless other ways. There really aren't many things it can't provide insights for.



Example#3

- **Leading bulk tanker transportation company**
- Driving innovation in bulk tanker transportation with advanced analytics and optimization
- Bulk tanker logistics is such a complex problem that many trucking companies can't offer a comprehensive service: they have to specialize in only a small set of products. Optimization software from IBM is helping this leading tanker carrier to buck that trend by finding efficient routes in near-real time—enabling huge cost savings and supporting growth.

Business challenge

To transport bulk products safely and profitably, this carrier needs to manage hundreds of constraints on tankers, drivers and cargos. How can it help its planners make optimal routing decisions?

Transformation

This leading bulk carrier embedded IBM optimization software into its operational systems and developed a sophisticated solution that provides insight to optimize driver and route planning every 10 minutes.

Prescriptive Analytics



Results of Prescriptive Analytics

Results

Millions

of dollars saved annually by
eliminating miles of unnecessary
driving

Millions

more dollars saved annually by
improving driver retention

Million-dollar

revenue boost achieved by
increasing driver productivity



Healthcare Case Study

Healthcare Analytics – Case Study



Healthcare Practices

- Private Practice
- Group Practice
- Large HMOs
- Hospital Based
- Locum Tenens



Healthcare Analytics – Case Study



Healthcare Practice

- Private Practice

- A physician practices alone without any partners and typically with minimal support staff
- Ideally works for physicians who wish to own and manage their own practice
- Benefits
 - Individual freedom, closer relationships with patients, and the ability to set their own practice's growth pattern
- Drawback
 - Longer work hours, financial extremes, and a greater amount of business risk.



Healthcare Analytics – Case Study



Healthcare Practice

- Group Practice
 - A group practice involves two or more physicians who all provide medical care within the same facility
 - They utilize the same personnel and divide the income in a manner previously agreed upon by the group
 - Group practices may consist of providers from a single specialty or multiple specialties
 - Benefits
 - Shorter work hours, built-in on-call coverage, and access to more working capital
 - All of these factors can lead to less stress
 - Drawback
 - Less individual freedom, limits on the ability to rapidly grow income, and the need for a consensus on business decisions.



Healthcare Analytics – Case Study



Healthcare Practice

- Large HMOs (Health Maintenance Organizations)
 - HMO, employs providers to care for their members and beneficiaries
 - The goal of HMOs is to decrease medical costs for those consumers
 - There are two main types of HMOs
 - The prepaid group practice model and the medical care foundation (MCF), also called individual practice association.
 - Benefits
 - More stable work life and regular hours for providers (physicians)
 - Less paperwork and regulatory responsibilities and a regular salary along with bonus opportunities
 - These bonuses are based on productivity or patient satisfaction
 - Drawback
 - The main drawback for physicians working for an HMO is the lack of autonomy
 - HMO's required physicians to follow their guidelines in providing care.
 - Examples:
 - The Kaiser Foundation Health Plan in California
 - The Health Insurance Plan of Greater New York
 - Group Health Cooperative of Puget Sound

Healthcare Analytics – Case Study



Healthcare Practice

- Hospital Based
 - In hospital based work, physicians earn a predictable income, have a regular patient base, and a solid referral network
 - Physicians who are employed by a hospital will either work in a hospital-owned practice or in a department of the hospital itself
 - Benefits
 - A regular work schedule, low to no business and legal risk, and a steady flow of income
 - Drawback
 - Lack of physician autonomy
 - Employee constraints and the expectation that physicians become involved in hospital committee work



Healthcare Analytics – Case Study



Healthcare Practice

- Locum Tenens

- Locum tenens is derived from the Latin phrase for "to hold the place of."
- In locum tenens, physicians re-home to areas hurting for healthcare professionals
- These types of positions offer temporary employment and may offer higher pay than more permanent employment situations
- Benefits
 - Physicians working in locum tenens scenarios enjoy the benefits of variety and the ability to experience numerous types of practices and geographic locations
 - They also enjoy schedule flexibility and lower living costs
- Drawback
 - The possibility that benefits are not included, and a potential lack of steady work
 - Physicians need to regularly uproot their families

Healthcare Analytics – Case Study



Healthcare Scenario – Sources of Data

- Common sources of practice data existing within the practice:
 - Revenue Cycle Management System (billing for a procedure, processing denials, collecting payments)
 - Electronic Health Records (EHRs) (provides services to a patient)
 - Scheduling & Information System (schedules an appointment)
 - Survey Results
 - Peer Review System
 - Other sources
- Analytics provides a platform to convert that data into actionable information for the practice
- As reimbursement shifts to a value-based care model, it is critical to have insight into both clinical and business metrics to prepare for the future

Healthcare Analytics – Case Study



Healthcare Scenario – Integrated Medical Partners (IMP)

- <https://www.youtube.com/watch?v=olpuyn6kemg>
- IMP
 - Primarily offers
 - Revenue Cycle Management (RCM) and Advanced Analytics among others
 - Helps maximize collections for independent physician practices
 - Provides practice performance analytics
 - Offers tailored solutions to partner with physicians and hospitals so that they can
 - Improve patient care, enhanced compliance, operational efficiency, and increase profitability

Healthcare Analytics – Case Study



Healthcare Scenario – Denial of Claims

- Consider that the healthcare practice has seen an increase in denied claims over the past several months
- Increased denials impact negatively on the financial performance of the organization
- Therefore, the company needs tools to help identify and resolve the root cause of the denials
 - This trend needs to be reversed as soon as possible, but how?
- The different types of analytics provide necessary insights into the data and the cause of the denial increase

Healthcare Analytics – Case Study



Healthcare Scenario – Descriptive Analytics

- Descriptive – *what happened?*
 - Descriptive analytics will tell you what is happening in the practice
 - In this example, there has been an increase in the number of denied claims over the past several months
 - Further research identifying a trend revealed that the increase in denials is specific to a particular denial code
 - What is this denial code?
 - This denial code is for a referring provider (doctor) who is not enrolled in the Medicare Provider Enrollment, Chain, and Ownership System (PECOS).
- Now, the question is why this provider is not enrolled in PECOS?

Healthcare Analytics – Case Study



Healthcare Scenario – Diagnostic Analytics

- Diagnostic – *why did it happen?*
 - The descriptive analytics has identified an increase in denials specific to a referring provider
 - The next step is to diagnose to understand why this change occurred
 - Identify why the referring provider is not enrolled in PECOS
 - So, we utilize diagnostic analytics to understand why this change occurred
 - Looking for changes in referring provider patterns at the time the increase in PECOS denials began would help to identify new referring providers
 - It will also identify whether these new referring providers are enrolled in PECOS
 - For the purpose of this example, assume we identify one referring provider who is new to referring to your practice that is not enrolled in PECOS.

Healthcare Analytics – Case Study



Healthcare Scenario – Predictive Analytics

- Predictive – what will happen?
 - Predictive Analytics allows us to learn from historical trends to predict what will happen in the future
 - Utilizing descriptive and diagnostic analytics, we can determine the historical referral pattern of the new referring provider not enrolled in PECOS
 - Assuming this provider remains not enrolled in PECOS and refers a steady stream of patients to the organization
 - predictive analytics will tell us the expected denials associated with these claims
 - The resulting impact of this situation is identified

Healthcare Analytics – Case Study



Healthcare Scenario – Prescriptive Analytics

- Prescriptive – what should I do?
 - Prescriptive Analytics assists in determining the best course of action from the information gathered from descriptive, diagnostic and predictive analytics
 - In this scenario, the best course of action would be to contact the new referring provider, express appreciation for the new referrals, but also evidence that the provider's referrals are resulting in denied claims due to PECOS enrollment
 - By working with this provider to become enrolled in PECOS, denied claims can now be rebilled
 - In addition, future claims are less likely to be denied due to the provider not being certified/eligible to be paid for the procedure/service on the claim date of service
 - The key to prescriptive analytics is implementing a solution that will prevent the same breakdown from occurring again in the future
 - By reducing the likelihood of claim denials, the overall financial health of the physician practice benefits



Thank You!