# FROM DATA TO DECISIONS: PREDICTIVE & PRESCRIPTIVE ANALYTICS IN THE INDUSTRY

BALAKRISHNAN ANANTHANARAYANAN

TECHNOLOGIST

INTEL TECHNOLOGY INDIA PVT. LTD.,

99MW04

# OBEISANCE

न गुरोरधिकं तत्त्वं न गुरोरधिकं तपः ।
तत्त्वज्ञानात् परं नास्ति तस्मै श्रीगुरवे नमः

– Guru Stotram #11

Na Guror-Adhikam Tattvam Na Guror-Adhikam Tapah |
Tattva-Jnyaanaat Param Naasti Tasmai Shrii-Gurave Namah ||

Meaning:
Neither is there any Reality Beyond the Guru,
Nor is there any Austerity Higher than the Guru,
There is no Knowledge beyond what comes
From the Guru; Salutations to that Guru.

குருட்டினை நீக்கும் குருவினை கொள்ளார்
குருட்டினை நீக்கா குருவினை கொள்வர்
குருடும் குருடும் குருட்டாட்டம் ஆடி
குருடும் குருடும் குழி வீழு மாறே

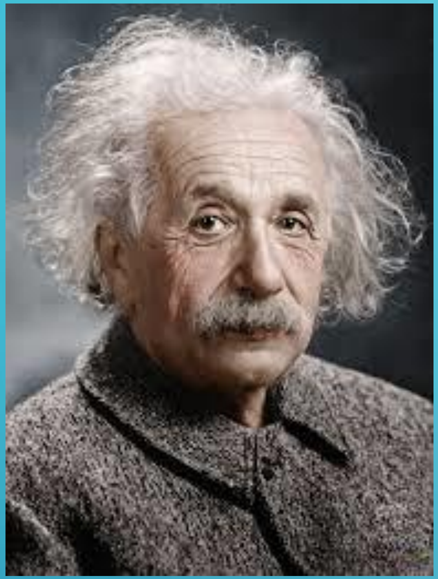
-திருமூலர்

kuruttinai neekkum guruvinai koLLaar
kuruttinai neekkaa guruvinai koLvar
kurudum kurudum kuruttaattam aadi
kurudum kurudum kuzhi veezhumaarE

Meaning:

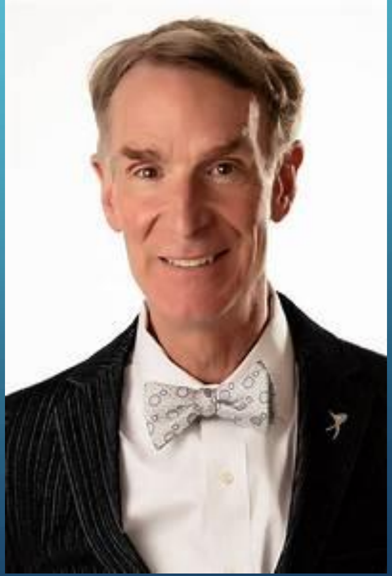
The ones who don't have a Guru who can eliminate their ignorance are blind. The ones who have a person as a Guru, who is not capable of removing the ignorance of their minds, are blind. Both these blind people live their life as the real blind people play a blindfolded game. Both these blind persons, fall into a quagmire where they cannot be recovered, as would the real blind people who would fall into a deep pit

# SOME QUIPS BEFORE WE START



The grand aim of all science is to cover the greatest number of empirical facts by logical deduction from the smallest number of hypotheses or axioms

– Albert Einstein



Everyone you will ever meet knows something you don't.
Respect their knowledge and learn from them. It will bring out the best in all of you.

– Bill Nye

# AGENDA

- About me
- What is analytics?
- About the topic
- From data to decisions – gaps and possible solutions
- Analytics Continuum / Conundrum
- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics
- A walk up the continuum with examples
- Key takeaways
- References and links

# ABOUT ME

- 14+ years of experience building analytics solutions primarily in Semiconductor Manufacturing and Supply Chain Management
- Currently leading a small team that does “Decision Engineering”
- Primary focus
  - Simulation in Operations Research => Descriptive Analytics!
  - Machine Learning / Data Science => Predictive Analytics!
  - Mathematical Optimization (Linear Programming) Algorithms => Prescriptive Analytics!
  - Metaheuristics (nature-inspired) Algorithms => Prescriptive Analytics!
- Interests
  - Internet of things (hacked around a bit) – Neurocognitive IoT, Smart Homes
  - Computer Vision – CNNs for object / character recognition
  - Competitive problem solving – Operations Research (www.puzzlor.com, https://careers.3ds.com/puzzle-quintiq) & Machine Learning (www.kaggle.com)

# TERMS DEMYSTIFIED

| Data Analysis | Data Analytics | Data Science |
| --- | --- | --- |
| The process of collection, manipulation, transformation, arrangement and examination of data. Can be broken down into:<br><br>1) Gathering data<br>2) Manipulate and transform the data into comprehensible formats<br>3) Separating data into component parts<br>4) Data exploration & evaluation<br>5) Examining the data for insights | A broader term that includes data analysis as necessary subcomponent. Analytics defines the science behind the analysis. The science means understanding the cognitive processes (the mechanics or algorithms) an analyst uses to understand problems and explore data in meaningful ways to derive insights. Analytics also includes data extract, transform, and load; specific tools, techniques, and methods; and how to successfully communicate results. | A combination of mathematics, statistics, programming, the context of the problem being solved, ingenious ways of capturing data and extraction of valuable knowledge / information from data. It is geared toward helping individuals and organizations make better decisions from stored, consumed and managed data. Data science was formerly known as datalogy. |
| *Analysis* is separating out a whole entity into parts, study the parts individually and their relationships with one another | *Analytics* is the principle or logic that drives the analysis | *Science* is the systematic knowledge of a system gained through observation and experimentation using stipulated methods or techniques |

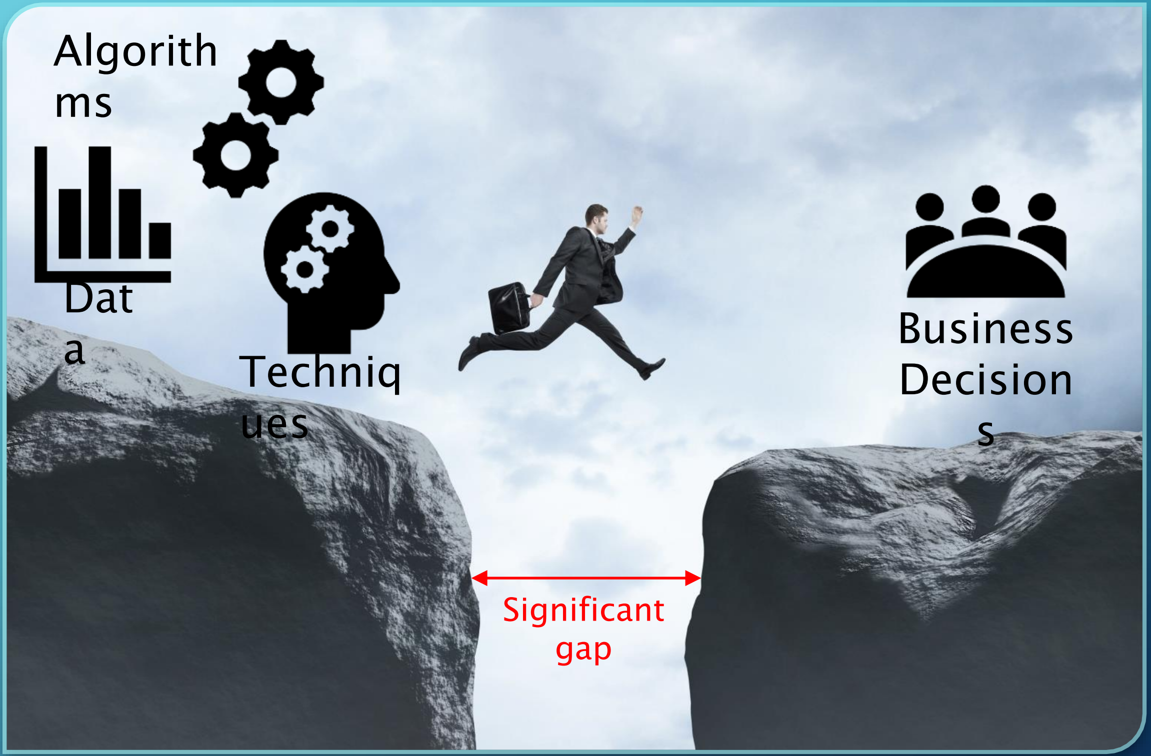
# THE TOPIC – FROM DATA TO DECISIONS #1/3

- Decision / Management Science – how do I answer these questions?
  - Who will be my customers and how many?
  - What will be my demand for the next year (and beyond)?
  - What should be my product(s) lines?
  - How many facilities should I open and where?
  - For supply chains –
    - What transportation option should I use?
    - How should I trade-off service and cost?
    - Where should I source my raw material from?
    - How much inventory should I have?
- We have abundant historical data, a plethora of techniques, algorithms, research and latest developments; still unable to answer the above questions “confidently”

# THE TOPIC – FROM DATA TO DECISIONS #2/3

## Reasons for this decision gap?

- **Models and algorithms are mostly blackbox** – specifically neural networks and complex ensemble models. Unexplainable predictions or un-interpretable results
- **Causal inference** – Models are usually based on correlation. Correlation does not imply causation. For business decisions, I need "cause & effect"; read as Sensitivity Analysis (SA)
- **Magnitude of uncertainty** – I added regularization, dropouts, etc., but did my model generalize well for actual scenarios? Point decisions are arrived at either using Maximum Likelihood Estimates (MLE) or optimizing a cost function
- **External factors** – Models mostly account for known, linear, direct interactions
- **Not comprehending all restraints** – Most business decisions are made under cost, time and effort restraints. Models don't account for these unknowns




LINKS
MLE – https://medium.com/@rrfd/what-is-maximum-likelihood-estimation-examples-in-python-791153818030
SA – http://dpannell.fnas.uwa.edu.au/dpap971f.htm

**BASE IMAGE SOURCE:** THINKSTOCKPHOTOS
**SOME IDEAS ADAPTED FROM:** DR. KARTHIKEYAN SANKARAN, DIRECTOR, LATENTVIEW ANALYTICS

# THE TOPIC – FROM DATA TO DECISIONS #3/3

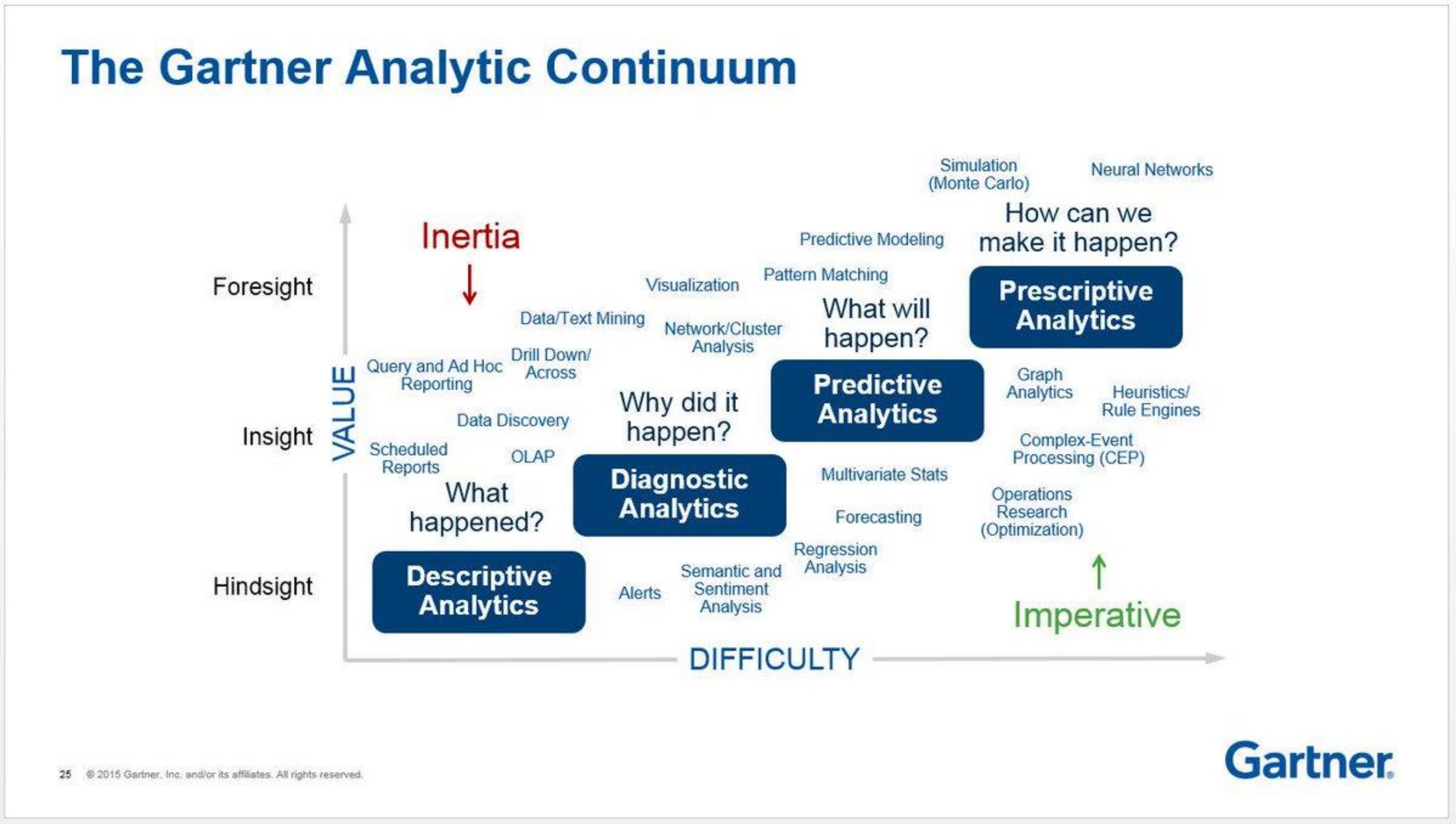
## What are the smart engineers/scientists doing to bridge the gap?

- **Model interpretability**
  - Local Interpretable Model-Agnostic Explanations (LIME) – https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime
  - Interpretable machine learning – https://christophm.github.io/interpretable-ml-book/
  - Deep Learning Feature Visualization – https://distill.pub/2017/feature-visualization/
  - Shapley Values – https://towardsdatascience.com/one-feature-attribution-method-to-supposedly-rule-them-all-shapley-values-f3e04534983d
- **Causal inference**
  - Causal Machine Learning – https://www.youtube.com/watch?v=rkpNZ6Sf7NM ; http://www.unofficialgoogledatascience.com/2017/01/causality-in-machine-learning.html
  - Bayes Nets – https://www.youtube.com/watch?v=gMQZq2O8yDA; https://inst.eecs.berkeley.edu/~cs188/fa18/assets/slides/lec15/FA18_cs188_lecture15_bayes_nets_III_inference_1pp.pdf
  - Structural Equation Modeling – https://www.youtube.com/watch?v=eKkESdyMG9w ; https://www.kdnuggets.com/2017/03/structural-equation-modeling.html
- **Magnitude of uncertainty**
  - Quantification of uncertainty – https://en.wikipedia.org/wiki/Uncertainty_quantification
  - Probabilistic programming/Bayesian Inference – http://jakevdp.github.io/blog/2014/03/11/frequentism-and-bayesianism-a-practical-intro/
- **External factors**
  - Reinforcement Learning techniques like Q-Learning helps in learning the environment / ecosystem.
- **Not comprehending all restraints**
  - Constrained Optimization to the rescue

**SOME IDEAS ADAPTED FROM:** DR. KARTHIKEYAN SANKARAN, DIRECTOR, LATENTVIEW ANALYTICS

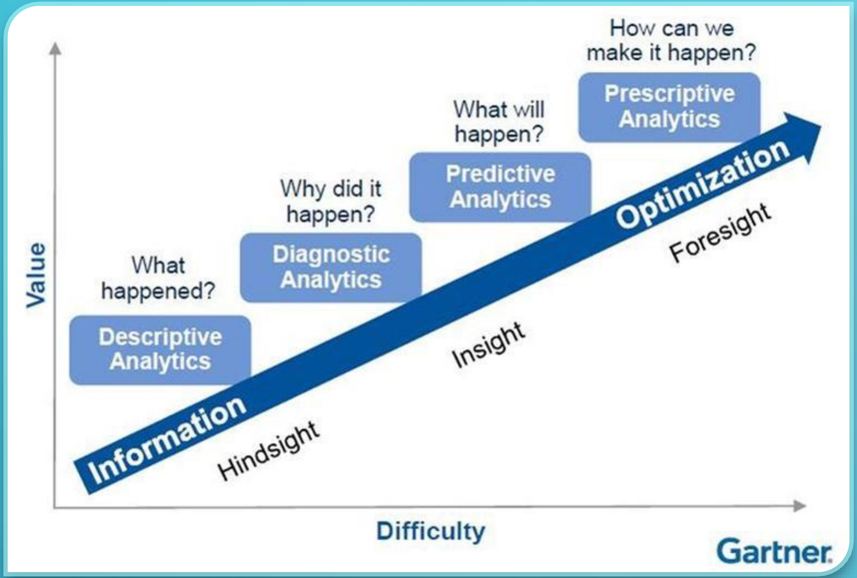
# THE GAP STILL EXISTS!
# WHAT IS THE SOLUTION?
# THAT IS OUR TOPIC FOR TODAY

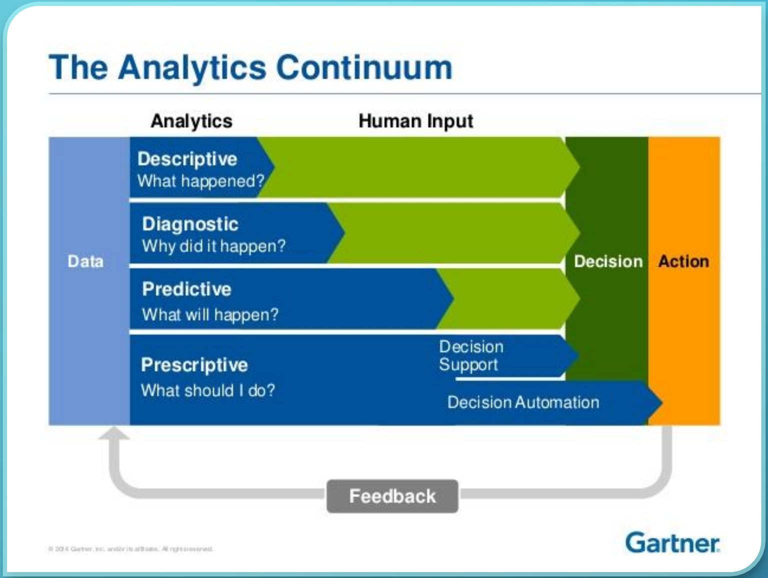
# THE ANALYTICS CONTINUUM #1 / 4

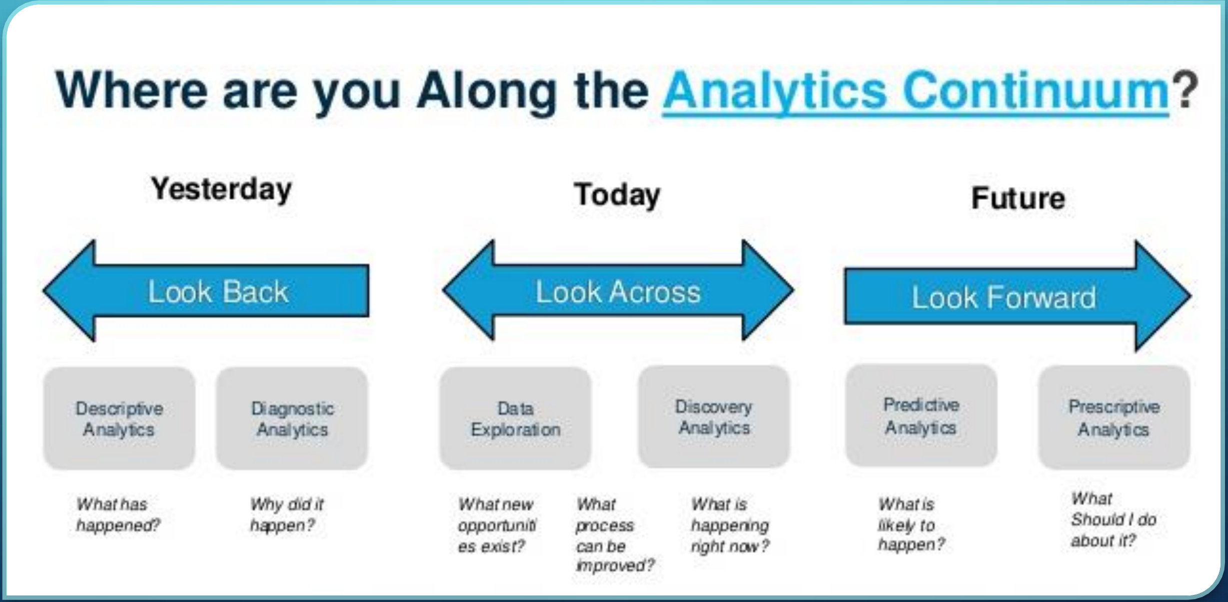



SOURCE: GARTNER © 2015

# THE ANALYTICS CONTINUUM #2/4

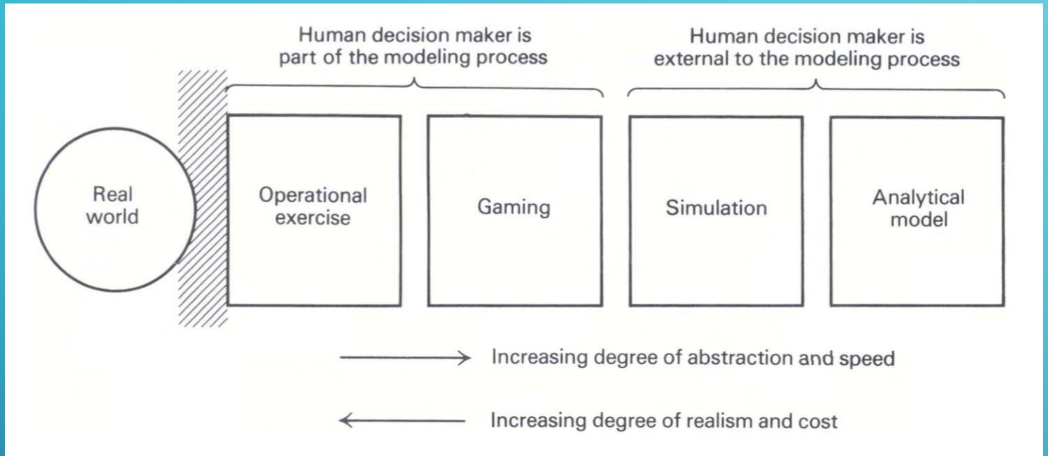


# THE ANALYTICS CONTINUUM – LOOK BACK #3/4

## TYPES OF MODEL REPRESENTATION




## CLASSIFICATION OF MODELS

| | *Strategy evaluation* | *Strategy generation* |
|---|---|---|
| *Certainty* | Deterministic simulation<br>Econometric models<br>Systems of simultaneous equations<br>Input-output models | Linear programming<br>Network models<br>Integer and mixed-integer programming<br>Nonlinear programming<br>Control theory |
| *Uncertainty* | Monte Carlo simulation<br>Econometric models<br>Stochastic processes<br>Queueing theory<br>Reliability theory | Decision theory<br>Dynamic programming<br>Inventory theory<br>Stochastic programming<br>Stochastic control theory |

Statistics and subjective assessment are used in all models to determine values for parameters of the models and limits on the alternatives.

**SOURCE:** APPLIED MATHEMATICAL PROGRAMMING BY BRADLEY, HAX, AND MAGNANTI, 1977 **LINK:** HTTP://WEB.MIT.EDU/15.053/WWW/AMP.HTM

# THE ANALYTICS CONTINUUM – LOOK BACK #4/4

| | Model Characteristics: | | |
|---|---|---|---|
| **Category** | **Form of $f(\cdot)$** | **Values of Independent Variables** | **Management Science Techniques** |
| Prescriptive Models | known, well-defined | known or under decision maker's control | Linear Programming, Networks, Integer Programming, CPM, Goal Programming, EOQ, Nonlinear Programming |
| Predictive Models | unknown, ill-defined | known or under decision maker's control | Regression Analysis, Time Series Analysis, Discriminant Analysis |
| Descriptive Models | known, well-defined | unknown or uncertain | Simulation, Queuing, PERT, Inventory Models |

- Descriptive models start with data. They may not give any decisions at all
- Prescriptive models start with functions/algorithms. They "may" depict the expected decisions, but what about data?

**SOURCE:** SPREADSHEET MODELING AND DECISION ANALYSIS BY CLIFF T. RAGSDALE, 6TH EDITION, 2010

# DESCRIPTIVE ANALYTICS – WHAT IS IT?

- Information needed for actionable decisions
- Principles for systematically collecting and interpreting data that can aid decision makers

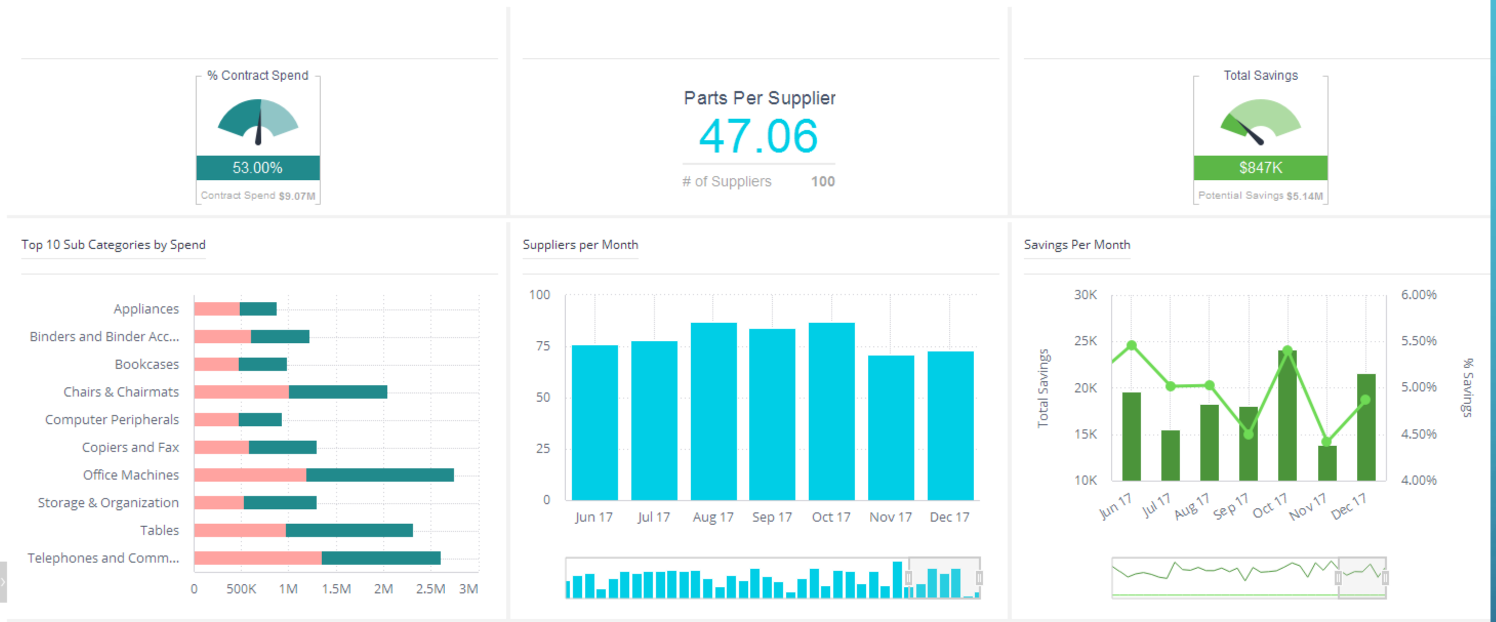
| Exploratory Research | Descriptive Research | Causal Research |
|---|---|---|
| (Ambiguous Problem) | (Aware of Problem) | (Problem Clearly Defined) |
| "Our sales are declining …why." | "What kinds of people are buying our products? Who buys our competitors' products?" | "Will buyers purchase more of our product with a change of our website?" |

SOURCE: BUSINESS ANALYTICS, WHARTON BUSINESS SCHOOL

# DESCRIPTIVE ANALYTICS – INDUSTRY EXAMPLE #1

Procurement Optimization

Source : Procurement Data. July 11th at 7:09PM




Top Savings Opportunities

| Product Name | Min Price | Last Price Paid | Max Price | Quantity | Potential Savings | Current Savings | Spend |
|---|---|---|---|---|---|---|---|
| Bretford CR8500 Series Meeting Room Furniture | $3.20 | $373.31 | $665.63 | 407 | $150,663 | $8,036 | $174,442 |
| Polycom ViewStation™ ISDN Videoconferencing Unit | $6,783.02 | $9,937.8 | $11,463.3 | 41 | $146,032 | $18,990 | $416,041 |
| Global Adaptabilities™ Conference Tables | $2.24 | $344.34 | $474.86 | 321 | $109,866 | $6,427 | $96,878 |
| Sharp AL-1530CS Digital Copier | $379.99 | $722.64 | $819.98 | 260 | $96,991 | $8,432 | $160,764 |
| Global Troy™ Executive Leather Low-Back Tilter | $395.77 | $542.51 | $841.65 | 550 | $89,412 | $16,401 | $341,104 |
| Canon PC940 Copier | $346.49 | $491.39 | $746.98 | 474 | $85,105 | $10,212 | $221,408 |
| Hon 4700 Series Mobuis™ Mid-Back Task Chairs with Adjustable Arms | $24.34 | $289.55 | $555.33 | 309 | $82,403 | $6,600 | $118,008 |
| Hewlett Packard LaserJet 3310 Copier | $455.99 | $665.99 | $1,007.98 | 334 | $80,800 | $19,161 | $230,656 |
| SAFCO PlanMaster Heigh-Adjustable Drafting Table Base, 43w x 30d x 30-... | $265.58 | $478.05 | $555.63 | 331 | $73,842 | $6,714 | $121,853 |
| Bretford Rectangular Conference Table Tops | $289.62 | $563.29 | $586.76 | 256 | $73,767 | $3,334 | $110,141 |
| Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish | $801.69 | $1,074.09 | $1,427.19 | 236 | $71,854 | $10,439 | $259,173 |
| Hewlett-Packard Deskjet 1220Cse Color Inkjet Printer | $328.80 | $532.73 | $657.59 | 292 | $61,469 | $8,427 | $150,599 |
| Okidata ML184 Turbo Dot Matrix Printers | $247.97 | $462.46 | $508.19 | 260 | $58,344 | $4,930 | $91,866 |
| Hon 94000 Series Round Tables | $233.98 | $367.5 | $500.54 | 381 | $52,450 | $7,783 | $127,606 |
| Hon 2090 “Pillow Soft” Series Mid Back Swivel/Tilt Chairs | $210.74 | $380.98 | $466.43 | 264 | $49,952 | $4,424 | $74,574 |
| Office Star - Mid Back Dual function Ergonomic High Back Chair with 2-Wa... | $120.74 | $209.60 | $272.06 | 458 | $41,804 | $4,356 | $82,476 |
| Hewlett-Packard cp1700 [D, PS] Series Color Inkjet Printers | $475.93 | $591.56 | $726.42 | 269 | $37,559 | $7,177 | $146,855 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |

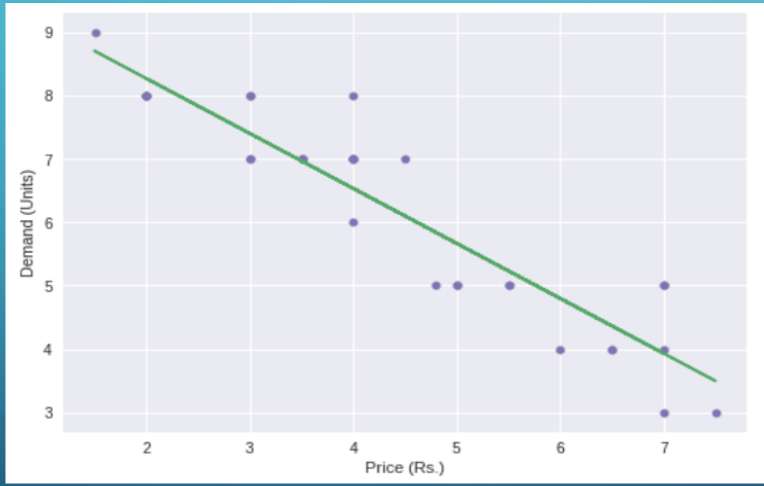
# DESCRIPTIVE ANALYTICS – EXAMPLE FROM SCRATCH

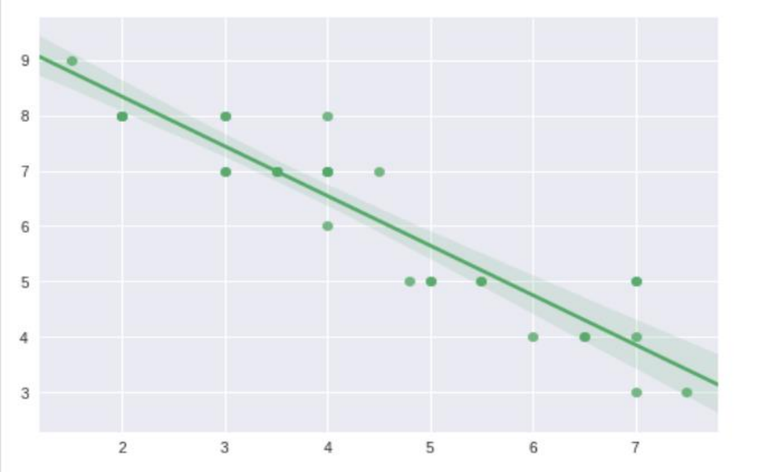
| Price (₹) | Demand (Units) |
|---|---|
| 4.0 | 7 |
| 3.5 | 7 |
| 5.0 | 5 |
| 6.0 | 4 |
| 6.5 | 4 |
| 7.0 | 4 |
| 2.0 | 8 |
| 4.0 | 6 |
| 5.5 | 5 |
| 3.0 | 7 |
| 3.5 | 7 |
| 2.0 | 8 |
| 2.0 | 8 |
| 3.0 | 8 |
| 3.0 | 7 |
| 1.5 | 9 |
| 3.0 | 8 |
| 4.8 | 5 |
| 5.0 | 5 |
| 4.0 | 7 |
| 4.5 | 7 |
| 4.0 | 8 |
| 7.5 | 3 |
| 4.0 | 7 |
| 6.5 | 4 |
| 4.0 | 7 |
| 7.0 | 3 |
| 5.5 | 5 |
| 7.0 | 5 |
| 3.5 | 7 |
| 7.0 | 5 |
| 2.0 | 8 |

A simple data set with price as an independent variable and demand as a dependent variable

| Price | | Demand | |
|---|---|---|---|
| Mean | 4.4 | Mean | 6.1875 |
| Standard Error | 0.307860725 | Standard Error | 0.295932 |
| Median | 4 | Median | 7 |
| Mode | 4 | Mode | 7 |
| Standard Deviation | 1.741523249 | Standard Deviation | 1.674043 |
| Sample Variance | 3.032903226 | Sample Variance | 2.802419 |
| Kurtosis | -0.99506987 | Kurtosis | -1.04812 |
| Skewness | 0.212632356 | Skewness | -0.35996 |
| Range | 6 | Range | 6 |
| Minimum | 1.5 | Minimum | 3 |
| Maximum | 7.5 | Maximum | 9 |
| Sum | 140.8 | Sum | 198 |
| Count | 32 | Count | 32 |

Descriptive_Analytics_Linear_Regression_Example.ipynb



Estimated coefficients:
b_0 = 10.004799510742398
b_1 = -0.867568070623272

# DESCRIPTIVE ANALYTICS – ENTER SIMULATION

A **simulation** is an imitation of the operation of a real-world process or system.[1] The act of simulating something first requires that a model be developed; this model represents the key characteristics, behaviors and functions of the selected physical or abstract system or process. The model represents the system itself, whereas the simulation represents the operation of the system over time. Source: Wikipedia

I came across a very powerful tool from a company called **Anylogic**. You can get a Personal Learning Edition (PLE) installed on any computer for free. https://www.anylogic.com/downloads/personal-learning-edition-download/ . Some examples below:

1) Bank ATM / Cashier example – https://cloud.anylogic.com/assets/embed?modelId=7fb3e1d9-acc8-4b91-b674-47f325408a64

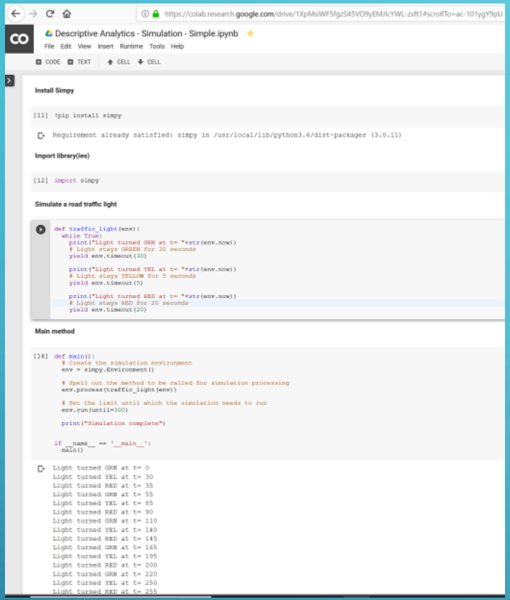
2) Baggage Claim in an airport – https://cloud.anylogic.com/assets/embed?modelId=5e8ff5b8-6ade-46d8-a26a-50729d1780d7

3) A manufacturing job shop factory – https://cloud.anylogic.com/assets/embed?modelId=8a1fecf4-6f48-4461-96c1-a0d5a1327cda

Industries use simulation mostly to see how systems or real processes would behave over time. We could also use simulation to quantify the uncertainty in a system – example Monte Carlo simulation

# DESCRIPTIVE ANALYTICS – SIMULATION WITH PYTHON SIMPY

- We can see how the system would behave over time even though we have a faint idea ahead of time. In the below example I have simulated a traffic signal for 5 minutes



https://colab.research.google.com/drive/1XpMsiWF5fgzS45VO9yEMJIcYWL-zxft1#scrollTo=ac-101ygY9pU

Descriptive Analytics - Simulation - Simple.ipynb

File Edit View Insert Runtime Tools Help

CODE TEXT CELL CELL

**Install Simpy**

```
[11] !pip install simpy
```

```
Requirement already satisfied: simpy in /usr/local/lib/python3.6/dist-packages (3.0.11)
```

**Import library(ies)**

```
[12] import simpy
```

**Simulate a road traffic light**

```
def traffic_light(env):
  while True:
    print("Light turned GRN at t= "+str(env.now))
    # Light stays GREEN for 30 seconds
    yield env.timeout(30)

    print("Light turned YEL at t= "+str(env.now))
    # Light stays YELLOW for 5 seconds
    yield env.timeout(5)

    print("Light turned RED at t= "+str(env.now))
    # Light stays RED for 20 seconds
    yield env.timeout(20)
```

**Main method**

```
[14] def main():
  # Create the simulation environment
  env = simpy.Environment()

  # Spell out the method to be called for simulation processing
  env.process(traffic_light(env))

  # Set the limit until which the simulation needs to run
  env.run(until=300)

  print("Simulation complete")


if __name__ == '__main__':
  main()
```

```
Light turned GRN at t= 0
Light turned YEL at t= 30
Light turned RED at t= 35
Light turned GRN at t= 55
Light turned YEL at t= 85
Light turned RED at t= 90
Light turned GRN at t= 110
Light turned YEL at t= 140
Light turned RED at t= 145
Light turned GRN at t= 165
Light turned YEL at t= 195
Light turned RED at t= 200
Light turned GRN at t= 220
Light turned YEL at t= 250
Light turned RED at t= 255
```

Descriptive_Analytics_Simulation_Simple.ipynb

- As you can see – within 2 seconds I can see what is going to happen over the next 5 minutes – “based on historical information”. This is very powerful to understand the system better.

- In the process, I can generate data which is not otherwise available. This is data augmentation which is a key subject when enough data is not available

# DESCRIPTIVE ANALYTICS – SIMULATION ADVANCED EXAMPLE

A fuel station has a limited number of fuel pumps that share a common fuel reservoir. Cars randomly arrive at the fuel station, request one of the fuel pumps and start refueling from that reservoir. A fuel station control process observes the fuel station's fuel level and calls a tank truck for refueling if the station's level drops below a threshold

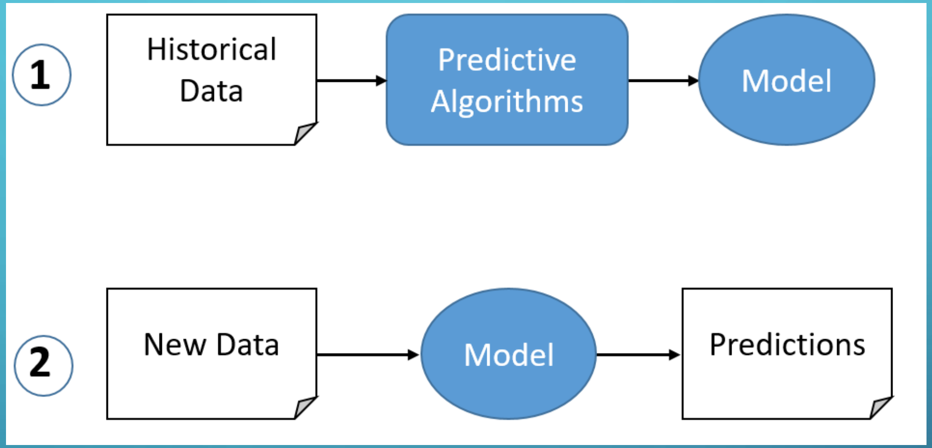
```
Fuel Station refuelling
Car 0 arriving at fuel station at 87.0
Car 0 finished refueling in 18.5 seconds.
Car 1 arriving at fuel station at 129.0
Car 1 finished refueling in 19.0 seconds.
Car 2 arriving at fuel station at 284.0
Car 2 finished refueling in 21.0 seconds.
Car 3 arriving at fuel station at 385.0
Car 3 finished refueling in 13.5 seconds.
Car 4 arriving at fuel station at 459.0
Calling tank truck at 460
Car 4 finished refueling in 22.0 seconds.
Car 5 arriving at fuel station at 705.0
Car 6 arriving at fuel station at 750.0
Tank truck arriving at time 760
Tank truck refuelling 188.0 liters.
Car 6 finished refueling in 29.0 seconds.
Car 5 finished refueling in 76.5 seconds.
Car 7 arriving at fuel station at 891.0
Car 7 finished refueling in 13.0 seconds.
```

Descriptive_Analytics_Simulation_Advanced.ipynb

ADAPTED FROM SIMPY DOCUMENTATION:
https://simpy.readthedocs.io/en/3.0.11/examples/gas_station_refuel.html

# PREDICTIVE ANALYTICS – DEFINITION & EXAMPLES

**Predictive analytics** encompasses a variety of statistical and mathematical techniques that analyze current and historical data/facts to make predictions about future or otherwise unknown events. In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding decision making. The model is then applied to current data to predict what will happen next. At its heart, predictive analytics answers the question, "What is most likely to happen based on my current data?"




- Which product the customer is likely to buy in his next purchase ? (recommender system)
- Which customer is likely to default in his/her loan payment ? (credit risk)
- Who is likely to cancel the product that was ordered through e-commerce portal ? (Logistics cost savings)
- Which warehouse will run out of capacity in the next quarter? (Supply Chain)
- What would be my product demand in the next month? (Manufacturing)

# PREDICTIVE ANALYTICS – WORKED OUT EXAMPLE

**Predicting Future Temperatures**

The problem we will tackle is predicting the max temperature for tomorrow in our city using one year of past weather data. We are going to act as if we don’t have access to any weather forecasts. What we do have access to is one year of historical max temperatures, the temperatures for the previous two days, and an estimate from a friend who is always claiming to know everything about the weather.

This is a supervised, regression problem. It’s supervised because we have both the features (data for the city) and the targets (temperature) that we want to predict.

We will use the Random Forest Regression technique for this problem. During training, we give the random forest both the features and targets and it must learn how to map the data to a prediction.

Predictive_Analytics_Temperature.ipynb

# PRESCRIPTIVE ANALYTICS – DEFINITION & EXAMPLES

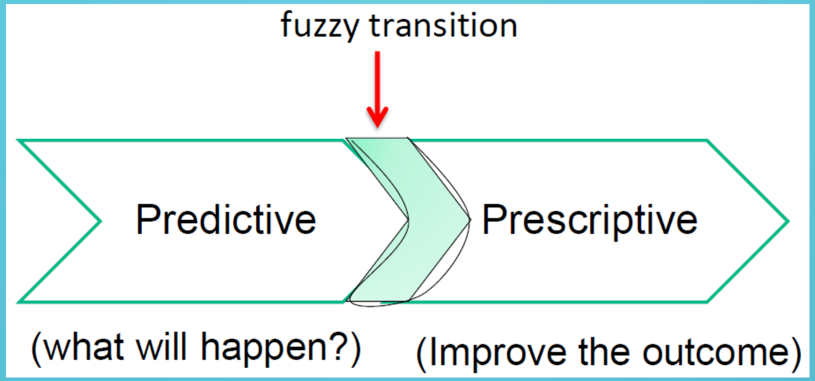
**Prescriptive analytics** is referred to as the "final frontier of analytic capabilities," prescriptive analytics entails the application of mathematical and computational sciences. Prescriptive analytics not only anticipates what will happen and when it will happen, but also why it will happen. Further, prescriptive analytics suggests decision options (answers questions like "what should I do next") on how to take advantage of a future opportunity or mitigate a future risk and shows the implication of each decision option. Prescriptive analytics can continually take in new data to re-predict and re-prescribe, thus automatically improving prediction accuracy and prescribing better decision options. Prescriptive analytics ingests hybrid data, a combination of structured (numbers, categories) and unstructured data (videos, images, sounds, texts), and business rules to predict what lies ahead and to prescribe how to take advantage of this predicted future without compromising other priorities. It encompasses the following:

- **Define the problem**
- **Define the Objectives and Goals**
- **Optimization** – how to find the best action to take

**Examples**:

1) How many units of inventory should I release from an inline store or warehouse in my factory? (Manufacturing)
2) How many trucks or air carriers should I subscribe to for maintaining customer service levels (Supply Chain Logistics)
3) A recommender system suggesting the most optimal recommendations for you based on your constraints (Retail)

# PRESCRIPTIVE ANALYTICS – MORE THOUGHTS & EXAMPLES




**Price optimization:** A particularly good example is airline ticket pricing. Especially in the case of airline pricing, the best solutions are those that can occur in near real time. The addition of the time constraint however does not differentiate predictive from prescriptive

**Optimizing parcel delivery routes and schedules**

**Mechanical preventive maintenance:** The optimization here is the tradeoff between leaving the device in service (presumably revenue generating) and proactively removing it from service prior to breakdown (presumably reducing repair cost and time).

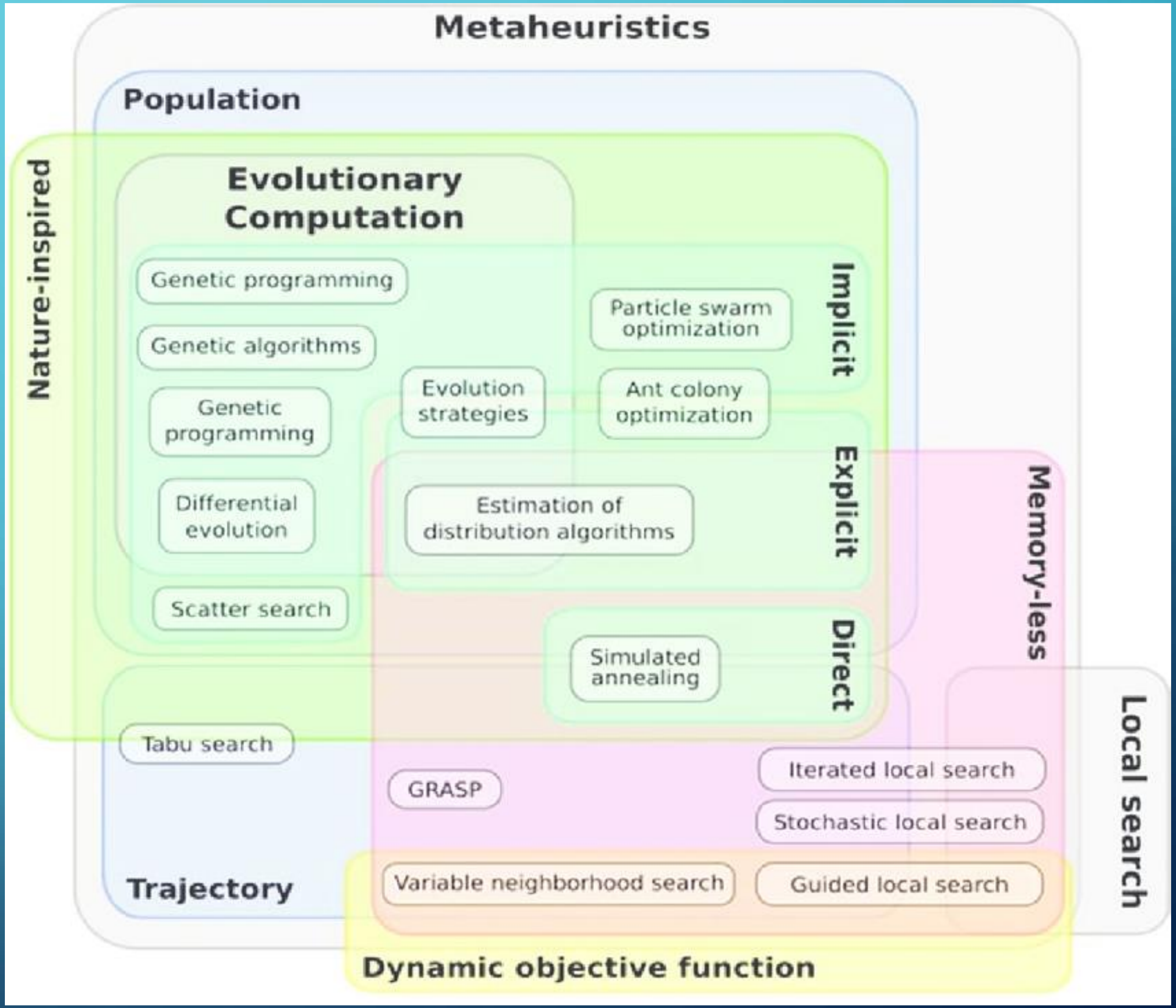
**Staffing optimization:** This is a hot topic for hospitals and other health care providers but could equally apply to the factory floor or service industries.

# PRESCRIPTIVE ANALYTICS – WORKED OUT EXAMPLE

The Department of Applied Mathematics and Computational Sciences (DAMCS) is planning a day long Login trip for 400 students. The company providing the transportation has 10 buses of 50 seats each and 8 buses of 40 seats, but only has 9 drivers available. The rental cost for a large bus is Rs. 800 and Rs. 600 for the small bus. Calculate how many buses of each type should be used for the trip with the least possible cost

# PRESCRIPTIVE ANALYTICS – METAHEURISTICS

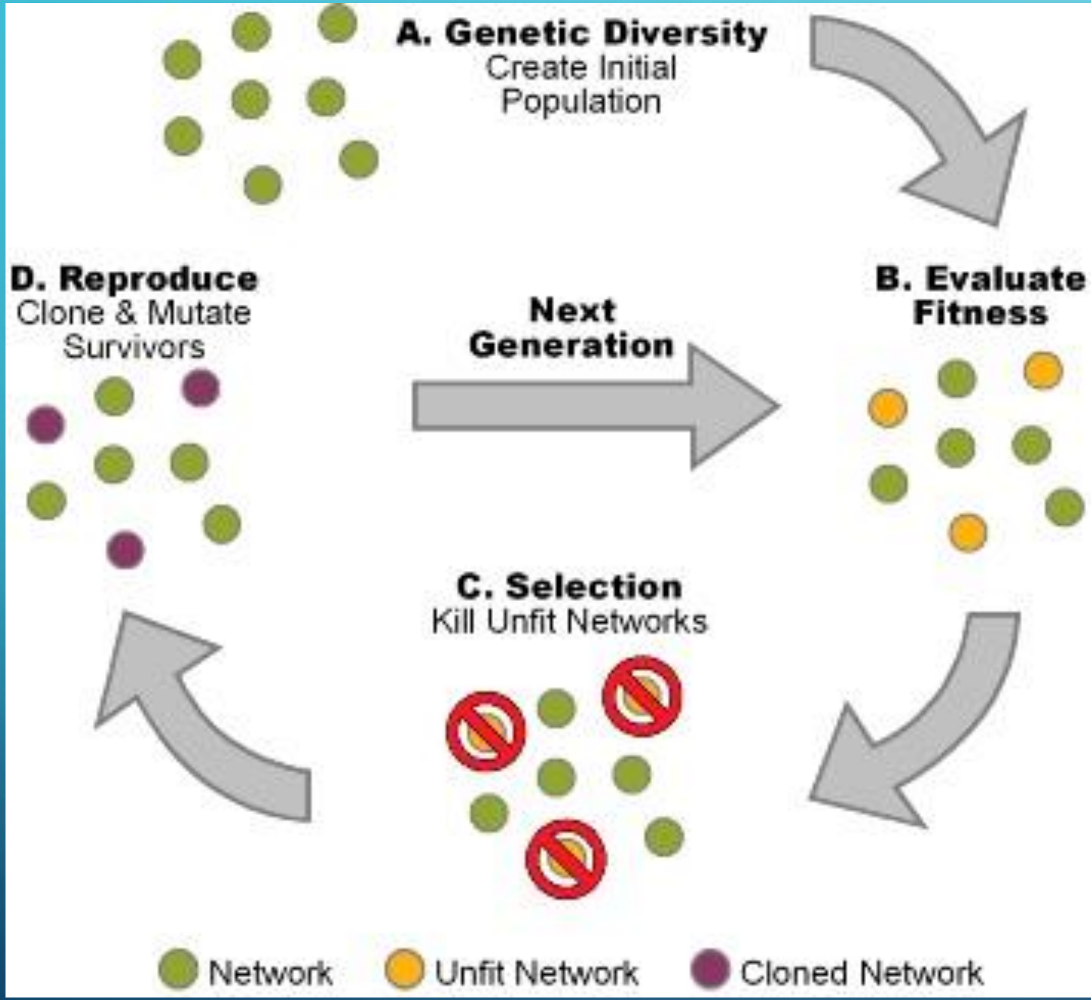
A metaheuristic is a procedure designed to find, generate, or select a heuristic (partial search algorithm) that may provide a sufficiently good solution to an optimization problem, especially with incomplete or imperfect information or limited computation capacity.

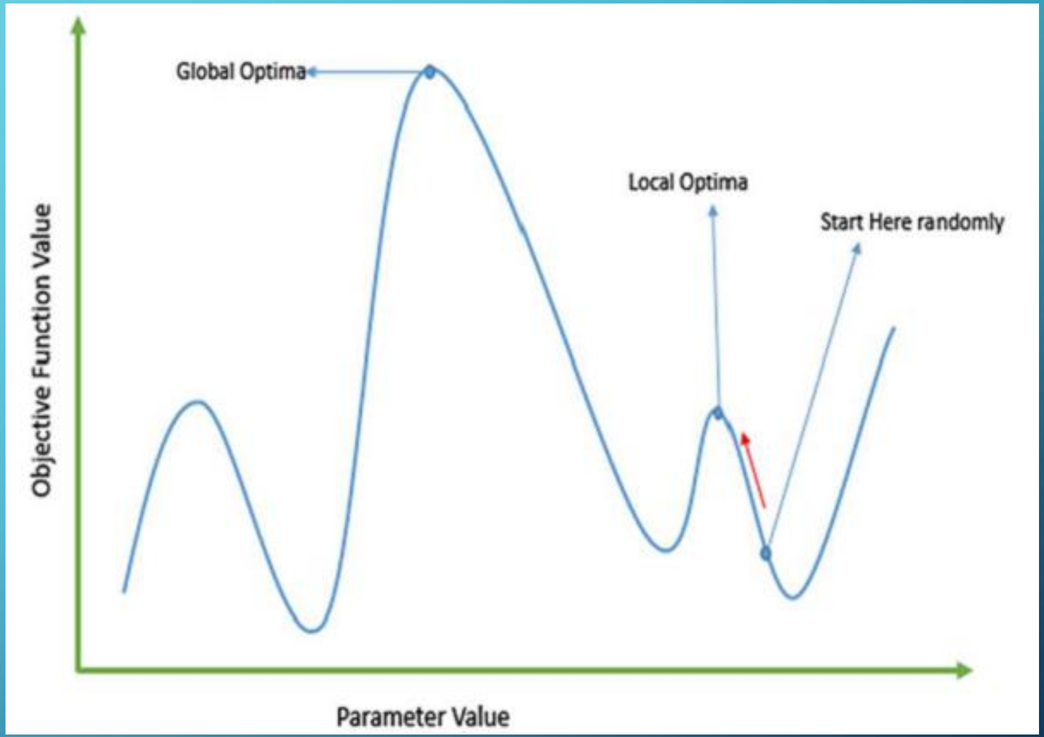


# PRESCRIPTIVE ANALYTICS – METAHEURISTICS

1. Stochastic Algorithms
    1. Random Search
    2. Adaptive Random Search
    3. Stochastic Hill Climbing
    4. Iterated Local Search
    5. Guided Local Search
    6. Variable Neighborhood Search
    7. Greedy Randomized Adaptive Search
    8. Scatter Search
    9. Tabu Search
    10. Reactive Tabu Search
2. Evolutionary Algorithms
    1. Genetic Algorithm
    2. Genetic Programming
    3. Evolution Strategies
    4. Differential Evolution
    5. Evolutionary Programming
    6. Grammatical Evolution
    7. Gene Expression Programming
    8. Learning Classifier System
    9. Non-dominated Sorting Genetic Algorithm
    10. Strength Pareto Evolutionary Algorithm
3. Physical Algorithms
    1. Simulated Annealing
    2. Extremal Optimization
    3. Harmony Search
    4. Cultural Algorithm
    5. Memetic Algorithm
4. Probabilistic Algorithms
    1. Population-Based Incremental Learning
    2. Univariate Marginal Distribution Algorithm
    3. Compact Genetic Algorithm
    4. Bayesian Optimization Algorithm
    5. Cross-Entropy Method
5. Swarm Algorithms
    1. Particle Swarm Optimization
    2. Ant System
    3. Ant Colony System
    4. Bees Algorithm
    5. Bacterial Foraging Optimization Algorithm
6. Immune Algorithms
    1. Clonal Selection Algorithm
    2. Negative Selection Algorithm
    3. Artificial Immune Recognition System
    4. Immune Network Algorithm
    5. Dendritic Cell Algorithm
7. Neural Algorithms
    1. Perceptron
    2. Back-propagation
    3. Hopfield Network
    4. Learning Vector Quantization
    5. Self-Organizing Map

# PRESCRIPTIVE ANALYTICS – GENETIC ALGORITHM WORKED EXAMPLE




C:\Users\
Downloads\FromD

# KEY TAKEAWAY FROM THE TALK

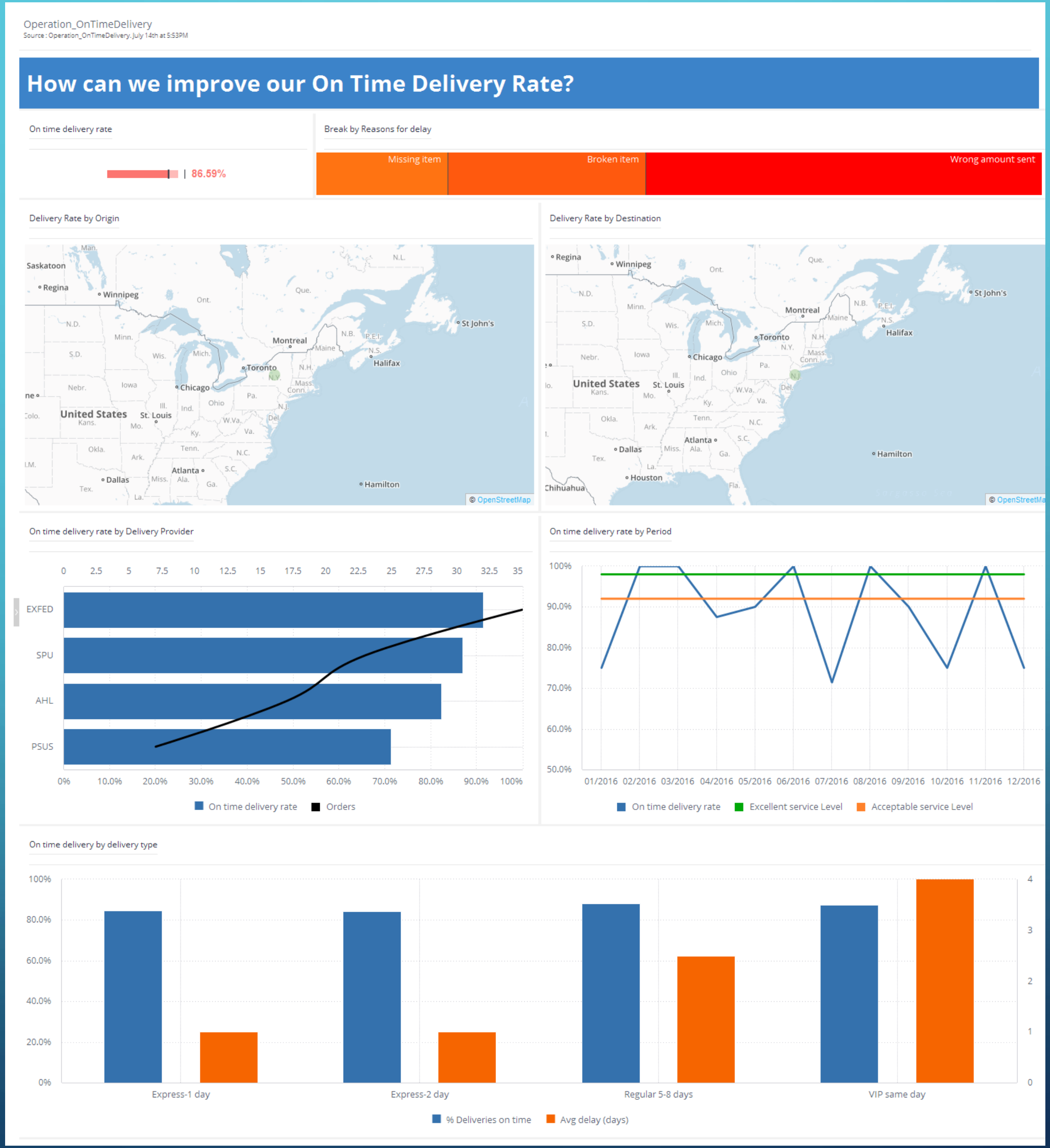
A CAREFUL WALK UP THE ANALYTICS CONTINUUM CAN GIVE YOU “WHAT DECISION” YOU NEED TO TAKE AND “WHEN” FOR YOUR BUSINESS’ COMPETITIVE ADVANTAGE

# KEY POINTERS FOR YOU

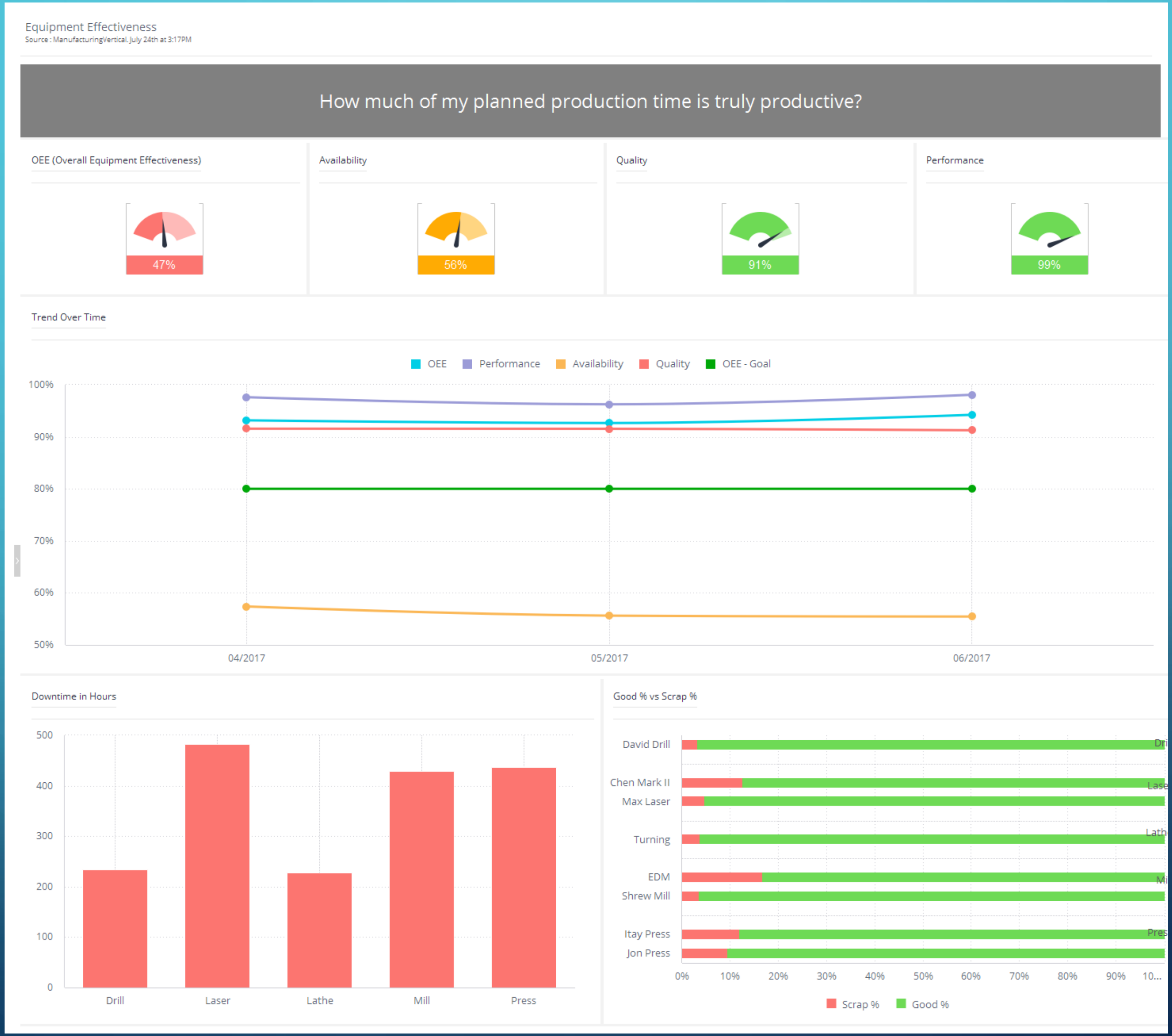
- Stick to the basics. Follow your lecturers/professors
- Maths, algorithms, programming, stats – focus and spend a lot of time. Practise and solve a lot of problems of your own.
- Pick one programming language and stick to it. But know in and out of it. Internals of how it works, memory management, etc.,
- Take up one idea. Make that one idea your life – think of it, dream of it, live on that idea. Let the brain, muscles, nerves, every part of your body, be full of that idea, and just leave every other idea alone. This is the way to success – Swami Vivekananda

# BACKUP

# DESCRIPTIVE ANALYTICS – INDUSTRY EXAMPLE #2



# DESCRIPTIVE ANALYTICS – INDUSTRY EXAMPLE #3



# DESCRIPTIVE – DESCRIPTIVE STATISTICS

| | Year | Month | # of students | # of small buses | # of big buses |
|---|---|---|---|---|---|
| 0 | 2001 | 2 | 284 | 7 | 2 |
| 1 | 2002 | 3 | 277 | 7 | 2 |
| 2 | 2003 | 2 | 317 | 8 | 1 |
| 3 | 2004 | 3 | 313 | 9 | 0 |
| 4 | 2005 | 1 | 318 | 8 | 1 |
| 5 | 2006 | 2 | 374 | 6 | 3 |
| 6 | 2007 | 1 | 413 | 3 | 6 |
| 7 | 2008 | 3 | 405 | 4 | 5 |
| 8 | 2009 | 1 | 355 | 8 | 1 |
| 9 | 2010 | 1 | 306 | 7 | 2 |
| 10 | 2011 | 2 | 271 | 7 | 2 |
| 11 | 2012 | 2 | 306 | 8 | 1 |
| 12 | 2013 | 3 | 315 | 8 | 1 |
| 13 | 2014 | 3 | 301 | 8 | 1 |
| 14 | 2015 | 2 | 356 | 7 | 2 |
| 15 | 2016 | 1 | 348 | 7 | 2 |

```
--------------------------------------------------------------------
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 5 columns):
Year                    16 non-null int64
Month                   16 non-null int64
# of students           16 non-null int64
# of small buses        16 non-null int64
# of big buses          16 non-null int64
dtypes: int64(5)
memory usage: 720.0 bytes
--------------------------------------------------------------------
```

| | Year | Month | # of students | # of small buses | # of big buses |
|---|---|---|---|---|---|
| **count** | 16.000000 | 16.000000 | 16.000000 | 16.000000 | 16.000000 |
| **mean** | 2008.500000 | 2.000000 | 328.687500 | 7.000000 | 2.000000 |
| **std** | 4.760952 | 0.816497 | 42.640698 | 1.549193 | 1.549193 |
| **min** | 2001.000000 | 1.000000 | 271.000000 | 3.000000 | 0.000000 |
| **25%** | 2004.750000 | 1.000000 | 304.750000 | 7.000000 | 1.000000 |
| **50%** | 2008.500000 | 2.000000 | 316.000000 | 7.000000 | 2.000000 |
| **75%** | 2012.250000 | 3.000000 | 355.250000 | 8.000000 | 2.000000 |
| **max** | 2016.000000 | 3.000000 | 413.000000 | 9.000000 | 6.000000 |

# PRESCRIPTIVE – PROBLEM 1: PRODUCT MIX PROBLEM

A telephone company processes and sells two kinds of products:
Desk phones
Cellular phones

Each type of phone is assembled and painted by the company, which wants to produce
at least 100 units of each product.

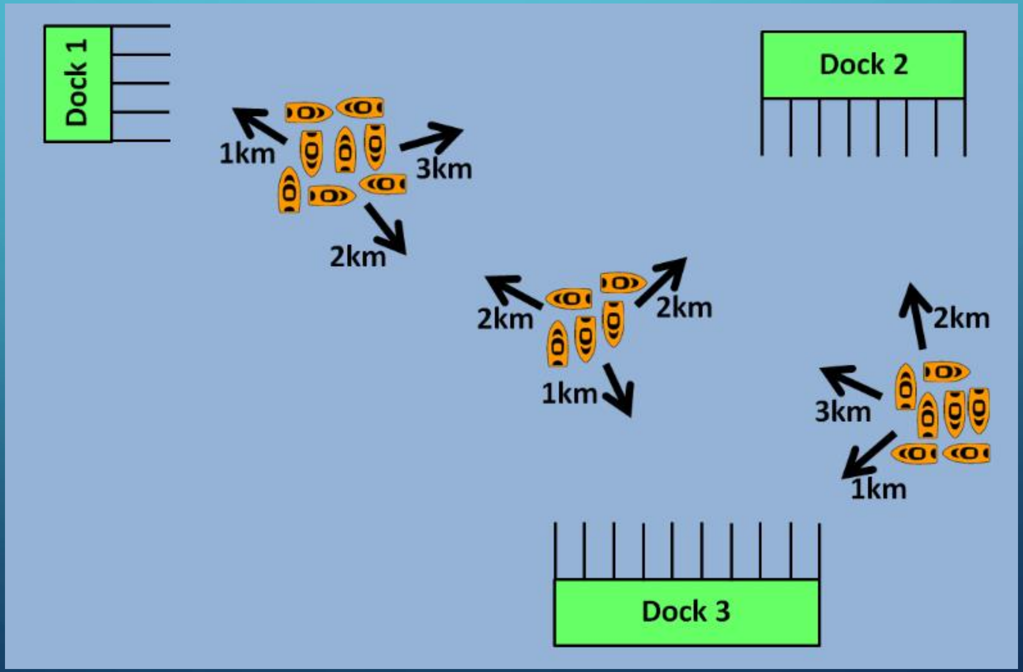
A desk phone's processing time is:
12 min. on the assembly machine and
30 min. on the painting machine
•
A cellular phone's processing time is:
24 min. on the assembly machine and
24 min. on the painting machine

- The assembly machine is available for only 400 hours.
- The painting machine is available for only 490 hours.
- Desk phones return a profit of $12 per unit.
- Cellular phones return a profit of $20 per unit.

The objective is to maximize profit.

# PRESCRIPTIVE – PROBLEM 2: ANY PORT IN A STORM

- There is significant danger to boats caught out in the open sea during a storm. Ideally, boats will dock before the storm hits and wait it out.
- The map below shows 20 orange boats out at sea. With a storm approaching, each boat needs to be directed to one of three docks. Docks have a limited number of spaces available for boats (indicated by the rectangular spaces). Altogether, there are 20 boat spaces available. The boats are clustered into three areas and each area varies in distance to the docks (indicated by black arrows). All boats must be assigned to 1 space in a dock.
- **What is the minimum possible total distance traveled by all boats?** [http://puzzlor.com](http://puzzlor.com)



Bala Ananthanarayanan receiving his new iPad mini from Marijn van Helvoort, Deputy Director at Quintiq, Kuala Lumpur.

# June 2016 Winner
# Bala Ananthanarayanan

"The software industry today expects the most optimal algorithms/solutions that are real-time, fast and scalable. Solving the Quintiq ComBUStion puzzle certainly sets you up for that glaring reality. I have been solving production scale mathematical optimization problems for a little over a decade now. So when one of my colleagues introduced me to the ComBUStion puzzle, it sounded quite familiar. I quickly wrote a solution roughly within an hour, but there was definitely more work to be done to get a perfect score. It was on my things-to-do list for a while before I actually got to it in June of 2016. I formulated the problem as a Binary Integer Program and implemented that using my favorite optimizer to achieve a 100% score. However the real challenge was achieving my personal target of taking less than 7 mins to solve the puzzle. I built another piece of software that would accept the parameters for that problem instance, call my Binary Integer Program and display the output in a way that can be easily fed into the Quintiq user interface. I finally managed to solve it in 7:02 mins. Overall, it was a thoroughly engrossing experience."

**- Bala Ananthanarayanan**