PUG CHALLENGE EXCHANGE AMERICAS

UMASS LOWELL Learning with Purpose

Insight is 20/20: The Importance of Analytics

June 6, 2017

Amit Deokar

Department of Operations and Information Systems

Manning School of Business

University of Massachusetts Lowell

Email: Amit_Deokar@uml.edu

Agenda

- Analytics / Data Science: What it is and Why Do We Care?
- Analytics and Business Strategy
- Analytics: A Closer Look
- Embedded to Predictive to Cognitive Analytics

Agenda

- Analytics / Data Science: What it is and Why Do We Care?
- Analytics and Business Strategy
- Analytics: A Closer Look
- Embedded to Predictive to Cognitive Analytics

Analytics is the scientific process of transforming data into insight for making better decisions.

Reference: INFORMS

Analytics: Why Now?

- **Ubiquity of data opportunities**
  - E.g., operations, manufacturing, supply chain management, customer behavior, marketing campaign performance, workflow procedures, …
- **Computing technology advances**
  - E.g., distributed/cloud computing, mobile computing, IOT/sensors, …

Reference: Provost & Fawcett (2013) Data Science for Business

The New York Times

BUSINESS DAY

What Wal-Mart Knows About Customers' Habits

By CONSTANCE L. HAYS NOV. 14, 2004

Correction Appended

HURRICANE FRANCES was on its way, barreling across the Caribbean, threatening a direct hit on Florida's Atlantic coast. Residents made for higher ground, but far away, in Bentonville, Ark., executives at Wal-Mart Stores decided that the situation offered a great opportunity for one of their newest data-driven weapons, something that the company calls predictive technology.

A week ahead of the storm's landfall, Linda M. Dillman, Wal-Mart's chief information officer, pressed her staff to come up with forecasts based on what had happened when Hurricane Charley struck several weeks earlier. Backed by the trillions of bytes' worth of shopper history that is stored in Wal-Mart's computer network, she felt that the company could "start predicting what's going to happen, instead of waiting for it to happen," as she put it.

Hurricane Frances

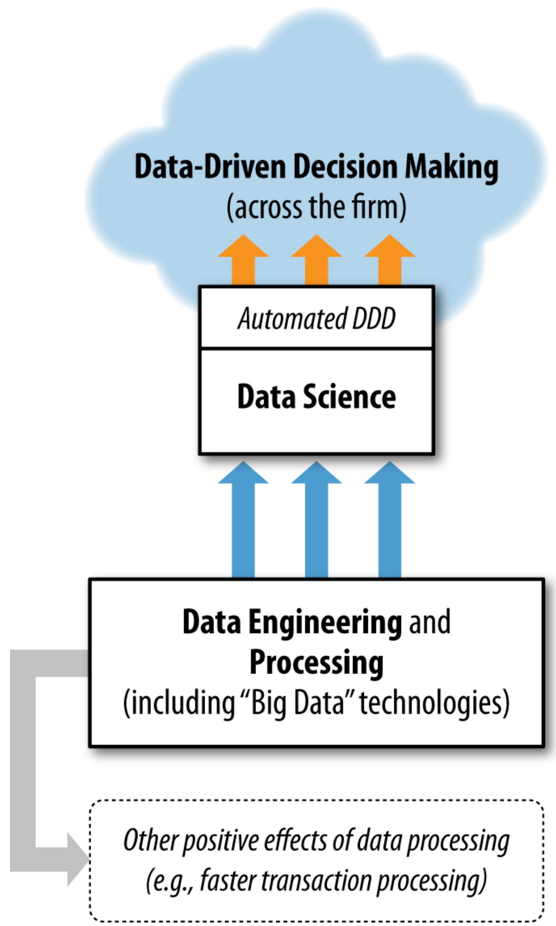
BUSINESS DAY | What Wal-Mart Knows About Customers' Habits

The experts mined the data and found that the stores would indeed need certain products -- and not just the usual flashlights. "We didn't know in the past that strawberry Pop-Tarts increase in sales, like seven times their normal sales rate, ahead of a hurricane," Ms. Dillman said in a recent interview. "And the pre-hurricane top-selling item was beer."

Thanks to those insights, trucks filled with toaster pastries and six-packs were soon speeding down Interstate 95 toward Wal-Marts in the path of Frances. Most of the products that were stocked for the storm sold quickly, the company said.

Reference: Provost & Fawcett (2013) Data Science for Business

Data Science (Analytics), Engineering & Data-Driven Decision Making



Practice of basing decisions on the analysis of data, rather than purely on intuition.

Reference: Provost & Fawcett (2013) Data Science for Business

Agenda

- Analytics / Data Science: What it is and Why Do We Care?
- Analytics and Business Strategy
- Analytics: A Closer Look
- Embedded to Predictive to Cognitive Analytics

Business & IT Today

Tight integration of modern business analytics efforts with business strategies

&

Proactive rather than reactive approach

Distinction using analytics

Modern Analytics Principles – A Strategy Perspective

- **Deliver business value and impact**
  - Building and continuous evolving analytics for high-value business impact
- **Focus on the last mile**
  - Deploying analytics into production to attain repeatable, ongoing business value
- **Leverage Kaizen**
  - Starting small and building on success
- **Accelerate learning and execution**
  - Doing, learning, adapting, and repeating
- **Differentiate your analytics**
  - Exploiting analytics to produce new results
- **Embed analytics**
  - Building analytics into business processes to gain repeatability and scalability
- **Establish modern analytics architecture**
  - Leveraging commodity hardware and next generation technology to drive out costs
- **Build on human factors**
  - Maximizing and grooming talent
- **Capitalize on consumerization**
  - Leveraging choices to innovate
    - BYOD (combine open/noncompetitive data with your data to discover patterns), BYOT (mix and match open source and proprietary tools), BYOM (leverage app stores and crowdsourcing)

Reference: Chambers and Dinsmore (2015) Advanced Analytics Methodologies

How to Distinguish Your Analytics?

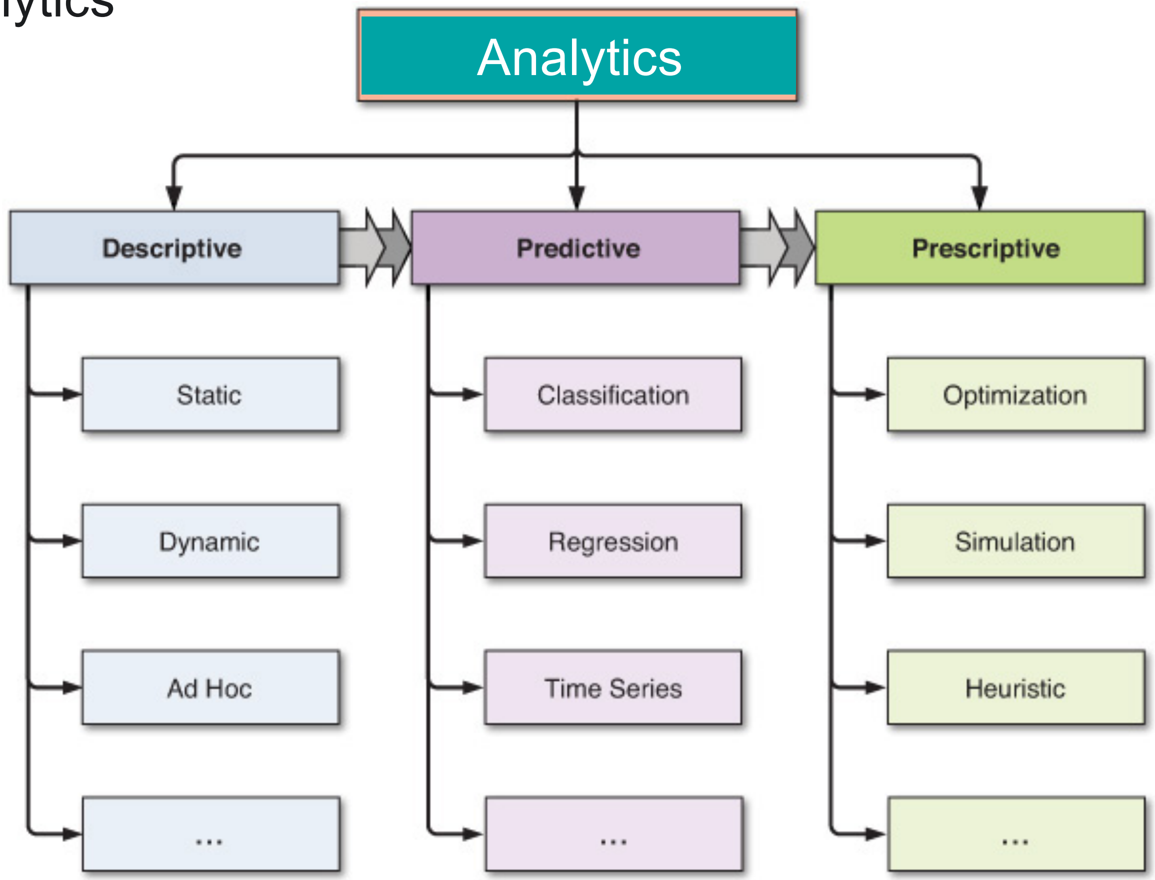
- **Business area**
  - Applying analytics to a new business area or problem
- **Data**
  - Leveraging or inventing new data to enrich analytical insights
- **Approach**
  - Employing a combination of analytical approaches in an innovative way to discover new patterns and value
- **Precision**
  - Increasing granularity of analytics by focusing on individuals (people, transaction, resources, etc.) rather than segments or groups
- **Algorithms**
  - Developing or using new groundbreaking mathematical of scientific approaches to gain advantage
- **Embedding**
  - Systematically inserting analytics into operational processes to gain deeper insights
- **Speed**
  - Accelerating the pace of business to stay ahead of the competition

Reference: Chambers and Dinsmore (2015) Advanced Analytics Methodologies

Agenda

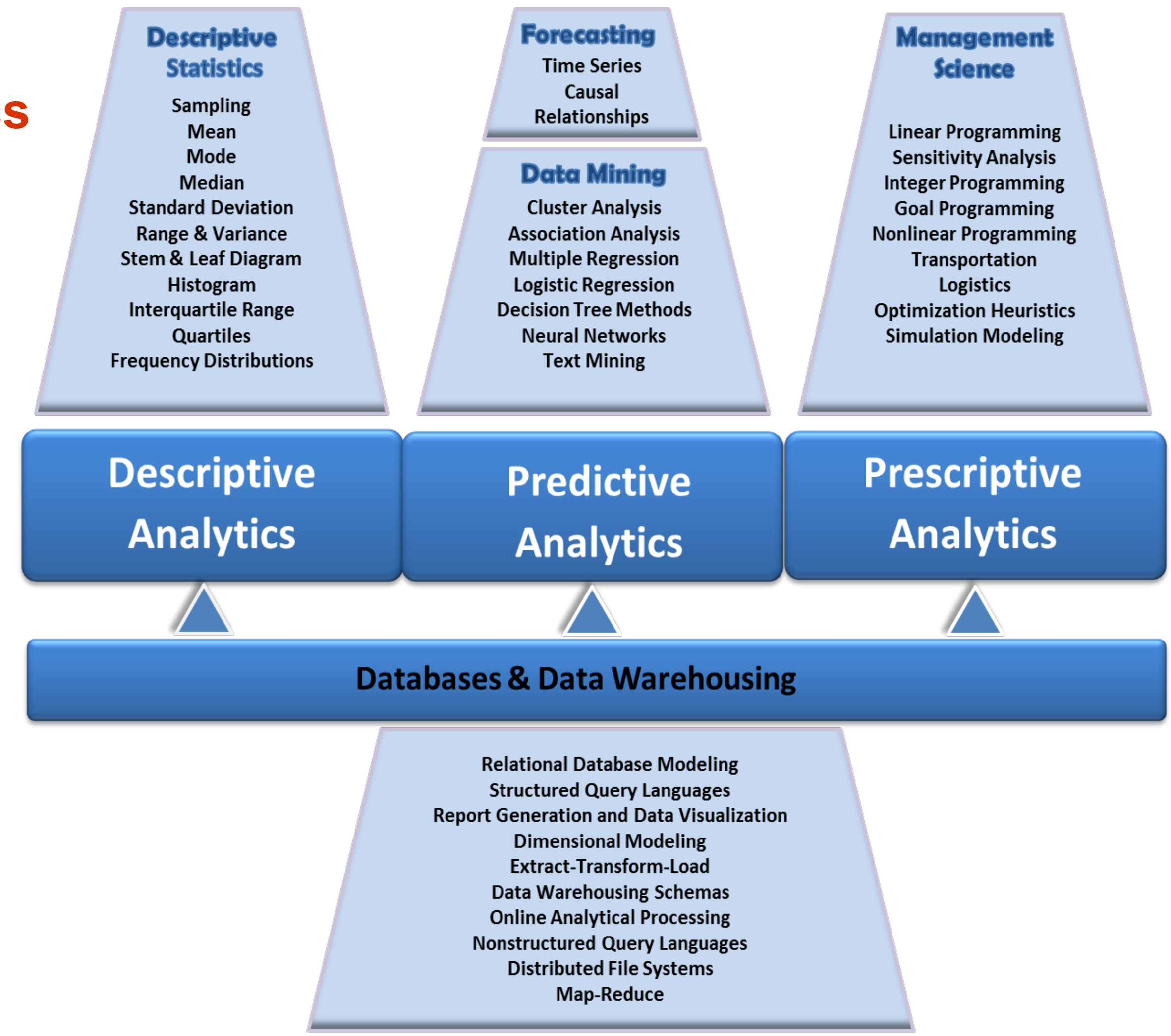
- Analytics / Data Science: What it is and Why Do We Care?
- Analytics and Business Strategy
- Analytics: A Closer Look
- Embedded to Predictive to Cognitive Analytics

Types of Analytics



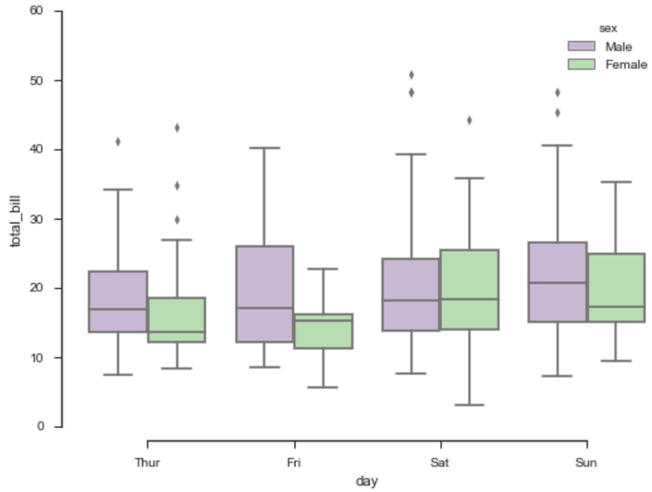
Reference: Delen (2015) Real-World Data Mining

Techniques Used in Different Types of Analytics

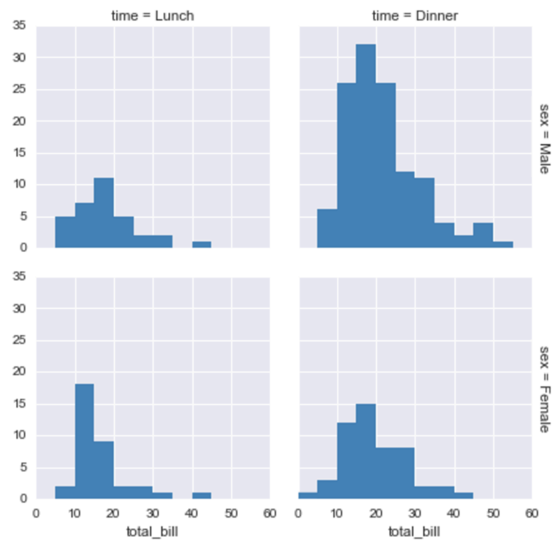


Reference: Asllani (2015) Business Analytics with Management Science Models and Methods

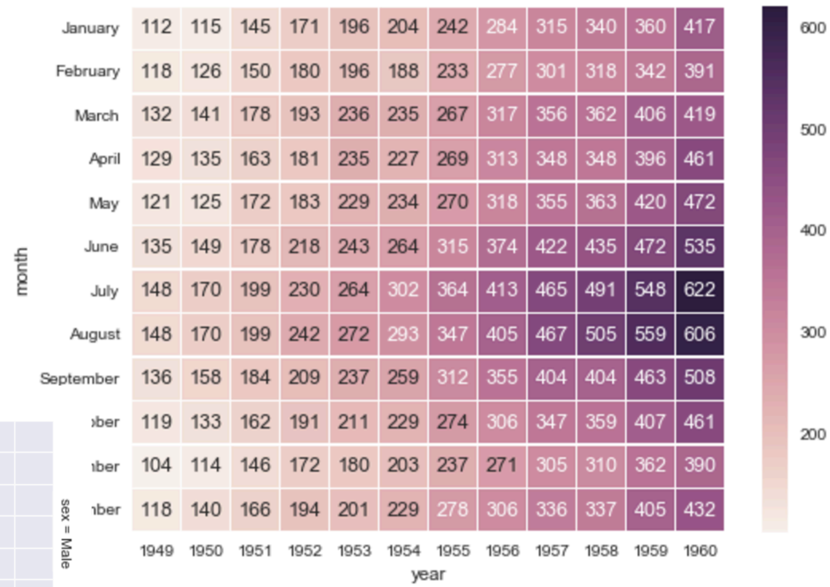
Descriptive Analytics



Boxplot



Histograms



Heatmap

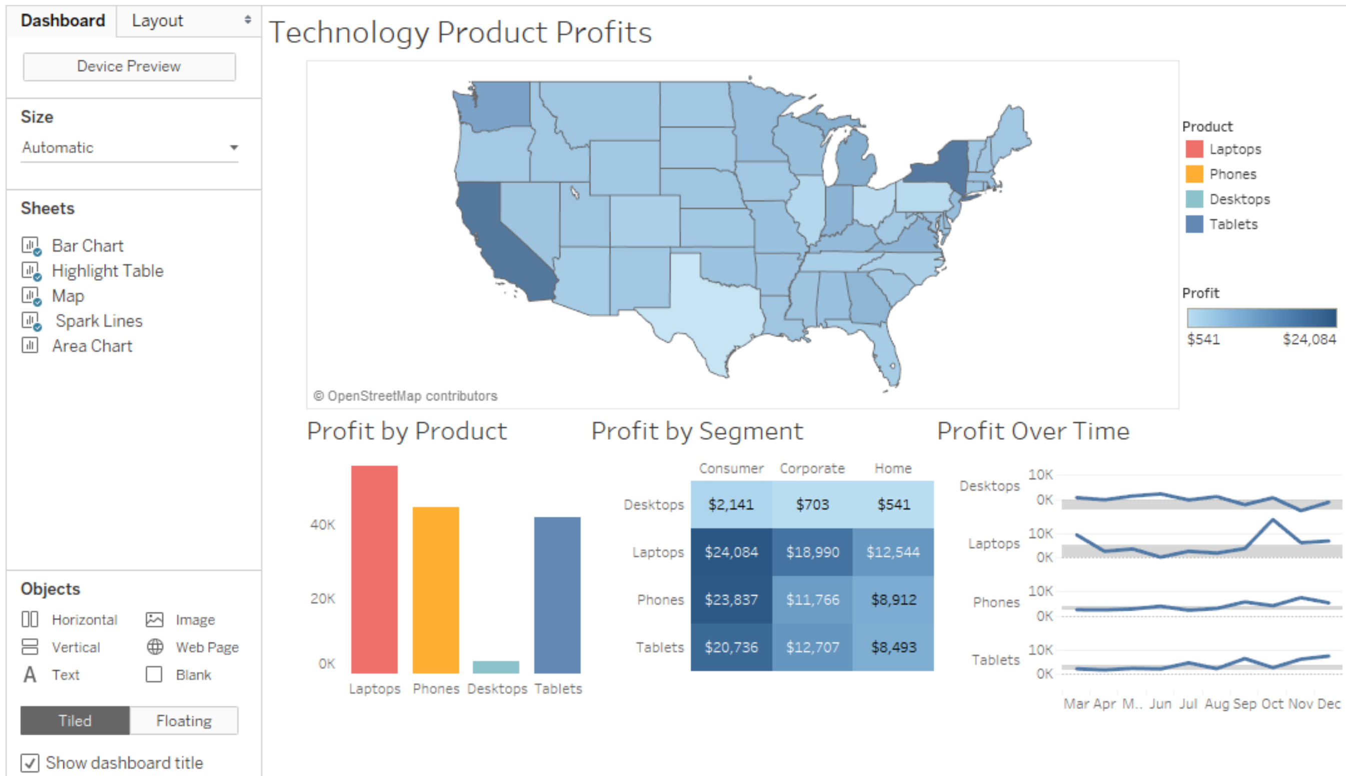
Embedding
Descriptive Analytics

- Dashboards & KPIs
- Example: **Customer Relationship Management**

Reference: SalesForce.com

Embedding
Descriptive Analytics

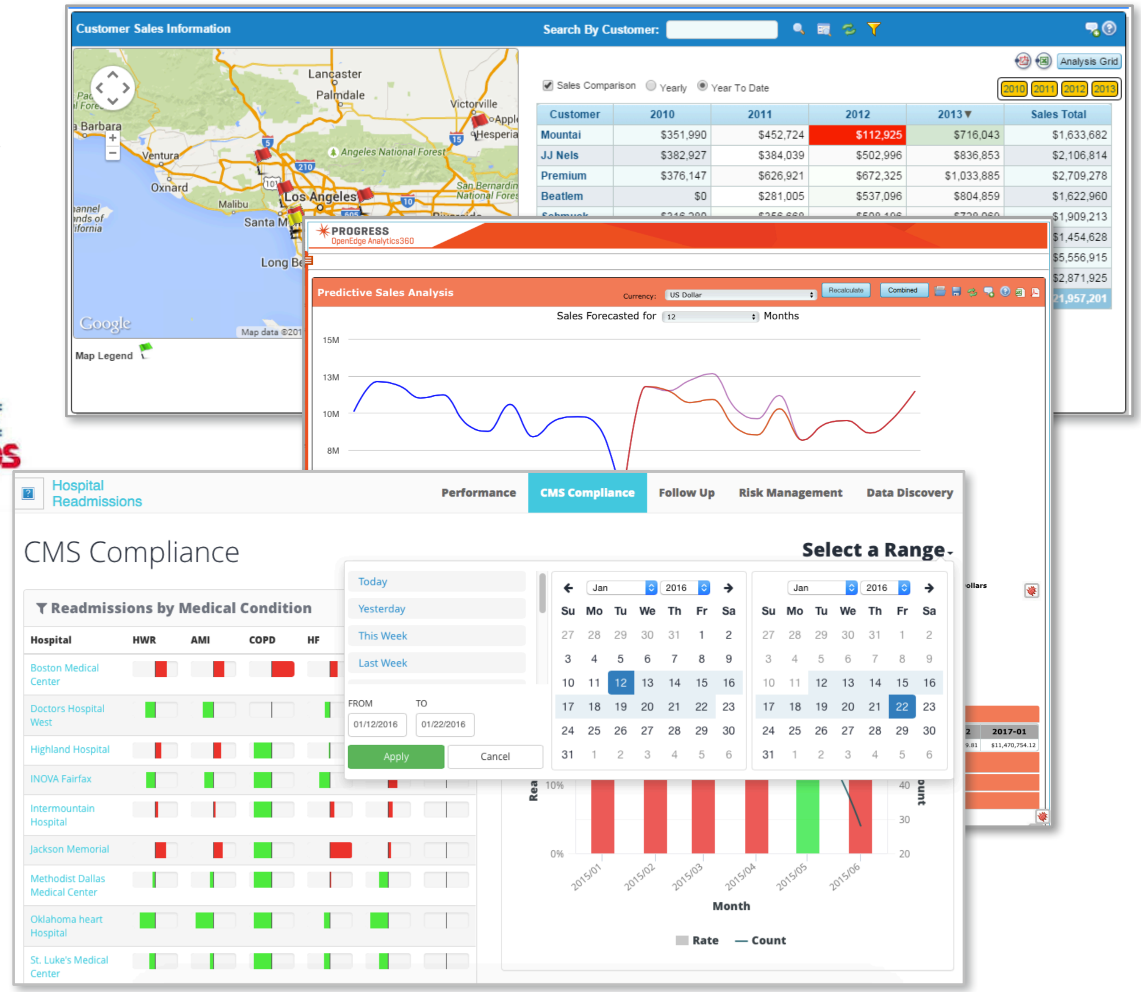
- Descriptive Analytics
- Dashboards & KPIs
- Example: **Product Management**



Reference: Tableau.com

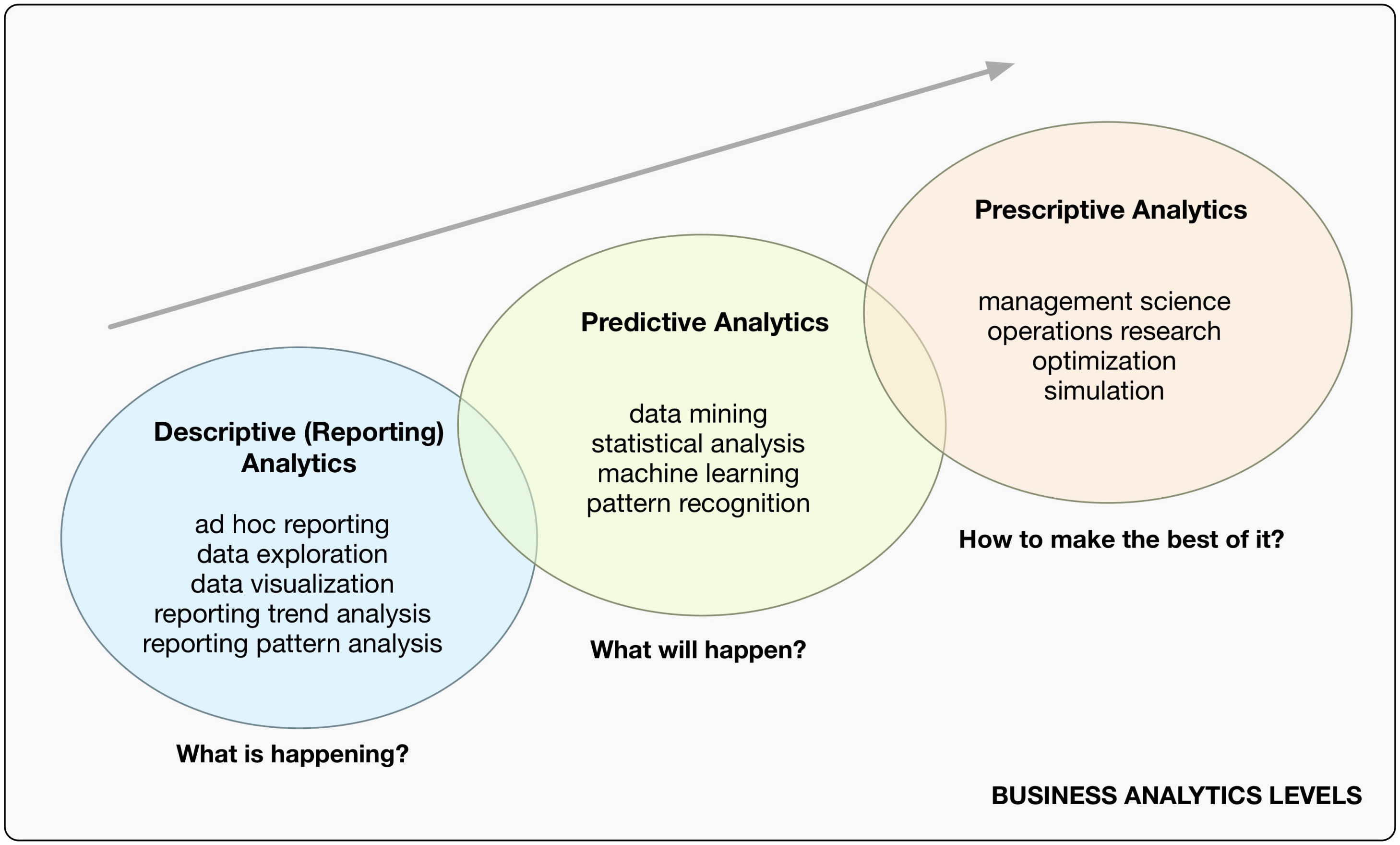
Descriptive Analytics: Vendors / Tools

- Off the Shelf Software Tools
  - Progress OpenEdge 306 Progress
  - Tableau +ableau
  - Qlik Qlik
  - SAS Visual Analytics sas
  - IBM Watson Analytics, Cognos IBM COGNOS
  - Tibco Spotfire TIBCO Spotfire
  - …



- Open Source Platforms / Languages
  - Python
  - R
  - …

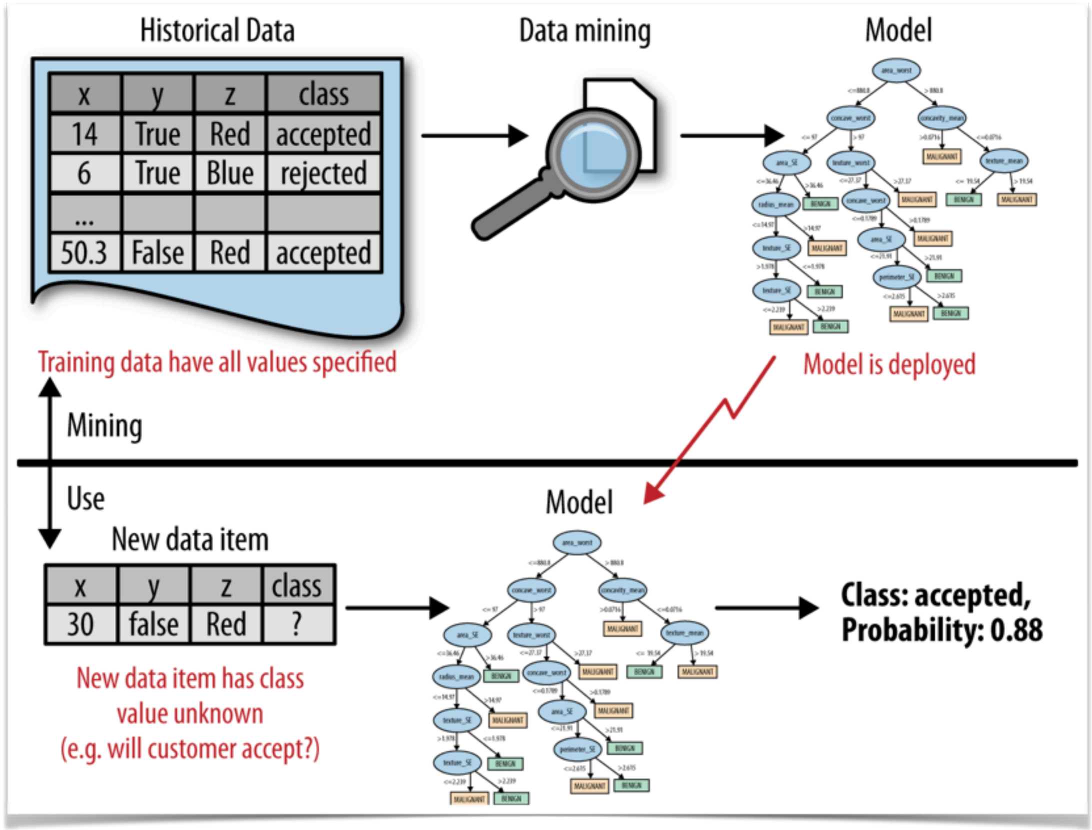
Types of Analytics: A Maturity Model Perspective



<table>
<tr><th>Competitive Advantage & Degree of Intelligence</th><th>Data-Driven Questions</th><th>Tools / Techniques</th><th>Source: INFORMS and IBM</th><th colspan="2">Source: SAS</th></tr>
<tr><td rowspan="8"></td><td>How to do thing’s better? What’s the best that can happen?</td><td>Optimization, Simulation</td><td>Prescriptive Analytics</td><td rowspan="4">Analytics</td><td rowspan="8">Business Intelligence Technologies</td></tr>
<tr><td>What will happen next?</td><td>Predictive modeling</td><td rowspan="3">Predictive Analytics (Data Mining)</td></tr>
<tr><td>What if these trends continue?</td><td>Forecasting / Extrapolation</td></tr>
<tr><td>Why is this happening? What</td><td>Statistical analysis</td></tr>
<tr><td>What actions are needed? When should I react?</td><td>Alerts</td><td rowspan="4">Descriptive / Reporting Analytics (Business Intelligence)</td><td rowspan="4">Data Access and Reporting</td></tr>
<tr><td>Where exactly is the problem?</td><td>Query/drill down</td></tr>
<tr><td>How many, how often, where?</td><td>Ad hoc reports</td></tr>
<tr><td>What happened? When did it happen?</td><td>Standard reports</td></tr>
</table>

Predictive Analytics: Data Mining

- Data mining is a **process** with well-understood stages based on:
  - application of information technology
  - analyst's creativity
  - business knowledge
  - common sense
- **Decompose** a data analytics problem into pieces such that you can solve a known task with a tool
- There is a **large number** of **data mining algorithms** available, but only a **limited number** of **data mining tasks**



Reference: Provost & Fawcett (2013) Data Science for Business

Predictive Analytics: Key Data Mining Tasks

- **Supervised Learning** – focus is on predicting a specific “target”
  - Classification
  - Regression
- **Unsupervised Learning** – focus is on discovering patterns
  - Clustering
  - Co-occurence grouping
  - …
  - Profiling
  - Link prediction
  - Data reduction
  - …

Predictive Analytics: Key Data Mining Tasks

Supervised Learning

- **Classification**
- Regression

Unsupervised Learning

- Clustering
- Co-occurence grouping

Classification attempts to ***predict***, for each individual in a population, which class this individual belongs to.

"Among all the customers of *RetailCo*, which are likely to respond to a given offer?"

will respond *will not respond*

"Among all the parts of the machine, which is likely to fail within the next X days?"

will fail *will not fail*

Classification algorithms provide models that determine which class a new individual belongs to.

Predictive Analytics: Key Data Mining Tasks

Supervised Learning

- Classification
- **Regression**

Unsupervised Learning

- Clustering
- Co-occurence grouping

Regression (value estimation) attempts to estimate or predict, for each individual, the numerical value of some variable for that individual

“How much will a given customer use the service?”

Predicted variable: service usage

“How much useful life of machine is remaining?

Predicted variable: useful life

Regression models are generated by algorithms that analyze other, similar individuals in the population and their historical service usage / useful life

Regression procedures produce a model that, given an individual, estimates the value of the particular variable specific to that individual

Predictive Analytics: Key Data Mining Tasks

Supervised Learning

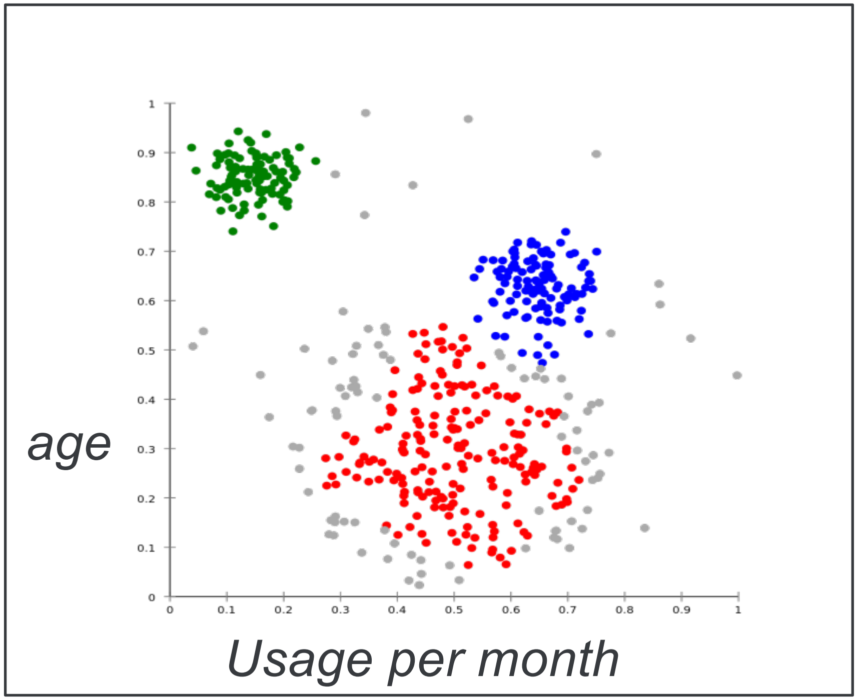
- Classification
- Regression

Unsupervised Learning

- **Clustering**
- Co-occurence grouping

Clustering

- Do my customers form natural groups?



Predictive Analytics: Key Data Mining Tasks

Supervised Learning

- Classification
- Regression

Unsupervised Learning

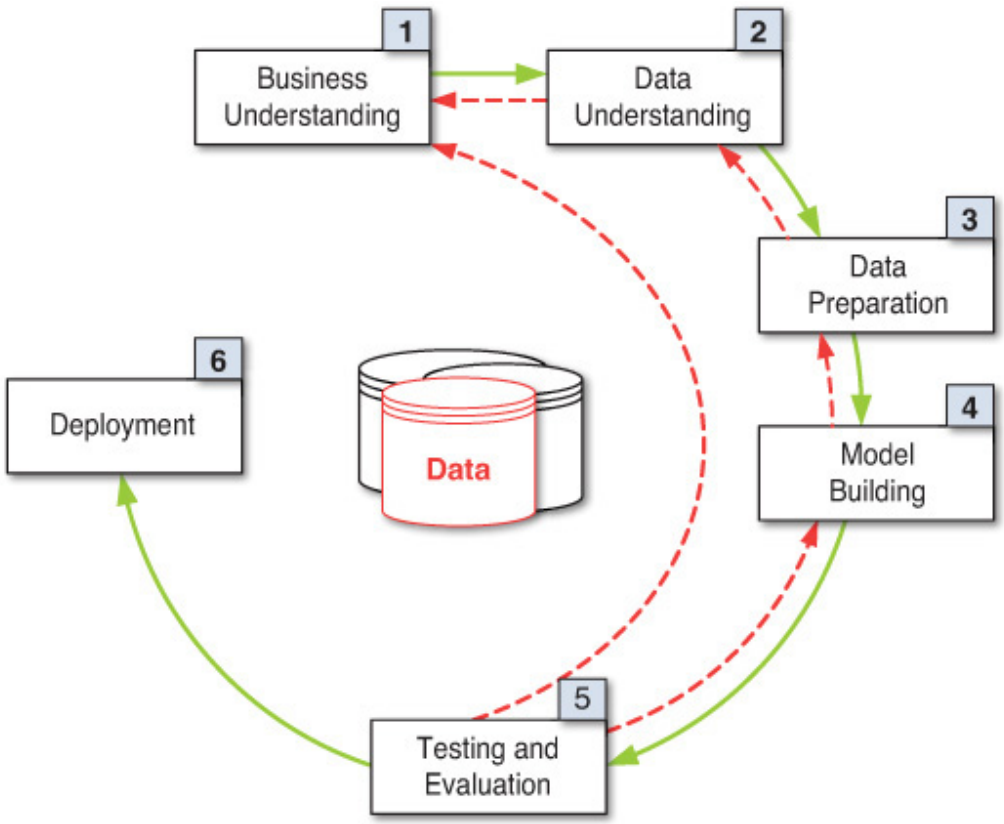
- Clustering
- **Co-occurence grouping**

Co-occurrence grouping / Association Rule Mining / Market Basket Analysis

- Find associations between entities based on the transactions they are involved in.

Predictive Analytics: Case of Predictive Maintenance of Assets

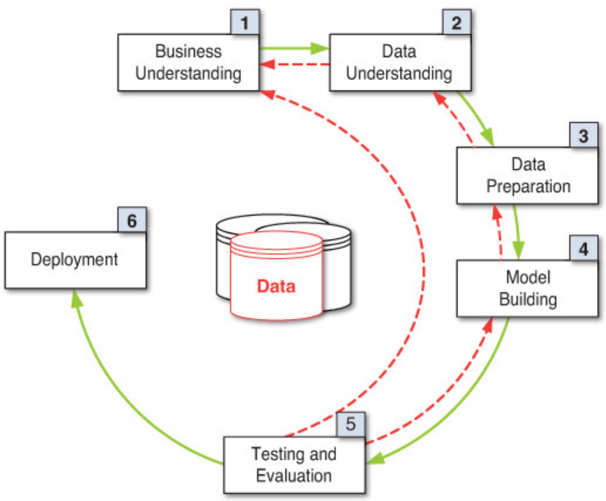
- **Regression:** Predict the Remaining Useful Life (RUL), or Time to Failure (TTF).
- **Binary classification:** Predict if an asset will fail within certain time frame (e.g. days).
- **Multi-class classification:** Predict if an asset will fail in different time windows: E.g., fails in window [1, *w0*] days; fails in the window [*w0*+1,*w1*] days; not fail within *w1* days



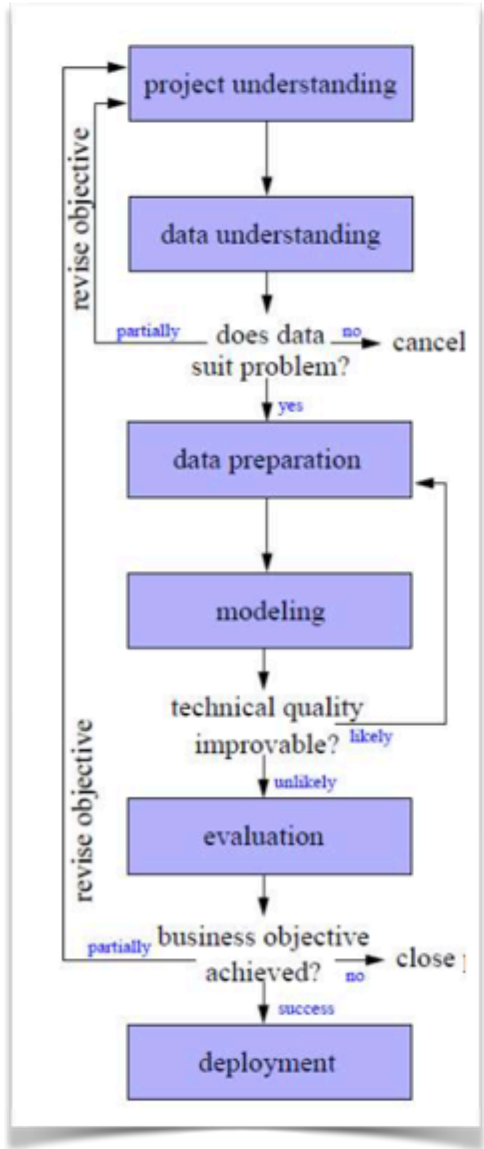
The CRISP-DM Data Mining Process

CRISP-DM

- Iteration as a rule
- Process of data exploration



The CRISP-DM Data Mining Process



What exactly is the problem, the expected benefit?
How would a solution look like?
What is known about the domain?

What data do we have available?
Is the data relevant to the problem?
Is it valid? Does it reflect our expectations?
Is the data quality, quantity, recency sufficient?

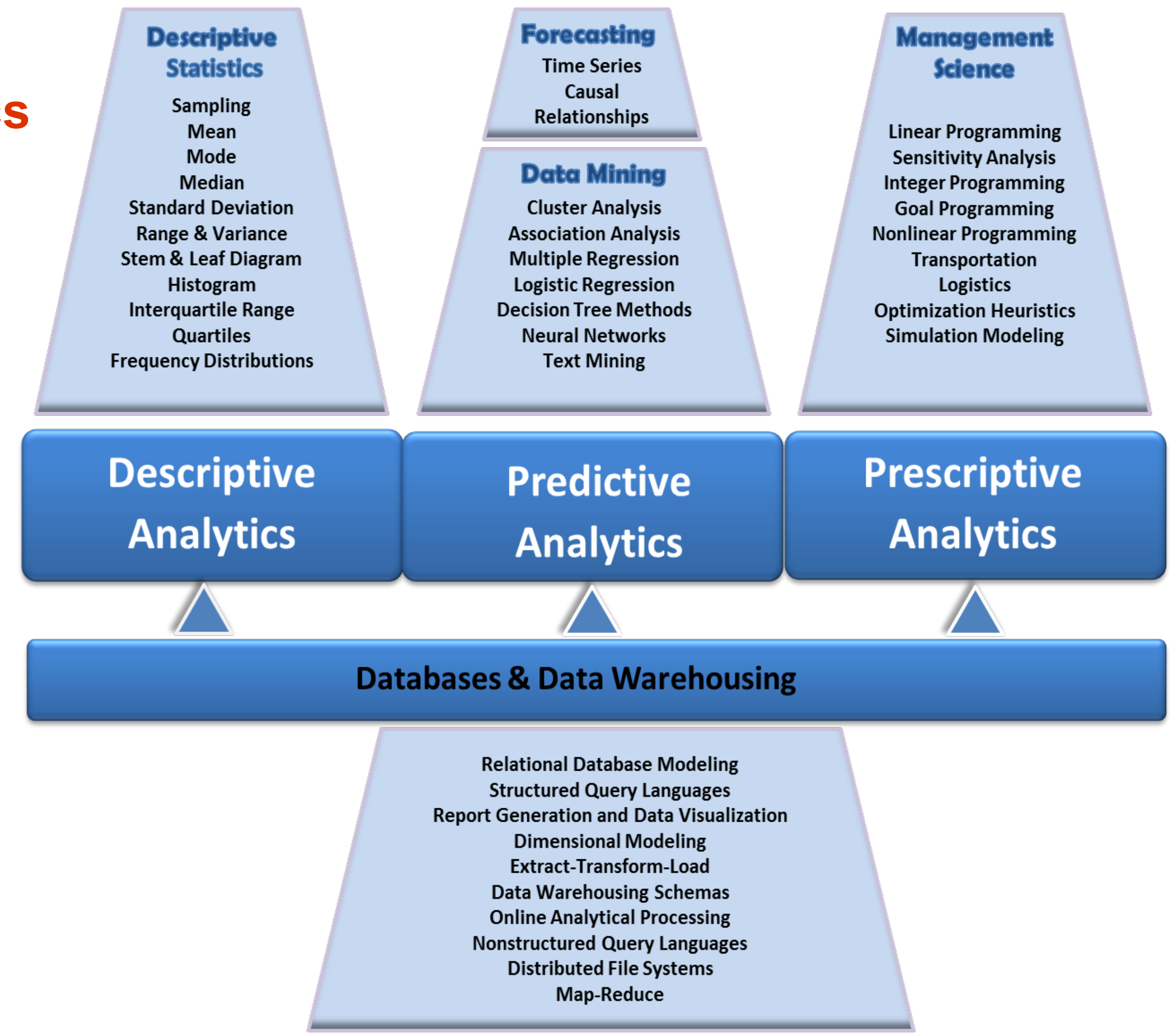
Which data should we concentrate on?
How is the data best transformed for modeling?
How may we increase the data quality?

What kind of model architecture suits the problem best?
What is the best technique/method to get the model?
How good does the model perform technically?

How good is the model in terms of project requirements?
What have we learned from the project?

How is the model best deployed?
How do we know that the model is still valid?

Techniques Used in Different Types of Analytics



Reference: Asllani (2015) Business Analytics with Management Science Models and Methods

Prescriptive Analytics: Case of Transportation Models in Logistics

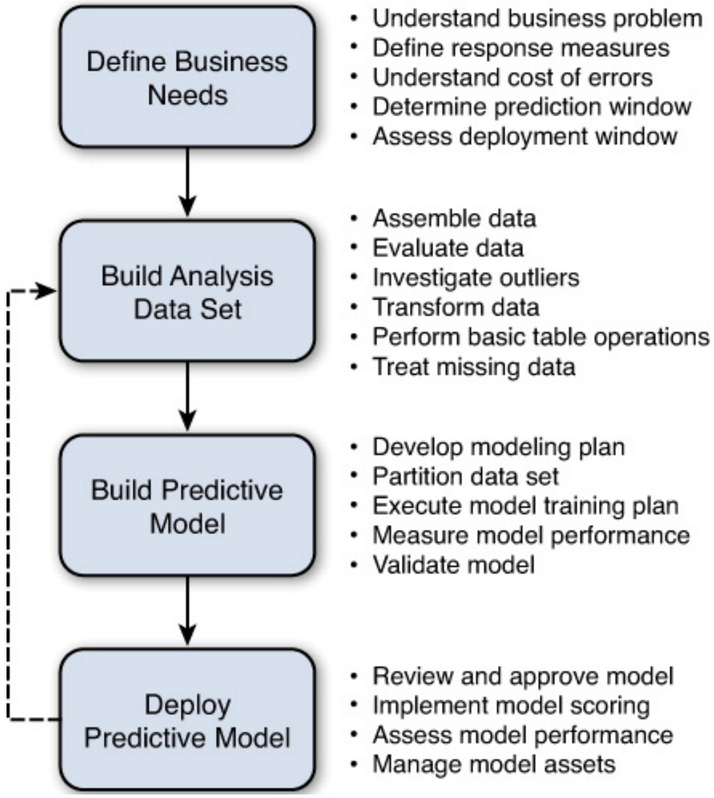
- Data Sources
  - Radio Frequency Identification (RFID) technology
  - Mobile devices
  - Sensors
  - External databases
- Data
  - Delivery times
  - Resource utilizations
  - Geographical coverages
  - Delivery statuses in real-time
- Optimization procedures (e.g., transhipment models) embedded to reroute vehicles on the go, and instant direction updates to drivers on their onboard navigation system to the next “best” destination.

Agenda

- Analytics / Data Science: What it is and Why Do We Care?
- Analytics and Business Strategy
- Analytics: A Closer Look
- Embedded to Predictive to Cognitive Analytics

Embedded Analytics Using Predictive & Prescriptive Analytics

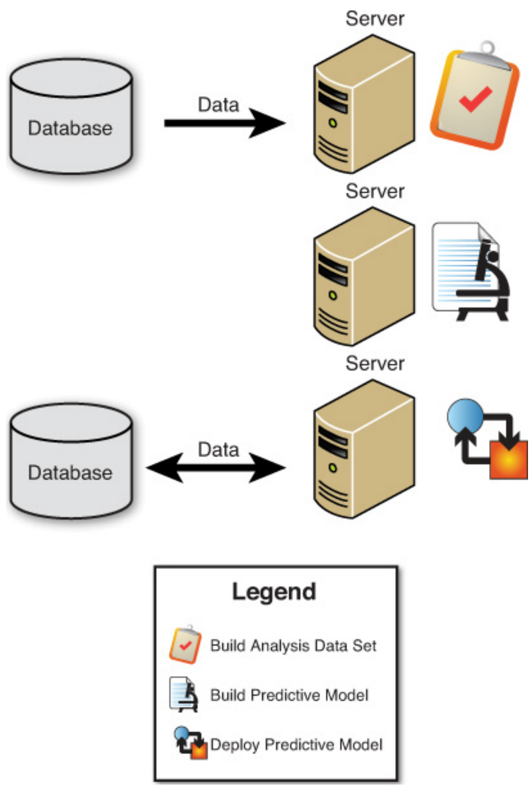
- Operationalizing / Deploying
  - E.g., deploy web services with model scoring
- Integrating Software Applications with Analytics
  - E.g.,
    - specific application functionality driven by analytics models
    - application meta-data incorporated in analytics models to predict application behavior
- Best Practices / Trends
  - Importance of deploying analytic models into environments that mimic real environment (not idealistic) environments
  - Deployment used to be "post-process"
  - Deployment and continuous improvement are now part of the full life-cycle analytic process



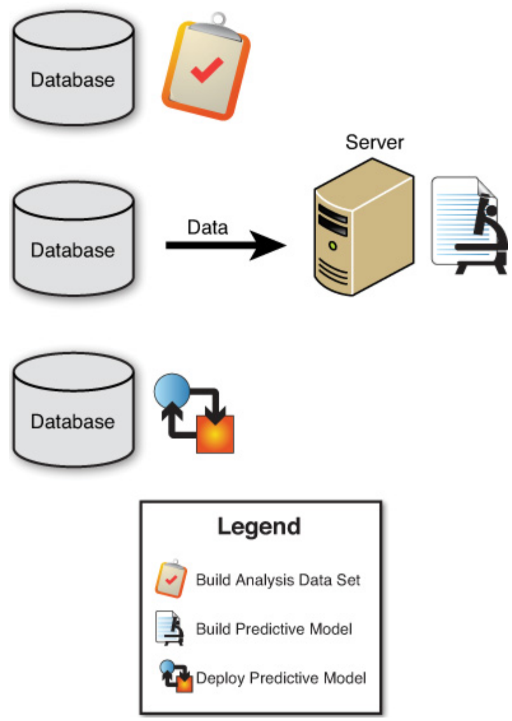
PUGCHALLENGE AMERICAS EXCHANGE

Reference: Chambers and Dinsmore (2015) Advanced Analytics Methodologies

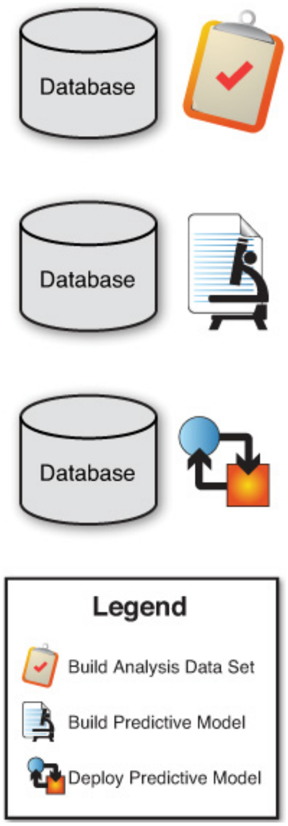
Freestanding Analytics



Partially-Integrated Analytics



In-Database Analytics

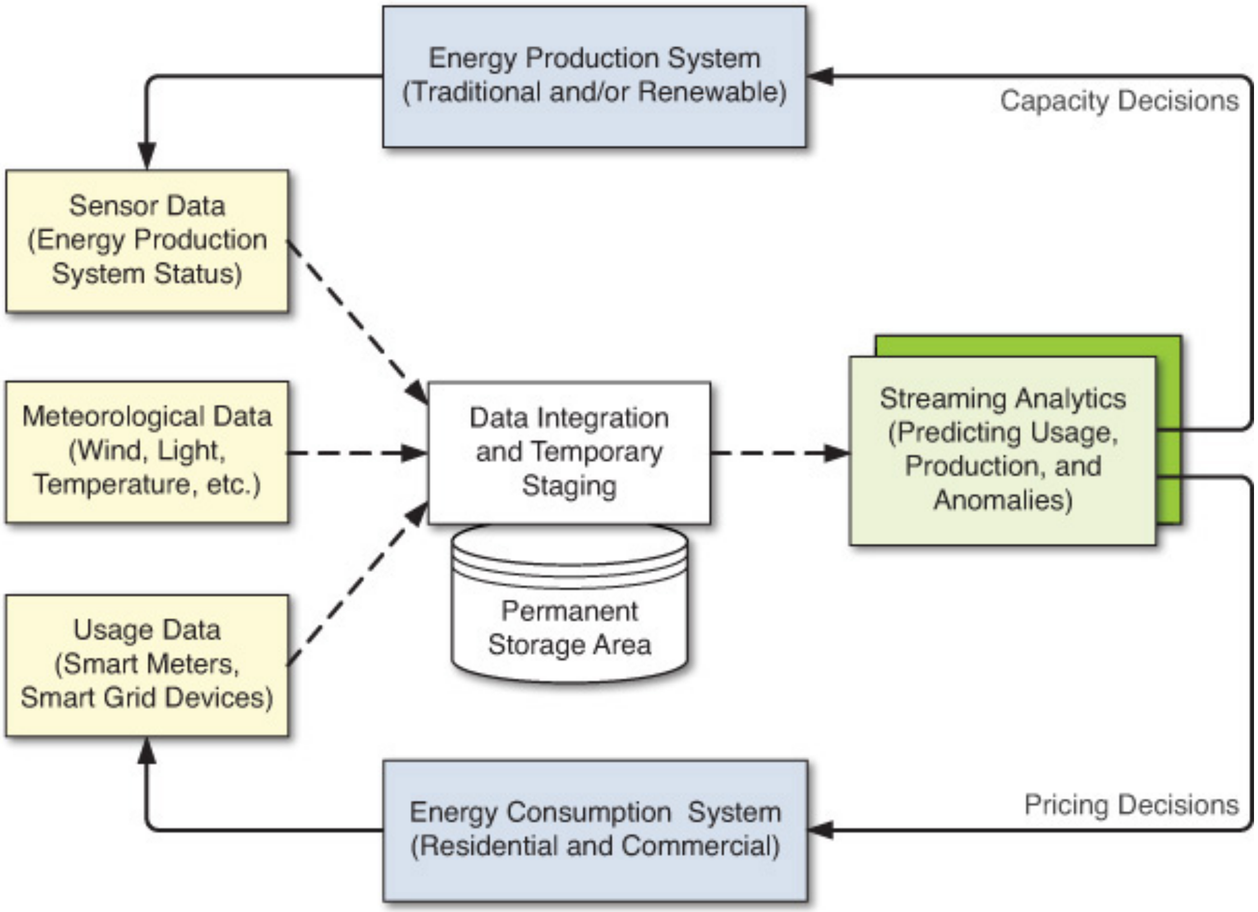


Reference: Chambers and Dinsmore (2015) Advanced Analytics Methodologies

Big Data & Cognitive Analytics

- Big Data
  - Volume
  - Variety
  - Velocity
  - Value !
- Examples of Big Data Sources
  - Web data (e.g., text web pages)
  - Social media data (e.g., blogs, tweets, pictures)
  - Event data (e.g., clickstreams, web logs)
  - Machine-generated data (e.g., sensors, RFID, IOT, IIOT)
- Cognitive Analytics
  - Scaling up predictive analytics to deal with big data
    - IBM's Watson – natural language processing
    - General Electrics' Predix for Industrial IOT
  - Example techniques
    - Deep learning – extensions of artificial neural networks
    - Text mining
    - Graph mining

Big Data Analytics: Case of Streaming Analytics in the Energy Industry

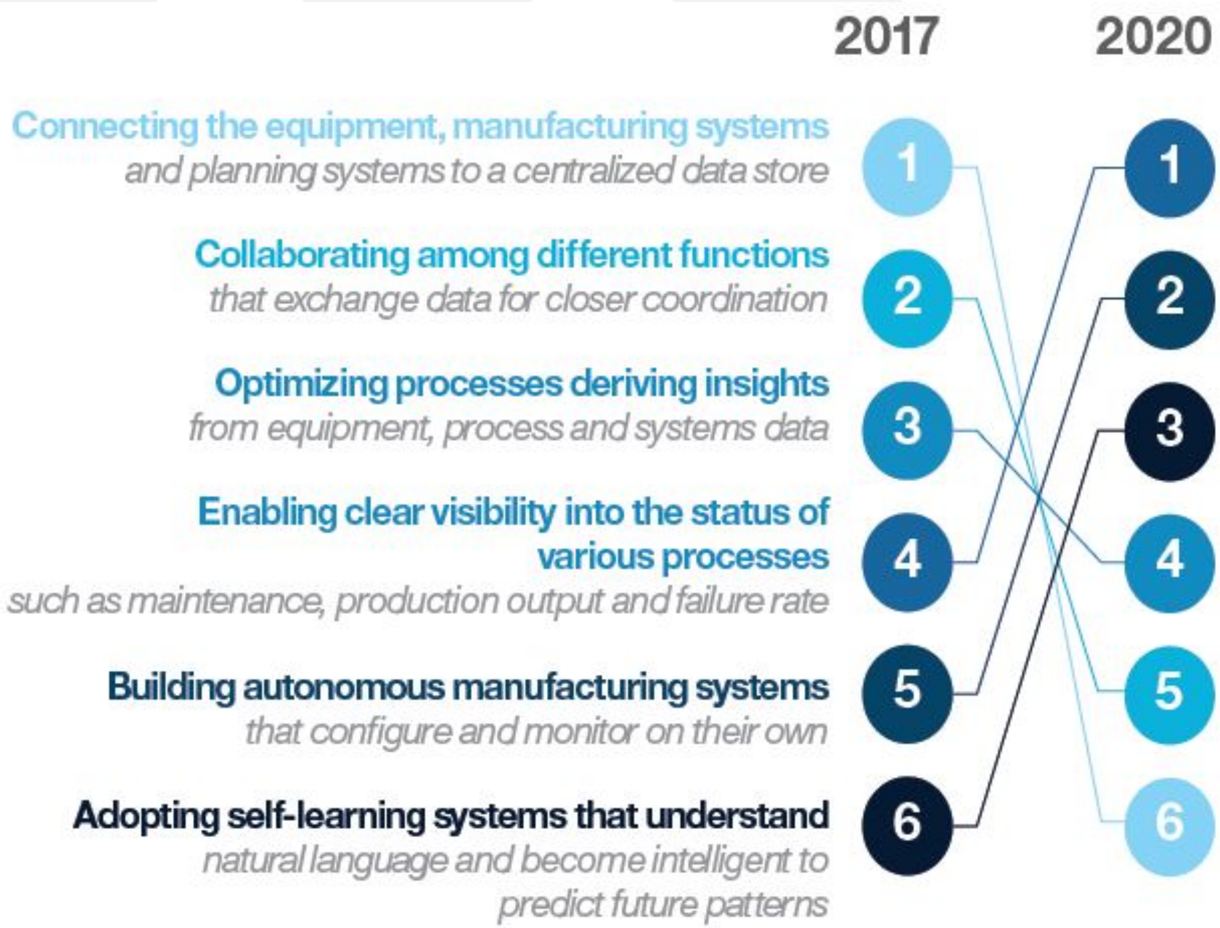


Reference: Delen (2015) Real-World Data Mining

Current Trends: Integrated Applications with Cognitive Analytics

- Traditional Application Development
  - Front-end tooling
  - Back-end services
  - Data connectivity
  - Business rules
- Integration with Cognitive Analytics
  - E.g., OpenEdge + DataRPM
  - Real-time data ingestion integrated in the application
  - Streaming analytics (high velocity)
  - Meta-learning – parameter tuning based on errors, i.e., learning on the go

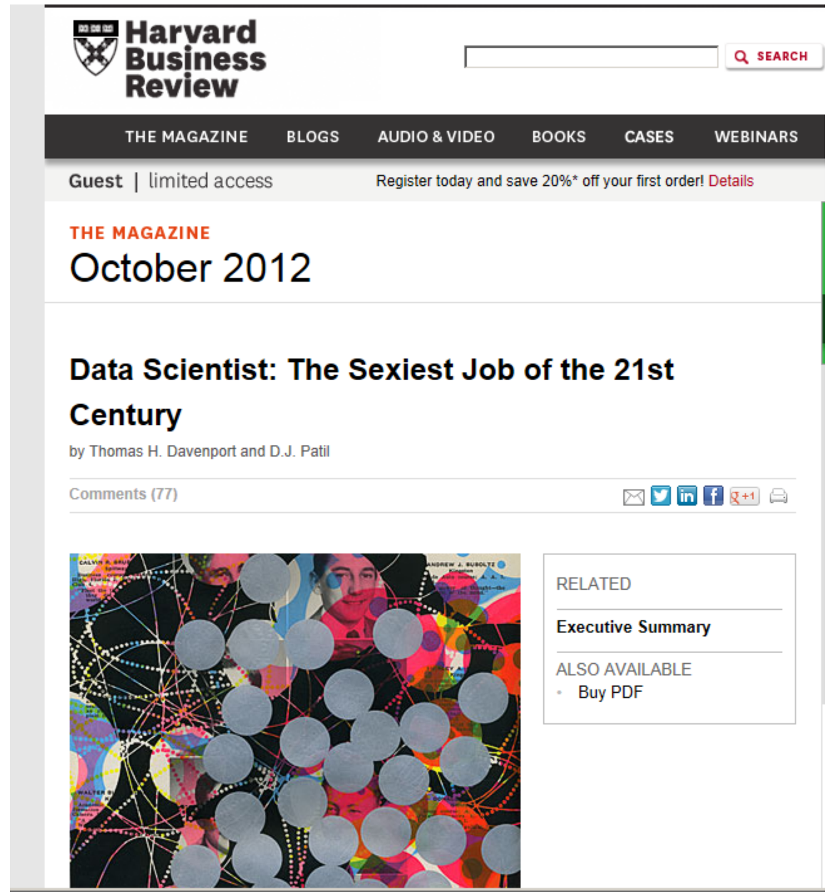
Current Trends: Cognitive Manufacturing

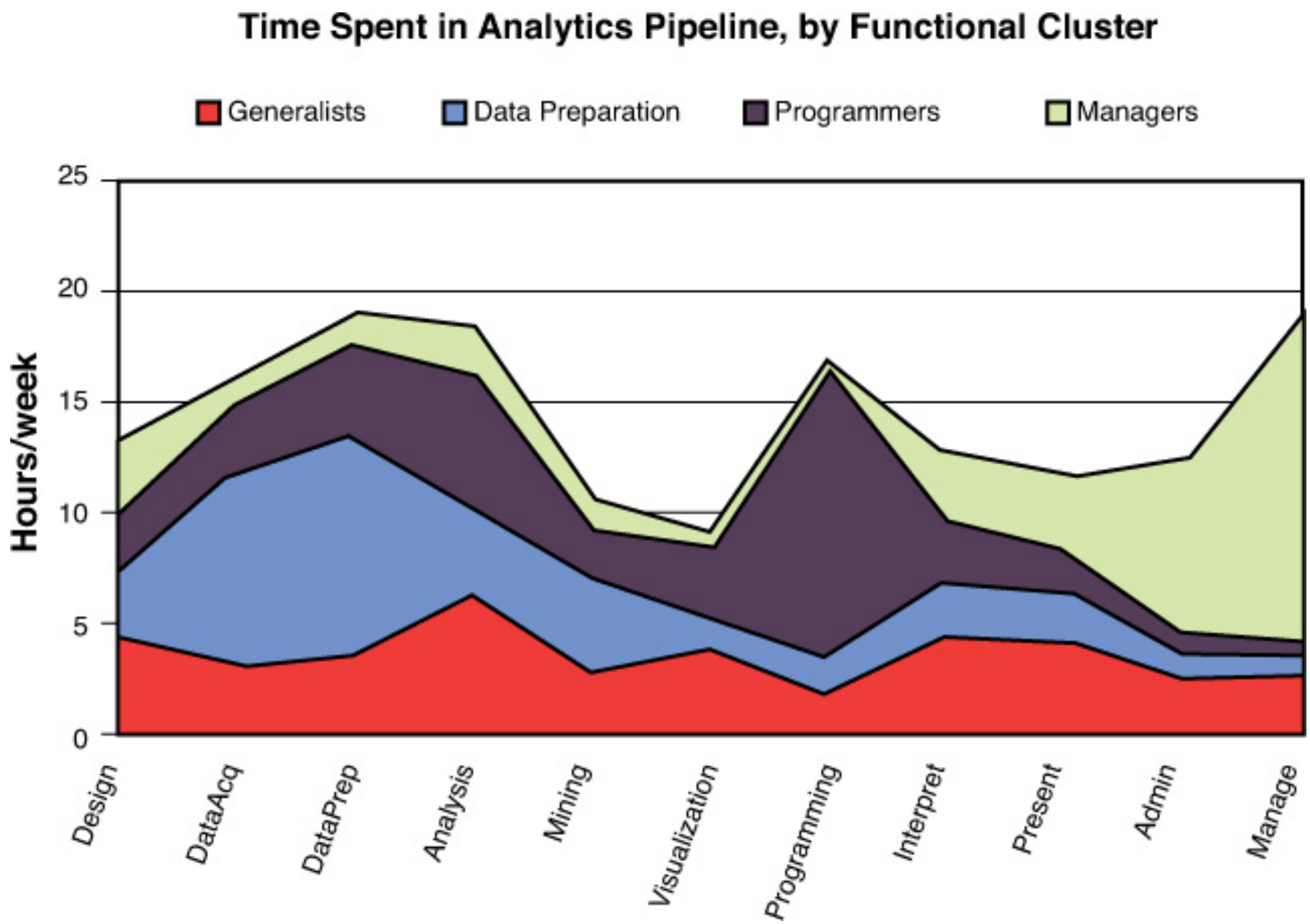


Reference: IBM Institute of Business Value

Roles in Analytics / Data Science

- **Data Scientist / Data Analyst**
  - Understanding the potential
  - Can translate from business to execution
  - Ability to evaluate proposal and execution
  - Can do the actual modeling
  - Applied statistician X computer scientist / info systems
  - Business knowledge X analytical tools knowledge
- **Collaborator in an analytics/data-science project**
- **Managing an analytics/data-science project**
- **Investing in an analytics/data-science project**
  - Understanding the potential
  - Can translate from business to execution
  - Ability to evaluate proposal and execution





Reference: Chambers and Dinsmore (2015) Advanced Analytics Methodologies

Summary

- Analytics has clearly gained traction in various industry sectors
- Transformation from embedded to predictive to cognitive analytics
- Key to align business strategy with analytics strategy
- Need for investment in data sources that can improve decision-making
- Integration of analytics in variety of applications

Thank You

Questions or Comments?