ADAX

DAY 1 – Lecture 1
Brief Introduction to Data Science

Data Driven Organization (DDO)

Today's session is delivered by …

Dr G. Rishwaraj - Data Scientist

From

The Center of Applied Data Science

PROFESSIONAL EXPERIENCE

Functional Expertise:

- Intelligent system study and research.
- Teaching and training.

Industry Expertise:

- Industry (training and research)
- Academia (teaching and research).
- Manufacturing (process engineer).

Qualifications:

- PhD in Engineering (2017)
- B.Eng Manufacturing Engineering major in Process Engineering (Hons) (2010)

Overview

1. Introductions
2. The Data Driven Organization (DDO) concept
3. Analytics Organization Design
4. Roles and team structures within the data eco-system
5. Wrap up

Samira

All sessions will be punctuated with interactive activities and examples of analytics applications.

Original content credits goes to ADAX.

Let’s Get to Know Each Other…

- Where do you come from? Bachelor’s, Masters, PhD?
- Data and analytics capability – statistics? programming? (scale of 1-5)
- What do you want to get from the Data Star program?


So like, Why are we here again?

So why?

#16	3,433	$105,395	#1
Highest Paying Job in Demand	Number of Job Openings	Average Base Salary	Best Job in America for 2016

Sources: 25 Best Jobs in America and 25 Highest Paying Jobs in America for 2016

"This hot new field promises to revolutionize industries from business to government, health care to academia."

— *The New York Times*

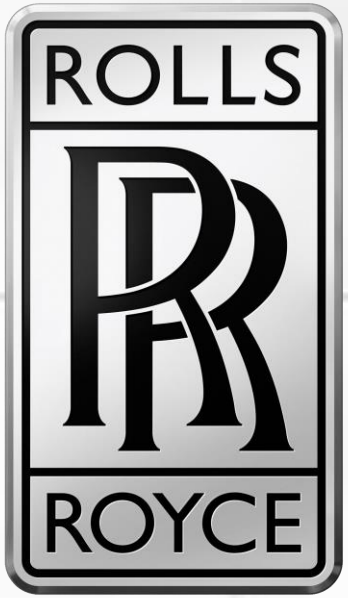
According to Linda Burtch, managing director at executive recruiting firm Burtch Works, "Within 10 years, if you're not a data geek, you can forget about being in the C-suite."

INTRODUCTION

An outline of the need for data and analytics.

STORY TIME

- What does Dickey's Barbecue Pit, Rolls Royce and Transport for London in common?



- Reacting to data outcomes within 12 to 24 hours instead of waiting for the end of the week.

By the end of this lecture, we aim to …

- Provide an understanding of the driver and demand for data analytics.
- Provide an understanding on the foundations for a data driven organisation.
- Provide an overview of the fundamentals of data and data science.
- Outline roles within the data and analytics eco-system and typical organisational structures.
- Show examples and applications of analytics.

The motivation for data and analytics

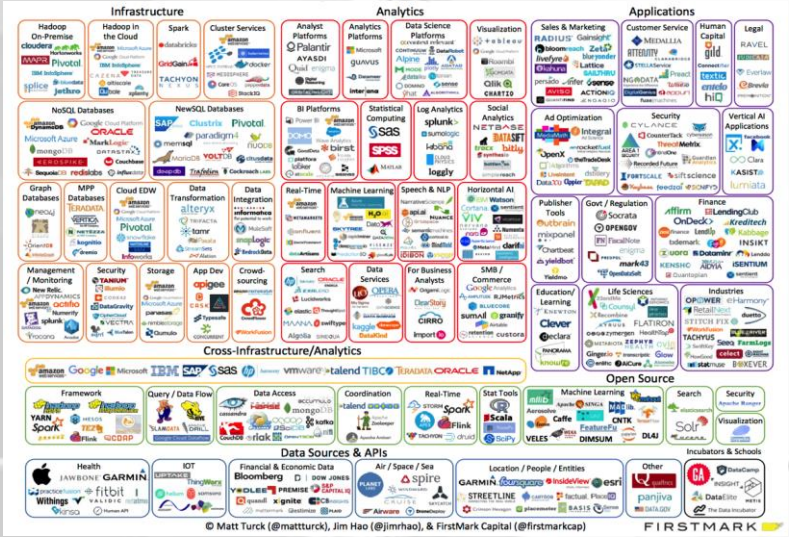
DRIVING FACTORS OF DATA & ANALYTICS

Rise in data

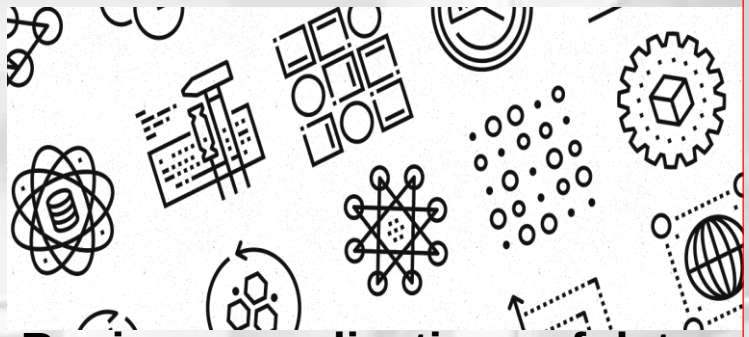
Yesterday	Today
Work	Life
Goal	Purpose
Managers	Moderator
Instructions	Values and culture
Avoid mistakes	Experiment, trial and error
Findings	Generating actionable insights
National	Global
Hierarchical	Meritocracy
Departments	Networking
Tell me	Inform me
Segmentation	Segmentation of 1
Blanket marketing	Targeted communications
Mass messaging	Personalised messaging

Changing world

Drivers of Data & Analytics



Changing Technologies

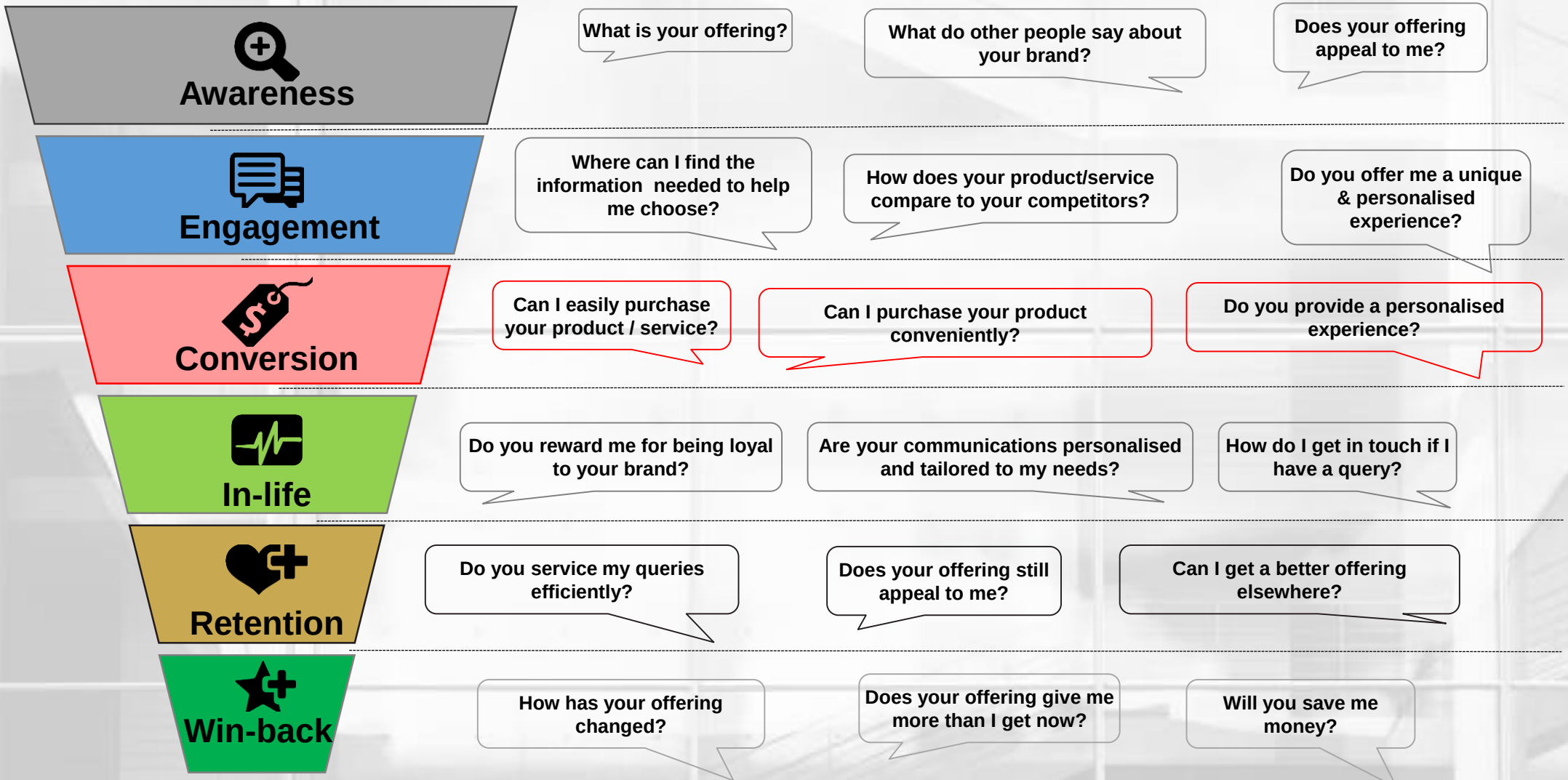


Business applications of data science

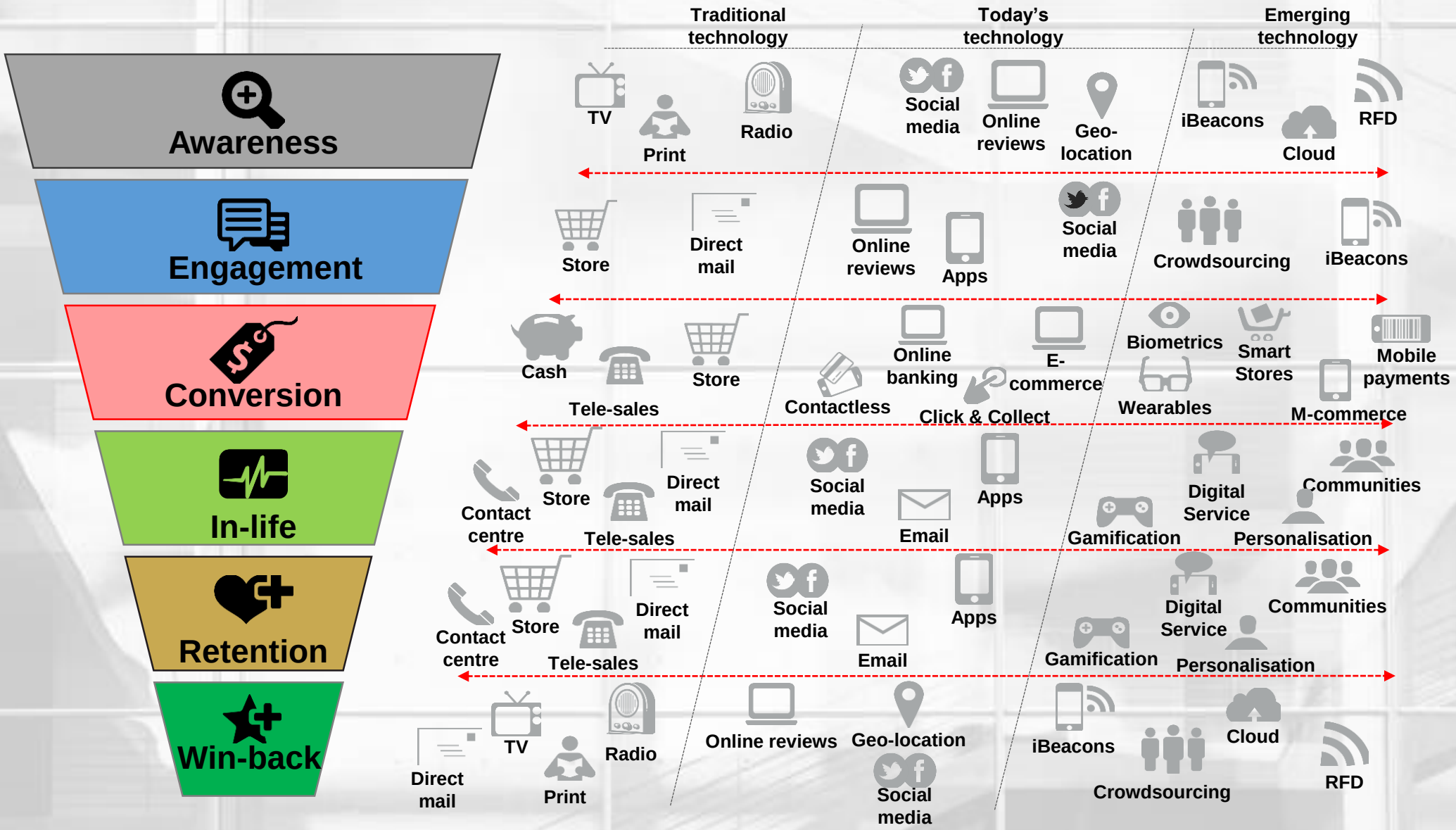
THE CHANGING WORLD

Yesterday	Today
Work	Life
Goal	Purpose
Managers	Moderator
Instructions	Values and culture
Avoid mistakes	Experiment, trial and error
Findings	Generating actionable insights
National	Global
Hierarchical	Meritocracy
Departments	Networking
Tell me	Inform me
Segmentation	Segmentation of 1
Blanket marketing	Targeted communications
Mass messaging	Personalised messaging

WHAT DOES YOUR CUSTOMER WANT?



CHANGING TECHNOLOGY




DATA
LIGHTYEAR
DATA EVERYWHERE

Data, data everywhere…

This rapidly evolving landscape is both enabling and challenging organisations to better understand their data and their customers

Google caches the internet twice daily

Google stores 15 Exabyte of data

Google receives **2 million search** requests every **minute**

Tesco has data for **15 million customers** – spanning over **20 years…**

90% of the data in the world today has been created in the last 2 years

The volume of e-commerce transactions in **2016** reached **$326 billion**…

The world's **500,000+ data centres** are large enough to fill **5,955 football fields….**

Business applications of data science?

Organisations are searching for answers to meet the customer challenges of today and tomorrow

I need to engage with my customer more effectively

I need to know more about my Customer and how to contact them

I want to engage with my customer more effectively

I want to base my decisions based on data and insight

I want to price my offering more competitively

I want to build a customer focussed organisation

My marketing campaigns are ineffective

I want to understand my data better and use it more efficiently

I'm losing customers to my competitors

My stores are not driving sales anymore

My reporting is ineffective

My cost to serve is increasing

I am looking to expand my store network

Can you identify your competitors?

Data Driven Organizations

What do we mean by DDO?

A mature Data Driven Organisation (DDO)

- ✓ Has analytics embedded in all decision making activities
- ✓ Do not treat analytics as a project
- ✓ Treat analytics is a core capability and is used to create insight
- ✓ DDOs drive trusted insight from internal and external data, analysis and reasoning in order to support organization-wide decisions and take actions accordingly

DDOs have 3 distinctive analytical characteristics;

- ➢ They are focusing on right questions.
- ➢ Doing the right analysis.
- ➢ Taking the right actions.

HiPPOs

A data-driven culture is meaningless without the support of the CEO and executive team -- and their own willingness to challenge assumptions that they hold dear. A top-down approach is the only way to break the tyranny of the "HiPPOs" (Highest Paid Person's Opinions).

~David Selinger (Amazon.com, RichRelevance)

Importance of DDO.

Let's look at numbers:

- ✓ 3 x more likely to acquire customers.
- ✓ 6 x as likely to retain those customers.
- ✓ 19 x as likely to be profitable as a result.

- Consider these aspects:
  - The Whys' and Hows'.
  - Overwhelming amount of information/data.

- Data from CrowdFlower recently showed that 83% of organizations are struggling to meet the data skills requirements they need.

ADAX Data Driven Organisation Maturity Model:

How would becoming a DDO help us and how to get there?

How capable is our analytics practice?

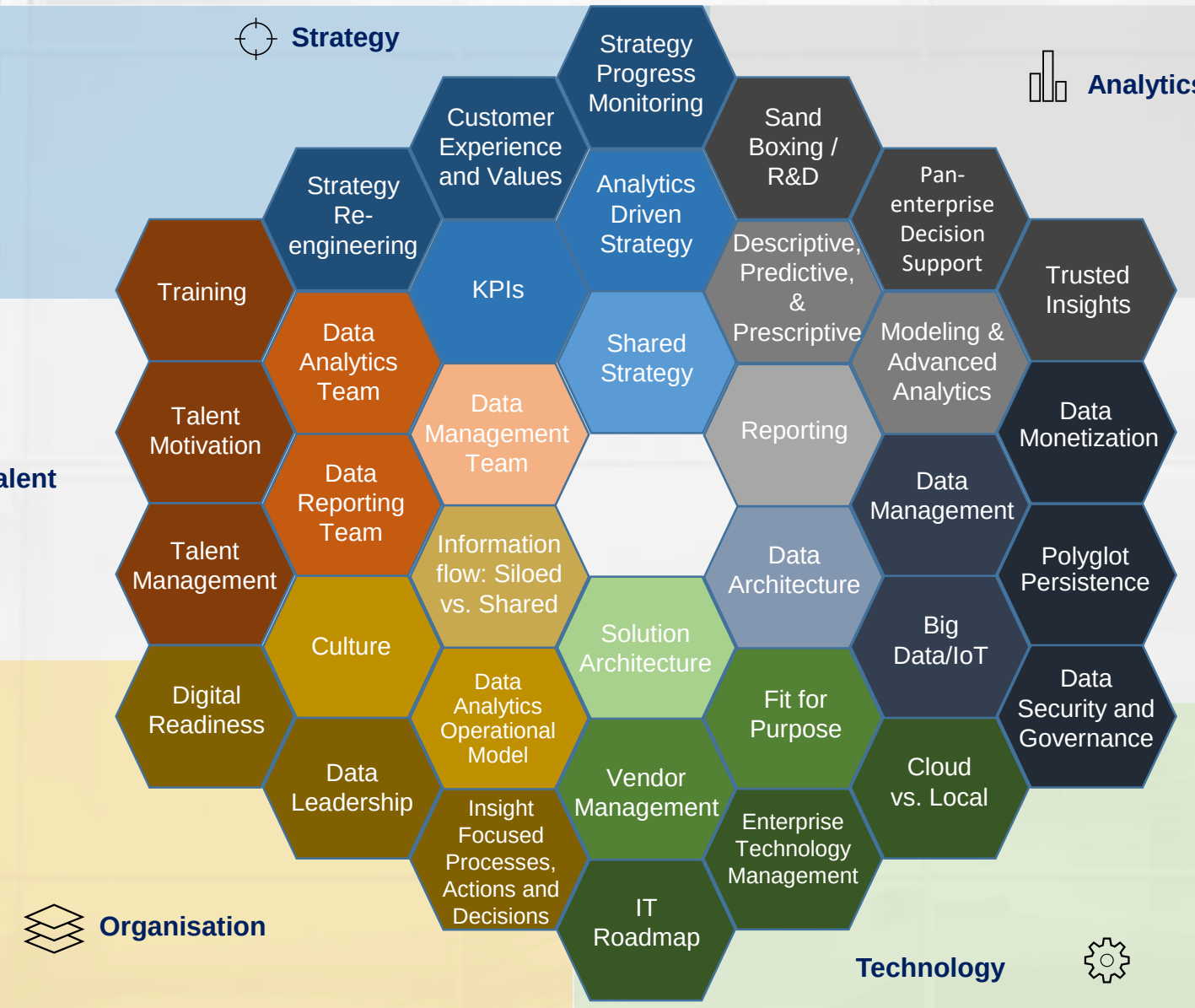
Does our people have the right skills and take the right actions?

How does our data support achieving our strategic objectives?

Does our culture and processes effectively support and control the delivery of insights?

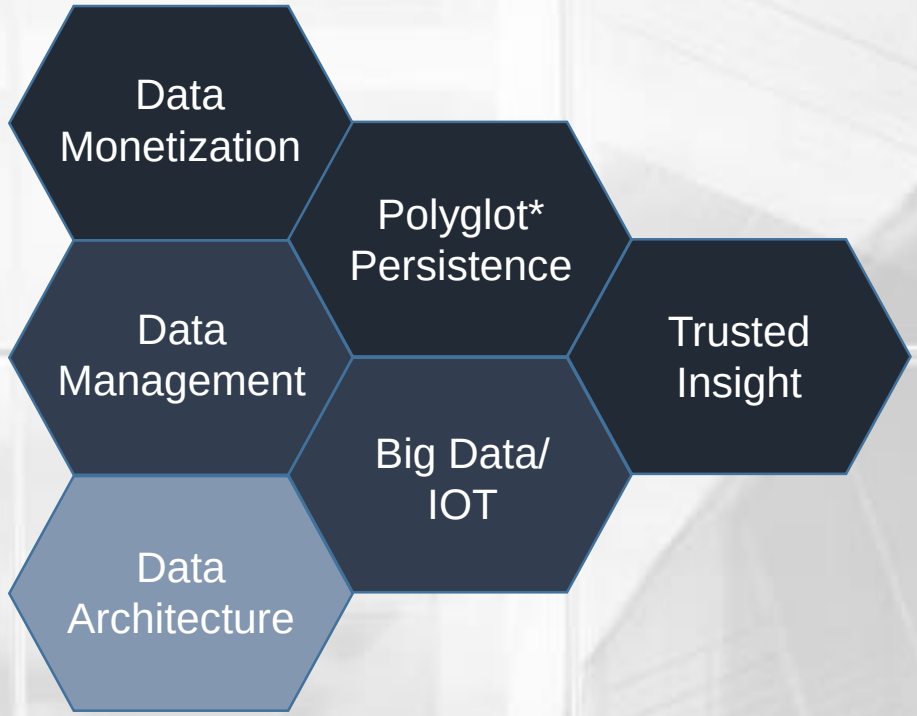
Do we have an integrated scalable technology platform in place?

ADAX DDO MATURITY MODEL:



DATA

- Have you completed the MoSCoW exercise across your organisation?
  - **Must** – Project will fail or company will not be able to operate at top level
  - **Should** – Internal data that will enable you to draw out additional insights
  - **Could** – External data that you could purchase or acquire to support the analysis
  - **Want** – Internal or external data that is typically used for exploratory analysis



Characteristics of different maturity model:

Mature in data and technology

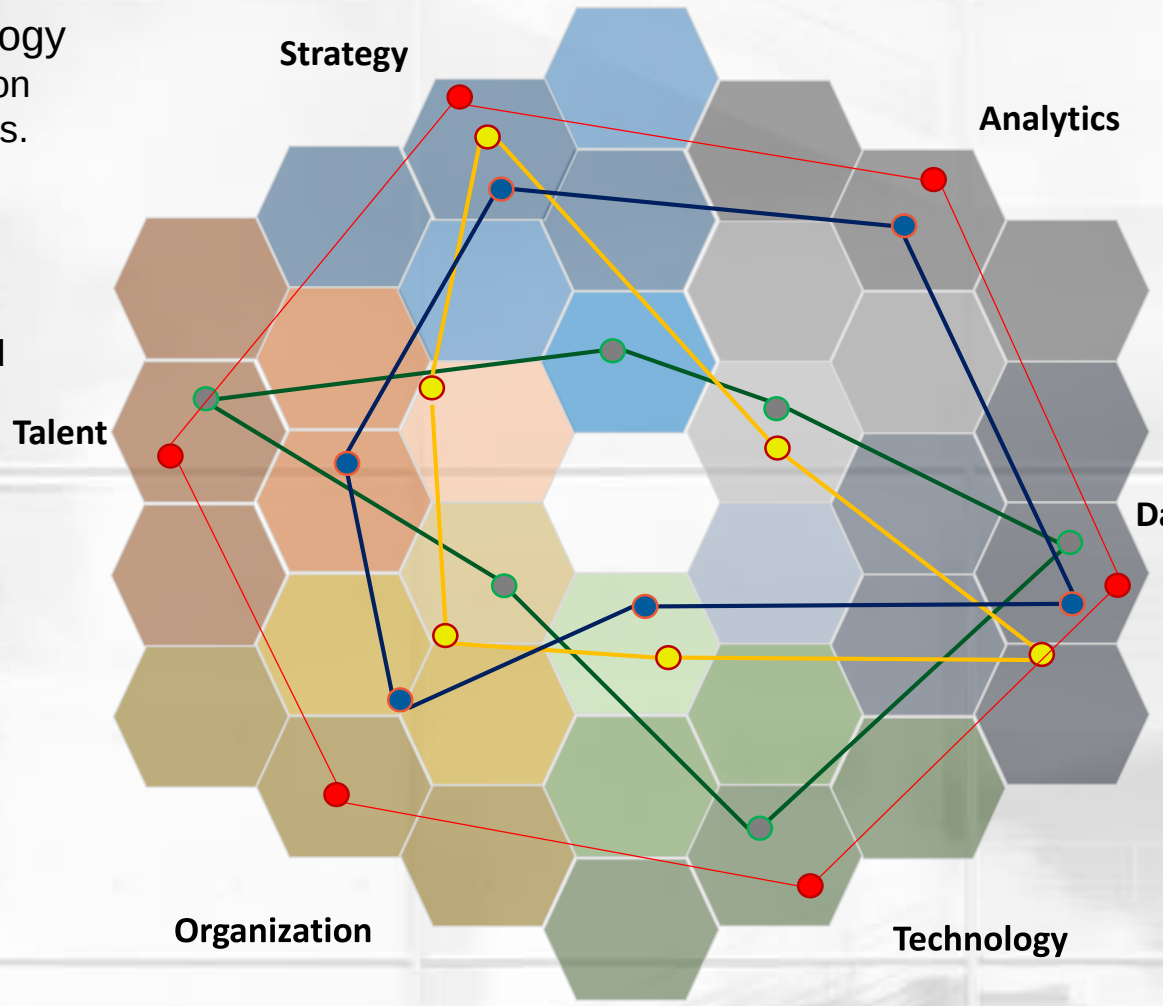
These organisations rely heavily on technology to deliver their services. As a result have access to large quantum of data.
Being technology led, they are slow to movers from a strategy, organisational culture change and analytics adoption perspective

Typical sectors: Telco, media.

Mature strategy

Traditional monoliths with large amount of data. The appetite to understand and make use of the data is there but the execution at the organisational level is low. Limited internal technology knowledge is seen as a limitation

Typical sectors: Healthcare, Newsprint.



Mature data and strategy

Organisations that operate in heavily regulated industries are motivated to manage and use their data as an asset but at the same time crippled by legacy systems and data regulations.

Typical sectors: Retail banks, Utilities.

Mature across all areas

Newly formed in the information age with little or no legacy systems and regulations.
Innovation based on data and analytics culture drives the organisation.

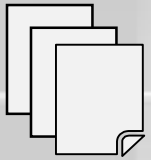
Typical sectors: Online gaming, consumer services (Uber etc).

MATURITY LEVELS

5th Degree: DDO	Trusted insight created by enterprise wide analytics supports decision making. The enterprise is reaping benefits and is focused on optimisation of analytics	Enterprise moving into artificial intelligence where certain internal and customer experience processes are replaced with AI e.g. customer contact strategy.
4th Degree: Analytical Enterprise	Analytics is viewed as the enterprise priority. The company is developing enterprise wide analytics capability to create meaningful contents and ideas	Decision making and business planning processes are trusted insight driven, execution plans/activities are tracked and monitored.
3rd Degree: Analytical Aspirations	Expanding siloed functional analytics to shared operational level analytics with support and commitment from C-suite	Enterprise-wide data, insights and information sharing with a strategy in execution stage.
2nd Degree: Localised Analytics	Pockets of analytics across the company, however in silos and no overarching data and analytics strategy	Focus on the data and analytics teams and decide on an effective structure for the organisation e.g. centralised or federated structure works best
1st Degree: Analytically Impaired	Minimal analytics activities and infrastructure across the company, with ambiguous data and analytics strategy	Acknowledge current situation through a root and branch review of data and analytics capabilities across strategy, analytics, data, technology, organisation culture and talent.

THE STRUCTURES

Data Sources



Data Platform

NoSQL Solutions

Data Warehouse

Cloud Solutions

Local Solutions

Analytics Platform

Demographic Analytics

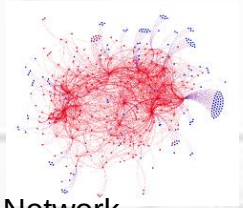
Web Analytics

$$FP_t = a_0 + \sum_{i=1}^{n} (\beta_i X_{it}) + e_t$$

Modelling

$$FP_{t+n} = \widehat{a_0} + \sum_{i=1}^{n} (\widehat{\beta_i} X_{it})$$

Forecasting



Network Analysis

Machine Learning

Trusted Actionable Insight and Automated Actions

Trusted Insight and Intelligence

Automated Actions

DATA IS AN ASSET

- The following must be considered key strategic assets:
  - Data
  - Data science capability

- Derive value from them:
  - Understanding customer needs to improve customer experience subsequently increasing retention
  - Product development and diversification
  - Market sizing – potential new market growth and defending current market share
  - Operational efficiencies

- “Assets” because we occasionally invest in them
  - Acquire data at cost
  - Hire, train, and retain a data science team and data-aware management

Roles and Team Structures within Data Eco System

DATA SCIENCE / ANALYTICS MANAGER OR DIRECTOR

- Member of the senior management in data-heavy firms or reporting into CEO in large organisation
- Understands business needs intimately
  - This includes cost/benefit analysis of data projects
- Understands basic data science (possibly more)
  - Evaluates proposals for data science projects
  - Determines if a project is plausible
- Coordinates storing and sharing data across departments
- Hires, trains, retains data scientists and data engineers

DATA SCIENTIST

Data Scientist

"As rare as Jedi"

Role

Cleans, massages, and organizes (big) data

Mindset

Curious data wizard

Skills & Talents

- Distributed computing
- Predictive Modeling
- Story-telling and visualizing
- Math, Stats, Machine Learning

Languages

R, SAS, Python, Matlab, SQL, Hive, Pig, Spark

Data Camp, 2015

DATA ANALYST

Data Analyst

“Data Detective”

Hired by:

Role

Collects, processes, and performs statistical data analyses

Mindset

Intuitive data junkie with high “figure-it-out” quotient

Skills & Talents

- Spreadsheet tools (e.g. Excel)
- Database systems (e.g. SQL and NO SQL based)
- Communication and visualization
- Math, Stats, Machine Learning

Languages

R, Python, HTML, Javascript, C/C++, SQL

Data Camp, 2015

DATA ARCHITECT

Data Architect

“The contemporary data modeler”

Hired by:

Role

Creates blueprint for data management systems to integrate, centralize, protect and maintain data sources.

Mindset

Inquiring ninja with a love for data architecture design patterns

Skills & Talents

- Data warehousing solutions
- In-depth knowledge of database architecture
- Extract Transformation and Load (ETL), spreadsheet, and BI tools
- Data modeling
- Systems development

Languages

SQL, XML, Hive, Pig, Spark

Data Camp, 2015

DATA ENGINEER

Data Engineer

“Software Engineer by Trade”

Hired by:

Role

Develops, constructs, tests and maintains architectures (such as database and large-scale processing system)

Mindset

All-purpose everyman

Skills & Talents

- Database systems (SQL & NOSQL based)
- Data modeling & ETL tools
- Data APIs
- Data warehousing solutions

Languages

SQL, Hive, Pig, R, Matlab, SAS, SPSS, Python, Java, Ruby, C++, Perl

Data Camp, 2015

STATISTICIAN

Statistician

“Historic leaders of data”

Hired by:

Role

Collect, analyzes and interprets qualitative as well as quantitative data with statistical theories and methods

Mindset

Logical and enthusiastic statistical genius

Skills & Talents

- Statistical theories & methodology
- Data mining & machine learning
- Distributed computing (Hadoop)
- Database systems (SQL & NOSQL based)
- Cloud tools

Languages

R, SAS, SPSS, Matlab, Stata, Python, Perl, Hive, Pig, Spark, SQL

Data Camp, 2015

Database Administrator

Database Administrator

"Database Caretaker"

Hired by:

Role

Ensures that database is available to all relevant users, is performing properly and is being kept safe

Mindset

Master of disaster prevention

Skills & Talents

- Backup & recovery
- Data modeling & design
- Distributed computing (Hadoop)
- Database systems (SQL & NOSQL based)
- Data security
- ERP & business managements

Languages

SQL, Java, Ruby on Rails, XML, C#, Python

Data Camp, 2015

Business Analyst

Business Analyst

“Change agent”

Hired by:

ORACLE

DELL

Role

Improves business processes as intermediary between business and IT

Mindset

Resilient project juggler

Skills & Talents

- Basic tools (e.g. MS Office)
- Data visualization tools (e.g. Tableau)
- Conscious listening and storytelling
- Business Intelligence and understanding
- Data modeling

Languages

SQL

Data Camp, 2015

Data and Analytics Manager

Data and Analytics Manager

"Data Science Team Leaders"

Hired by:

slack MOTOROLA coursera

Role

Manages a team of analysts and data scientists

Mindset

Data Wizard's Cheerleaders

Skills & Talents

- Database systems (SQL & NOSQL based)
- Leadership & project management
- Interpersonal communication
- Data mining & predictive modeling

Languages

SQL, R, SAS, Python, Matlab, Java

Data Camp, 2015

DDO Design principles

Strategic	Insight and data teams will deliver requirements and outputs (reports & insight) which help the organisation understand progress towards the vision and strategy and manage day-to-day business
Leadership	An individual (or group of individuals) will lead data governance processes and be responsible for driving effective data management across the organisation. Business owners will be accountable for definition and use of insight that they need
Governed	There will be clear data governance processes that drive efficiency and effectiveness of data and insight processes and embed data management standards across the organisation
Consolidated	Inconsistent data sources and technologies will, as far as is appropriate, be joined up and synchronised to create a single version of the truth, with clear master data management rules
Collaborative	Insight and data teams will work across organisational siloes to deliver the insight needed to support strategic decision making
Supported	End users will be provided with the required level of training, subject matter experts' input and technical support to action the insight generated, as well as create their own insights
Capability	The analytics function will have the required technical and business skillsets to deliver reports & insights as needed, both regularly and ad-hoc. These people may be embedded in departments or in a centre of excellence
Independent	Central insight and data activities will be independent of organisation silos, to ensure that requirements are appropriately prioritised and data and insight is not biased
Versatility	The insight and data capability will be flexible enough to adapt to the needs of each function and evolve with changing demands, whilst also ensuring ad-hoc needs can be met
Simple	The processes and outputs for accessing and interpreting insight should be simple, intuitive and efficient for the consumers and producers

Key: ***Imperative*** ***Required*** ***Advisable***


Data Stories :
Success or
Failure?

Data: Importance vs Usage

- 90% of the world's data comes from the past few years.
- There two types of data:
  - Quantitative Data.
  - Qualitative Data.

- Data telling you WHAT it is happening (or NOT), usually numerical data.
- Data telling you WHY it is happening.
- Good data scientist/analysist/engineer/etc need to be able to see the data, recognize its importance and when the data should be used.
- Not only about the number. EMPATHY.
- The important saying.

Data: Importance vs Usage

With great power
comes great responsibility.

- Batman

Understand how to use data – examples:

Netflix

- Gathered lots of data by looking into the habits of subscribers.
- Studied play/pause patterns, genre viewing patterns, weekday vs. weekend viewing patterns, time viewed, devices used, etc.
- Also studied piracy market.

How they used Data and Analytics?

- Built recommender systems based preferences.

- Used analytics to study the market:
  - Most watched director + collaboration.
  - Most watched actor + genre.
  - Most watched actor + director.

Understand how to use data– examples:

Amazon

- Maximizing business benefits by improving quality, preventing theft and gaining 360-degree customer profiles
- Massive amounts of customers data available to Amazon.
- They became one of the earliest adopters and through the use of Big Data, went on to transform the way they did business.

- **How are they using (Big) Data and Analytics?**
  - Patenting the shipping of goods
  - Improving customer care quality
  - 360-degree customer profiles
  - Preventing theft
  - Amazon Webstore

Mercedes-AMG

- Acquiring real-time data and new analytics techniques to automate and increase productivity
- Realized that a core function of engineering could benefit tremendously from Big Data and Analytics.
- The company began to reinvent their key business processes.

- **How are they using Data and Analytics?**
  - Identifying when to stop a testing procedure to save significant time and money.
  - Gaining real-time insights, improving quality and increasing productivity with Vmax platform.

What happened here?– examples:

Target

- Assigns every customer a Guest ID number, tied to their credit card, name, or email address that becomes a bucket.
- Stores a history of everything they've bought and any demographic information Target has collected from them or bought from other sources.
- Ran test after test, analyzing the data, and before long some useful patterns emerged.

How they used Data and Analytics (wrongly)?

- Studied the purchase pattern of people.
- E.g. unscented lotions, blue rug, calcium-magnesium-zinc vitamins, a large purse. Meaning?
- Why wrong? What is wrong?

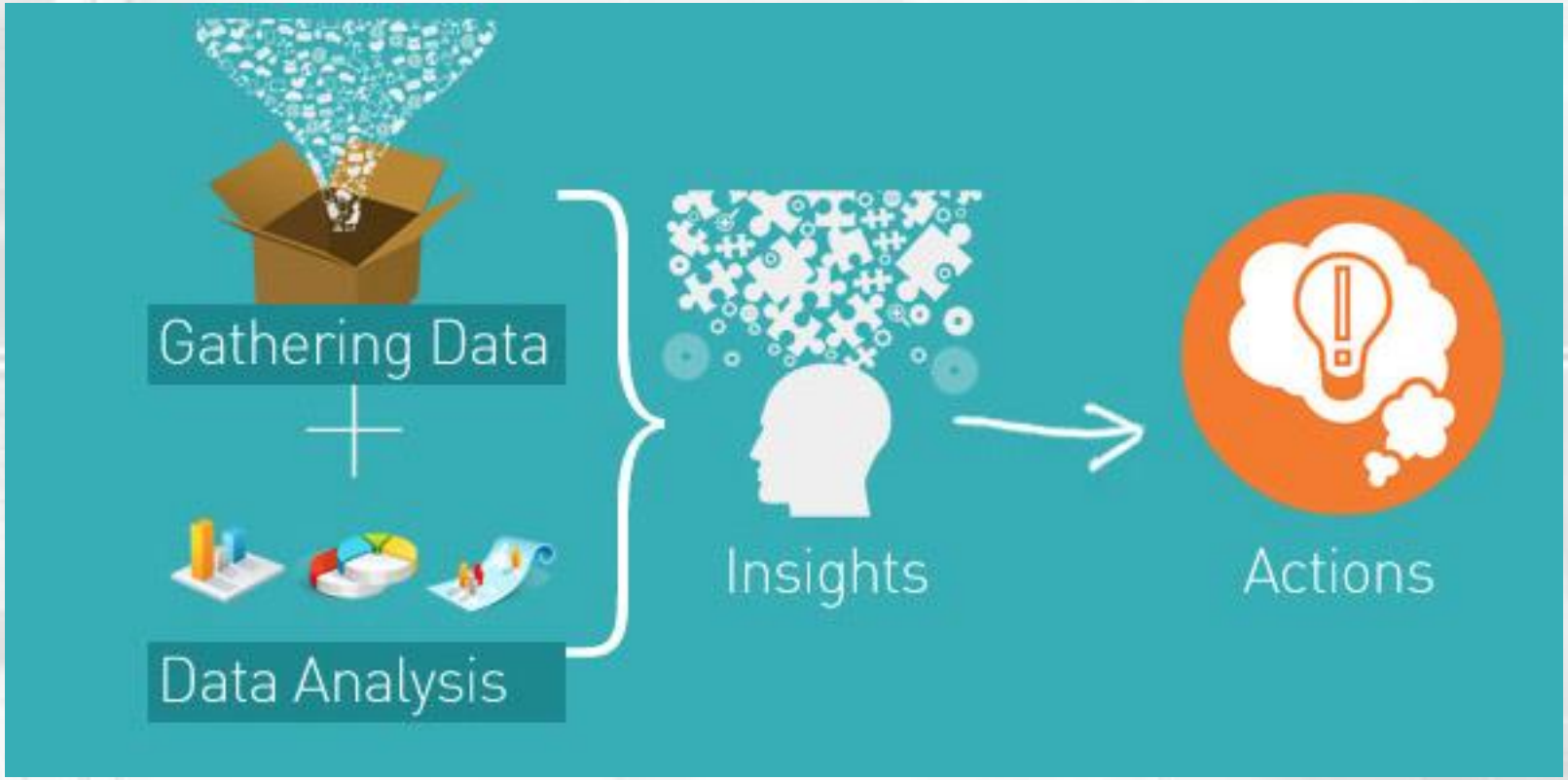
Pinterest

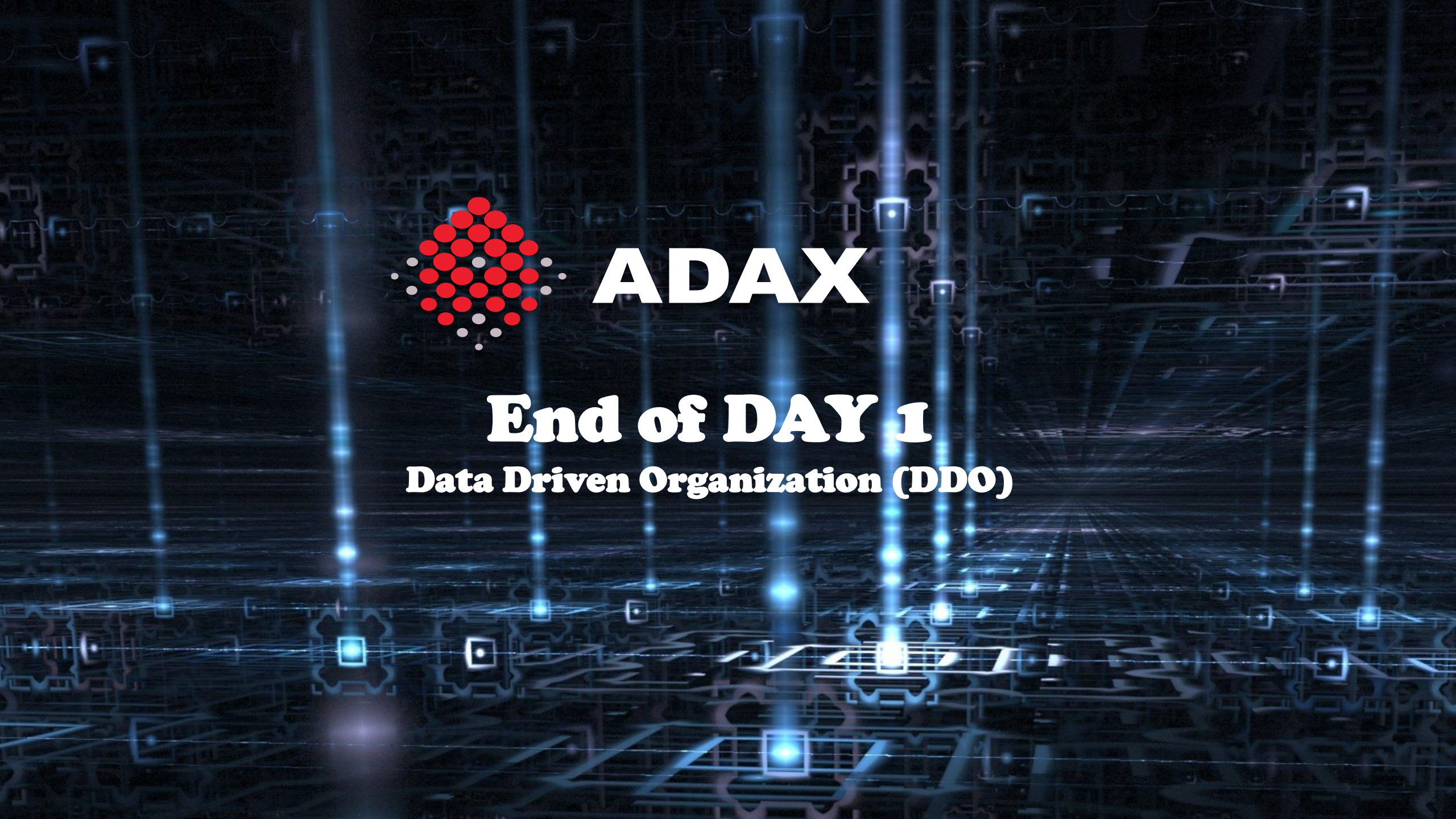
- A website specializes of pictures, images and photos.
- Captures the data after users click into the website to browse pictures, photos, etc.

How are they using Data and Analytics?

- Identifying what is most browsed or looked at image/photo/picture.
- Data is used for content and advertising.
- What could go wrong?
- Pinterest accidentally congratulates single women on their weddings.

Conclusion




ADAX
End of DAY 1
Data Driven Organization (DDO)

Key function of data scientist?

Data scientists are big data wranglers. They take an enormous mass of messy data points (unstructured and structured) and use their formidable skills in math, statistics and programming to clean, massage and organize them. Then they apply all their analytic powers – industry knowledge, contextual understanding, skepticism of existing assumptions – to uncover hidden solutions to business challenges.

Data Scientist Responsibilities

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

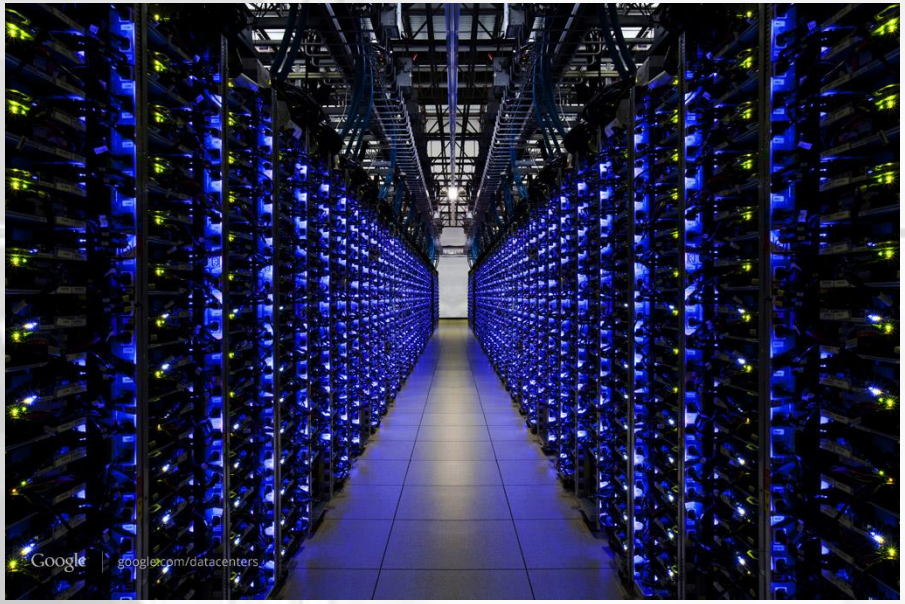
On any given day, a data scientist may be required to:

- Conduct undirected research and frame open-ended industry questions
- Extract huge volumes of data from multiple internal and external sources
- Employ sophisticated analytics programs, machine learning and statistical methods to prepare data for use in predictive and prescriptive modeling
- Thoroughly clean and prune data to discard irrelevant information
- Explore and examine data from a variety of angles to determine hidden weaknesses, trends and/or opportunities
- Devise data-driven solutions to the most pressing challenges
- Invent new algorithms to solve problems and build new tools to automate

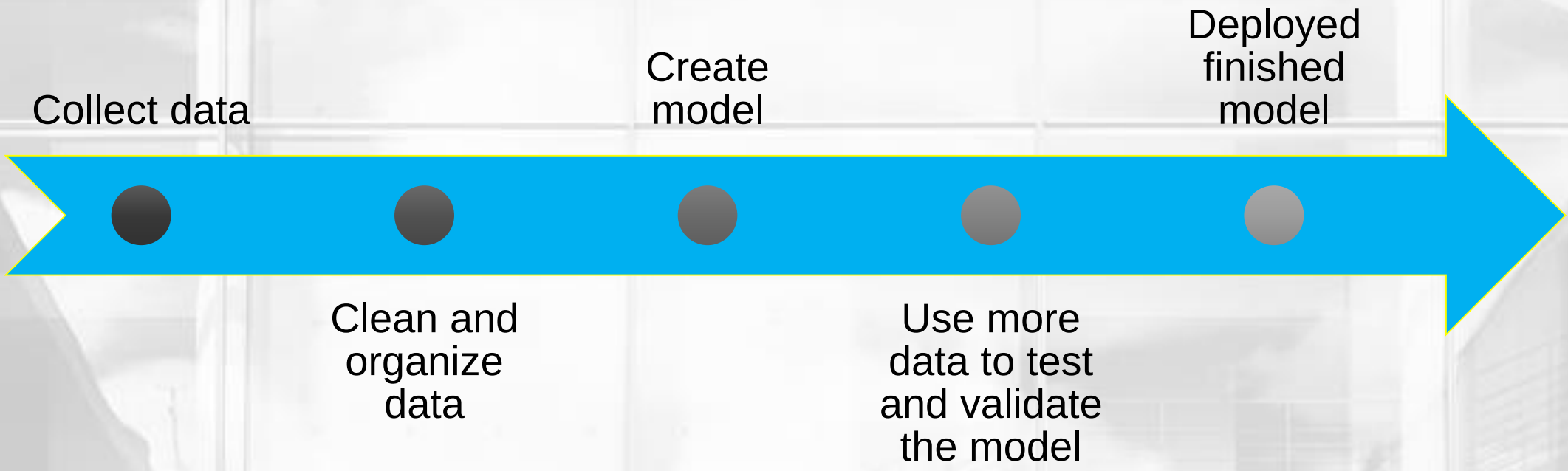
Responsibilities

- The janitor, the cleaner, the analyst and the reporter.

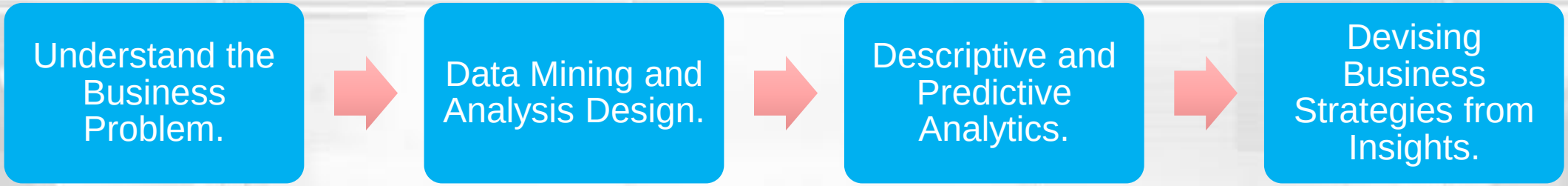
It is not…



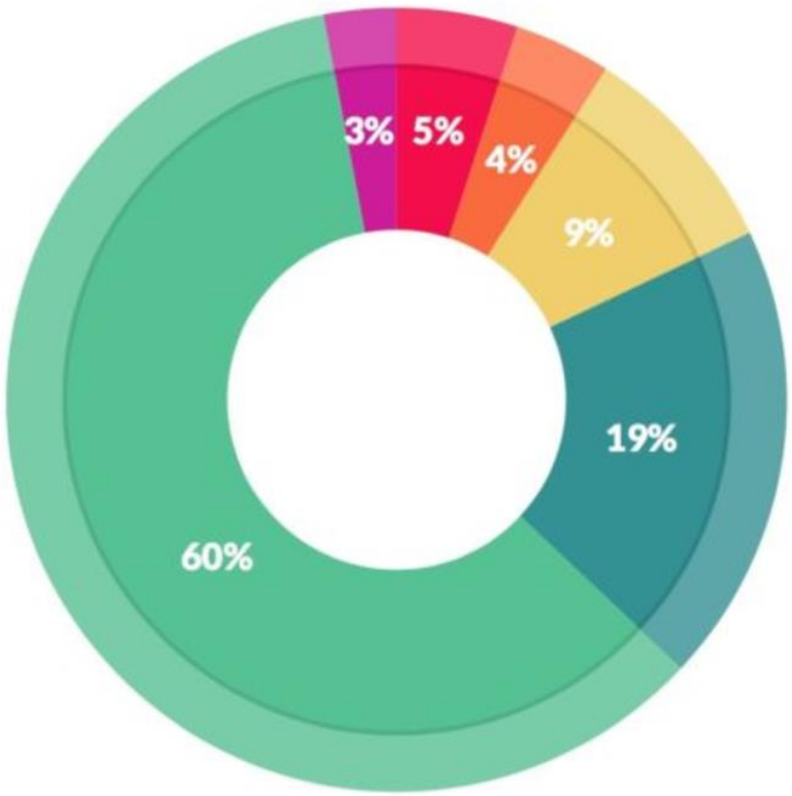
The real job…


Collect data
Clean and organize data
Create model
Use more data to test and validate the model
Deployed finished model

Key functions



Type of Work

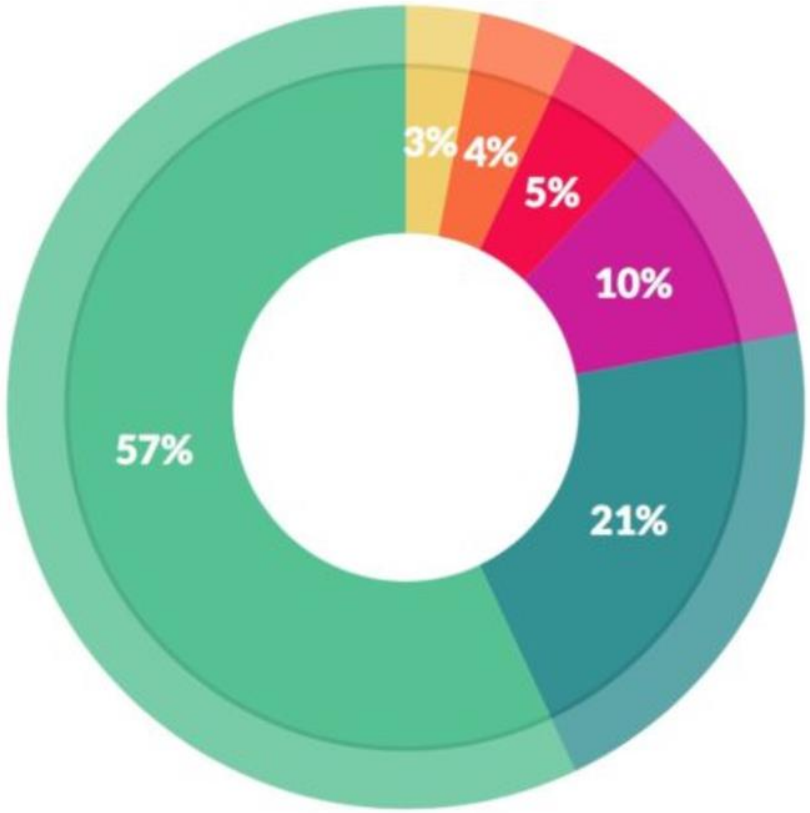


What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#719196f63

More work



What's the least enjoyable part of data science?

- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

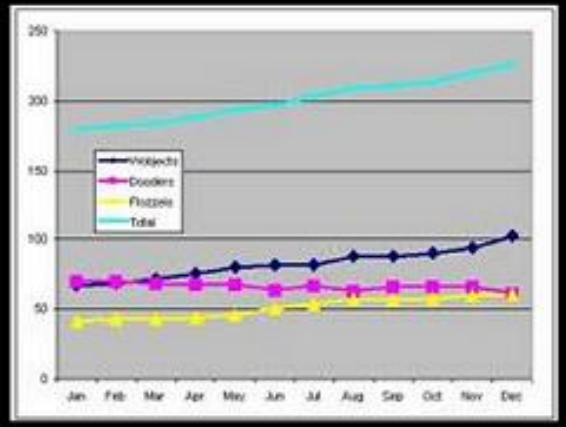
https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#f1396f63

Data Scientist

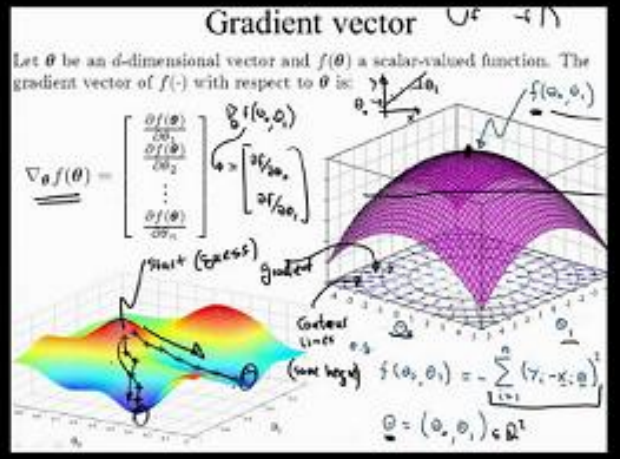
What my friends think I do

What my mom thinks I do

What society thinks I do



What my boss thinks I do



What I think I do

What I actually do

Aptitudes of a data scientist?

The skills.

CREATIVITY

- Common Sense.
- Curious Mind.
- Clear and Simplified Thoughts.

COMMUNICATION

- Good Listening Skills.
- Excellent Communication Skills.
- Visualization Capabilities.

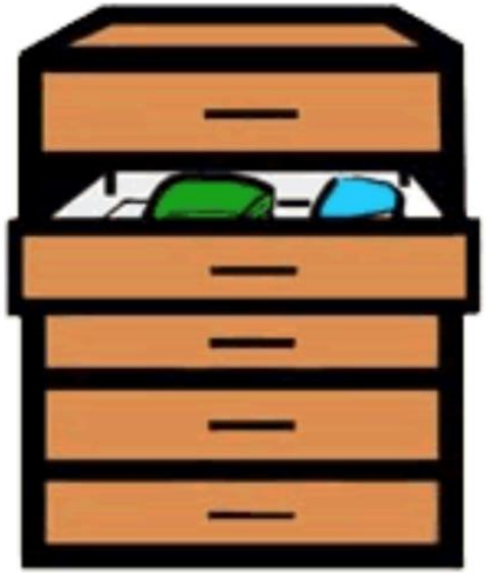
ANALYTICS

- Mathematics and Statistics.
- Machine Learning.

INSIGHTS

- Puzzle Solving.
- Pattern Recognition.
- Business Insights.

Let’s try your skills.



I have **4 red**, **18 black** and **8 brown** socks in my sock drawer. If it is completely dark and I cannot see the colour of the socks that I am picking, how many socks do I need to take from the drawer to be sure that I have at least one pair of socks that are the same colour?

What is the number under the car?

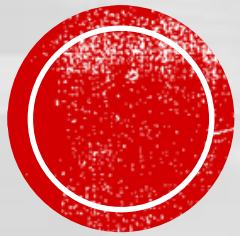
Birthday exercise

- In a math class, there is a a random group of 23 people, what is the chance of that two people have the same birthday (same day and month, e.g. 1st June)?

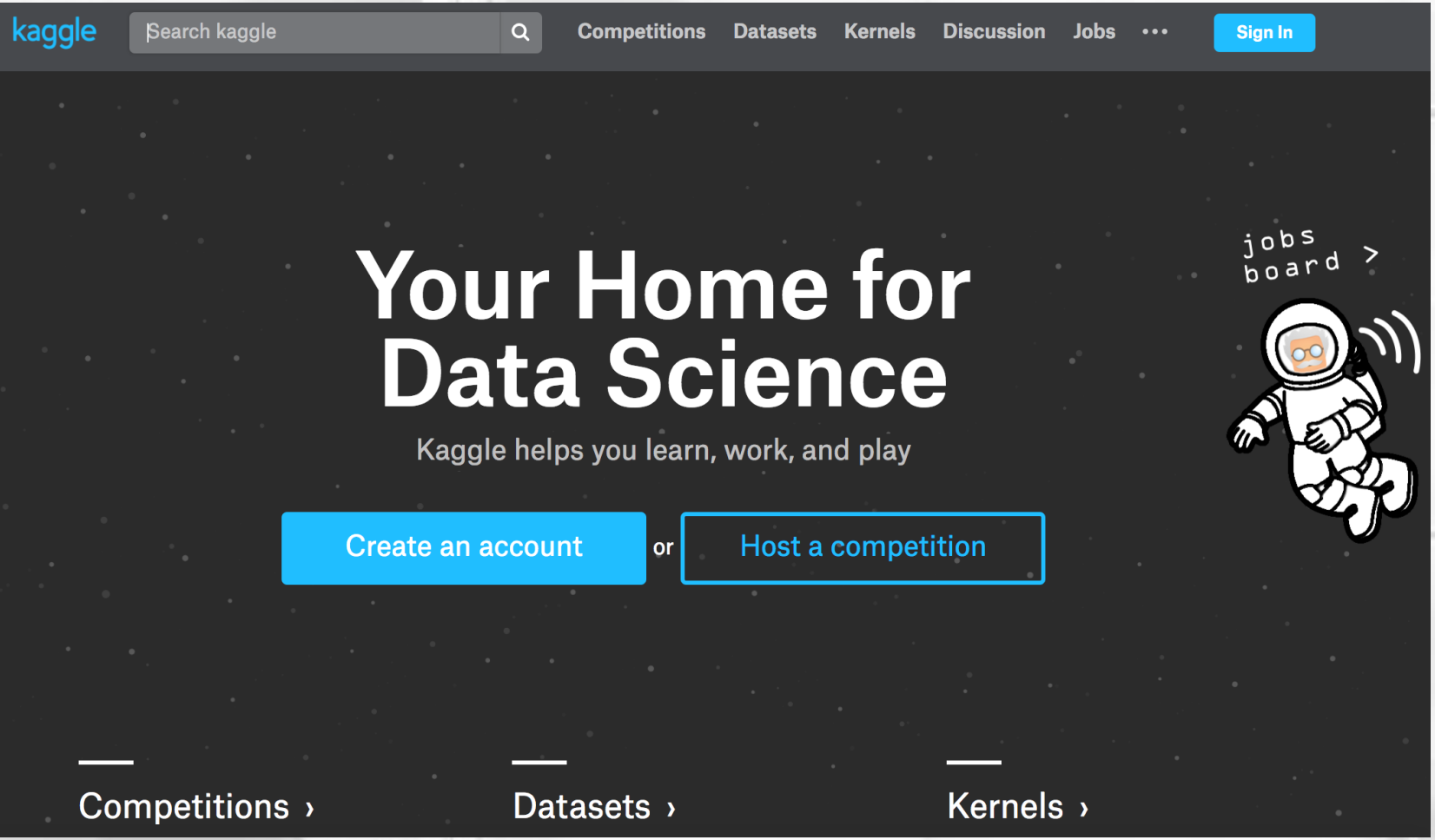
50%

- How about a random group of 50 people or more?

97%



Where and what are we doing?

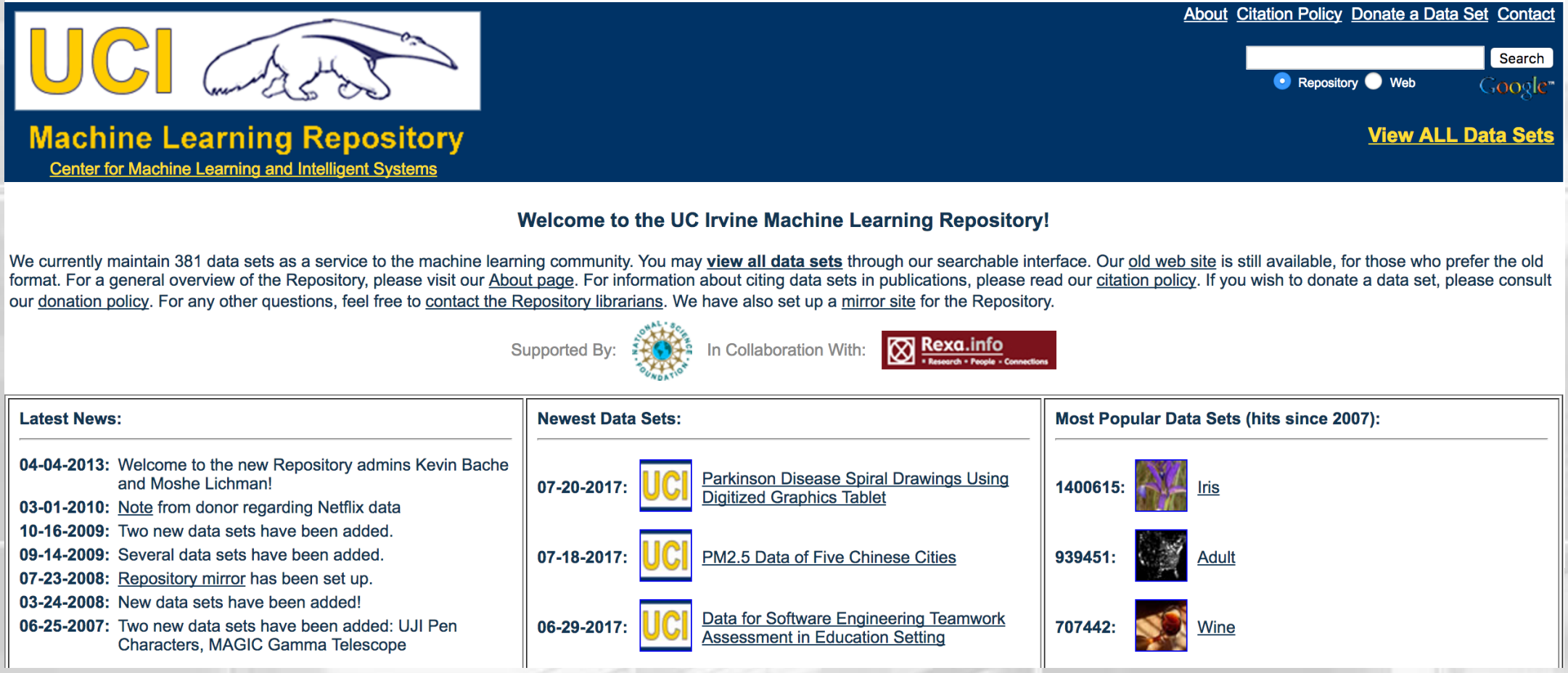


kaggle

- Data sets, real problems, in unprocessed manner.
- Recommend to go through past competitions.
- Read through the forums with particular competitions to find out useful discussion and tips/hints that will be useful for solving future problems.
- https://www.kaggle.com/

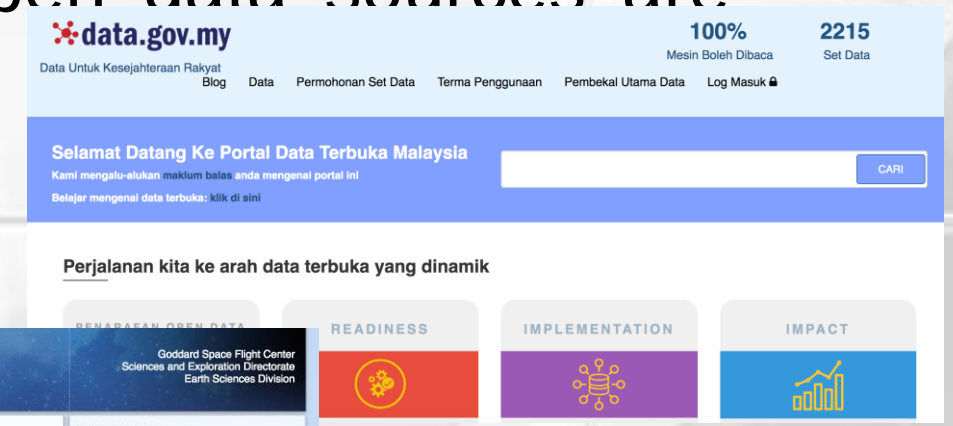
UC Irvine Machine Learning Repository

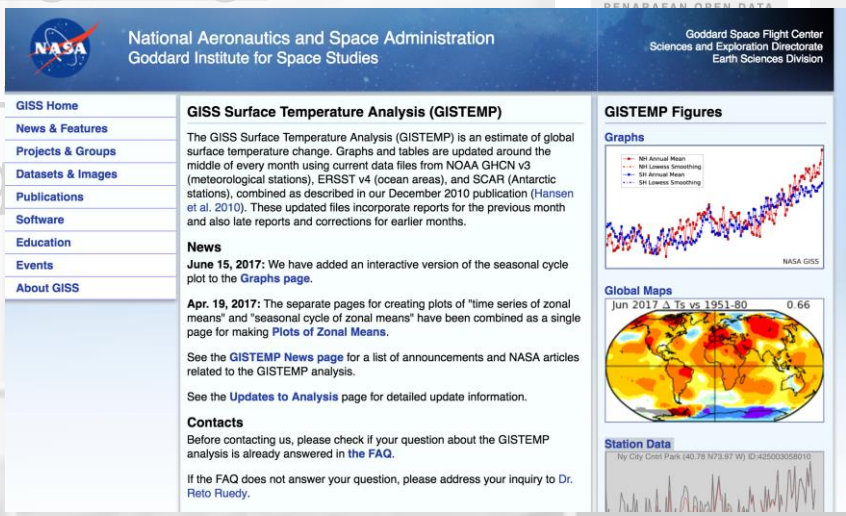
- 360 data sets as a service to the machine learning community
  http://archive.ics.uci.edu/ml/



Open data

- Where else can we get the data? Open data sources are available in several sites.
- Malaysia - http://www.data.gov.my/
- Singapore - https://data.gov.sg/
- Nasa - https://data.giss





MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Databases SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query

Welcome to Enterprise Data Science.
Your 42 Days journeys starts now.

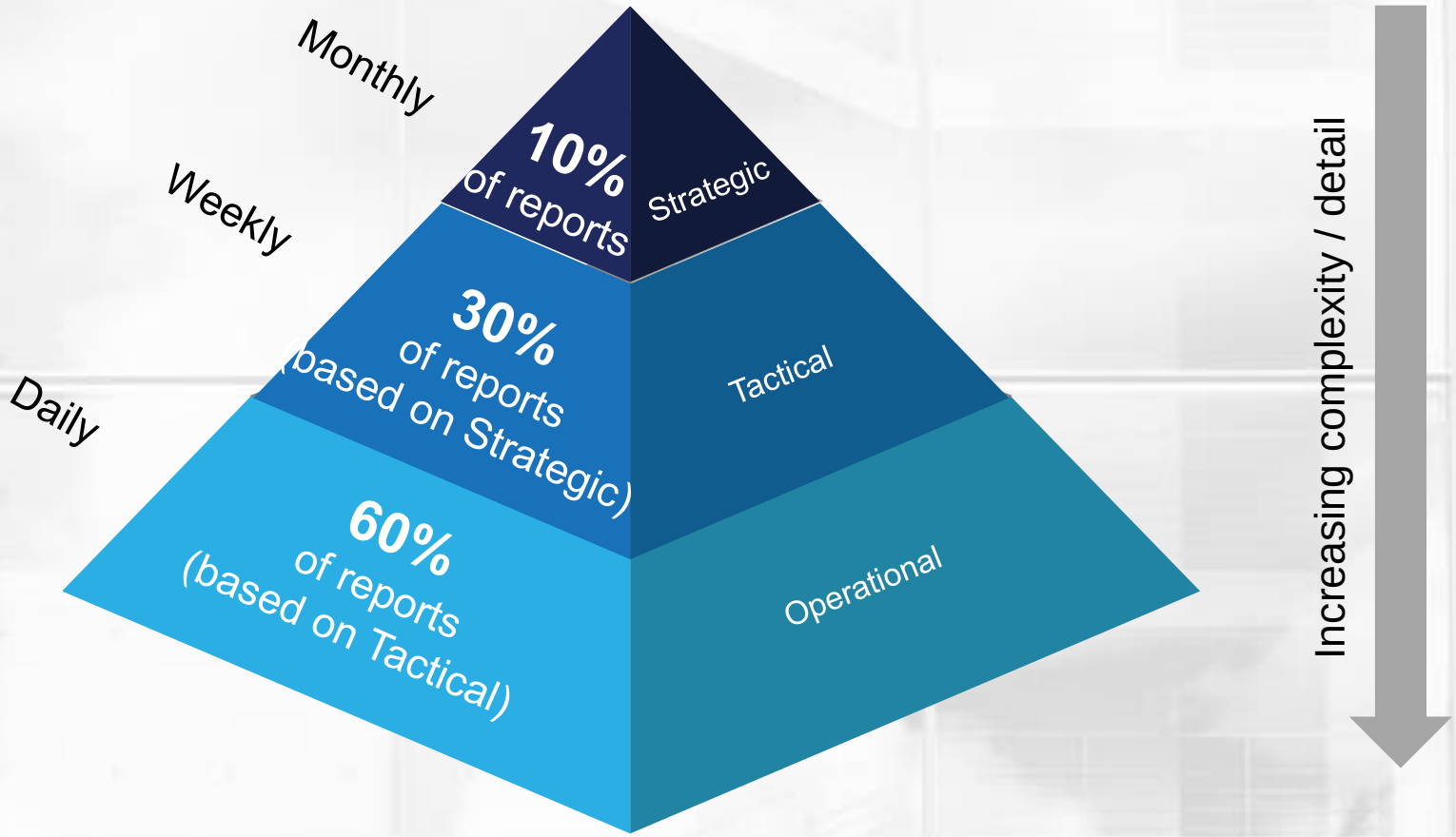
DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

KPIs are a mechanism to measure business success on the operational, tactical and strategic level



EXTRAS - examples:

Gatwick Airport

- Using biometric data and predictive technology to improve operational efficiency and customer experience.
- Gatwick Airport has moved into the predictive technology domain and has begun using Big Data for both operational efficiency and the improvement of customer experience.

How are they using Big Data and Analytics?

- Airport customer traffic predictions and crisis prevention enabled by machine-generated data.
- Improving passenger experience using social media analysis.
- Using biometric data to track passenger movements across the airport
- Third party collaboration.

How are they using Data and Analytics?