Quantitative study on Barriers of adopting Big Data Analytics for UK and Eire SMEs

M. Willetts, A. S. Atkins, C. Stanier

School of Digital, Technologies and Arts, Staffordshire University, College Road, Stoke-on-Trent, ST4 2DE United Kingdom
matthew.willetts@research.staffs.ac.uk

Abstract – Big Data Analytics has been widely adopted by large companies, enabling them to achieve competitive advantage. However, SMEs (small and medium-sized enterprises) are underutilising this technology due to a number of barriers including financial constraints and lack of skills. Previous studies have identified a total of 69 barriers to SMEs adoption of Big Data Analytics, rationalised to 21 barriers categorised into five pillars (Willetts, Atkins and Stanier, 2020b). To verify the barriers identified from the literature, an electronic questionnaire was distributed to over 1000 SMEs based in the UK and Eire using the snowball sampling approach during the height of the COVID-19 pandemic. The intention of this paper is to provide an analysis of the questionnaire, specifically applying the Cronbach's alpha test to ensure that the 21 barriers identified are positioned in the correct pillars, verifying that the framework is statistically valid.

Keywords: Big Data Analytics, SMEs, barriers to Big Data Analytics adoption, strategic framework, COVID-19.

1. Introduction

SMEs account for 99.9% of all businesses in the UK, employ 60% of the workforce and generate £2,168 billion; this represents 52% of the turnover of all businesses in the UK (Rhodes, 2019). Similarly in Eire, SMEs consist of 99.8% of all businesses, 70.1% of employment and contribute € 91.9 billion, *41.5%* to value added (European Commission, 2020). This paper discusses the use of a questionnaire to collect primary data for use in the validation of the Big Data Analytics adoption framework for SMEs proposed by Willetts, Atkins and Stanier (2020a). The resulting data is then used to assess the internal consistency of the pillars of the strategic framework, using Cronbach's Alpha statistical analysis, to test the validity of the framework. This will allow poor internal consistency to be addressed by restructuring the framework. The individual barriers can then be further assessed, and ranked in order of relative importance, in order to identify those barriers that present challenging issues to the implementation of Big Data Analytics at SMEs.

The structure of this paper is as follows: Section 2 provides a literature review. Section 3 describes the construction, revision and distribution of the questionnaire. Section 4 outlines the data analysis, statistical techniques employed and the revision to the strategic framework. Section 5 provides a conclusion to the paper and discusses future work.

2. Literature Review

Big Data is defined as: 'an umbrella term used to describe a wide range of technologies that capture, store, transform and analyse complex data sets which can be of a high volume, generated at a high velocity in a variety of formats' (Willetts, Atkins and Stanier, 2020a, p. 3034). Big Data Analytics refers to the variety of software tools and techniques which are used to extract insights from Big Data sources. Mikalef et al. (2019, p. 262) state that a widely used definition of Big Data Analytics is: 'a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high velocity capture, discovery and/or analysis'. There are many case studies of large companies achieving a variety of benefits from the adoption of Big Data Analytics including significant savings from unplanned downtime (Mathew, 2016) and increased efficiencies (Sena *et al.*, 2019) but there are also case studies of SMES utilising the technology resulting in increased sales (Walsh, 2017). However, there are a number of barriers to SMEs adoption including lack of understanding, shortage of in-house data analytic expertise and financial barriers (Coleman *et al.*, 2016). Sixty nine barriers to Big Data Analytics have been identified through a previous literature review which were rationalised to 21 barriers through the utilisation of the thematic analysis process (Willetts, Atkins and Stanier, 2020b, 2020a). This study outlines a quantitative analysis and statistical validation of the results to develop a holistic assessment framework to assist SMEs in adopting Big Data analytics to provide competitive advantage.

3. Proposed Work

3.1. Research Design

A questionnaire was developed to validate the barriers identified from a thematic analysis (Willetts, Atkins and Stanier, 2020a). The questionnaire development approach documented in Moore and Benbasat (1991) was followed, which divides the process into three stages: item construction, reviewing process and testing.

1. **Stage One: Item Construction**

Sixty-nine barriers to SMEs adopting Big Data Analytics were identified from a literature review. A thematic analysis was performed which rationalised the barriers from 69 to 21 barriers and grouped them into 5 pillars: Business, Environmental, Human, Organisational and Technological. The pillars originated from three overlapping theoretical frameworks: Technology-Organisation Environment (TOE) (Tornatzky, Fleischer and Chakrabarti, 1990); Human, Technology, Organisation-fit (HOT-fit) (Yusof *et al.*, 2008) and the Information Systems Strategy Triangle (ISST) (Pearlson, 2001). TOE provides the Technological, Organisational and Environmental pillars, HOT-fit provides the Human pillar and ISST provides the Business pillar. Therefore, these were used as measurement items for the questionnaire.

One of the most widely reported barriers to SMEs adopting Big Data Analytics in the literature is lack of awareness. Therefore, it is anticipated that some SMEs will not know what Big Data Analytics is (Soroka *et al.*, 2017), for this reason an "I do not know" option was added to the Likert items. Although the "I do not know" or "not applicable" options are not recommended for online surveys, as it may increase the number of "I do not know" responses (de Leeuw, Hox and Boevé, 2016), "I do not know" responses may provide valuable information in this case as it may suggest that the lack of Big Data Analytics awareness could be one of the most significant barriers to SMEs adopting the technologies. Similarly, it was considered that failing to provide an "I do not know response" option could have resulted in a lower response rate as, if participants did not understand Big Data or Big Data Analytics, they may have chosen not to complete the questionnaire.

2. **Stage Two: Reviewing Process**

The purpose of the reviewing process is to evaluate the content validity of the questionnaire. Content validity is the extent to which the items of the questionnaire provide adequate coverage of the investigative questions (Saunders, Lewis and Thornhill, 2012). This can be achieved through reviewing the literature available to identify content items and consulting experts in the field (Saunders, Lewis and Thornhill, 2012). To provide further validity, the questionnaire was reviewed by subject matter experts to ensure that the questions represented the barriers to Big Data Analytics adoption by SMEs. Five IT professionals reviewed the questionnaire during this stage to check the content of the questionnaire from a technical perspective. They suggested that additional questions were added to the Likert questions to ensure participants understood the difference between presenting data and choosing a suitable Big Data Analytics solution, as this distinction may not be clear to the participant, depending on their understanding of Big Data Analytics.

3. **Stage Three: Testing**

The final stage of the questionnaire development is testing. The minimum sample size for a student pilot questionnaire is suggested as 10, due to the lack of financial or time resources required for large-scale field trials (Saunders, Lewis and Thornhill, 2012). Isaac and Michael (1995) suggest that small sample sizes are suitable when it is not economically feasible to collect a large sample, and that sample sizes of 10 to 30 are sufficient (Hill, 1998). Therefore, a pilot sample size of 10 was considered sufficient for this study. A pilot study of the questionnaire was distributed to two groups. The first group consisted of five IT professionals who were asked to review the content of the questionnaire. The second group consisted of five non-IT professionals working for SMEs; the aim was to test the usability of the Qualtrics questionnaire system and obtain their feedback. The pilot questionnaire was successful, as all participants completed the questionnaire without encountering any technical issues and the content of the questionnaire was understood, with minor amendments suggested, such as formatting changes, which were subsequently implemented.

3.2 Questionnaire Design

The questionnaire consisted of 42 questions divided into five parts. The first part acted as a coversheet, stating that participants remained anonymous, participation was voluntary and that they were not required to answer every question if they did not wish to. Information was provided regarding the storage and use of the data provided, following the University's ethical guidance. The second part consisted of demographic questions and the third section contained questions relating to data captured and analysed, software applications, IT support and the IT budget. The fourth part consisted of Likert questions relating to the 21 barriers to SMEs adopting Big Data Analytics and the final part provided a thank you message, and the author's contact details. A Likert scale 1 to 5 (Boone and Boone, 2012) was adopted for the where 1 is strongly agree and 5 is strongly disagree.

3.3 Population and Sample of the Study

The research population is the quantity of items, people, objects or organisation which will be the subject of the study (Walliman, 2017). However, depending on the nature of the study, it is rarely feasible to collect data from the entire population, for example due to limitations of time, money or access (Saunders, Lewis and Thornhill, 2012). Therefore, a sample that represents the research population needs to be selected (Saunders, Lewis and Thornhill, 2012). This aim of this study is to develop a strategic framework to assist SMEs in adopting Big Data Analytics, therefore the research population are all SMEs based in the UK and Eire. The study focuses on the UK and Eire for several reasons. Firstly, the definition of an SME can vary between countries; for example, in Australia, a business which employs up to 200 staff is regarded as an SME while in the United States is up to 499 people (Alkhoraif, Rashid and MacLaughlin, 2018). In addition, SMEs in different countries may encounter different challenges, hence the barriers encountered by UK SMEs may not be applicable to SMEs in other countries, raising issues of consistency. Similarly, the trading conditions may vary from country to country, including legislation such as the General Data Protection Regulation in the European Union. As the researcher was located in the UK and their SME contacts are all located in the UK and Eire, it was more feasible to limit the study to SMEs based in the UK and Eire.

3.4 Administration and Distribution of the Questionnaire

The questionnaire was designed and developed using an online surveying platform, Qualtrics. Evans and Mathur (2018) outline a number advantages to utilising an online questionnaire, including physical reach, as participants can be located anywhere; the flexibility offered by survey applications allows questionnaires to be developed relatively easily without the need to write programming or mark-up code; convenience; speed and timeliness; question diversity, as multiple question formats can be utilised; and large sample sizes are easy to obtain. Physical reach was a key advantage of using online questionnaires, as the questionnaire was distributed during the COVID-19 pandemic, and it was not possible to meet face to face with interview subjects during the lockdown period. Online questionnaires are recommended when interviewer interaction with respondents is not required or desirable, therefore interviewer bias and errors are eliminated (Evans and Mathur, 2005). Despite the advantages of online questionnaires, there are also a number of weaknesses including: perception that the emails distributed to participants are junk mail; the surveys may be seen as impersonal; privacy issues; and low response rates (Evans and Mathur, 2018).

The FluidSurveys (2020) sample size calculator was utilised to calculate a sample size using the population of all UK SMEs reported in 2019, which is stated as 4.86 million (Rhodes, 2019), using a confidence level of 95% and a margin of error of 5%. The recommended a sample size generated by this calculation was 385. However, Gorsuch (1983) and Kline (1979) recommended that 100 is a sufficient sample size (MacCallum *et al.*, 1999) and this has been recommended by other authors for statistical techniques, including factor analysis (Williams, Onsman and Brown, 2010). Due to the constraints of time and the COVID-19 pandemic, it was decided that Gorsuch (1983) and Kline (1979) sample size of 100 provided a sufficient sample for the statistical analysis and was more feasible to acquire than 385. The questionnaire was distributed at the height of the COVID-19 pandemic, when many businesses were closed or had non-essential staff furloughed, resulting in businesses not being able to, or not prioritising, completing a questionnaire over other work.

The questionnaire was distributed using ”snowball sampling”, which is a social-chain approach to sampling, whereby participants assist in identifying further participants to grow the sample size (Saunders, Lewis and Thornhill, 2012). This technique is employed for studying hard-to-reach populations and has been utilised in a variety of disciplines (Heckathorn, 2011). The snowball approach has been utilised in other Big Data Analytics studies (Côrte-Real, Oliveira and Ruivo, 2017). However, a potential disadvantage of the snowball approach is that participants are likely to invite other participants who have similar characteristics to themselves, introducing the possibility of bias (Saunders, Lewis and Thornhill, 2012). Kirchherr and Charles (2018) provide a number of recommendations for increasing sample diversity in snowball samples including utilising personal contacts, issuing reminders and ensuring that the initial sample seed is diverse. Therefore, to promote diversity in the sectors represented in the sample, a wide range of SMEs operating in different sectors, including manufacturing, retail, financial services and business services, were initially contacted, in an attempt to maximise diversity at the genesis of the “snowballs”.

Invitation emails containing a link to the online questionnaire were distributed to participants who worked for UK based SMEs from the researcher’s personal contacts in May 2020. The researcher also utilised contacts on their personal LinkedIn profile, who were sent messages containing a link to the online questionnaire. Where appropriate, contacts were asked to invite members of their own network of SME contacts to participate in the study. The British Computer Society was contacted in May 2020 for assistance in distributing the questionnaire. They agreed to help, and distributed the details and link to the questionnaire in an email to all members of the Data Management Special Interest Group. The Chambers of Commerce for the Black Country, Staffordshire, Shropshire and Birmingham were also contacted for assistance disrtibuting the questionnaire to local businesses. Other charities located in the West Midlands which support SMEs were contacted, but they were unable to assist with the distribution of the questionnaire.

To further increase the number of respondents, businesses were randomly selected using Google Maps in the West Midlands, East Midlands, London, Glasgow and Staffordshire areas. This had the additional benefit of increasing the geographical distribution of SMEs covered in the questionnaire. Each business selected was reviewed on the Company’s House website (Companies House, 2020), which displays the accounts for each business submitted at the end of each financial year. Using this report, it could be determined whether the selected companies SMEs, based on the turnover, assets and number of staff documented in the annual accounts report.

4. Result

4.1 Data Analysis

A total of 224 questionnaire responses were received. The results were coded and analysed utilising IBM’s Statistical Package for Social Sciences (SPSS) version 27.

102 fully completed responses from SMEs were received, however an additional 5 responses identified as large companies were excluded from the analyses. Similarly, 46 responses were mostly complete or had not completed the Likert questions, therefore these were utilised for the initial analysis. Appendix 1 shows a flowchart detailing the exclusions made for the various phases of the analysis.

The included responses were then assessed for validity. The questionnaire utilised predominantly closed questions and Likert questions, which restricted the data that could be entered in response to each question. Hence, it was not possible for respondents to give unrelated responses. In addition, comparisons were made between questions, to identify any combinations of responses that would anomalous. For example, if a participant that stated that their business utilises Big Data Analytics, then it should not also have been reasonable for them to state that they did not know what Big Data is, as it would be assumed that to utilise Big Data Analytics, the user would need to understand what this is. No such anomalies were detected.

4.2 Initial Analysis

The majority of respondents were in senior roles at the business (owner: 38.0%, director: 24.8%), with the remainder generally in managerial or IT-based roles. As such, it is likely that most respondents would have been in a position of sufficient knowledge to accurately complete the questionnaire. A breakdown is shown in Figure 1.

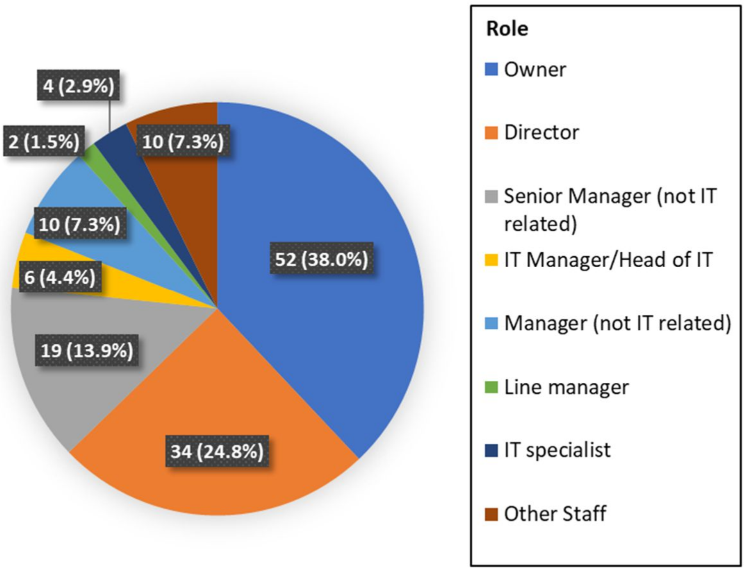


Figure 1 - Number of staff

A diverse range of business sizes were also represented, with 39.4%, 37.2% and 23.4% of respondents from companies employing 1-9 (micro), 10-49 (small) and 50-249 (medium-sized) staff, respectively. A total of 24 sectors were represented in the respondents (with two additional respondents stating "other" sectors), indicating that some degree of diversity had been achieved by the sampling methodology. However, there was a clear preponderance of respondents based in the technology (21.9%) and business services (20.4%) fields, displayed in Figure 2. Appendix 2 provides a full breakdown of the demographics.

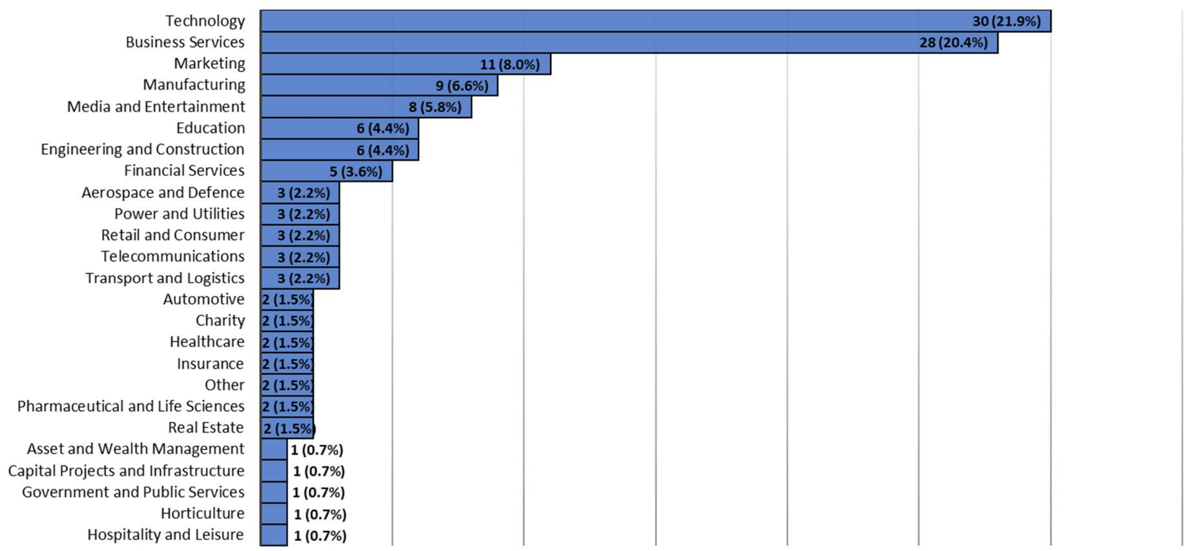


Figure 2 - Participants by sector

As Big Data Analytics can be utilised to analyse a variety of data stored in different formats, therefore it was important to understand the data captured and analysed by SMEs. The majority of participants analyse customer data (72.3%), sales data (62.0%) and website data (51.1%).

Other categories of data were less widely utilised including supplier data (38.0%), competitor data (23.4%), social media data (40.9%), images (21.2%). and sensor data (5.1%). It was also important to identify the software applications currently utilised by SMEs to analyse data. The majority of businesses utilised spreadsheet applications (83.9%), and over half utilised Google Analytics (51.8%). However, Twitter Analytics (18.2%), Microsoft Power BI (15.3%) and Data Warehouses (13.1%) are not widely adopted by the SMEs surveyed. An option for the participants to input other data analytics software was provided in the form of a free text input box. Some of the applications utilised by SMEs to analyse data include: SPSS, Sage, GDS, Tableau, Snowflake, Zoho Analytics, QuickBooks, Bullhorn (a recruitment system), Xero, Survey Monkey, Qualtrics, SNAP, Askia, Crystal Knows, Qlik, Salesforce, Python, R, Mailchimp, Cube19, Snap Surveys and QuenchTec.

As there are a number of Technical barriers to the adoption of Big Data Analytics, there were several questions relating to IT support and the IT budget. A total of 37.2% of participants stated that their business had their own IT department; however, with 30 of the businesses taking part in the survey being based in the technology sector, it would be expected that many of these will have their own IT department. 35.8% of the businesses outsource their IT support and 7.3% combine IT support with another role. 19.7% of businesses do not have any dedicated IT support, as shown in Figure 3. This suggests that the skills required to implement a Big Data Analytics solution may be lacking without dedicated IT staff.

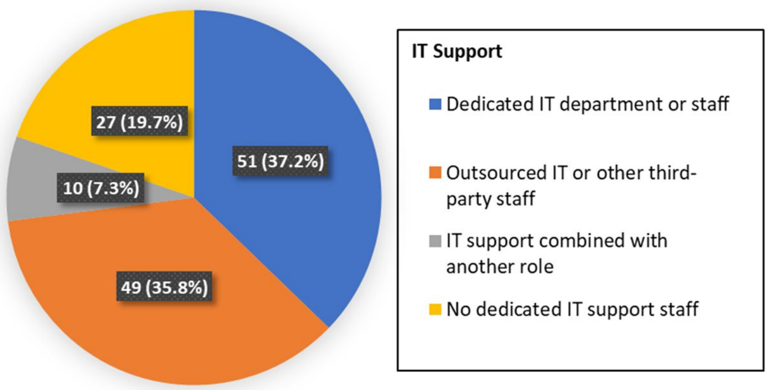


Figure 3 - IT Support

Only 45.3% of the participants stated that their business has an IT budget, 42.3% stated that they do not have a budget and the remaining 12.4% did not know. Sixty-one of the 137 sample completed the follow-on question asking how much their IT budget is. Of the 61, 24.6% stated that their IT budget was more than £50,000, 23.0% have an IT budget between £10,000 and £50,000, 16.4% have an IT budget between £5,000 and £10,000 and 13.1% have an IT budget of less than £5,000. 23.% did not know if their business has a dedicated IT budget. Figure 4 displays a breakdown for the participants who answered the follow-on question regarding the percentage of the IT budget of their business' annual turnover.

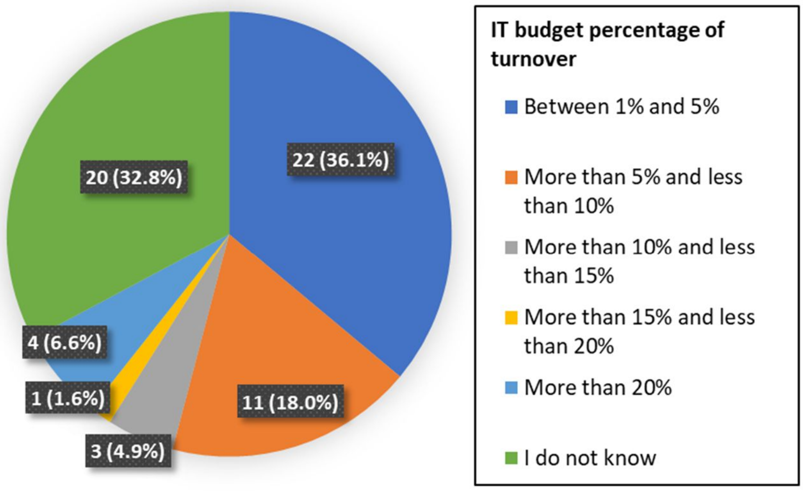


Figure 4 - IT budget percentage of turnover

Participants were asked if they knew what Big Data and Big Data Analytics are. Most participants (63.5%) understood what Big Data is, 14.6% were unsure and 21.9% did not know. Similarly, 61.1% of participants understood what Big Data Analytics was but only 9.6% of businesses were using it. A recent study reported that one in ten SMEs in the European Union using Big Data Analytics (Bianchini and Michalkova, 2019) confirmed similar results. A survey of 15 manufacturing SMEs based in South Wales revealed that only 46.7% were aware of Big Data Analytics, of which 75% had a vision of how they would use it (Soroka *et al.*, 2017), which suggests that the level of knowledge of Big Data Analytics amongst SMEs has increased. However, 28.7% did not know what Big Data Analytics was and 14% were unsure, suggesting that the lack of awareness may be a barrier as shown in Figure 5.

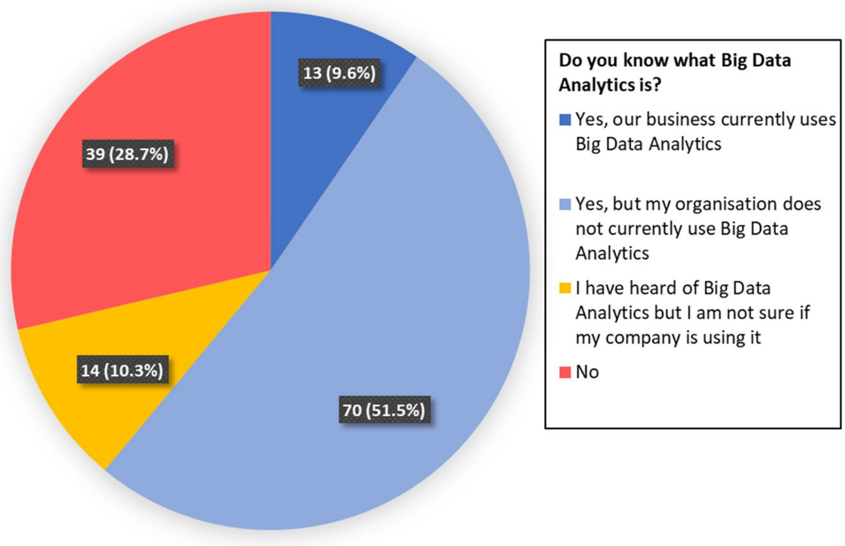


Figure 5 - Awareness of Big Data Analytics

4.3 Associations Between Demographics and Understanding of Big Data Analytics

The "*Do you know what Big Data Analytics is?*" question was correlated against the questions from the first two parts of the questionnaire to identify if there is a relationship between the knowledge of Big Data Analytics and other factors such as the sector the business operates in or the role of the participant. Appendix 3 displays the results for the significance tests. The relationship between the role of the participant and knowledge of Big Data Analytics was insignificant, with a p-value of 0.492 reported, suggesting that there not a significant relationship between these values. However, there was a very significant relationship between the sector the business operates in and the participant's knowledge of Big Data Analytics with a p-value < 0.001 calculated. The majority of participants in the sectors Communications and Technology (87.9%%) have an understanding of Big Data Analytics, with a large proportion in Business Services (85.7%) and Marketing and Media (78.9%). The relationship between number of staff and knowledge of Big Data Analytics was insignificant with a p-value of 0.627.

The relationship between IT support and Big Data Analytics knowledge was very significant with a p-value of p< 0.001. The majority of participants with a dedicated IT department or staff (88.2%) or where IT support was combined with another role (100%) stated that they know what Big Data Analytics is, suggesting that when IT support is provided in-house, there is a greater knowledge computer technology internally than for firms who outsource their IT support. The relationship between knowledge of Big Data Analytics and IT budget was also significant with a p-value of 0.019 reported. However, the relationship between Big Data Analytics and either IT decision making, or IT budget amount was insignificant as both categories scored p-values greater than 0.05.

There was a very significant relationship between whether a firm analyses data, and knowledge of Big Data Analytics, with a p-value of p<0.01 reported. Of the businesses which analyse data, 81.6% of participants report that they know what Big Data Analytics is, suggesting that if a business analyses data then there is a high probability they will be aware of Big Data Analytics. Very strong relationships were reported between Big Data Analytics knowledge and social media (0.020) and images (0.020). Both types of data are classified as Big Data Analytics; this suggests that if businesses analyse these categories of data, they are likely to understand what Big Data Analytics is. However, there was no significant relationship between Big Data Analytics knowledge and analysis of customer data (0.149), sales data (0.186), supplier data (0.722), competitor data (0.599), website data (0.054) and sensor data (0.085).

4.4 Barriers to Big Data Analytics

The final set of questions were the Likert questions representing the barriers to Big Data Analytics adoption. The final section of the questionnaire contained 23 Likert questions, 21 of which represented the 21 barriers identified to SMEs adopting Big Data Analytics. 102 SME participants fully competed the Likert questions. The "I do not know" responses were removed from the Cronbach's Alpha calculation used to test the internal consistency. Table 1 shows the data used to calculate the Cronbach's Alpha scores, where the difference between N and 102 represent the "I do not know answers".

A further Likert question in the final stage of the analysis discussed issues unrelated to the barriers to SMEs adopting Big Data Analytics. The 102 complete responses were utilised for this Likert question analysis. The first question asked the participants how strongly they agreed with the statement: '*My business would benefit from Big Data Analytics'.* Almost half of the participants agreed that their business would benefit from Big Data Analytics with 11.8% strongly agreeing and 33.3% moderately agreeing. There were 18.6% of participants were neutral, 7.8% strongly disagreed and 7.8% moderately disagreed. 20.6% of participants did not know, suggesting that they do not know what Big Data Analytics is or how it would benefit their business.

Pillar	Barrier	N	Mean	Strongly disagree	Moderately disagree	Neutral	Moderately agree	Strongly agree
Business	Financial barriers	84	2.5	17 (20.2%)	28 (33.3%)	23 (27.4%)	13 (15.5%)	3 (3.6%)
Business	Lack of business cases	82	3.4	8 (9.8%)	10 (12.2%)	20 (24.4%)	29 (35.4%)	15 (18.3%)
Environmental	Ethical concerns in data use	86	3.4	10 (11.6%)	10 (11.6%)	15 (17.4%)	34 (39.5%)	17 (19.8%)
Environmental	Inability to assess and address digital risks	79	3.5	4 (5.1%)	12 (15.2%)	15 (19.0%)	35 (44.3%)	13 (16.5%)
Environmental	Regulatory issues	81	3.4	6 (7.4%)	17 (21.0%)	10 (12.3%)	34 (42.0%)	14 (17.3%)
Environmental	The lack of common standards	73	3.2	6 (8.2%)	14 (19.2%)	19 (26.0%)	27 (37.0%)	7 (9.6%)
Human	Lack of in-house data analytics expertise	87	3.2	10 (11.5%)	10 (11.5%)	24 (27.6%)	35 (40.2%)	8 (9.2%)
Human	Shortage of consultancy services	87	3.5	12 (13.8%)	7 (8.0%)	13 (14.9%)	35 (40.2%)	20 (23.0%)
Organisational	Change management	81	3.5	5 (6.2%)	5 (6.2%)	24 (29.6%)	39 (48.1%)	8 (9.9%)
Organisational	Cultural barriers	86	3.4	8 (9.3%)	6 (7.0%)	25 (29.1%)	35 (40.7%)	12 (14.0%)
Organisational	Insufficient volumes of data to be analysed	84	2.5	24 (28.6%)	21 (25.0%)	16 (19.0%)	18 (21.4%)	5 (6.0%)
Organisational	Lack of managerial awareness and skills	87	2.7	15 (17.2%)	22 (25.3%)	24 (27.6%)	22 (25.3%)	4 (4.6%)
Organisational	Lack of top management support	84	3.3	8 (9.5%)	11 (13.1%)	26 (31.0%)	26 (31.0%)	13 (15.5%)
Organisational	Management of technology	80	3.2	11 (13.8%)	15 (18.8%)	16 (20.0%)	25 (31.3%)	13 (16.3%)
Organisational	Talent management	82	2.8	19 (23.2%)	21 (25.6%)	8 (9.8%)	24 (29.3%)	10 (12.2%)
Technological	Complexity of data	75	2.8	17 (22.7%)	15 (20.0%)	13 (17.3%)	23 (30.7%)	7 (9.3%)
Technological	Data scalability	81	2.4	27 (33.3%)	19 (23.5%)	14 (17.3%)	14 (17.3%)	7 (8.6%)
Technological	Data silos	84	3.0	16 (19.0%)	14 (16.7%)	20 (23.8%)	24 (28.6%)	10 (11.9%)
Technological	Infrastructure readiness	78	2.9	19 (24.4%)	10 (12.8%)	17 (21.8%)	20 (25.6%)	12 (15.4%)
Technological	Lack of suitable software	83	2.7	21 (25.3%)	15 (18.1%)	19 (22.9%)	20 (24.1%)	8 (9.6%)
Technological	Poor data quality	82	3.2	11 (13.4%)	13 (15.9%)	20 (24.4%)	25 (30.5%)	13 (15.9%)

Table 1 - Data used for the Cronbach's Alpha test

4.5 Cronbach's Alpha

The reliability of the questionnaire was tested by examining the internal consistency between the questionnaire items. In the study described in this paper, the analysis based on Cronbach's Alpha utilised a complete cases approach in that only those barriers in a pillar where respondents gave an affirmative response (i.e. excluding "I do not know" responses) to all of the barriers within a pillar were included in the analysis of that pillar and as such, the number of the respondents included in the analysis of each pillar ranged from 64 to 85. The Cronbach's Alpha test was performed on each of the five pillars to test the relationship between the barriers. Appendix 4 displays the results of the test. Where there are more than two barriers in a pillar, the Cronbach's Alpha score for each pillar is displayed if one of the barriers is removed.

The internal reliability was highest on the Technological pillar with a Cronbach's Alpha score of 0.91. The removal of barriers in this pillar had minimal affect with a Cronbach's Alpha for items removed ranging from 0.88 to 0.90. Similarly, high Cronbach's Alpha was observed for the Organisational pillar at 0.86 with a Cronbach's Alpha of items removed ranging from 0.82 to 0.85. The Cronbach's Alpha for the Environmental pillar was 0.65. There was some evidence that the Ethical concerns in data use barrier was less consistent with the other barriers in this pillar as removing this improved the alpha to 0.71, however this was deemed to be acceptable. However, there were two pillars which did not meet the acceptable threshold of 0.5, namely Human with 0.46 and Business at 0.37. As such, these were investigated further to assess whether the barriers populating these pillars could be rearranged to improve the Cronbach's Alpha of the pillars as a whole.

The four barriers forming the Business and Human pillars were moved into each of the other pillars to test the Cronbach's Alpha again to determine where they would fit. Moving *Financial barriers* to the Environmental pillar increased the Cronbach's Alpha by 0.014. Moving *Lack of Business Cases* to the Organisational pillar slightly decreased the Cronbach's Alpha score by -0.020 but from a theoretical perspective, business cases are required by the organisation's decision makers to influence their decision to adopt Big Data Analytics. Moving the *Lack of in-house data analytics* expertise barrier to the Organisational pillar increased the Cronbach's Alpha score by 0.024. The *Shortage of Consultancy Services* remained in the Human pillar on its own as moving this to any of the other pillars reduced the Cronbach's Alpha. Table 2 shows the breakdown of the Cronbach's Alpha test following the relocation of the four barriers.

	N Included	Cronbach's Alpha	Alpha (if Item Removed)
Environmental	**63**	**0.67**	
Ethical concerns in data use			0.66
Financial barriers			0.66
Inability to assess and address digital risks			0.63
Regulatory issues			0.58
The lack of common standards			0.55
Human	**N/A***	**N/A***	
Shortage of consultancy services			N/A*
Organisational	**62**	**0.87**	
Change management			0.85
Cultural barriers			0.84
Insufficient volumes of data to be analysed			0.85
Lack of business cases			0.89
Lack of in-house data analytics expertise			0.84
Lack of managerial awareness and skills			0.86
Lack of top management support			0.84
Management of technology			0.86
Talent management			0.86
Technological	**64**	**0.91**	
Complexity of data			0.88
Data scalability			0.89
Data silos			0.90
Infrastructure readiness			0.89
Lack of suitable software			0.89
Poor data quality			0.89

*Table 2 - Cronbach's Alpha test on the four pillars of the revised Big Data Analytics Strategic Framework for SMEs *- not applicable because Cronbach's Alpha requires at least two barriers to be calculable*

4.6 Framework Refinement

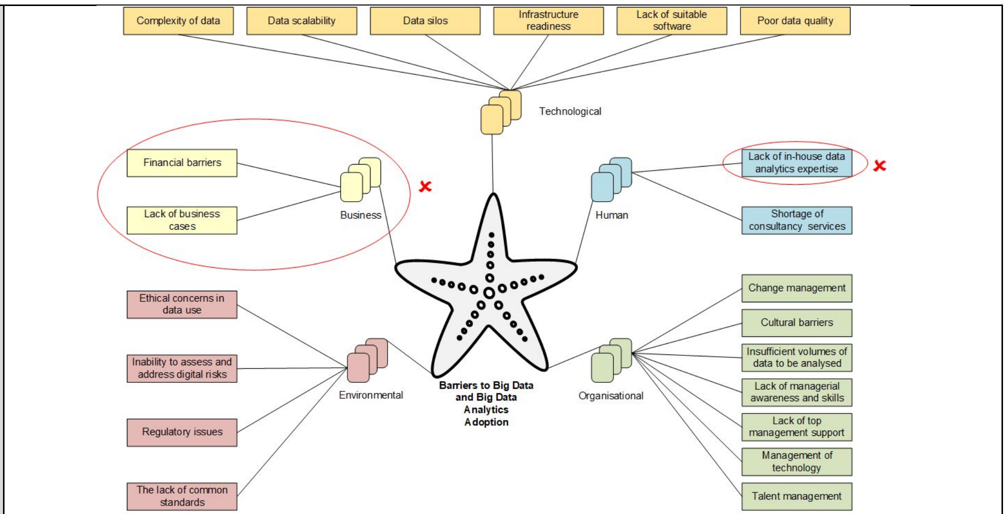
The Big Data Analytics Strategic Framework for SMEs has been refined utilising the feedback received from SMEs participating in the questionnaire. The barriers which had a low Cronbach's Alpha scored have been moved to pillars which increased their score. This suggests that statistically, the barriers are in their correct position alongside barriers which they are related, ensuring that the framework is intuitive. The Cronbach's Alpha test has been widely utilised in studies across a variety of fields, therefore it provides confidence in its results. Using the Cronbach's Alpha test also helps with the validation and evaluation of the strategic framework. Figure 6 shows displays how the strategic framework has been revised based on the questionnaire feedback and statistical analysis, which are:

1. The original version of the strategic framework was developed by undertaking a literature review to identify the barriers to SMEs adopting Big Data Analytics (Willetts,

Atkins and Stanier, 2020a). The barriers were refined utilising a thematic analysis and the barriers were categorised into pillars identified from theoretical frameworks.

2. The second version of the framework was developed from the feedback received from the questionnaire. The Cronbach's Alpha suggested that three of the barriers needed to be relocated, for example *Financial Barriers* moved from the Business pillar to the Environmental pillar. The Business pillar was removed because both of its barriers were moved to other pillars, therefore the revised version of the framework contains four pillars. The Environmental, Organisational and Technological pillars can be considered as internal as all of the barriers contained within these pillars relate to the internal constraints of the organisation. The Human pillar, which contains *Shortage of Consultancy Services* can be considered an External pillar as it refers to factors outside of an organisation's control.

Version 1 of the Big Data Analytics Strategic Framework for SMEs based on the literature (Willetts, Atkins and Stanier, 2020a)



Version 2 of the Big Data Analytics Strategic Framework for SMEs revised utilising the questionnaire feedback received from SMEs

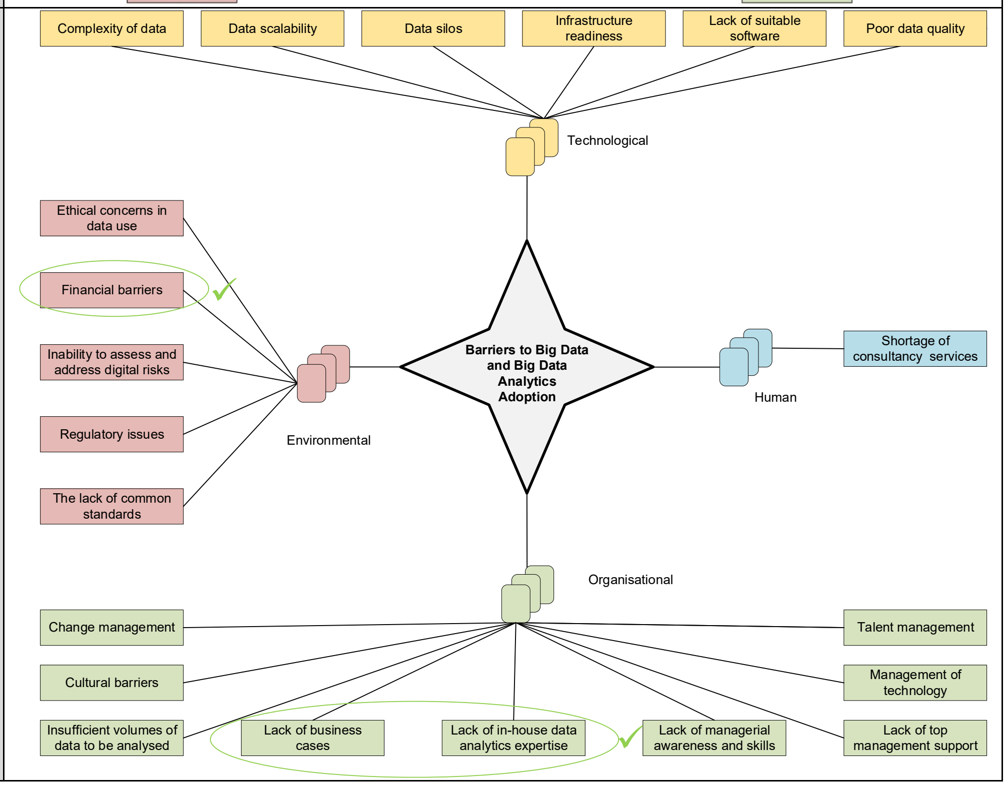


Figure 6 - Framework refinement

4.7 Limitations

One limitation of this study is the sample seed. The initial participants selected were the author's contacts, primarily based in the Technology and Business Services sectors, therefore the questionnaire was not evenly distributed amongst sectors. As the snowball technique was used to distribute the surveys to the participants' contacts, it is likely that their contacts were also located in the same sector in which they operate. Additionally, as the data collection was conducted during the COVID-19 pandemic, this may have resulted in a lower response rate than if the questionnaire had been distributed prior to the pandemic.

5. Conclusion

The qualitative analysis of the questionnaire has demonstrated despite the majority of the participants understand the concept of Big Data and Big Data Analytics, less than 10% of the participants have adopted Big Data Analytics. It has also shown that SMEs in the UK are diverse, with some businesses having dedicated IT staff and utilising software for the analysis of data, suggesting that they may be more receptive to Big Data Analytics. Similarly, only 45.1% of businesses stated that they believe Big Data Analytics would be beneficial for their businesses, which may suggest that the relevance of this technology may depend on the nature of the business or the participants may not be aware of the potential benefits. The 21 barriers to Big Data Analytics have also been verified.

This study has resulted in a revised strategic framework for SMEs adoption of Big Data Analytics utilising the feedback from a statistical analysis. Future work will require qualitative data to be capture from SME practitioners to provide further verification of the barriers identified. The intention of this framework is to help make SMEs aware of the barriers outlined and assist them in overcoming these to provide competitive advantage.

Acknowledgements

The authors would like to thank James Hodson for his statistical advice and assistance.

6. References

Alkhoraif, A., Rashid, H. and MacLaughlin, P. (2018) 'Lean Implementation in small and medium enterprises: Literature review', p. 100089. doi: 10.1016/j.orp.2018.100089.

Bianchini, M. and Michalkova, V. (2019) *OECD SME and Entrepreneurship Papers No. 15 Data Analytics in SMEs: Trends and Policies*. doi: 10.1787/1de6c6a7-en.

Boone, H. N. and Boone, D. A. (2012) 'Analyzing Likert data', *Journal of Extension*, 50(2).

Coleman, S. *et al.* (2016) 'How Can SMEs Benefit from Big Data? Challenges and a Path Forward', *Quality and Reliability Engineering International*. Wiley-Blackwell, 32(6), pp. 2151–2164. doi: 10.1002/qre.2008.

Companies House (2020) *Companies House - GOV.UK*. Available at: https://www.gov.uk/government/organisations/companies-house (Accessed: 1 May 2020).

Côrte-Real, N., Oliveira, T. and Ruivo, P. (2017) 'Assessing business value of Big Data Analytics in European firms', *Journal of Business Research*. Elsevier Inc., 70, pp. 379–390. doi: 10.1016/j.jbusres.2016.08.011.

European Commission (2020) '2019 SBA Fact Sheet: Ireland', *European Commission*, pp. 1–19. Available at: https://ec.europa.eu/docsroom/documents/32581/attachments/21/translations/en/renditions/native.

Evans, J. R. and Mathur, A. (2005) 'The value of online surveys', *Internet Research*. Emerald Group Publishing Limited, 15(2), pp. 195–219. doi: 10.1108/10662240510590360.

Evans, J. R. and Mathur, A. (2018) 'The value of online surveys: a look back and a look ahead', *Internet Research*. Emerald Group Publishing Ltd., pp. 854–887. doi: 10.1108/IntR-03-2018-0089.

FluidSurveys (2020) *Survey Sample Size Calculator - FluidSurveys*. Available at: http://fluidsurveys.com/university/survey-sample-size-calculator/ (Accessed: 18 August 2020).

Heckathorn, D. D. (2011) 'Comment: Snowball Versus Respondent-Driven Sampling', *Sociological Methodology*. [American Sociological Association, Wiley, Sage Publications, Inc.], 41, pp. 355–366. Available at: http://www.jstor.org/stable/41336927.

Hill, R. (1998) 'What sample size is "enough" in internet survey research?', *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century*, 6(3), pp. 1–10. Available at: http://cadcommunity.pbworks.com/f/what sample size.pdf (Accessed: 20 August 2020).

Kirchherr, J. and Charles, K. (2018) 'Enhancing the sample diversity of snowball samples: Recommendations from a research project on anti-dam movements in Southeast Asia', *PloS one*. Public Library of Science, 13(8), pp. e0201710–e0201710. doi: 10.1371/journal.pone.0201710.

de Leeuw, E. D., Hox, J. J. and Boevé, A. (2016) 'Handling Do-Not-Know Answers: Exploring New Approaches in Online and Mixed-Mode Surveys', *Social Science Computer Review*, 34(1), pp. 116–132. doi: 10.1177/0894439315573744.

MacCallum, R. C. *et al.* (1999) 'Sample size in factor analysis', *Psychological Methods*. American Psychological Association, 4(1), pp. 84–99. doi: 10.1037/1082-989X.4.1.84.

Mathew, B. (2016) *How Big Data is reducing costs and improving performance in the upstream industry*. Available at: https://www.worldoil.com/news/2016/12/13/how-big-data-is-reducing-costs-and-improving-performance-in-the-upstream-industry (Accessed: 5 December 2019).

Mikalef, P. *et al.* (2019) 'Big data analytics and firm performance: Findings from a mixed-method approach', *Journal of Business Research*, 98, pp. 261–276. doi: https://doi.org/10.1016/j.jbusres.2019.01.044.

Moore, G. C. and Benbasat, I. (1991) 'Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation', *Information Systems Research*. INFORMS Inst.for Operations Res.and the Management Sciences, 2(3), pp. 192–222. doi: 10.1287/isre.2.3.192.

Pearlson, K. (2001) 'Managing and using information systems: a strategic approach'. New York: John Wiley & Sons, pp. xviii, 278 p. Available at: file://catalog.hathitrust.org/Record/004209999.

Rhodes, C. (2019) 'Briefing Paper 2019: Business Statistics', *UK Statistics: A guide for business users*. doi: 10.4324/9780429397967-7.

Saunders, M., Lewis, P. and Thornhill, A. (2012) *Research methods for business students*. 6th edn. Harlow: Pearson. Available at: https://www-dawsonera-com.focus.lib.kth.se/readonline/9780273750802 (Accessed: 7 May 2020).

Sena, V. *et al.* (2019) 'Big Data and Performance: What Can Management Research Tell us?', *British Journal of Management*. Blackwell Publishing Ltd, 30(2), pp. 219–228. doi: 10.1111/1467-8551.12362.

Soroka, A. *et al.* (2017) 'Big Data Driven Customer Insights for SMEs in Redistributed Manufacturing', in *Procedia CIRP*. Elsevier, pp. 692–697. doi: 10.1016/j.procir.2017.03.319.

Tornatzky, L. G., Fleischer, M. and Chakrabarti, A. K. (1990) 'The processes of technological innovation. Issues in organization and management series', *Lexington Books*. doi: 10.1016/j.cca.2013.05.002.

Walliman, N. (2017) *Research Methods: the Basics*. Florence, United States: Taylor & Francis Group. Available at: http://ebookcentral.proquest.com/lib/staffordshire/detail.action?docID=5015633.

Walsh, J. (2017) *How a small company used big data to increase its sales an Australian pool company mined customer insights on a budget and turned its tides*. Available at: www.imd.org (Accessed: 2 May 2020).

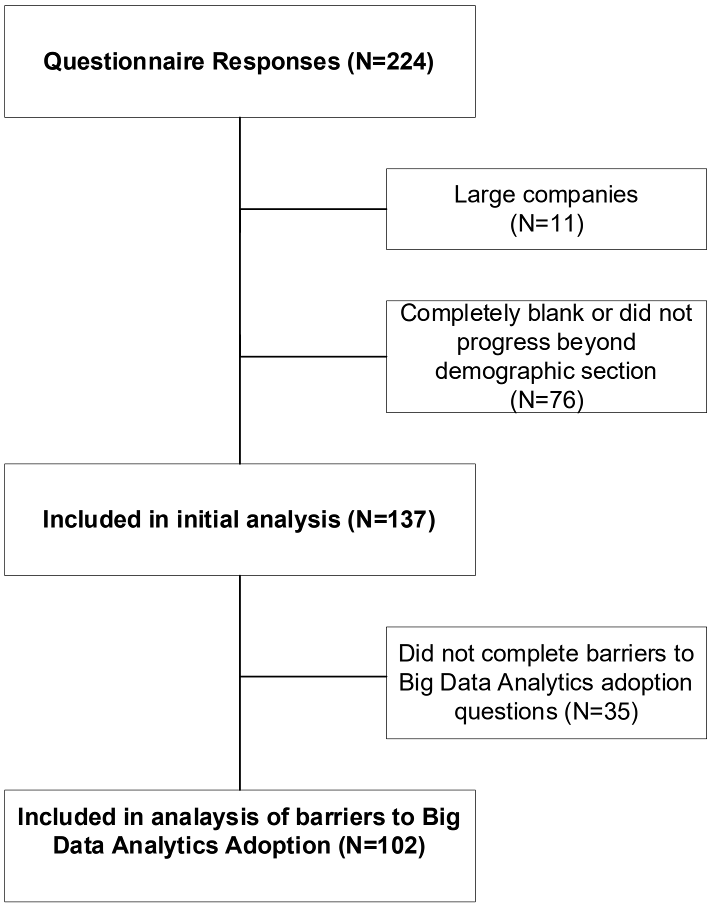
Willetts, M., Atkins, A. S. and Stanier, C. (2020a) 'A Strategic Big Data Analytics Framework to Provide Opportunities for SMEs', in *14th International Technology, Education and Development Conference*, pp. 3033–3042. doi: 10.21125/inted.2020.0893.

Willetts, M., Atkins, A. S. and Stanier, C. (2020b) 'Barriers to SMEs Adoption of Big Data Analytics for Competitive Advantage', in *The Fourth International Conference on Intelligent Computing in Data Sciences ICDS2020*. Fez, Morocco.

Williams, B., Onsman, A. and Brown, T. (2010) 'Exploratory factor analysis: A five-step guide for novices', *Journal of Emergency Primary Health Care*, 8(3), pp. 1–13. doi: 10.33151/ajp.8.3.93.

Yusof, M. M. *et al.* (2008) 'An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit)', *International Journal of Medical Informatics*. Elsevier, 77(6), pp. 386–398. doi: 10.1016/J.IJMEDINF.2007.08.011.

Appendix 1 – Flowchart of the number of responses included in the analysis



Appendix 2 – Demographics

	Total Responses	N (%)
Role	137	
Owner		52 (38.0%)
Director		34 (24.8%)
Senior Manager (not IT related)		19 (13.9%)
IT Manager/Head of IT		6 (4.4%)
Manager (not IT related)		10 (7.3%)
Line manager		2 (1.5%)
IT specialist		4 (2.9%)
Other Staff		10 (7.3%)
Sector	137	
Aerospace and Defence		3 (2.2%)
Asset and Wealth Management		1 (0.7%)
Automotive		2 (1.5%)
Business Services		28 (20.4%)
Capital Projects and Infrastructure		1 (0.7%)
Charity		2 (1.5%)
Education		6 (4.4%)
Engineering and Construction		6 (4.4%)
Financial Services		5 (3.6%)
Government and Public Services		1 (0.7%)
Healthcare		2 (1.5%)
Horticulture		1 (0.7%)
Hospitality and Leisure		1 (0.7%)
Insurance		2 (1.5%)
Manufacturing		9 (6.6%)
Marketing		11 (8.0%)
Media and Entertainment		8 (5.8%)
Pharmaceutical and Life Sciences		2 (1.5%)
Power and Utilities		3 (2.2%)
Real Estate		2 (1.5%)
Retail and Consumer		3 (2.2%)
Technology		30 (21.9%)
Telecommunications		3 (2.2%)
Transport and Logistics		3 (2.2%)
Other		2 (1.5%)
Number of staff	137	
1 to 9		54 (39.4%)
10 to 49		51 (37.2%)
50 to 249		32 (23.4%)

Appendix 3 – Significance Testing

Chi-squared test comparing the relationship between knowledge of Big Data Analytics, demographics and IT* Where P < 0.050 text is bold

	N	Know Big Data Analytics Yes	No	p-Value
Role	136			0.492
Owner/Director		62 (72.1%)	24 (27.9%)	
Manager (Not IT-Related)		19 (63.3%)	11 (36.7%)	
IT Manager/Head of IT		6 (100.0%)	0 (0.0%)	
IT Specialist		3 (75.0%)	1 (25.0%)	
Other		7 (70.0%)	3 (30.0%)	
Sector	136			**p < 0.001**
Communications and Technology		29 (87.9%)	4 (12.1%)	
Business Services		24 (85.7%)	4 (14.3%)	
Financial		6 (60.0%)	4 (40.0%)	
Construction and Manufacturing		8 (36.4%)	14 (63.6%)	
Marketing and Media		15 (78.9%)	4 (21.1%)	
Others		15 (62.5%)	9 (37.5%)	
Number of staff	136			0.627
1 to 9		41 (75.9%)	13 (24.1%)	
10 to 49		34 (68.0%)	16 (32.0%)	
50 to 249		22 (68.8%)	10 (31.3%)	
How is IT supported	136			**p < 0.001**
Dedicated IT department or staff		45 (88.2%)	6 (11.8%)	
Outsourced IT or other third-party staff		25 (52.1%)	23 (47.9%)	
IT support combined with another role		10 (100.0%)	0 (0.0%)	
No dedicated IT support staff		17 (63.0%)	10 (37.0%)	
IT decision makers	134			0.253
The owner		40 (71.4%)	16 (28.6%)	
Senior management		38 (67.9%)	18 (32.1%)	
IT Manager/Head of IT		19 (86.4%)	3 (13.6%)	
IT budget?	119			**0.019**
Yes		51 (82.3%)	11 (17.7%)	
No		36 (63.2%)	21 (36.8%)	
IT budget amount	47			0.801
Less than £5,000		7 (87.5%)	1 (12.5%)	
£5,000 to £10,000		8 (80.0%)	2 (20.0%)	
Between £10,000 to £50,000		12 (85.7%)	2 (14.3%)	
More than £50,000		14 (93.3%)	1 (6.7%)	

Chi-squared test comparing the relationship between knowledge of Big Data Analytics, software and data analysed * Where P < 0.050 text is bold

	N	Know Big Data Analytics Yes	 No	p-Value
Analyse data?	136			**p < 0.001**
No		17 (44.7%)	21 (55.3%)	
Yes		80 (81.6%)	18 (18.4%)	
Data warehouse	136			**0.020**
No		80 (67.8%)	38 (32.2%)	
Yes		17 (94.4%)	1 (5.6%)	
Spreadsheet applications	136			0.608
No		14 (66.7%)	7 (33.3%)	
Yes		83 (72.2%)	32 (27.8%)	
Google Analytics	136			0.202
No		43 (66.2%)	22 (33.8%)	
Yes		54 (76.1%)	17 (23.9%)	
Twitter Analytics	136			0.567
No		78 (70.3%)	33 (29.7%)	
Yes		19 (76.0%)	6 (24.0%)	
Microsoft Power BI	136			0.113
No		79 (68.7%)	36 (31.3%)	
Yes		18 (85.7%)	3 (14.3%)	
Customer data	136			0.149
No		23 (62.2%)	14 (37.8%)	
Yes		74 (74.7%)	25 (25.3%)	
Sales data	136			0.186
No		33 (64.7%)	18 (35.3%)	
Yes		64 (75.3%)	21 (24.7%)	
Supplier data	136			0.722
No		59 (70.2%)	25 (29.8%)	
Yes		38 (73.1%)	14 (26.9%)	
Competitor data	136			0.599
No		73 (70.2%)	31 (29.8%)	
No		24 (75.0%)	8 (25.0%)	
Social media	136			**0.020**
No		51 (63.8%)	29 (36.3%)	
Yes		46 (82.1%)	10 (17.9%)	
Website data	136			0.054
No		42 (63.6%)	24 (36.4%)	
Yes		55 (78.6%)	15 (21.4%)	
Sensor data	136			0.085
No		90 (69.8%)	39 (30.2%)	
Yes		7 (100.0%)	0 (0.0%)	
Images	136			**0.020**
No		83 (77.6%)	24 (22.4%)	
Yes		14 (48.3%)	15 (51.7%)	

Appendix 4 - Cronbach's Alpha test on the five pillars of the Big Data Analytics Strategic Framework for SMEs

	N Included	Cronbach's Alpha	Alpha (if Item Removed)
Business	**76**	**0.37**	
Financial barriers			N/A*
Lack of business cases			N/A*
Environmental	**65**	**0.65**	
Ethical concerns in data use			0.71
Inability to assess and address digital risks			0.56
Regulatory issues			0.49
The lack of common standards			0.56
Human	**85**	**0.46**	
Lack of in-house data analytics expertise			N/A*
Shortage of consultancy services			N/A*
Organisational	**67**	**0.86**	
Change management			0.84
Cultural barriers			0.84
Insufficient volumes of data to be analysed			0.84
Lack of managerial awareness and skills			0.85
Lack of top management support			0.82
Management of technology			0.85
Talent management			0.85
Technological	**64**	**0.91**	
Complexity of data			0.88
Data scalability			0.89
Data silos			0.90
Infrastructure readiness			0.89
Lack of suitable software			0.89
Poor data quality			0.89

*- not applicable because Cronbach's Alpha requires at least two barriers to be calculable