

5BUIS019W Business Analytics

Lecture 1

Introduction to Business Analytics

Content

- What is Business analytics?
- Problem solving approach
- Dataset: basic concepts
- Module content

PART 1

EXPLORING BUSINESS ANALYTICS

What is Business Analytics?

- Business analytics (BA) consists in the continuous iterative **exploration and analysis of business data** in order to measure business **performance** and drive an appropriate **planning**.
- It refers to **quantitative methods and techniques** that help making informed decisions to meet defined business objectives, i.e. **achieve desired outcomes**.
- Business analytics is used to understand the system by evaluating business operations, identify weaknesses, define solution alternatives and predict future states.
- Business Analytics Includes Statistics; Management Science/Operations Research techniques; Data Mining

What is Business Analytics?



Business Analytics popularity using Google Trends

Worldwide ▼ 2004 - present ▼ All categories ▼ Web Search ▼

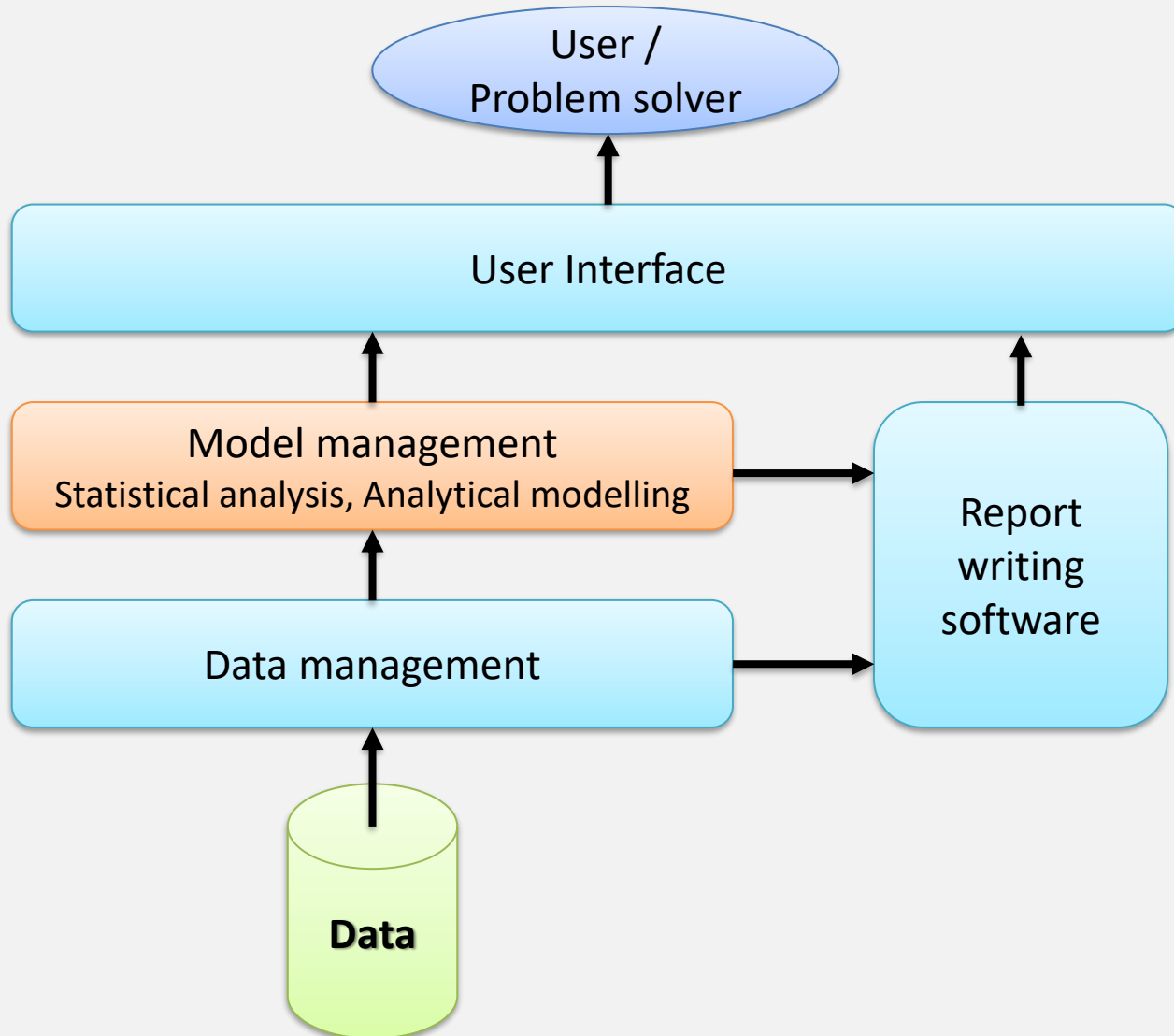
Interest over time ?



Business Analytics & Decision Support systems

- **Decision support system (DSS)** is a computer-based information system that aids the process of decision making.
- A DSS should
 - **Assist** users in making decisions to solve problems
 - **Support** their judgement rather than replace it
 - **Improve** their decision making effectiveness rather than its efficiency.

Business Analytics & Decision Support systems



Emerging Trends: Big Data

40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

Volume SCALE OF DATA



2005

2020

It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day



Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session



Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
— almost 2.5 connections per person on earth



Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States.



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[151 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT
are shared on Facebook every month



Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS



4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month



400 MILLION TWEETS
are sent per day by about 200 million monthly active users



1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

Veracity UNCERTAINTY OF DATA

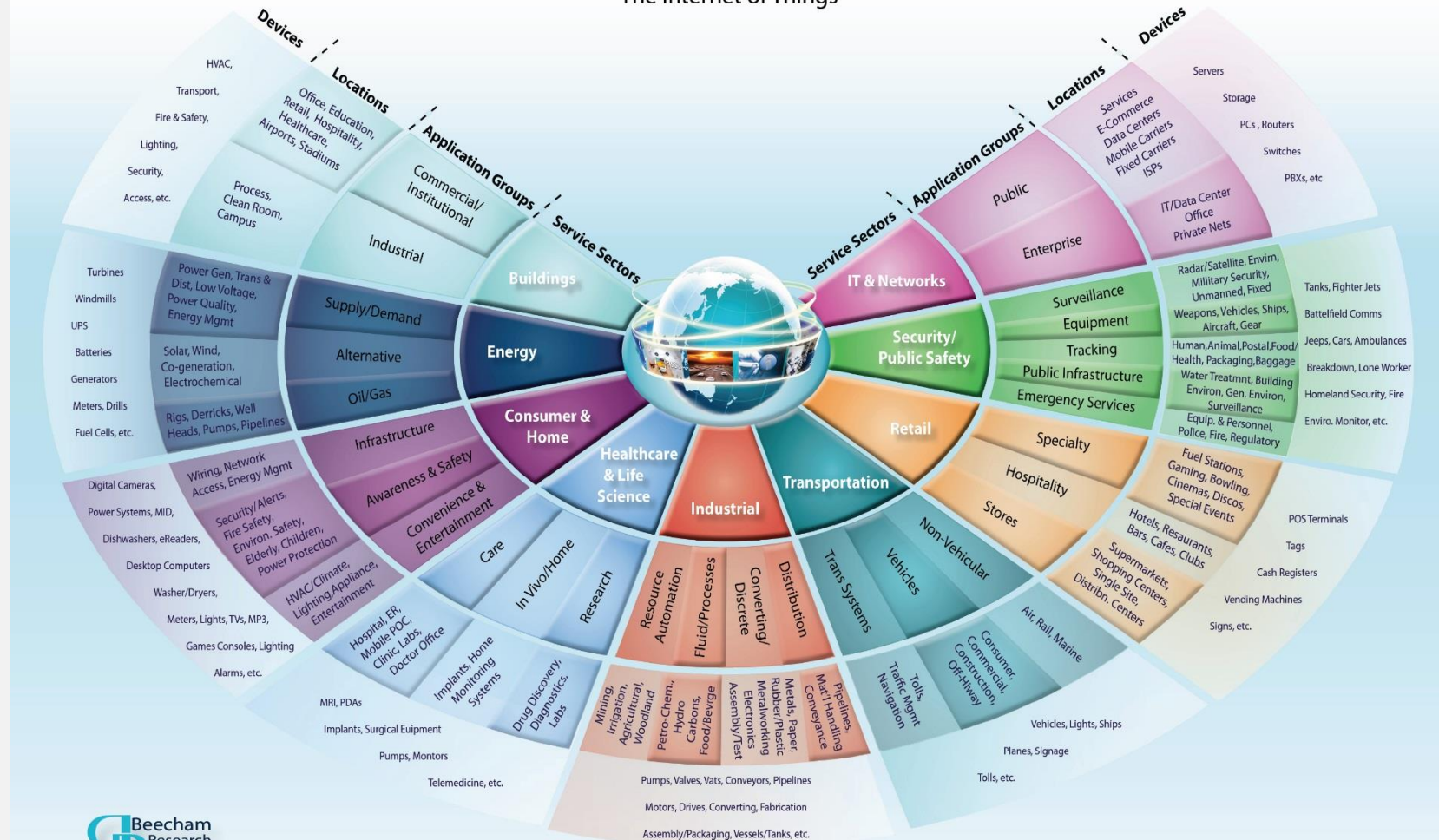
Poor data quality costs the US economy around
\$3.1 TRILLION A YEAR



Emerging Trends: The Internet of Things

M2M World of Connected Services

The Internet of Things



Boston | London

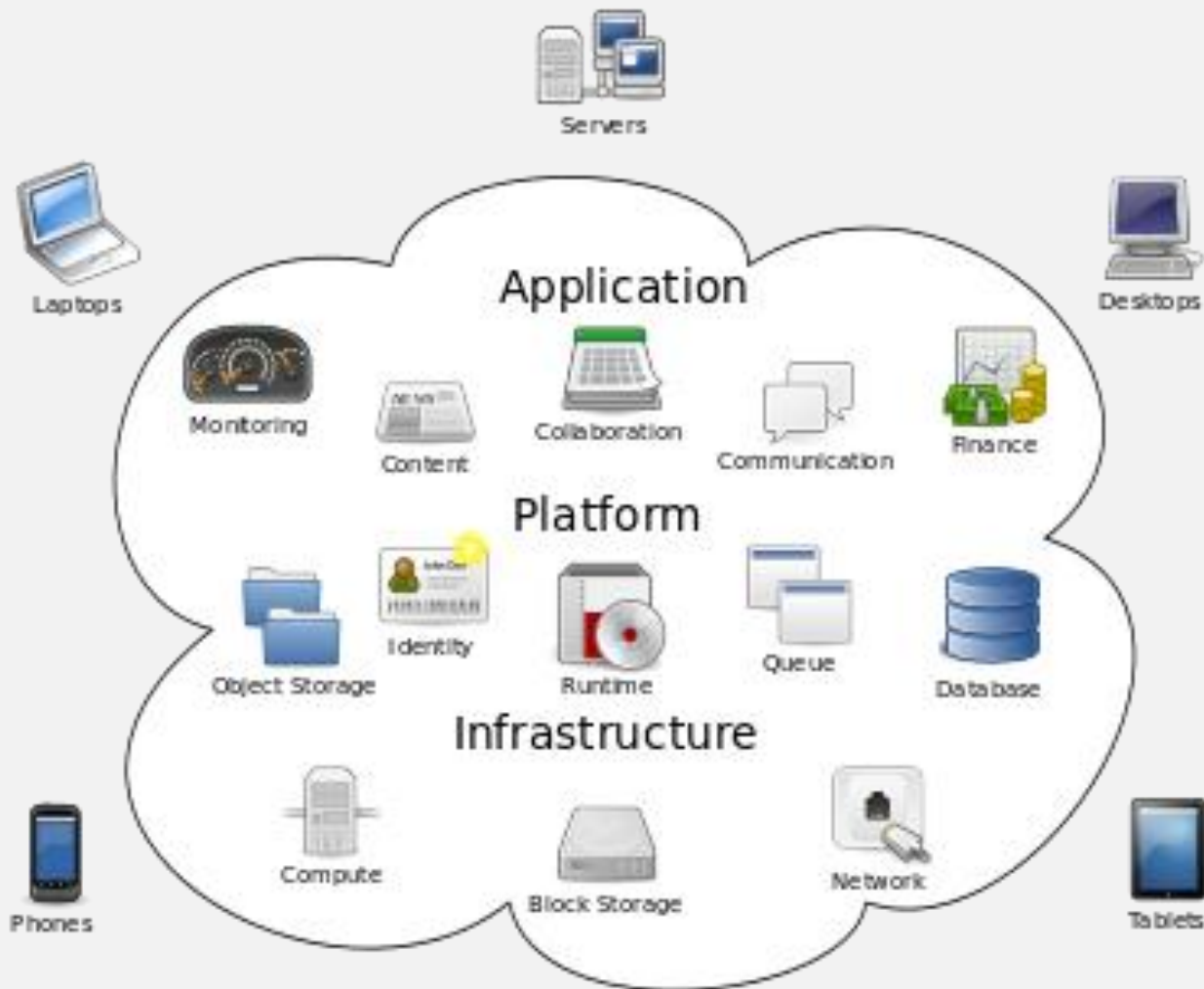
info@beechamresearch.com

+44 (0)845 533 1758

www.beechamresearch.com

© 2009 Beecham Research Ltd.

Emerging Trends: Cloud Computing



What career path?

- A **Data Scientist** use **Data Analytics** to solve business problems.
- Examples of related job titles:
 - Data analyst/manager, Operation Research analyst/manager, Business Intelligence analyst/manager, Insight analyst/manager, Performance analyst/manager...
- To find out more on Business Analytics career paths:
 - *“Data scientist: the sexiest job of the 21st century”* by T.H. Davenport and D.J. Patil
 - *“Career paths in Business Analytics – Plan your Next Best Role in the Data Science World”* by T. Srivastava
 - *“Dynamics of Data Science skills: how can all sectors benefit from data science talent?”* The Royal Society



PART 2

PROBLEM SOLVING APPROACH

Basic concepts

- What is a **system**?
 - It is a **group of elements or objects** that interact to accomplish a specified purpose.
- What is a **model**?
 - It is a representation of a system;
 - A model usually takes the form of a **set of assumptions** concerning the operation of the system.
 - Advantages of models (compared to experimenting with the real situation) → requires less time, is less expensive, involves less risk.
- A **mathematical model**
 - Represent real world problems through a system of **mathematical formulas and expressions** based on key assumptions, estimates, or statistical analyses.

Problem Solving

- Main Steps:
 - Understand the system;
 - Define the problem and the objectives;
 -  – Collect and Prepare the data;
 -  – Develop a model;
 - Model solution;
 - Model testing/validation;
 - Result analysis and discussion;
 - Generate report;
 - Implement the selected solution.

Data Preparation

- Data preparation is not a trivial step, due to the time required and the possibility of data collection errors.
- A model with 50 decision variables and 25 constraints could have over 1300 data elements!
- Often, a fairly large data base is needed.
- Information systems specialists might be needed.

Model Solution

- The analyst attempts to **identify the alternative** (the set of decision variable values) that **provides the “best” output** for the model.
- The “best” output is the **optimal solution**.
- If the alternative does not satisfy all of the model constraints, it is rejected as being **infeasible**, regardless of the objective function value.
- If the alternative satisfies all of the model constraints, it is **feasible** and a candidate for the “best” solution.

Model Solution

- One solution approach is **trial-and-error**.
 - Might not provide the best solution
- Special solution procedures have been developed for specific mathematical models.
 - Most practical applications require using a computer

Computer Software

- A variety of computer tools are available for solving statistical and mathematical models:
 - Matlab, AIMMS, Mathematica, Maple, LINDO/LINGO, CPLEX...
 - SIMUL8, AnyLogic, Arena, Flexsim, ExtendSim...
 - Palisade Decision tools (@Risk, StatTools...)...
 - R, Python...
- In this module we will be using
 - Microsoft Excel
 - R (Statistical programming language)

Model Testing and Validation

- Often, goodness/accuracy of a model cannot be assessed until solutions are generated.
- Small test problems having known, or at least expected, solutions can be used for model testing and validation.
- If the model generates expected solutions, use the model on the full-scale problem.
- If inaccuracies or potential shortcomings inherent in the model are identified, take corrective action such as:
 - Collection of more-accurate input data
 - Modification of the model

Report Generation

- A managerial report, based on the results of the model, should be prepared.
- The report should be easily understood by the decision maker.
- The report should include:
 - the **recommended decision**
 - other pertinent information about the results (for example, how sensitive the model solution is to the assumptions and data used in the model)

Implementation and Follow-Up

- Successful implementation of model results is of critical importance.
- Secure as much **user involvement** as possible **throughout the modeling process**.
- Continue to monitor the contribution of the model.
- It might be necessary to refine or expand the model.

PART 3

DATASET: BASIC CONCEPTS

Dataset

- A **data set** is usually a rectangular array of data, with variables in columns and observations in rows.
- A **variable** (or **field** or **attribute**) is a characteristic of members of a population, such as height, gender, or salary.
- An **observation** (or **case** or **record**) is a list of all variable values for a single member of a population.

| Person | Age | Gender | Regions | Children | Salary |
|--------|-----|--------|-----------------|----------|----------|
| 1 | 30 | Male | South East | 1 | £ 65,400 |
| 2 | 56 | Female | London | 2 | £ 62,000 |
| 3 | 39 | Male | North West | 0 | £ 63,200 |
| 4 | 32 | Male | West Midlands | 2 | £ 52,000 |
| 5 | 30 | Female | South West | 3 | £ 81,400 |
| 6 | 35 | Female | East of England | 3 | £ 46,300 |
| 7 | 68 | Female | Yorkshire | 2 | £ 49,600 |
| 8 | 48 | Male | North East | 1 | £ 45,900 |
| 9 | 42 | Male | London | 3 | £ 47,700 |
| 10 | 33 | Female | East Midlands | 1 | £ 59,900 |

Types of Data

- A **numerical** variable can represent:
 - A **measurement**, e.g. individual's weight, waiting time.
 - A **count**, e.g. number of pages you read from the textbook, number of transactions per day.
- A **categorical** variable can take on a finite number of values. It can be:
 - **Ordinal** if there is a natural ordering of its possible values, e.g. degree award.
 - **Nominal** if there is no natural ordering, e.g. gender, marital status.

Binary variable (or **dichotomous** variable) can take on exactly two values.
- There is also a third data type, a **date** variable.

| Person | Age | Gender | Regions | Children | Salary |
|--------|-----|--------|---------------|----------|----------|
| 1 | 30 | Male | South East | 1 | £ 65,400 |
| 2 | 56 | Female | London | 2 | £ 62,000 |
| 3 | 39 | Male | North West | 0 | £ 63,200 |
| 4 | 32 | Male | West Midlands | 2 | £ 52,000 |

Types of Data

- A numerical variable is
 - **discrete** if it results from a count, e.g. number of children.
 - **continuous** if it is the result of an essentially continuous measurement, e.g. weight or height.

| Person | Age | Gender | Regions | Children | Salary |
|--------|-----|--------|-----------------|----------|----------|
| 1 | 30 | Male | South East | 1 | £ 65,400 |
| 2 | 56 | Female | London | 2 | £ 62,000 |
| 3 | 39 | Male | North West | 0 | £ 63,200 |
| 4 | 32 | Male | West Midlands | 2 | £ 52,000 |
| 5 | 30 | Female | South West | 3 | £ 81,400 |
| 6 | 35 | Female | East of England | 3 | £ 46,300 |
| 7 | 68 | Female | Yorkshire | 2 | £ 49,600 |
| 8 | 48 | Male | North East | 1 | £ 45,900 |
| 9 | 42 | Male | London | 3 | £ 47,700 |
| 10 | 33 | Female | East Midlands | 1 | £ 59,900 |

Types of Data

- **Cross-sectional** data are data on a cross section of a population at a distinct point in time.
- **Time series** data are data collected over time.

| Person | Age | Gender | Regions | Children | Salary |
|--------|-----|--------|-----------------|----------|----------|
| 1 | 30 | Male | South East | 1 | £ 65,400 |
| 2 | 56 | Female | London | 2 | £ 62,000 |
| 3 | 39 | Male | North West | 0 | £ 63,200 |
| 4 | 32 | Male | West Midlands | 2 | £ 52,000 |
| 5 | 30 | Female | South West | 3 | £ 81,400 |
| 6 | 35 | Female | East of England | 3 | £ 46,300 |
| 7 | 68 | Female | Yorkshire | 2 | £ 49,600 |
| 8 | 48 | Male | North East | 1 | £ 45,900 |
| 9 | 42 | Male | London | 3 | £ 47,700 |
| 10 | 33 | Female | East Midlands | 1 | £ 59,900 |

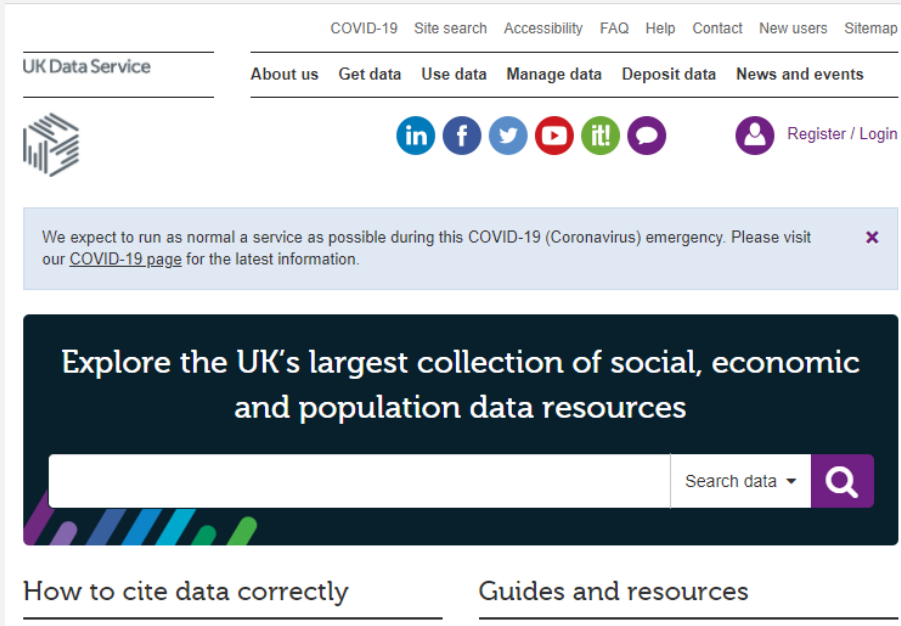
Cross-sectional data

| Quarter | Revenue (in 000s) |
|---------|----------------------|
| Q1-2010 | £ 1,333.80 |
| Q2-2010 | £ 1,372.80 |
| Q3-2010 | £ 1,536.60 |
| Q4-2010 | £ 3,719.30 |
| Q1-2011 | £ 1,523.60 |
| Q2-2011 | £ 1,623.70 |
| Q3-2011 | £ 1,749.80 |
| Q4-2011 | £ 4,422.60 |
| Q1-2012 | £ 1,671.80 |
| Q2-2012 | £ 1,712.10 |
| Q3-2012 | £ 1,883.70 |
| Q4-2012 | £ 5,060.90 |
| Q1-2013 | £ 1,900.60 |
| Q2-2013 | £ 1,887.60 |
| Q3-2013 | £ 2,120.30 |
| Q4-2013 | £ 5,460.00 |

Time series data

Data Sources

- **Primary data:** Data collected first-hand for a specific purpose.
 - Through surveys, interviews, experiments ...
- **Secondary data:** collected by a third party for a different purpose.
 - Available online freely or through a subscription
 - Official statistics, technical reports, scholarly journals, online platform ...



The screenshot shows the UK Data Service website. At the top, there is a navigation bar with links for COVID-19, Site search, Accessibility, FAQ, Help, Contact, New users, and Sitemap. Below this is a secondary navigation bar with links for About us, Get data, Use data, Manage data, Deposit data, and News and events. The main content area features a large dark blue banner with the text "Explore the UK's largest collection of social, economic and population data resources" and a search bar. A small blue box with a white 'x' icon contains a message about the service's status during the COVID-19 emergency. At the bottom, there are links for "How to cite data correctly" and "Guides and resources".

UK Data Service

COVID-19 Site search Accessibility FAQ Help Contact New users Sitemap

About us Get data Use data Manage data Deposit data News and events

We expect to run as normal a service as possible during this COVID-19 (Coronavirus) emergency. Please visit our [COVID-19 page](#) for the latest information.

Explore the UK's largest collection of social, economic and population data resources

Search data

How to cite data correctly Guides and resources



[This Photo](#) by Unknown Author is licensed under [CC BY-NC-ND](#)

PART 4

MODULE CONTENT


In this module, we will focus on

- Forecasting
- Simulation
- Decision making
- Linear programming

Forecasting

Met Office

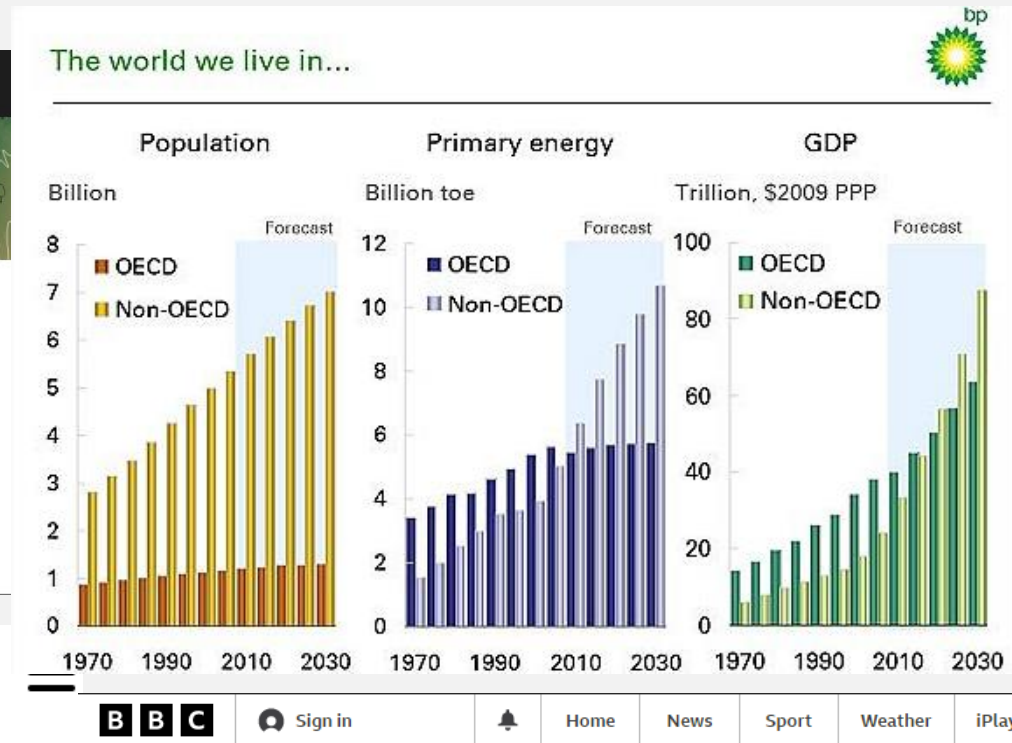
MENU



UK Climate Projections (UKCP)

The UK Climate Projections (UKCP) provides the most up-to-date assessment of how the climate of the UK may change over the 21st century. Find information to help with your climate change risk assessments and adaptation plans.

The Health Foundation



Home > News and comment > Blogs and long reads

How can the NHS make the most of risk prediction tools?

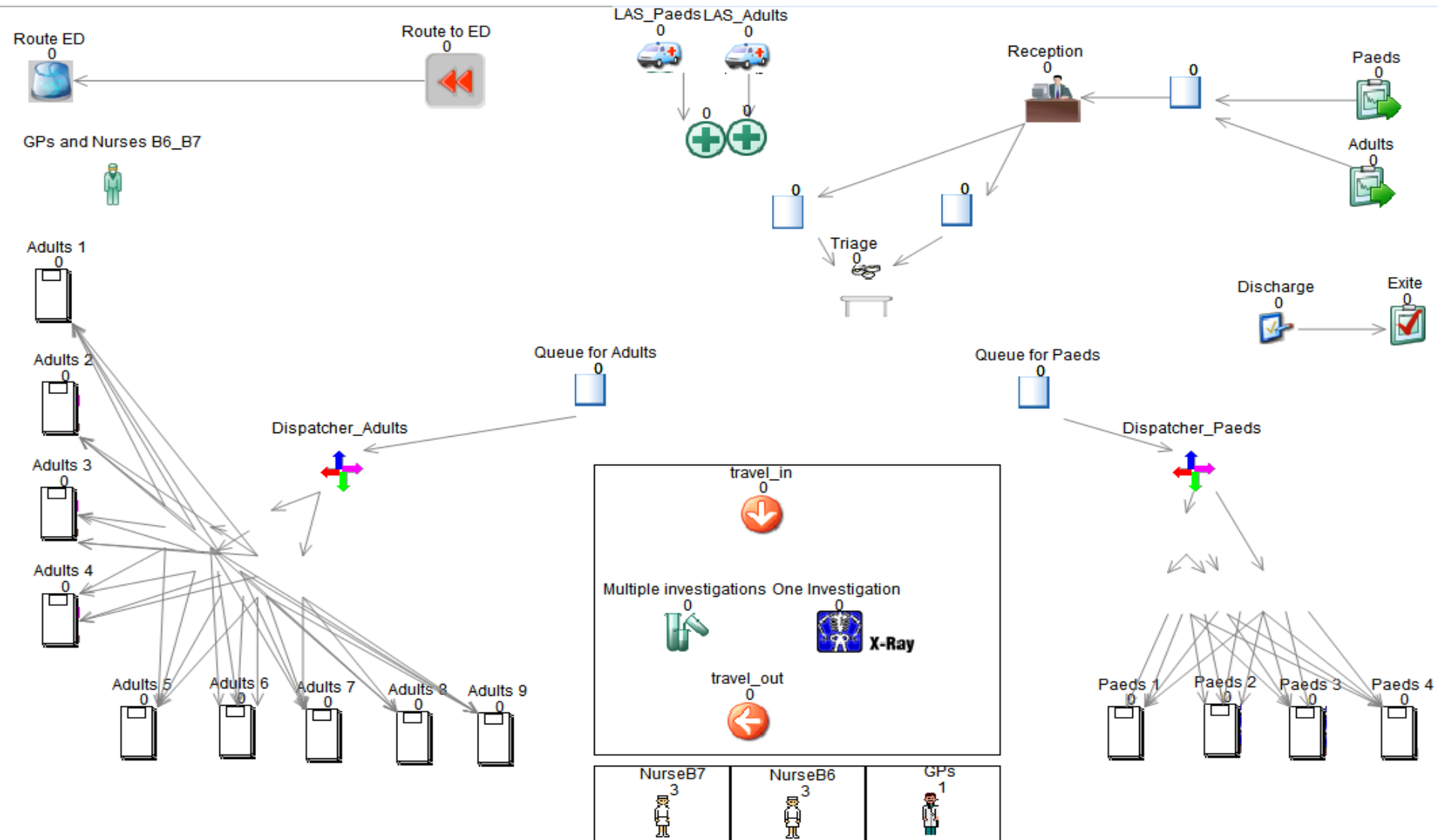
NEWS

Home | Coronavirus | Climate | UK | World | Business | Politics | Tech | Science | Health |

Business | Your Money | Market Data | Companies | Economy | Global Car Industry | Business

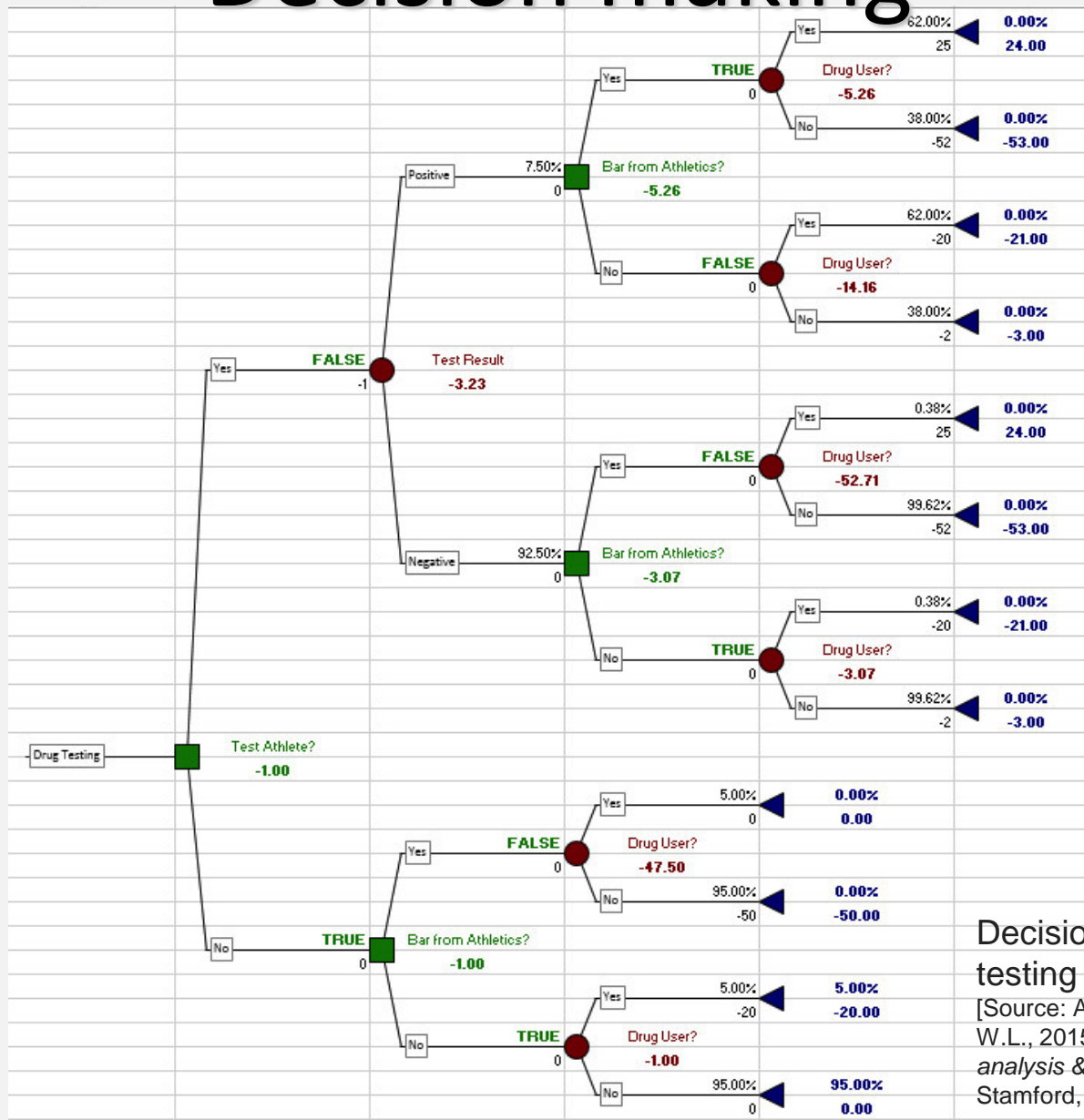
Amazon predicts slower sales growth as Covid boost eases

Simulation



[Source: Tadjer, M., Chaussalet, T.J., Fouladinajed, F. and Chahed, S., 2012. Using data mining and simulation for health system understanding and capacity planning: an application to urgent care. In *High Tech Human Touch: Proceedings of the 38th ORAHS conference*.]

Decision making

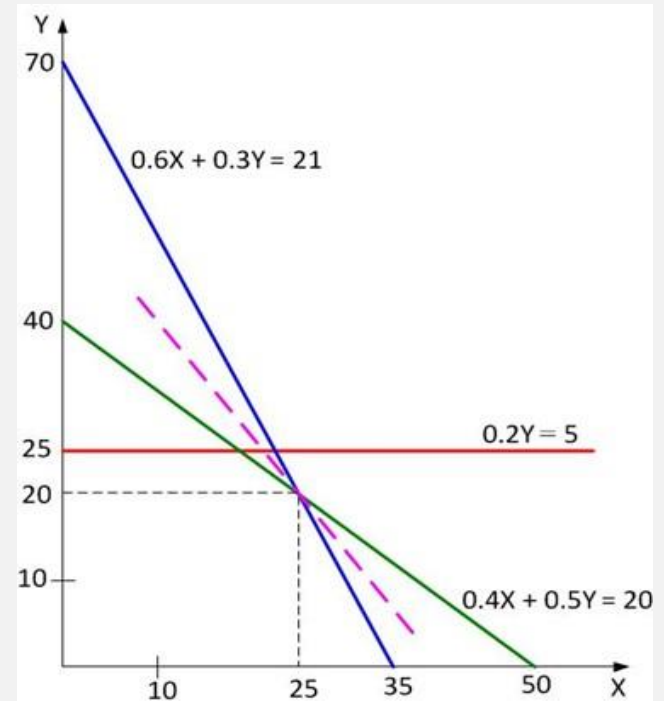


Decision Tree for Drug testing athletes

[Source: Albright, S.C. and Winston, W.L., 2015. *Business analytics: Data analysis & decision making*. 5th Edition. Stamford, CT, USA: Cengage Learning.]

Linear Programming (LP)

Maximise $Z = c_1 x_1 + c_2 x_2 + \dots + c_k x_k$
subject to $a_{11} x_1 + a_{12} x_2 + \dots + a_{1k} x_k \leq b_1$
 $a_{21} x_1 + a_{22} x_2 + \dots + a_{2k} x_k \leq b_2$
 \vdots
 $a_{n1} x_1 + a_{n2} x_2 + \dots + a_{nk} x_k \leq b_n$
 $x_1, x_2, \dots, x_k \geq 0$



A steam and electric power plant example by Alan Weiss, MathWorks
[<https://uk.mathworks.com/videos/mathematical-modeling-with-optimization-part-1-68973.html>]]