



BITS Pilani

Pilani | Dubai | Goa | Hyderabad

INTRODUCTION TO DATA SCIENCE

MODULE # 2 : DATA ANALYTICS

IDS Course Team

BITS Pilani

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

TABLE OF CONTENTS



- 1 ANALYTICS
- 2 BIG DATA
- 3 DATA ANALYTICS
- 4 CASE STUDIES ON DATA ANALYTICS

DEFINITION OF ANALYTICS – DICTIONARY



OXFORD Analytics is the systematic computational analysis of data or statistics.

CAMBRIDGE Analytics is a process in which a computer examines information using mathematical methods in order to find useful patterns.

DICTIONARY.COM Analytics is the analysis of data, typically large sets of business data, by the use of mathematics, statistics, and computer software.

Analytics is treated as both a noun and a verb.

DEFINITION OF ANALYTICS – WEBSITES



ORACLE Analytics 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.

EDUREKA Data Analytics refers to the techniques used to analyze data to enhance productivity and business gain.

INFORMATICA Data analytics is the pursuit of extracting meaning from raw data using specialized computer systems.

I couldn't find a definition from popular websites such as Google, IBM, Microsoft, Amazon, etc.

Source: Big Data Analytics – A Hands-on Approach by Arshdeep Bahga & Vijay Madisetti



- 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.

GOALS OF DATA ANALYTICS



- To predict something
 - ▶ whether a transaction is a fraud or not
 - ▶ whether it will rain on a particular day
 - ▶ whether a tumor is benign or malignant
- To find patterns in the data
 - ▶ finding the top 10 coldest days in the year
 - ▶ which pages are visited the most on a particular website
 - ▶ finding the most searched celebrity in a particular year
- To find relationships in the data
 - ▶ finding similar news articles
 - ▶ finding similar patients in an electronic health record system
 - ▶ finding related products on an e-commerce website
 - ▶ finding similar images
 - ▶ finding correlation between news items and stock prices



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

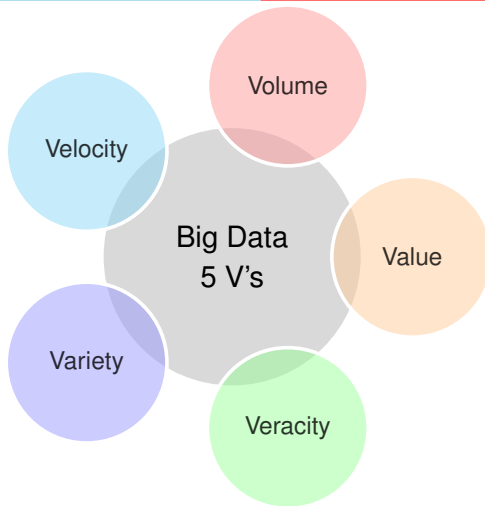
TABLE OF CONTENTS



- 1 ANALYTICS
- 2 BIG DATA
- 3 DATA ANALYTICS
- 4 CASE STUDIES ON DATA ANALYTICS

- Big data is defined as collections of datasets whose volume, velocity or variety is so large that it is difficult to store, manage, process and analyze the data using traditional databases and data processing tools.
- Big Data analytics deals with collection, storage, processing and analysis of this massive scale data.
- Specialized tools and frameworks are required for big data analysis when:
 - 1 the volume of data involved is so large that it is difficult to store, process and analyze data on a single machine
 - 2 the velocity of data is very high and the data needs to be analyzed in real-time
 - 3 there is variety of data involved, which can be structured, unstructured or semi-structured, and is collected from multiple data sources,
 - 4 various types of analytics need to be performed to extract value from the data

CHARACTERISTICS OF BIG DATA



CHARACTERISTICS OF BIG DATA



① Volume

- ▶ Volume of data involved is so large that it is difficult to store, process and analyze data on a single machine.
- ▶ Volumes of data generated by IT / IoT systems is growing exponentially.
 - ★ lowering costs of data storage and processing architectures
 - ★ need to extract valuable insights from the data to improve business processes, efficiency and service to consumers.

② Velocity

- ▶ Velocity of data refers to how fast the data is generated.
- ▶ High velocity of data results in the volume of data accumulated to become very large, in short span of time.

3 Variety

- ▶ Variety refers to the 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.

4 Veracity

- ▶ Veracity refers to how accurate is the data.
- ▶ To extract value from the data, the data needs to be cleaned to remove noise.

5 Value

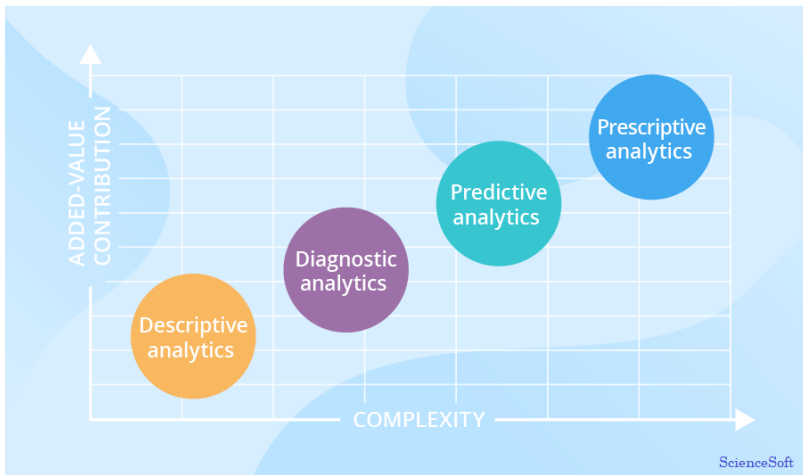
- ▶ Value of data refers to the usefulness of data for the intended purpose.
- ▶ The value of the data is also related to the veracity or accuracy of the data.
- ▶ For some applications value also depends on how fast we are able to process the data.

TABLE OF CONTENTS



- 1 ANALYTICS
- 2 BIG DATA
- 3 DATA ANALYTICS
- 4 CASE STUDIES ON DATA ANALYTICS

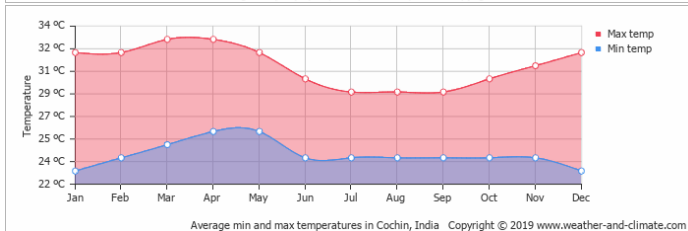
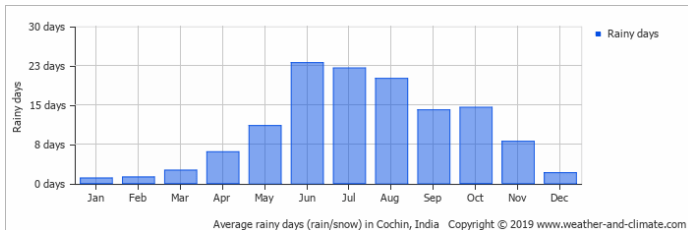
- Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making.
- 4 different types of analytics
 - 1 Descriptive Analytics
 - 2 Diagnostic Analytics
 - 3 Predictive Analytics
 - 4 Prescriptive Analytics





- Answers the question of **what happened**.
- Summarize past data usually in the form of dashboards.
- **Insights into the past**.
- Also known as **statistical analysis**.
- Raw data from multiple data sources.

DESCRIPTIVE ANALYTICS EXAMPLE



- Techniques:
 - ▶ Descriptive Statistics - histogram, correlation
 - ▶ Data Visualization
 - ▶ Exploratory Analysis

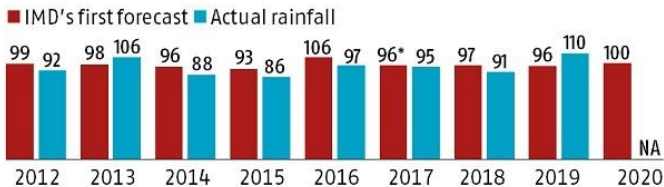


- Answers the question of **what is likely to happen**.
- **Predict future trends**.
- Being able to predict allows one to make better decisions.
- Analysis based on machine or deep learning.
- Accuracy of the forecasting or prediction highly depends on data quality and stability of the situation.

PREDICTIVE ANALYTICS EXAMPLE



MONSOON FORECAST As % of long-period average



*Updated forecast on May 24; NOTE: All the forecasts are with a model error of plus and minus 5%. The first forecast is issued in April every year

LPA is the average rainfall received in the last 50 years, estimated to be 887 mm

Source: IMD



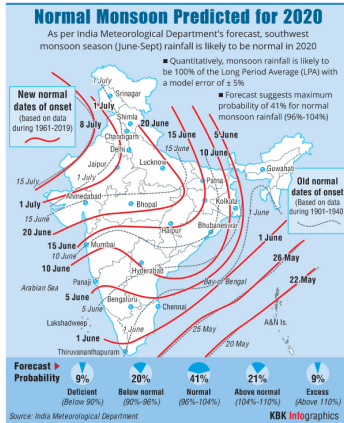
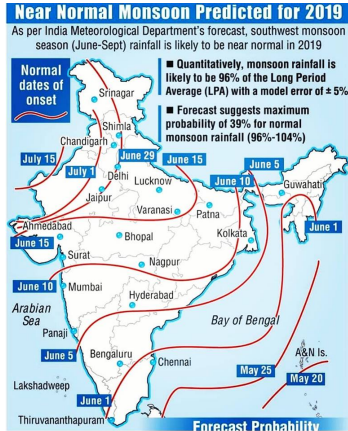
- Techniques / Algorithms:
 - ▶ Regression
 - ▶ Classification
 - ▶ ML algorithms like Linear regression, Logistic regression, SVM
 - ▶ Deep Learning techniques

- Answers the question of **why something happened**.
- Gives in-depth insights into data.
- **Identify relationship** between data and **identify patterns** of behavior.

DIAGNOSTIC ANALYTICS EXAMPLE



What is the effect of global warming in the Southwest monsoon?



- Pattern recognition to identify patterns.
- Linear / Logistic regression to identify relationship.
- Neural Network
- Deep Learning techniques

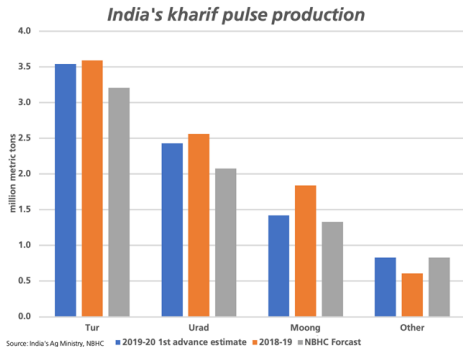


- Answers the question of **what might happen**.
- **Data-driven decision making and corrective actions**
- Prescribe what action to take to eliminate a future problem or take full advantage of a promising trend.
- Need historical internal data and external information like trends.
- Analysis based on machine or deep learning, business rules.
- Use of AI to improve decision making.

PRESCRIPTIVE ANALYTICS EXAMPLE

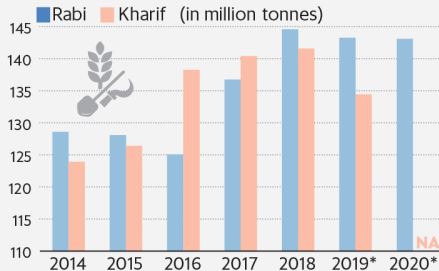


How can we improve the crop production?



Food for thought

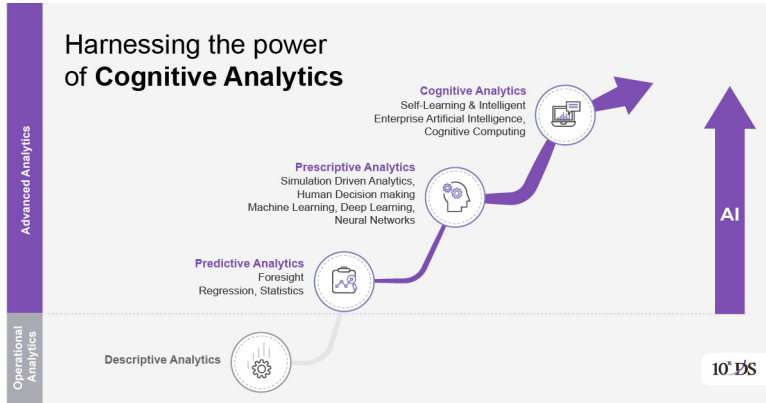
Erratic monsoon likely to take a toll on kharif crop.



Note: Estimate for Kharif of 2019 are Crisil Research estimates, 2019-20 rabi production is government target

Source: Ministry of agriculture

Cognitive Analytics – What Don't I Know?



- Next level of Analytics
- 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."

- 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.
- According to Jean Francois Puget:
"It extends the analytics journey to areas that were unreachable with more classical analytics techniques like business intelligence, statistics, and operations research."

<https://www.ecapitaladvisors.com/blog/analytics-maturity/>

<https://www.xenonstack.com/insights/what-is-cognitive-analytics/>

Types of analytics according to the domain

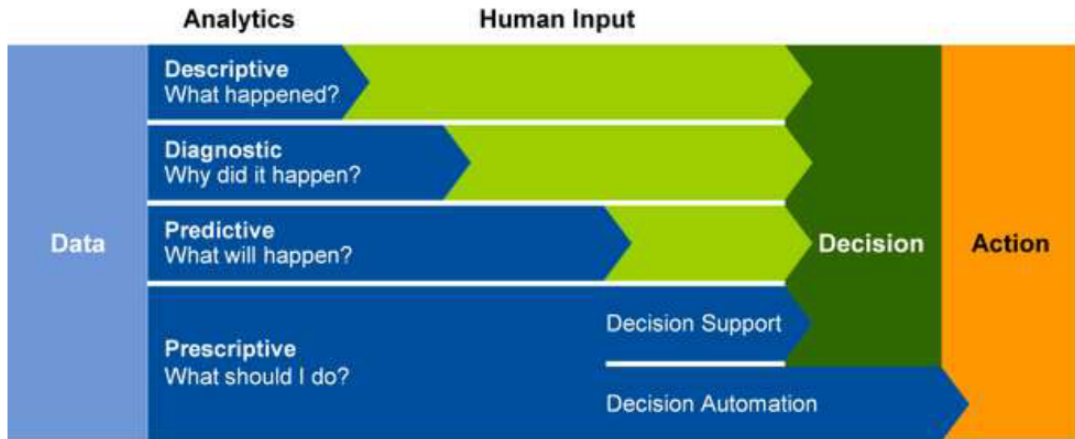
- ① Marketing Analytics
- ② Financial Analytics
- ③ Healthcare Analytics
- ④ Sports Analytics
- ⑤ HR Analytics
- ⑥ Customer Analytics
- ⑦ Web Analytics
- ⑧ Social Analytics
- ⑨ Cultural Analytics
- ⑩ Political Analytics



Types of analytics according to the type of data

- ① Text analytics
- ② Real-time data analytics
- ③ Multimedia analytics
- ④ Geo analytics
- ⑤ Mobile analytics

DATA ANALYTICS & LEVEL OF AUTOMATION



DATA ANALYTICS & COMPUTATIONAL TASKS

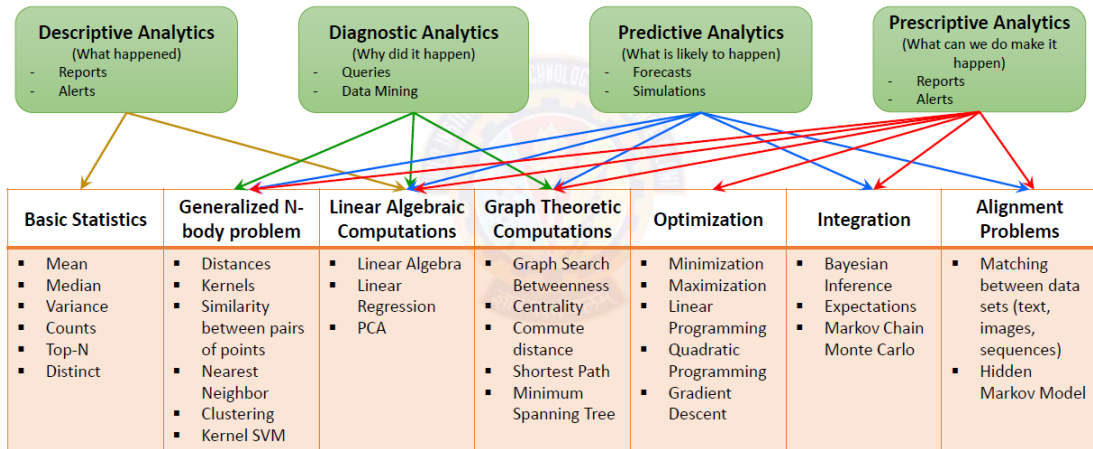


TABLE OF CONTENTS



- 1 ANALYTICS
- 2 BIG DATA
- 3 DATA ANALYTICS
- 4 CASE STUDIES ON DATA ANALYTICS

DESCRIPTIVE ANALYTICS – EXAMPLE #1



Problem Statement :

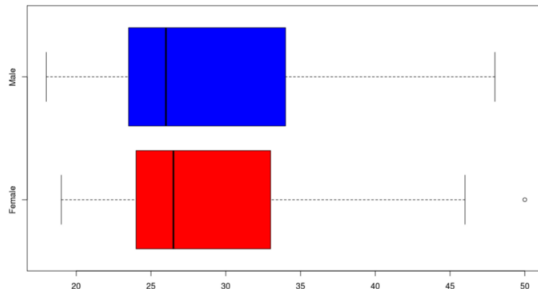
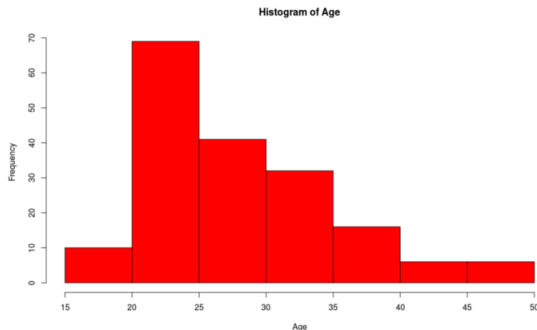
“Market research team at Aqua Analytics Pvt. Ltd is assigned a task to identify profile of a typical customer for a Digital fitness band that is offered by Titanic Corp. The market research team decides to investigate whether there are differences across the usage patterns and product lines with respect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

Data captured

- 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

<https://medium.com/@ashishpahwa7/first-case-study-in-descriptive-analytics-a744140c39a4>

DESCRIPTIVE ANALYTICS – EXAMPLE #1



	Usage	2		3		4		5		6		7	Totals
	Gender	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Male	
Product													
DX87					1	2	10	3	9	2	5	2	40
IQ75		13	6	19	18	7	15	1	1				60
MZ65		7	7	14	12	5	7	3					60
Totals		20	13	33	36	14	38	7	10	2	5	2	180

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.

DIAGNOSTIC ANALYTICS – EXAMPLE #1

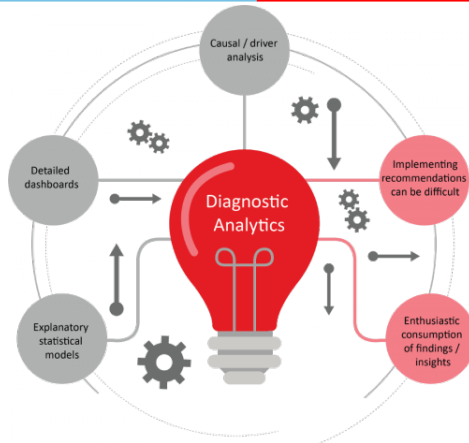


Problem Statement :

“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?”

<https://medium.com/parrotai/understand-data-analytics-framework-with-a-case-study-in-the-business-world-15bfb421028d>

DIAGNOSTIC ANALYTICS – EXAMPLE #1



<https://www.sganalytics.com/blog/change-management-analytics-adoption/>

PREDICTIVE ANALYTICS – EXAMPLE #1



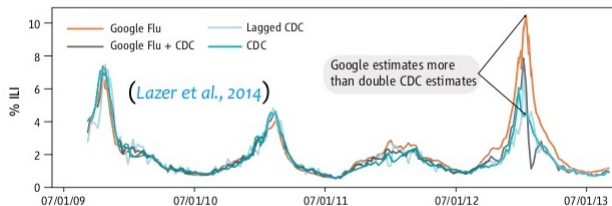
- Google launched **Google Flu Trends (GFT)**, to collect predictive analytics regarding the outbreaks of flu. It's a great example of seeing big data analytics in action.
- So, did Google manage to predict influenza activity in real-time by aggregating search engine queries with this big data and adopting 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.

PREDICTIVE ANALYTICS – EXAMPLE #1



Google Flu Trends: Failure

$$\text{logit}(P) = \beta_0 + \beta_1 \times \text{logit}(Q) + \epsilon \quad (\text{Ginsberg et al., 2009})$$



The estimates of the online Google Flu Trends tool were approx. **two times larger** than the ones from the CDC in 2012/13

<https://www.slideshare.net/VasileiosLamos/usergenerated-content-collective-and-personalised-inference-tasks>

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.
- 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.

<https://www.footsmart.com/pages/health-resource-center>

PREDICTIVE ANALYTICS – EXAMPLE #2



Predictive Policing (Self study)

- <https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained>
- <https://www.youtube.com/watch?v=YxvyeaL7NEM>

PRESCRIPTIVE ANALYTICS – EXAMPLE #1



- 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.

PRESCRIPTIVE ANALYTICS – 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.

<https://accent-technologies.com/2020/06/18/examples-of-prescriptive-analytics/>

HEALTHCARE ANALYTICS – CASE STUDY



Self study

- <https://integratedmp.com/4-key-healthcare-analytics-sources-is-your-practice-using-them/>
- <https://www.youtube.com/watch?v=olpuyn6kemg>

References:

- Big Data Analytics – A Hands-on Approach by Arshdeep Bahga & Vijay Madisetti

THANK YOU