**University of Nottingham**

**Understanding and Creating the Conditions to Acquire a Data Analytics Capability at Lincoln College**

**Graham Harrison**

**MBA**

─────────────────────────────────────────

***How can the conditions be created for Lincoln College to acquire a data analytics capability that delivers the benefits and gains that are evident in other sectors but are thus far unproven in the UK education sector?***

**by Graham Harrison**

**2020**

A Management Project presented in part consideration for the degree of

Master of Business Administration - Executive.

─────────────────────────────────────────

Executive Overview: The Idea in Brief

The Dilemma

Data analytics is the process of analysing data with the objective of identifying patterns, obtaining insight, extracting meaning and drawing conclusions. Business analytics is the use of data analytics by firms to support and inform decision making in order to improve business outcomes and to gain competitive advantage.

Many commercial firms have adopted data analytics and embedded it within their operations and there is a large and accepted body of evidence that firms with a mature data analytics capability are high performers relative to competitors who lack this capability.

There is a recognised gap between the “haves” and “have-nots” and Further Education in the UK and small-medium sized enterprises are amongst the sectors have not traditionally been able to secure the resources and expertise to acquire the capabilities and benefit from the outcomes.

The Solution

Lincoln College can realise the benefits by developing and growing a small subsidiary company called Lincolnshire Data Analytics Ltd. that exists primarily to drive data-analytics associated outcomes through all aspects of its curriculum, administrative and management functions.

The sub-company will be staffed by a small number of homegrown data scientists by attracting individuals demonstrating strong potential, commitment and enthusiasm who will be given the support, opportunity and resources necessary to rapidly acquire high-end data science skills.

As well as internal, educational delivery Lincolnshire Data Analytics Ltd. will seek external sponsors and funding and will deliver external commercial services locally to contribute to its operating costs.

The Benefits

The benefits to Lincoln College will be the adoption and embedding of data analytics within its operations, thereby building a strong culture of data-driven decision making.

This will ensure highly effective management decisions that avoid conscious and unconscious biases thereby enabling exceptional decision making that includes the optimum triangulation of experience and evidence.

Tangible and measurable improvements in organisational outcomes will accrue including improved processes, improved competitive position, better curriculum products resulting from evidence-backed market data, improved human resource capabilities and better risk management.

This capability will contribute to the delivery of improvements in all four cornerstones by which Lincoln College measures itself –

* The capability, capacity & well-being of our people.
* The quality & relevance of our education provision.
* The contribution from our UK commercial and international activities.
* Ensuring that the use of our information, finances & estate enables great learning.

The delivery of commercial data analytics will distil these benefits to local employers and the local economy whilst a collaboration with software suppliers will deliver benefits to the wider FE sector.

Table of Contents

[Chapter 1: Initial Management Project Problem 1](#_Toc39319540)

[Chapter 2 - Literature Review 2](#_Toc39319541)

[A. Research Questions 3](#_Toc39319542)

[Chapter 3: Findings and Objectives 36](#_Toc39319543)

[A. Key Findings 36](#_Toc39319544)

[B. Objectives 44](#_Toc39319545)

[Chapter 4: Defining the Research Question 52](#_Toc39319546)

[Chapter 5: Methodology 53](#_Toc39319547)

[A. Elements of the Methodology 53](#_Toc39319548)

[Chapter 6: Proposal and Recommendations 70](#_Toc39319549)

[A. Proposal 70](#_Toc39319550)

[B. Recommendations 71](#_Toc39319551)

[Chapter 7: Conclusion and Next Steps 91](#_Toc39319552)

[References 92](#_Toc39319553)

[Acknowledgements 101](#_Toc39319554)

[Index of Figures and Tables 102](#_Toc39319555)

# Chapter 1: Initial Management Project Problem

Lincoln College has an established and mature Management Information Services function which has seen significant improvements in recent years. However, it has now reached a plateau and the improvements have slowed down and become more incremental than revolutionary.

The CEO would like to see Lincoln College develop a data science and business analytics capability which could see a revolutionary new approach with the potential to significantly accelerate improvements in this area.

The introduction of this capability would be something completely new to Lincoln College and a preliminary investigation has suggested that there is nothing similar within the Further Education (FE) sector.

The novelty of the proposal combined with the significant financial, resource, regulatory and demographic challenges within the FE sector make the introduction of such a leading-edge management capability particularly challenging and interesting.

The overall objectives of this report are –

* To explain what data analytics and the many associated terms mean, to develop these themes and to provide clear definitions.
* To explore and understand the potential of data analytics particularly within the Further Education sector
* To make a proposal as to how Lincoln College can acquire a data analytics capability within the context of the significant challenges and constraints currently faced by the sector.

# Chapter 2 - Literature Review

“A research literature review is a systematic, explicit and reproducible method for identifying, evaluating and synthesizing the existing body of completed and recorded work produced by researchers, scholars and practitioners“ (Fink, 2020).

The purpose of this literature review is to apply the principles defined by (Fink, 2020) to the management problem in order to “build an understanding of [the] topic, of what has already been done on it, how it has been researched, and what the key issues are” (Hart, 2014).

However, according to Dr. David Stillwell of the University of Cambridge ,the field of Data Analytics is still relatively new and hence more emphasis than would be usual will need to be placed on practitioners including sources like blogs, podcasts, social media posts, web-sites etc. (Stillwell (1), 2019).

## Research Questions

### What do “big data”, “data analytics”, “business analytics” and other related terms mean and how do they relate?

The objective of this part of the literature review is to explore the jargon within the field and to establish clear and easily understood definitions and explanations that can be used as frames of reference within the rest of the report.

#### Big Data

Big data is one of the key terms that is used frequently in books, journal articles, web sites and the media.

There is agreement across much of the literature in the use of the acronym of the “V’s” to define what big data is although the 4th and 5th V are not common to all the sources –

**Volume**: volume refers to the size of big data.

For example, 4 billion video-hours are generated on YouTube every month, 400 million tweets are created every day, and 30 billion posts are shared on Facebook every month. (Cambridge (3), 2019).

A succinct and easily understood definition offered by (Cambridge (3), 2019) is that “big data is anything that will not fit in an Excel spread-sheet”.

**Velocity**: big data is typically generated at high speed and it needs to be captured, analysed and processed in real-time and with no delays.

The company Just Eat decides whether to purchase Google AdWords campaigns during Champions League matches by monitoring real-time activity. By the end of the match that data has no further relevance or value (BBVA, 2017).

**Variety**: Traditionally data was generated by user input and stored in tabular database systems but big data is typically much more varied in format.

YouTube Videos, Instagram Images, Facebook posts, Tweets, GPS co-ordinates generated by smart-phones and data generated by sensors inside turbine engines are all examples of variety.

**Veracity**: veracity refers to the potential for big data to be highly accurate because it is based on actual, real-time behaviour.

For example, web site logs, smart phone records, and data captured from sensors and wearable devices will have high accuracy when compared to users keying data into traditional data processing systems which could suffer from delay, lower speed and transcription errors.

**Value**: value refers to the ability to build or transform a business based on data and the opposite where firms that ignore big data can lose value relative to their competitors.

The “FANG” companies (Facebook, Amazon, Apple, Netflix, Google) demonstrate the potential for big data to generate corporate value whilst companies like Uber, Spotify and Strava have business models that are completely reliant on their big data (Shaw, 2018), (Cambridge (3), 2019), (McAfee & Erik, 2012), (BBVA, 2017).

Big data is also surprisingly easy for firms to acquire as there is a wealth of publicly available, low cost big data that is easy to access and use.

For example, it is quite straight-forward to access tweets that are in the public domain and to use that data in the IBM Watson Artificial Intelligence Personality Profile service to acquire highly accurate customer profiles.

The example shown in Figure 1 was developed by the author in the Python programming language and demonstrates how sophisticated customer profiling can be developed quickly and easily using free resources.

However, there are many firms that have not developed and understanding of big data and its potential to generate value as the oft-quoted Facebook post by Dan Ariely, Professor of Psychology and Behavioural Economics at Duke University, so eloquently describes –

*“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it…”*

(Ariely, 2013)

Big data then, and the more traditional tabular data stored in databases and spreadsheets, are the raw materials and inputs of data analytics.

#### Data Analytics and Business Analytics

There are no commonly accepted definitions of the terms “data analytics” and “business analytics” in the literature hence the following definitions are proposed by the author through the synthesis of several sources including (Cambridge Dictionary, 2020), (Informatica, 2020) and (Cambridge (0), 2019) –

**“Data Analytics is the process of analysing data with the objective of identifying patterns, obtaining insight, extracting meaning and drawing conclusions.”**

**“Business Analytics is the use of data analytics by firms to support and inform decision making in order to improve business outcomes and to gain competitive advantage”**

Hence a data analytics capability will enable a firm to extract insight and meaning from data whilst a business analytics capability will translate those findings into business outcomes and competitive advantage.

Also “Data Analytics” could be applied in a scientific context to extract meaning in the pure pursuit of knowledge acquisition whereas business analytics applies that insight in a commercial and business context.

#### Descriptive, Predictive and Prescriptive Analytics

These are all Subsets of Data Analytics. Each represents a set of techniques that are useful in different contexts to obtain and extract the desired patterns, meanings and insights.

**Descriptive Analytics** refers to reports and other visualisations explaining what happened in the past (Davenport & Harris, 2017). It is backwards looking and historical. Descriptive analytics can tell you what just happened but cannot offer any insight into what might happen next.

**Predictive Analytics** is used to predict the future (Davenport & Harris, 2017). However, (Wu (2), 2013) states that predicting the future through analytics is not possible, rather predictive analytics will forecast what might happen and is probabilistic in nature.

|  |  |
| --- | --- |
|  |  |
| *Figure 1 – IBM Watson AI Service used with Python and Jupyter to profile Barack Obama from his public Internet Quotes* | |

**Prescriptive Analytics** is the application of the findings, insights and models produced by descriptive and predictive analytics in order to make business decisions (Cambridge (7), 2019) or to do a job better (Davenport & Harris, 2017).

|  |
| --- |
|  |
| *Figure 2 – Visual Representation of Descriptive, Predictive and Prescriptive Analytics* (Slide Team, 2020) |

#### Diagnostic Analytics and Indicative Analytics

**Diagnostic analytics** expands the “what happened?” of descriptive analytics to find explanations as to why it happened by visualising, drilling down and correlating descriptive data (Bekker, 2017). Diagnostic analytics ascribes meaning, understanding and explanation to descriptive analytics.

**Indicative analytics** isthe use of descriptive analytics to make predictions about what is about to happen (Howitt, 2020). For example, current data about 15 year-old school children can be used to predict the size of the 16 year-old market next year. This technique has some of the benefits of predictive analytics without the associated complexity[[1]](#footnote-1).

#### Visualising the Relationships between the Concepts

*Figure 3* shows how all the concepts relate and highlights how the subsets of data analytics vary in respect of complexity and potential value.

This suggest that firms could benefit by adopting the higher value types of analytics. However, the available evidence suggests that the opposite is actually the case.

The author’s observation and the interview with (Howitt, 2020) suggest that most firms are firmly stuck in the descriptive analytics space and have little or no capabilities in the others. This is triangulated by (Wu (1), 2013) who states that “over 80% of … business analytics … are descriptive”.

|  |
| --- |
|  |
| *Figure 3 – Visual Representation of Analytics* (Bekker, 2017)*,* (Mc.ai, 2018) |

#### Artificial Intelligence, Machine Learning and Deep Learning

Artificial Intelligence, Machine Learning and Deep Learning are all techniques that are used within predictive analytics to create models that are used to make probabilistic predictions about future behaviour.

**Artificial Intelligence** is a blanket term for all the techniques that are used within this space. AI uses computing techniques that often mimic human behaviour to make probabilistic predictions about future outcomes.

Models are built using sets of training data but crucially the models can then make forecasts about data they have never encountered before which differentiates AI from more traditional computing and statistical techniques.

**Machine Learning** is a subset of artificial intelligence. Techniques in this category include regression analysis, tree induction and support vector machines. They are referred to as “supervised” learning because human experts typically use their technical and domain expertise to build, encode and tune the models.

**Deep Learning** is a subset of machine learning. Techniques like neural networks can perform complex tasks including speech and image recognition. The key difference to machine learning is that these models, once created, can train themselves and hence are often referred to as “unsupervised” learning.

#### Prescriptive Analytics in Detail

In prescriptive analytics decision makers and managers come together with data analysts and data scientists. The insights gained from the models and outputs of the data analytics processes will be reviewed, discussed and triangulated with other sources such as the experience of the managers as the basis for making operational and strategic decisions.

Because an understanding of prescriptive analytics is critically important for decision makers it is explored in more detail in the sections below.

##### Correlation vs. Causation

Correlation indicates that there is some type of relationship between two variables (Bock, n.d.). Causation indicates a stronger relationship where one variable is causing the other i.e. student absence is causing poor exam grades.

Figure 4 shows a scatter plot of height vs. weight for females. The data-points cluster along the diagonal axis which suggests a correlation between height and weight (correlation).

Figure 5 shows a scatter plot of the attendance of students vs. the attendance of their peers. The diagonal clustering again suggests a correlation.

In this case it would be more reasonable to assume a potential causal link (causation) between the absence of a student’s peer group and the student’s own absence.

If it could be demonstrated that a student’s absence causes their peers to be absent or vice versa then this would provide an actionable insight: if a student’s peers change their behaviour or a student moves peer groups then their attendance could be positively influenced.

Analytics techniques often look for correlation as a pattern that can be exploited. For example, if I know the height of a female, I can use this data to predict their weight. Height and weight are correlated but it is not true to say that height causes weight which is causation.

However, datasets can often demonstrate correlation where there is no meaning and the link is co-incidental and Figure 6 provides a clear example of a spurious correlation. Clearly forming management action based on data whose patterns are co-incidental would be dangerous and in-correct.

|  |
| --- |
|  |
| *Figure 4 – Scatter Plot of the Relationship between Height and Weight in Females* |

|  |
| --- |
|  |
| *Figure 5 – Scatter Plot of the Relationship between a student’s attendance and the attendance of their peers / contacts* (Kassarnig, et al., 2017) |

|  |
| --- |
| A close up of a map  Description automatically generated |
| *Figure 6 – The Correlation Between Margarine Consumption and Divorce Rates* (Vigen, 2020) |

Discounting genuinely spurious correlations there are different views in the literature about whether action and intervention requires proven causality or whether correlation alone is a valid basis for action.

The Khan Academy uses a case study correlating skipping breakfast and being inactive with obesity (Khan Academy (1), 2012) to argue that acting on this link would be wrong because the there is no causal link.

Chris Anderson offered a contrary view by stating that the huge volumes of available big data mean that “correlation is enough” (Anderson, 2008) giving rise to a popular meme in the big data community that “causality is dead” (Ritter, 2014).

For example, if you know that sales always increase on rainy days do you need to know why or should you just increase stock if tomorrow’s weather forecast predicts a downpour?

(Ritter, 2014) expresses a third view and proposes a simple model based on confidence and risk for deciding when to act on a correlation and when to ignore it –

|  |
| --- |
|  |
| *Figure 7 – When to Act on a Data Correlation* (Ritter, 2014) |

Although there are differences in the literature decision makers need to understand the terms and concepts and then consider the context, risk, data volumes and other factors when deciding whether to act on a correlation, when to ignore it or when to seek further evidence.

##### Experimentation / AB Testing

Experimentation (or AB Testing) is a common theme found throughout the literature (Cambridge (4), 2019), (Kohavi & Thomke, 2017), (Gallo, 2018) which can be very useful to decision makers during prescriptive analytics –

1. It is accepted that Experimentation can ascertain causality i.e. that changing one factor has a causal link to another.
2. It can be used to “try out” a conclusion that has been drawn from the data analytics process. If the conclusion would be very costly to roll out across a firm, it can be verified on a smaller scale commitment of resources and full-scale adoption.

Experimentation is simply the design of a controlled experiment to prove or disprove a hypothesis about correlation.

One “control” or “B group” contains customers that are excluded from the initial trial whilst members of the “intervention” or “A group” are tested against the hypothesis and the results are compared.

Figure 9 shows a correlation between the introduction of police shoulder cameras and a simultaneous drop in complaints and a reduction of use-of-force incidents. A subsequent AB test proved that there was no causation and that the changes were attributable to other factors. This prevented a widespread rollout of expensive cameras that would have had no impact.

Clearly decision makers need to be aware of experimentation and AB testing so that, when appropriate, causality can be evaluated before significant decisions are taken based on descriptive data.

##### Decision Biases

(Cambridge (1), 2019) dedicates an entire module to data biases and decision traps and asserts that firms can have good data and smart people but still make poor decisions –

|  |
| --- |
|  |
| *Figure 8 – Decision Traps* (Cambridge (1), 2019) |

Decision makers need to know about decision traps so that they can be aware of biases they might have consciously or un-consciously applied and hence make better decisions and improve outcomes derived from data analytics.

5 decision traps are proposed and defined by (Cambridge (1), 2019) –

**Framing**: this relates to the inability to frame the question correctly.

For example, Pepsi tried to out-design the new curved Coca-Cola bottle which had increased their market share but framing the problem around curved bottles proved to be false. The answer turned out to be introduction of a basic-design 2 litre plastic bottle after the frame of the question was changed to “how do we generate more sales?”.

**Over-Confidence bias**: over-confidence can be either an overly optimistic or pessimistic bias about the outcome.

Over-confidence can be illustrated by the following quote - “A severe depression like that of 1920-21 is outside of the range of probability” (Harvard Economic Society, 1929).

|  |
| --- |
|  |
| *Figure 9 – Rialto Police Shoulder Camera Experiment* (Cambridge (4), 2019) |

**Confirmation bias**: confirmation bias is where the decision maker(s) use the data to confirm an existing view or opinion and become blind to other views.

An example is a US survey that used research data to confirm a pre-held belief that the lowest rates of kidney cancer were in small towns. The highest rates of kidney cancer were later found to also occur in small towns.

**Availability bias**: availability bias is where the data that is currently available is used even if the question being asked requires other data. The data that is not “available” can often be much harder to acquire.

For example, surveys on the effectiveness of internal recruitment invariably conclude that those appointed are successful but rarely investigate the success of the 2nd placed, rejected candidate as a benchmark.

Another example is a study of bullet holes in World War II bomber planes returned from operations to inform where to make structural re-enforcements.

The patterns of bullet holes in the bombers that were shot down and did not make it back were never considered. Presumably obtaining this data would have involved retrieving wreckage from bombers on the ocean floor (Bragdon, n.d.).

|  |
| --- |
| A picture containing ax, table  Description automatically generated |
| *Figure 10 – Availability Bias of Bullet Hole Patterns in Surviving Planes* (Bragdon, n.d.) |

**Short-termism**: this is a short-term focus on immediate out-comes and overlooking the medium and long term considerations.

Pressure from managers and a need for an immediate public relations boost is attributed with over-looking what in hindsight were clear safety warnings in the Challenger Space Shuttle disaster (Cambridge (1), 2019).

##### Collective Biases and Cognitive Diversity

Homogeneity (when a group is made up of similar individuals) and homophily (seeking out similar individuals) in groups, teams and organisations can lead to collective blindness in decision making.

Even if individuals have avoided decision biases their groups and teams can still be caught by biases, blindness and poor decision making (Reynolds & Lewis, 2017), (Syed, 2019).

It has been proposed that homogeneity within the CIA was a significant contributory factor to the failure to join the dots necessary to have anticipated the terrorist attacks of 9/11 even though the CIA had some of the best analysts in the world and access to good data (Syed, 2019).

The theory can be visualised as follows –

|  |  |
| --- | --- |
| Image result for rebel ideas matthew syed an intelligent individual" | Image |
| Image | |
| *Figure 11 –Cognitive Diversity and Homogeneity in Problem Solving* (Syed, 2019) | |

In the 1st figure David is an intelligent individual, but the problem is complex and the problem space (represented by the rectangle) is large so David only understands part of it.

In the 2nd figure David is part of a team of clever individuals but because they are homogenous their perceptions of the problem overlap; hence these clever individuals are part of an unintelligent team.

The team in the 3rd figure has the same number of members as in the 2nd but because this team is made up of cognitively diverse individuals, they cover much more of the problem space than the unintelligent team.

In order to avoid collective biases arising from interpreting data analytics the groups analysing and deciding should be diverse.

Diversity in this context does not relate to age, ethnicity, gender or other protected characteristics, rather to the way individuals process knowledge and have different perspectives which is cognitive diversity (Reynolds & Lewis, 2017).

The implication for decision makers interpreting and acting on data analytics is to build cognitively diverse teams or to seek the views and opinions of others who are cognitively diverse.

##### Triangulation

|  |
| --- |
|  |
| *Figure 12 –The Importance of Triangulating Sources of Information* (Holden, 2019) |

In his “Determined to Lead” seminar in October 2019, Trevor Holden - Managing Director of Broadland and South Norfolk Councils - highlighted the criticality of triangulating sources of information before acting (Holden, 2019).

The University of Cambridge Business Analytics course frequently re-enforces the view management intuition is a valid tool to use in shaping decision making whilst emphasising that data analytics should be used to triangulate that intuition. If the data view matches then it re-enforces the intuitive view and if it differs then more analysis is required (Cambridge (1), 2019).

When using data analytics to triangulate intuition and other sources it is critical to avoid individual and group biases to ensure the objectivity which will drive good decision making and lead to optimally improved outcomes.

### Can a mature data analytics capability add value to firms?

The definition of “firm” is assumed to be sector agnostic and not specific to the education sector. The theme of sector-specific value will be explored in the following sections.

In an often-quoted report by Bain & Company (Wegener & Sinha, 2003) the authors carried out research across over 400 firms to evaluate the relative performance of companies who were “top performers” in terms of analytics capability which they define along 4 axes. The report is summarised in Figure 13 –

|  |
| --- |
|  |
| Figure 13 - Axes of Analytics Capability (Wegener & Sinha, 2003) |

The report presented detailed evidence that top performers in big data and analytics out-perform the others along several axes –

|  |
| --- |
|  |
| Figure 14 – Top Performers Decision Making Capability (Wegener & Sinha, 2003) |

The research then went on to quantify the extent to which firms with the best analytic capabilities outperform the competition (see Figure 15).

(Wegener & Sinha, 2003) present an overwhelming case for the positive impact that data analytics capabilities have on company performance and state that opportunities exist in almost every industry and that competitive-edge is not limited to technical or data-sensitive industries.

However, there are some aspects of the report that should be challenged –

* Bain & Company are a firm that deliver data analytics consultancy hence there could be confirmation bias or homogeneity bias in the report.
* The report was written in 2003 hence things could have changed in this fast-paced sector.
* The methodology used for the research and the choice of firms for the survey is not known and hence cannot be validated.

|  |
| --- |
|  |
| Figure 15 – Quantification of Effectiveness of firms with the Best Analytic Capabilities (Pearson & Wegener, 2013) |

A more recent view is the McKinsey online global survey on data and analytics. The survey defines a high-performing company as “one that, according to respondents, has seen annual rates of organic revenue growth and growth in earnings before interest and taxes of 10 percent or more in the past three years” (McKinsey & Company (1), 2019).

Respondents from these high-performing organizations are three times more likely than others to say that their data and analytics initiatives have contributed at least 20 percent to earnings before interest and taxes (EBIT) over the past three years (McKinsey & Company (1), 2019).

The report is also useful in describing the pace of recent change relating to industry competition brought about by data and analytics (see Figure 16).

The literature review identified a lot of corroborating evidence for the positive aspects of investing in data analytics but it is more challenging to find a balanced view and very hard to find any literature which suggests that an investment in data analytics produces any negative effects which suggests that the potential for data-biases to exist.

One piece of research that presents a more balanced view is “Betting on Big Data” (Forbes Insights, 2015) and shown in Figure 17. The title suggests that big data is a gamble and some of the findings go on to suggest a more conservative impact.

For example, if “two-thirds of respondents report that big data and analytics initiatives have had a significant, measurable impact on revenues” then one-third have observed no impact.

The findings also suggest a relatively modest impact for the two-thirds and challenges to what extent causality can be proven between investing in analytics and the improvements the firms observed. Perhaps those firms were on an upward trend anyway or it was other internal initiatives that had causality with or contribution to the positive performance.

|  |
| --- |
|  |
| Figure 16 – Recent Changes in the Nature of Industry Competition Brough About by Analytics (McKinsey & Company (1), 2019) |

|  |  |
| --- | --- |
|  |  |
| Figure 17 - Effect of Big Data and Analytics on Revenues and Costs (Forbes Insights, 2015) | |

Another source presenting an alternative view is “When Data Creates Competitive Advantage” (Hagiu & Wright, 2020) which suggests a set of questions to determine the potential degree and sustainability of competitive advantage inferred by data-enabled learning –

1. How much value is added by data?
2. How quickly does the marginal added value drop off?
3. How fast does the relevance of the data drop off?
4. Is the data proprietary (i.e. difficult for competitors to obtain or replicate)?
5. How hard is it to imitate the improvements that are based on the data?
6. Does data for one customer help improve the experience for others?
7. How fast can data-driven insight be incorporated into the business?

“Competing on Analytics” (Davenport & Harris, 2017) conducted a review of firms by rating their analytical maturity against their financial performance concluding that there was a “significant correlation” –

|  |
| --- |
|  |
| Figure 18 – Importance of Analytical Orientation in High vs. Low Performing Firms (Davenport & Harris, 2017) |

Research by the International Data Corporation (Morris, 2003) found a return on investment for analytical projects as follows –

* 277% for projects improving production.
* 139% for projects improving financial management.
* 55% for projects aimed at customer relationship management.
* 145% for predictive projects (those using data to make predictions about the future).
* 89% for descriptive projects (those looking at existing or historical data to drive actionable insight).

Nucleus Research, a company that researches IT value, reported in 2014 that for every dollar a company spends on analytics, it gets back $13.01 – 1.2 times more than they got just three years ago (Nucleus Research, 2014).

In the search for a practical, concrete example of data analytics adding value to firms, the literature on the Netflix Prize provides valuable insight.

The Netflix Prize challenged participants to use data analytics to optimise the way the firm made movie recommendations to subscribers and can be summarised as follows –

|  |
| --- |
|  |
| THE NETFLIX PRIZE TIMELINE  NETFIIX  Oct 08 2006  WXYZ Consulting was the first to beat NetflixS algorithm  Oct-Dec 2006  ML at University of Toronto A  May 2007  BellKor: wins the first year's progress prize of S50000 by  beating Nettlix% algorithm by 8.5%  sep 2009  BellKorS Pragmatic Chaos: wins the SIM prize (This team  identically on theTestSet as their competitor,  but submitted their algorithm 20 minutes earlier)  20071  2008  Oct 02 2006  Netfiix announces anyone who can improve their  movie recommendation algorithm by 1096 will win $1 M  Nov-Dec 2006  WXYZ Consulting  Jan-May 2007  Gravity  2008  BellKOr in Big Chaos: Wins the second years  prize Of |
| Figure 19 – Netflix Prize Details and Timeline (Cambridge (5), 2019) |

The competition ended on 26th June 2009, having received submissions from 5,169 teams. The $1m prize was awarded to a team called “BellKor's Pragmatic Chaos” who had achieved a 10.1% improvement over NetFlix original algorithm.

The prize money cost Netflix $1m. However, in the year following the launch of the new algorithm the share price increased by 292% and the market capitalisation increased by 264% or $3.85bn to $6.19bn (Figure 20 and **Error! Reference source not found.**). The prize money was just 0.0003% of the increase in market capitalisation.

|  |
| --- |
|  |
| Figure 20 - Change in Netflix market capitalisation (Macrotrends, 2020) |
|  |
| Figure 21 - Change in Netflix share price (Google Nasdaq, 2020) |

In summary there is a wide body of literature suggesting that data analytics can add value to firms across many sectors. Firms making high-end use of data analytics are more likely to be financial high-performers and will execute decisions more quickly and as intended and will create competitive advantage.

There are some sources that temper that view but even the more conservative sources suggest that having a data analytics capability adds value and enhances financial key performance indicators like share price and market capitalisation.

### Can data analytics add value in education?

The question about whether data analytics can add value to the education sector specifically is more difficult than considering its application to firms generically, primarily because there is much less literature and evidence available.

To illustrate this point, the McKinsey global survey presents a matrix of sector vs. function to highlight the areas where data analytics is having the most (and least) impact –

|  |
| --- |
|  |
| Figure 22 – Changes to Industry Practices through Data Analytics (McKinsey & Company (2), 2017) |

Education does not feature as one of the sectors that are being most heavily impacted. The detail of each sector was not included in the survey but even if the assumption is made education is part of “Public, social sector” it is still one of the sectors with the lowest impact.

There is a scarcity of sources describing where data analytics has positively impacted the UK education sector but there are many sources that discuss the high potential for future use and several sources for current and historic use in the United States of America.

One example describing the future potential is a recent post on the FE Week online news site where John Gray (Director of Further Education at EMSI) makes a case for colleges to embrace big data to enhance organisational outcomes.

(Gray, 2019) makes the case for using big data and analytics for colleges to develop an understanding of local employers and local skills need and that big data should be used because directly engaging with all local employers would be a “Sisyphean task” (i.e. endless and ineffective).

The case is then developed to suggest how the FE sector could use data analytics to shape curriculum and skills design to ensure that their outputs are aligned with employer and curriculum demand which also links to a recent Ofsted requirement for educators to demonstrate the “intent of the curriculum” (Ofsted, 2019), (Gray, 2019). EMSI’s core product is economic data for further education, hence further triangulation of this view is required.

In another FE Week post Richard Alberg makes a more rounded case for introducing big data and analytics into Further Education. Alberg states that FE is yet to embrace data analytics but that potential does exist across many functions including recruitment, retention, student completion, career choice, course innovation, business development and curriculum design (Alberg, 2019).

(Alberg, 2019) also describes existing examples of the use of publicly available apprenticeship data to overlay apprenticeship delivery and market opportunity and a view that the FE sector is on the threshold of using machine learning and artificial intelligence to deliver competitive advantage.

A research-based source was identified in the “Handbook of Contemporary Education Economics” which dedicates a full chapter to “Data Analytics and Decision-Making in Education” (Agasisti & Bowers, 2017).

(Agasisti & Bowers, 2017) suggests the combination of private, internal data and publicly available data sets to create big data that can leverage new insight and this triangulates the views expressed by (Gray, 2019).

The projects and case studies detailed by (Agasisti & Bowers, 2017) are all US based showing that the use of data analytics in education is much more mature and widespread in the US than in the UK.

The research demonstrates that applying analytics to education has the potential to add value but that the UK significantly lags behind the US in this respect.

For example, in the US the Harvard University Strategic Data Project (SDP) has created a range of programs with the aims of developing leaders, building and sharing analytic tools and placing analytic resources in educational institutions (Harvard University (1), 2020).

The SDP cites numerous completed projects delivering tangible outcomes that include –

* Improving inclusion and diversity.
* Improving recruitment.
* Informing turn-around strategies.
* Measuring teacher impact to develop insights into improving learning.
* Measuring societal impact.
* Improving student perceptions.
* Tracking progress of early reading skills.
* Accurately identifying learners who need additional support.
* Evaluating and comparing educational institutions with significant demographic differences.

Each of these projects includes an overview and detailed report of the challenge, intervention and impact providing detailed evidence that data analytics is being applied in a systematic and organised way to add genuine value to the US education sector (Harvard University (2), 2020).

In another US source (Holt & Schools, 2015) wrote about a project that applied data analytics techniques including composite indicators, regression and tree induction to predict student success factors at Grade 3. The outcomes included –

* Identification of a significant correlation between pre-Kindergarten (pre age 5) classes and Grade 3 success suggesting that policy makers need to invest much more in early learning.
* Development of a dashboard that can be used to monitor predicted and actual student progress that could be used to direct intervention and support.

(Holt & Schools, 2015)

In another industry-based source Loginworks are a “13+ year old premium data analytics and business intelligence” business based in Virginia, US.

They have blogged on the “Top 10 Advantages of Data Analytics for [the] Education Industry”, some of which have not been suggested in the other sources, or have been presented in a different light –

* Assemblage, aggregation, comparison and tracking of very large datasets.
* Tailoring an intelligent curriculum based on individual learning ability, approach, preference and performance.
* Advanced attendance data including patterns of library utilisation etc.
* Predicting “at risk” students to efficiently target intervention and support.
* Predicting and suggesting successful options for further study and career choice.
* Identifying and distilling best teaching and learning practice.

(Radhika, 2018)

In summary the lack of literature sources describing current and tangible value derived from data analytics in the UK education sector demonstrates that its use is not widespread. Much of the literature discusses future potential and opportunity rather than current or historical success.

However, there is evidence in support of the successful application of data analytics in the education sector in the United States.

### Why is data analytics not more commonly used in UK education and what are the potential barriers?

The conclusion so far is that application of data analytics can add value to firms across diverse sectors but that for the UK education sector there is opportunity and potential but not widespread uptake.

This raises the need to address a supplementary question as to why data analytics is not used in the UK education sector and what the potential barriers to adoption might be.

#### Legal Considerations

The key legal consideration within the UK and EU is the General Data Protection Regulation (GDPR) (Gov.uk (1), 2018) which applies to the data of all EU citizens and any companies processing their data.

GDPR has strict rules called data protection principles used to ensure that information is:

* Used fairly and lawfully.
* Used for limited and specifically stated purposes.
* Used to determine how the data owner uses the information.

A much wider range of data (for example computer IP addresses) is now considered to be “personal”, consent to use must be obtained by affirmative action and customers have wide-ranging legal rights relating to data accuracy and data portability.

Also, under GDPR regulation the maximum penalty for a firm has increased to 4% of global annual turnover or €20 million – whichever is higher.

Such high penalties and the limitations and obligation of firms may be a barrier to entry into the area of big data and data analytics for the UK education sector (Cambridge (9), 2019).

#### Ethical and Moral Considerations

A firm’s moral and ethical approach should go further and deeper than the legal obligations which should be considered as the minimum because behaviour that is legal but still unethical may drive customers away and damage company reputation (Cambridge (9), 2019).

For example, in 2000 Amazon used big data to offer differential pricing to individual customers. This is a clear example of first-degree price discrimination where the seller of a product is able to charge each customer the maximum price they are willing to pay (Sloman, et al., 2016).

Due to a customer backlash the Amazon CEO made a formal press release apologising for the mistake and thousands of customers were refunded to help rebuild corporate reputation.

(Agasisti & Bowers, 2017) also propose ethical and moral considerations that are education specific describing how triangulation of data with students’ personal characteristics, whilst driving value, raises serious privacy concerns.

Education institutions and the Further Education sector set a high bar for ethical and moral considerations hence these may also have formed a barrier to entry.

#### Cyber Attacks and Cyber Security

Having an increased reliance on big data and data analytics increases the potential impact of a cyber-attack (Cambridge (9), 2019).

In 2015 TalkTalk was subject to a cyber-attack that affected 157,000 users with 15,600 sets of customer bank records and 28,000 credit and debit card numbers were stolen.

As a direct result of the attack TalkTalk shares lost 1/3rd of their market value and they were fined £400k under Data Protection regulations. If the attack had happened post GDPR the maximum fine could have been £59m (BBC, 2015), (Cambridge (9), 2019).

The increased impact of cyber security could be a barrier to entry but very few FE colleges have been attacked (or published the details) and the sector does not appear to be a target for organised cyber crime.

#### Cost and Affordability

(Alberg, 2019) and (Agasisti & Bowers, 2017) refer to cost as a potential barrier to entry into data analytics for education.

(Agasisti & Bowers, 2017) go on to propose that that a data analytics capability requires a socio-technical system consisting of experts, data scientists and hardware and software, all of which require substantial investment.

Publicly available data shows a trend of falling income and reduced expenditure within the FE sector. Hence it can be reasonably concluded that the sector is caught in a wave of cost reduction and increasing budget deficits (see Figure 23).

In contrast an investment in a data analytics capability would be an investment in innovation which has some special considerations when developing a case –

* Investment in innovation requires a fundamentally different approach to traditional cost cases based on Net Present Value.
* It is nearly impossible to predict with any accuracy which innovations are likely to succeed and which are not.

(Gunther-McGrath, 2008).

It is likely that innovation projects like data analytics are likely to be deferred or over-looked in a sector in a severe and long-term financial down-turn and that cost and affordability are barriers to entry.

#### Complexity

(Alberg, 2019) and (Agasisti & Bowers, 2017) discuss complexity as a potential barrier to adopting data analytics.

(Agasisti & Bowers, 2017) proposes that technical complexity arises from the need for highly developed data integration that is required for reliability, validity, accuracy and accessibility of data.

(Agasisti & Bowers, 2017) go on to discuss whether educational institutions will have the necessary competencies or be able to solve this complexity and propose a list of competencies that would be required –

* Data expertise.
* Analysis expertise.
* Evaluation competencies.
* Visualisation / communication of results.

Clearly the complexity of developing or acquiring a non-core data analytics capability may be a barrier to adoption of data analytics by the UK educational sector because data analytics will be a non-core competency.

|  |  |
| --- | --- |
|  |  |
|  |  |
| Figure 23 – Financial Performance of UK FE Sector 2014 - 2017[[2]](#footnote-2) (ESFA, 2019) | |

#### Competitive Strategy

Several frameworks and sources were used to explore the potential for competitive strategy to act as a barrier including Porter’s 5 Forces (Figure 24), data on market demographics (Figure 25) Kotter’s 8 Step Change Model (Figure 26).

|  |
| --- |
|  |
| Figure 24 – External Industry Analysis using Porter’s 5 Forces (Porter, 2008) |

|  |
| --- |
|  |
| Figure 25 – Gradually Declining Population Demographic (Office for National Statistics (1), 2020) |

FE Colleges, Secondary Schools, Sixth Forms and Apprenticeship Training Providers are competing over a gradually declining customer population / demographic (Figure 25).

However, the highly regulated nature of education makes it very difficult for new entrants to effectively threaten the core market and until recently FE colleges could not legally become insolvent (Association of Colleges, 2019).

This market protection could create an absence of “urgency for change” (Kotter, 2018) contributing to low take-up of innovation like data analytics.

|  |
| --- |
| A screenshot of a cell phone  Description generated with very high confidence |
| Figure 26 – Kotter’s 8 Steps (Kotter, 2018) |

Based on these analyses it is highly likely that competitive strategy, market forces and external industry conditions have created a historic barrier to entry.

### What are the general and sector-specific issues relating to acquiring or building a data analytics callability?

In “Competing on Analytics” (Davenport & Harris, 2017) propose a road-map for becoming an analytical competitor whilst “Data Science for Business” (Provost & Fawcett, 2013) discusses and details the challenges of attracting and retaining world-class data scientists.

Turning to practical sources from industry professionals Monica Rogati, Data Science Advisor and former VP of Data, posted a view on Quora.com which is comprehensive and cohesive –

* **Goals**: why do you need a data team and what will the team’s immediate goals be?
* **Roles**: what roles will be required within the team and what will the composition of those roles be?
* **Leadership**: has a strong leader been identified who can attract and recruit the team, define the strategy and work with the founders or shareholders?
* **Timing**: when is the best time to start building the team and in what order?
* **Organisation**: how will the data team fit into the organisation?
* **Data culture**: does the company understand data analytics, and will it prioritise it?
* **Infrastructure**: does the company have the necessary infrastructure to support a data team? If not what skills, tools, and capabilities are needed to provide it and how long will it take?
* **Recruiting and retention**: how will talented data scientists be attracted in an environment of high demand and skills shortages?

(Rogati, 2017)

Some of the key issues highlighted in these sources will now be explored in more detail.

#### Recruitment and Retention

Recruitment of data analytics professionals will certainly be an issue for firms generally and more specifically for the FE sector and the Lincolnshire area.

Figure 27 shows how data science and analytics skills and roles compare on the axes of “projected growth” and “hardness to fill” –

|  |
| --- |
| Analysis of Jobs and Skills Demand |
| Figure 27 – Data Science and Analytics Skills and Jobs (Forbes, 2017) |

The roles that are key to a data analytics capability include “Data Scientist”, “Data Engineer” and “Business Intelligence Architect” and these are in the top right-hand quadrant.

The same pattern exists for the required skills. “Machine Learning”, “Big Data”, and “Data Visualisation” are all in their top right-hand quadrant showing that they are the fast growing skills that are the hardest to find.

The salaries required to attract the required roles is also a potential issue. The analysis in Figure 28 from (Glassdoor, 2020) shows the average London salaries for the main data analytics roles –

|  |  |
| --- | --- |
|  |  |
|  |  |
| *Figure 28 – Data Analytics Jobs and Salaries in London* (Glassdoor, 2020) | |

Another potential issue is that Glassdoor reports no job postings for any of the 4 key roles in the Lincoln area. The nearest location where they can be found is Nottingham and although there are fewer roles available than in London the salaries are similar.

This evidence shows that there is a combination of growing skills demands, job vacancies that cannot be filled, high salaries and an absence of jobs in the Lincoln area.

There are further challenges that are specific to the FE Sector as shown in Figure 29 –

|  |
| --- |
| A screenshot of a cell phone  Description automatically generated |
| A screenshot of a cell phone  Description automatically generated |
| *Figure 29 – Median Salaries and Ages of Staff in the Further Education Sector* (Frontier Economics, 2017) |

The median gross salary for technical staff in Further Education in 2017 was £20,000 in 2017 (Frontier Economics, 2017) and there have only been three 1% pay-rises since the report was published. The median age of staff in Further Education was 50-54.

This suggests further challenges that Lincoln College will face in the recruitment of data analytics professionals. Data analytics average salaries are more than twice the FE sector median pay for technical staff and the age profile also suggests a poor fit as most data analytics professionals are recent graduates.

The challenges may not end there. In “Data Science for Business” (Provost & Fawcett, 2013) state that “certain [data analytics] individuals have the combination of innate creativity, analytical acumen, business sense, and perseverance that enables them to create remarkably better solutions than their peers”. This suggests that the pool of top talent is very small and very difficult to access.

#### Understanding the Data Analytics Roles

Understanding the role definitions that are required is another area where the available literature has many differing and often contradictory views.

However, (Saltz & Grady, 2017) state the case for a common understanding of workforce descriptions -

*“The main motivation for workforce descriptions is the need to identify, recruit, train, develop, and maintain an appropriately skilled workforce by providing a common language to categorize and describe the type of data science work that needs to be done”*

(Saltz & Grady, 2017).

(Saltz & Grady, 2017) go on to explore case studies of standards organisations, industry organisations and an advisory & consultancy firm and go on to produce clear, visual representations of the roles and expertise associated with data analytics -

|  |  |
| --- | --- |
|  |  |
| Figure 30 – Data Analytics Roles and Team Expertise (Saltz & Grady, 2017) | |

The following are the author’s definitions of the key roles based on a synthesis the findings of (Saltz & Grady, 2017) with practitioner literature sources, for example the (European Leadership University, 2020) blog –

* **Data Engineer**: takes the data questions from the data scientists and then locates, curates, extracts, translates, joins and cleans the data to make it accessible and available for the data scientists to use in their models.
* **Data Scientist**: translates business problems provided by data analysts into data questions. Liaises with data engineers to obtain the required data. Creates advanced data models to answer the data questions and liaises with the data analysts to understand the findings.
* **Data Analyst**: uses the data and models obtained by the engineers and analysts to create reports, visualisations and “analytics stories” (EdX, Microsoft, 2020). Uncovers insights and headlines hidden in the data and translates and communicates the findings to managers and decision makers.

There is no intention to infer that the roles must be fulfilled by different people. In fact, the absence of consensus in the literature suggests that the roles are commonly mixed together with individuals who possess over-lapping skill-sets often fulfilling two or more of the three key roles.

# Chapter 3: Findings and Objectives

This chapter identifies, extracts and synthesises the key findings from the literature review and defines a set of objectives that need to be met in order for the introduction of a successful data analytics capability at Lincoln College.

## Key Findings

### Firms that possess a mature data analytics capability will have a competitive advantage over those that do not

There is a significant body of evidence supporting the view that a strong data analytics capability positively impacts company performance –

* (Wegener & Sinha, 2003) show data analytic firms are twice as likely to be top-quartile financial performers.
* (McKinsey & Company (1), 2019) show data analytic firms increase their earnings before tax by 20%.
* (Forbes Insights, 2015) show that two-thirds of data analytic firms deliver significant, measurable impact on revenues.
* (Davenport & Harris, 2017) show a link between higher levels of analytical maturity and robust five-year compound growth rates.
* (Morris, 2003) shows a 277% return-on-investment for analytics projects focused on improving production.
* (Nucleus Research, 2014) show that every dollar spent on analytics returns $13.01.

|  |
| --- |
|  |
| Figure 31 – Link Between Analytic Capabilities and Financial Performance (Pearson & Wegener, 2013) |

### There is a gap in the market for provision of data analytics capabilities to the UK education sector that is not currently being exploited.

Despite the evidence supporting the potential competitive advantage conferred by data analytics there is no evidence demonstrating its adoption or uptake within the UK Further education sector.

Most sources evaluating the impacting on different sectors do not even include the education sector. For example, the McKinsey Global Survey (McKinsey & Company (2), 2017) evaluates the impact of analytics on 10 key industries but does not mention education.

The application of data analytics to the education sector is mature in the United States as can be evidenced by the Harvard Universities Strategic Data Plan (SDP). The aims of the SDP are developing leaders, acquiring and sharing analytic tools and placing analytic resources in educational institutions (Harvard University (1), 2020).

There are many specific examples of tangible outcomes in the US including the correlation shown in Figure 32 used to inform policy makers that strengthening pre-age 5 kindergarten participation will have a direct and positive impact on improving academic outcomes of school children later in life.

|  |
| --- |
|  |
| Figure 32 – High Performance of the “Composite Success Factors” Analytics Technique (Holt & Schools, 2015) |

Figure 32 is a detailed scientific data chart taken from the US study; the key take-aways from this chart are –

1. There is a clear correlation between pre-age 5 kindergarten participation and later academic outcomes.
2. In the US it is common to see advanced data analytics techniques applied to education.

Figure 33 provides further evidence from the US. This is a documented example of the machine learning technique known as “tree induction” being used to build a model to make future predictions about academic outcomes.

|  |
| --- |
|  |
| Figure 33 – Using Tree Induction to Predict Grade 3 Student Performance (Holt & Schools, 2015) |

In contrast the UK focus is on what could potentially be achieved in the future rather than what is actively being worked on and developed in the present. However, the range of potential benefits and uses that have been suggested are significant including –

* Understanding local employers, local skills need and labour market intelligence to shape curriculum and drive curriculum “intent” via -
  + Recruitment
  + Retention
  + Completion
  + Career choice
* Using this capability to demonstrate to Ofsted that education providers align curriculum intent with local employer and economic needs

(Gray, 2019)

* Improving teaching effectiveness
* Optimising learning outcomes
* Predicting student progress and performance
* Improving student achievement
* Improving decision making

(Alberg, 2019)

* Tailoring an intelligent curriculum based on individual learning ability, approach, preference and performance.
* Advanced attendance data including patterns of library utilisation etc.
* Predicting “at risk” students to efficiently target intervention and support.
* Predicting and suggesting successful options, further study and career choices.
* Identifying and distilling best teaching and learning practice.

(Radhika, 2018)

* Improving inclusion and diversity
* Improving recruitment
* Informing turn-around strategies
* Measuring teacher impact to develop insights into improving learning
* Measuring societal impact
* Improving student perceptions
* Tracking progress of early reading skills
* Accurately identifying learners who need additional support
* Evaluating and comparing educational institutions with significant demographic differences

(Agasisti & Bowers, 2017)

* Identifying, managing and evaluating policy interventions
* Improving outcomes of the macro-level education system

(Harvard University (2), 2020)

Using labour market intelligence to shape curriculum intent (Gray, 2019) is of significant interest to schools and colleges. The new Ofsted Education Inspection Framework gives specific direction that “the provider’s curriculum is coherently planned and sequenced towards cumulatively sufficient knowledge and skills for future learning and employment” (Ofsted, 2019). Curriculum intent was introduced in the May 2019 Education Inspection Framework and all providers will be concerned about how to respond to this clear and regulated directive.

Despite clear evidence that data analytics can add value to firms generally and many examples of the potential value for the UK education sector there is scarce evidence of the UK Further Education sector currently using data analytics capabilities to drive improved outcomes.

The evidence for the positive impact of data analytics on firms generally, the existing applications to education in the US, the absence of current application within UK education sector and the wealth of potential future applications all suggest a significant potential gap in the market for provision of these capabilities.

### Legal, moral, ethical, cost and complexity issues all represent significant challenges to the acquisition of data analytics capabilities within education, but they are not preventing it.

There are considerable legal considerations to leveraging data, specifically the introduction in 2018 of the General Data Protection Regulation (GDPR) which introduced far stricter controls on firms using customer data, with associated increases in penalties and fines of up to 4% of turnover or €20 million, whichever is higher (Gov.uk (1), 2018).

However, the educational sector is already familiar with GDPR compliance for many aspects of its delivery and many firms in other sectors successfully leverage data analytics whilst operating within the GDPR legislation.

Ethical and moral considerations should go beyond GDPR and be more rigorous and the education sector places a high value on these considerations.

(Cambridge (9), 2019) outline and then discuss four principles which form the basis of an approach to addressing the legal, ethical and moral considerations –

1. Transparency – are customer aware of how their data is going to be used?
2. Control – customers must default out and opt into a service and be in control throughout.
3. Face validity – does what you are doing make sense to customers? i.e. using Facebook profile data from 10 years ago to inform credit rating may not make sense.
4. Customer benefit – do customers genuinely feel that the service is of benefit to them and not just delivering profit or other outcomes for the business?

The conclusion drawn in this research is that whilst legal and regulatory concerns are genuine and significant the application and adherence to the four principles will ensure not only compliance but clear water between a firm’s position and its legal obligations.

Legal, moral or ethical considerations, whilst posing some challenges, are not absolute barriers to the acquisition of data analytics capabilities by the UK education sector if they are acknowledged, understood and addressed.

In examining cost as a barrier-to-entry, the macro-economic climate within the FE sector has been shown to be contributing to a lack of adoption with (Alberg, 2019) and (Agasisti & Bowers, 2017) triangulating this view.

Certainly, the highly challenging budget circumstances will reduce the investment available for innovation projects in favour of cost-reduction initiatives, bottom-line protection and survival.

In addition to cost the complexity of data analytics could also be playing a part in the lack of take-up however there are clear instances of colleges working with outsourced partners to invest in other types of technology that confer competitive advantage.

For example the blended learning consortium uses a subscription model with over 50% of UK colleges paying £5k each to generate £400k per year for the production and distribution of shared blended learning resources (Blended Learning Consortium, 2016).

In conclusion legal, moral, ethical, cost and complexity issues represent significant challenges to the acquisition of data analytics capabilities within education. These factors are likely to have been barriers-to-entry that will need addressing in any solution or proposal.

### Government legislation and regulation provide protection to state-funded education, reducing competitive forces and providing dis-incentives for innovations like data analytics

At a macro level, external market forces are playing a part in the slowing of innovation initiatives in the UK education sector.

The FE sector is in a cycle of financial decline as shown in Figure 23 but legal and regulatory frameworks make it very difficult for new entrants to threaten core education markets.

Despite the new insolvency regime (Association of Colleges, 2019) and declining numbers of colleges (ESFA, 2019), government under-writing of the sector provides a large degree of protectionism.

When the threat of new market entrants, substitutes & alternatives and industry rivalry are all low the effect will be to reduce the motivation for implementing transformations and innovations like data analytical capabilities.

### MOOCs are education sector disruptors who are actively using data analytics and complacency by incumbents could be causing declining recruitment in their adult, higher education and full cost markets.

There are clear warning signs that new market entrants are having an impact on further education and that these new entrants are using data analytics and other technologies to disrupt established markets.

Figure 35 shows the “long-tail” of reducing innovation for incumbent products and shows how invading products are initially out-competed by incumbents before ultimately disrupting and taking over (Utterback, 1994) –

|  |
| --- |
| Image result for utterback product process lifecycles non assembled products |
|  |
| Figure 35 – Incumbent vs. Entrant Product Innovation (Utterback, 1994) |

Also (McKinsey & Company (2), 2017) shows that 39% of “digital native” firms attribute data analytics to 20%+ of their pre-tax earnings compared to 7% of incumbents (see Figure 36).

|  |
| --- |
|  |
| Figure 36 – An interactive look at organizations’ use of data and analytics (McKinsey & Company (2), 2017) |

There are many new entrants into the “Massive Open Online Course” (MOOC) and digital education space where providers deliver a wide range of high-quality education using low-cost, accessible, online business models including –

* Khan Academy – online courses in Maths, Science, Economics, Finance, Arts, Humanities and Computing, free with donation requests (Khan Academy (2), 2020).
* edX – 2500 online courses from 140 institutions, a wide array of subject areas, usually free with optional payments for recognised certification (EdX, 2020).
* Coursera – online courses from 190+ leading universities and companies (Coursera, 2020).
* LinkedIn Learning - over 14,000+ online courses taught by real-world professionals, corporate subscription (LinkedIn Learning, 2020).
* Open University – Long standing provider of online undergraduate and post graduate degrees, full cost (Open University, 2020).
* The Digital College – provision of low cost, 100% online Further Education courses in Health & Safety, Construction, Healthcare and Safeguarding (The Digital College, 2020).
* InterHigh – A leading UK provider of 100% online primary, secondary and 6th form education (InterHigh, 2020).

Given some of the key uses for data analytics like customised marketing and recommendations (Cambridge (3), 2019), these new entrants are already placing data analytics capabilities front-and-centre of their offerings and business models.

(Utterback, 1994) developed many case studies to demonstrate the impact of innovation and invasion on incumbent firms and products. One example is the advent of electrical lighting in 1877 which initially struggled to compete with the established gas-lighting providers. Typically, the incumbents respond to such threats by improving their products and by blocking the innovation politically, but innovation history is littered with examples of the incumbents’ eventual demise.

There are clear parallels with the current state of play in further education which suggests that MOOCS and providers of technology-based curricula stand poised to threaten traditional providers if they do not respond.

There is no empirical evidence to quantify the timing or the financial impact, but adult education, higher education and full cost markets are those that best lend themselves to digital delivery supported by a data analytics capability and it is likely that tailing recruitment in these areas is linked to disruption.

If the historical patterns are repeated it is not only possible but likely that the adoption of data analytics and other technology solutions could be a matter of long term survival as well as medium term optimisation of outcomes for further education.

### Colleges are good at FE specific information services but do not systematically develop data into intelligence and knowledge. There are some centralised initiatives, but to date they are slow to deliver and are specific to Universities.

It is useful to re-challenge the assumption that the UK focus is all on future potential rather than current capability.

Certainly Further Education (FE) institutions do have mature Information Services (IS) departments, the key purpose of which is to collect, store and process data into an electronic report called an Individualised Learner Return (ILR) (Gov.uk (2), 2020).

The regular submission of the ILR to central government drives the calculation, generation and distribution of income for the courses delivered. Without a good IS capability FE institutions would not generate any income, hence the IS capabilities tend to be mature and effective.

FE IS skills have limited transferability outside of the sector because knowledge of government and ILR funding rules are highly complex and sector specific. Expert practitioners take years to develop and maintain the relevant skills and can then be “locked in” to the sector.

In this context heavily FE-specific information services capabilities flourish but more recent data analytics and data science capabilities do not.

FE IS teams are often criticised internally for being data-rich but knowledge-poor. In the visualisation of the knowledge ladder seen in Figure 37, further education IS capabilities are commonly considered to be operating towards the bottom left corner where the top right had corner is the space that can confer the most value and the best outcomes.

|  |
| --- |
|  |
| Figure 37 – The Knowledge Ladder; adapted from (Gottschalk, 2013) |

Looking beyond local, internal capabilities the Joint Information Systems Committee (JISC) and the Higher Education Statistics Agency (HESA) have collaborated on “Analytics Labs” where volunteers from education come together to work on data analytics problems that will then be rolled out to member institutions (JISC, 2020).

To date the results have been of limited value. The projects have taken a long time to deliver, the deliverables are relatively static, and they use traditional descriptive rather than predictive analytics.

Also, all the deliverables to date have been specific to Universities and Higher Education with nothing produced for Further Education colleges.

## Objectives

### Raising of Management Awareness, Understanding, Confidence and Buy-in

It has been established that FE colleges are caught in a cycle of financial decline and that such a climate will provide a disincentive for investment in innovation initiatives. However, due to its diverse portfolio of international business Lincoln College could be poised to exit from that cycle and the timing for making an innovation investment case is fortuitous.

At the beginning of the literature review and investigation the Lincoln College management team did not have an wide awareness of data analytics or an understanding of its potential and according to (Cambridge (9), 2019) the following are the two most significant barriers to the adoption of data analytics initiatives –

1. Inconsistent senior management understanding and buy in
2. Inconsistent and generally low middle management understanding and buy in

Also, the Kotter 8 Step change model presented in Figure 26 shows that successful change initiatives require the creation of a climate for change which in turn requires the creation of a guiding coalition which must include senior stakeholder support.

A clear objective therefore is to raise awareness, understanding confidence and buy-in of the executive leadership team and the wider college management team.

Without delivery of this objective it is unlikely that a data analytics capability will gain the buy-in, aupport and approval necessary to succeed.

### Adopt an Innovation Mindset

The nature of a data analytics capability within education is that it is complex, innovative, novel, risky and disruptive and therefore an innovative mindset is a much better fit than a conventional business case and cost-benefit approach (Lamb, 2018).

The problem is broadly known i.e. ***an absence of data analytics capability in the presence of a market gap*** but there are many unknowns associated with the solution.

The solution could potentially be a mixed model of delivery to internal and external commercialised markets but the potential of these markets is largely unknown.

As there is no clear precedent for data analytics in UK education there are unknowns around how to acquire such a capability, whether it can generate a budget surplus, whether it will be taken up by the intended customer base, what the potential for growth is, whether suitable resources can be found etc.

In this context, the proposed capability appears in the bottom right hand corner of the problem / solution grid in Figure 38. It is a “new way of doing something” and as such application of traditional management techniques is unlikely to be appropriate or successful (Lamb, 2018).

|  |
| --- |
|  |
| Figure 38 – Managing Risk in Innovation (Furr, 2011), (Lamb, 2018) |

There are several key reasons why innovation “start-ups”[[3]](#footnote-3) often fail –

* Fundamentally [a] lack of recognition that start-ups are an experiment.
* [A lack of understanding that] the vision is the hypothesis and that validating the hypothesis is the task.
* [A lack of understanding that] the goal is to find a sustainable business model and that only when validated should it be scaled.

(Lamb, 2018)

Conventional business models can be managed using traditional accounting methods, cost optimisation, return-on-investment, accounting rate-of-return and net present value whilst innovation requires different methods of evaluation.

An alternative that is more applicable and relevant to innovation initiatives is “innovation accounting” defined as “how well innovation teams are doing in their search for profitable business models” (Viki, et al., 2017).

(Viki, et al., 2017) also propose that the business model for innovation start-ups be validated through an iterative process of testing and learning DURING execution of the business model whereas traditional management techniques would seek to fully validate the business model BEFORE execution begins.

“The job of the innovators is to continually validate the value hypothesis (does the product meet customer needs) and the growth hypothesis (how will customers and revenue grow)”

(Viki, et al., 2017).

However, the absence of a pre-approved business and cost case does not mean that criteria are not applied to consistently review the success of the project and to evaluate whether it should continue or be stopped.

(Taymor, 2020) proposes four tests which, if positive, indicate that an innovation project should be stopped –

1. Your team loses enthusiasm.
2. The project is missing growth targets.
3. Customers are not engaging.
4. A loss of internal support.

Adopting an innovation mind-set is a clear objective for the proposed solution within the constraints of constant, critical, and objective evaluation of success (or failure).

### Develop a Business Model

This objective defines the need to develop an initial business model using various techniques including a business model canvas (see Figure 39) as explained in “9 Steps to Creating a Successful Business Model” (The Business Channel, 2016).

The business model will need to be constantly reviewed during the “Innovation Lifecycle” and updated frequently during the constant iterations of testing and learning.

This objective then is the clear need to identify, compare and contrast a range of potential business models, then to select the one that best fits and to develop the chosen model using a business model canvas.

|  |
| --- |
|  |
| Figure 39 – An Example Business Model Canvas |

### Develop an Expertise in Data Analytics

It has been established that Lincoln College has a mature Further Education Information Services capability but no prior experience or knowledge of data analytics.

The MBA “Business Intelligence in the Digital Economy” (BIDE) module provided some knowledge and expertise of big data, customer journeys, cloud computing, Web 2.0 and digital business models. However, BIDE did not explore the key definitions and concepts of data analytics (Shaw, 2018).

Evidence uncovered during the literature review and investigation re-enforced the requirement for any manager within the field of data analytics to have a broad and deep understanding of it.

(Provost & Fawcett, 2013) state that “They [good data science managers] need to understand the fundamentals of data science well, possibly even being competent data scientists themselves”.

|  |
| --- |
|  |
| Figure 40 – Driving the Success of Data Science Solutions: Skills Roles and Responsibilities (Granville, 2017) |

Figure 40 maps the over-lapping roles and skills based on a Gartner report and clearly the “Analytics Leader” skillset is significant, incorporating many of the skills seen in the technical and business roles.

In “Data Science for Business” (Provost & Fawcett, 2013) propose that good management of the data science team is possibly more critical to success than having superior data scientists and provide a long list of skills for managers of data analytics –

* An understanding of data science including being competent practitioners [of data science].
* A good understanding of the needs of business.
* An ability to communicate effectively with both senior management and data scientists and being able to “translate” between the two.
* The co-ordination of technically complex activities including technical architecture and software.
* The need to be able to anticipate the outcomes [i.e. the success or failure] of data science projects.

In fact (Provost & Fawcett, 2013) go further by suggesting that the only reliable predictor of success of a data analytics project is the prior success of the investigator (manager).

In early conversations between the CEO of Lincoln College and the author it was quickly established that the CEO wanted the author to play a key role in acquiring and managing any data analytics capability that might later be agreed.

Linking the need for the author to be closely involved, the broad-brush knowledge imparted by the MBA BIDE module, and the extensive set of skills and knowledge required by managers of data analytics capabilities led to the formulation of this objective for the author to acquire and develop a deeper expertise in the field.

### Identify the Required Resources

The of-quoted report by Bain & Company provides a clear visual representation of the four elements that are important for developing advanced analytical capabilities seen in Figure 13 – people, intent, tools and data (Wegener & Sinha, 2003).

The report emphasised the importance for a firm to capture and store its own data but also in the era of big data it becomes important for a data analytics capability to be able to utilise data from many sources including –

* Private, internal datasets.
* Public, external datasets, for example the Office for National Statistics (Office for National Statistics (2), 2020).
* Hybrid, subscription datasets, for example EMSI Economic Modelling Data (EMSI Economic Modelling, 2020).
* Advanced APIs, for example the IBM Watson Personality Profile Service (IBM Watson, 2020).

Clearly data is a key resource and the key datasets will need to be identified for any data analytics capability to start and then be constantly reviewed, sometimes on a project-by-project basis.

Identification of the people necessary to staff a data analytics capability is also critical to success.

Identification of a standard, flexible toolset to enable the people to explore the data to identify patterns and to derive insight is key according to (Wegener & Sinha, 2003). Adoption of a standard toolset will enable the acquisition of expertise and facilitate rapid solution development but also the standard tools may need to expand and grow to deliver against diverse problems and projects.

This objective is the need for the solution to clearly identify the people, data, tools and any other resources necessary for the acquisition of a data analytics capability to be successful.

### Explore and Understand the Potential Market and Customer Base

The initial assumption is that the primary market for the data analytics capability will be the internal Lincoln College market and that the customers would be the academic managers, executive leadership team and non-curriculum areas like business development, marketing and commercial etc.

It has also been shown that the education sector has been dis-incentivised to invest in innovation like data analytics and that hence it can be assumed that the target customers may not understand or appreciate the potential of the product.

The intended internal market will need to be engaged with in order to lift their awareness and understanding and to ascertain the scale of the internal market and the potential demand.

Although the proposed capability is going to be developed for Lincoln College (and not the wider education sector in the first instance) there may be an opportunity for some external, local, commercial customers and this market also needs to be explored and understood.

The market evaluation is unlikely to be perfect in the innovation lifecycle model; it will need to be understood to the point that the capability can start and then constantly reviewed, refined, adjusted and improved.

This objective is to ensure that, as far as possible, the potential market and customer base are explored and understood.

In support of this objective there is a need to develop a social media presence to develop links with potential markets and other key stakeholders.

Figure 41 shows some examples of the social media blogs and posts that the author has engaged in throughout the project to begin this task –

|  |  |
| --- | --- |
|  |  |
| *Figure 41 – Sample Blogs and Posts on by the author on Data Analytics to External and Internal Professional Networks* | |

### Minimise Cost of Start-up

There are several drivers to minimising the cost of starting-up.

Firstly, it has already been shown that the UK FE sector is in a cycle of financial decline. It has also been suggested that there is evidence for Lincoln College being poised to exit from that cycle but even if that happens it is unlikely that investment cases with a high cost are likely to succeed as there will be many calls on limited capital resources.

(Lamb, 2018) proposes that innovation projects require small, iterative steps with each individual step being low cost and low risk.

(Viki, et al., 2017) go on to advise that innovators should specifically aim to “lower the cost of innovation” and that “if they do this successfully, they will hardly ever need high level budget approvals”.

If Lincoln College successfully emerges from the cycle of financial decline it is unlikely they will proceed with any investment without rigorous and robust review but certainly the advice of (Viki, et al., 2017) holds; the lower the cost and risk of each iteration of the innovation cycle the greater the chances will be that Lincoln College will chose to invest.

The cost of start-up could be further mitigated by seeking alternative sources of investment and funding or by exploring options like utilising existing office space and re-using or sharing existing resources in order to keep costs down.

### Establish Standard Frameworks for a Data Analytics Capability and Individual Projects

The final objective is to develop a standard framework for establishing the data analytics capability and a separate framework for approaching individual data analytics projects.

This would enable new data scientists to understand how to approach what can often seem to be daunting and unstructured problem spaces.

It can also be used to communicate to potential customers, funders and other key stakeholder how Lincoln College would approach development of a data analytics capability and individual projects which will aide understanding and help to achieve buy-in and support.

There is no agreed or widely adopted approach in the body of literature but establishing a loose framework and most significantly establishing principles for how results and insight are translated and transmitted from data experts to managers and hence to customers are essential objectives to maximise the chances of success.

# Chapter 4: Defining the Research Question

Based on the initial problem statement, literature review, key findings, defined objectives and on discussions with the CEO of Lincoln College and other key contributors and stakeholders, the research question and project title were refined and synthesised into the following –

***“How can the conditions be created for Lincoln College to acquire a data analytics capability that delivers the benefits and gains that are evident in other sectors but are thus far unproven in the UK education sector?”***

# Chapter 5: Methodology

No single or pre-existing methodology for addressing the research question was selected.

Rather a series of simultaneous initiatives were enacted that when added together address the key findings and objectives and inform the proposal and recommendations in order to address the research question.

## Elements of the Methodology

### Run a Pilot Project

There are many different examples throughout the literature of a process for running an analytics project but very little on how to acquire or build a data analytics capability.

One notable exception is (Davenport & Harris, 2017) who dedicate a whole chapter including proposing a clear framework with definitions for each stage of the journey and how to move from one stage to the next –

|  |
| --- |
|  |
| Figure 42 – Roadmap to becoming an analytical competitor (Davenport & Harris, 2017) |

The current state of Lincoln College best fits with stage 3 “Analytical Aspirations” although Lincoln College demonstrates some attributes of stages 1 and 2 as well which raised the question as to whether a pilot project (the “prove-it” path) would help prove the applicability of data analytics to the context of Lincoln College.

(Davenport & Harris, 2017) warn that the “prove-it detour” can add 1-3 years onto the timeline to becoming analytically competitive but cite advantages including obtaining evidence, gaining valuable insights, acquiring momentum, getting started and lower costs with the following recommend steps for a pilot project –

1. Find a sponsor.
2. Implement a localised project that adds value and produces measurable benefits.
3. Document the benefits and share with key stakeholders.
4. Continue with a string of localised successes until the organisation is ready to progress to the next stage.

In March 2019 an ideal opportunity presented itself for running a pilot data analytics project.

Lincoln Bishop Grosseteste University were running a semi-commercial small, experimental research group called LORIC which had the potential to help Lincoln College run a pilot project.

By June 2019 funding and commercial terms were agreed and the pilot project proceeded along the following timeline –

* March 2019: Initial meeting between Lincoln College and LORIC.
* June 2019: Lincoln College CEO gave authorisation to proceed and budget agreed.
* July 2019: Initial stakeholder workshops with Marketing, Student Services, Information Services, Learning Resources and Curriculum Staff.
* September 2019: Data analytics work commences.
* September 2019 – February 2020: Iterations of the following “OODA” framework (Enck, 2012)
  + Observe: what are we seeing and what does that tell us?
  + Orientate: Analysis and investigation of the observation?
  + Decide: what have we learned? What are the actionable insights?
  + Act: If we can we act, what should we do or do we need to go back to “observe”?

At the strategic level the pilot project developed mission and vision statements and adhered to these throughout the project –

**MISSION**: To use out-sourced expertise to explore the idea that data analytics can be used to obtain actionable insight that can drive business outcomes.

**VISION**: To combine internal, private data with external, public data to understand why learner numbers have gradually declined and hence to build an evidence-backed view as to what actions could increase learner numbers and recruitment in future.

The mission statement describes organizational purpose; it addresses “Why we exist.” The vision statement projects “What we want to be.” (Grant, 2015)

As the pilot project progressed four key hypotheses were identified and defined which if answered would achieve the vision and deliver the vision –

* **Hypothesis 0 (null hypothesis):** Declining recruitment is due entirely to a fall in the number of 16-18 year olds in the catchment area.
* **Hypothesis 1:** Removal of the education maintenance grant[[4]](#footnote-4) reduced recruitment/retention.
* **Hypothesis 2:** Competition from local schools has adversely impacted learner numbers.
* **Hypothesis 3:** Removal of the travel grant led to a shrinking of the student catchment area.

During the literature review evidence was uncovered from (Agasisti & Bowers, 2017) suggesting the combination of private, internal data and publicly available data sets can create big data that can leverage new insight and the pilot project demonstrated this in practice.

On 9th February 2020 the key findings were communicated to the Executive Leadership Team –

|  |
| --- |
|  |
|  |
| Figure 43 – Hypotheses and Findings from the Pilot Project |

After several months of data science work, no data analytical evidence had been uncovered that supported any of the positive (i.e. non-null) hypotheses.

In order to try to extract some actionable insight from the body of work an additional hypothesis was developed - “Hypothesis 4: We have lost market share in Sleaford because of a combination of travel times, loss of the travel allowance and increased local competition.”

No actionable insights were uncovered but several of the theories relating to data analytics were observed in practice –

* The 4th hypothesis was created by a stakeholder wishing to use data to prove what they thought they already knew. This is a clear manifestation of a “Confirmation Bias” (Cambridge (1), 2019) and helped by providing a concrete example of the theory.
* In addition, the dis-proving of hypothesis 4 helped to avoid making a significant investment in a trial for assisted transport which the analytical evidence suggested would have no impact just like the example of the Rialto Police department uncovered in the literature review (see Figure 9).
* The data showed that the withdrawal of the transport allowance had not adversely impacted learner numbers but that this is not a proof that introducing a new transport allowance would not positively impact recruitment.

This conclusion provides another concrete example of the theoretical literature evidence from (Cambridge (1), 2019); that evidence produced by data analytics must be used in triangulation with managers and leaders observational evidence and experience and not acted on in isolation.

The Executive Leadership Team found the myth-busting of the stakeholder hypotheses interesting and useful. The pilot had not uncovered actionable insight but had helped by avoiding investment in operational initiatives based on intuitive opinions that had no basis in data.

There were many other valuable “lessons learnt” from the pilot project –

* Data analytics can provide results that are contrary to pre-held beliefs (but sometimes knowing what is unlikely to work can add value).
* It re-enforced the best-practice approach that the framing of the problems and questions is key to driving analytics outcomes as potentially the original framing was too broad and not sufficiently specific and targeted.
* Working with an external data analytics consultancy has its challenges – they know about analytics, but FE data is very complex for them to understand.
* In a data analytics project, it is data trends rather than 100% accuracy that guide outcomes. Outcomes are more likely to be probabilistic than absolute.
* The pilot helped in the acquisition of expertise and in identification of publicly available datasets.
* The pilot identified a significant data gap in respect of access to high-end labour market intelligence data like EMSI (EMSI Economic Modelling, 2020).
* There is no existing capability for data analytics in Lincolnshire. LORIC has a weak offering that is focused on academic research and has moved away from data analytics.
* An excellent working relationship was established and developed with “The Data Place” who are a social enterprise data science company based in Plymouth who were originally a sub-contractor of LORIC.
* The Executive Leadership Team gained understanding and knowledge because the pilot project attempted to solve a problem which they understood well.
* Internal stakeholder commitment with an external provider of data analytics was inconsistent and generally poor.
* The pilot raised the profile of data analytics with the ELT.

### Knowledge Acquisition

It has already been demonstrated that good management of any data science team is critical to success and that the only reliable predictor of success of a data analytics project is the prior success of the investigator (manager) (Provost & Fawcett, 2013).

Also, the MBA BIDE module had provided some high-level awareness of the area but not enough expertise and knowledge to enable an informed acquisition of a data analytics capability.

To achieve knowledge acquisition and to gain the required skills the author undertook a course in “Business Analytics: Decision Making Using Data” with the University of Cambridge Judge Business School between September and December 2019 (see Figure 44).

The course was delivered online and included regular live webinars with experienced data scientists and academics who are leading experts in their field. It was specifically targeted for managers and executives who are who will be leading the data analytics teams and capabilities in their organisations.

The course modules were structured as follows (Emeritus, 2020) –

1. Decision biases.
2. Descriptive analytics.
3. Big data opportunities.
4. Experimentation.
5. Predictive analytics.
6. Prescriptive analytics.
7. Organisational, ethical and legal issues.

The course provided significant learning which was all relevant and appropriate to the potential acquisition of a data analytics capability for Lincoln College.

Much of the learning, knowledge and citations etc. have been used (with references) throughout this report in the literature review, findings, objectives and methodologies as well as being used to inform the thinking that shaped the proposal and recommendations.

The official certification and the credibility that conveys is also being put to useful effect for building a professional network online via LinkedIn.

The Judge Business School course provided an excellent understanding of what can be achieved by data science and an in-depth explanation of the terms, but this did not extend to any technical knowledge of data science, hence the author undertook further self-taught education into how data science is implemented.

A brief review of the available technologies suggested that the Python programming language, particularly when embedded in a Jupyter Notebook, is one of the most popular and powerful toolsets currently being used by data science practitioners (see Figure 45).

The author then self-taught these skills using learning resources and tools that are freely available online. The highest quality learning was obtained by taking exploratory examples from towardsdatascience.com (Towards data science, 2020) and then using the Anaconda, Microsoft VS Code, Python and Jupyter software development tools to develop working examples.

|  |
| --- |
|  |
| Figure 44 – Certificate of Completion for Business Analytics: Decision Making Using Data |

|  |
| --- |
|  |
| Figure 45 – Most Popular Programming Languages Over Time (Stack Overflow, 2019) |

Using this approach, the author has completed a wide range of fully working, informative data science projects covering the following areas –

* Use of big data and application programming interfaces.
* Supervised machine learning and linear regression.
* Advanced regression including polynomial regression and logistical regression.
* Identifying and avoiding over-fitting of machine learning models.
* Support vector machine techniques for prediction of classification.
* Tree induction to make linear and classification predictions.
* Practical application of these techniques to predict student attendance patterns.

Figure 46 and Figure 47 are included as examples of some of the data analytics and data science outputs produced during the learning and discovery process –

|  |
| --- |
| A picture containing object  Description automatically generated |
| Figure 46 – Using tree induction to predict if patients will develop diabetes |
| A screenshot of a cell phone  Description automatically generated |
| Figure 47 – Using a scatter-plot matrix or “SPLOM” to uncover patterns in English & Maths student attendance |

This learning approach has enabled the author to gain a good understanding of hands-on data analytics. It has already proven its worth with the application of the techniques to a real-world problem with student attendance and will prove invaluable in informing and steering the future acquisition of a data analytics capability.

### Raise Awareness, Understanding and Buy-in of Senior and Middle Management

Raising management awareness has been identified as being key to success, both through the earlier application of Kotter’s 8 step change framework (Figure 26) and by Dr David Stillwell, University Lecturer in Big Data Analytics & Quantitative Social Science at Cambridge University Judge Business School who discussed the criticality of senior management buy-in (Cambridge (0), 2019).

The author has designed a series of briefing sessions during the project that have been delivered directly to the Executive Leadership Team (ELT) at Lincoln College –

**On 30th September 2019** the first session to ELT was delivered about data biases and decision traps. Explanations and examples of framing bias, over-confidence, confirmation bias, availability bias and short-termism were provided through a visual presentation (see Figure 48) and open discussion.

|  |  |
| --- | --- |
|  |  |
|  |  |
| Figure 48 – Example Slides from the ELT Data Bias Presentation of 30th October 2019 | |

This session was designed to begin the ELT engagement process, to start raising awareness and to build confidence levels and interest and was pitched at a level that is appropriate for senior leaders.

The (Cambridge (1), 2019) “Carter Racing” case study and the instance of the Discovery Space Shuttle disaster were used to help the explanations and to visualise the concepts.

The introductory session was well received by the ELT and the feedback was very positive.

**On the 21st January 2020 a second session** was held on how supervised machine learning and the tree induction technique can be used to make probabilistic predictions about future patients who are at risk of developing diabetes and what factors can be changed to move them to a lower risk.

This session was designed specifically to show how other sectors can and are using data analytical techniques to drive business outcomes and the question was posed – “if so many other sectors can and are using data analytics to drive outcomes, why is it that the education sector is not using these techniques?”

Feedback was again very positive, specifically that the author was able to take an extremely complex topic, strip out the jargon and make it easy to understand. For example, the principle of entropy was explained by mixing a glass of orange squash with a glass of water and then discussing how it might be possible to separate them out again.

**On the 3rd February 2020** the next session was held where the results, findings and lessons learnt of the pilot project were fed back to ELT.

The detail can be found in the “[Run a Pilot Project](#_Run_a_Pilot)” section and hence are not repeated here.

**On 11th February** EMSI Economic Modelling were invited into the ELT to present on their subscription data set and analytics tool for driving business insight from economic and employment related forecasting.

This session was also well received and has resulted in an investment in the service and an associated plan to roll out the expertise to the middle management tier and into the curriculum planning processes.

In addition to Executive Leadership Team engagement, the CEO has worked with the author to embed an awareness of data analytics directly into the management development training programme for all the senior and middle managers.

This is a concrete example of sharing tools, skills and responsibility in respect of data analytics that is collectively referred to as “The Democratization of Data” (Cornelissen, 2018).

**On the 13th February 2020** the author presented on “Data Analytics: How Managers and Leaders can use data to improve decision making and outcomes” into a session of the “Determined to Lead” management development programme which is delivered to all 60 managers across the entire college management team.

Topics and areas covered included:

* Terminology: Descriptive, Diagnostic, Predictive and Prescriptive Analytics.
* How Managers can make Good Decisions Using Data (or how to avoid bad ones).
* Correlation and Causation.
* Using Data Analytics in the future of Education.

There were several participatory exercises including a final exercise where attendees authored their own “user stories” envisaging what could be achieved with access to any public, private, or even imaginary datasets.

Again, feedback was very positive indicating that the goals of raising awareness, understanding and buy-in for the college management team made good progress.

### Identify, Compare and Evaluate High Level Business Models

Figure 49 shows visual representations of all available permutations of high level business models identified during a “brain-storming” session between the author and the CEO of Lincoln College on 21st January 2020.

In each instance the blue circle represents where the data analytics capability is provided.

The rectangles represent the scope and boundaries of the organisations involved and the black dotted rectangle represents the scope and boundary of Lincoln College in each diagram.

|  |  |
| --- | --- |
| 1. Internal, dedicated | 1. Collaboration, shared |
|  |  |
| 1. External, consultancy | 1. External, shared service |
|  |  |
| 1. Internal, commercial | 1. Hybrid |
|  |  |
| *Figure 49 – High Level Business Models for a Data Analytics Capability at Lincoln College* | |

**Internal, dedicated**: In this business model the data analytics capability is wholly delivered internally by Lincoln College and the staff would be full-time-employed staffing resources.

The advantages are the simplicity of the model, speed of implementation and having full control of the capability and capacity. The disadvantages are that Lincoln College would bear the full cost of all the resources and if the innovation-venture were to fail it would be difficult to remove or redeploy the staff.

Also, it is likely to prove difficult to attract and retain the high-expertise staffing resources required when constrained by existing standard contracts, pay-scales, terms and conditions.

**Collaboration, shared**: This model represents launching the capability in collaboration with another institution, potentially another FE college with which Lincoln College has close working ties.

The advantages are that the start-up costs are halved and both organisations can contribute resources and expertise into the capability. The disadvantages are a loss of control and a loss of 50% of the output.

The governance could also be complex and could lead to disputes with both organisations having a say in which projects to prioritise and how to divide up the resources and deliverables.

**External, consultancy**: This model represents a scaling up of the approach taken in the pilot project where an external consultancy like The Data Place provides the specialist data analytics resources.

The advantages are that the consultancy will already have the resources, knowledge, experience and expertise and a proven track record.

The disadvantages are higher costs, how the consultancy will deal with the complexities of FE-specific data and achieving internal stakeholder commitment and engagement.

**External, shared service**: In this model the data analytics capability is provided by a centralised, external, shared service like the analytics labs run by JISC and HESA.

The advantages are no cost or very low cost to the consumers of the service and no requirement to build or develop a data analytics capability.

The disadvantages are that the projects are all selected by committee and this process as well as the actual production of deliverables has been shown to be slow and unwieldly in practice.

Also, the JISC/HESA lab deliverables have all been very specific to Universities who control the committees and there is no evidence of any advanced or predictive analytics outputs.

**Internal, commercial:** This is an extension of the “internal, dedicated” model. As well as delivering data analytics services internally this model also delivers external, commercial services to local firms.

The advantages are that the commercial work will generate income to contribute towards the costs or even generate a surplus. It could also enhance the reputation and profile of Lincoln College with local employers and other key stakeholders.

The disadvantages are that the outputs would be shared between the internal and external markets and there could be conflicting priorities in choice of projects. There could also complications attributing a notional income value when committing resources to developing deliverables for the internal market.

**Hybrid:** This is an extension of the “Internal, commercial” model which adds in a flexible relationship with an external consultancy to enhance delivery capability and capacity.

The advantages are that the external consultancy will already have data analytics expertise which would be useful in the start-up period whilst the internal resources were identified, acquired and trained. Also, the external consultancy would be able to provide extra resources during periods of peak demand.

The disadvantages are the additional costs associated with engaging with the external consultancy and time spent managing the external relationship.

### Research the Market

#### The Internal Market

The primary focus for any proposed Lincoln College data analytics capability will be the internal market hence a methodology was required for assessing the internal market demand.

As part of the data analytics session which the author presented to the full management team on 13th February 2020 an exercise was designed and delivered to capture the internal appetite for data analytics products and deliverables.

Following the presentation each group was set the task of imagining free and easy access to any real or imaginary dataset and to capture requirements in the form of “user stories” (AoC / ETF Data Science, 2019) –

|  |
| --- |
| ووى ■  User Story  » AS a:  - :When  l• want :to  So l :can  كم هلـعماء-  ممت  < user >  وا)) € SR X مىممهم  سهدهم  ءلسمه .عسهلمأ) سماعما ليه  <context>  هه يساره، ه -همه: ندمم أه،  مره نه بم . نمه ممالمم . بلممهاس مذ  <feature>  ين ممهد ناب- اذ  ماسلمد مه مام  <benefit>  مبي بج مهدنلمد.٠ لهام) (الأمه  د ء.. ام، آنهه :٧نف--  تؤديه ايبا ) 1- اك |
| Figure 50 – Example User Story for Capturing Data Analytics Requirements (AoC / ETF Data Science, 2019) |

The features and attributes of the 31 user stories captured were then collated and analysed using Microsoft Excel and RapidMiner (RapidMiner, 2020) as shown in Figure 51 –

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
| I want to …  A screenshot of a cell phone  Description automatically generated | So that I can …  A screenshot of a cell phone  Description automatically generated |
| *Figure 51 – Statistical Analysis of the Management User Stories* | |

The interpretation of the analysis is as follows –

Engagement with the exercise was excellent with 75% of the attendees submitting a total of 31 user stories.

The datasets already exist for 58% of the user stories, the other 42% would require data that does not currently exist or is not currently collected.

71% of the stories would require backwards-looking descriptive analytics with 29% requiring predictive capabilities. 29% is high relative to the 80/20 figure for wider industry (Wu (1), 2013) suggesting a high level of innovative thinking.

Just under 50% of the stories could be delivered with existing expertise with the other 50% requiring new expertise. Even for the 50% which could be delivered there are still issues with capacity to deliver within the existing teams.

50% of the user stories were judged to be complex or very complex indicating that a high expertise data analytics capability would be required to deliver them.

75% of the user stories would require a significant investment of resources clearly showing a need for additional data analytics capacity if they were to be delivered.

Only 23% of the stories were judged to deliver high or very high benefits indicating that requests for data analytics work needs to be carefully evaluated and selected based on a range of criteria including resource needs vs. potential benefits.

The word cloud analysis was insightful. This textual analysis of the user stories submissions shows that college managers want to understand more about their data so that they can improve their outcomes. This analysis suggested a strong alignment between the managers’ requirements and some of the key purposes of data analytics.

The user-story exercise strongly suggested that there is an internal market need for a data analytics capability.

Given that the high-level business models include the potential to deliver some external, commercial services the secondary market would be local firms. Hence a further analysis was required to evaluate this market.

#### The External Market

The best data that could be obtained to investigate the external market was found in the EMSI Economic Modelling datasets (EMSI Economic Modelling, 2020)[[5]](#footnote-5) and can be seen in Figure 52 –

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
| *Figure 52 – External Market Analysis using EMSI Economic Modelling Data* (EMSI Economic Modelling, 2020) |

Data Scientist, Data Engineer and Data Analyst as a cluster of jobs has the lowest “posting intensity” of all occupations in the Lincolnshire region and this re-enforces the earlier findings from Glassdoor (Glassdoor, 2020) suggesting that these types of roles will be very hard to recruit for conventionally.

This indicates a lack of data analytics activity within the Greater Lincolnshire region. Interpreting this finding is difficult in the context of an innovation project. It could mean that there is no appetite amongst local businesses or it could mean that there is a latent demand which is yet to be fulfilled.

If the latter is true it could indicate that a part-commercial capability would be able to find customers for its services.

The predicted regional trend shows 6.9% growth between now and 2025 which is ahead of the national trend but from a starting position in 2020 which is 66% below the national average.

The analysis also showed that 60% of data analytics posts are filled by males whilst the 25-34 age range is the largest representing 43.9% of all posts filled.

In contrast to the strong positive indications for an internal market the evidence for an external market of local firms wanting to consume the data analytics capabilities in a mixed internal / external delivery model is unclear and inconclusive.

# Chapter 6: Proposal and Recommendations

The purpose of this chapter is to pull together all the information from the literature review, findings, objectives and methodology in order to make a proposal and recommendation that will answer the research question i.e.

***“How can the conditions be created for Lincoln College to acquire a data analytics capability that delivers the benefits and gains that are evident in other sectors but are thus far unproven in the UK education sector?”***

## Proposal

As of May 2020, much of the background work and foundations described in the preceding sections was either fully completed or well advanced.

This included completion of a pilot project, a high degree of knowledge acquisition and a significant programme of raising awareness, understanding and buy-in of senior and middle managers at Lincoln College.

The CEO of Lincoln College Group was bought into the concept throughout, enthusiastic about the opportunities and had played an active part in facilitating and developing the ideas.

Building on the completed elements of the methodology the proposal for creating the conditions for Lincoln College to acquire a data analytics capability is to develop and grow, rather than acquire, a data analytics capability.

To begin with that capability will be modest and small-scale but in keeping with “The Innovation Lifecycle Model” (Viki, et al., 2017) it will constantly iterate through cycles of testing and learning with the ability to scale up, scale down, close down or pivot at any review point based on the 4 criteria for the closing of innovation projects proposed by (Taymor, 2020) –

1. Your team loses enthusiasm.
2. The project is missing growth targets.
3. Customers are not engaging.
4. A loss of internal support.

The proposed name of the venture is “Lincolnshire Data Analytics Ltd.” (LDA Ltd.) which will be a subsidiary of the main holding company of Lincoln College.

The proposal is that the data science capability be split between delivering internal, education-related outcomes and external, commercial delivery.

This internal, education-related focus will enable delivery of data analytics outcomes into Lincoln College.

The objective of the external, commercial activity is not to focus on pure commercial profit, rather it is to generate some income to contribute to start-up and ongoing costs in keeping with the “Minimise Cost of Start-up” objective .

The following sections extend the proposal by describing the individual recommendations in detail.

## Recommendations

### Business Model and Structure

|  |
| --- |
|  |
| *Figure 53 – Hybrid Business Model* |

In Chapter 5: Methodology, the section, “Identify, Compare and Evaluate High Level Business Models” defined and described a range of potential high-level business models.

The recommendation is to adopt and implement the hybrid model shown in Figure 53 and broken down in detail in the proposed business model canvas shown in Figure 56.

The hybrid model is an extension of the internal, dedicated model which has the advantages of simplicity, quick implementation, and a high degree of control.

The disadvantage of fixed costs (salaries, resources, over-heads etc.) is mitigated by the proposal to generate external, commercial income to contribute towards those costs.

The other main disadvantage of the need to redeploy staff if the venture fails is mitigated by keeping the corporate structure distinctly separate from the core business which will enable a novel, commercial employment contract and terms and conditions (see “Staff Recruitment, Development and Retention”).

A key advantage of the “hybrid” extension to the internal, dedicated model is the addition of a formal relationship with an external consultancy to enhance capability and capacity.

During the pilot project a positive, trusting and collaborative relationship was forged with social enterprise data science firm The Data Place. Managing director and co-owner Martin Howitt has already been approached in respect of being part of the proposed business model which he is open and enthusiastic about.

There are several advantages to collaboration with an external, established but geographically distant data science company –

1. During the early stages when the staffing resources are being recruited and trained The Data Place can provide an immediate capacity should the commercial activity increase before the internal data science staff are ready to deliver.
2. Throughout the life of the venture The Data Place can provide additional capacity to guard against peaks in demand that cannot be met solely by internal staffing resources.
3. The opposite also applies; The Data Place have agreed that a two-way partnership would include them approaching Lincolnshire Data Analytics Ltd. during their peak work-loads with commercial activity they cannot accommodate.
4. The data scientists at The Data Place and at Lincolnshire Data Analytics can work together to share and develop best practice and technical knowledge.
5. Larger projects that neither team can deliver individually may be successful if delivered in partnership.
6. The geographic distancing (Lincoln and Plymouth) mean that Lincolnshire Data Analytics Ltd. and The Data Place are highly unlikely to compete directly over customers or contracts.

Clearly out-sourcing commercial activity to The Data Place will be financially disadvantageous because they will need to be paid for their activity, but the trade-off is that this is better than turning away potential work and there could still be room for a small profit margin.

Even in the early stages the internal staff will be able to learn from The Data Place and successful partnership projects will help to generate market-awareness and positive public relations.

Any local out-sourcing to The Data Place will be white-labelled and delivered through Lincolnshire Data Analytics Ltd. (and vice versa) to retain intellectual property and develop a positive reputation.

Lincolnshire Data Analytics Ltd. will be formed as a separate limited company within the corporate governance structure of Lincoln College and will start with a part time Director of Data Analytics and 2 full time data scientists (see Figure 54 and Figure 55).

It should be noted that in any sub-contracting arrangement with The Data Place the ownership of the data and all associated rights would be with Lincoln College and the customer and that explicit data sharing arrangements will be put in place for each project.

The business model and other aspects of the proposal make a case for creation of a low-cost, high innovation model and it has been demonstrated that traditional accounting methods and cost cases are not part of the initial innovation lifecycle.

However, by way of an indication of the potential benefits the following illustration is provided –

Lincoln College has 2,700 16-18 ESFA learners that generate £4,700 each totalling £12.69m of annual income.

Given that many of the proposed academic uses of data analytics are related to attendance and retention, if the analytics capability could provide insights leading to improved student retention each 1% improvement would be worth £127k per year.

|  |
| --- |
|  |
| *Figure 54 – Proposed Governance Structure of Lincolnshire Data Analytics Ltd.* |

|  |
| --- |
|  |
| *Figure 55 – Proposed Initial Staffing of Lincolnshire Data Analytics Ltd.* |

### Funding

In line with the objectives that have already been discussed, specifically to “Minimise Cost of Start-up” an innovation project is more likely to be successful if cost and risk are reduced as fat as possible (Lamb, 2018) and if the innovators have sought to “lower the cost of innovation” (Viki, et al., 2017).

The recommendation is low cost in so far as the initial staffing proposal is modest (with future potential for growth) but the costs will be further reduced and could potentially be removed if external funding can be used to seed Lincolnshire Data Analytics.

There are several potential funding sources that could be used and the recommendation is to identify and rigorously investigate all potential sources of funding.

The Lincolnshire Chamber of Commerce web-site provides details of two potential local funding sources (The Lincolnshire Chamber of Commerce, 2020).

|  |
| --- |
| A screenshot of a cell phone  Description automatically generated |
| *Figure 56 – Business Model Canvas for Lincolnshire Data Analytics Proposal* (The Business Channel, 2016) |

Collaboration 4 Growth (C4G) could potentially be used to provide funding directly to local businesses who wanted to engage commercially with Lincolnshire Data Analytics for analytics and data science services. If this funding could be obtained then the local firms would effectively receive the data analytics services “free-of-charge” as the costs could be fully covered by the C4G funding.

Lincolnshire Data Analytics could also complete all the administration and form-filling necessary to obtain the funding thereby removing any barriers for local firms in accessing the funding.

Another potential source of funding and support on the Chamber web-site is access to Paula Finch, a local business coach and mentor, who has helped many local businesses obtain funding for consultancy, coaching, marketing, strategy and business growth development (The Lincolnshire Chamber of Commerce, 2020).

These potential funding sources would significantly help in reducing the costs and increasing the attractiveness of the investment case.

There is also a potential funding source for the internal, education focused data analytics. A company called Advanced provide the core student records management information systems to Lincoln College and are one of the leading providers in the UK Further Education sector –

|  |
| --- |
|  |
| Figure 57 – Market Analysis of UK FE Student Records System (Lincoln College STEP Review Team, 2016) |

The author has already approached senior management at Advanced to enter into a data analytics partnership and the proposal is currently subject to an internal business case that is being championed by several key staff at Advanced.

As part of the pitch data from the Advanced Pro Suite was extracted and displayed in a series of scatter charts with simple linear regressions and correlations and then visualised using the principles of user story-telling.

This showed the potential to obtain fresh insights using data analytics approaches (see Figure 58).

The basis of the proposal to Advanced is that Lincoln College are a high-end user of the ProSolution student record system and are fast becoming sector leading in their understanding and adoption of data analytics and much of the source data is contained in the Advanced Pro suite systems.

The proposition put to Advanced was that in return for direct sponsorship to cover start-up costs like staffing and equipment, Lincolnshire Data Analytics will openly share the results of any machine learning algorithms that operate on data stored partially or wholly in the Advanced systems.

Advanced can then implement the machine learning algorithms in future releases of the Pro suite and sell them as standard features.

Lincoln College will not suffer any competitive disadvantage as their direct competitors are local sixth forms for vocational learning, independent training providers for apprenticeships and local universities for higher education whilst the Pro suite is specific to Further Education and not relevant or appropriate to the competitors.

The benefit to Lincolnshire Data Analytics is seed funding and ongoing sponsorship and the benefit to Advanced is the fast acquisition of machine learning algorithms that leverage Lincoln College’s significant knowledge of FE sector data and association with the college’s excellent reputation for innovative use of management information systems and data.

This could also be a potential route by which advances in machine learning and data analytics in at Lincoln College are disseminated to the wider Further Education sector.

It should be noted that in any commercial relationship with Advanced the ownership of the data would remain with the college; it is the ownership and design of the machine learning algorithms that would be the key deliverable in this relationship.

Also, there are a range of additional opportunities for potential funding including –

* The Greater Lincolnshire Local Enterprise Partnership (GLLEP)
* The GLLEP Skills Advisory Panel (SAP) sub-committee
* The Lincolnshire Chamber of Commerce
* The Lincolnshire Federation of Small Businesses (FSB)
* Lincoln Bishop Grosseteste University
* JISC (the Joint Information Systems Committee)
* The Collab Group of Colleges
* Selenity (a local software development company)
* Crowd Funding

All these sources need to be rigorously investigated to maximise external funding at start-up and beyond, thereby minimising the cost of innovation and maximising the probability that the project will be authorised to move forward to the implementation stage.

|  |
| --- |
| A screenshot of a cell phone  Description automatically generated |
| Figure 58 – Market Analysis of UK FE Student Records System (Lincoln College STEP Review Team, 2016) |

### Internal / External Market Split

The recommendation for the relative split between the internal education market and the external commercial market cannot be finalised until the results of the funding search are known.

The funders, whilst not being shareholders, will certainly become important stakeholders and this will influence the approach in respect of how the resources are allocated across internal and external activity.

Within this caveat there are some principles and guidelines that are recommended.

If Advanced become a sponsor and stakeholder then this will direct the early projects and influence them to be focused on internal education with a sub-objective of delivering machine learning algorithms that can then be implemented within via the Pro suite across the wider sector.

Alternatively, if Advanced decide not to become sponsors but some funding is obtained from a local funder like the Chamber of Commerce this would focus the early initiatives externally.

Also, there is the consideration of paying bonuses for the recruited data scientists which are a core aspect of the proposed remuneration package (see “Staff Recruitment, Development and Retention”).

Additional bonuses can only be paid in respect of exceptional commercial performance which generates a surplus from which the bonuses can be paid.

However, a balance must be struck to ensure the data scientist staffing resources do not avoid or under-deliver internal projects in order to focus on external projects which could potentially be more lucrative for them.

A potential solution is to structure the end-of-year bonus based on both aspects. The performance related bonuses could mandate commercial success to generate the income and delivery against internal key performance indicators to distribute it.

### Staff Recruitment, Development and Retention

“Data Scientist”, “Data Engineer” and “Director of Analytics” are the 3 roles that are both hardest to fill and have the highest projected growth according to (Forbes, 2017) placing these roles into the “Disrupters” quadrant.

The median salaries for these roles are far higher than those paid to Further Education technical roles (Glassdoor, 2020), (Frontier Economics, 2017) and according to (EMSI Economic Modelling, 2020) these roles have the lowest posting intensity of all occupations in the Lincolnshire region indicating that they are extremely rare locally.

This combination of factors indicates that a standard approach to recruitment is unlikely to succeed and that a radical and innovative approach is required.

A recent post on the popular question-and-answer blog site Quora.com posed an interesting question – “If data science is so in-demand why does it appear to be almost impossible to land a data science role?” (Wayland, 2018).

|  |
| --- |
|  |
| Figure 59 – Skewed distribution of the attractiveness (to employers) of data scientists in the jobs market (Wayland, 2018) |

The data in Figure 59 is fictitious and has been made up to demonstrate the point. The discussion has no empirical basis, however the conclusion that high-end companies fight over a tiny pool of highly skilled data scientists is corroborated by other academic and empirical sources.

(Provost & Fawcett, 2013) cite a tiny pool of top data scientists who consistently win all the high-value data science competitions (for example the NetFlix Prize). They surmise that this pool is almost impossible to break into and that the scientists already in that pool are very difficult to recruit as they can be very exacting when accepting job offers.

Conversely there are plenty of sources corroborating the view that many firms value soft skills and attitude above pure technical skills in their recruitment processes including (Anderson, 2019).

If the attitude approach is correct it will certainly be possible to identify candidates in the “Everyone Else” section of the bar graph in Figure 59 with high potential and to support, liberate and enable them to work their way towards the “Top 5%”.

This theory is the basis of the proposal for staff recruitment, development and retention for Lincolnshire Data Analytics Ltd.

The proposed governance structure creates Lincolnshire Data Analytics Ltd. as a separate sub-company within the overall Lincoln College governance structure enabling the flexibility to create and adopt non-standard approaches to employment contracts and terms and conditions.

The basic pay will be in line with FE sector averages for technical staff. This will ensure that existing technical staff do not feel demoralised or under-valued through perceived or actual lack of equity in basic salary as shown in Figure 60 (Jain, 2018).

|  |
| --- |
| A picture containing knife, table  Description automatically generated |
| Figure 60 – The “Equity Equation” (Jain, 2018) |

However, the data scientists will be incentivised with performance related pay linked to commercial income which will enable their remuneration to significantly exceed basic pay for an exceptional combination of external, commercial and internal, education outcomes and deliverables.

Novel advertising will appeal to potential data scientists who have either a formal or self-taught background in the basics of data analytics and whose demonstrable enthusiasm and commitment for learning the subject indicate “Top 5%” potential.

The job postings will also make clear that these individuals must be prepared to become part of a small, innovative venture, thereby accepting and sharing not only the rewards for success but also the risks of failure which would lead to closing the venture and effective redundancy.

Successful candidates would need to be confident in backing themselves by accepting a relatively low basic package with performance related pay as well as taking responsibility for their own constant self-development and improvement.

In return for accepting these conditions successful candidates will be given the opportunities to fast track their development to becoming credible data scientists.

The author’s own experience strongly suggests that this approach is possible.

The author achieved an online qualification in “Business Analytics: Decision Making Using Data” with the University of Cambridge Judge Business School between September and December 2019 which provided a solid foundation for understanding the principles of data analytics.

From that foundation the author has expanded his knowledge into the practical application of data analytics and data science using free or low-cost online resources leading to the design and development of a professional development path for data scientists (see Table 1).

(Raval, 2018) suggests that these resources could be completed in 3 months. In reality that time-frame is more likely to be at least 12 months with the learning ordered and sequenced to reflect the prior experience of the staff and the current operational and delivery priorities of Lincolnshire Data Analytics Ltd.

The view of the author is that candidates with potential and commitment can be found in the region and then trained and developed using a range of freely available, high-quality, low-cost learning resources combined with real-world commercial and educational experience.

A final point to note is retention. As the staff approach the threshold of the “Top 5%” (Wayland, 2018) their market attractiveness will be such that staff turn-over is likely or even inevitable even where staff are valued, rewarded and well managed.

Staff turn-over will be managed by being prepared to make the recruitment and training process ongoing rather than one-off and by maintaining the relationship with The Data Place to provide an option for resourcing if a resignation were to coincide with a peak of activity.

### Toolsets, Datasets and Resources

Selection of appropriate datasets, powerful toolsets to manipulate those datasets and other resources required to deliver outcomes is a critical success factor.

Making the wrong choices will lead to an inability to deliver outcomes, long delivery timescales, an inability to explain how outcomes have been arrived and excessive costs.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Course / Resource | URL / Link | Provider | Platform | Cost | | Introduction to Python for Data Science | <https://www.edx.org/course/introduction-to-python-for-data-science-2> | Microsoft | edX | Free | | Statistics and probability | <https://www.khanacademy.org/math/statistics-probability> | Khan Academy | Khan Academy | Free | | Computing for Data Analysis | <https://www.edx.org/course/computing-for-data-analysis> | Georgia Tech | edX | Free | | Kaggle Learning | <https://www.kaggle.com/learn/overview> | Kaggle | Kaggle | Free | | Kaggle Titanic Competition | <https://www.kaggle.com/c/titanic> | Kaggle | Kaggle | Free | | Machine Learning for Data Science and Analytics | <https://www.edx.org/course/machine-learning-for-data-science-and-analytics> | Columbia University | edX | Free | | Python Data Science Handbook | <https://jakevdp.github.io/PythonDataScienceHandbook/index.html> | O’Reilly | Github | Free | | Deep Learning | <https://www.deeplearningbook.org/> | MIT | MIT | Free | | Introduction to Relational Databases | <https://www.udacity.com/course/intro-to-relational-databases--ud197> | Udacity | Udacity | Free | | Introduction to NoSQL Data Solutions | <https://www.edx.org/course/introduction-to-nosql-data-solutions> | Microsoft | edX | Free | | Introduction to Hadoop and MapReduce | <https://www.udacity.com/course/intro-to-hadoop-and-mapreduce--ud617> | cloudera | Udacity | Free | | Analytics Storytelling for Impact | <https://www.edx.org/course/analytics-storytelling-for-impact-2> | Microsoft | edX | Free | | Towards Data Science | <https://towardsdatascience.com/> | Towards Data Science | Medium | £40pa | | Quora Data Science | <https://www.quora.com/search?q=data%20science> | Quora | Quora | Free | | QGIS 3.0 for GIS Professionals | <https://www.udemy.com/course/qgis-for-gis-professionals/> | Udemy | Udemy | £25 | |
| Table 1 – 12 High quality, online, low cost resources for learning data analytics and data science (Raval, 2018)[[6]](#footnote-6) |

According to Gartner, a world-leading research and advisory company, there are two key categories of toolsets relating to Data Analytics –

* Data Science and Machine Learning Platforms (Gartner (1), 2020)
* Analytics and Business Intelligence Platforms (Gartner (2), 2020)

Gartner frequently publish “Magic Quadrants” that separate the key platform providers into Niche Players, Visionaries, Challengers and Leaders together with the details of the strengths and weaknesses of each product -

|  |
| --- |
| A screenshot of a map  Description automatically generated |
| Figure 61 – Gartner Magic Quadrant for Data Science and Machine Learning Platforms (Gartner (1), 2020) |
| A screenshot of a cell phone  Description automatically generated |
| Figure 62 – Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (Gartner (2), 2020) |

Note: the author has overlaid the 2020 data on top of the 2019 data and added the trajectories of the major players to highlight recent changes in Gartner’s evaluations.

The positioning on the Gartner magic quadrant is a useful indicator when selecting platforms and toolsets, however there are other considerations. For example, some of the platforms are free, others are free to education and other are very expensive.

In addition to features, reputation and cost a transparent provider of data analytics will need to be able to show adherence to “Ethical and Moral Considerations” (Cambridge (9), 2019) including demonstrating how the results were calculated and obtained.

In this context it is important to note that many of the higher order analytics platforms, for example the RapidMiner platform (RapidMiner, 2020), do not have the capability to explain or justify their results because of their higher automation and lower control.

Taking all these factors into consideration the recommendations for the toolsets and platforms to be adopted by Lincolnshire Data Analytics Ltd. are –

* The **Python programming language** for data analytics, data science and machine learning.
* **Jupyter Notebooks** for visualising results and for collaborating on shared analytics projects.
* **QGIS** for geographical analysis, mapping and visualisation.
* **Microsoft Power BI** for high-quality visualisation, presentation and analytical story telling.
* **SQL Server** for relational database work.
* **Microsoft Excel** for small scale analyses, basic and exploratory visualisation.
* **RapidMiner for Education** to compare models with the high-end automated tools for the purposes of validation and verification of results.

Another category of resource are the data sources that will be identified and used in the delivery of education and commercial work. Table 3 provides details of a range of high-value datasets that will provide significant value to any education or commercial data analytics project.

Lincolnshire Data Analytics Ltd. will require premises and the intention was to house the data scientists in the business incubation hub within the Lincoln College Gibney Building however that project was paused in March 2020 due to the COVID-19 outbreak. Office space will be found at no cost within the Lincoln College main campus but the precise details could not be finalised at the time of writing.

The remaining resource needs are modest. As a sub-company within the Lincoln College structure, Lincolnshire Data Analytics Ltd. will be able to access the Microsoft Campus Agreement which provides access to almost all Microsoft applications software including E-Mail, Office 365, Power BI and SQL Server.

All the selected data science and business intelligence platforms are free or free to education, hence the only capital resources required are £2,564.95 per data scientist for computer hardware –

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Item | Cost | | [Microsoft Surface Laptop 13.5” screen, i7 Processor, 16 GB RAM, 500 GB Hard Drive](https://www.johnlewis.com/microsoft-surface-laptop-3-intel-core-i7-processor-16gb-ram-512gb-ssd-13-5-pixelsense-display/p4755463?sku=238340314&colour=Black&s_ppc=2dx92700046622704901&tmad=c&tmcampid=2&gclid=Cj0KCQjwybD0BRDyARIsACyS8muwp5VxUJLpCCkSPf7DNSfy06a2lxuGYwVyuOcIZlVXUiBQ1sAJGmAaApEmEALw_wcB&gclsrc=aw.ds) | £1,929.00 | | [Microsoft Surface Laptop Carrying Case](https://www.microsoft.com/en-gb/p/maroo-premium-leather-sleeve-for-surface-book-2/8vk7lw90t8fk?cid=msft_web_collection) | £59.99 | | [Microsoft Surface Docking Station](https://www.microsoft.com/en-gb/p/surface-dock/8QRH2NPZ0S0P/DJV0?source=googleshopping&ef_id=Cj0KCQjwybD0BRDyARIsACyS8ms0ORD27Ialr4e95qsybqjHyKnE-ZVQd1BN-3XdZRKpFoPJL9LqWG4aAiP9EALw_wcB%3aG%3as&s_kwcid=AL!4249!3!357185421760!!!g!816389144294!&OCID=AID2000007_SEM_Cj0KCQjwybD0BRDyARIsACyS8ms0ORD27Ialr4e95qsybqjHyKnE-ZVQd1BN-3XdZRKpFoPJL9LqWG4aAiP9EALw_wcB%3aG%3as&activetab=pivot%3aoverviewtab) | £189.99 | | [Bluetooth External Keyboard](https://www.amazon.co.uk/Microsoft-WS2-00003-Surface-Bluetooth-Keyboard/dp/B06WLLTDYF), [Bluetooth External Mouse](https://www.amazon.co.uk/TECKNET-Bluetooth-Adjustable-Wireless-Cordless/dp/B07H4JBRV1/ref=sr_1_4?dchild=1&keywords=bluetooth+mouse+for+laptop&qid=1586271669&s=computers&sr=1-4) and [28" 4K LED Monitor](https://www.amazon.co.uk/AOC-U2879VF-Inches-LED-Monitor/dp/B0163JLIWU) | £385.97 | | TOTAL | **£2,564.95** | |
| Table 2 – Computer Requirements for Each Data Scientist in Lincolnshire Data Analytics Ltd. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Dataset | URL / Link | Type | Provider | Purpose | | Indices of Multiple Deprivation (IMD) 2019 | <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019> | Public | Gov.uk | Demographic modelling | | Population Estimates | <https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=2010> | Public | Office for National Statistics | Population modelling | | Lower Super Output Area Boundaries | <https://geoportal.statistics.gov.uk/datasets/lower-layer-super-output-areas-december-2011-boundaries-ew-bfe> | Public | Office for National Statistics | Geographic modelling | | LSOA to Postcode Mapping | <https://data.gov.uk/dataset/9b090605-9861-4bb4-9fa4-6845daa2de9b/postcode-to-output-area-to-lower-layer-super-output-area-to-middle-layer-super-output-area-to-local-authority-district-february-2018-lookup-in-the-uk> | Public | Gov.uk | Geographic modelling | | Administrative Boundaries | <https://www.ordnancesurvey.co.uk/business-government/products/boundaryline> | Public | Ordnance Survey | Geographic modelling | | Income | <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/adhocs/11426annualsurveyofhoursandearningsasheestimatesofgrossannualandweeklyearningsforlocalauthorityby2digitoccupationuk2019> | Public | Office for National Statistics | Demographic modelling | | People in Work | <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork> | Public | Office for National Statistics | Demographic modelling | | National Public Transport Access Nodes | <https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan> | Public | Gov.uk | Travel modelling | | Office for National Statistics Home | <https://www.ons.gov.uk/> | Public | Office for National Statistics | Various | | Educational Establishments | <https://get-information-schools.service.gov.uk/> | Public | Gov.uk | Competition modelling | | Further Education and Skills | [https://www.gov.uk/government/collections/further-education-and-skills-statistical-first-release-sfr?utm\_source=c4f422c4-6479-41f6-b2fd-4052ba5080bb&utm\_medium=email&utm\_campaign=govuk-notifications&utm\_content=immediate#latest-releases](https://www.gov.uk/government/collections/further-education-and-skills-statistical-first-release-sfr?utm_source=c4f422c4-6479-41f6-b2fd-4052ba5080bb&utm_medium=email&utm_campaign=govuk-notifications&utm_content=immediate%23latest-releases) | Public | Gov.uk | Sector modelling | | FE Research and Statistics | [https://www.gov.uk/education/further-and-higher-education-skills-and-vocational-training#research\_and\_statistics](https://www.gov.uk/education/further-and-higher-education-skills-and-vocational-training%23research_and_statistics) | Public | Gov.uk | Sector modelling | | EMSI Economic and Education | [https://e.economicmodeling.com/analyst/?t=2zvj2#h=NjRt8&page=home](https://e.economicmodeling.com/analyst/?t=2zvj2%23h=NjRt8&page=home) | Subscription | EMSI | Curriculum modelling | | Lincoln College Local MIS Data | Not applicable | Private | Advanced | Internal modelling | | Kaggle Datasets for Beginners | <https://www.kaggle.com/m2skills/datasets-and-tutorial-kernels-for-beginners> | Public | Kaggle | Training & Learning | |
| Table 3 – 15 key datasets for delivery of education and commercial data analytics projects |

### Project Approach and Methodology

#### The Recommended Approach and Methodology for Acquiring and Operating the Overall Data Analytics Capability

The recommended high-level framework to acquire, operate and evaluate the proposed data analytics capability is the “Innovation Lifecycle Model” (Viki, et al., 2017).

|  |
| --- |
|  |
| Figure 63 – Innovation Lifecycle Model (Viki, et al., 2017) |

The reason for choosing this model over a more conventional business and cost case are the high degree of innovation, the high number of unknowns, the novelty of the solution and the risk reward profile. The (Viki, et al., 2017) model is suited to an environment exhibiting these attributes.

However this does not mean that the scrutiny associated with conventional business cases is absent but rather that the value of the new capability will be constantly evaluated and the options to continue, scale up, scale down, close down or to pivot the business model will be frequently and robustly reviewed.

The issues identified proposed by (Rogati, 2017) identified in research question 5 are now revisited and applied to validate the proposal and recommendations -

**Goals**: why do you need a data team and what will the team’s immediate goals be?

A data analytics capability is needed to enable Lincoln College to leverage the benefits that are evident in other sectors.

The immediate goals will be to deliver early projects that align with the needs of the funding stakeholders and to provide rapid, tangible, measurable benefits and outcomes including –

1. Improved processes.
2. Improved competitive position.
3. New and improved products, stemming from better customer and market data.
4. Informationalization, or building data into products and services.
5. Improved human capabilities.
6. Improved risk management.

(Ladley & Redman, 2020)

**Roles**: what roles will be required within the team and what will the composition of those roles be?

Initially two data scientists and a part-time Director of Data Analytics will be recruited with further capacity provided by an out-sourced provider.

These roles have been defined in detail in the “Business Model and Structure” section and visualised in Figure 54 and Figure 55.

**Leadership**: has a strong leader been identified who can attract and recruit the team, define the strategy and work with the founders or shareholders?

Yes, the Director of Data Analytics role will be fulfilled by the author.

**Timing**: when is the best time to start building the team and in what order?

The exact timing has become less clear following the outbreak of COVID-19 which has interrupted normal business activity. However, the intention is still to begin building the team following allocation of 2020/21 budgets on 1st August 2020.

**Organisation**: how will the data team fit into the organisation?

The team will be organised as a standalone company within the group company structure (see Figure 54) which will enable a flexible approach to recruitment and reward whilst providing a governance structure that simultaneously allows for educational and commercial engagements.

**Data culture**: does the company understand data analytics, and will it prioritise it?

Through a sustained period of raising management awareness the company does have a good grounding in understanding data analytics. If authority is given to proceed there is support from the CEO and the leadership team that will facilitate prioritisation.

**Infrastructure**: does the company have the necessary infrastructure to support a data team? If not what skills, tools, and capabilities are needed to provide it and how long will it take?

The company has a solid base of management information systems, databases and access to software and systems. The “Toolsets, Datasets and Resources” section has provided details of further infrastructure and how to acquire it.

**Recruiting and retention**: how will talented data scientists be attracted in an environment of high demand and skills shortages?

The “Staff Recruitment, Development and Retention” section has provided the details of how recruitment and retention will be addressed.

The (Rogati, 2017) principles overlap with and augment the key activities defined in the business model –

* Establish ELT authorisation.
* Establish budget.
* Obtain funding.
* Recruit and train staff.
* Establish key stakeholder relationships.
* Marketing and sales.
* Prioritise, agree and deliver initial projects.
* Develop local presence.

The two can be taken together to develop an initial project plan of the key tasks and activities (see Figure 64).

The project plan, like the overall business model, must be flexible and adaptive and will change constantly but as with the business case this does not mean an absence of rigorous and robust commercial standards and principles.

#### The Recommended Approach and Methodology for Running and Operating Individual Data Analytics Projects

(Agasisti & Bowers, 2017) proposed a model for delivering data analytics projects within the education context which is also applicable to the commercial context –

|  |
| --- |
|  |
| Figure 65 – A Project Model for Data Analytics within Education (Agasisti & Bowers, 2017) |

This model can be improved through some augmentation and extension identified during the research and the pilot project.

Prior to “Collection and acquisition (of data)” a successful data analytics project requires that the problem be clearly framed (Cambridge (1), 2019) and that a set of hypotheses be defined and agreed that the project will set out to either prove or disprove using data including a null hypothesis[[7]](#footnote-7) (Howitt, 2020).

|  |
| --- |
| A screenshot of a social media post  Description automatically generated |
| Figure 64 – Lincolnshire Data Analytics Initial Project Plan and Timeline |

Hence prior to stage 1 the key project stakeholders will agree the hypotheses, for example –

* **Hypothesis 0 (null hypothesis):** Declining recruitment is due entirely to a fall in the number of 16-18 year olds in the catchment area.
* **Hypothesis 1:** Removal of the education maintenance grant reduced recruitment/retention.

The (Agasisti & Bowers, 2017) stages 3 – 5 work well where the data analytics project is utilising descriptive analytics (i.e. looking backwards).

However, where the project is a forwards-looking, predictive analytics project utilising machine learning algorithms and techniques stages 3-5 are too simplistic to be effective.

In these cases, steps 3-5 will be substituted for the (Koehrsen, 2018) proposal for a machine learning workflow –

1. Data cleaning and formatting
2. Exploratory data analysis
3. Feature engineering and selection
4. Compare several machine learning models on a performance metric
5. Perform hyperparameter tuning on the best model
6. Evaluate the best model on the testing set
7. Interpret the model results
8. Draw conclusions and document work

(Koehrsen, 2018)

A further augmentation is that within the (Koehrsen, 2018) lifecycle step 4 can be guided by a set of principles defined in the RapidMiner data mining platform’s online tutorials –

|  |
| --- |
|  |
| Figure 66 – RapidMiner Tutorial Guidance for Machine Learning Model Selection (RapidMiner, 2017) |

Finally (Agasisti & Bowers, 2017) step 6 can be significantly expanded and improved as it is during this stage where the findings are presented back to the customers, hence it will be critical in establishing the success of the project.

This will also become one of the unique selling points of Lincolnshire Data Analytics. All feedback to customers will be delivered in the form of “User Storytelling” as defined in the Microsoft / EdX course “[Analytics Storytelling for Impact](https://www.edx.org/course/analytics-storytelling-for-impact-2)”.

This analytics delivery method ensures the findings are communicated in a way that is informative, enlightening, engaging, entertaining and easy-to-understand.

The Microsoft course sates that analytics story telling is a way to connect, transfer knowledge and win hearts-and-minds and highlights that the entire data analytics process will not achieve its outcomes unless the key stakeholders understand and are bought into the findings.

|  |
| --- |
| Machine generated alternative text: A story is a way to...  • Connect  • Transfer knowledge  • Win hearts and minds  2:00 / 2:17  Speed  1.50x  o  * |
| Figure 67 – The Microsoft / EdX free online course “[Analytics Storytelling for Impact](https://www.edx.org/course/analytics-storytelling-for-impact-2)” |

# Chapter 7: Conclusion and Next Steps

Through application of the scientific method this report has addressed the management problem of how to acquire a data analytics capability for Lincoln College despite a starting position of little knowledge or experience and in the context of significant financial, logistical and operational constraints in the FE sector.

The next steps are to secure the internal support and authorisation to proceed from the Chief Executive Officer and the Executive Leadership Team and crucially to secure any funding that may be available to contribute to the start-up and running costs.

The author has invested significant effort in learning about data analytics and data science and how it can be applied to inform and improve outcomes for Lincoln College.

I sincerely believe in the opportunity that this project and report have uncovered and I will now endeavour to move it from a theoretical report into a real-world part commercialised venture with the potential to deliver benefits for Lincoln College, the wider Further Education sector and the employers and economy of the Lincolnshire region.

# References

Advanced, 2020. *Suite of Further Education Solutions.* [Online]   
Available at: https://www.oneadvanced.com/solutions/prosuite/  
[Accessed 6th April 2020].

Agasisti, T. & Bowers, A. J., 2017. Data Analytics and Decision-Making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions. In: *Handbook of Contemporary Education Economics.* Cheltenham, UK: Edward Elgar Publishing Inc., pp. 184-210.

Alberg, R., 2019. *Why FE colleges need “big data” in the battle to recruit apprentices.* [Online]   
Available at: https://feweek.co.uk/2019/07/06/why-fe-colleges-need-big-data-in-the-battle-to-recruit-apprentices/

Anderson, C., 2008. *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.* [Online]   
Available at: https://www.wired.com/2008/06/pb-theory/  
[Accessed 28 January 2020].

Anderson, D., 2019. *Technical Skills Are Overrated. Focus on Your Attitude..* [Online]   
Available at: https://medium.com/s/please-advise/attitude-over-aptitude-4e8363795ff8  
[Accessed 6th April 2020].

AoC / ETF Data Science, 2019. *Data Science Conference.* London, s.n.

Ariely, D., 2013. *Dan Ariely's Facebook Page.* [Online]   
Available at: https://www.facebook.com/dan.ariely/posts/904383595868  
[Accessed 28 January 2020].

Association of Colleges, 2019. *The new college insolvency regime.* [Online]   
Available at: https://www.aoc.co.uk/news/the-new-college-insolvency-regime  
[Accessed 27 January 2020].

BBC, 2015. *TalkTalk hack 'affected 157,000 customers'.* [Online]   
Available at: https://www.bbc.co.uk/news/business-34743185  
[Accessed 27 January 2020].

BBVA, 2017. *The five V’s of big data.* [Online]   
Available at: https://www.bbva.com/en/five-vs-big-data/  
[Accessed 28 January 2020].

Bekker, A., 2017. *4 Types of Data Analytics to Improve Decision-Making.* [Online]   
Available at: https://www.scnsoft.com/blog/4-types-of-data-analytics  
[Accessed 28 January 2020].

Bertolucci, J., 2013. *Big Data Analytics: Descriptive Vs. Predictive Vs. Prescriptive.* [Online]   
Available at: https://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-descriptive-vs-predictive-vs-prescriptive/d/d-id/1113279  
[Accessed 28 January 2020].

Blended Learning Consortium, 2016. *Blended Learning Consortium.* [Online]   
Available at: http://www.blc-fe.org/  
[Accessed 17 February 2020].

Bock, T., n.d. *What is Correlation?.* [Online]   
Available at: https://www.displayr.com/what-is-correlation/  
[Accessed 6 February 2020].

Bragdon, T., n.d. *When data gives the wrong solution.* [Online]   
Available at: https://www.trevorbragdon.com/blog/when-data-gives-the-wrong-solution  
[Accessed 02 May 2020].

Cambridge (0), U. o., 2019. *Business Analytics: Decision Making Using Data - Introductory Overview,* Cambridge: University of Cambridge.

Cambridge (1), U. o., 2019. *Business Analytics: Decision Making Using Data, Module 1 - Decision Biases,* Cambridge: University of Cambridge.

Cambridge (3), U. o., 2019. *Business Analytics: Decision Making Using Data, Module 3 - Big Data Opportunities,* Cambridge: University of Cambridge.

Cambridge (4), U. o., 2019. *Business Analytics: Decision Making Using Data, Module 4 - Experimentation,* Cambridge: University of Cambridge.

Cambridge (5), U. o., 2019. *Business Analytics: Decision Making Using Data, Module 5 - Predictive Analytics,* Cambridge, UK: University of Cambridge.

Cambridge (7), U. o., 2019. *Business Analytics: Decision Making Using Data, Module 7 - Prescriptive Analytics I,* Cambridge: University of Cambridge.

Cambridge (9), U. o., 2019. *Business Analytics: Decision Making Using Data, Module 9 - Ethics, Legal and Organisational Issues,* Cambridge, UK: University of Cambridge.

Cambridge Dictionary, 2020. *Cambridge Dictionary.* [Online]   
Available at: https://dictionary.cambridge.org/dictionary/english/analytics  
[Accessed 28 January 2020].

Cornelissen, J., 2018. *The Democratization of Data Science.* [Online]   
Available at: https://hbr.org/2018/07/the-democratization-of-data-science  
[Accessed 9th April 2020].

Coursera, 2020. *Your Course to Success.* [Online]   
Available at: https://www.coursera.org/  
[Accessed 17 February 2020].

Davenport, T., 2018. *DELTA Plus Model & Five Stages of Analytics Maturity: A Primer,* s.l.: International Institute for Analytics.

Davenport, T. H. & Harris, J. G., 2017. *Competing on Analytics: The Science of Winning.* Boston, MA: Harvard Business Review Press.

EdX, Microsoft, 2020. *Analytics Storytelling for Impact.* [Online]   
Available at: https://courses.edx.org/courses/course-v1:Microsoft+DAT248x+1T2020/course/  
[Accessed 9 February 2020].

EdX, 2020. *EdX Home Page.* [Online]   
Available at: https://www.edx.org/  
[Accessed 17 February 2020].

Emeritus, 2020. *Business Analytics: Decision Making Using Data.* [Online]   
Available at: https://emeritus.org/management-certificate-programs/business-analytics/  
[Accessed 19 February 2020].

EMSI Economic Modelling, 2020. *Maximise Your College's Potential.* [Online]   
Available at: https://www.economicmodelling.co.uk/further-education/  
[Accessed 18 February 2020].

Enck, R. E., 2012. The OODA Loop. *Home Health Care Management & Practice,* Vol 24(Issue 3), pp. 123-124.

ESFA, 2019. *ESFA financial management: college accounts.* [Online]   
Available at: https://www.gov.uk/guidance/esfa-financial-management-college-accounts  
[Accessed 27 January 2020].

European Leadership University, 2020. *Careers in Data Science: Data Analyst vs Data Engineer vs Data Scientist.* [Online]   
Available at: https://elu.nl/careers-in-data-science-data-analyst-vs-data-engineer-vs-data-scientist/  
[Accessed 9 February 2020].

Fink, A., 2020. *Conducting Research Literature Reviews From the Internet to Paper, Fifth Edition.* Los Angeles: SAGE Publications Ltd.

Forbes Insights, 2015. *Betting on Big Data,* Unknown: Forbes.

Forbes, 2017. *IBM Predicts Demand For Data Scientists Will Soar 28% By 2020.* [Online]   
Available at: https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#72f34ceb7e3b  
[Accessed 29 January 2020].

Frontier Economics, 2017. *Further Education Workforce Data for England,* s.l.: Frontier Economics.

Furr, N., 2011. *Who Cares About Startups: Why Everyone Must Be an Entrepreneur.* [Online]   
Available at: https://www.forbes.com/sites/nathanfurr/2011/04/11/who-cares-about-startups-why-everyone-must-be-an-entrepreneur/#22116e293915  
[Accessed 18 February 2020].

Gallo, A., 2018. The Fundamentals of AB Testng. In: *Data Analytics Basics for Managers.* Boston, MA: Harvard Business Review Press, pp. 59-70.

Gartner (1), 2020. *Magic Quadrant for Data Science and Machine Learning Platforms.* [Online]   
Available at: https://www.gartner.com/doc/reprints?id=1-1YCTPMUL&ct=200213&st=sb  
[Accessed 7th April 2020].

Gartner (2), 2020. *Magic Quadrant for Analytics and Business Intelligence Platforms.* [Online]   
Available at: https://www.gartner.com/doc/reprints?id=1-1XYUYQ3I&ct=191219&st=sb  
[Accessed 7th April 2020].

Gigabit, 2019. *What we can learn from the biggest hacks in history.* [Online]   
Available at: https://www.gigabitmagazine.com/big-data/what-we-can-learn-biggest-hacks-history  
[Accessed 27 January 2020].

Glassdoor, 2020. *Your Glassdoor Dashboard.* [Online]   
Available at: https://www.glassdoor.co.uk/index.htm  
[Accessed 29 January 2020].

Google Nasdaq, 2020. *Netflix Share Price.* [Online]   
Available at: https://tinyurl.com/thtfok8

Gottschalk, P., 2013. Policing White-Collar Crime. *Political Science.*

Gov.uk (1), 2018. *Guide to the General Data Protection Regulation.* [Online]   
Available at: https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation

Gov.uk (2), 2020. *Individualised Learner Record (ILR).* [Online]   
Available at: https://www.gov.uk/government/collections/individualised-learner-record-ilr  
[Accessed 17 February 2020].

Grant, R. M., 2015. *Contemporary Strategy Analysis.* Text and Cases Edition, 9th Edition, Kindle Edition. ed. Padstow, Cornwall: Wiley.

Granville, V., 2017. *Difference between Machine Learning, Data Science, AI, Deep Learning, and Statistics.* [Online]   
Available at: https://www.datasciencecentral.com/profiles/blogs/difference-between-machine-learning-data-science-ai-deep-learning  
[Accessed 18 February 2020].

Gray, J., 2019. *Colleges won’t see the wood for the trees without big data.* [Online]   
Available at: https://feweek.co.uk/2019/11/02/colleges-wont-see-the-wood-for-the-trees-without-big-data/

Grossman, R. L. & Siegel, K. P., 2014. Organisational Models for Big Data and Analytics. *Journal of Organization Design,* 3(1), pp. 20-25.

Gunther-McGrath, R., 2008. *A New Approach to Innovation Investment.* [Online]   
Available at: https://hbr.org/2008/03/a-new-approach-to-innovatio  
[Accessed 27 January 2020].

Gupta, M. & Georgeb, J. F., 2016. Toward the development of a big data analytics capability. *Information & Management,* 53(8), pp. 1049-1064.

Hagiu, A. & Wright, J., 2020. When Data Creates Competitive Advantage. *Harvard Business Review,* Issue January-February 2020, pp. 94-101.

Hart, C., 2014. *Doing a Literature Review: Releasing the Social.* London: SAGE Publications Ltd.

Harvard Economic Society, 1929. *1929 US Economic Crash,* s.l.: Harvard Economic Society.

Harvard University (1), 2020. *Strategic Data Project - Transform Your Use of Education Data.* [Online]   
Available at: https://sdp.cepr.harvard.edu/

Harvard University (2), 2020. *Strategic Data Project - Our Impact.* [Online]   
Available at: https://sdp.cepr.harvard.edu/our-impact

HBR Press, 2018. *Data Analytics Basics for Managers.* Boston, MA: Harvard Business Review Press.

Holden, T., 2019. *Leadership Lecture, Determined to Lead Workshop.* Lincoln: s.n.

Holt, A. & Schools, T. P., 2015. *Early Elementary On-Track Indicators Leading to Third-Grade Reading Proficiency,* Cambridge, MA: Strategic Data Project Capstone Report.

Howitt, M., 2020. *Interview with the CEO of "the Data Place"* [Interview] (16 January 2020).

IBM Watson, 2020. *Personality Insights.* [Online]   
Available at: https://www.ibm.com/watson/services/personality-insights/  
[Accessed 18 February 2020].

Informatica, 2020. *What is Data Analytics?.* [Online]   
Available at: https://www.informatica.com/gb/services-and-training/glossary-of-terms/data-analytics-definition.html#fbid=Kvegj0acGTB  
[Accessed 28 January 2020].

InterHigh, 2020. *The UK's Leading Online School.* [Online]   
Available at: https://interhigh.co.uk/  
[Accessed 17 February 2020].

Jain, A., 2018. *MBA Managing People Module - Lecture 2: Individuals in the Organisation.* Nottingham: University of Nottingham Business School.

JISC, 2020. *Analytics labs.* [Online]   
Available at: https://www.jisc.ac.uk/analytics-labs  
[Accessed 17 February 2020].

Johnes, G. J. ,. J., Agasisti, T. & Lopez-Torres, L., 2017. *Handbook of Contemporary Education Economics.* Cheltenham, UK: Edward Elgar Publishing Inc..

Kassarnig, V. et al., 2017. Class attendance, peer similarity, and academic performance in a large field study. *Plos One.*

Khan Academy (1), 2012. *Correlation and causality.* [Online]   
Available at: https://www.khanacademy.org/math/probability/scatterplots-a1/creating-interpreting-scatterplots/v/correlation-and-causality  
[Accessed 28 January 2020].

Khan Academy (2), 2020. *Khan Academy.* [Online]   
Available at: https://www.khanacademy.org/  
[Accessed 17 February 2020].

Koehrsen, W., 2018. *A Complete Machine Learning Project Walk-Through in Python: Part One.* [Online]   
Available at: https://towardsdatascience.com/a-complete-machine-learning-walk-through-in-python-part-one-c62152f39420  
[Accessed 8th April 2020].

Kohavi, R. & Thomke, S., 2017. The Surprising Power of Online Experiments. *Harvard Business REview,* September-October 2017(1).

Kotter, J. P., 2018. *Accelerating Change: A PowerPoint Presentation Based on the Work of John P. Kotter.* [Online]   
Available at: https://hbr.org/product/accelerating-change-a-powerpoint-presentation-based-on-the-work-of-john-p-kotter/9467TL-PPT-ENG  
[Accessed 5th May 2018].

Ladley, J. & Redman, T. C., 2020. *Use Data to Accelerate Your Business Strategy.* [Online]   
Available at: https://hbr.org/2020/03/use-data-to-accelerate-your-business-strategy  
[Accessed 8th April 2020].

Lamb, R., 2018. *MBA Lecture: Innovation Management - Managing Uncertainty and Failure.* Nottingham, University of Nottingham, Business School.

Lincoln College STEP Review Team, 2016. *Options Appraisal, Business Case, Design - Management Information Systems and Services,* Lincoln: Lincoln College.

LinkedIn Learning, 2020. *LinkedIn Learning Courses | Official Site‎.* [Online]   
Available at: https://www.linkedin.com/learning/me  
[Accessed 17 February 2020].

MacGarvie, M. & McElheran, K., 2018. Pitfalls of Data-Driven Decisions. In: *Data Analytics Basics for Managers.* Boston, MA: Harvard Business Review Press, pp. 155-164.

Macrotrends, 2020. *Netflix Market Cap 2006-2019 | NFLX.* [Online]   
Available at: https://www.macrotrends.net/stocks/charts/NFLX/netflix/market-cap

Mc.ai, 2018. *Deep Learning Demystified.* [Online]   
Available at: https://mc.ai/deep-learning-demystified/  
[Accessed 28 January 2020].

McAfee, A. & Erik, B., 2012. Big Data: The Management Revolution. *Harvard Business Review,* October 2012(1).

McKinsey & Company (1), 2019. *Catch them if you can: How leaders in data and analytics have pulled ahead,* Unknown: McKinsey & Company.

McKinsey & Company (2), 2017. *Fueling growth through data monetization - December 2017 Survey,* Unknown: McKinsey & Company.

McNulty, K., 2018. *How to build an analytics team for impact in an organization.* [Online]   
Available at: https://towardsdatascience.com/how-to-build-an-analytics-team-for-impact-in-an-organization-21bb05925587  
[Accessed 9 February 2020].

Morris, H., 2003. *Predictive Analytics and ROI: Lessons from IDC's Financial Impact Study,* Framingham, MA: International Data Corporation.

Nucleus Research, 2014. *Analytics Pays Back £13.01 for Every Dollar Spent,* Boston, MA: Nucleus Research Inc..

Office for National Statistics (1), 2020. *16-17 year old population: All persons.* [Online]   
Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/timeseries/jn5p/lms  
[Accessed 27 January 2020].

Office for National Statistics (2), 2020. *Explorable Datasets.* [Online]   
Available at: https://www.ons.gov.uk/searchdata?q=explorable%20datasets  
[Accessed 18 FEbruary 2020].

Ofsted, 2019. *The Education Inspection Framework.* [Online]   
Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/801429/Education\_inspection\_framework.pdf

Open University, 2020. *The Open University.* [Online]   
Available at: https://www.open.ac.uk/  
[Accessed 17 February 2020].

Pearson, T. & Wegener, R., 2013. *Big Data: The Organizational Challenge.* [Online]   
Available at: https://www.bain.com/insights/the-value-of-big-data/  
[Accessed 17 February 2020].

Porter, M. E., 2008. The Five Competitive Forces That Shape Strategy. *Harvard Business REview,* January, January 2008(1), p. Unknown.

Provost, F. & Fawcett, T., 2013. *Data Science for Business.* First Edition, Second Release ed. Sebastopol, CA: O'Reilly.

Radhika, G., 2018. *Top 10 Advantages of Data Analytics for Education Industry.* [Online]   
Available at: https://www.loginworks.com/blogs/top-10-advantages-of-data-analytics-for-education-industry/

RapidMiner, 2017. *Finding the Right Model.* [Online]   
Available at: https://www.youtube.com/watch?v=C8Ko3-2f-pA&list=PLssWC2d9JhOZLbQNZ80uOxLypglgWqbJA&index=16  
[Accessed 8th April 2020].

RapidMiner, 2020. *Lightning Fast Business Impact with AI.* [Online]   
Available at: https://rapidminer.com/  
[Accessed 20 February 2020].

Raval, S., 2018. *Learn Data Science in 3 Months.* [Online]   
Available at: https://www.youtube.com/watch?v=9rDhY1P3YLA&t=3s  
[Accessed 7th April 2020].

Reynolds, A. & Lewis, D., 2017. *Teams Solve Problems Faster When They’re More Cognitively Diverse.* [Online]   
Available at: https://hbr.org/2017/03/teams-solve-problems-faster-when-theyre-more-cognitively-diverse  
[Accessed 29 January 2020].

Ritter, D., 2014. *When to Act on a Correlation, and When Not To.* [Online]   
Available at: https://hbr.org/2014/03/when-to-act-on-a-correlation-and-when-not-to  
[Accessed 28 January 2020].

Rogati, M., 2017. *What are the challenges of building a data team at a startup?.* [Online]   
Available at: https://www.quora.com/What-are-the-challenges-of-building-a-data-team-at-a-startup  
[Accessed 29 January 2020].

Saltz, J. S. & Grady, N. W., 2017. *The ambiguity of data science team roles and the need for a data science workforce framework.* s.l., s.n., pp. pp.2355-2361.

Seltzer, G., 2014. *Competing on Analytics.* [Online]   
Available at: https://www.slideshare.net/gseltzer/competing-on-analytics-38030796  
[Accessed 09 February 2020].

Shaw, D., 2018. *MBA Business Intelligence in the Digital Economy Module.* Nottingham, University of Nottingham Business School.

Slide Team, 2020. *Descriptive Diagnostic Predictive And Prescriptive Data Types.* [Online]   
Available at: https://www.slideteam.net/descriptive-diagnostic-predictive-and-prescriptive-data-types.html  
[Accessed 28 January 2020].

Sloman, J., Garratt, D., Guest, J. & Jones, E., 2016. *Economics for Business.* Harlow, UK: Pearson.

Stack Overflow, 2019. *Most Popular Programming Languages on Stack Overflow.* [Online]   
Available at: https://www.youtube.com/watch?v=cKzP61Gjf00  
[Accessed 19 February 2020].

Stillwell (1), D., 2019. *Big Data Opportunities (Webinar).* [Online]   
Available at: https://student.emeritus.org/courses/899/pages/webinars

Syed, M., 2019. *Rebel Ideas.* 1 ed. London: John Murray (Publishers).

Taymor, E., 2020. *When Should You Pull the Plug on an Innovation Project?.* [Online]   
Available at: https://www.innovationexcellence.com/blog/2017/10/25/when-should-you-pull-the-plug-on-an-innovation-project/  
[Accessed 18 February 2020].

The Business Channel, 2016. *The Business Model Canvas - 9 Steps to Creating a Successful Business Model - Startup Tips.* [Online]   
Available at: https://www.youtube.com/watch?v=IP0cUBWTgpY  
[Accessed 18 February 2020].

The Digital College, 2020. *Learn anywhere, anytime.* [Online]   
Available at: https://www.thedigitalcollege.co.uk/  
[Accessed 17 February 2020].

The Lincolnshire Chamber of Commerce, 2020. *Funding and Support for Lincolnshire Businesses.* [Online]   
Available at: https://www.lincs-chamber.co.uk/funding-and-support-for-lincolnshire-businesses/  
[Accessed 6th April 2020].

Towards data science, 2020. *Towards Data Science Home Page.* [Online]   
Available at: https://towardsdatascience.com/  
[Accessed 19 February 2020].

Utterback, J. M., 1994. *Mastering the Dynamics of Innovation.* Boston, MA: Harvard Business School Press.

Vigen, T., 2020. *Spurious Correlations.* [Online]   
Available at: https://www.tylervigen.com/spurious-correlations  
[Accessed 29 January 2020].

Viki, T., Toma, D. & Gons, E., 2017. *The Corporate Startup: How Established Companies Can Develop Successful Innovation Ecosystems.* Deventer, Netherlands: Vakmedianet.

Waller, D., 2020. *10 Steps to Creating a Data-Driven Culture.* [Online]   
Available at: https://hbr.org/2020/02/10-steps-to-creating-a-data-driven-culture  
[Accessed 9 February 2020].

Wayland, J., 2018. *If data science is in demand, why is it so hard to get a data scientist job?.* [Online]   
Available at: https://www.quora.com/If-data-science-is-in-demand-why-is-it-so-hard-to-get-a-data-scientist-job  
[Accessed 6th April 2020].

Wegener, R. & Sinha, V., 2003. *The Value of Big Data: How Analytics Differentiates Winners.* [Online]   
Available at: https://www.bain.com/contentassets/5672af3b82f84aa2a80ca732fa8ea06c/bain20\_brief\_the\_value\_of\_big\_data.pdf

Wu (1), M., 2013. *Big Data Reduction 1: Descriptive Analytics.* [Online]   
Available at: https://community.khoros.com/t5/Science-of-Social-Blog/Big-Data-Reduction-1-Descriptive-Analytics/ba-p/77766  
[Accessed 28 January 2020].

Wu (2), M., 2013. *Big Data Reduction 2: Understanding Predictive Analytics.* [Online]   
Available at: https://community.khoros.com/t5/Science-of-Social-Blog/Big-Data-Reduction-2-Understanding-Predictive-Analytics/ba-p/79616  
[Accessed 28 January 2020].

Wu (3), M., 2013. *Big Data Reduction 3: From Descriptive to Prescriptive.* [Online]   
Available at: https://community.khoros.com/t5/Science-of-Social-Blog/Big-Data-Reduction-3-From-Descriptive-to-Prescriptive/ba-p/81556  
[Accessed 28 January 2020].

Wu (4), M., 2013. *Are Your Big Data Analytics Actionable?.* [Online]   
Available at: https://community.khoros.com/t5/Science-of-Social-Blog/Are-Your-Big-Data-Analytics-Actionable/ba-p/129029  
[Accessed 28 January 2020].

# Acknowledgements

**Gary Headland**, Chief Executive Officer of Lincoln College; senior sponsor who has provided inspiration and enthusiasm and has been unfailingly generous with time to discuss and generate ideas for the project.

**David Corke**, Director of Policy for the Association of Colleges; David is an acknowledge “guru” or management information and funding in the FE sector and has provided views, advice and mentoring.

**Martin Howitt**, Co-founder and technology lead at The Data Place; Martin played a key role in the pilot project, has given many hours of mentoring support and has provided advice, views and help to shape the project.

**Lucy Knight**, Data Scientist at The Data Place; Lucy also played a key role in the pilot project and has helped to coach, train and develop, in particular spending time helping to develop skills in geographical analytics.

**Phil Wescott**, CEO of Filament Artificial Intelligence; Phil has spent time discussing and advising on the project proposal, provide help and advice from his vast experience in AI and analytics.

**Rob Carroll**, Director Catapult Ventures Group, Professor at Nottingham University Business School; Rob has been the management project tutor and has always been on hand with meetings, phone calls and emails to help steer the project.

**Dawn Barron**, Founder at Human Alchemy, Co-Founder at MyDogMyStyle and Trustee at YMCA Lincolnshire; Dawn has helped me with time and advice on many projects over the years and has helped to inspire confidence and self-belief.

**Karla Hoff**, Account Director, Further Education at Emsi; Karla has provided help in the form of the detailed economic modelling and analysis that underpinned the external market research.

**Matt Trinder**, Account Director at Advanced; Matt has provided a lot of help and support and has been useful to test ideas and to understand the sector.

# Index of Figures and Tables

[*Figure 1 – IBM Watson AI Service used with Python and Jupyter to profile Barack Obama from his public Internet Quotes* 5](#_Toc39319471)

[*Figure 2 – Visual Representation of Descriptive, Predictive and Prescriptive Analytics* (Slide Team, 2020) 6](#_Toc39319472)

[*Figure 3 – Visual Representation of Analytics* (Bekker, 2017)*,* (Mc.ai, 2018) 7](#_Toc39319473)

[*Figure 4 – Scatter Plot of the Relationship between Height and Weight in Females* 9](#_Toc39319474)

[*Figure 5 – Scatter Plot of the Relationship between a student’s attendance and the attendance of their peers / contacts* (Kassarnig, et al., 2017) 9](#_Toc39319475)

[*Figure 6 – The Correlation Between Margarine Consumption and Divorce Rates* (Vigen, 2020) 10](#_Toc39319476)

[*Figure 7 – When to Act on a Data Correlation* (Ritter, 2014) 11](#_Toc39319477)

[*Figure 8 – Decision Traps* (Cambridge (1), 2019) 12](#_Toc39319478)

[*Figure 9 – Rialto Police Shoulder Camera Experiment* (Cambridge (4), 2019) 13](#_Toc39319479)

[*Figure 10 – Availability Bias of Bullet Hole Patterns in Surviving Planes* (Bragdon, n.d.) 14](#_Toc39319480)

[*Figure 11 –Cognitive Diversity and Homogeneity in Problem Solving* (Syed, 2019) 15](#_Toc39319481)

[*Figure 12 –The Importance of Triangulating Sources of Information* (Holden, 2019) 16](#_Toc39319482)

[Figure 13 - Axes of Analytics Capability (Wegener & Sinha, 2003) 17](#_Toc39319483)

[Figure 14 – Top Performers Decision Making Capability (Wegener & Sinha, 2003) 17](#_Toc39319484)

[Figure 15 – Quantification of Effectiveness of firms with the Best Analytic Capabilities (Pearson & Wegener, 2013) 18](#_Toc39319485)

[Figure 16 – Recent Changes in the Nature of Industry Competition Brough About by Analytics (McKinsey & Company (1), 2019) 19](#_Toc39319486)

[Figure 17 - Effect of Big Data and Analytics on Revenues and Costs (Forbes Insights, 2015) 20](#_Toc39319487)

[Figure 18 – Importance of Analytical Orientation in High vs. Low Performing Firms (Davenport & Harris, 2017) 20](#_Toc39319488)

[Figure 19 – Netflix Prize Details and Timeline (Cambridge (5), 2019) 21](#_Toc39319489)

[Figure 20 - Change in Netflix market capitalisation (Macrotrends, 2020) 22](#_Toc39319490)

[Figure 21 - Change in Netflix share price (Google Nasdaq, 2020) 22](#_Toc39319491)

[Figure 22 – Changes to Industry Practices through Data Analytics (McKinsey & Company (2), 2017) 23](#_Toc39319492)

[Figure 23 – Financial Performance of UK FE Sector 2014 - 2017 (ESFA, 2019) 28](#_Toc39319493)

[Figure 24 – External Industry Analysis using Porter’s 5 Forces (Porter, 2008) 29](#_Toc39319494)

[Figure 25 – Gradually Declining Population Demographic (Office for National Statistics (1), 2020) 29](#_Toc39319495)

[Figure 26 – Kotter’s 8 Steps (Kotter, 2018) 30](#_Toc39319496)

[Figure 27 – Data Science and Analytics Skills and Jobs (Forbes, 2017) 31](#_Toc39319497)

[*Figure 28 – Data Analytics Jobs and Salaries in London* (Glassdoor, 2020) 32](#_Toc39319498)

[*Figure 29 – Median Salaries and Ages of Staff in the Further Education Sector* (Frontier Economics, 2017) 33](#_Toc39319499)

[Figure 30 – Data Analytics Roles and Team Expertise (Saltz & Grady, 2017) 34](#_Toc39319500)

[Figure 31 – Link Between Analytic Capabilities and Financial Performance (Pearson & Wegener, 2013) 36](#_Toc39319501)

[Figure 32 – High Performance of the “Composite Success Factors” Analytics Technique (Holt & Schools, 2015) 37](#_Toc39319502)

[Figure 33 – Using Tree Induction to Predict Grade 3 Student Performance (Holt & Schools, 2015) 38](#_Toc39319503)

[Figure 35 – Incumbent vs. Entrant Product Innovation (Utterback, 1994) 41](#_Toc39319504)

[Figure 36 – An interactive look at organizations’ use of data and analytics (McKinsey & Company (2), 2017) 42](#_Toc39319505)

[Figure 37 – The Knowledge Ladder; adapted from (Gottschalk, 2013) 44](#_Toc39319506)

[Figure 38 – Managing Risk in Innovation (Furr, 2011), (Lamb, 2018) 45](#_Toc39319507)

[Figure 39 – An Example Business Model Canvas 47](#_Toc39319508)

[Figure 40 – Driving the Success of Data Science Solutions: Skills Roles and Responsibilities (Granville, 2017) 48](#_Toc39319509)

[*Figure 41 – Sample Blogs and Posts on by the author on Data Analytics to External and Internal Professional Networks* 50](#_Toc39319510)

[Figure 42 – Roadmap to becoming an analytical competitor (Davenport & Harris, 2017) 53](#_Toc39319511)

[Figure 43 – Hypotheses and Findings from the Pilot Project 56](#_Toc39319512)

[Figure 44 – Certificate of Completion for Business Analytics: Decision Making Using Data 59](#_Toc39319513)

[Figure 45 – Most Popular Programming Languages Over Time (Stack Overflow, 2019) 59](#_Toc39319514)

[Figure 46 – Using tree induction to predict if patients will develop diabetes 60](#_Toc39319515)

[Figure 47 – Using a scatter-plot matrix or “SPLOM” to uncover patterns in English & Maths student attendance 60](#_Toc39319516)

[Figure 48 – Example Slides from the ELT Data Bias Presentation of 30th October 2019 61](#_Toc39319517)

[*Figure 49 – High Level Business Models for a Data Analytics Capability at Lincoln College* 63](#_Toc39319518)

[Figure 50 – Example User Story for Capturing Data Analytics Requirements (AoC / ETF Data Science, 2019) 65](#_Toc39319519)

[*Figure 51 – Statistical Analysis of the Management User Stories* 66](#_Toc39319520)

[*Figure 52 – External Market Analysis using EMSI Economic Modelling Data* (EMSI Economic Modelling, 2020) 69](#_Toc39319521)

[*Figure 53 – Hybrid Business Model* 71](#_Toc39319522)

[*Figure 54 – Proposed Governance Structure of Lincolnshire Data Analytics Ltd.* 73](#_Toc39319523)

[*Figure 55 – Proposed Initial Staffing of Lincolnshire Data Analytics Ltd.* 73](#_Toc39319524)

[*Figure 56 – Business Model Canvas for Lincolnshire Data Analytics Proposal* (The Business Channel, 2016) 74](#_Toc39319525)

[Figure 57 – Market Analysis of UK FE Student Records System (Lincoln College STEP Review Team, 2016) 75](#_Toc39319526)

[Figure 58 – Market Analysis of UK FE Student Records System (Lincoln College STEP Review Team, 2016) 77](#_Toc39319527)

[Figure 59 – Skewed distribution of the attractiveness (to employers) of data scientists in the jobs market (Wayland, 2018) 79](#_Toc39319528)

[Figure 60 – The “Equity Equation” (Jain, 2018) 79](#_Toc39319529)

[Figure 61 – Gartner Magic Quadrant for Data Science and Machine Learning Platforms (Gartner (1), 2020) 82](#_Toc39319530)

[Figure 62 – Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (Gartner (2), 2020) 82](#_Toc39319531)

[Figure 63 – Innovation Lifecycle Model (Viki, et al., 2017) 85](#_Toc39319532)

[Figure 65 – A Project Model for Data Analytics within Education (Agasisti & Bowers, 2017) 87](#_Toc39319533)

[Figure 64 – Lincolnshire Data Analytics Initial Project Plan and Timeline 88](#_Toc39319534)

[Figure 66 – RapidMiner Tutorial Guidance for Machine Learning Model Selection (RapidMiner, 2017) 89](#_Toc39319535)

[Figure 67 – The Microsoft / EdX free online course “Analytics Storytelling for Impact” 90](#_Toc39319536)

[Table 1 – 12 High quality, online, low cost resources for learning data analytics and data science (Raval, 2018) 81](#_Toc39319537)

[Table 2 – Computer Requirements for Each Data Scientist in Lincolnshire Data Analytics Ltd. 83](#_Toc39319538)

[Table 3 – 15 key datasets for delivery of education and commercial data analytics projects 84](#_Toc39319539)

1. No evidence could be found to corroborate this practitioner interview finding in the literature. However, “indicative analytics” has proved extremely useful in practice warranting its definition and inclusion. [↑](#footnote-ref-1)
2. Due to changes in the ESFA reporting a like-for-like comparison with 2017/18 data could not be calculated [↑](#footnote-ref-2)
3. “Start-ups” can be genuinely new firms or a start-up venture within an existing firm. [↑](#footnote-ref-3)
4. The educational maintenance grant is an historical government which ceased in 2010/11 which paid students a case incentive to stay in learning [↑](#footnote-ref-4)
5. EMSI agreed to help with the research by providing access to the key datasets [↑](#footnote-ref-5)
6. The resources suggested by (Raval, 2018) have been extended and augmented by the author. [↑](#footnote-ref-6)
7. It is common practice within the data science community to always add a null hypothesis that states that the status quo is correct and that nothing should change [↑](#footnote-ref-7)