MSCI152: Introduction to Business Intelligence and Analytics

Lecture 1: Introduction

Lancaster University Management School

Overview

- Module Organisation
- Teaching and Learning
- Business Analytics and Intelligence
- Types of Data
- Collecting Data
- Data and Statistics used in examples

Module Organisation: People

- **Dr Christopher Kirkbride:** Charles Carter, Room D42
  - c.kirkbride@lancaster.ac.uk
- **Dr Anna Sroginis:** Charles Carter, Room D30
  - a.sroginis@lancaster.ac.uk
- **Tutors: Workshops and Drop-ins***
  - Chris/Anna* & Shayan
- **Administration/Undergraduate Coordinator:**
  - **Helena Greenwood:** Charles Carter, Room D25
  - h.greenwood@lancaster.ac.uk
- **Q & A Forum:** Available on Moodle

Module Organisation: Moodle

All required material will be provided here:

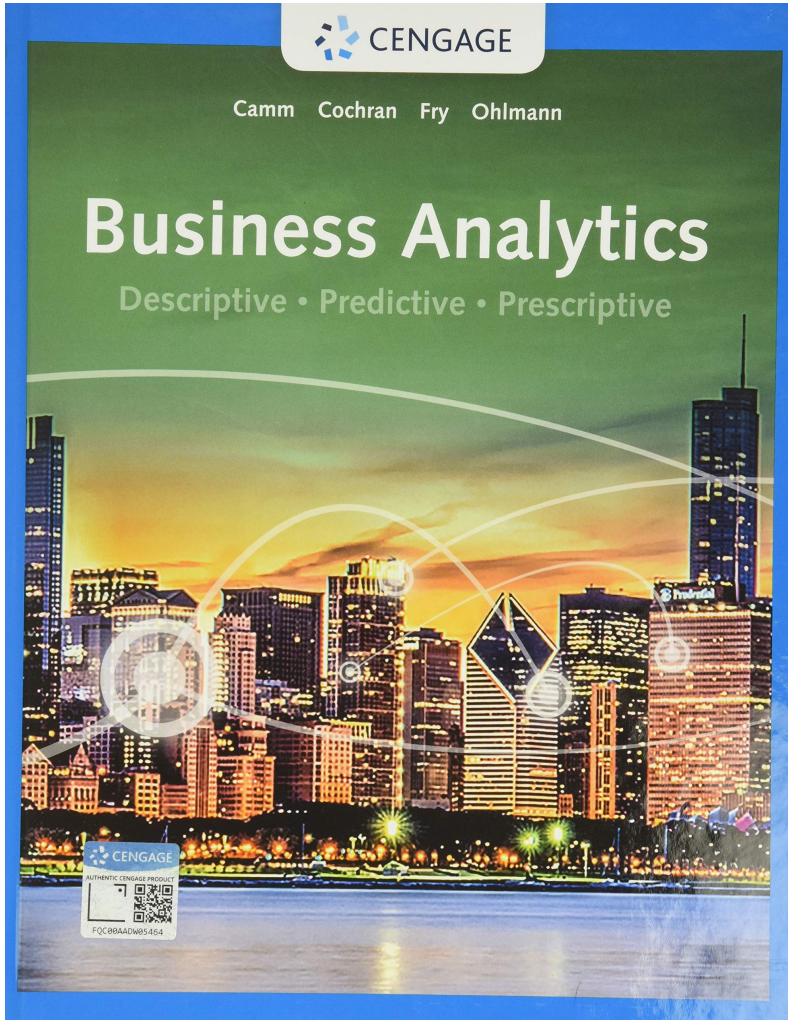
- Important announcements, supporting materials, etc.
- Lecture slides (in advance where possible)
- Tutorial and Workshop exercises (in advance)
  - Solutions posted at the end of the week
- Coursework instructions and upload
- Check frequently for updates/notifications

Module Organisation: Textbook

Camm, Cochran, Fry, Ohlmann

Business Analytics

Cengage, 2021, 4th Edition



Teaching and Learning: MSCI152

Lectures:

- **Thursday 13:00–14:00**, LEC Biology LT1
- **Friday 11:00–12:00**, LUMS(WP) LT15

Workshops: See personal timetable for allocation

- **Workshops:** 1 hours per week, Weeks 2–10
- **Online Drop-in:** Thursday & Friday: 16:00–17:00
    - MS Teams link will be available on Moodle

Administrative issues – e.g., change workshop slot, etc., see Helena Greenwood

Week 1 Workshop (Self-Study)

While there is no (physical) workshop this week, you have been provided with some self-study exercises on Moodle in the Workshop section.

These are meant as introductory and/or refresher exercises. They include:

- (Short) online training courses in MS Word and Excel
- Some refresher maths exercises

Try these out in your own time during the week!

We'll be on hand in the Week 2 workshops and online drop-in sessions if you have any questions about these

Teaching and Learning: Syllabus

Approximate order of topics:

- Introduction to Business Analytics and Intelligence
- Data, storage and collection
- Data visualisation
- Summarising data
- Identifying relationships in data (Regression)
- Forecasting
- (Possibly) Data Mining

Aim of the module is to apply these methods using Excel

Teaching and Learning: Assessment

Mini-tests: (10% total weighting)

- Short online MCQ test
- Week 5 covering weeks 1-4 (5%)
- Week 8 covering weeks 4-7 (5%)

Dashboard: (10% weighting)

- Individual Excel prototype Dashboard
- Week 7 submission

Final Test (40% weighting)

- Longer online MCQ test in Week 10
- Covering weeks 1-9

Coursework: (40% weighting)

- Individual 10-page analytical report
- Week 11 submission

Teaching and Learning: Advice

Prepare:

- Identify what you do and don't understand – be armed with questions

Attend:

- It's compulsory – make the most of these learning opportunities
- Don't forget to sign-in via iLancaster!
- Workshops – can leave early if you show the tutor you have (successfully) completed the workshop exercises

You are responsible for your own learning!

- Studying (Slides, textbook, videos)
- Doing (Workshop exercises, Textbook exercises, etc.)

Teaching and Learning: Support

As a cohort you will have a **broad range of mathematical backgrounds**:

- Some may not have studied (taught) mathematics for 2 years!
- The aim of the module is to get you all to a common level of understanding

Learning mathematics can be challenging and, at times, frustrating:

- **Persevere** – you **can** be comfortable and own this material
- **Seek advice** – when you are stuck sometimes you need a nudge!

Where can you get support?

- (Briefly) at the end of a lecture
- From tutors in Workshops/Drop-ins
- Q & A Forum

Teaching and Learning: LUMS MASH

Outside of the taught elements of the module Lancaster University provide others opportunities to develop your mathematics and statistics learning skills

LUMS MASH (Mathematics and Statistics Hub) provides Workshops and 1-to-1 advice sessions throughout the term. It has been very popular since its inception and many students have benefited from this service.

Details can be found here:

https://www.lancaster.ac.uk/library/learning-development/maths-and-stats-hub/

Become an Ambassador?

Lancaster University
Management School

LUMS Undergraduate Management Science Student Ambassadors

Are you a Business Analytics or MORSE student?

Take on a flexible paid role that works to your timetable:

- Earn money
- Represent LUMS to prospective students, parents and teachers at Open Days, Offer Holder Events, school visits and more
- Make friends, build skills and expand your network

Apply by 20 October 2024

lancaster.ac.uk/LUMS-ambassadors

Any questions, comments?

Recommender systems

Recommender systems

- are information filters with the aim of providing users with relevant items
- Used in online stores, streaming services, social media, search engines, e.g., Netflix
- A huge amount of user data and a range of complex algorithms are employed to display content/recommendations to users

Questions:

- What are positive aspects of these systems?
- What are the negative aspects?

Business Intelligence/Business Analytics/Data Science

My perspective: Each of these is an umbrella term that means roughly the same thing, in that there is a strong degree of overlap in their definitions.

- **Business Intelligence:** Wikipedia / Tableau
- **Business Analytics:** Wikipedia / Oracle
- **Data Science:** Wikipedia / IBM

Big Data

A common theme is the large growth in computing power and storage. This is, naturally, in conjunction with an of explosion data.

The four (or five) V's of Big Data (IBM):

- Volume
- Velocity
- Variety
- Veracity
- Value

[**Note:** Having a small amount data does not mean we cannot solve business problems!]

Business Analytics

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

- **Descriptive Analytics:** Use techniques on data that describe and summarize what has happened in the past.
- **Predictive Analytics:** Use techniques that the predict the future or describe relationships and interactions.
- **Prescriptive Analytics:** Use techniques that aid decision making or prescribe a course of action.

Applications: Health Analytics, Sports Analytics, Web Analytics, Marketing Analytics,...

Decision Making Process

Can be defined by the following process:

1. Identify and define the problem.
2. Determine the criteria that will be used to evaluate alternative solutions.
3. Determine the set of alternative solutions.
4. Evaluate the alternatives.
5. Choose an alternative.

In this module we focus on descriptive and predictive analytics, primarily "statistical" techniques, that we can use to inform decision making.

Statistics as a Process and Cycle

Collect data:

- What data is needed
- Find or collect data
- Examine and Question data

Discover and Communicate:

- Summarize data
- Categorise data
- Analyse data

Find meaning:

- Interpret the results in context
- Question the results
- Present the analysis

Once and done? Or repeated analysis (cycle process)?

Statistics: Population and Sample

Let's introduce some new terminology: **Population** and **Sample**

(Full) Population

- *Set* of all subjects or elements that we wish to study – about which we want to make inferences

Sample:

- *Subset* of the (full) population from which we collect data

Question: Why not simply gather data from the population?

Question: What makes a good sample?

[Inference: a conclusion reached on the basis of evidence and reasoning]

Population vs. Sample

Population	**Sample**
All University students in the UK	MSCI152 students at LU
All voters in the UK	An opinion poll
All potential video streamers	Users on Netflix
All potential customers	Customers included in Market Survey
All shares on the stock market	Shares included in the Dow Jones

Question: Are these good samples?

Question: What makes a good (or at least representative) sample?

Question: But what do we want to know?

Types of Data

The type of data affects how best to present and analyse it

Qualitative = Categorical data:

- A subjective qualification (image, country, brand, type,. . .)
- Can often be put into categories
    - Usually not much use if it isn't!

Quantitative data = Numerical data

- Measurements that are recorded on a naturally occurring numerical scale
- A numerical quantity (number, age, length, price. . .)

Nominal Data

Qualitative data only

Just **names or identifiers**, no obvious order. e.g.

- Type of product
- Favourite football team
- Country of birth

There can be an **agreed order**. e.g.

- Alphabetically, colours often ordered by lightness, ...
- Categories are also usually nominal

Ordinal Data

Data with obvious (objective) **order**

Can be both qualitative **and** quantitative. e.g.

- rating of a service:
- rankings of all types

Must be **cautious** when using numbers. e.g.

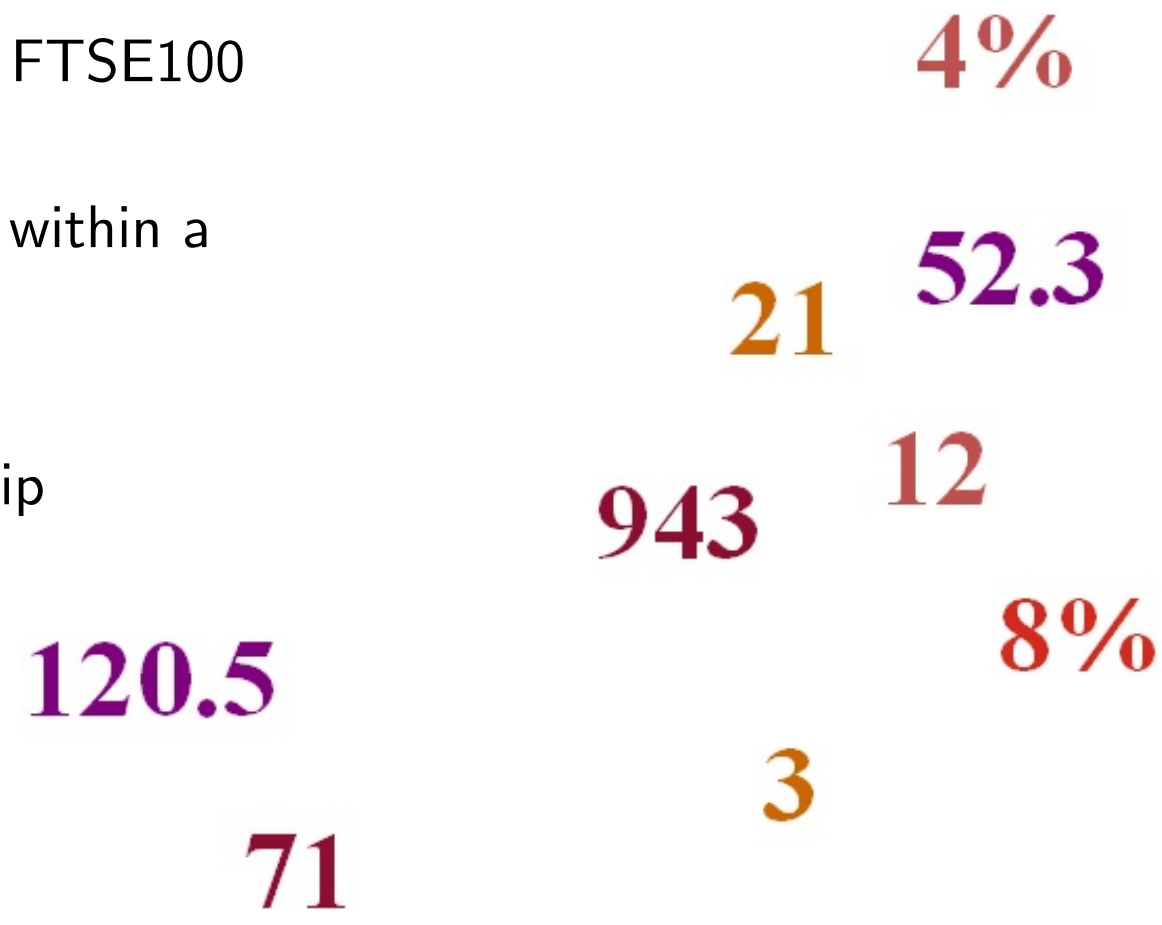
- very good = 5, good = 4, average = 3, poor = 2, very poor = 1
- Difference between 5 and 4 might not be the same as between 4 and 3 here!
- Can we calculate a mean value?

very good
good
average
poor
very poor

Quantitative: Numerical Data

Examples:

- Number of biscuit tins sold
- Salaries of CEO in FTSE100 companies
- Ages of employees within a company
- Mileage of Lorries
- Containers on a ship



Types of Numerical Data

Discrete

Data that assume **discrete values** (integer numbers, gaps between numbers, countable)

For example:

- number of stocks
- number of students in class
- number of cars
- number of typos

Continuous

Data that can assume **any value** within an interval (i.e., is measurable)

For example:

- time
- money
- distance
- weight

Note: Collected data may need to be put into categories

- less than £15k, between £15k and £30k, ...
- such categories are then **ordinal** data

Categories of data

By collection:

- Cross-sectional data
- Time-series data

By structure:

- structured data (transaction data, . . .)
- unstructured data (contracts, texts, voices, video, . . .)

By characteristics:

- descriptive data (attributes, characteristics, geo/demographics)
- behavior data (orders, transactions, payments, credit)
- interaction data (email, chat transcripts, click-streams)
- attitudinal data (opinions, preferences, needs and desires)

Big Data

Analysing high-volume semi- or unstructured data requires specialised storage and programming approaches

- Single location storage can be risky, expensive, time-consuming to process and, potentially, simply not possible.

Hadoop (HDFS): Open-source programming environment that breaks data down into parts that are stored (and replicated) across many nodes (or computers) in a cluster.

MapReduce: A programming model that: Maps data to nodes in the cluster for storing and processing; then, Reduces by collecting answers from the nodes and combining them into a response to the original query.

Big Data issues is a subset of analytics: Still scope for business insight from more manageabe databases!

How do we obtain data?

If you are lucky (or perhaps unlucky), it is given to you

- **Check** exactly what it is, how reliable, etc.

Sometimes there may be **too much** data

- For example, all records kept for a company
- Think about which bits you really want/are relevant to the task
- You cannot analyse all of it in a reasonable time!

At other times there may **not be enough** data or **non at all**

- Must collect it
- Often requires sampling (next week's lectures)
- important to consider the value of this

Data that is given to you

May not be **self-explanatory**

May not be exactly **what you want**

May be **suspect**

May **hide things** (hidden columns?)

May be **summary data** when what you really need is **raw data**.
e.g.

- you need to analyse waiting times at hospitals and all you are given is a set of averages

May **miss** lots of **important information**

May include **bugs**, typos, **errors**, spelling mistakes, . . .

Examples: In the news

A newspaper headline may state:

Company X's profits fall by 50%

That doesn't sound good for Company X:

- Was it from £100 million to £50 million?
- Or from £1 million to £500k?
- What was the sector performance? Competitor analysis?

The headline does not really tell us very much!

Examples: In advertising

Advertisers are not allowed to lie

- But they do not have to tell the whole truth
- And they will try to mislead you as much as they think they can get away with

Flights to the Canaries from £29.99 one way!

- What about the other way?
- Are there a lot of expensive taxes and charges?
- What level of service will I receive?
- Will it take me to the island I want to go to?
- And then there is the word "**from**" ...

Examples: Another from advertising

A well known slogan (in the UK):

Domestos kills 99% of all known germs

Assume that you are obsessively interested in household hygiene

- What do you think of this slogan?
- What is it telling you?
- And what is it not telling you?

Data: Questions we should ask

- **Who** says so?
- **Why** are they saying this?
- **What** does it mean?
- **How** do they (claim to) know?
- So what?

Wrap up

Today we:

- Over-viewed the general structure of the module
- Gave a bit of flavour to what we are doing and why
- Discussed **data**:
  - The different types of data
  - Thoughts about where data comes from
  - Challenges with data
  - Examples of data communication (and why we need to question it)

Next time:

- Approaches to data collection: **Sampling Methods**