

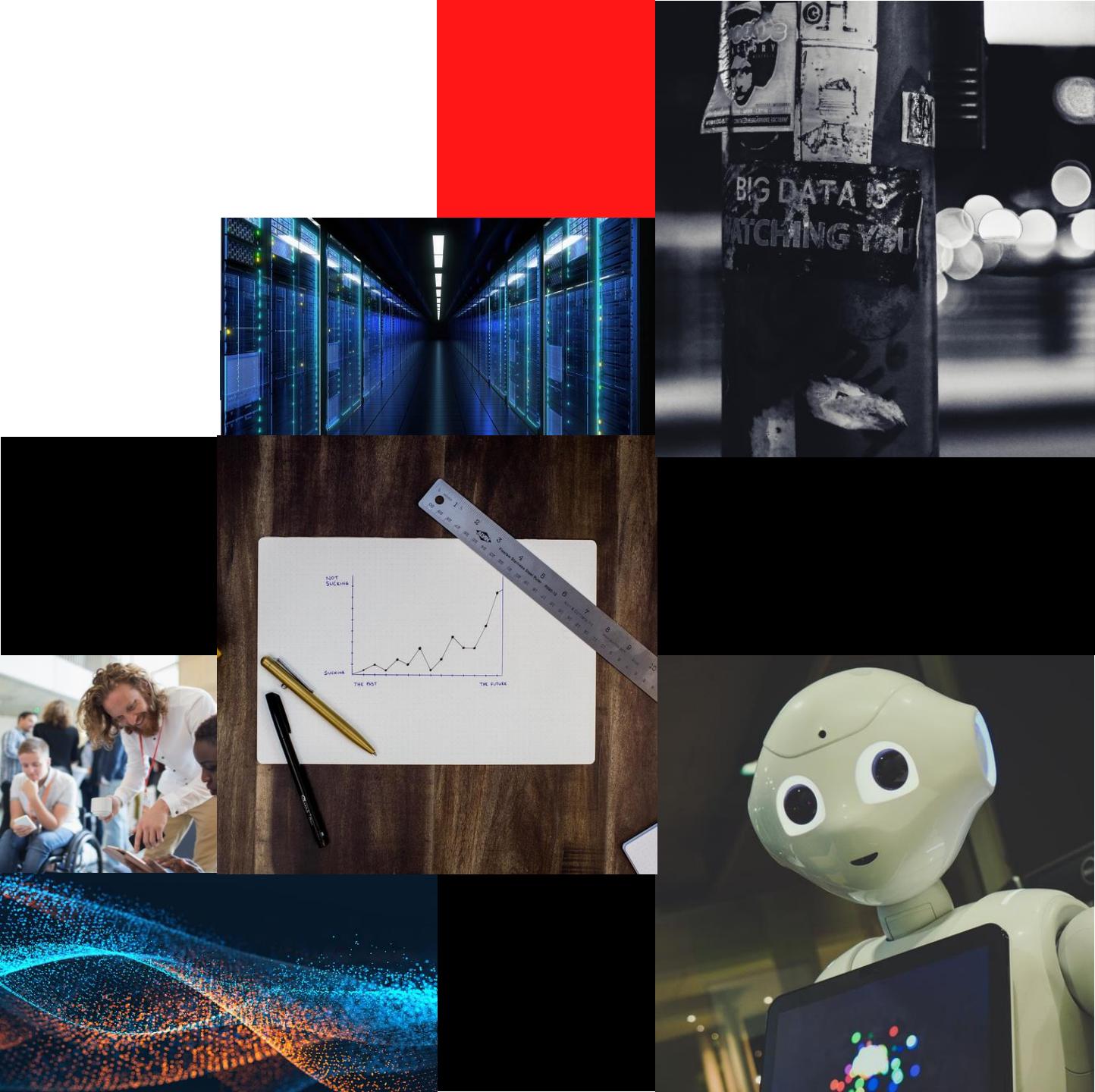
Big Data

Presentation by Maheshkumar Paik



Maltem
Consulting Group

Strictly private and confidential



<https://www.maltem.com>

DATA TO ACTIONABLE INSIGHTS





“In God we trust, all others bring data”

Day 1

Staff Training for Practical Machine Learning

Date: Monday, 27 September 2021

Time: 9:00 am to
05:00am

ITE Campus

	Agenda	Presenter	Time	Duration (mins)
1	Introduction of Myself	Mahesh	9:00 AM 9:15 AM	15
2	Data To Actionable Presentation	Mahesh	9:15 AM 10:15 AM	60
3	Break	Ben	10:15 AM 10:45 AM	30
4	Background on data engineering	Mahesh	10:45 AM 12:15 PM	90
5	Lunch Break	Ben	12:15 PM 1:15 PM	60
6	Lab On SQL	Ben	1:15 PM 2:15 PM	60
7	The process flow of data engineering	Mahesh	2:15 PM 2:45 PM	30
8	Break	Ben	2:45 PM 3:15 PM	30
9	Overview of Data Engineering components	Mahesh	3:15 PM 3:45 PM	30
10	Lab On SQL	Ben	3:45 PM 4:45 PM	60

"In God we trust, all others bring data"

Profile

Maheshkumar Paik

HEAD OF DATA & ANALYTICS

Goals and Ambitions

I am passionate about data and experienced in data my key motivator is all around analysing and searching for insights from data to improve processes and create a competitive advantage for organisations and my client. Leveraging information technology solutions as a competitive differentiator, driving innovation, growth and customer value. This has been achieved for a number of different sectors such as Bank, Insurance, Retail, Telcom, Gov etc.



"In God we trust, all others bring data"

Profile

<https://www.linkedin.com/in/maheshkumarpaik/>



Maheshkumar Paik

Data Strategy | Data Architecture | Data Engineer | Data Science | MLOps | Adjunct Lecture | Public Speaker |

Talks about #data, #mlops, #datacloud, #dataanalytics, and #dataarchitecture

Singapore · [Contact info](#)

8,274 followers · 500+ connections



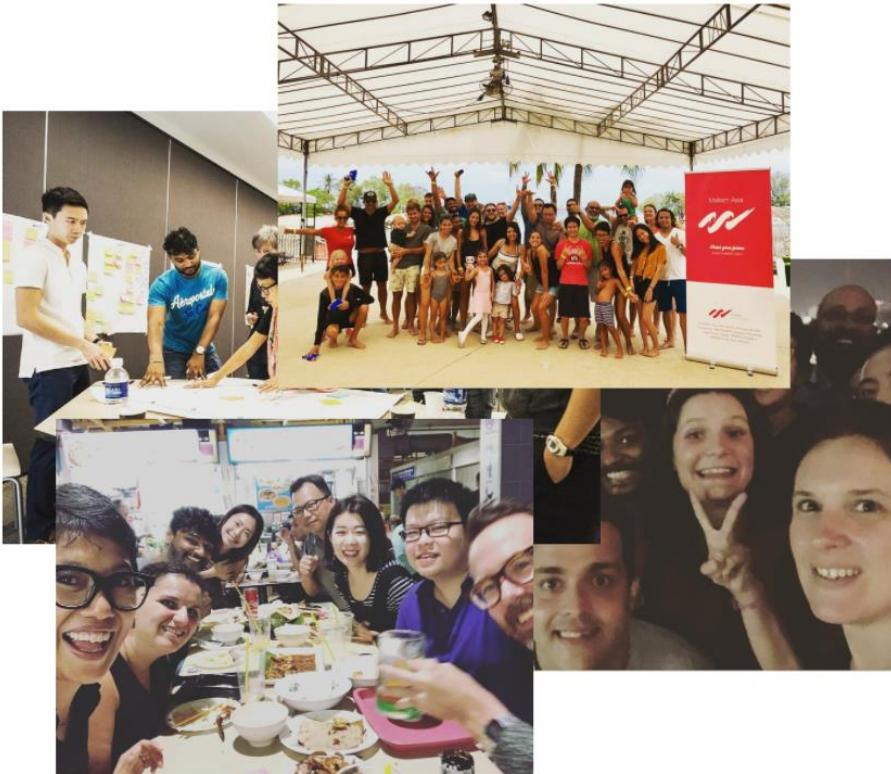
Maltem Consulting Group



SP Jain School of Global
Management - Dubai,...

*"TO BE BETTER
EACH DAY"*

Maltem Asia



Maltem Community

WHY WE ARE AMAZING?

Its rapid growth is representative of its development dynamic. To choose Maltem is to give yourself the choice to collaborate with its entities located on all continents around the globe. Maltem facilitates and encourages its employees by accompanying them in their process of mobility, both professionally and personally, matching them to projects that are fully adapted to their profiles.

*"The price of light
is less than the
cost of darkness."*
- Nielsen

Workshop Objectives

- Why
 - Research
 - Todays Market
- What
 - Understanding the Data Strategy
 - Companies Competing
 - Analytics
 - Modelling Concepts
- How
 - Python Notebook
- Real world example
 - Descriptive / Predictive analysis Actionable report

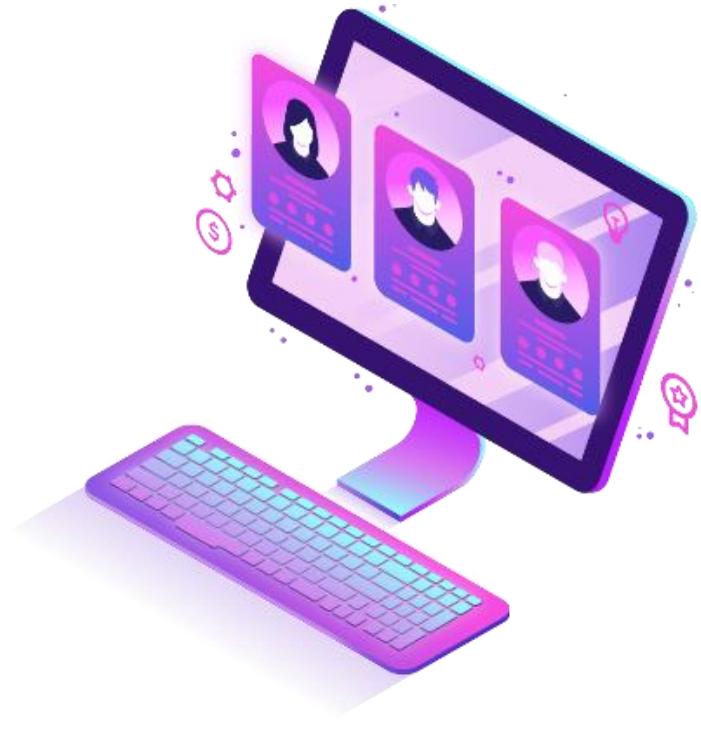


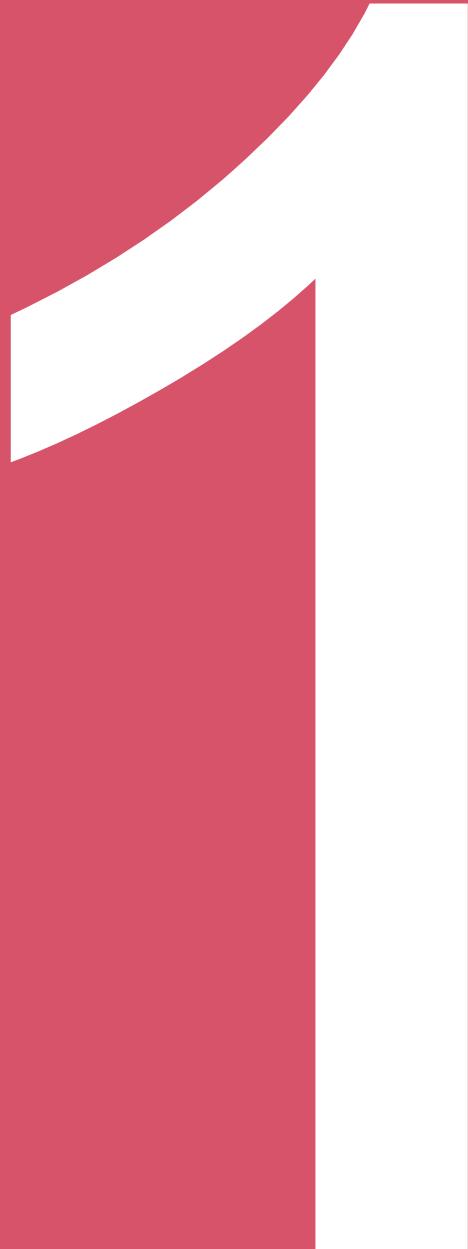
m

"You can have data without information, but you cannot have information without data"

Ice Breaker

- What is your background?
- What you expect from Workshop?
- What does the term **Data Engineering** mean to you?
- Do you use data for day-to-day work





Why
Research
Todays Market

"Without a systematic way to start and keep data clean, bad data will happen"

Data Science

- Data Science is the science of **analysing raw data** in order to make conclusions about that information.
- Many of the techniques and processes of data analytics have been **automated into mechanical processes and algorithms** that work over raw data for human consumption.
- Data analytics techniques can **reveal trends and metrics** that would otherwise be lost in the mass of information.
- This information can then be used to optimize processes to increase **the overall efficiency** of a business or system.

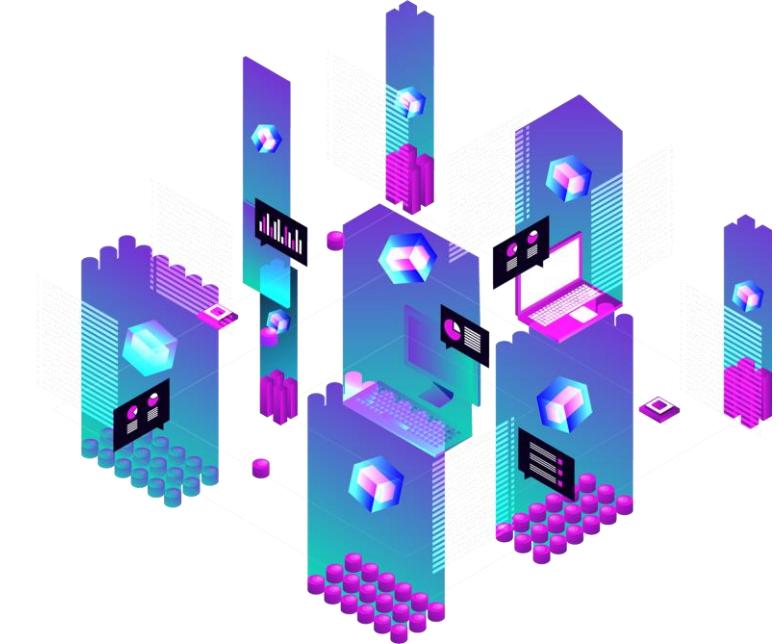


22

*"Torture the data,
and it will confess
to anything"*

Why Data ?

How big do you think
the data economy is?



m

*“Torture the data,
and it will confess
to anything”*

Why Data ?

ASEAN CEOs believe that Data and analytics are critical to remaining competitive and relevant in the digital future.

Business value of data and analytics

A critical challenge in our increasingly digitised and connected society is the **proliferation of data**. Every organisation, regardless of size, is bursting with information. The real power of data, however, comes from the informed decisions it enables people to make, and investing in **technology and data and analytics to make decisions better, faster and cheaper** is the next competitive advantage organisations seek.

Research shows that 72% of ASEAN CEOs agree that artificial intelligence will have a larger impact on the world than the internet revolution.

Opportunities from analytics



Identify bottlenecks in processes



Predict customer needs



Share information while minimising security threats

Data and Analytics in Business

By using data and analytics it would be possible to predict future outcomes and automate non routine tasks

- Opportunities for the profession to deliver greater value and to help businesses transform their decision-making.
- Business can enhance their understanding of their operations, customers, risks and markets
- Applying analytics to gain new insights and interrogate entire data sets



What does it really mean to be data-driven?

Organisations tend to focus their efforts on analytics technology only. We believe there is so much more to it:

- ✓ **Business decisions.** Analytics is used in the organisation's priority business decisions.
- ✓ **Data and information.** Both internal and external data sources are leveraged and managed to ensure quality and security.
- ✓ **Tech and infrastructure.** Flexible, scalable platforms (on-premise or on cloud) that enable end-to-end analytics
- ✓ **Organisation and governance.** Operating model with clear roles and responsibilities, and addresses internal and external compliance
- ✓ **Process and integration.** Agility, with a focus on value creation and streamlined interaction between analytics teams and business users
- ✓ **Culture and talent.** Analytics skills are prioritised in hiring and L&D



"Big data isn't about bits, it's about talent"

\$143b
by 2023

Reuters
(2019)

Growing demand for data management, rising regulatory compliance and regulations, improved adoption rate of massive parallel processing architectures & columnar databases and enhanced streamline business operations are some key factors triggering the data market size.

However, data validation issues and presence of data silos are some factors limiting the market growth.



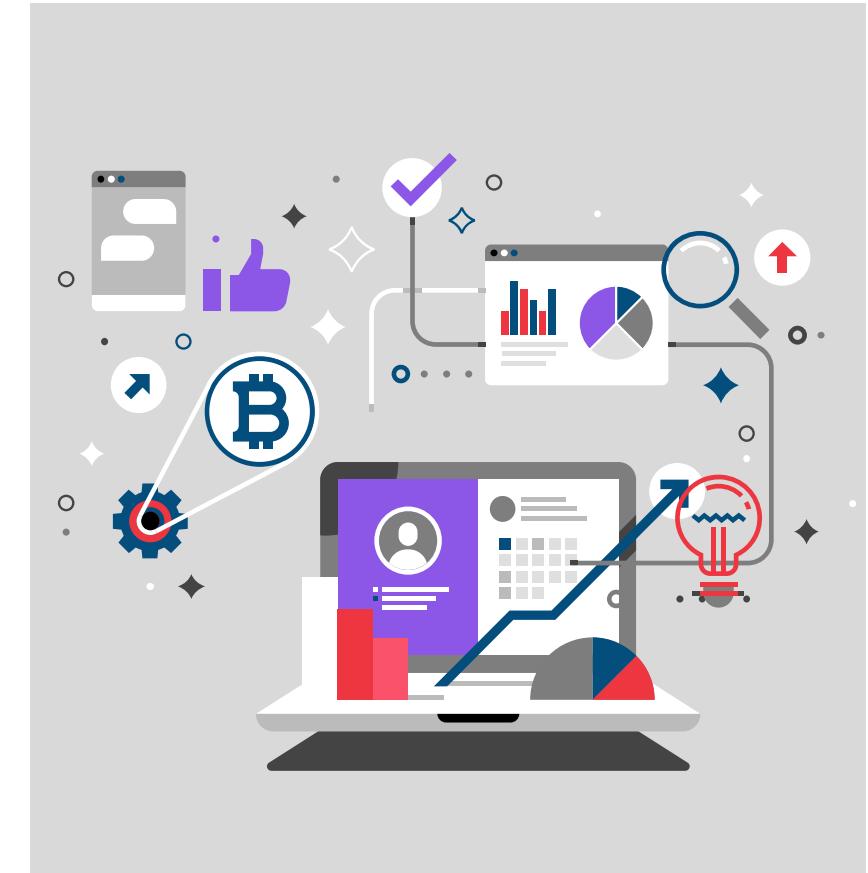
"No great marketing decisions have ever been made on qualitative data"

The data economy is growing at an unprecedented rate, driven by the following factors:

- Improved technology and lowered data storage/processing cost
- Companies' need to generate new revenue
- Companies are seeing a market demand for data
- Data-driven companies ahead of their counterparts
- Disruptions to established industries generated by technology

\$400b
by 2025

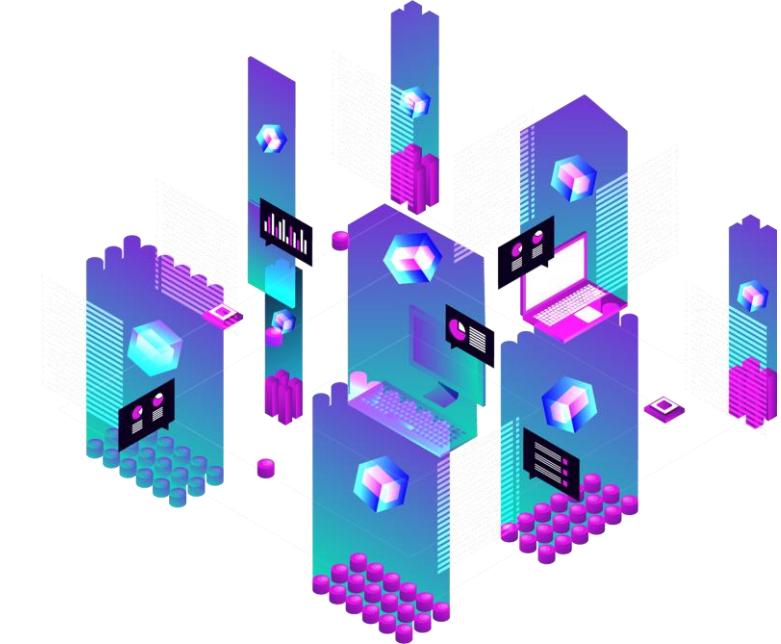
PwC Strategy& (2019)



*"Torture the data,
and it will confess
to anything"*

Why Data Analytics ?

What is the need of Data Science or Data Analytics?





"With data collection, 'the sooner the better' is always the best answer"

Data is transforming the world!!!



CONNECTED WORLD



MORE DATA



DIGITAL DEMOCRACY & SECURITY

INTERNET OF THINGS INDUSTRY 4.0

- Smart Things / Devices
- New Use Cases
- New Architectures
- By 2020: 20.8b devices
- IoT: \$1.7trillion in 2020

DATA AS PRODUCTION FACTOR

- New Analytics
- Machine Learning & AI
- New data sources
- Data Virtualization
- Data Science

RISE OF OPEN SOURCE

- Digital Sharing Economy: Open Data & Algorithms
- Enterprise ready Open Source (e.g. Apache)
- Digital (distributed) Trust (esp. Blockchain)



“Data is the new oil”

TOP INDUSTRY VERTICALS



BANKING



TELCO



HEALTHCARE

USE CASES

- Fraud detection
- Anti-money laundering
- Spend analytics

- Churn analysis
- Customer care
- Network optimization

- Patient care (IoT)
- Genomics research
- Regulatory compliance

CUSTOMERS

- Barclays
- Bank of America
- Citi
- RBS

- Comcast
- British Telecom
- Deutsche Telekom
- Vodafone

- GlaxoSmithKline
- Symphony Health
- Sharp Healthcare
- IQVIA

m

"It is a capital mistake to theorize before one has data"

Job Prospects

LinkedIn search results for "analytics" in Singapore:

- Jobs: 7,430 results
- Sort by: relevance - date
- Location: Singapore
- Page 1 of 4,366 jobs



mycareersfuture.sg search results for "analytics" in Singapore:

- 1,365 jobs found
- Show 270 jobs with government support

"It is a capital mistake to theorize before one has data"

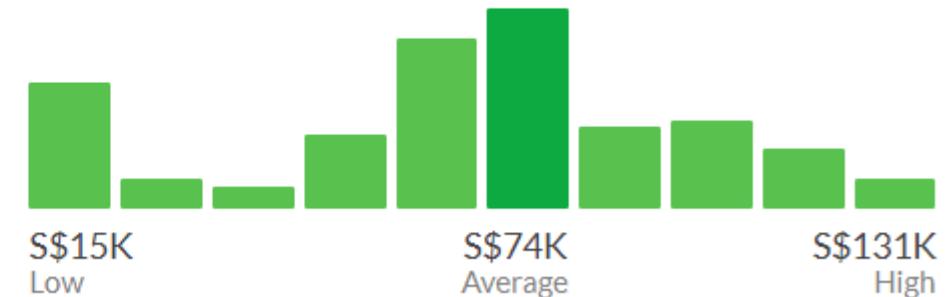
Salary

Data Scientist Salaries in Singapore Area

15 Salaries Updated 30 Jan 2020

Average Base Pay

S\$73,645 /yr



www.glassdoor.sg

Singapore (S\$)

Job Title	Mean	Median	Avg. increase*	Avg. bonus**
Student	26,000	24,000	No change	-
Researcher	74,310	72,000	0-5%	6,687
Data Scientist	82,779	76,400	0-5%	12,365
Senior Data Scientist	128,158	110,500	11-15%	17,261
Chief Data Scientist	170,313	144,000	6-10%	33,550
Machine Learning Engineer	71,818	68,000	0-5%	9,960
Data Engineer	81,715	72,000	6-10%	8,298

www.peoplemattersglobal.com





"Without a systematic way to start and keep data clean, bad data will happen"

Job Profile



Data Engineer / ETL Developer

Expert in SQL and NOSQL, Database and Big Data Knowledge



Data Analyst / Business Intelligence Consultant

Idea of Data and Big Data, Analyze the data business ask for



Data Scientists / Report Designer

Knowledge of Algorithms, High Aptitude and Reporting



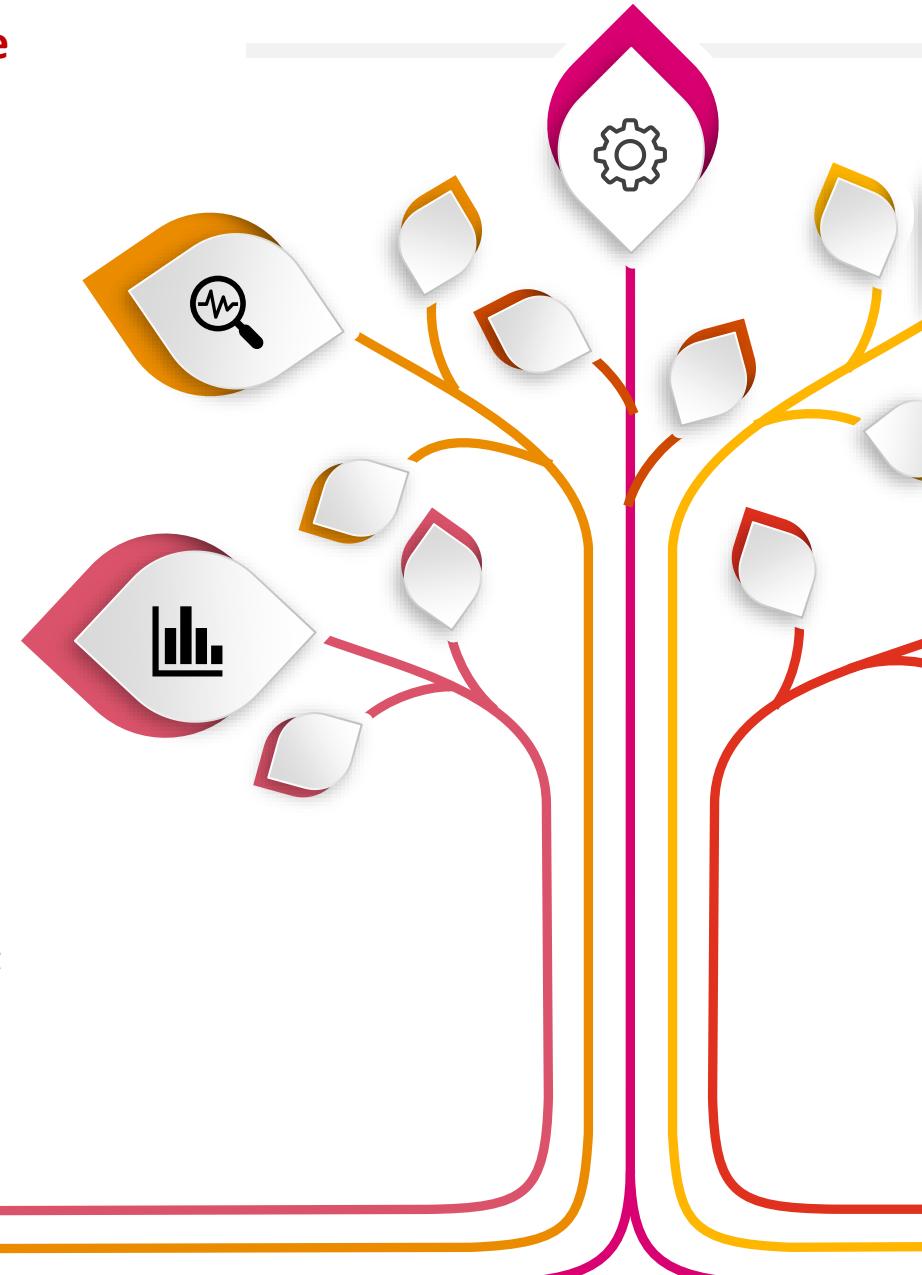
Data Architect / All Rounder

Data Modeling, Data Quality and Data Governance



Data Support and Data Trainer

Support Jobs and Schedule, has a idea of whole data project in and out



*"Torture the data,
and it will confess
to anything"*

Companies that hire?



m

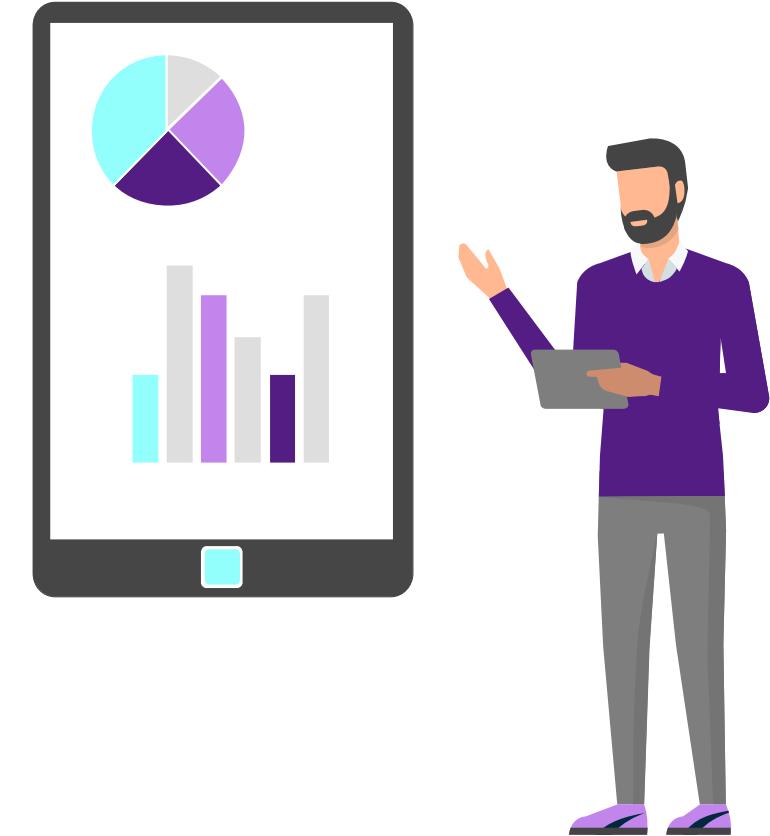


What
Understanding the Data
Strategy
Companies Competing
Analytics
Modelling Concepts

*"Torture the data,
and it will confess
to anything"*

What is Strategy?

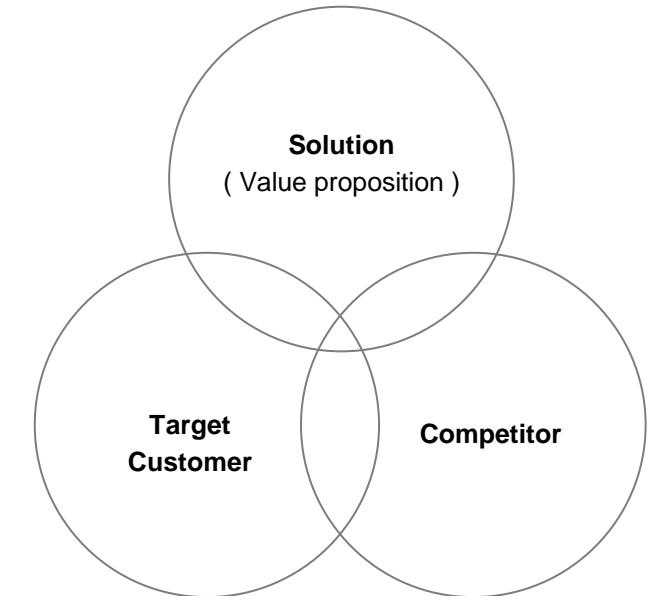
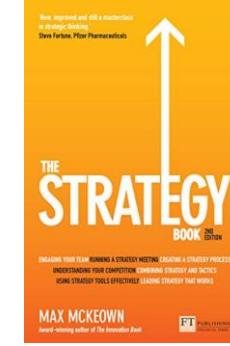
What is strategy for you?



“Only the fittest will survive”

Strategy

Strategy is the **process** of creating a set of **well aligned activities** with the aim of occupying a **valuable position** in a **competitive landscape**.



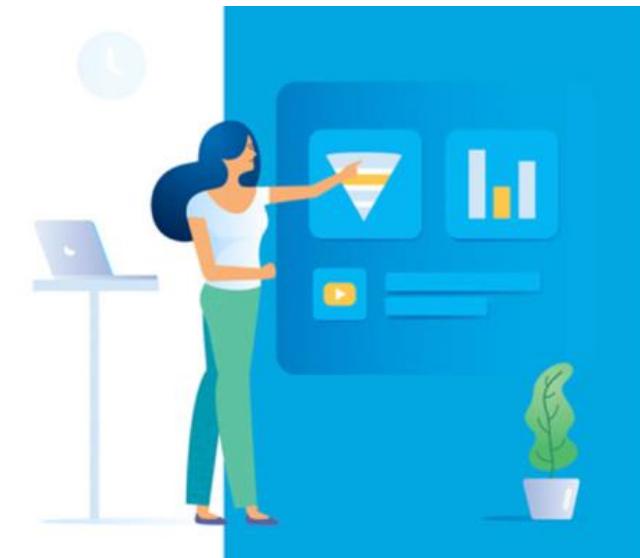
Stratēgos = Master plan for a General



*"TO BE BETTER
EACH DAY"*

Business Competing

- ✓ Low Cost Product
- ✓ Good Product in specification
- ✓ Fast moving product
- ✓ Fantastic Service



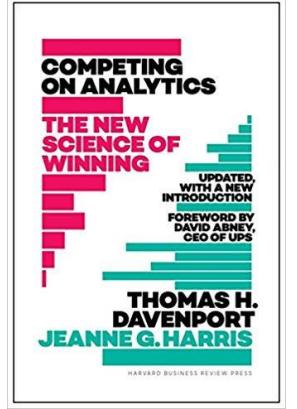
m

"You can have data without information, but you cannot have information without data"

Competing on Analytics

- ✓ last remaining points of differentiation
- ✓ Know what Customer want, what Price they are willing to pay, volume they will buy
- ✓ Actionable decision makers

“
Best decisions: big and small, every day, over and over and over.
”



"In God we trust, all others bring data"

Data Strategy

“

Aligns and prioritizes data and analytics activities with key organizational priorities, goals and objectives.
- Micheline Casey

”



m

"In God we trust, all others bring data"

Data Stories

Robo Advisor are a class of financial adviser that provide financial advice or Investment management online with moderate to minimal human intervention.



Only data-driven investment decisions

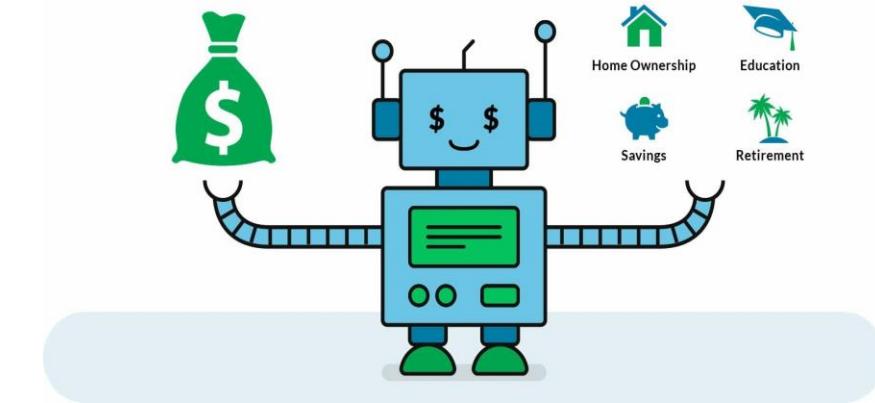


Automated rebalancing



Low annual management fees
0.2% to 0.8%

“
Investing strategy is
Economic Regime-based
Asset Allocation
”



*"The price of light
is less than the
cost of darkness."*
- Nielsen

Data Stories

Kyra Hair Analytics



KYRA hair questionnaire and schedule your hair sample collection, gathering information



KYRA shampoo and conditioner will be formulated based your hair's specific needs and your hair questionnaire



Receive your Kyra haircare package and wave goodbye to your hair problems



“

www.kyrabeauty.com

”

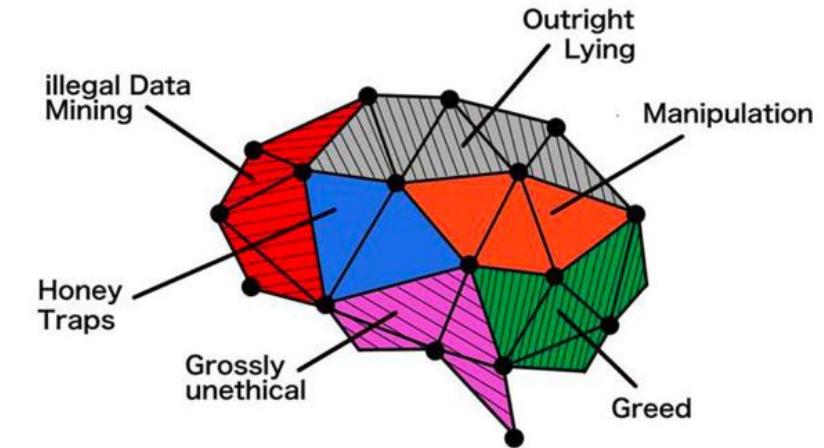
“

"Data is the new oil"

Data Stories

Cambridge Analytica

- Data Mining and Computation
- Machine learning algorithm to predict the personality type
- Focused Campaign
- Influence to change the vote



Cambridge Analytica

“ “

Alexander Nix said he was not doing anything wrong in court and US Congress and Court too couldn't prove anything wrong

F

©2018
FORBES
.COM/
CARTOONS

forbes.com/cartoons

m

” ”



"No great marketing decisions have ever been made on qualitative data"

Data Analytics



Information that exists in a variety of formats and sizes



Name: Gus the Cat

Age: 4

Parents: Adam Webb

City: Toronto

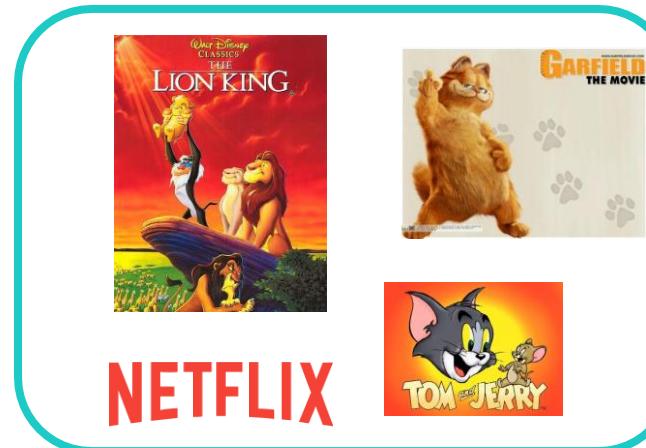
Phone: 416-555-MEOW

Twitter: @guscat13

Hobby: Nap, Nip and **NETFLIX**



Process of examining data to draw conclusions about that information.



←If you likethis
you will also like this →



"Big data isn't about bits, it's about talent"

Analytics Value

Broadly speaking there are three key business values that can be realised by enabling better access to relevant data and providing fit-for-purpose data through effective data analytics.

Drive Revenue



Bringing together data points that are currently siloed within your systems and **unlocking the value** through **analytics** and **data monetisation**

Reduce Costs



Intelligently managing business-critical data in integrated IT systems to **lower operating costs** and **minimise inefficiencies**

Manage Risks

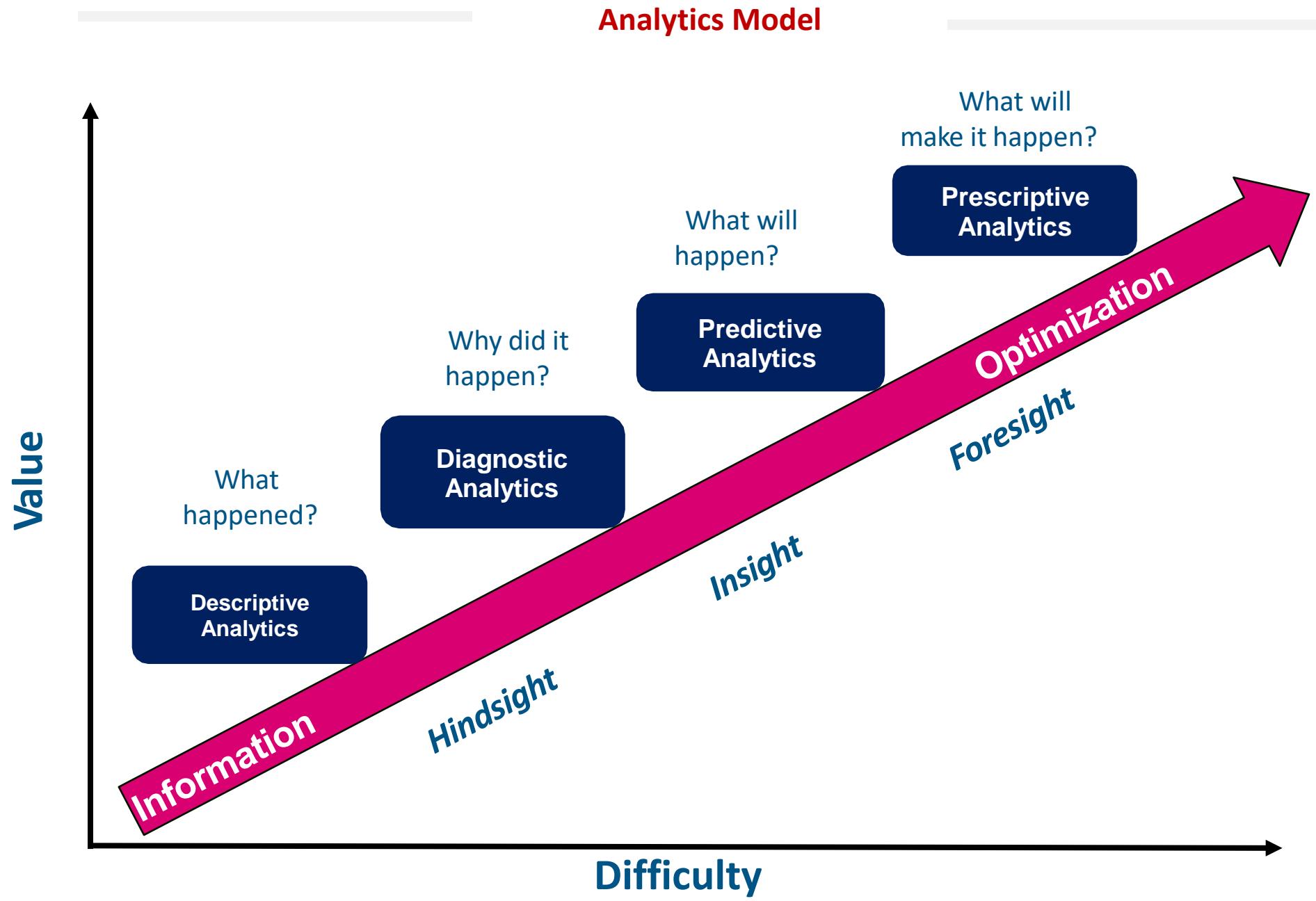


Initiatives across **data discovery**, **data quality**, and **data processing** critical to comply with regulations





"It is a capital mistake to theorize before one has data"



"Without a systematic way to start and keep data clean, bad data will happen"

Predictive Analytics

- Predictive analytics is used to make predictions about **unknown future events**.
- Predictive analytics uses many techniques from **data mining, statistics, modeling, machine learning, and artificial intelligence** to analyze current data to make predictions about future.
- Predictive modeling is a process used in predictive analytics to create a **statistical model** of future behavior.
- Predictive analytics is the area of **data mining** concerned with **forecasting probabilities** and trends.

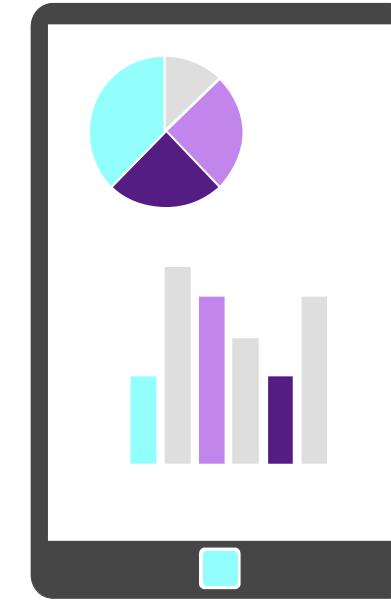
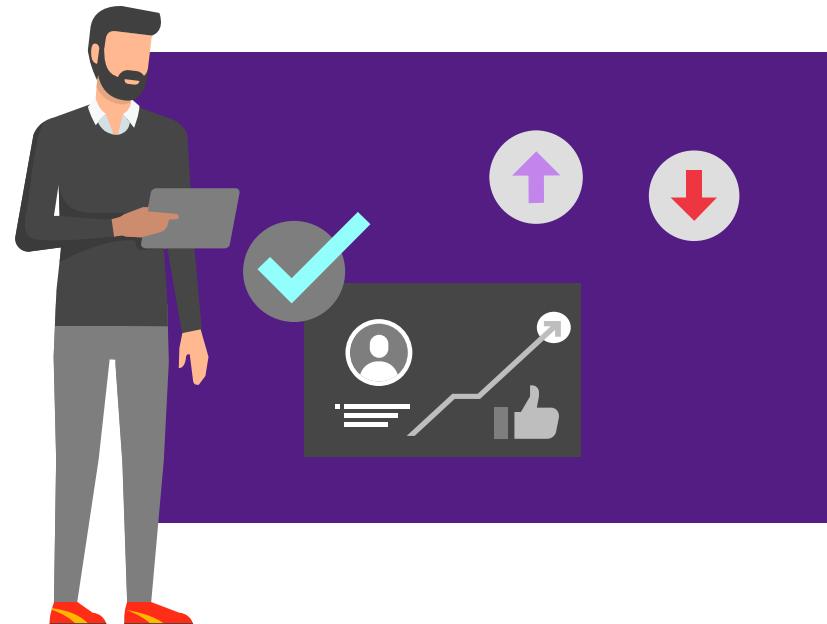


22

*"Torture the data,
and it will confess
to anything"*

Breakout

1. Take one Company
2. Build its data strategy
3. List its data analytics process
4. What is it going to achieve





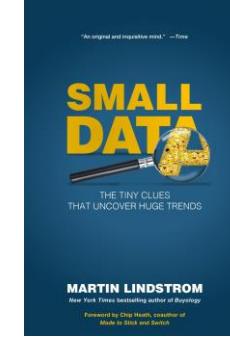
How
Use case of Silver
Oak
Use Case of
Advertising Company
Python Notebook

"Only the fittest will survive"

History Repeats

I believe that the future is only the past again, just entered through another gate!

-Arthur Wing Pinero



m



"You can have data without information, but you cannot have information without data"

Liner regression

Business Plan of Silver Oak

A unique food outlet which has a unique plan of having only home delivery of food

Partners

In the first fortnight of operations Silver Oak used the services of three of the courier companies -

- Alembic Associates [A],
- Barry's Buffets [B]
- Careful Couriers [C].

Build a new Data Strategy

With help of the data Silver Oak wants to

- Increase the revenue
- Improve the operational efficiency
- Manage the risk



SILVER OAK

"Without a systematic way to start and keep data clean, bad data will happen"

Data Project Process

1. Define the problem.
2. Obtain the data.
3. Explore the data.
4. Model the data.
5. Evaluate the model.
6. Answer the problem.

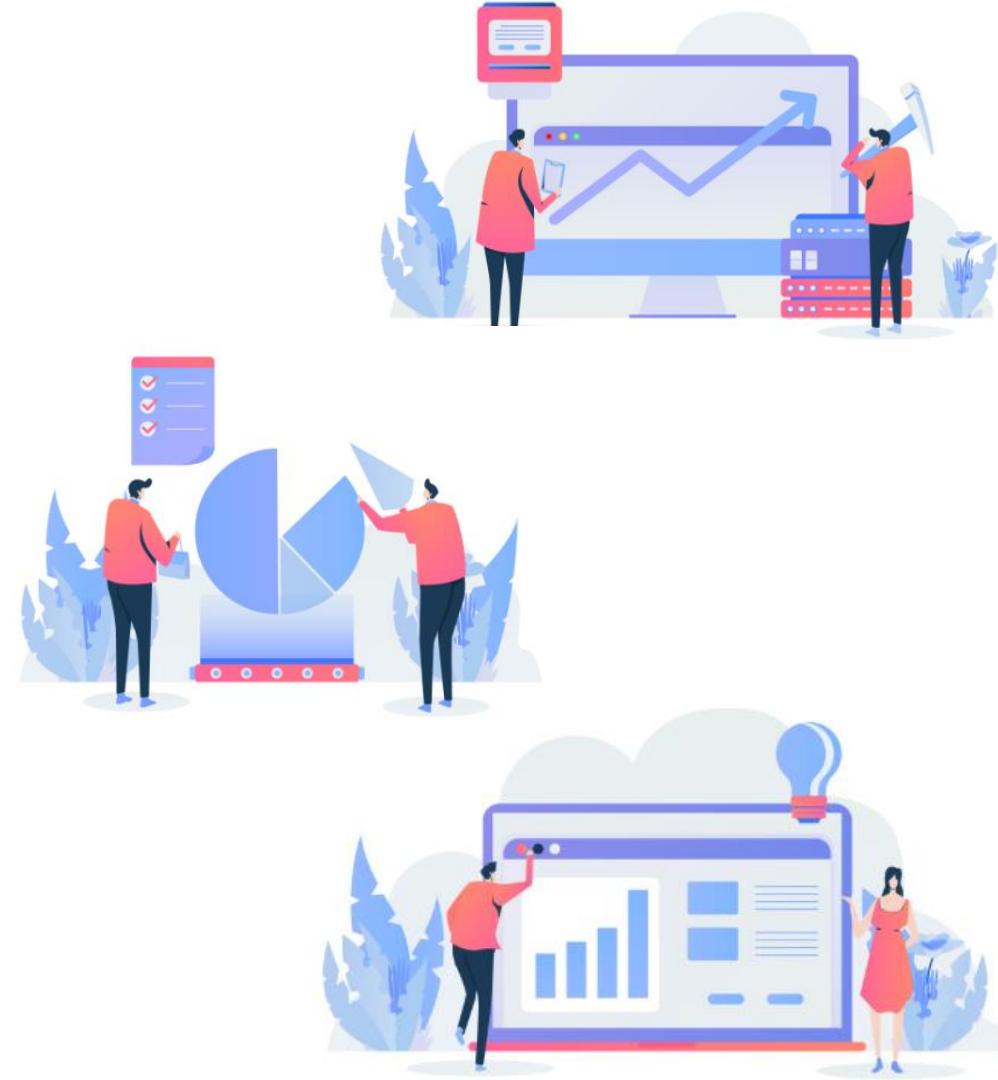


“Data is the new oil”

Predictive Models

Ideally, these techniques are widely used:

- Linear regression
- Multiple regression
- Logistic regression
- Regression with regularization
- Neural networks
- Support vector machines
- Naive Bayes models
- K-nearest-neighbors classification
- Decision trees
- Ensembles of trees





"You can have data without information, but you cannot have information without data"

Reporting Steps

1. **Identifying** the key data assertions associated with business policies
2. **Determining** how those data assertions relate to quantifiable business impact
3. **Evaluating** how the identified data flaws are categorized in Dimension and Measure within a set of data
4. **Quantifying** the each rules to conformance with each business policy
5. **Articulating** and storytelling the data state with a drillable report framework



You can have data without information, but you cannot have information without data

Python

The image shows a Jupyter Notebook interface with several open notebooks:

- Linear Regression.ipynb**: A notebook titled "In Depth: Linear Regression" with code for plotting a linear regression model.
- R.ipynb**: A notebook showing R code and a scatter plot of Sepal.Length vs Sepal.Width.
- Lorenz.ipynb**: A notebook titled "Julia" showing Julia code and a Lorenz attractor plot.
- Data.ipynb**: A notebook titled "python notebook" showing Python code for solving differential equations.
- Altair.ipynb**: A notebook titled "Seattle Weather: 2012-2015" showing an Altair visualization of Seattle weather data.
- Fasta.ipynb**: A notebook titled "R" showing R code and a scatter plot of Sepal.Length vs Sepal.Width.
- Julia.ipynb**: A notebook titled "R" showing R code and a scatter plot of Sepal.Length vs Sepal.Width.
- Cpp.ipynb**: A notebook titled "R" showing R code and a scatter plot of Sepal.Length vs Sepal.Width.
- Untitled.ipynb**: An empty notebook.

The interface includes a sidebar with icons for Python 3, C++11, C++14, C++17, Julia 1.1.0, phylogenetics (Python 3.7), R, and a launcher for other languages like Python 3, C++11, C++14, C++17, and Julia 1.1.0.

22



Real world example
Descriptive / Predictive
analysis Actionable report

Project: Implementing Process Efficiency by Analytics for one of the Asia biggest Stock Exchange



Problem Statement

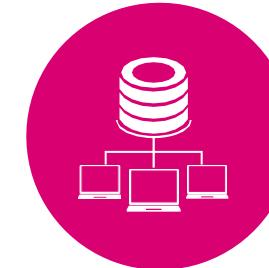
- Due to poor efficiency, low productivity, and customer dissatisfaction, Which resulted in delays and revenue loss for the company.
- Traditionally, Stock Market Business has been a
 - **Labor-intensive**
 - **Complex task**
 - **Requiring many hours of deep manual analysis**



"Without a systematic way to start and keep data clean, bad data will happen"

Solution

Collecting and Analysing Information



Identifying & Engaging Stakeholders



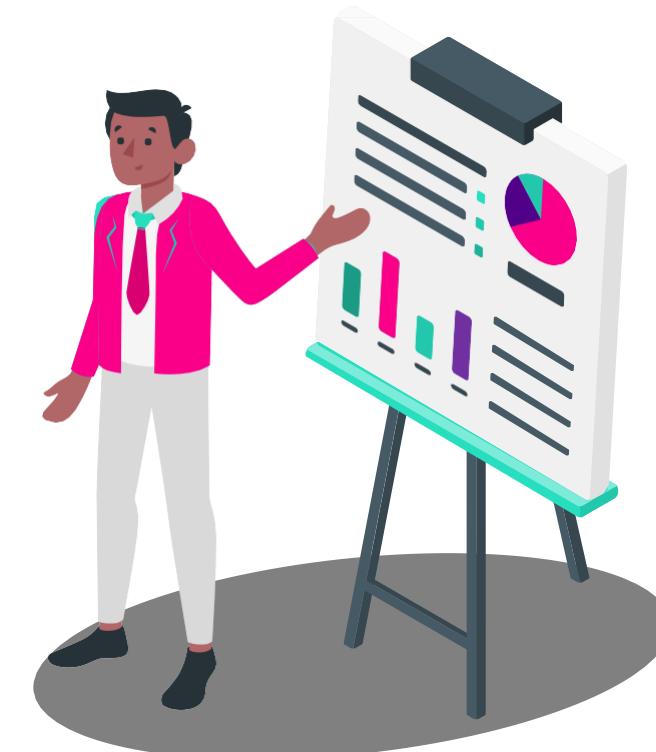
Analyzing Business Problem



Creating business rule engine platform



Creating Dashboard



Actionable Communication Platform

"You can have data without information, but you cannot have information without data"

Descriptive / Predictive analysis Actionable report

Use Case - Budget Forecasting

Report Efficiency Stats

- Improved Forecasting Efficiency by 60%
- Better Feedback from Management / Better Experience finance report got consistent 4 – 5 X
- Report acceptance increased by 35%
- Replaced the Legacy System



"In God we trust, all others bring data"

Benefits

Project: Implementing Process Efficiency by Analytics for one of the Asia biggest Stock Exchange



Business Benefits

- Ensure consistency by automating acceptability and Financial guidelines.
- Business rules assure that all Business guidelines are reviewed for each policy.
- Automated processes significantly improve review consistency and enhance Business results
- Enables collaboration between intermediaries, Finance Analyst , and senior and junior Analyst.
- Helps Finance team make quick, informed, and consistent underwriting decisions



"In God we trust, all others bring data"

DSIF

Data Science Immersive Flex



Topics

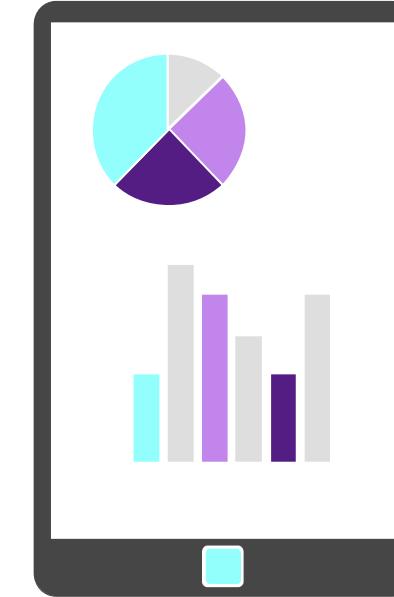
- Collect, extract, query, clean, and aggregate data for analysis
- statistical analysis on data using Python
- Build, implement, and evaluate data science problems
- Data visualization tools to communicate findings
- Create clear and reproducible reports to stakeholders
- Identify big data problems
- Understand how distributed systems and parallel computing technologies are solving these challenges
- Apply question, modeling, and validation problem-solving processes to datasets from various industry



*"TO BE BETTER
EACH DAY"*

Thank You

Your Questions?



"TO BE BETTER
EACH DAY"

Thank You

Thank you.

Maheshkumar Paik

mpaik@maltem.com

+65 82047512

m

