

Article

The feasibility of measuring the sharing economy: November 2017 progress update

A proposed working definition of the UK sharing economy and a framework for identifying its businesses, new information on characteristics of sharing economy businesses and individuals, and experimental data science research identifying whether a business is in the sharing economy or not.

Contact:
Pauline Beck
sharing.economy@ons.gsi.gov.
uk
+44 (0)1633 455856

Release date:
9 November 2017

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [Introduction](#)
3. [Defining the sharing economy](#)
4. [Background to this publication and literature review](#)
5. [Potential future sources of information](#)
6. [Findings from current data sources](#)
7. [ONS data science project](#)
8. [Acknowledgments](#)
9. [Annex A: Methodology related to the definitional work](#)
10. [Annex B: Challenges in measuring and defining the sharing economy](#)
11. [Annex C: Statistical analysis](#)
12. [Annex D: K-means clustering](#)

1 . Main points

- Office for National Statistics (ONS) has developed a conceptual framework to support the collection and dissemination of statistics on sharing economy activities.
- New estimates on individuals who use an intermediary app or website to book accommodation or transport from another individual, provide an insight into sharing economy activity.
- New descriptive statistics on business data shed a light on characteristics of sharing-economy businesses, such as employment costs, advertising and marketing, such as employment costs, advertising and marketing, and turnover
- Data science techniques investigate using a K-means clustering analysis to determine whether businesses can be classified into the sharing economy based on their responses to ONS business surveys.

2 . Introduction

The sharing economy is not a new concept. Yet its scope and the pace at which it is growing make it a striking new phenomenon. Sharing-economy platforms account for an increasing proportion of the UK economy and consequently it is important to develop the appropriate statistics for use by stakeholders, policy-makers, researchers, organisations and the media. Yet while traditional ways of producing and consuming goods and services are changing with new technologies and social media use, there are many challenges associated with measuring the sharing economy. In particular, there is no internationally-agreed definition and no official conceptual framework.

This article proposes a working definition of the sharing economy, as well as a framework for identifying businesses that form part of it. It presents new information on characteristics of sharing-economy businesses and individuals, and descriptive statistics. This article also highlights experimental data science research, which aims to identify the characteristics that can indicate whether a business is in the sharing economy or not. We published articles in April 2016 and October 2016 on the [feasibility of measuring the sharing economy](#) and a [progress update](#). This article updates this workstream and provides an account of the work that we have carried out so far.

3 . Defining the sharing economy

There is no widely-accepted definition of the sharing economy. However, the main characteristics of such activities are:

- operating through an online platform through a website or an app
- enabling consumer-to-consumer transactions
- temporarily providing access to a good or service with no transfer of ownership; this excludes the second-hand economy in which goods are resold
- utilising an under-used asset

These characteristics can blur the lines between personal and commercial: for example, peer-to-peer activities such as car-sharing used to be regarded as “personal”, yet using an online platform to organise this activity can commercialise it. The platforms can also blur the difference between full-time and casual labour, as well as between employee and self-employed; indeed some traditional full-time employment is being replaced by platform work that permits different levels of time commitment.

We have developed a conceptual framework to support the collection and dissemination of statistics on sharing-economy activities. We define the sharing economy as, “the sharing of under-used assets through completing peer-to-peer transactions that are only viable through digital intermediation, allowing parties to benefit from usage outside of the primary use of that asset.” However, it is important to highlight that even though the criteria we use are useful for our purposes, the creation of a definition will be subjective. It is also worth noting that this definition is likely to evolve alongside our understanding of how to measure sharing-economy activities.

There are also some concepts in this definition¹ that require further consideration and development. These include:

- “peer-to-peer” – when to differentiate between individuals and commercial activity; there are currently ambiguities, such as when a person considers themselves to be a business, in the case of self-employment; currently, surveys use concepts that are left to respondents to self-define, especially where respondents are employees of a business that they own
- “viability” – identifying which businesses could operate without a digital platform
- “primary use” – there will be differences in opinion on the primary use of an asset, particularly with assets that are not used regularly by the owner

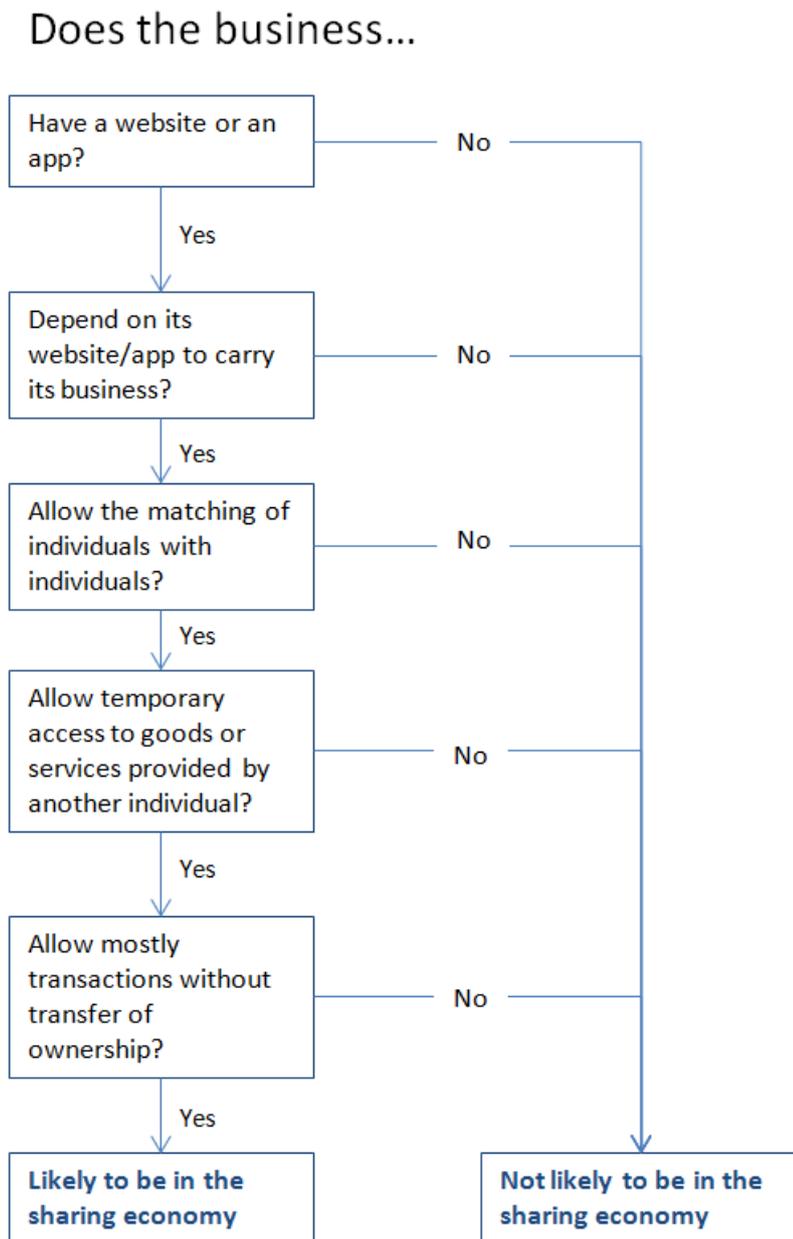
Using a decision tree to identify sharing-economy businesses

Our previous update highlighted ways to categorise different types of sharing-economy activities. Three broad categories were identified, as follows:

- property rental and access
- peer-to-peer services
- collaborative finance

In addition to our definition, we also used a decision tree to help identify sharing-economy businesses based on a set framework. This decision tree (Figure 1) helps to determine whether a given business is likely to be part of the sharing economy.

Figure 1: Decision tree for identifying sharing economy businesses



Notes for: Defining the sharing economy

1. The methodology behind this definition is explained in the technical annex A.

4 . Background to this publication and literature review

Office for National Statistics (ONS) pledged to assess the feasibility of developing statistics on the sharing economy, which was prompted by a range of publications on the sharing economy. Recent reports on [The sharing economy in the UK](#) (PDF, 1.0MB) (Coyle, 2016) and Chapter 3 of The [Independent review of UK economic statistics](#) (PDF, 5.36MB) in March 2016 (Bean, 2016) called for the development of statistics on the sharing economy. These reports noted that activity associated with the sharing economy is neglected to some extent by current official statistics. As an answer to these reports and due to the increasing interest from policy-makers, industry experts and trade bodies, the sharing economy has become an important aspect of our work programme.

Wider research highlights the potential for economic growth in sharing activities

In order to understand wider research and concepts around the sharing economy, we carried out a literature review on sharing-economy publications.

PriceWaterhouseCoopers (PwC) forecast (2014) that [the sharing economy would be worth £9 billion to the UK economy by 2025](#), estimating that it was worth £0.5 billion in 2014. A more recent publication from PwC in April 2016 estimated [the size of the “collaborative economy” in Europe](#) (PDF, 1.0MB) (the term here is used as a synonym of “sharing economy”). It identified over 275 “collaborative platforms”. These platforms facilitated transactions of an estimated €28 billion in 2015. While the European Commission’s exploratory study (2017) estimated that [there are 485 platforms in Europe](#), only 4% are very large with over 100,000 unique daily visitors.

Academic and media writers are also increasing their coverage of the sharing economy

There is also an increasing interest in the sharing economy from academics and the media. Much of the media coverage around prominent sharing-economy businesses has focused on their impact on traditional businesses and economies, particularly in relation to employment. There has been a lot of interest in the legal status of people working in the sharing economy and their working conditions, including the [Taylor Report](#) (2017). The impact on employment patterns and on incumbent businesses needs to be considered alongside the potential. The 2016 [Coyle report](#) (PDF, 1.0MB) argues that the sharing economy can match the demand and supply of goods and services quickly and cheaply.

Several studies are concerned with consumer thought and behaviour, and conceptual arguments about the character of sharing. Researchers have also studied [the reasons motivating participation in the sharing economy](#) (PDF, 130KB) (Hamari et al. 2016). This study has shown that different factors are influential in different sectors. For example, sustainability is important for users of businesses that encourage ecological consumption (lift-sharing websites, for example), while attitude and reputation are important for transportation and accommodation platforms. While establishing trust is inherent to the success of the platform, the report from the European Commission (2017) summarised research indicating that [“peer review and rating systems” were not entirely reliable](#) as fewer than half of the consumers wrote reviews. Moreover very few consumers actually left low ratings or bad reviews, suggesting a positive bias.

The literature also includes a terminological debate evident around the definition of “sharing economy”. There are various terms surrounding this debate that are used interchangeably and sometimes wrongly. The most popular terms are “sharing economy”, “collaborative economy” and “gig economy”, however, they are often confused with other terms such as “digital economy”, or “collaborative consumption”. There are discussions surrounding which is the most appropriate term and whether or not a given business fits within this category. We have opted for the term “sharing economy” as it is perhaps the term which is the most popular and representative of the activities we are looking to measure.

5 . Potential future sources of information

Questions in the Opinions and Lifestyle Survey collect information on the sharing economy

Information on activities related to the sharing economy were collected as part of the [Internet access](#) module in this survey. Two new questions were introduced in 2017 to fulfill a Eurostat requirement, relating to the use of intermediary websites or apps to arrange accommodation and/or transport. The questions asked are in Table 1.

Table 1: Sharing economy questions added to the Internet Access module

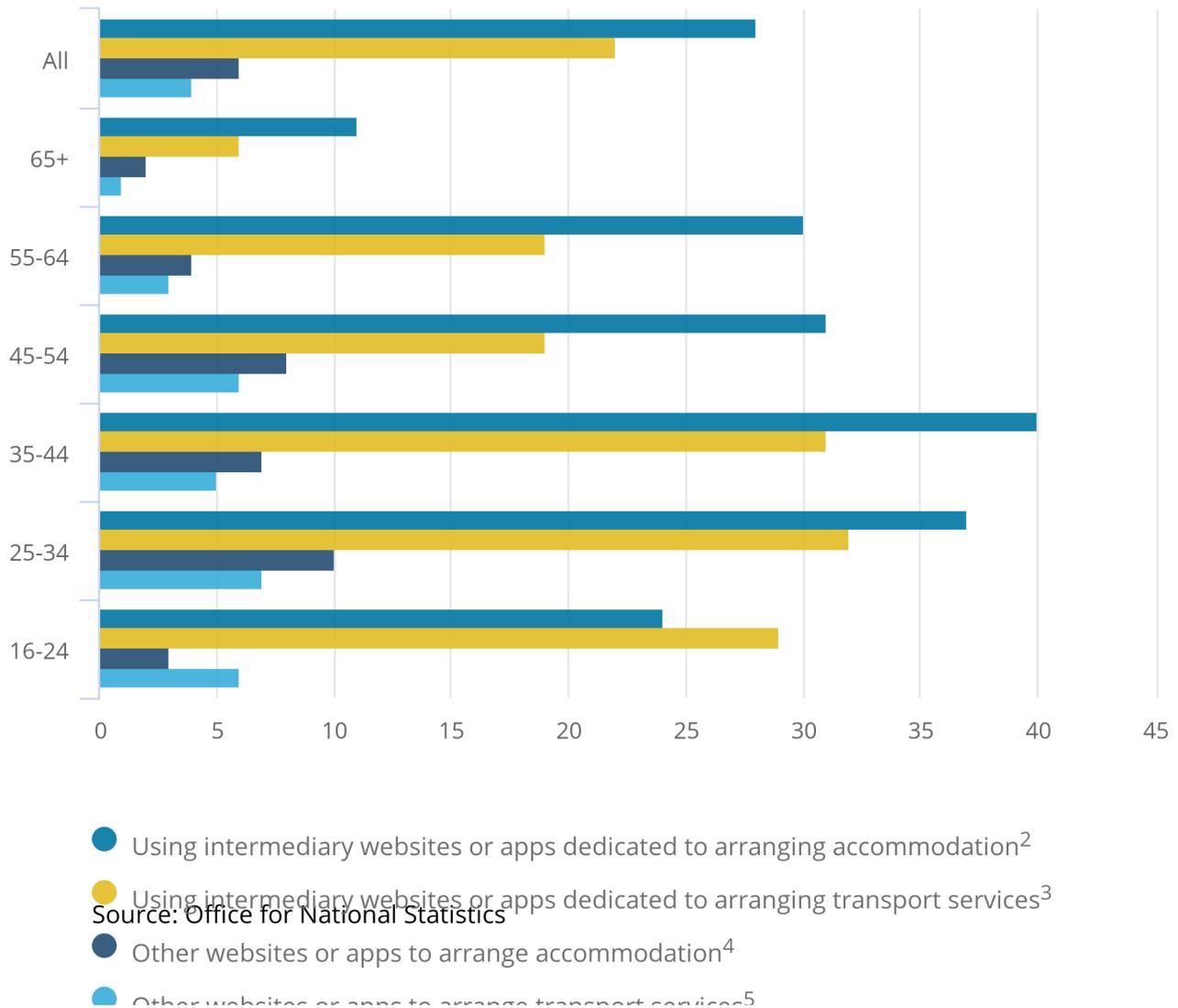
Transport	<p>In the last 12 months have you used any website or 'app' to arrange transport services (e.g. car travel) from another private individual?</p> <p>(1) Yes, intermediary websites or 'apps' dedicated to arranging transport services (such as Uber, Lyft, BlaBlaCar, Liftshare etc)</p> <p>(2) Yes, other websites or 'apps' (including Facebook, Twitter etc)</p> <p>(3) No, I have not.</p>
Accommodation	<p>In the last 12 months have you used any website or 'app' to arrange accommodation (room, apartment, house, holiday cottage, etc.), from another private individual?</p> <p>(1) Yes, intermediary websites or 'apps' dedicated to arranging accommodation (such as Airbnb, HomeAway, Onefinestay, SpareRoom etc)</p> <p>(2) Yes, other websites or 'apps' (including Facebook, Twitter etc)</p> <p>(3) No, I have not</p>

Source: Office for National Statistics

The main findings from these sharing-economy data were that 28% of adults used intermediary websites or apps to arrange accommodation, in a year (Figure 2). Likewise for transport, 25% of men had used an intermediary website or app, whereas 18% of women arranged transport through similar means. These questions will be asked again next year, allowing analysis of changes regarding the use of these services and to provide an indication on how the sharing economy is changing.

Figure 2. Use of the internet to arrange accommodation or transport from another individual, by age group¹, 2017, Great Britain

Figure 2. Use of the internet to arrange accommodation or transport from another individual, by age group¹, 2017, Great Britain



Source: Office for National Statistics

Notes:

1. Base: Adults (aged 16 and over) in Great Britain.
2. Such as Airbnb, HomeAway, Onefinestay, SpareRoom etc.
3. Such as Uber, Lyft, BlaBlaCar, Liftshare etc.
4. Including Facebook, Twitter etc.
5. Including Facebook, Twitter etc.

Sharing economy questions tested in the Labour Force Survey need further development

We are exploring different avenues for data collection for sharing-economy activities. We are in conversations with [Labour Force Survey](#) (LFS) stakeholders to consider adding new questions to the questionnaire. Currently the LFS provides information, such as employment status, hours worked, earnings, type of work, or whether one is looking for work. However, the survey does not collect any information on the use of digital platforms, nor could it identify which ones are used.

New potential questions were tested in the annual LFS pilot. These new questions were related to whether respondents have used a digital platform to find work and whether it was their main source of earnings. The feedback as part of this pilot was mixed, which will result in the need to change the wording of the questions, however, it was thought the questions fitted well within the questionnaire. We will be collaborating with one of the LFS's stakeholders, the [Department for Business, Energy and Industrial Strategy](#) (BEIS) to carry out further work on these questions.

Redesign of the Living Costs and Food Survey provides a future opportunity to gather sharing-economy information

We are also exploring the use of the [Living Costs and Food](#) (LCF) Survey, which covers income and expenditure. This survey collects essential information on household spending patterns, which feeds into the Consumer Prices Index (CPI) and gross domestic product (GDP) figures; it also collects detailed information of food consumption. The survey consists of a questionnaire and a diary. The questionnaire collects information on income and main expenditures, while the diary collects all expenditures made over a period of two weeks.

In early 2017, the LCF underwent a [National Statistics Quality Review](#) (NSQR): one of its main recommendations was to ensure that the questionnaire design keeps pace with ongoing changes in consumer spending and behaviours. As a result, we have been working closely with the LCF team to review their questionnaire to see where questions or categories related to the sharing economy could be added. We are in the process of collecting evidence from main users to support the addition of these type of questions or new categories to the survey.

A time-use study is another option

Another avenue that we have been exploring is an adapted [time-use study](#). Time-use surveys are used to collect information on how much time individuals spend undertaking different activities and the [latest UK study](#) was run between April 2014 and December 2015.

In recent months, we have worked to explore a new form of time-use survey specially adapted to capture activities performed for sharing-economy purposes and other aspects of the modern digital economy. Time-use data are normally collected through a diary over a period of days. Each respondent records what they are doing at specified intervals throughout the day. For sharing-economy activities, a time-use survey might identify how much time respondents take preparing their spare room, or how long they spend using sharing-economy websites.

The information about time spent on the activities could be used as the basis for estimates of the value of the sharing economy. The viability of this approach will be explored over the coming months and if successful, results are expected towards the end of 2018. If you would like to feed your views into how these time-use data might inform on measuring the sharing economy, you are encouraged to contact us at sharing_economy@ons.gov.uk.

6 . Findings from current data sources

New information on sharing economy businesses available in two current ONS surveys

Office for National Statistics (ONS) sampled sharing-economy businesses as part of the [E-commerce Survey](#) and [Annual Business Survey](#) (ABS) in 2016 to produce some descriptive statistics. The responses from these businesses have been compared with responses from non-sharing economy businesses in the sample, matched for similar levels of employment and industrial classification. E-commerce Survey information from 81 sharing-economy and 152 non-sharing-economy businesses, and ABS information from 45 sharing-economy and 6,451 non-sharing-economy businesses were used in the analysis. Twenty-three questions across the surveys were deemed relevant for this analysis and have been listed in Annex C, Table 6.

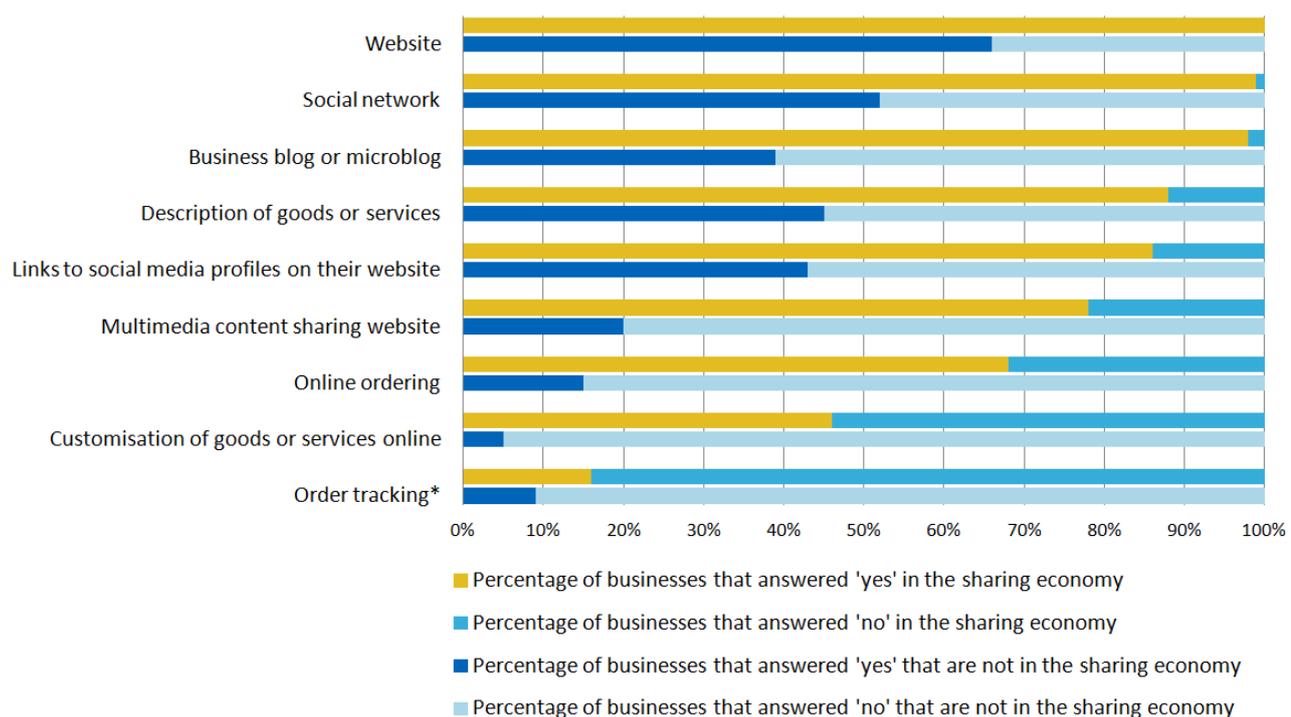
Sharing-economy businesses were identified using a number of sources, including articles written on [the size and presence of the collaborative economy in Europe](#) (PDF, 1.0MB) by PricewaterhouseCoopers (April 2016) and [digital matching firms: a new definition](#) (PDF, 632KB) by the US Department of Commerce (June 2016), both of which have provided lists of sharing-economy businesses across the world. Members of UK trade body [Sharing Economy UK](#) were also included where appropriate.

Some members of Sharing Economy UK are said to be “interested parties”, which does not always agree with our definition of the sharing economy. It is important to note that this is a sample of sharing-economy businesses and not a comprehensive list, as we are not aware of all the businesses in the sharing economy. Businesses were tested against the decision tree presented in Figure 1 to confirm the designation of sharing-economy businesses. For the businesses in the sharing economy that did not respond to the surveys, extensive research was undertaken to predict how they might respond to the E-commerce Survey. While this is experimental research, the analysis of these data is useful to give an overview of the characteristics of a typical sharing-economy business.

The E-commerce Survey reveals some significant differences between sharing and non-sharing economy businesses

The E-commerce Survey results take the form of Yes or No answers. In the majority of questions there was a significant difference between the two groups; Figure 3 gives a breakdown of these results ¹.

Figure 3. Comparison of e-commerce survey variables for sharing economy and non-sharing economy businesses.



Businesses in the sharing economy are ultra-connected

The sharing economy is relatively new and therefore [a large proportion of the businesses have been established during the last decade](#) (ONS, 2016). The sharing economy is heavily dependent on a range of technology that has recently become readily available to consumers, such as smartphones, mobile applications, global positioning systems and a multitude of other functionalities. The e-commerce Survey asks a range of questions about the technical capability of the business, which allows a sharing-economy business's dependence on technology to be investigated.

All sharing-economy businesses have stated that they have a website, which is significantly larger than their sampled non-sharing-economy counterparts. This is not surprising as sharing-economy businesses have a website or app to process the peer-to-peer transactions. Sharing-economy businesses are also more visible on social media, with almost 100% having a social network, compared with 52% for non-sharing-economy businesses. Included within "social media"² are websites such as Facebook, LinkedIn and Yammer.

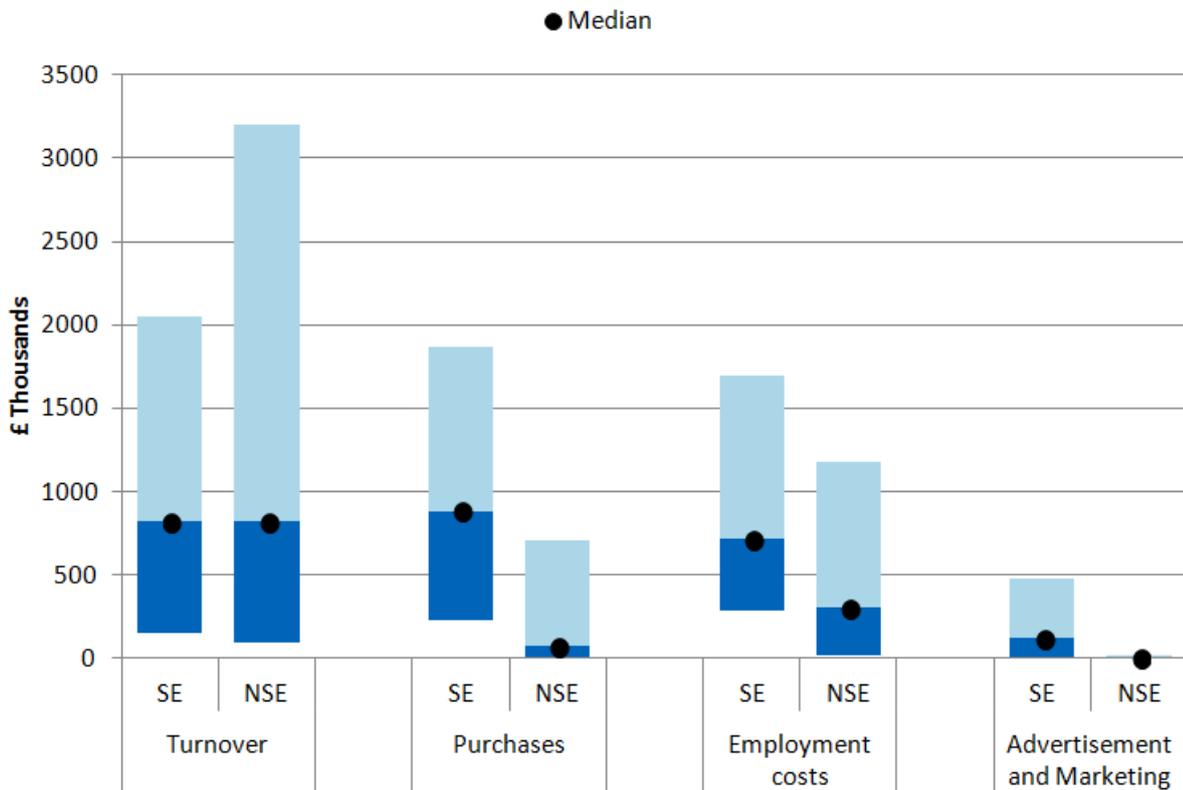
This trend continues for business blogs, microblogs and multimedia content-sharing websites. A business blog is a written blog that can usually be found on the business's website; on the other hand a microblog is one where a website such as Twitter is used by the business. In addition, YouTube, Flickr, Picasa and Instagram are all websites included under multimedia content-sharing websites. Furthermore, sharing-economy businesses are more likely to have links to social media profiles on their websites.

Sharing economy comparisons using Annual Business Survey results are more mixed

In the Annual Business Survey, the distribution of businesses is strongly weighted towards smaller businesses, with only a few larger businesses. This means many of the standard hypotheses tests were not feasible. Instead, we present the inter-quartile ranges to compare each variable across both groups of businesses; Figure 4 summarises these results.

Figure 4. Comparison of Annual Business Survey variables for sharing economy and non-sharing economy businesses.

Median and Inter-quartile ranges for the Annual Business Survey



Most businesses in the sharing economy are start-ups

Technological innovation has enabled the efficient and cost-effective matching of sharing-economy participants. It is clear from analysing the median and inter-quartile ranges of the Annual Business Survey that these businesses have invested more in advertising and marketing, possibly to raise the awareness of the service that they provide and establish a market share, given that many of them are relatively new. Participants in an engagement workshop with Sharing Economy UK suggested that this was the case.

The median value spent on advertising and marketing costs for sharing-economy businesses was £124,000 in 2016, compared with £1,000 for the sampled non-sharing-economy businesses. The inter-quartile range is also considerably higher, meaning the spread of money spent on advertising and marketing is much wider in the sharing economy. This could be because of the difference in maturity of the sharing-economy businesses sampled – newer businesses need to spend a higher proportion of their turnover on advertising to become known. However, the resources available to invest in advertising are also likely to be larger for businesses that are established and profitable, such as Airbnb.

Sharing-economy businesses also spend more on their total purchases; the median value for these businesses was £886,000 in 2016, compared with £72,000 for the sampled businesses not facilitating sharing-economy activity. As with expenditure on advertising, the spread of total purchases for companies in the sharing economy was greater.

Again, contextual evidence provided during a workshop with Sharing Economy UK suggested that the larger expenditure could be because a number of these businesses were in their infancy. Advertisement and marketing costs, energy and water costs, subcontractor costs and other services costs, are included within total purchases. Many sharing-economy businesses have relatively high subcontractor costs. This could also be explained by the fact that these businesses are in their infancy, or that they are more likely to hire IT/Tech contractors. It is difficult to establish whether this is more generally the case as subcontractor costs have not been investigated in detail in this statistical analysis.

Businesses in the sharing economy have higher employment costs

Sharing-economy businesses have higher employment costs. This is probably because the majority of employees are likely to be more technically skilled and therefore command higher wages, compared with many other companies. "Employment costs" here relates to employees who work for the platform, rather than individuals who find work through the platform.

The median employment cost was £717,000 for businesses in the sharing economy in 2016, compared with £306,000 for non-sharing-economy businesses. That said, the contribution to pension funds is lower in sharing-economy businesses.

Notes for: Findings from current data sources

1. Chi-squared tests confirmed the differences were significant; the results from the chi-squared tests can be found in Annex C, Table 2
2. The e-commerce survey forms give instructions on what to include under social media (Facebook, LinkedIn, Xing, Viadeo, Yammer), as well as business blogs and microblogs (Twitter, Present.ly) and multimedia content sharing websites (YouTube, Flickr, Picasa).

7 . ONS data science project

Unsupervised versus supervised learning

This section explores whether variables from the Annual Business Survey and E-commerce Survey allow for differentiation between sharing- and non-sharing-economy businesses. This will also help identify the variables that will likely be relevant for future models predicting whether a business operates in the sharing economy. It is important to note that the following analysis is exploratory and will likely undergo further development in future work.

Based upon previous work, it has proven difficult to get an accurate and robust classification of sharing- and non-sharing-economy businesses. As a result of this, unsupervised models will be discussed in this section. Further work to classify businesses is ongoing, once complete, the more powerful supervised learning approach will be investigated.

K-means clustering allows data to be grouped according to characteristics

K-means clustering analysis compares the characteristics of multiple entities, resulting in similar entities being clustered together and dissimilar entities being clustered apart. This has been identified as an appropriate technique for our purposes as it allows the investigation of characteristics which may, in combination, differentiate between sharing and non-sharing economy businesses⁴ .

If sharing-economy businesses do share common characteristics that are distinct from non-sharing-economy businesses, then good discriminatory variables should be identified. If they do not, further work would be required to determine whether it is the quality or quantity of the variables used that needs to be enhanced to allow for accurate classification of sharing economy businesses. The methodology used for this analysis is outlined further in Appendix D. The method of this analysis differs to that used for the Chi square tests⁵, where rather than analysing each e-commerce variable independently, clustering analyses in all 10 variables simultaneously.

This research finds that it is not possible to conclusively discriminate between sharing- and non-sharing-economy businesses using Annual Business Survey (ABS) numeric economic data for individual businesses. Better results were achieved using binary Yes or No e-commerce survey variables, with results further improved by combining the two datasets; however, none of the results are considered conclusive.

Cluster analysis using ABS data

Based on the ABS data alone, cluster analysis was unable to distinguish between sharing- and non-sharing-economy businesses

Cluster analysis was applied to the ABS data, which grouped the data according to similarity based on the 11 variables used⁶. ABS data were split into six clusters, as shown in Table 2. The results suggest that the ABS data cannot successfully discriminate between sharing- and non-sharing-economy businesses.

The 151 businesses used for this analysis, 81 sharing and 70 non-sharing, are expressed as proportions in Tables 2 to 5. The analysis has successfully distinguished between sharing- and non-sharing-economy businesses if there is a notable difference in the proportion of either in a particular cluster. This is not the case as most businesses are clustered together (Group 3). The only group shown to discriminate between sharing- and non-sharing-economy businesses was Group 6, but this group only had five observations.

Table 2: Results of clustering the Annual Business Survey data into six groups

Group	Units: %						Total
	One	Two	Three	Four	Five	Six	
Proportion of sharing economy businesses	1.2	7.4	88.9	1.2	1.2	0	100
Proportion of non-sharing economy businesses	0	5.7	82.9	4.3	0	7.1	100

Source: Office for National Statistics

As Group 3 contained a large number of businesses, these were subjected to a further round of cluster analysis to split into a further six clusters. The results were unsuccessful again in differentiating between sharing- and non-sharing-economy businesses, as most observations are clustered into Group 4, see Table 3. This provides strong evidence that, according to the ABS variables alone, the majority of sharing- and non-sharing-economy businesses have similar characteristics.

Table 3: Results of clustering Group 3 of the Annual Business Survey data into six groups

Units: %

Group	One	Two	Three	Four	Five	Six	Total
Proportion of sharing economy businesses	0	6.1	12.2	81.7	0	0	100
Proportion of non-sharing economy businesses	1.7	15.5	5.2	74.1	1.7	1.7	100

Source: Office for National Statistics

Cluster analysis using e-commerce data

E-commerce data better distinguishes between sharing-economy and non-sharing-economy businesses

Cluster analysis was also applied to the binary Yes or No e-commerce data, which showed a better distribution of businesses between clusters compared with the ABS cluster analysis, as shown in Table 4. As with the ABS data, e-commerce businesses were clustered into six groups. In this analysis, Groups 2 and 5 show a good split between sharing- and non-sharing-economy businesses.

It appears that the number of Yes responses to the E-commerce Survey provides a broad indication of whether a business operates in the sharing economy

It was expected that typical sharing-economy businesses would fall into Group 1, as it only contained businesses where all 10 sharing-economy e-commerce variables⁵ were answered positively. However, the data show 8.6% businesses in this group were non-sharing-economy businesses, see Table 4. This suggests that utilising a wide range of technologies is not necessarily a strict requirement for being a sharing-economy business.

Group 2 contains businesses who responded positively to most e-commerce questions. Of businesses within this cluster 48.1% were classified as operating in the sharing economy, see Table 4; therefore, this group appears to be a better indicator of whether a business enables sharing-economy activity.

Table 4: Results of clustering e-commerce data into six groups

Units: %

Group	One	Two	Three	Four	Five	Six	Total
Proportion of sharing economy businesses	14.8	48.1	18.5	11.1	2.5	4.9	100
Proportion of non sharing economy businesses	8.6	8.6	15.7	21.4	44.3	1.4	100

Source: Office for National Statistics

Group 5 contains businesses who responded negatively to the majority of the e-commerce variables, with the exception of whether the business has a website. This cluster appears to be typical of non-sharing-economy businesses, with 44.3% of businesses within it classified as being non-sharing-economy businesses.

Based on these results, it appears that the number of Yes responses to the E-commerce Survey provides a broad indication of whether a business operates in the sharing economy.

Cluster analysis using ABS and e-commerce data

Combining ABS and e-commerce data gives a better indication of cluster groupings

Cluster analysis of the combined ABS and e-commerce data was also undertaken, to determine whether the inclusion of variables from both surveys can further improve the ability of the model to discriminate between sharing- and non-sharing-economy businesses. Combining the two surveys provided 21 variables for analysis, 11 from the ABS⁶ and 10 from the E-commerce Survey⁷.

Six clusters were again produced using variables from both surveys, see Table 5. Group 6 appeared to differentiate effectively between sharing- and non-sharing-economy businesses, containing 1.2% of sharing-economy business and 41.4% of non-sharing-economy businesses. Group 4 also appeared to provide a fair degree of discrimination, containing 69.1% sharing-economy businesses and only 18.6% non-sharing-economy businesses.

The results suggest that even though the ABS data on its own contain little discriminatory power, it does contain important discriminatory attributes, which, when combined with the more powerful discriminatory variables from the e-commerce data, provide better results than using either dataset in isolation. While providing promising results, this analysis also highlights that not all businesses can be accurately classified.

Table 5: Results of clustering Annual Business Survey and e-commerce variables together into six groups

Group	Units: %						Total
	One	Two	Three	Four	Five	Six	
Proportion of sharing economy businesses	7.4	1.2	3.7	69.1	17.3	1.2	100
Proportion of non sharing economy businesses	20	0	4.3	18.6	15.7	41.4	100

Source: Office for National Statistics

Future work points to Chi-squared Automatic Interaction Detectors (CHAID) and additional data sources

Overall, the cluster analysis suggests that ABS variables alone cannot successfully differentiate between sharing- and non-sharing-economy businesses. While providing slightly better categorisation, e-commerce variables alone are also unable to accurately classify the majority of businesses. The best results are produced when combining both surveys; however, even then, not all businesses are correctly categorised as operating in either the sharing or non-sharing economy.

Upon completion of the sharing-and non-sharing-economy classification, the next step will be to apply the more powerful supervised learning technique. It is believed that such technique will be able to better identify the characteristics associated with sharing-economy businesses.

Furthermore, the use of additional variables describing businesses, such as from other ONS surveys and administrative data, is also recommended, as these will likely improve the discriminatory power of the analysis.

Notes for: ONS data science project

1. All conclusions drawn from tabulated results are significant at 95% confidence limit, as defined by an appropriate statistical test. (see: Kanji, G. K. (2006). 100 statistical tests. Sage.)
2. See Section 6 for discussion on results from the Chi square tests.
3. Employment, and the following scaled by employment: total turnover, amounts payable for telecommunications services, amounts payable for advertising and marketing services, gross wages and salaries, contributions to pension funds, total employment costs, total purchases of goods materials and services, total value of all stocks at beginning of period, total value of work in progress at end of period, computer software developed by staff for 1 or more years of use
4. The allocation of group number in each analysis is random. There is no relationship between the same numbered groups at different stages of the analysis.
5. Website, Online ordering or reservation or booking, Description of goods or services, Order tracking, Possibility for visitors to customise or design the goods or services online, Links or references to social media profiles, Social networks, Business blogs or microblogs, Multimedia content sharing websites, Personalised online content for repeat customers.
6. Employment, and the following scaled by employment: total turnover, amounts payable for telecommunications services, amounts payable for advertising and marketing services, gross wages and salaries, contributions to pension funds, total employment costs, total purchases of goods materials and services, total value of all stocks at beginning of period, total value of work in progress at end of period, computer software developed by staff for 1 or more years of use.
7. Website, Online ordering or reservation or booking, Description of goods or services, Order tracking, Possibility for visitors to customise or design the goods or services online, Links or references to social media profiles, Social networks, Business blogs or microblogs, Multimedia content sharing websites, Personalised online content for repeat customers.

8 . Acknowledgments

Authors: Pauline Beck, Michael Hardie, Natalie Jones, and Ash Loakes.

The authors would like to acknowledge the contributions from Diane Coyle, Jon Gough, Paul Richards, Chloe Gibbs, Andrew Jowett, and the ONS Data Science Campus.

9 . Annex A: Methodology related to the definitional work

This annex relates to the definitional work done by Office for National Statistics (ONS), highlighted in section 3. ONS held a “Sharing-economy definitions hack day” with the aim to debate and agree on an ONS definition of the sharing economy. This day brought together experts from various areas, with representation of teams from labour market, social surveys, national accounts and economic statistics. In order to help the discussions, we framed the day using exercises: looking at businesses that are more difficult to classify as being part of the sharing economy or not.

For example, we considered Uber as opposed to a local taxi firm. We found that they both had a website or an app; these apps both allowed ordering a vehicle, rating your driver and both had location-based functionality. It was then clear that functionality and innovation might not be the best indicator for classifying sharing-economy businesses. However, we posed the question whether the business might still be able to operate without a digital platform to facilitate their business: the local taxi firm was likely to continue to operate, while Uber might not.

We also considered what the primary use of the asset was; in this case the vehicle. We discussed that if the car was bought with the primary aim to work for Uber or the taxi firm full-time, then we would not consider this sharing-economy activity. While if the primary use of the vehicle was for personal use, then it would. Similarly, we would not classify all activity on the Uber website as being part of the sharing economy. Indeed, if Uber is used solely as a source of main income, then this would be captured within labour market statistics. Although it would be very useful to determine how many people are using these platforms for full-time work.

The challenge will be to actually collect this information from businesses and individuals. This survey data would also need to be compared with commercial data to give a full picture. These discussions helped to determine that sharing-economy activity can be seen as exploiting slack, or avoiding waste of assets within the economy, which is different to when the primary aim of owning assets is intent to profit from renting.

There are many debates around defining the sharing economy; we sought feedback from other government departments to refine our concepts. Following this hack-day, a cross-government sharing-economy group to consider our ideas and gather feedback took place. Our group was made up of HM Revenue and Customs (HMRC), the Department for Business, Energy and Industrial Strategy (BEIS), the Department for Digital, Culture, Media and Sport (DCMS), Bank of England, Welsh Government, Scottish Government and Northern Ireland Office.

The group was very receptive to ONS's progress, but most importantly it helped us to understand policy needs better and which data we should be looking to capture. There was a strong interest to look at how many businesses are enabling sharing-economy activities and whether trust and accessibility are driving the use of platforms. There was also a high interest in understanding who the individuals participating in the sharing economy are and whether they are being disadvantaged or vulnerable. It was also considered important to collect information on income (whether they are using it as primary or secondary source of income) and whether users are making saving plans towards their pension.

We have also actively consulted with other entities, such as Statistics Canada (STATCAN), France's National Institute of Statistics and Economic Studies (INSEE), National Bureau of Statistics of China (NBS), [VVA policy](#), [Open Data Institute](#) (ODI), [Sharing Economy UK](#) (SEUK), who provided very useful feedback to help us shape the definition and open various routes for further collaboration.

10 . Annex B: Challenges in measuring and defining the sharing economy

The sharing economy is fundamentally based on peer-to-peer transactions: "platforms" enable the matching of the supply and demand. With this effective, quick and cheaper matching, individuals no longer need a business to act as an intermediary to match their supply and demand. Faster internet speed and mobile access have also broadened the number of potential participants, creating markets that otherwise would not be viable. The digital platforms make transactions possible almost instantly. For example, you could not advertise an ad-hoc taxi service via a newspaper every time you have a spare couple of hours, which is what certain platforms essentially allow individuals to do.

This business model creates new challenges when trying to measure the sharing economy. A number of individuals use sharing-economy platforms due to the increased ease of access to the supply and lower cost of transactions or services. Furthermore, individuals may also be using sharing economy platforms to supplement their existing income. In 2014, Nesta estimated that [25% of the UK adult population is sharing services and goods online](#) in some way and Coyle estimated that 3% of the UK workforce is already providing a service through the sharing economy. Certain activities involving exchanges between individuals are not included in gross domestic product (GDP) and the Consumer Prices Index (CPI), particularly as the latter only includes purchases made by consumers from businesses. As a result, Coyle suggests there may be under-reporting of lower prices, which consumers gain from sharing economy platforms.

Another challenge resides in how we classify activities and businesses. As previously mentioned, the sharing-economy platforms have an innovative business model that facilitates peer-to-peer transactions complicating how we classify businesses. ONS uses the [Standard Industrial Classification: SIC 2007](#) as a basis for collecting data and publishing official business statistics. Sharing-economy platforms do not easily fit this classification system, as it is possible for a business in any given industry to contribute to the sharing economy. Similarly, ONS uses the [Standard Occupational Classification: SOC 2010](#), which might not always be representative of some sharing-economy occupations.

Another challenge is that transactions are not always financial: individuals might be swapping goods or services, with no money being transferred. This complicates the situation, as the internationally-agreed definition of GDP excludes non-monetary transactions – such as the provision of free goods or services, by donating, lending or swapping. ONS is still interested in these activities, to understand economic well-being better.

11 . Annex C: Statistical analysis

Using the Annual Business Survey (ABS) and the E-commerce Survey, 23 questions were of interest to support our investigations into the sharing economy, see Table 6.

Table 6: Variables used for analysis

Annual Business Survey Questions	E-Commerce Survey Questions
Total turnover	Does this business have a website, either its own or third party?
Amounts payable for advertisement and marketing services	Does this business' website have online ordering or reservation /booking
Amounts payable for telecommunication services	Does this business' website have description of goods or services
Contribution to pension funds	Does this business' website have order tracking
Gross wages and salaries	Does this business' website have the possibility for visitors to customise or design the goods or services online
Services purchased for resale without further processing	Does this business' website have personalised content for regular /repeat visitors
Total employment costs	Does this business' website have links or references to this business' social media profiles
Total purchases of energy, goods, materials and services	Does this business use social networks, for example Facebook, LinkedIn, Xing, Viadeo, Yammer etc.
Work in progress at the beginning of the period	Does this business use business blogs or microblogs, for example Twitter, Present.ly, etc.
Work in progress at the end of the period	Does this business use multimedia content sharing websites, for example YouTube, Flickr, Picasa
Computer software developed by own staff	Does this business use wiki based knowledge sharing tools
N/A	During 2016, of total turnover, what percentage resulted from orders received via a website or app?

Source: Office for National Statistics

Methodology

Business selection

A list of around 100 businesses was compiled, using the technique outlined in Section 6. This is the list of businesses that we have used throughout our article. Each business was checked to ensure compliance with our new definition of the sharing economy and our decision tree. It is important to remember that the decision tree only implies that a business is likely or unlikely to be in the sharing economy if they possess these traits.

Prior to sampling for the 2016 survey, 107 businesses that enable sharing-economy activities were included in samples for both the ABS and the E-commerce Survey. Out of the total 107 businesses in ABS, 24 were not sampled. This was due to the following three reasons:

- businesses being reclassified to a different standard industrial classification (SIC) that are out of scope for the two surveys
- companies ceased trading
- osmotherly – this is when a small business can't be sampled if they've been selected by another survey prior to being selected by the ABS or e-commerce

This implies that businesses in the sharing economy may start-up and then cease trading reasonably regularly. Out of the remaining businesses, 45 responded to ABS and 24 responded to the E-commerce Survey. To overcome this, data were imputed manually for the remaining businesses. Each business that had not returned data was researched extensively to predict how they would respond to each Yes or No question. When data could not be imputed manually (when uncertain), the cell was left blank. The final dataset had a total of 81 businesses from the E-commerce Survey and 45 from the ABS.

For comparison, businesses that do not enable sharing-economy activities were also researched. For the ABS, responses from 6,451 other businesses were extracted from the database. They had similar Standard Industry Codes (SIC) (ranging from SIC 62011 to SIC 82990 and then SIC 96010 to SIC 96090) and also have similar employments to the businesses in the sharing economy.

For the E-commerce Survey, responses from 152 other businesses were collected. A similar technique to matched pairs was used, but all businesses with the same SIC and similar employments were selected. This meant we had no bias as to which business was chosen.

Descriptive analysis techniques

E-commerce Survey

The E-commerce Survey produces categorical results (Yes or No). To analyse the data Chi-squared tests were performed. The Chi-squared tests test for independence between two groups. In this case, a test of independence between businesses that enable sharing-economy activities and those sampled businesses that do not enable sharing-economy activities was executed.

Table 7: Chi-square values and p-values for e-commerce variables

E-Commerce Question	Chi-Squared value	p-value
Website	33.726	0
Online ordering or reservation/booking	63.725	0
Description of goods or services	37.606	0
Order tracking	2.287	0.131
Possibility for visitors to customise or design the goods or services online	55.546	0
Links or references to social media profiles	39.554	0
Social networks	51.245	0
Business blogs or microblogs	73.025	0
Multimedia content sharing websites	71.823	0

Source: Office for National Statistics

Annual Business Survey

The data for the Annual Business Survey (ABS) in both groups follow a Pareto distribution. This means many of the standard hypotheses tests were not feasible. Inter-quartile ranges were used to compare the spread of the data in each variable across both groups of businesses. Medians and means were also used to make comparisons.

12 . Annex D: K-means clustering

ABS variables must be scaled and normalised

To ensure data are measured on the same scale, the variables are scaled and normalised. In their raw form, some numeric Annual Business Survey (ABS) are orders of magnitude higher than others. Data are scaled across variables by using employment as a denominator for each variable, for example, turnover becomes turnover or employment, resulting in a percentage. This ensures contributions are not skewed in favour of larger businesses. Normalisation then scales data vertically, again to prevent a variable having too much influence over the analysis and is done as follows:

Val = the original value prior to normalisation

Max = the maximum value of the variable

Min = the minimum value of variable

High = the highest value of the chosen range of the normalised data

Low = the lowest value of the chosen range of the normalised data

Val' = the new value after normalisation

$$Scale = \frac{High - Low}{Max - Min}$$

$$Offset = \frac{(Max \times Low) - (Min \times High)}{Max - Min}$$

$$Val' = (Val \times scale) + Offset$$

K-means clustering is conducted on the data

K-means clustering is a method of partitioning observations into clusters, whereby each observation corresponds to the cluster with the nearest mean; the mean serves as a cluster centre. This is achieved using the following algorithm:

$$J = \sum_{j=1}^K \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2,$$

where:

$\|x_i^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j , and is an indicator of the distance between the n data points and their respective cluster centres

The algorithm is composed of the following steps:

1. Place K cluster points in the space ¹ represented by the variables that are being clustered, where K equals the number of clusters. These points represent initial group cluster centres or means.
2. Assign each observation to the group that has the closest cluster centre.
3. When all observations have been assigned, recalculate the cluster means to be the new positions of the K cluster centres.
4. Repeat steps 2 and 3 until the cluster centres no longer move. This produces a separation of the objects into groups from which the metric to be minimised can be calculated.

Eleven ABS variables and ten e-commerce variables relevant to the sharing economy were analysed, which describe the economic characteristics of 151 businesses: 81 assigned to the sharing economy and 70 to the non-sharing economy. The purpose of the analysis is to see if the combination of variables resulted in the sharing-economy businesses clustering together because their measured characteristics are more similar to each other than to the measured characteristics of the non-sharing-economy businesses.

The number of clusters required must be determined before the analysis is carried out. As there are two groups labelled within the data, sharing economy and non-sharing economy, it may be thought that two clusters are required. However, this is not correct because most datasets have unique observations which cluster by themselves, leaving all remaining observations in the other group. Experimental clustering was initially undertaken with six and ten clusters; from this it was assessed that six clusters provided more understandable interpretation than ten clusters. As a result, six clusters are used throughout this analysis.

The distance is measured between each business plotted in conceptual space and the cluster centres; this distance is known as the Euclidian distance. The software maximises the sum of squares between clusters and minimises the sum of squares within clusters to allocate individual businesses to the optimal cluster.

Notes for: Annex D: K-means clustering

1. The number of dimensions in this space is equivalent to the number of variables used in the cluster analysis.