Classifying vacancy data at 6-digit level SOC 2020: A feasibility study

A report prepared for

Gatsby Charitable Foundation

By

Dr Jeisson Cárdenas-Rubio

September 2021

Warwick Institute for Employment Research

University of Warwick

Dr Jeisson Rubio Cárdenas

Warwick Institute for Employment Research

University of Warwick

Coventry CV4 7AL

Tel: +44(0)24 76574751

Jeisson.Cardenas-Rubio@warwick.ac.uk

Contents

List of Figures

List of Tables

Acknowledgements

The institute for Employment Research at the University of Warwick would like to sincerely thank the Gatsby Charitable Foundation for supporting this feasibility study, in particular Daniel Sandford Smith.

The IER would like to acknowledge that the vacancy database used in this feasibility study was partly created with funding from the Department for Education as part of the LMI for All service.

1. Setting the context

This report sets out the process and findings of a feasibility study to code and check the statistical consistency of vacancy data at 6-digit level of the UK Standard Occupational Classification (SOC) 2020. It concludes with a summary of the analyses and some key learning from the process.

An occupational classification (e.g., SOC) is a framework used to group jobs based on the tasks and duties undertaken in the job. This framework is a key input for conducting standardised statistical labour market analysis since it allows the measurement, for instance, of the type of jobs (task complexity) demanded or offered in a country. As responsible for maintaining the occupational classification, the Office for National Statistics (ONS) has recently produced the UK Standard Occupational Classification (SOC) 2020 at 4-digit level. However, as mentioned by the GSS (2021), the current 4-digit level SOC disaggregation does not provide enough detail for some SOC users (e.g., Vocational Education and Training (VET) programmes need a higher level of disaggregation to properly provide the skills required for a certain job).

Given this issue, the ONS has started developing a more detailed version of the SOC 2020, named The SOC Extension Project (SOC EXT). This project aims to disaggregate the 4-digit SOC to a fifth level (6-digit SOC 2020). This fifth level identifies in more detail the occupational groups that share similar tasks (GSS, 2021). Although this work remains experimental, there is a first draft of the SOC 2020 extended framework[1]. This draft is particularly useful because it allows testing the classification at 6-digit digit level in different databases.

Among the different databases that can potentially be classified to 6-digit SOC, database with data from job portals (or vacancy websites) offer a promising source which the SOC Extension could be applied to. This could produce more detailed information and data on a labour market. Job portal information has notably attracted researchers and policymakers' attention because it is a source that can offer (at low cost) detailed labour demand information, which would otherwise be too complex to gather via other means.

Different companies, such as Burning Glass, Emsi, Adzuna etc., have started collecting and organising vacancy information from online sources. These data have been valuable in identifing skills and occupational demand in the UK. Nevertheless, some of these companies have coded the vacancy information to SOC 2010 at a 4-digit level (in the best case), while others use different classifications and aggregations. Moreover, there is a lack of transparency: in most instances, companies do not reveal how information is collected, cleaned, classified, deduplicated and organised, making it difficult to evaluate the scope and consistency of the vacancy data at different disaggregation levels (e.g., at SOC 2020 6-digit levels).

[1] First draft of SOC 2020 extended framework, https://gss.civilservice.gov.uk/wp-content/uploads/2020/10/draft-standard-occupational-classification-2020-extended-framework.ods

As part of developing and extending the data available through LMI for All[2], the LMI for All team have piloted a web-scraping technique[3] to collect UK vacancy data from online job portals. This work was funded by the Department for Education. The collection period started in and has continued from February 2019 for the Guardian and Reed, while for the NHS job portal the collection period began in May 2020[4]. This dataset is coded to UK SOC 2010 4-digit and SOC 2020 4-digit level primarily using CASCOT (Computer Assisted Structured Coding Tool[5], Jones and Elias, 2004). The LMI for All team have checked the quality and consistency of the coding and data (see Barnes et al. 2021). Additionally, the LMI for All team standardised variables such as skills and qualification requirements, wages, etc. which could be helpful for coding these data at 6-digit SOC level. Thus, the data collected and processed by IER and the use of a tested tool, such as CASCOT, are a good starting point to evaluate the possibility of coding job portal information to SOC 6-digit level.

Given these sources of information available, this research sets out a possible method to code the vacancy database developed as part of the LMI for All service and to assess this coding using a statistical approach to evaluate the consistency of the vacancy (volatility) data at 6-digit SOC level. The aim of this proposed project was to undertake a feasibility study to assess and evaluate whether information from job portals can be classified to 6-digit SOC level and, therefore, provide a new source of information at this disaggregated level.

Following this introduction, the next section presents an overview of the methodology and the approach which includes:

- Using CASCOT outputs for coding job portals information at 6-digit level;
- Applying word-based matching methods that use job titles and CASCOT's outputs to code vacancy observations into SOC 2020 at 6-digit level;
- Mapping the skills information in the vacancy database to classify "hard-to-code" job titles;
- Piloting a machine learning algorithm that uses job titles and skills for coding vacancy observations to SOC 2020 at 6-digit level; and
- Testing the consistency of the vacancy database for statistical inference at 6-digit level SOC 2020.

[2] For more information on the LMI for All service, see https://www.lmiforall.org.uk/

[3] "Web scraping" consists of a computerised method to automatically collect information from across the internet; in this instance job portals are being scraped to collect information on advertised vacancies

[4]These job portals are defined as significant in terms of the number of visits per day and volume of information available.

[5] Developed at the University of Warwick by the Institute for Employment Research.

2. Approach to coding vacancy information to SOC 2020 at 6-digit level

Job portal information contains different variables that might allow occupational coding at 6-digit level. The job title variable is the backbone for coding a vacancy database to SOC 2020. This variable indicates the type of tasks required for a certain job position and, thus, serves as an important element to ascertain the occupational code of an observation in the vacancy database (Cárdenas-Rubio, 2020). The second component is the skills demanded (Section 5 details how the skill variable is built). This variable might serve to confirm or check the vacancy classification when job titles are not sufficient (or too ambiguous) to determine a SOC code at 6-digit level. The third main component comes from the SOC 2020 extended framework. This contains SOC 2020 at 4-digit level and also the (preliminary) Sub-Unit Group (codes) and Group Title (labels) of the occupational groups at 6-digit level. This last variable was particularly useful for coding the vacancy data, as reported in the following sections.

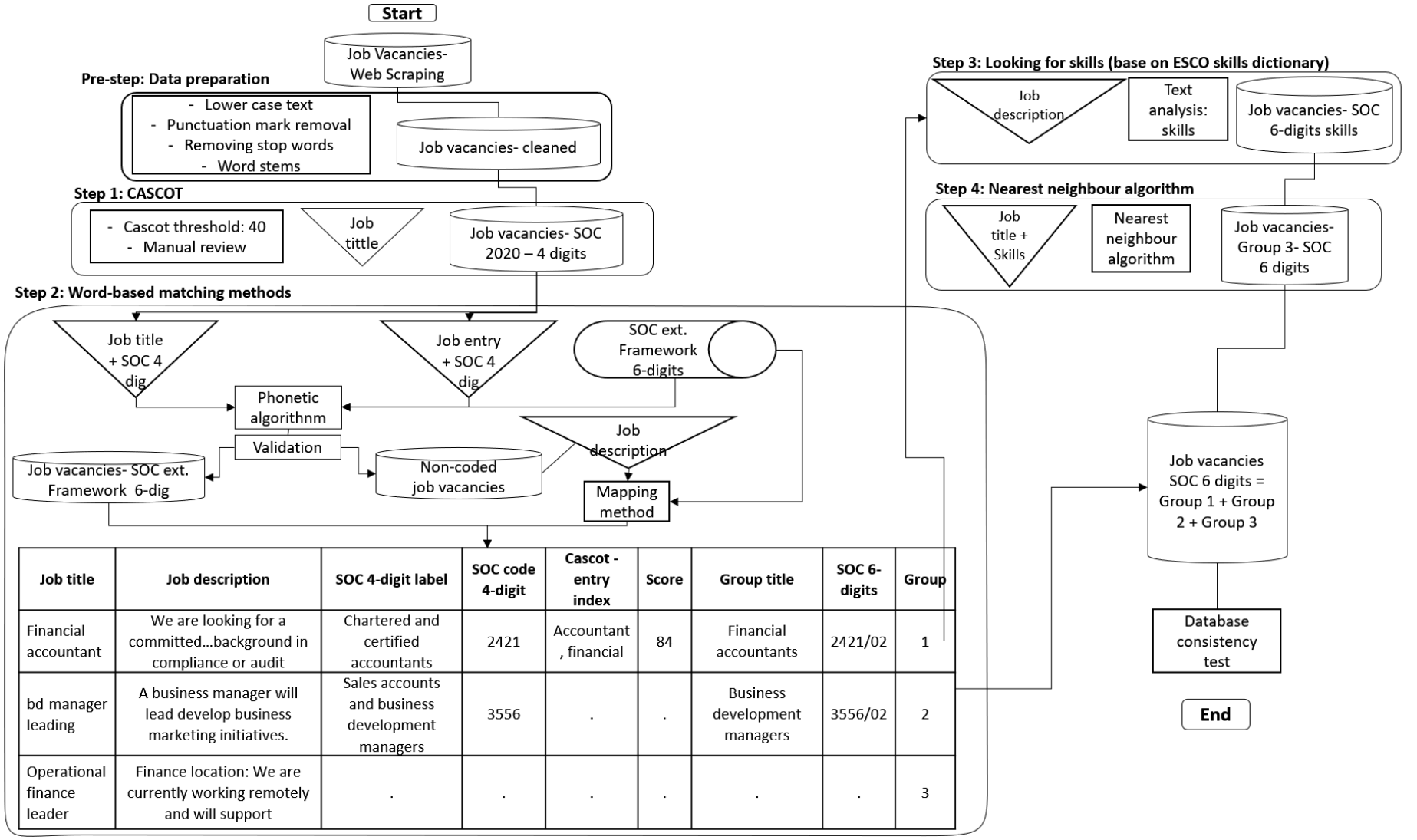
In terms of classification methods, CASCOT, word-based matching methods (better known as "fuzzy merge" methods) and machine learning algorithm (nearest neighbour algorithm)[6] were used. Figure 1 shows the steps for coding and testing the vacancy database at 6-digit level SOC 2020. This methodology starts with the vacancy database scraped and created for the LMI for All service. Before data processing, the vacancy database was prepared for this project. Specifically, this process involved the implementation of text mining techniques (such as stop words, stemming, etc.) to clean "noisy" information (such as prepositions, symbols etc.). Once the data were cleaned, the methodology followed the next four steps:

- CASCOT classification;
- Word-based matching methods;
- Mapping skills in the vacancy database;
- Application of nearest neighbour algorithm; and
- Consistency testing: data volatility.

Next, each of these steps to classify and test the vacancy database at 6-digit level SOC 2020 are explained.

[6] These methods have been tested on different databases to classify job titles into occupational codes with high accuracy results (Jones and Elias, 2004; Gweon et al. 2017; Lima and Bakhshi, 2018).

Figure 1. General scheme for coding vacancy data at 6-digit level SOC 2020



Job title	Job description	SOC 4-digit label	SOC code 4-digit	Cascot - entry index	Score	Group title	SOC 6-digits	Group
Financial accountant	We are looking for a committed…background in compliance or audit	Chartered and certified accountants	2421	Accountant , financial	84	Financial accountants	2421/02	1
bd manager leading	A business manager will lead develop business marketing initiatives.	Sales accounts and business development managers	3556	.	.	Business development managers	3556/02	2
Operational finance leader	Finance location: We are currently working remotely and will support	.	.	.	.	.	.	3

2.1. Data preparation

As previously noted, the IER through the LMI for All project has been developing a large vacancy database from the main UK job portals since February 2019. These data have been compiled using web scraping techniques[7] collecting huge volumes of information automatically from the three UK job portals, IER then cleans and standardises the information to ensure its consistency. As information is downloaded, the job vacancies are deduplicated and coded for statistical analysis using a number of variables, including: text pattens within the vacancy posting (e.g. 'experience required') and wage offered; geographical areas (such as Local Enterprise Partnerships (LEP)); the skills demanded[8]; and the Standard Industry Classification (SIC) for sector[9] (Barnes et al., 2021).

Given the data availability, this study uses the data from February 2019 to April 2021. Each observation in the database is a vacancy. A vacancy can require one or more people (the total number of jobs or job placements available). The total number of observations (vacancies) in the database is 2,449,076, while the number of jobs is 2,856,947.

Table 1 lists the primary variables included in the dataset. The job titles and the vacancy description are the primary input variables of the vacancy database at 6-digit level SOC 2020. However, it is necessary clean these variables before starting the automatic codification. The classification algorithms use the characters in variables, such as the job titles and descriptions, to associate each observation with the most similar occupational category, in this case to SOC 2020. As shown by Cárdenas-Rubio (2020), the performance of these algorithms is sensible to the type and number of words used as inputs. The higher number of unrelated words to occupational information in the input variables, the lower the algorithms' performance (accuracy).

Employers usually write unrelated information to the tasks to be performed when posting a vacancy. For instance, it is common to observe job titles such as "Event Coordinator - *London*", "hr coordinator *12 month ftc*", "sen teaching *assistant part time*". Moreover, the use of prepositions, punctuations, special characters and so on, also affects the classification processes. Consequently, a cleaning process was conducted on the input variables in order to improve classification accuracy. These methods included:

- **Lowercase text:** transformation of all the characters in the input variables into lowercase.
- **Punctuation mark removal:** deletion of all punctuation marks.

---

7 Computerised methods to automatically collect information from the Internet.

8 We use European Skills, Competencies, Qualifications and Occupations (ESCO). This widely used multilingual classification system provides a list of 13,485 skills.

9 In addition to vacancy information, job portals also provide company names. Alternatively, Companies House (https://www.gov.uk/government/organisations/companies-house) provides information such as the name and SIC code of companies registered in the UK. We use these names to combine both the vacancy and Companies House information and, by doing so, we identify the sector of the vacancy observations.

- **Removing stopwords:** “stopwords” refers to a list of words that do not provide relevant information to identify the occupational group. This list includes words such as prepositions, pronouns, etc. Importantly, through a manual verification process, stopwords have also been included to this list of specific words that are not related to occupational information and frequently appears in the job titles and job descriptions (e.g., “experienced”, “opportunity”, “part time”, “full time”, year, city or region names, etc.).
- **Word stem:** Words are reduced to their grammatical root. Words that are in plural are transformed to their singular form. This process helps, for instance, to compare words regardless of whether they were written in the plural or singular forms.

These standard procedures were implemented on the vacancy database information and on the SOC 2020 extended framework. After these steps, the data were processed by the different classification algorithms.

Table 1. Main variables of the vacancy dataset

Variable	**Definition**
Job title	Short description of the job title offered
Job link	Hyperlink to vacancy on job portal
Vacancy description	Detailed information about the profile required to fill the vacancy (e.g., skills, job experience, type of contract, etc.)
Location	Place where the vacancy is available, coded to county and LEP level
Wage	Continuous variable which indicates the amount of money that the hired person will receive
Job openings	The number of positions open for the vacancy.

Source: Barnes et al., 2021.

2.2. CASCOT classification

CASCOT is a tool classification to assign an occupational (or industrial) code to texts. Over the last two decades and partnership arrangements with labour market experts, this tool has designed a set of tested rules (such as downgraded words, equivalent word ends, abbreviations, replacement words etc.) and job indexes to reveal the best matches between job titles and an occupational classification (Jones and Elias, 2004; Wageindicator, 2009; IER, 2018). Importantly, IER has closely worked with the ONS on updating the SOC 2010 to SOC 2020 (ONS, 2021). Because of this work, CASCOT has been updated and can code job titles to SOC 2020 at 4-digit level. CASCOT was,

therefore, a good starting point in classifying vacancy job titles to SOC 2020 4-digit level and potentially disaggregate the vacancy information at SOC 2020 6-digit.

Consequently, the first step of this methodology coded the vacancy database to SOC 2020 at 4-digit level using CASCOT. The codification proceeded as follows: the cleaned job titles from the vacancy database were the input variable that CASCOT used to assign the SOC 2020 at 4-digit best match with corresponding similarity scores and the closest index entry[10].

A total of ("non-duplicated") 789,208 job titles[11] were found in the UK vacancy database. These 789,208 job titles were processed through CASCOT. By doing so, around 88.6% of observations (699,528) in the vacancy database received an occupational code at the SOC 2020 4-digit level (see Technical Appendix. Table A1). Table 2 shows an example of the CASCOT results. The first column indicates a list of job titles from the vacancy database, while the other columns show the outputs after running the data through the CASCOT tool. The first row of this table indicates that a company required a draughtsperson. CASCOT suggested that the best match for "Draughtsperson" is Draughtsman, with a similarity score of 57[12]. The result indicates that this observation's occupational code is "3120 CAD, drawing and architectural technicians". Similar examples are provided in rows two and three of Table 2. The last row of Table 2 shows an example where the CASCOT similarity score is relatively low. As Jones and Elias (2004) mentioned, a reasonable threshold to accept the result from CASCOT is 40 or more.

[10] Index entry refers to a comprehensive list of job denominations (or job titles) associated with one occupational code at four-digit level.

[11] This list of job titles does not indicate that there are 789,208 different/unique job titles in the UK labour market. A job title in this list might slightly differ from another title due to "extra" words that were not removed in the data preparation steps (e.g., "IT Procurement Manager - OJEU", "IT Procurement Manager").

[12] On a scale of 0 to 100, the higher the number, the higher level of certainty that a match code is the correct one.

Table 2. Example of CASCOT coding results

Job titles (From vacancy dataset)	Index entry	Similarity score	SOC 2020 4-digit
Draughtsperson	Draughtsman	57	3120 CAD, drawing and architectural technicians
Primary SENCO	SENCO	75	2316 Special needs education teaching professionals
system engineer	Engineer, systems	83	2133 IT business analysts, architects and systems designers
operational finance leader	Leader, team, operations, computer	36	.

The accuracy of the CASCOT tool has been tested in various studies (see for example Jones and Elias, 2004; Wageindicator, 2009; IER, 2018). However, a visual inspection of the vacancy database was conducted in order to provide more evidence of the CASCOT accuracy for coding job vacancies to SOC 2020 4-digit level.

The visual inspection consisted of a general review of the titles and the occupational codes. Subsequently, a random sample was selected (as the visual inspection is a time-consuming task, around 500 non-duplicated observations or job titles) to verify the accuracy level of the CASCOT classification and eventually correct the misclassified cases. According to this manual check, around 95% of observations had the correct occupational code (SOC 2020) at a four-digit level. Moreover, it was found that some job titles were misclassified, such as “CSCS labourer”, “patient appointments clerk clerical officer”, “assistant professor law”, among others. These cases were corrected.

Three points are worth noting from the visual inspection. First, CASCOT codes with a relatively high precision level with most of the job titles available in the vacancy dataset. This tool provided a solid input to group the vacancy information at SOC 2020 6-digit level. Second, the misclassified cases were mainly due to “noisy” information present in the job advertisements (e.g., companies or name of places). Thirdly, similarly, the uncoded observations tended to be job titles with a high present of “noisy” words or job titles that do not indicate a particular occupational group (e.g. “trustee student”, “academic specialist”, “graduates wanted”, etc.). After the visual inspection, these results were used in the next step of this methodology for coding the vacancy database at 6-digit SOC 2020.

2.3. Word-based matching methods

The second step of this methodology consisted of applying word-based matching methods between the vacancy database and the SOC 2020 extended framework. Word-based matching methods are the most straightforward way to code vacancy data to SOC 2020 at 6-digit level. Broadly speaking, these methods are a set of algorithms that compare words and match phrases that are above a certain threshold matching score. In this case, the algorithms could use the job title or Index entry (variables available in the vacancy database) and match them with the Group Title from the SOC 2020 extended framework. The result from this merging process could assign an occupation code at 6-digit level to a set of observations in the vacancy database.

As shown in Figure 1, there are two approaches to merging the vacancy data with the SOC 2020 extended framework. The first approach was using the job title and the SOC 4-digit as key variables from the vacancy database (coded with CASCOT) and the Group Title and SOC 4-digit variables from the SOC 2020 extended framework. The use of the variable SOC at 4-digit level reduces the number of possible matches produced by “jobtitle - GroupTitle”.

The second approach was to combine the vacancy database and the SOC 2020 extended framework to use the entry index and the SOC 4-digit variables from the vacancy database (coded with CASCOT) and the Group Title and SOC 4-digit variables from the SOC 2020 extended framework. The use of the variable SOC at 4-digit level reduced the number of possible matches produced by “index entry – GroupTitle” and thus helped to produce more precise results.

Merging the databases with the CASCOT entry index served two purposes: First, it helped to classify the observations not coded by the “jobtitle - GroupTitle” process. Second, this process was used to validate, to some extent, the “jobtitle-GroupTitle” merge. The same results from the “jobtitle - GroupTitle” and “index entry - GroupTitle” processes suggested a high certainty level in the vacancy database’s codification. When there were discrepancies between the combination processes, the one with the higher word-based similarly score was considered as the final output[13]. The following subsection explains

2.3.1. “Jobtitle – GroupTitle” fuzzy merge

The first approach to code the vacancy dataset to SOC 2020 6-digit level was using the job title and the SOC 4-digit as key variables from the vacancy database (coded with CASCOT) and the Group Title and SOC 4-digit variables from the SOC 2020 extended framework. There are various approaches to match string variables. In general, word-based matching methods compare sentences and match phrases that are above a certain distance score. The lower the score (less distance between the

[13] Visual inspections will be conducted to, some extent, guarantee the accuracy level of the results and evaluate complex cases (e.g., similar similarity scores).

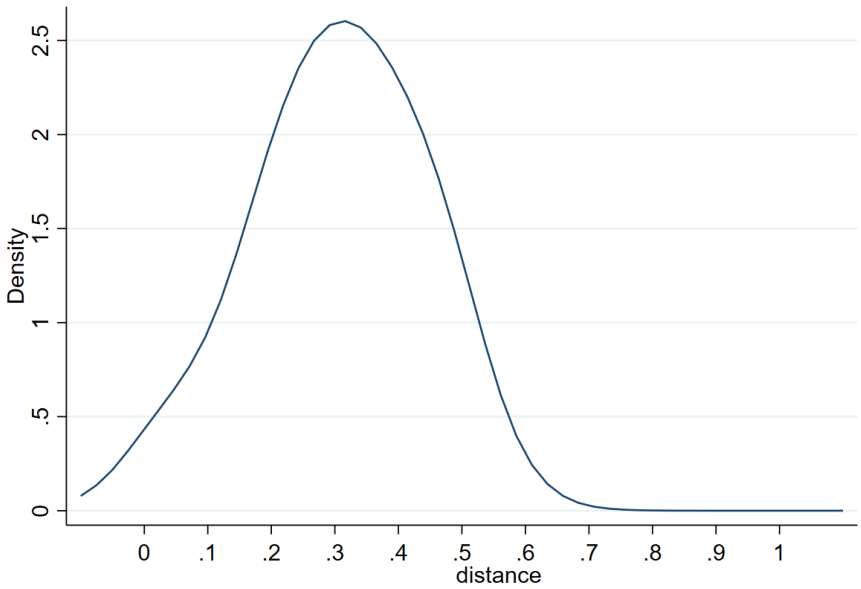
words), the more precise the results, and possibly fewer observations are matched; conversely, the higher the distance score, the less accurate the results will be, but it is likely that more observations are matched. The differences between one method and another are due to the rules used to obtain a similarity score. Testing all the word-based matching methods was beyond the scope of this feasibility study. However, Cárdenas-Rubio (2020) has shown that a set of algorithms might be particularly useful to match vacancy data with other sources of information: Jaro-Winkler, Levenshtein, Cosine similarity and Soundex.

These algorithms do not require a high computational capacity and do not take long to execute. Each of these methods has particular advantages and disadvantages. For this case and simplicity, this feasibility study focused on the Jaro-Winkler algorithm (Winkler, 1999). Unlike, for instance, the cosine similarity that only evaluates the number of same words between sentences to estimate the distance score, the Jaro-Winkler algorithm considers three elements: length of the string, number of matching characters, and number of transpositions to calculate the distance score. These characteristics are particularly useful for this exercise since the job titles from the vacancy database and group titles from the SOC 2020 extended framework might differ by a few characters (e.g., “sales admin” and “sales administrators”).

It is also possible to combine these methods to increase the number of observations matched. However, as will be noted below, the Jaro-Winkler algorithm performed relatively well. Additional combinations with other algorithms might provide only a marginal increase in the number of observations coded at the 6-digit level. Thus, this feasibility study applied the Jaro-Winkler algorithm.

Figure 2 shows the distribution of the distance score. This score is (slightly) right-skewed with a mean of 0.30 and a standard deviation of 0.10. A visual inspection of the job titles classified at SOC 2020 6-digit level showed that the matching algorithm worked fairly well when the distance score was below 0.5. For illustrative purposes, Table A2 in the Technical Appendix presents a sample of the fuzzy merge classification at various distance levels. It is almost certain that a job title was correctly classified at a distance of 0 (first ten rows of Table A2). Similar results were found when the distance is below or equal to 0.4. Above the threshold of 0.4, the fuzzy merge result was less accurate. For instance, the job title “microsoft uc consultant” was coded as 2134/01 computer games designers. The job title does not correspond to this category at SOC 6-digit level. The fuzzy merge fails to code job titles because: the CASCOT classification did not provide an accurate result; or the words in the job title did not provide enough information to distinguish the group title at SOC 2020 6-digit level (e.g., see the last row of Table A2). Although it is challenging to define a distance threshold, the visual inspection suggests that a threshold of 0.4 could provide consistent results.

Figure 2. Distribution of the "Jobtitle – GroupTitle" distance score

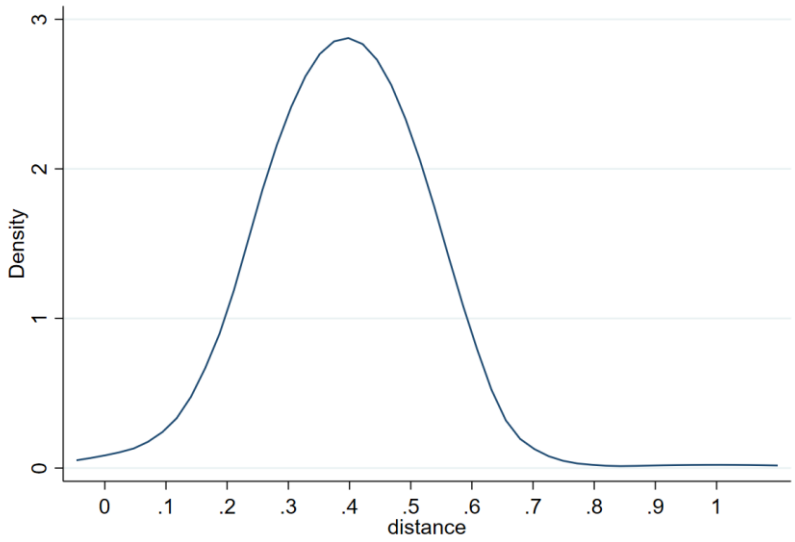


Source: IER-LMI. Own calculations

2.3.2. "Index entry - GroupTitle" fuzzy merge

Alternatively, it is also possible to merge the vacancy data with the SOC 2020 6-digit level using the CASCOT entry index (from the vacancy dataset) and the "Group Title" (from the SOC 2020 extended framework). This subsection presents the main result of conducting this exercise for 699,528 job titles. Figure 3 shows the distribution of the distance score. This variable tends to follow a normal distribution with a mean of 0.39 and a standard deviation of 0.10. A visual inspection was conducted to indicate the fuzzy merge performance. The revision showed that the algorithm performed relatively well when the score was below or equal to 0.40 score distance.

Figure 3. Distribution of the "Index entry – Group Title" distance score



Source: IER-LMI. Own calculations

Compared with Figure 2, Figure 3 tends to show higher distance scores. A closer comparison between “Jobtitle – GroupTitle” and “Index entry - GroupTitle” exercise shows that both exercises provide the same output (6-digit code) for 385,639 job titles (55.1%) (Table 3).

Table 3. Comparison between word-based matching methods

6-digit code comparison	Number of job titles	Percentage
6-digit code using entry index = 6-digit code using job titles	385,639	55.1%
6-digit code using entry index ≠ 6-digit code using job titles	313,889	44.9%
Total	699,528	100.0%

Source: IER-LMI. Own calculations

The differences between these two exercises are due to the following reasons:

- **Cases in which the “Index entry - GroupTitle” provides a lower distance score:** Given that Index entry from CASCOT is a cleaned list of job titles (without “noisy” words), in occasions the results of matching the Index entry from the vacancy dataset and the Group title from the SOC 2020 extended framework was more precise. For instance, Table 4 shows some examples of this case. The first job title in this table is “consultant luminal gastroenterologist”. When the “Jobtitle – GroupTitle” fuzzy merge was applied the result at 6-digit level was “specialist medical practitioners n e c”. While the result is different and more precise when the “Index entry - GroupTitle” fuzzy merge was applied. In this case, the 6-digit output was gastroenterologists. As mentioned above “noisy” information in the job title variable makes the “Jobtitle – GroupTitle” fuzzy merge less precise.

Table 4. Examples which the "Index entry - GroupTitle" matching provides a lower distance score

Job title	Index entries	6-digit label using index entries	6-digit label using job titles	Distance using index entries	Distance using job titles
consultant luminal gastroenterologist	Gastroenterologist	gastroenterologists	specialist medical practitioners n e c	0.05	0.36
specialist physiotherapist womens mens health	Physiotherapist	physiotherapists	electrotherapists	0.06	0.37
summer internship audit reading	Auditor	auditors	bookkeepers payroll managers wage clerks n e c	0.13	0.50
trainee butcher	butchers	butchers	slaughterers	0.13	0.51

Source: IER-LMI. Own calculations

- **Cases in which the "Index entry - GroupTitle" provides a higher distance score:** This occurred when the index entries did not provide enough information to identify the occupational group at 6-digit level or add extra information that affects the distance score. For instance, Table 5 shows that the entry index for "commercial property solicitors" was "solicitor", consequently, the fuzzy merge did not have enough information to code this observation accurately ("solicitors lawyers n e c"). In these cases, the results are more precise when it used the job title to match the vacancy database with the SOC 2020 extended framework.

Table 5. Examples which the "Index entry - GroupTitle" matching provides a higher distance score

Job title	Index	6-digit label using index entries	6-digit label using job titles	Distance using index entries	Distance using job titles
contract managers	Manager, contracts	risk managers	contract managers	0.38	0.00
commercial property solicitors	Solicitor	solicitors lawyers n e c	commercial solicitors lawyers	0.26	0.11
vehicle tech mot tester	Tester, MOT	auto electricians	vehicle testers	0.55	0.11
cover supervisors urgently	Supervisor, cover (educational establishments)	school teaching assistants	cover supervisors	0.36	0.07
ux ui design	Designer, UI	web designers	ui ux designers	0.41	0.09

Source: IER-LMI. Own calculations

Thus, combining these two methods can help classify the maximum number of job titles accurately. Given the advantages and disadvantages of these methods, it was necessary to define a decision rule to select the most accurate output for each observation. A straightforward way to set this rule was to choose the output (6-digit code) with the minimum distance score. However, as previously discussed, the word-based matching methods tend to be less accurate when the distance score is above 0.4. Job titles with a higher distance score of 0.4 should not be coded at 6-digit level with these methods. Thus, the decision rule can be formally defined as follows:

$$6-digit\ occupationa\ code_i = \min\{\text{"Jobtitle – GroupTitle"}\ \ distance\ score,\ \text{"Index – GroupTitle"}\ \ distance\ score \mid \leq 0.40\}$$

Observations in which the minimum distance score is above the 0.4 threshold require further data management to assign them a proper occupational code. With this decision rule, around 88.6% of job titles (699,528) were coded at SOC 2020 6-digit level.

The following subsection develops another approach to identify occupations at the 6-digit for those observations that remain without an occupation code at SOC 2020 6-digit level (89,680; 11.3% of job titles).

2.4. Identifying job titles in the job descriptions

Another alternative to identifying the occupational code at the 6-digit level in the vacancy dataset is the job description. This variable is a rich source of information that provides detailed information on the job vacancy. There might be cases where the employer provides clearer information of the job title required in the job description variable rather than in the job title variable. However, processing the job description is challenging because this variable contains a large volume of words describing the job's characteristics. Word-based matching methods might not provide accurate results due to the high number of words in this variable.

Consequently, a procedure that mapped the (exact) words of the Group titles from the SOC 2020 extended framework in the job description provided accurate results at 6-digit level. Although this exercise coded relatively few observations, since it required finding the same words in the job description and in the SOC 2020 extended framework to assign an occupational code, the result complemented the fuzzy merge outputs from the previous step.

At this point, around 11.3% of job titles (89,680) remained without an occupational code, and 20,005 out of 89,680 job titles were coded using the job descriptions. A visual inspection was conducted to verify the accuracy of this mapping. The algorithm generally worked well since it looked for the exact words in the job description and in the SOC 2020 extended framework. However, there were cases where mapping misclassified some job titles. This misclassification mainly occurred when the Group Title variable in the SOC 2020 extended framework was composed of only one word (e.g., "tailor", "model", "producer", etc.). Given the short title, it is more likely that the algorithm considered words in the job description that did not describe the job title required[14].

Table 6 shows the most frequent occupations identified with this method. 2,441 job titles were coded as "Mental health workers", followed by "DevOps engineers" and "Operations managers".

[14] For instance, an employer posted a job vacancy with the following job description: "The role will be client facing, providing independent and expert IT operating ***model*** and sourcing advice to support programmes and projects initially in the public sector…". As can be seen, the job description contains the word "model" but this word does not refer to the occupation "3413/07" "actor, entertainers and presenters"/ "models"

Table 6. Most frequent occupations identified using the job vacancy descriptions

SOC 2020 6-digit	Job titles	Percentage
Mental health workers	2,441	12.2%
DevOps engineers	896	4.5%
Operations managers	873	4.4%
Project managers	853	4.3%
Auditors	470	2.3%
Cover supervisors	448	2.2%
Learning support assistants	391	2.0%
Chief executives	387	1.9%
Business analysts	274	1.4%
Business development managers	272	1.4%

Source: IER-LMI. Own calculations

By applying the word-based matching methods and the mapping of occupational groups in the job descriptions, 719,553 job titles were coded at SOC 2020 6-digit level. This number represents around 91.1% of the total non-duplicated job titles available in the vacancy database. 1,530 occupational groups at 6-digit level are identified.

Table 2 shows an example of the vacancy database after applying the first two steps of this methodology (word-match methods and mapping job titles in the job descriptions). The first two columns of Table 2 indicate the variables from the job portals (job titles, job descriptions, among others). Columns three to six show the variables created after CASCOT being applied. The seventh and eighth columns ("Group title" and "SOC 6-digits") indicate the variables created after the methods being applied. The last variable ("Group") serves to explain the following steps of this methodology.

The group variable indicates the different groups that result after applying the first two steps of this methodology over the vacancy database. Group 1 indicates those observations that were coded with CASCOT and the word-based matching methods find a good match in the SOC 2020 extended framework (either with "jobtitle - GroupTitle" or the "index entry - GroupTitle" merges). This group of observations can be considered coded at 6-digit level SOC 2020 (699,528 observations). Group 2 indicates those observations that were not coded with the word-based matching methods, but the job descriptions provided enough information to determine the occupational code at 6-digit level (20,025 observations). Group 3 represents the observations that the word-based matching methods and the job title mapping in the job description did not code. Observations in group 3 will required further analysis to code them accordingly to SOC 2020 at 6-digit level.

Table 7. Example of the vacancy database after applying CASCOT and word-based matching methods

Job title	Job description	SOC 4-digit label	SOC code 4-digit	Cascot - entry index	Score	Group title	SOC 6-digits	Group
Financial accountant	We are looking for a committed…background in compliance or audit	Chartered and certified accountants	2421	Accountant, financial	84	Financial accountants	2421/02	1
Fundraising director	Fundraising manager bishops gate institute is a home for ideas	Marketing, sales and advertising directors	1132	Director (advertising)	93	Advertising directors	1132/01	1
Manager extraction fossils	As a planning director, you will work closely with director and other members	Production managers and directors in mining and energy	1123	Manager (quarrying and extraction)	53	Managers and directors in the extraction of fusil fuels	1123/04	1
bd manager leading	An regional role arisen leading accountancy firm. business development manager will lead develop business development marketing initiatives.	Sales accounts and business development managers	3556	.	.	Business development managers	3556/02	2
Operational finance leader	Finance location: Dunstable, Bedfordshire. We are currently working remotely and will support	.	.	.	.	.	.	3

Source: IER-LMI. Own calculation

2.5. Looking for skills

The previous steps used the job title variable to make a first codification approach at 6-digit level. However, more detailed information will be needed to code a set of observations that are not coded by the first steps of this methodology.

Specifically, the job vacancy descriptions might use to help to code those “hard-to-code” observations. This variable is an important input to identify labour market requirements such as qualifications, experience and skills requirements etc. The skills requirements might be crucial information for coding the vacancy database at 6-digit level SOC 2020 (see Lima and Bakhshi, 2018; Cárdenas-Rubio, 2020).

This section details how coding “hard-to-code” observations in the UK vacancy base at 6-digit level was undertaken using skills requirements. To do so, it was necessary first to identify the skills demanded by employers. Generally, employers use the job description to list the skills needed for a certain job. However, skills requirements are not usually organised under separated variables nor categorised under the same typology (Cárdenas-Rubio, 2020).

Given that the European Skills, Competences, Qualifications and Occupations (ESCO[15]) is an internationally used and publicly available dictionary, this methodology used ESCO dictionary to identify and standardise the skill information in the vacancy database. This dictionary provides a comprehensive list of 13,485 skills used in the labour market. These skills are grouped into four categories: 1) Transversal – knowledge, skills and competences that are important to a broad range of occupations and sectors. Also known as “core skills”, “basic skills”, etc.; 2) Cross-sector – knowledge, skills and competences that are necessary for different sectors. For instance, knowledge in “mechanics” is relevant for the automotive and textile industries; 3) Sector-specific, skills and competences “specific to one sector but are relevant for more than one occupation within that sectors. E.g., monitoring of livestock is specific for the sector husbandry and breeding”; and 4) Occupation-specific – These skills tend to be used within one occupation or specialism. For instance, knowledge in “bioethics” is a relevant skill for a biomedical scientist (ESCO, 2017, p.6). The key advantage of ESCO over other skills classification (such as the O*NET) is that ESCO is adapted to identify the skills used in the European labour market, while O*NET is a US-based classification.

As explained by Cárdenas-Rubio (2020), the skills pattern identification with ESCO dictionary and a vacancy dataset proceeded as follows:

- Common words (such as prepositions, stop words) were removed, letters are transformed to lower case and words are reduced to their grammatical root in both the ESCO dictionary and in the description of the vacancy database;
- Each word or phrase in the skills dictionary was searched for across each job vacancy advertisement. This exploration of words was encoded into unigram variables (n-gram) (a maximum of 13,485 variables will be created);

[15] Developed by the European Commission (EC).

- Manual checks were undertaken to ensure a certain accuracy level of the skill mapping.

2.6. Nearest neighbour algorithm

Machine learning methods[16] have been recently used as an alternative for coding job titles into occupations at 4-digit level (Gweon et al., 2017; Lima and Bakhshi, 2018; Cárdenas-Rubio, 2020). These methods could potentially be used to help classify the vacancy database into SOC 2020 at 6-digit level, so were tested in this feasibility study. As shown in Table 7, the vacancy observations were categorised into three groups. Group 1 and 2 corresponds to all observation coded at 6-digit level using CASCOT, word-based matching methods, visual inspections and job titles mapping in the job descriptions. This set of observations could be used as the pre-coded database that serves as an example to calibrate a machine learning model.

Most of these machine learning methods use the job titles to learn how to code observations into occupations. However, Lima and Bakhshi (2018) and Cárdenas-Rubio (2020) have demonstrated that the skills variables could be used along with the job titles to improve the machine learning algorithms' results. Particularly, Cárdenas-Rubi (2020) has shown that the "nearest neighbour algorithm" provides consistent results at 4-digit level using both the job titles and skills information. Moreover, this algorithm is relatively easy to interpret and does not require a high computational capacity[17].

In the case of coding the vacancy database to SOC 2020 6-digit level, the nearest neighbour algorithm used the skill patterns to identify the nearest neighbour for each observation. The higher number of skills in common the nearest the neighbour. An observation (job title) might have *n* neighbours with *m* skills in common. The occupational code assigned will correspond to the most frequent neighbour within a neighbourhood with *m* skills in common.

As previously mentioned, the observations in Group 1 and 2 (See Table 7) in the job titles data could be used as the learning database. These data are divided into the training and testing databases to calibrate the neighbour algorithm parameters. Once the algorithm is tested, these parameters are used to code the remaining job titles (Group 3).

Table 8 shows a reduced example of how the nearest neighbour algorithm worked in practice. On the one hand, let's suppose that we have a training database. In this database, four jobs are being demanded and coded: web designer (SOC code

[16] Machine learning can be defined "set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty" (Murphy, 2012, p.1).

[17] Broadly speaking, this algorithm could use the job titles, the skill variables (from the pre-coded database) and an improved version of the cosine similarity score to determine which patterns might indicate a certain occupational group (nearest neighbour) (for more details, see Cárdenas-Rubio, 2020; Gweon et al., 2017).

2141/03 Web designers), graphic designer (2142/99 Graphic and multimedia designers n.e.c.) and jewellery designers (3422/03 Jewellery designers). On the other hand, there is an observation in Group 3 in which the job title is “designer”. With only this input, it is not possible to determine the SOC code at 4- and 6-digit levels. However, the skills information from the job vacancy description could provide extra information to determine the corresponding occupational codes. Columns 6 to 8 indicate the skills demanded. As can be noted, uncoded observation (Group 3) has three neighbours (it shares the “2D painting” with):

- “designer specialist” (6-digit SOC 2020: 2142/99 Graphic and multimedia designers n.e.c.) (second row of Table 8).
- “Graphic designer”, (6-digit SOC 2020: 2142/99 Graphic and multimedia designers n.e.c.) (third row of Table 8).
- “Jewellery designer”, (6-digit SOC 2020: 3422/03 Jewellery designers) (fourth row of Table 8).

Since the more frequent number of nearest neighbours correspond to “2142/99 Graphic and multimedia designers n.e.c.”, this code will be assigned to the Group 3 observation. It is important to note that defining a minimum number of m skills in common is necessary to consider an observation as a neighbour of another observation. However, there is not a universal rule to set up this parameter. This definition is dependent on the data distribution, which is discussed in the following section.

Table 8. Example nearest neighbour algorithm using job titles and skills

Source	Job title					Skills			SOC 4 digit	SOC 6-digit
	Web	Graphic	Jewellery	designer	specialist	ASP.NET	2D painting	cut gem stones		
Training dataset (Group 1 and 2)	1	0	0	1	0	1	0	0	2141 Web design professionals	2141/03 Web designers
	0	1	0	1	1	0	1	0	2142 Graphic and multimedia designers	2142/99 Graphic and multimedia designers n.e.c.
	0	0	0	1	0	0	1	0	2142 Graphic and multimedia designers	2142/99 Graphic and multimedia designers n.e.c.
	0	0	1	1	0	0	1	1	3422 Clothing, fashion and accessories designers	3422/03 Jewellery designers
	0	0	1	1	0	0	0	1	3422 Clothing, fashion and accessories designers	3422/03 Jewellery designers
New record (Group 3)	0	0	0	1	0	0	1	0	2142 Graphic and multimedia designers	2142/99 Graphic and multimedia designers n.e.c.*

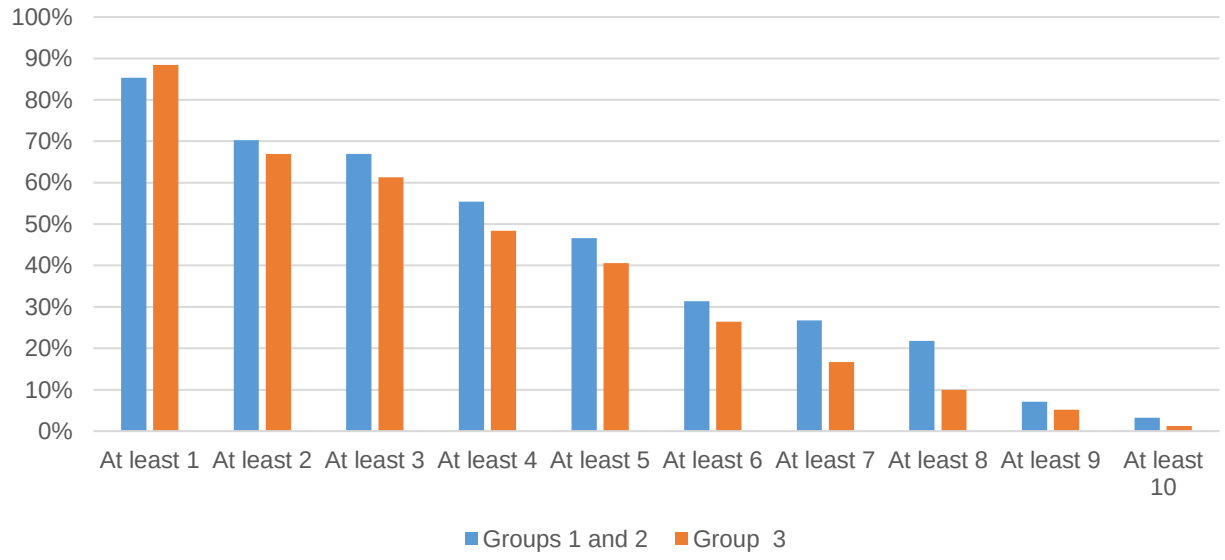
Source: IER-LMI. Own calculations

Alternatively, instead of using the ESCO dictionary to identify skills demanded and use them as input to code vacancy observations to SOC 2020 6-digit level, it is possible to use the "Group title" (label) variable from the SOC 2020 extended framework as a dictionary. These words could be mapped in the job vacancy descriptions to identify neighbours. Within this feasibility study, it was not possible to determine which option, the ESCO skills or the "Group title" variable from the SOC 2020 extended framework, could provide better results. For now, these options are complementary alternatives, and both are explored. Coding the remaining missing occupational values could be part of future work.

2.7. Implementing the nearest neighbour algorithm on the vacancy data

This section shows the main results nearest neighbour algorithm. According to Figure 4 around 85% of observations in the group 1 and 2 mentioned at least one skill in the job description, while this percentage is 88% for group 3. However, the percentage of observations in group 3 that mention at least two skills is lower than groups 1 and 2. This result suggests that observations in group 3 ("hard-to-code") tend to mention less frequently skills requirements than those vacancies in groups 1 and 2. In other words, observations in group 3 tend to provide fewer details about the job (i.e., skills requirements) than groups 1 and 2. Despite this difference between groups, the results suggest that skills information can be identified in the vacancy data and could be used as input for a machine learning model.

Figure 4. Percentage of observations that mention at least one skill in the job description



Source: IER-LMI. Own calculations

For illustration purposes, Table 9 shows the skills most demanded in groups 1-2 and 3. The two skills most required for all the groups are “work in teams” and “communication”. Additionally, there are no considerable differences in the skills demanded between groups. For instance, “customer service”, “lead a team”, “manage work” are in the top 15 of skills most demanded in all groups.

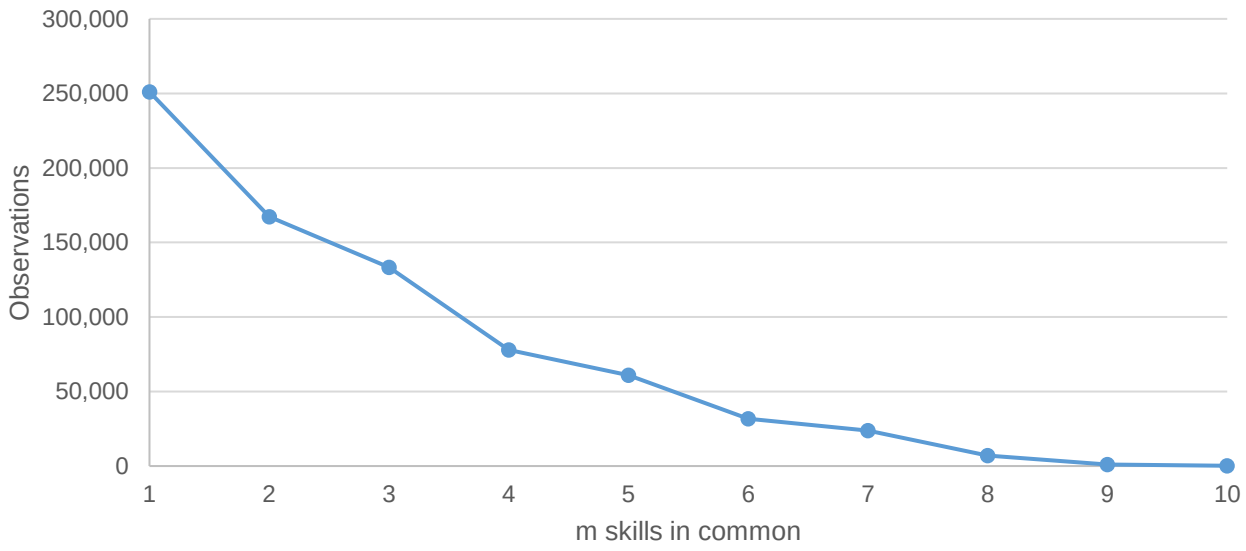
Table 9. Top 15 of skills most demanded across groups in the vacancy dataset

Group 1 and 2			Group 3		
Ranking	**Skills**	**Type**	**Ranking**	**Skills**	**Type**
1	work in teams	transversal	1	work in teams	transversal
2	communication	cross-sector	2	communication	cross-sector
3	lead a team	cross-sector	3	customer service	sector specific
4	manage work	cross-sector	4	manage a team	cross-sector
5	manage a team	cross-sector	5	project management	sector specific
6	provide information	cross-sector	6	lead a team	cross-sector
7	project management	sector specific	7	manage time	transversal
8	financial management	cross-sector	8	business processes	cross-sector
9	customer service	sector specific	9	focus on service	cross-sector
10	support colleagues	transversal	10	plan nursing care	occupation specific
11	manage time	transversal	11	provide information	cross-sector
12	marketing management	cross-sector	12	financial management	cross-sector
13	business processes	cross-sector	13	primary care	cross-sector
14	business analysis	cross-sector	14	manage work	cross-sector
15	work with nursing staff	sector specific	15	process data	cross-sector

Source: IER-LMI. Own calculations

With this skill mapping, the nearest neighbour algorithm was implemented on the vacancy data. The observations in groups 1 and 2 were randomly split in half[18] to create a training and testing dataset. As previously mentioned, it is necessary to define a minimum number of *m* skills in common to consider an observation as a neighbour of another observation. This definition depends on the data distribution. Figure 5 shows the number of observations coded with the nearest neighbour algorithm with different *m* skills in common. For instance, if the *m* parameter is set to 1 (which means that it is only necessary one skill in common to consider an observation as a neighbour of another observation), the number of observations coded in the training database is 251,060 (out of 253,596). With *m* set to a minimum of 2, the number of observations coded in the training database is 167,201. This result is because it is less likely to find observations with at least two skills in common.

Figure 5. Number observations coded by skills in common

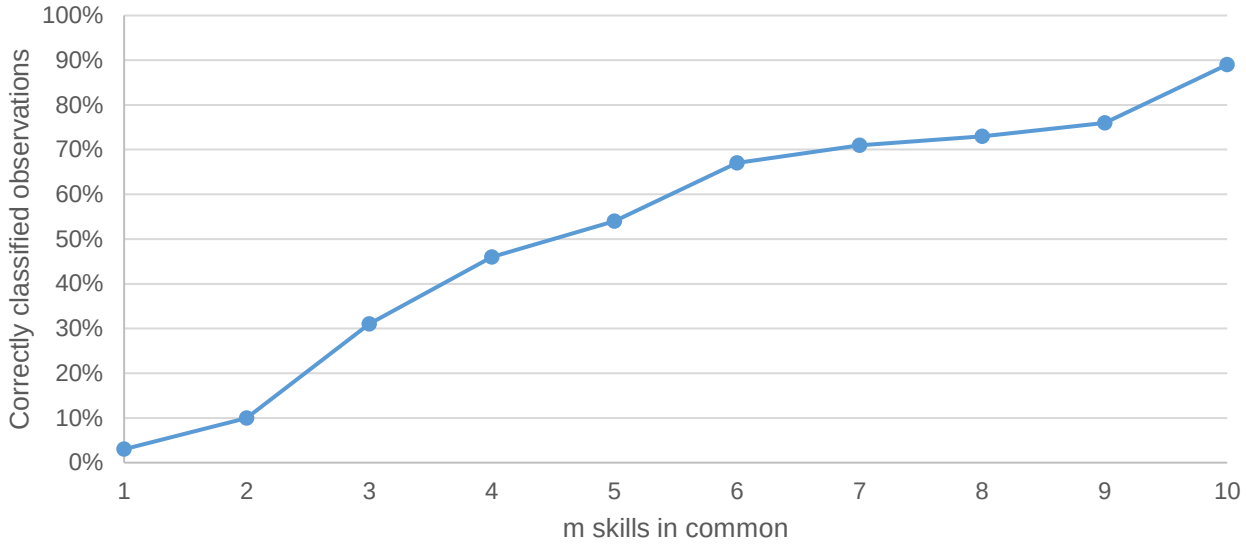


Source: IER-LMI. Own calculations

Figure 6 evaluates the algorithm performance (percentage of observation correctly coded at 6-digit level). There are two facts to highlight from this figure. First, the algorithm performance improves as the number of skills in common increases. The more skills in common, the more likely those observations belong to the same occupational group. Second, the algorithm performance is low. Even in the highest *m* parameter (at least ten skills in common), the algorithm performance is around 89%. This result suggests that the machine learning algorithm was not an appropriate tool to code the remaining uncoded observations.

[18] For simplicity, the database coded at SOC 2020 6-digit level was split in half. So, the training and the testing databases has around 253,596 observations. Depending on the data distribution, other parameters could be used (e.g., 40% - 60%).

Figure 6. Algorithm performance



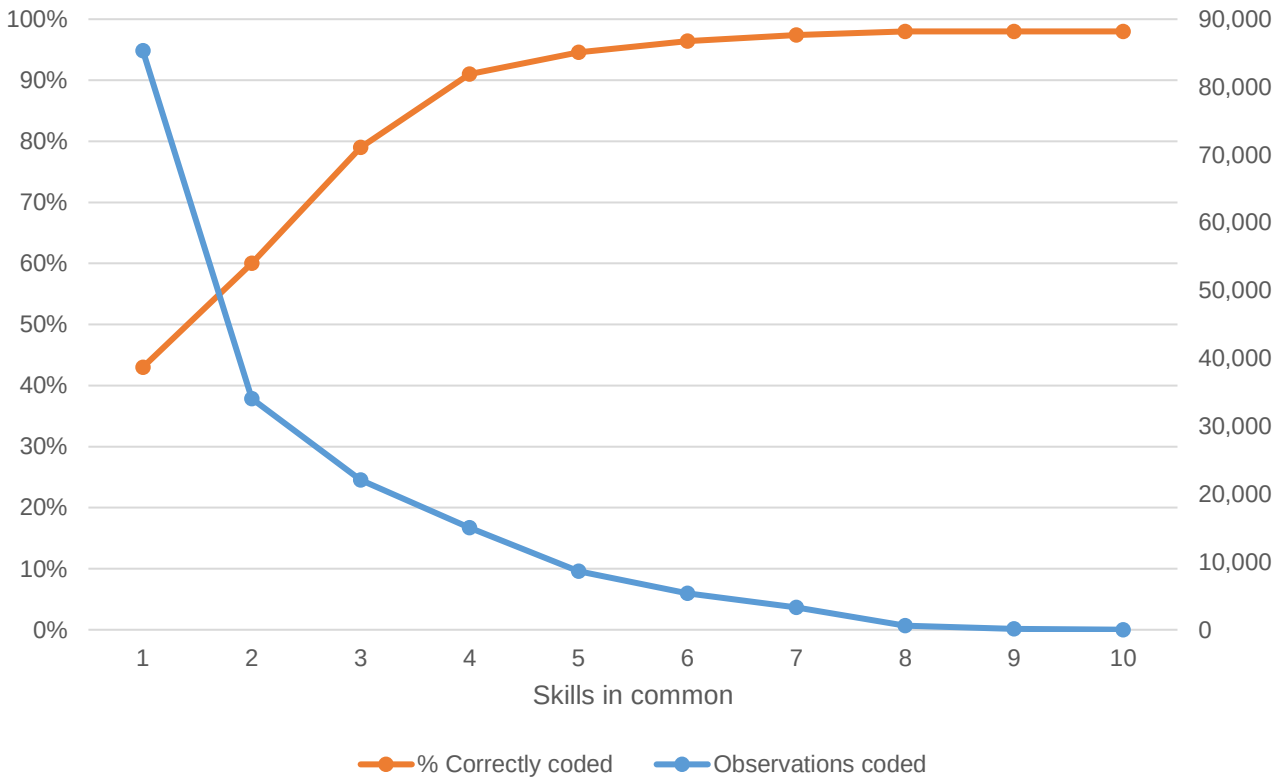
Source: IER-LMI. Own calculations

However, the algorithm performance could be improved by refining the input variables, including occupation-specific skills; this was tested. Specifically, as Table 9 showed, the skills most frequent demanded are cross-sector, sector-specific and transversal skills. These skills are used in a variety of occupations. For instance, “work in teams” is a skill that could be demanded in any occupation, “customer service” is a sector-specific that could be demanded in occupations such as residential childcare worker, receptionist, etc. Those skills, however, do not provide critical information that allows distinguishing between occupational groups. Thus, the type of skills used could explain the low algorithm performance.

Alternatively, only the occupation-specific skills as input for the machine learning algorithm could be used. As previously noted, these skills tend to be used within one occupation or specialism. For instance, knowledge in “bioethics” is a relevant skill for a biomedical scientist. Thus, this set of skills are the ones that could provide critical information on identifying the occupational groups at 6-digit level.

The machine learning algorithm was implemented using only occupation-specific skills as input. Figure 7 shows the number of observations coded and the algorithm performance by skills in common. Two points are worth noting from this figure. First, the number of observations coded is considerably lower than the previous exercise (see Figure 5). This result is because occupation-specific skills are less frequently mentioned in job descriptions. Around 24% of observations required at least one occupation-specific skill. Second, the algorithm performance improves considerably. With at least four skills in common, around 91% of observations in the testing are coded correctly, while this percentage increases to 96% with at least six skills in common. A threshold of at least four occupation-specific skills in common maximise both the number of coded observations and the algorithm performance. Thus, this threshold was used to code the remaining observations.

Figure 7. Number observations coded and the algorithm performance by skills in common



Source: IER-LMI. Own calculations

By implementing this machine learning algorithm, around 9,490 job titles were coded[19]. Consequently, 92.3% (729,043) of job titles were coded at 6-digit level[20]. Finally, a (short) manual codification was conducted. This last step rose the percentage of job titles coded to 93.0%. Table 10 summarises the methodological steps and the number and percentage of observations classified on each stage.

[19] A visual inspection was conducted to verify the algorithm performance. Common mistakes were manually corrected.

[20] Alternatively, it was used the "Group title" (label) variable from the SOC 2020 extended framework as a dictionary (instead of ESCO dictionary) in the machine learning model. These words could be mapped in the job vacancy descriptions to identify neighbours. However, the algorithm performance was relatively low (few observations correctly coded). This low performance is mainly because the "Group title" (label) variable does not provide enough information (words) to distinguish one job title from another job title.

Table 10. Summary of steps and codification results

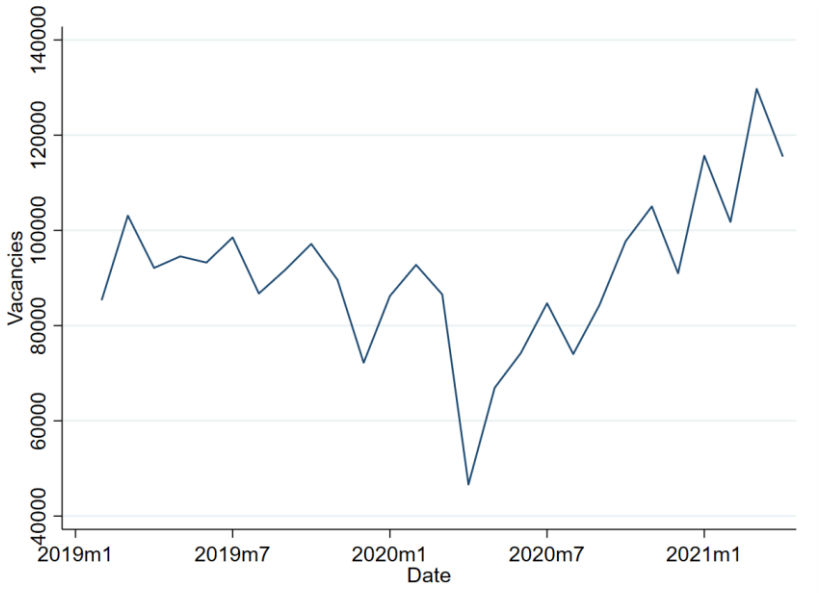
Method	Percentage and number of job titles coded
Word-based matching methods	88.6% (699,528)
Identifying job titles in the job descriptions	2.5% (20,025)
Nearest neighbour algorithm	1.2% (9,490)
Manual classification	0.7% (5,524)
Total job titles coded	**93.0% (734,567)**

Source: IER-LMI. Own calculations

3. Descriptive analysis

This section provides a descriptive analysis of the vacancy database coded at SOC 2020 6-digit level. Specifically, it shows occupations most frequently demanded and analyses data volatility (starting point for evaluating data consistency). Figure 8 indicates the number of vacancies posted in the main UK job portals during the study period (February 2019 – April 2021). The total number of observations (vacancies) in the database are 2,456,395. The monthly average of vacancies is 100,000 (2,456,395 vacancies). The national lockdown in March 2020 dramatically hit the number job vacancies with a decreased in the number of vacancies by 46.1% compared to February 2020.

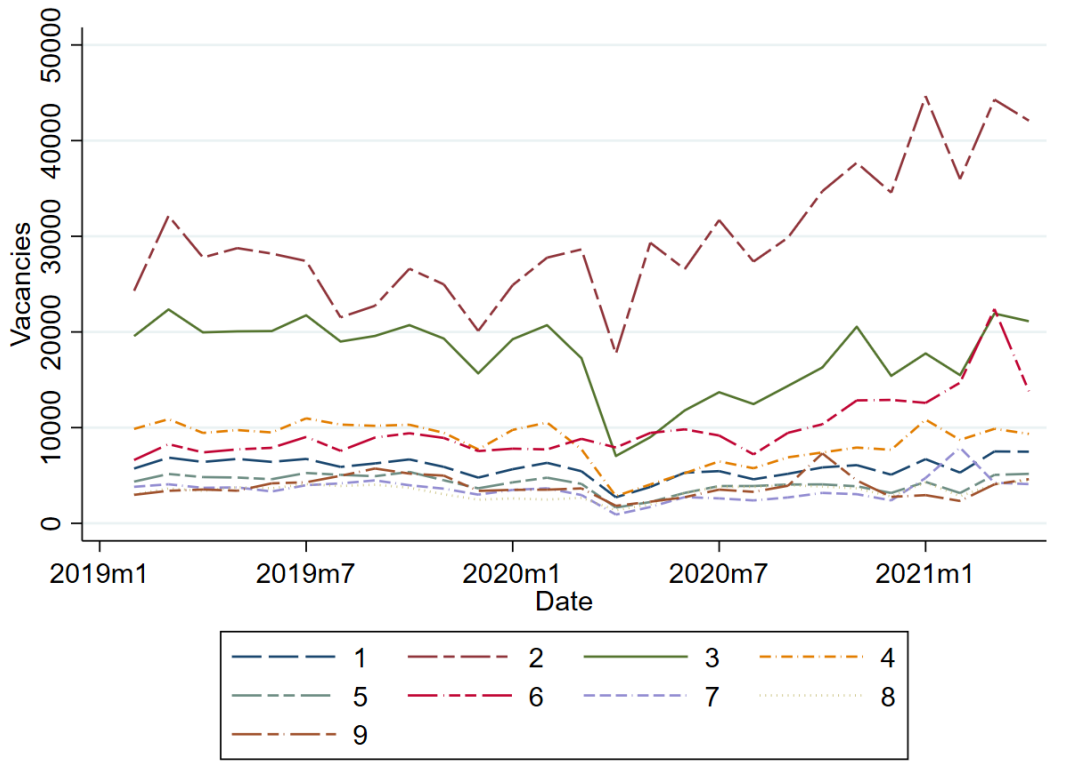
Figure 8. Number of UK vacancies over time (February 2019 – April 2021)



Source: IER-LMI. Own calculations

Figure 9 shows the number of vacancies by major occupational groups (one-digit level SOC-2020). The higher number of vacancies correspond to *Professional*, followed by *Associate professionals* and *Caring, leisure and other service workers*. The data suggest that the occupations most hit by the national lockdown were *Professional*, *Associate professionals* and *Administrative and secretarial workers*.

Figure 9. Number of vacancies for major occupational groups (February 2019 – April 2021)



Source: IER-LMI. Own calculations

1. Managers 2. Professional 3. Associate professionals 4. Administrative and secretarial workers 5. Skilled trades workers 6. Caring, leisure and other service workers 7. Sales and customer service workers 8. Process, plant and machine operatives 9. Elementary occupations

The total number of vacancies coded at 6-digit level is 2,325,671. This figure indicates that the percentage of observations without an occupational code at 6-digit level is relatively low (around 5.3%). The total number of occupational groups identified at 6-digit level is 1,530[21]. Figure 10 depicts the job vacancies distribution by SOC 2020 6-digit level. The most frequent demanded occupations are *Other nursing professionals* (2237/00), *Software developers* (2134/03), *Care workers and home carers n.e.c.* (6135/99), *Community support workers* (6135/01), *Primary education teaching professionals* (2314/00), 9252/00 *Warehouse operatives*, among others. Yet, this figure reveals two important points: the observations are

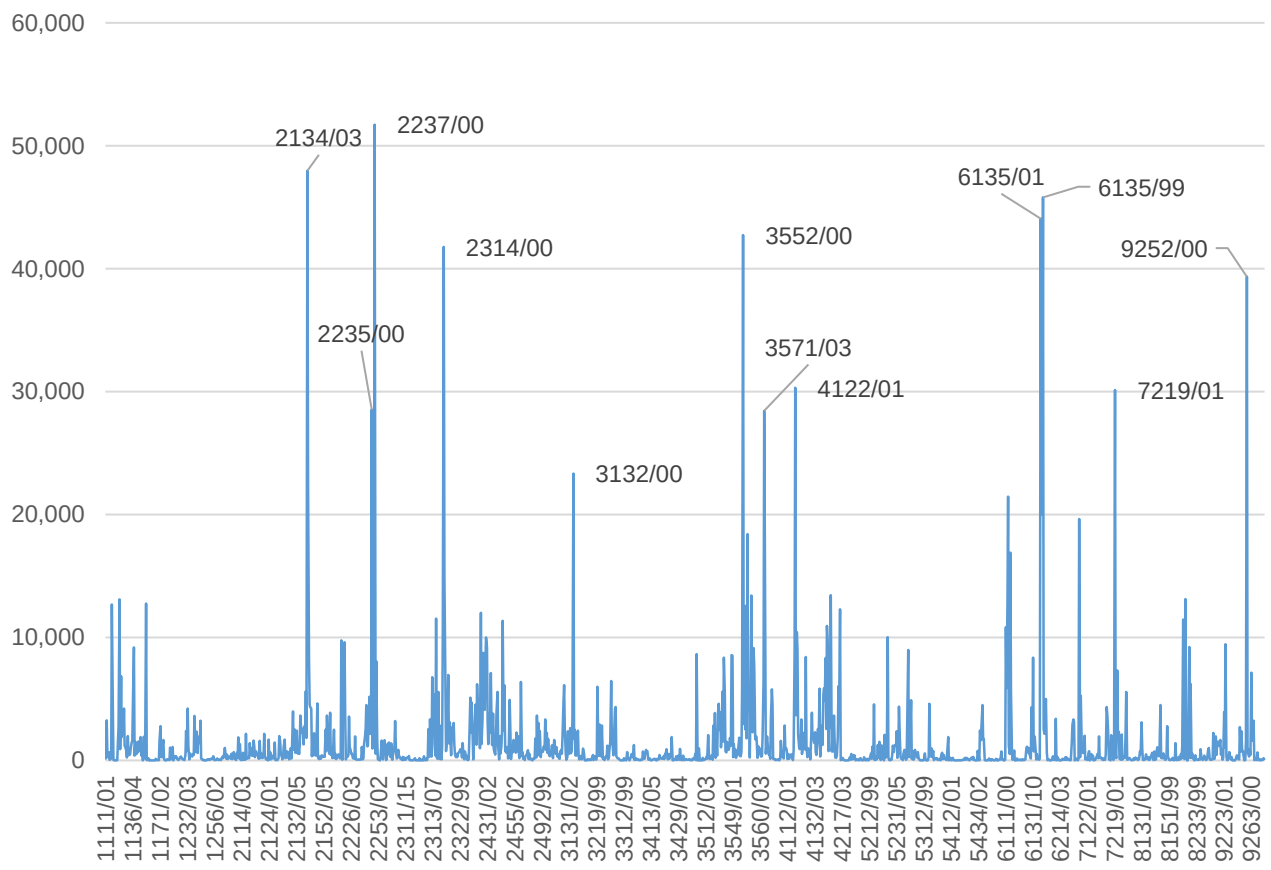
[21] The SOC 2020 extended framework is composed of 1,549 occupational groups.

distributed across different occupational groups; and the categories “not elsewhere classified” (/99) or “other” tend to be overrepresented. For instance, the “other nursing professionals (2237/00)” is the larger group in the UK vacancy database (around 51,700 vacancies). Nevertheless, the high frequency of this group is because this category includes job titles that could not be coded in the other occupational groups or job titles that did not provide enough information to be coded in other occupational groups.

In this example, it is common to find in the vacancy database job titles such as “registered general nurse”, “staff nurse”, “nurse”, etc. These job titles (and the job descriptions) do not provide enough or clear information about the type of nurse required (e.g., Mental health nurses - 2235/00, Neonatal nurses - 2236/01, Paediatric nurses - 2236/02, etc.). Thus, given the lack of detailed information, these job titles are coded “other nursing professionals (2237/00)”.

A similar case occurs with the Care workers and home carers n.e.c. (6135/99)” occupational group. It is common to find job titles such as “care assistant”, “personal care”, “care worker”, etc. These job titles (and the job descriptions) do not provide enough information to determine which specific type of care worker is being demanded (e.g., “Community support workers - 6135/01”, “Domiciliary care workers - 6135/02”, Residential care workers - 6135/03”, etc.).

Figure 10. Job vacancies distribution by SOC 2020 6-digit level



Source: IER-LMI. Own calculations

4. Testing the consistency of the vacancy database coded at 6-digit level SOC 2020

The section of the report details the process of testing and evaluating the consistency of the vacancy database at SOC 2020 6-digit level. As Cárdenas-Rubio (2020) and Kureková et al. (2014) note, it is difficult to test the vacancy data representativeness and consistency since (usually) countries do not have a vacancy census that supplies the total number of vacancies (the statistical universe), which can be used as a comparison point. Cárdenas-Rubio (2020) and Štefánik (2012) have conducted correlation tests with labour supply data to provide insights into the online job portals representativeness. Despite tests limitations, these authors conclude that online job portals tend to capture well the labour dynamics of a considerable set of formal, non-agricultural, non-governmental, non-military and non-self-employed ("business owners") occupations.

However, these tests have been made at ISCO 4-digit level. Currently, there has been no representative tests conducted and published at anything lower than 4-digit level. Testing data representativeness at a more disaggregated level is even more challenging since it is also necessary to have a vacancy census disaggregated at a lower level (i.e., a fifth level). Consequently, the last step of this feasibility study focused on testing the vacancy data volatility rather than the representativeness of these data[22].

The vacancy data volatility was used to determine the occupational groups, which show a relatively stable number of vacancies posted over time. The results provide an indication that data from those "stable" groups could be used to analyse mid and (eventually) long-term vacancy trends. For instance, a "stable" vacancy series at 6-digit level with a positive trend could indicate that type of job is increasingly demanded. High-volatile occupational series would suggest that the data could not be used for analysing labour market trends.

There are different approaches to measure data volatility (e.g., Standard deviation, variance, ARCH models etc.). However, details on which method better suits the vacancy database depend on results from the previous steps. Parameters to be considered when determining volatility would include:

- The proportion of missing values in the SOC 6-digit variable.
- The number of observations by occupational groups. Low-frequency occupations might have a higher volatility coefficient because they have a low share of the labour market rather than the vacancy database does not collect sufficient information.
- Time, as mentioned in the introduction, the LMI for All for has scraped data from job portals since February 2019. However, Covid-19 effects in March 2020 might have affected time series stability. An analysis will be conducted to determine the period to be considered for the volatility estimations.

4.1. The proportion of missing

As previously mentioned, around 5.3% of vacancies (130,556) have missing values in the variable SOC 2020 at 6-digit level. This result suggests that the missing values are not a major issue for analysing the data volatility at 6-digit level. Moreover, a visual inspection of

[22] Data volatility refers to the rate of change in the number of vacancies (at SOC 2020 6-digit level) posted over a period.

those vacancies without occupational code reveals that the uncoded observations correspond to a wide range of job titles such as “teacher”, “coordinator”, “manager”, “specialist”, etc. (Figure 11). Given that the uncoded job titles are not concentrated in a certain group of jobs, it is unlikely the analysis will be affected by not considering those missing observations.

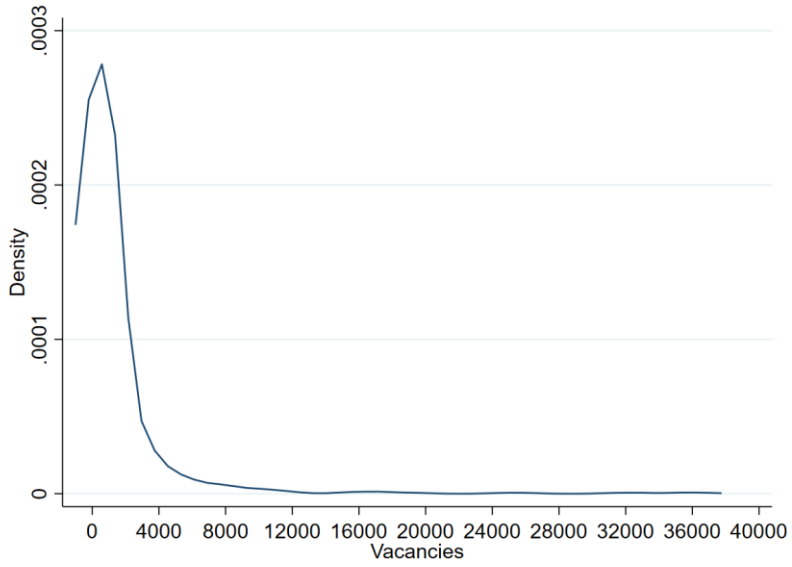
Figure 11. Representation of uncoded job titles

Source: IER-LMI. Own calculations

4.2. The number of observations by occupational groups

Figure 12 shows the distribution of vacancies by occupational groups at 6-digit level. The distribution analysis shows that 25% of the occupational groups (out of 1,530) in the vacancy database contains around 61 observations, while 50% (median) has at least 251 vacancies. There are relatively few occupational groups with more than 4,000 observations. Consequently, the distribution of vacancies by occupational groups at 6-digit level is right-skewed. There is a relatively high number of occupations with few vacancies, and a small number of occupations highly demanded. Occupations with a relatively low frequency might not represent well the corresponding occupational labour market. There is no rule on how to determine the low-frequency threshold. Figure 12 suggests that the first quartile could be used as a reasonable threshold to identify low-frequent occupations. Occupations with a lower frequency of 61 observations in the vacancy database might not represent the labour market properly. Occupations such as Crofters, Poultry dressers, Bowen therapists, Senior police officers n.e.c., Chief superintendent, Homeopaths etc, are in this group.

Figure 12. Vacancy density



Source: IER-LMI. Own calculations

4.3. Time

As shown in Figure 8, the social distance measures announced in March 2020 and other restrictions implemented as a result of the pandemic had a considerable impact on the number of vacancies available in the UK. The figure suggests that the highest impact of these measures took place between March 2020 to October 2020. During this period, the number of vacancies dramatically decreased and then increased during this period. Consequently, the volatility of the vacancy time series is substantially affected by this exogenous shock. Thus, the data from that period was not considered for the volatility analysis since the dramatic decrease/increase in the number of vacancies was due to an extraordinary shock.

4.4. Vacancy volatility test

Given the above considerations, the volatility analysis was conducted on 1,178 occupational groups for the period February 2019 - January 2020. Therefore, a total number of 12 periods were available for the volatility test. Due to the relatively short period and the frequency of the occupational groups, the test used was the coefficient of variation (CV)[23]. This well-known test measures the dispersion of data points in a data series around the mean. It is frequently used for comparing the level of variation from one data series to another. The higher the coefficient of variation, the higher heterogeneity of the values of the variable (volatility). Unlike other tests such as the deviation standard or the variance, the CV is a dimensionless number (i.e., allows comparisons between data series with different units or widely different means). As shown in the previous section, the average of vacancies by occupational groups

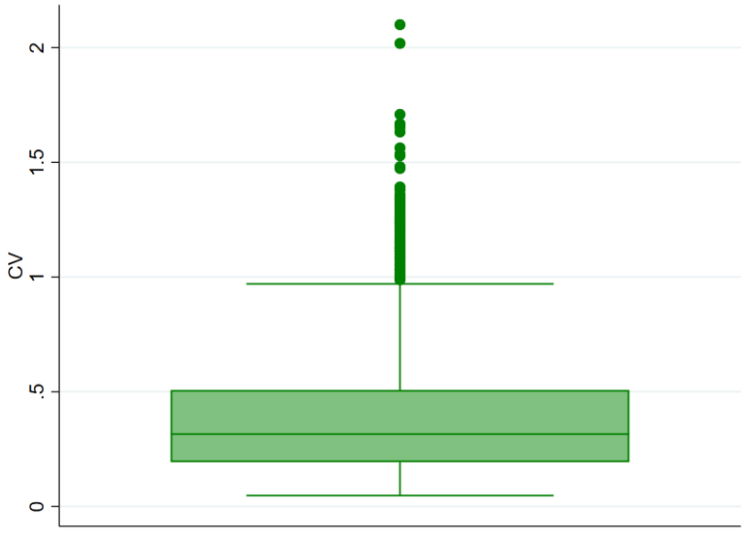
[23] Alternatively, an exercise for February 2019 - April 2021 (excluding March 2020 to October 2020) was conducted to calculate de CV. However, the results suggest that the social distancing measures considerably affects the CV estimations. Occupations more exposed by the pandemic outbreak (such as waitress, nurse staff, pharmacists etc.) tend to show a relatively high CV. In most cases, these high CV coefficients are due to the unpreceded pandemic effects rather than limitations of the vacancy database.

widely differs from each other. Thus, the CV is an adequate tool to measure and comparing data volatility in the UK vacancy database.

Figure 13 shows the box plot of the CV. The median of the CV is 0.333 and the third quartile is 0.473. This figure indicates that a considerable set of CV coefficients tend to be concentrated to the left of the distribution. However, some occupational groups have a relatively high coefficient of variation. Despite there is no general rule on how to determine the CV outliers, the results from Figure 13 could be used to set up rules to analyse the vacancy database at SOC 6-digit level adequately:

- Those occupational categories with a CV lower than the first quartile (0.212) are precise and could be used for analysing labour demand trends.
- Those occupations with a CV higher than the first quartile and lower than median (0.333) has an acceptable precision.
- Those occupations with a CV higher than the median and lower than third quartile (0.473) are less precise therefore these data should be used with caution.
- Occupational groups with a CV higher than the third quartile should not be used for analysing labour demand trends. At most, these data could be used for identifying possible emerging patterns from the job descriptions (e.g., new skill requirements, contracts arrangements, etc.).

Figure 13. Coefficient of variation distribution



Source: IER-LMI. Own calculations

Based on the above rules, the occupational categories were grouped into four groups, each of them composed of 295 categories (see Technical Appendix. Table A3). This list provides an indication of the occupations to which the vacancy database could provide more/less consistent results. Nevertheless, this test is not final proof to discard the analysis of certain occupations using the vacancy database.

5. Conclusions

A frequent claim from labour market information users (e.g., educational and training providers, policymakers, etc.) is that the lack of desegregated labour market information hampers the design of public policies and the provision of training and educational programmes aligned with the employers' requirements. This feasibility study has assessed whether sources of information such as vacancies from job portals can be classified according to the relatively new 6-digit (fifth level) SOC classification developed by the ONS. This fifth level identifies in more detail the occupational groups that share similar tasks. Thus, classifying vacancies to this disaggregated level could inform labour market information users about the current labour demand requirements in the UK.

The vacancy data used to conduct this feasibility study was collected from February 2019 to April 2021 by the IER funded as part of the LMI for All service by the Department for Education. These data are composed of 2,449,076 observations. As part of the LMI for All project these data have organised by IER using standardised variables, such as occupations (at SOC 2010 4-digit level), skills and qualification requirements, wages, etc. For this feasibility study, Vacancy job titles were coded to SOC 6-digit level using CASCOT. This tested tool is a key pillar in this study since it helped to classify the job titles to SOC 2020 at 4-digit with a high precision level for the LMI for All service.

Based on CASCOT outputs, this feasibility study used the job titles and the job descriptions to develop word-based and machine learning algorithms to identify the occupational codes at 6-digit level for vacancy. Along with visual inspections in different stages, the percentage of vacancies coded at 6-digit level is 94.7%.

Nevertheless, the high coding rate does not imply that the vacancy data provides consistent results for analysing the labour market requirements at SOC 6-digit level. The vacancy data might collect relatively few observations for a set of occupations. For this reason, statistical tests were necessary to validate the consistency of the vacancy database. Despite the methodological and data limitations, a volatility test to provide an indication of which the vacancy database could provide more/less consistent results was applied. This demonstrated that:

- by combining different techniques and variables inputs, it is possible to code a considerable proportion of the vacancy database at SOC 6-digit level with a high accuracy level;
- there is a considerable share of occupations with a relatively high frequency and low coefficient of variation. This result suggests that the vacancy database could be used to describe trends and labour market requirements for a considerable set of occupations at SOC 2020 6-digit level.

With additional resources further work on this field could be undertaken to improve the occupational coding rate from 94.7%. Firstly, further work could focus on: checking in more detail coding accuracy. Given that visual inspection is a time-consuming task and the large number of occupational groups at SOC 6-digit level, some of those occupations might be misclassified. Second, further work could be undertaken to increase the occupational coding rate. There are uncoded job titles that could be classified by identifying different word patterns in the job descriptions (e.g., certifications). Thirdly, the frequency of the categories “not

elsewhere classified" (/99) or "other" could be reduced. Like the previous point, different word patterns in the job vacancy descriptions could provide more precise information of occupational category. Fourthly, an evaluation of whether the uncoded job titles correspond to new job titles and potentially new occupations could be undertaken. Finally, more volatility and (if possible) representative tests are needed to identify the strengths and weakness of the vacancy database desegregated at SOC 2020 6-digit level (e.g., internal and external consistency tests, to what extent the wages information could be used at SOC 2020 6-digit level, etc.).

Overall, this feasibility study has proven that data collected from job vacancies can provide a source of information at SOC 6-digit level. Using these data and further analyses, it would be possible to examine labour demand, wage levels, and education and skills requirements (and trends) at this occupational level. With this disaggregated information, education and training providers could have a high-quality source of updated labour demand information to align their programmes with the employers' needs.

References

Barnes, Sally-Anne, Bimrose, Jenny, Cárdenas Rubio, Jeisson, Wilson, Rob, Owen, David, Hogarth, Terence, Bosworth, Luke, Day, Rosie, Attwell, Graham, and Rustemeier, Philipp. (2021). *Enhancing a labour market information database: LMI for All Stakeholder Engagement and Usage, Data and Technical Developments (2019-2020)*(Research report RR1096). London: Department for Education.

Cárdenas-Rubio, Jeisson 2020. "*A web-based approach to measure skill mismatches and skills profiles for a developing country: the case of Colombia*". Bogotá; D.C.: Editorial Universidad del Rosario.

ESCO 2017. *ESCO Handbook. European Skills, Competences, Qualifications and Occupations*. Brussels: European Union.

GSS (Government Statistical Service) 2021. "Standard Occupational Classifications (SOC) Extension Project." Retrieved April 10, 2020 (https://gss.civilservice.gov.uk/user-facing-pages/standard-occupational-classifications-soc/).

Gweon, Hyukjun, Matthias Schonlau, Lars Kaczmirek, Michael Blohm, and Stefan Steiner. 2017. "Three Methods for Occupation Coding Based on Statistical Learning." *Journal of Official Statistics* 33(1):101–22.

IER. 2018. "Cascot International." Institute for Employment Research, University of Warwick. Retrieved September 20, 2018 (https://warwick.ac.uk/fac/soc/ier/software/cascot/internat/).

Jones, Rob, and Peter Elias. 2004. *CASCOT: Computer-Assisted Structured Coding Tool.* Institute for Employment Research, University of Warwick.

Kureková, Lucia Mytna, Miroslav Beblavy, and Anna-Elisabeth Thum. 2014. *Using Internet Data to Analyse the Labour Market: A Methodological Enquiry*. IZA Discussion Paper 8555. Bonn.

Lima, Antonio, and Hasan Bakhshi. 2018. *Classifying Occupations Using Web-Based Job Advertisements: An Application to STEM and Creative Occupations*. ESCoE Discussion Paper 2018-08. London: Economic Statistics Centre of Excellence.

ONS. 2021. "Classifying the Standard Occupational Classification 2020 (SOC 2020) to the International Standard Classification of Occupations (ISCO-08)" Retrieved April 10, 2020 (https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020/classifyingthestandardoccupationalclassification2020soc2020totheinternationalstandardclassificationofoccupationsisco08).

Štefánik, Miroslav. 2012. "Internet Job Search Data as a Possible Source of Information on Skills Demand (with Results for Slovak University Graduates)." *Building on Skills Forecasts—Comparing Methods and Applications* 246.

Wageindicator.org. 2009. EurOccupations: CASCOT Software for Coding Job Titles. European Policy Brief No. 3. Accessed May 25, 2018. https://wageindicator.org/copy_of_documents/policy-briefs/European-Policy-Brief-no-3-CASCOT- coding-

program-EUROCCUPATIONS-20100104.pdf.

Winkler, W. E. 1999. The state of record linkage and current research problems. Statistics of Income Division, Internal Revenue Service Publication R99/04. Available from http://www.census.gov/srd/www/byname.html

Technical Appendix

Table A1: Vacancy distribution at SOC 2020 4-digit level

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
NA	NA	89,680	11.4%
2134	Programmers and software development professionals	30,793	3.9%
2237	Other nursing professionals	17,516	2.2%
4122	Bookkeepers, payroll managers and wages clerks	14,797	1.9%
3571	Human resources and industrial relations officers	14,127	1.8%
4159	Other administrative occupations n.e.c.	13,835	1.8%
3556	Sales accounts and business development managers	13,549	1.7%
3552	Business sales executives	13,405	1.7%
6135	Care workers and home carers	13,364	1.7%
3554	Marketing associate professionals	12,080	1.5%
7219	Customer service occupations n.e.c.	9,772	1.2%
2314	Primary education teaching professionals	9,682	1.2%
2422	Finance and investment analysts and advisers	9,362	1.2%
2440	Business and financial project management professionals	8,905	1.1%
9252	Warehouse operatives	8,157	1.0%
6112	Teaching assistants	8,127	1.0%
3132	IT user support technicians	8,106	1.0%
2133	IT business analysts, architects and systems designers	7,868	1.0%
2212	Specialist medical practitioners	7,775	1.0%
3534	Financial accounts managers	7,347	0.9%
2313	Secondary education teaching professionals	6,627	0.8%
5223	Metal working production and maintenance fitters	6,431	0.8%
1131	Financial managers and directors	6,238	0.8%
3574	Other vocational and industrial trainers	6,205	0.8%
6113	Educational support assistants	5,971	0.8%
1136	Human resource managers and directors	5,724	0.7%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
2431	Management consultants and business analysts	5,675	0.7%
2461	Social workers	5,656	0.7%
4129	Financial administrative occupations n.e.c.	5,639	0.7%
2434	Business and related research professionals	5,084	0.6%
2132	IT managers	5,034	0.6%
3119	Science, engineering and production technicians n.e.c.	4,985	0.6%
4131	Records clerks and assistants	4,858	0.6%
2412	Solicitors and lawyers	4,618	0.6%
9223	Cleaners and domestics	4,340	0.5%
2315	Nursery education teaching professionals	4,301	0.5%
2311	Higher education teaching professionals	4,221	0.5%
2319	Teaching professionals n.e.c.	4,219	0.5%
6131	Nursing auxiliaries and assistants	4,197	0.5%
1132	Marketing, sales and advertising directors	4,193	0.5%
2129	Engineering professionals n.e.c.	4,071	0.5%
4215	Personal assistants and other secretaries	4,056	0.5%
1251	Property, housing and estate managers	3,999	0.5%
1121	Production managers and directors in manufacturing	3,883	0.5%
3549	Business associate professionals n.e.c.	3,882	0.5%
3229	Welfare and housing associate professionals n.e.c.	3,833	0.5%
2316	Special needs education teaching professionals	3,827	0.5%
2421	Chartered and certified accountants	3,822	0.5%
3533	Financial and accounting technicians	3,717	0.5%
8219	Road transport drivers n.e.c.	3,689	0.5%
1139	Functional managers and directors n.e.c.	3,641	0.5%
2259	Other health professionals n.e.c.	3,633	0.5%
2432	Marketing and commercial managers	3,592	0.5%
8211	Large goods vehicle drivers	3,527	0.4%
7111	Sales and retail assistants	3,509	0.4%
3131	IT operations technicians	3,499	0.4%
2234	Nurse practitioners	3,466	0.4%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
5434	Chefs	3,456	0.4%
2482	Quality assurance and regulatory professionals	3,451	0.4%
8214	Delivery drivers and couriers	3,406	0.4%
5231	Vehicle technicians, mechanics and electricians	3,333	0.4%
4216	Receptionists	3,303	0.4%
2419	Legal professionals n.e.c.	3,297	0.4%
3543	Project support officers	3,175	0.4%
2423	Taxation experts	3,142	0.4%
6111	Early education and childcare assistants	3,082	0.4%
2433	Actuaries, economists and statisticians	2,985	0.4%
3544	Data analysts	2,907	0.4%
5242	Telecoms and related network installers and repairers	2,860	0.4%
3541	Estimators, valuers and assessors	2,839	0.4%
4134	Transport and distribution clerks and assistants	2,824	0.4%
3520	Legal associate professionals	2,809	0.4%
2235	Mental health nurses	2,808	0.4%
2232	Community nurses	2,805	0.4%
2233	Specialist nurses	2,723	0.3%
3582	Health and safety managers and officers	2,706	0.3%
3557	Conference and exhibition managers and organisers	2,649	0.3%
2222	Occupational therapists	2,643	0.3%
2161	Research and development (R&D) managers	2,630	0.3%
8222	Fork-lift truck drivers	2,597	0.3%
2221	Physiotherapists	2,581	0.3%
4133	Stock control clerks and assistants	2,573	0.3%
3133	Database administrators and web content technicians	2,519	0.3%
3555	Estate agents and auctioneers	2,449	0.3%
2493	Public relations professionals	2,392	0.3%
2211	Generalist medical practitioners	2,372	0.3%
4151	Sales administrators	2,362	0.3%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
9211	Postal workers, mail sorters and messengers	2,348	0.3%
6136	Senior care workers	2,324	0.3%
3221	Youth and community workers	2,267	0.3%
2115	Social and humanities scientists	2,238	0.3%
2251	Pharmacists	2,191	0.3%
2317	Teachers of English as a Foreign Language	2,179	0.3%
4132	Pensions and insurance clerks and assistants	2,119	0.3%
1232	Residential, day and domiciliary care managers and proprietors	2,109	0.3%
3551	Buyers and procurement officers	2,072	0.3%
4136	Human resources administrative occupations	2,026	0.3%
2142	Graphic and multimedia designers	2,001	0.3%
2454	Chartered surveyors	1,975	0.3%
8143	Routine inspectors and testers	1,926	0.2%
4143	Customer service managers	1,900	0.2%
2139	Information technology professionals n.e.c.	1,897	0.2%
4121	Credit controllers	1,885	0.2%
4141	Office managers	1,880	0.2%
3412	Authors, writers and translators	1,869	0.2%
7113	Telephone salespersons	1,863	0.2%
3531	Brokers	1,862	0.2%
1242	Managers in storage and warehousing	1,861	0.2%
2131	IT project managers	1,810	0.2%
5221	Metal machining setters and setter-operators	1,808	0.2%
4142	Office supervisors	1,805	0.2%
2135	Cyber security professionals	1,791	0.2%
2226	Other psychologists	1,762	0.2%
3113	Engineering technicians	1,754	0.2%
4123	Bank and post office clerks	1,741	0.2%
3222	Child and early years officers	1,693	0.2%
2136	IT quality and testing professionals	1,686	0.2%
1171	Health services and public health managers and directors	1,670	0.2%
8120	Metal working machine operatives	1,669	0.2%
5241	Electricians and electrical fitters	1,661	0.2%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
2494	Advertising accounts managers and creative directors	1,623	0.2%
2121	Civil engineers	1,608	0.2%
3532	Insurance underwriters	1,598	0.2%
2254	Medical radiographers	1,596	0.2%
7211	Call and contact centre occupations	1,591	0.2%
2137	IT network professionals	1,587	0.2%
2113	Biochemists and biomedical scientists	1,581	0.2%
3319	Protective service associate professionals n.e.c.	1,556	0.2%
5315	Plumbers & heating and ventilating installers and repairers	1,539	0.2%
1122	Production managers and directors in construction	1,535	0.2%
9263	Kitchen and catering assistants	1,517	0.2%
3111	Laboratory technicians	1,499	0.2%
7129	Sales related occupations n.e.c.	1,494	0.2%
3232	Early education and childcare practitioners	1,492	0.2%
8149	Assemblers and routine operatives n.e.c.	1,474	0.2%
3120	CAD, drawing and architectural technicians	1,436	0.2%
5246	Electrical service and maintenance mechanics and repairers	1,434	0.2%
4152	Data entry administrators	1,429	0.2%
1150	Managers and directors in retail and wholesale	1,415	0.2%
3213	Medical and dental technicians	1,342	0.2%
2119	Natural and social science professionals n.e.c.	1,329	0.2%
1134	Purchasing managers and directors	1,314	0.2%
2236	Childrens nurses	1,294	0.2%
6240	Cleaning and housekeeping managers and supervisors	1,285	0.2%
3572	Careers advisers and vocational guidance specialists	1,265	0.2%
3231	Higher level teaching assistants	1,261	0.2%
2112	Biological scientists	1,261	0.2%
4212	Legal secretaries	1,254	0.2%
9231	Security guards and related occupations	1,236	0.2%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
1241	Managers in transport and distribution	1,220	0.2%
2223	Speech and language therapists	1,196	0.2%
2321	Head teachers and principals	1,191	0.2%
2439	Business, research and administrative professionals n.e.c.	1,184	0.2%
3560	Public services associate professionals	1,166	0.1%
6212	Travel agents	1,163	0.1%
2469	Welfare professionals n.e.c.	1,148	0.1%
2141	Web design professionals	1,146	0.1%
5244	Computer system and equipment installers and servicers	1,144	0.1%
8113	Chemical and related process operatives	1,127	0.1%
3223	Housing officers	1,121	0.1%
1111	Chief executives and senior officials	1,118	0.1%
9132	Packers, bottlers, canners and fillers	1,112	0.1%
2122	Mechanical engineers	1,095	0.1%
4111	National government administrative occupations	1,065	0.1%
3212	Pharmaceutical technicians	1,057	0.1%
4211	Medical secretaries	1,056	0.1%
3416	Arts officers, producers and directors	1,050	0.1%
2452	Chartered architectural technologists, planning officers and consultants	1,035	0.1%
1137	Information technology directors	1,019	0.1%
6232	Caretakers	1,017	0.1%
3553	Merchandisers	1,015	0.1%
5245	Security system installers and repairers	1,003	0.1%
5316	Carpenters and joiners	1,000	0.1%
1259	Managers and proprietors in other services n.e.c.	995	0.1%
2453	Quantity surveyors	982	0.1%
2152	Environment professionals	981	0.1%
3115	Quality assurance technicians	978	0.1%
8159	Construction operatives n.e.c.	965	0.1%
2225	Clinical psychologists	964	0.1%
8112	Textile process operatives	957	0.1%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
2127	Engineering project managers and project engineers	955	0.1%
5249	Electrical and electronic trades n.e.c.	954	0.1%
2491	Newspaper and periodical editors	945	0.1%
2322	Education managers	944	0.1%
6231	Housekeepers and related occupations	927	0.1%
2114	Physical scientists	924	0.1%
2125	Production and process engineers	917	0.1%
5319	Construction and building trades n.e.c.	916	0.1%
5330	Construction and building trades supervisors	881	0.1%
7131	Shopkeepers and owners - retail and wholesale	872	0.1%
2455	Construction project managers and related professionals	869	0.1%
3417	Photographers, audio-visual and broadcasting equipment operators	862	0.1%
2492	Newspaper and periodical journalists and reporters	859	0.1%
6222	Beauticians and related occupations	833	0.1%
7220	Customer service supervisors	828	0.1%
5250	Skilled metal, electrical and electronic trades supervisors	824	0.1%
2229	Therapy professionals n.e.c.	809	0.1%
7122	Debt, rent and other cash collectors	806	0.1%
8111	Food, drink and tobacco process operatives	806	0.1%
2329	Other educational professionals n.e.c.	802	0.1%
2481	Quality control and planning engineers	800	0.1%
9139	Elementary process plant occupations n.e.c.	796	0.1%
3219	Health associate professionals n.e.c.	776	0.1%
2123	Electrical engineers	762	0.1%
7132	Sales supervisors - retail and wholesale	759	0.1%
1224	Leisure and sports managers	745	0.1%
9219	Elementary administration occupations n.e.c.	733	0.1%
2451	Architects	732	0.1%
3429	Design occupations n.e.c.	731	0.1%
1133	Public relations and communications directors	722	0.1%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
3432	Sports coaches, instructors and officials	716	0.1%
3116	Planning, process and production technicians	712	0.1%
1243	Managers in logistics	707	0.1%
9232	School midday and crossing patrol occupations	706	0.1%
2312	Further education teaching professionals	703	0.1%
3542	Importers and exporters	702	0.1%
4213	School secretaries	689	0.1%
7115	Vehicle and parts salespersons and advisers	675	0.1%
7114	Pharmacy and optical dispensing assistants	663	0.1%
9129	Elementary construction occupations n.e.c.	662	0.1%
4112	Local government administrative occupations	662	0.1%
2111	Chemical scientists	638	0.1%
2162	Other researchers, unspecified discipline	628	0.1%
5213	Welding trades	628	0.1%
7213	Communication operators	624	0.1%
5436	Catering and bar managers	618	0.1%
1222	Restaurant and catering establishment managers and proprietors	614	0.1%
2231	Midwifery nurses	584	0.1%
2323	Education advisers and school inspectors	569	0.1%
9251	Elementary storage supervisors	566	0.1%
5232	Vehicle body builders and repairers	564	0.1%
8144	Weighers, graders and sorters	549	0.1%
2224	Psychotherapists and cognitive behaviour therapists	549	0.1%
3415	Musicians	547	0.1%
9131	Industrial cleaning process occupations	541	0.1%
5449	Other skilled trades n.e.c.	536	0.1%
9253	Delivery operatives	530	0.1%
1255	Managers and directors in the creative industries	526	0.1%
1172	Social services managers and directors	525	0.1%
6132	Ambulance staff (excluding paramedics)	519	0.1%
9264	Waiters and waitresses	506	0.1%
8131	Paper and wood machine operatives	502	0.1%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
8139	Plant and machine operatives n.e.c.	500	0.1%
7212	Telephonists	499	0.1%
3224	Counsellors	495	0.1%
8141	Assemblers (electrical and electronic products)	495	0.1%
3573	Information technology trainers	481	0.1%
3112	Electrical and electronics technicians	477	0.1%
9266	Coffee shop workers	465	0.1%
3421	Interior designers	461	0.1%
8133	Energy plant operatives	459	0.1%
4113	Officers of non-governmental organisations	444	0.1%
1221	Hotel and accommodation managers and proprietors	444	0.1%
8115	Metal making and treating process operatives	441	0.1%
4124	Finance officers	435	0.1%
8134	Water and sewerage plant operatives	433	0.1%
3413	Actors, entertainers and presenters	431	0.1%
8212	Bus and coach drivers	429	0.1%
8142	Assemblers (vehicles and metal goods)	427	0.1%
3114	Building and civil engineering technicians	420	0.1%
8229	Mobile machine drivers and operatives n.e.c.	416	0.1%
2252	Optometrists	413	0.1%
5225	Air-conditioning and refrigeration installers and repairers	410	0.1%
6211	Sports and leisure assistants	407	0.1%
2124	Electronics engineers	406	0.1%
2151	Conservation professionals	399	0.1%
3581	Inspectors of standards and regulations	392	0.0%
2255	Paramedics	390	0.0%
8160	Production, factory and assembly supervisors	381	0.0%
5113	Gardeners and landscape gardeners	369	0.0%
5422	Printers	366	0.0%
2411	Barristers and judges	346	0.0%
8135	Printing machine assistants	343	0.0%
9265	Bar staff	339	0.0%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
5323	Painters and decorators	336	0.0%
5435	Cooks	327	0.0%
2463	Clergy	326	0.0%
6117	Playworkers	324	0.0%
9119	Fishing and other elementary agriculture occupations n.e.c.	321	0.0%
8239	Other drivers and transport operatives n.e.c.	319	0.0%
2253	Dental practitioners	317	0.0%
2324	Early education and childcare services managers	310	0.0%
6114	Childminders	303	0.0%
3311	Non-commissioned officers and other ranks	302	0.0%
6219	Leisure and travel service occupations n.e.c.	302	0.0%
5224	Precision instrument makers and repairers	299	0.0%
9224	Launderers, dry cleaners and pressers	292	0.0%
6133	Dental nurses	279	0.0%
1135	Charitable organisation managers and directors	274	0.0%
3313	Fire service officers (watch manager and below)	270	0.0%
9261	Bar and catering supervisors	264	0.0%
4214	Company secretaries and administrators	263	0.0%
2256	Podiatrists	260	0.0%
6134	Houseparents and residential wardens	257	0.0%
1223	Publicans and managers of licensed premises	253	0.0%
5222	Tool makers, tool fitters and markers-out	249	0.0%
4217	Typists and related keyboard occupations	243	0.0%
6213	Air travel assistants	237	0.0%
8152	Road construction operatives	235	0.0%
5317	Glaziers, window fabricators and fitters	235	0.0%
3411	Artists	232	0.0%
5212	Metal plate workers, smiths, moulders and related occupations	231	0.0%
9225	Refuse and salvage occupations	231	0.0%
9233	Exam invigilators	221	0.0%
9269	Other elementary services occupations n.e.c.	218	0.0%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
3512	Ship and hovercraft officers	218	0.0%
9267	Leisure and theme park attendants	216	0.0%
7124	Market and street traders and assistants	212	0.0%
3433	Fitness and wellbeing instructors	212	0.0%
6116	Nannies and au pairs	212	0.0%
6137	Care escorts	211	0.0%
8114	Plastics process operatives	211	0.0%
6312	Parking and civil enforcement occupations	210	0.0%
9259	Elementary storage occupations n.e.c.	209	0.0%
6129	Animal care services occupations n.e.c.	209	0.0%
5322	Floorers and wall tilers	208	0.0%
2483	Environmental health professionals	205	0.0%
8119	Process operatives n.e.c.	204	0.0%
2471	Librarians	194	0.0%
1231	Health care practice managers	185	0.0%
8145	Tyre, exhaust and windscreen fitters	185	0.0%
7214	Market research interviewers	184	0.0%
5119	Agricultural and fishing trades n.e.c.	183	0.0%
5243	TV, video and audio servicers and repairers	179	0.0%
3422	Clothing, fashion and accessories designers	176	0.0%
8146	Sewing machinists	173	0.0%
9241	Shelf fillers	169	0.0%
4135	Library clerks and assistants	168	0.0%
5314	Roofers, roof tilers and slaters	166	0.0%
5211	Sheet metal workers	166	0.0%
2462	Probation officers	161	0.0%
2472	Archivists and curators	161	0.0%
9112	Forestry and related workers	154	0.0%
2240	Veterinarians	152	0.0%
9226	Vehicle valeters and cleaners	144	0.0%
8132	Mining and quarry workers and related operatives	143	0.0%
5321	Plasterers	140	0.0%
1225	Travel agency managers and proprietors	140	0.0%
8233	Air transport operatives	138	0.0%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
9111	Farm workers	136	0.0%
3214	Complementary health associate professionals	135	0.0%
8221	Crane drivers	134	0.0%
9222	Street cleaners	134	0.0%
5423	Print finishing and binding workers	130	0.0%
1163	Senior officers in fire, ambulance, prison and related services	129	0.0%
6221	Hairdressers and barbers	129	0.0%
1212	Managers and proprietors in forestry, fishing and related services	126	0.0%
7121	Collector salespersons and credit agents	126	0.0%
5114	Groundsmen and greenkeepers	123	0.0%
5112	Horticultural trades	122	0.0%
5233	Vehicle paint technicians	121	0.0%
5313	Bricklayers	121	0.0%
8234	Rail transport operatives	115	0.0%
1112	Elected officers and representatives	114	0.0%
3414	Dancers and choreographers	112	0.0%
1258	Directors in consultancy services	110	0.0%
5442	Furniture makers and other craft woodworkers	109	0.0%
3431	Sports players	107	0.0%
9262	Hospital porters	101	0.0%
8231	Train and tram drivers	101	0.0%
5214	Pipe fitters	99	0.0%
1257	Hire services managers and proprietors	98	0.0%
1123	Production managers and directors in mining and energy	95	0.0%
1254	Waste disposal and environmental services managers	94	0.0%
5411	Upholsterers	93	0.0%
5234	Aircraft maintenance and related trades	91	0.0%
1161	Officers in armed forces	91	0.0%
1140	Directors in logistics, warehousing and transport	89	0.0%
3314	Prison service officers (below principal officer)	88	0.0%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
3240	Veterinary nurses	88	0.0%
5111	Farmers	84	0.0%
5431	Butchers	81	0.0%
3211	Dispensing opticians	81	0.0%
6121	Pest control officers	78	0.0%
5421	Pre-press technicians	78	0.0%
1256	Betting shop and gambling establishment managers	78	0.0%
5235	Boat and ship builders and repairers	78	0.0%
5312	Stonemasons and related trades	77	0.0%
2126	Aerospace engineers	77	0.0%
8151	Scaffolders, stagers and riggers	72	0.0%
3312	Police officers (sergeant and below)	72	0.0%
5441	Glass and ceramics makers, decorators and finishers	71	0.0%
7112	Retail cashiers and check-out operators	70	0.0%
5412	Footwear and leather working trades	69	0.0%
1211	Managers and proprietors in agriculture and horticulture	67	0.0%
1253	Hairdressing and beauty salon managers and proprietors	67	0.0%
2464	Youth work professionals	66	0.0%
6138	Undertakers, mortuary and crematorium assistants	66	0.0%
5432	Bakers and flour confectioners	65	0.0%
8213	Taxi and cab drivers and chauffeurs	61	0.0%
5311	Steel erectors	60	0.0%
5413	Tailors and dressmakers	58	0.0%
8232	Marine and waterways transport operatives	58	0.0%
1252	Garage managers and proprietors	56	0.0%
7123	Roundspersons and van salespersons	54	0.0%
3511	Aircraft pilots and air traffic controllers	53	0.0%
9121	Groundworkers	51	0.0%
5419	Textiles, garments and related trades n.e.c.	49	0.0%
7125	Visual merchandisers and related occupations	49	0.0%

SOC 2020 Code	Group title 4-digit	Frequency	Distribution
1233	Early education and childcare services proprietors	45	0.0%
9249	Elementary sales occupations n.e.c.	45	0.0%
8153	Rail construction and maintenance operatives	44	0.0%
9229	Elementary cleaning occupations n.e.c.	44	0.0%
8215	Driving instructors	39	0.0%
6214	Rail travel assistants	26	0.0%
5236	Rail and rolling stock builders and repairers	22	0.0%
6311	Police community support officers	21	0.0%
9221	Window cleaners	21	0.0%
2435	Professional/chartered company secretaries	19	0.0%
1162	Senior police officers	17	0.0%
5433	Fishmongers and poultry dressers	14	0.0%
5443	Florists	6	0.0%
6250	Bed and breakfast and guest house owners and proprietors	4	0.0%
	Total	789,208	100.0%

Source: IER-LMI. Own calculations

Table A2. Sample of the fuzzy merge classification at various distance levels

#	Job title	Code	SOC 4-digit title	Distance	SOC 6-digits	Group title SOC 6-digit
1	nursing associates	3213	Medical and dental technicians	0.00	3213/07	nursing associates
2	groundworkers	9121	Groundworkers	0.00	9121/00	groundworkers
3	credit controllers	4121	Credit controllers	0.00	4121/00	credit controllers
4	recruitment administrators	4136	Human resources administrative occupations	0.00	4136/01	recruitment administrators
5	warehouse operatives	9252	Warehouse operatives	0.00	9252/00	warehouse operatives
6	transport planners	2455	Construction project managers and related professionals	0.00	2455/02	transport planners
7	physiotherapy assistants	6131	Nursing auxiliaries and assistants	0.00	6131/06	physiotherapy assistants
8	bar staff	9265	Bar staff	0.00	9265/00	bar staff
9	warehouse operatives	9252	Warehouse operatives	0.00	9252/00	warehouse operatives
10	cover supervisors	6112	Teaching assistants	0.00	6112/01	cover supervisors
11	commercial litigation senior solicitor	2412	Solicitors and lawyers	0.13	2412/02	commercial solicitors lawyers
12	commercial property solicitor teesside	2412	Solicitors and lawyers	0.13	2412/02	commercial solicitors lawyers
13	speciality doctor paediatrics	2212	Specialist medical practitioners	0.13	2212/99	specialist medical practitioners n e c
14	automation warehouse maintenance engineer	5223	Metal working production and maintenance fitters	0.13	5223/01	automation maintenance technicians
15	health safety compliance	3582	Health and safety managers and officers	0.13	3582/99	health safety managers officers n e c

#	Job title	Code	SOC 4-digit title	Distance	SOC 6-digits	Group title SOC 6-digit
	manager occ					
16	customer service administrator sutton coldfield â mth ftc	7219	Customer service occupations n.e.c.	0.13	7219/99	customer service occupations n e c
17	secondary teacher design technology ascension island package â	2313	Secondary education teaching professionals	0.13	2313/99	secondary education teaching professionals n e c
18	education commercialisation manager mfd	2322	Education managers	0.13	2322/99	education managers n e c
19	food specification technologist	2129	Engineering professionals n.e.c.	0.13	2129/06	food technologists
20	logisitics optimisation manager	1243	Managers in logistics	0.13	1243/01	logistics managers
21	junior commercial litigation solicitor	2412	Solicitors and lawyers	0.31	2412/02	commercial solicitors lawyers
22	technology recruiter senior management	1136	Human resource managers and directors	0.31	1136/03	learning development managers
23	sen teachers north west	2316	Special needs education teaching professionals	0.31	2316/01	special educational teachers coordinators
24	contact centre sales advisor west	7211	Call and contact centre occupations	0.31	7211/00	call contact centre occupations

#	Job title	Code	SOC 4-digit title	Distance	SOC 6-digits	Group title SOC 6-digit
25	snr clinical fellow	2212	Specialist medical practitioners	0.38	2212/99	specialist medical practitioners n e c
26	junior software engineer c	2134	Programmers and software development professionals	0.38	2134/03	software developers
27	project export coordinator	3542	Importers and exporters	0.38	3542/00	importers exporters
28	export coordinator manager	3542	Importers and exporters	0.38	3542/00	importers exporters
29	reservations consultant telesales	7113	Telephone salespersons	0.38	7113/00	telephone salespersons
30	work home energy telesales advisors	7113	Telephone salespersons	0.38	7113/00	telephone salespersons
31	registered manager registered nurse	2237	Other nursing professionals	0.46	2237/00	nursing professionals
32	gp hmp durham	2211	Generalist medical practitioners	0.46	2211/99	generalist medical practitioners n e c
33	marketing administration assistant contract	4151	Sales administrators	0.47	4151/00	sales administrators
34	microsoft uc consultant	2134	Programmers and software development professionals	0.46	2134/01	computer games designers
35	health care assistant nhs	6135	Care workers and home carers	0.47	6135/03	residential care workers
36	c developer asp net mvc wpf wcf sql	2134	Programmers and software development professionals	0.47	2134/01	computer games designers

#	Job title	Code	SOC 4-digit title	Distance	SOC 6-digits	Group title SOC 6-digit
37	jt consultant psychiatrist	2212	Specialist medical practitioners	0.47	2212/01	anaesthetists
38	intermediate care assistants	6135	Care workers and home carers	0.47	6135/03	residential care workers
39	principal teacher fif	2313	Secondary education teaching professionals	0.47	2313/14	teachers physical education
40	lecturer public services	2311	Higher education teaching professionals	0.47	2311/03	teaching professionals biological sciences

Source: IER-LMI. Own calculations

Table A3. Coefficient of variations categories and vacancy frequency

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	3554/04	Promoters	1,800	0.07%
Low	3133/01	Database administrators	1,751	0.07%
Low	3416/04	Producers	520	0.02%
Low	2226/02	Educational psychologists	831	0.03%
Low	2124/99	Electronics engineers n.e.c.	395	0.02%
Low	5245/00	Security system installers and repairers	2,960	0.12%
Low	6240/00	Cleaning and housekeeping managers and supervisors	3,155	0.13%
Low	1241/01	Airport managers	582	0.02%
Low	2482/01	Compliance and regulatory professionals	3,654	0.15%
Low	4159/05	Warranty administrators	2,702	0.11%
Low	4143/00	Customer service managers	4,872	0.20%
Low	2493/02	Press officers	1,117	0.05%
Low	2461/99	Social workers n.e.c.	6,394	0.26%
Low	2440/03	Risk managers	2,462	0.10%
Low	3520/02	Compliance officers	2,056	0.08%
Low	4124/00	Finance officers	1,735	0.07%
Low	2421/04	Forensic accountants	4,968	0.20%
Low	3549/02	Business support officers	1,324	0.05%
Low	3571/02	Human resources advisors	6,600	0.27%
Low	1121/00	Production managers and directors in manufacturing	12,689	0.52%
Low	3549/08	Research coordinators	967	0.04%
Low	6135/99	Care workers and home carers n.e.c.	45,798	1.86%
Low	5434/04	Head chefs	2,422	0.10%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	4159/99	Other administrative occupations n.e.c.	13,431	0.55%
Low	2492/99	Newspaper and periodical journalists and reporters n.e.c.	754	0.03%
Low	4136/99	Human resources administrative occupations n.e.c.	1,245	0.05%
Low	5315/00	Plumbers & heating and ventilation installers and repairers	4,600	0.19%
Low	3534/03	Investment managers	1,530	0.06%
Low	3421/00	Interior designers	1,901	0.08%
Low	3549/09	Strategy consultants	1,864	0.08%
Low	3229/99	Welfare and housing associate professionals n.e.c.	2,064	0.08%
Low	1224/99	Leisure and sports managers n.e.c.	305	0.01%
Low	3582/99	Health and safety managers and officers n.e.c.	2,833	0.12%
Low	5319/99	Construction and building trades n.e.c.	641	0.03%
Low	2440/99	Business and financial project management professionals n.e.c.	6,111	0.25%
Low	2133/04	IT systems architects	2,105	0.09%
Low	3133/99	Database administrators and web content technicians n.e.c.	2,301	0.09%
Low	4122/01	Accounting clerks and bookkeepers	30,308	1.23%
Low	3551/00	Buyers and procurement officers	5,901	0.24%
Low	7212/00	Telephonists	2,650	0.11%
Low	3541/06	Vehicle damage assessors	1,078	0.04%
Low	2422/02	Financial advisors and planners	8,742	0.36%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	2432/02	Marketing managers	7,123	0.29%
Low	5221/00	Metal machining setters and setter-operators	4,554	0.19%
Low	2493/03	Public relations officers	973	0.04%
Low	4131/03	Court officers	1,002	0.04%
Low	5232/00	Vehicle body builders and repairers	1,645	0.07%
Low	2134/03	Software developers	47,960	1.95%
Low	3412/07	Technical writers	863	0.04%
Low	1131/99	Financial managers and directors n.e.c.	6,863	0.28%
Low	2133/99	IT business analysts, architects and systems designers n.e.c.	2,647	0.11%
Low	2421/06	Insolvency examiners	991	0.04%
Low	1131/03	Finance managers and directors	13,102	0.53%
Low	2133/03	Data engineers	2,569	0.10%
Low	4123/02	Post office clerks	1,127	0.05%
Low	2493/04	Social media managers	3,348	0.14%
Low	3571/01	Equality and diversity officers	1,917	0.08%
Low	4213/01	School administrators	1,606	0.07%
Low	2132/06	Systems managers	583	0.02%
Low	2133/01	Computer analysts and scientists	3,660	0.15%
Low	1139/03	Communications and policy managers	1,477	0.06%
Low	3544/00	Data analysts	8,508	0.35%
Low	1232/03	Residential care managers	4,228	0.17%
Low	2491/00	Newspaper and periodical editors	2,409	0.10%
Low	4121/00	Credit controllers	8,848	0.36%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	1136/01	Diversity and equality managers	996	0.04%
Low	4131/05	Examination officers	836	0.03%
Low	6136/03	Senior residential care workers	2,698	0.11%
Low	3556/01	Brand managers	4,085	0.17%
Low	2121/05	Structural engineers	1,590	0.06%
Low	7219/01	Customer service advisers	30,122	1.23%
Low	4215/02	Personal assistants	6,052	0.25%
Low	2134/01	Computer games designers	5,636	0.23%
Low	6137/00	Care escorts	622	0.03%
Low	2439/01	Policy officers	1,982	0.08%
Low	2129/05	Energy engineers	1,132	0.05%
Low	4132/03	Pensions administrators	2,298	0.09%
Low	1136/04	Recruitment managers	4,785	0.19%
Low	2440/02	Contract managers	11,354	0.46%
Low	3313/99	Fire service officers (watch manager and below) n.e.c.	679	0.03%
Low	3112/99	Electrical and electronics technicians n.e.c.	942	0.04%
Low	3553/00	Merchandisers	2,972	0.12%
Low	3557/02	Exhibition managers and organisers	1,893	0.08%
Low	3541/03	Estimators	1,509	0.06%
Low	2122/03	Mechanical design engineers	2,169	0.09%
Low	1136/03	Learning and development managers	1,599	0.07%
Low	5242/00	Telecoms and related network installers and repairers	8,991	0.37%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	5223/99	Metal working production and maintenance fitters n.e.c.	10,024	0.41%
Low	3571/03	Recruitment officers	28,440	1.16%
Low	5436/00	Catering and bar managers	1,746	0.07%
Low	5246/00	Electrical service and maintenance mechanics and repairers	4,889	0.20%
Low	4129/99	Finance administrative occupations n.e.c	8,423	0.34%
Low	2433/02	Data scientists	3,807	0.15%
Low	3554/01	Fundraisers	3,947	0.16%
Low	2431/03	Risk analysts	1,326	0.05%
Low	4212/01	Legal administrators	725	0.03%
Low	3416/05	Studio and stage managers	509	0.02%
Low	3119/01	Art teaching technicians	4,765	0.19%
Low	3571/99	Human resources and industrial relations officers n.e.c.	18,095	0.74%
Low	1251/01	Facilities managers	2,345	0.10%
Low	4129/04	Grants administrators	2,235	0.09%
Low	1224/02	Golf course managers	346	0.01%
Low	2129/06	Food technologists	1,687	0.07%
Low	2419/02	Conveyancers	2,224	0.09%
Low	2421/03	Financial controllers	4,486	0.18%
Low	1150/00	Managers and directors in retail and wholesale	12,766	0.52%
Low	2132/02	IT service delivery managers	2,522	0.10%
Low	2125/99	Production and process engineers n.e.c.	1,451	0.06%
Low	4141/00	Office managers	5,404	0.22%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	7211/00	Call and contact centre occupations	3,678	0.15%
Low	3557/01	Conference managers and organisers	1,719	0.07%
Low	2139/99	Information technology professionals n.e.c.	2,047	0.08%
Low	2127/03	Project managers	1,539	0.06%
Low	2142/99	Graphic and multimedia designers n.e.c.	4,627	0.19%
Low	3534/02	Credit control managers	5,595	0.23%
Low	7115/01	Vehicle parts advisers	1,149	0.05%
Low	4215/01	Farm secretaries	1,938	0.08%
Low	6136/02	Senior domiciliary care workers	2,152	0.09%
Low	2132/04	Product managers	1,110	0.05%
Low	2434/02	Researchers in national security and protective services	5,585	0.23%
Low	6113/99	Educational support assistants n.e.c.	5,595	0.23%
Low	2431/99	Management consultants and business analysts n.e.c.	3,839	0.16%
Low	3555/02	Lettings agents	5,408	0.22%
Low	1132/99	Marketing, sales and advertising directors n.e.c.	4,211	0.17%
Low	2431/02	Management consultants	3,160	0.13%
Low	5250/04	Metalwork supervisors	682	0.03%
Low	2223/00	Speech and language therapists	3,656	0.15%
Low	4216/00	Receptionists	12,285	0.50%
Low	2422/03	Financial analysts	6,128	0.25%
Low	8120/99	Metal working machine operatives n.e.c.	3,090	0.13%
Low	3534/04	Relationship managers	8,357	0.34%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	2453/00	Quantity surveyors	4,934	0.20%
Low	8112/99	Textile process operatives n.e.c.	2,239	0.09%
Low	2139/02	Webmasters and website managers	336	0.01%
Low	4215/99	Personal assistants and other secretaries n.e.c.	3,692	0.15%
Low	3132/00	IT user support technicians	23,340	0.95%
Low	4122/99	Bookkeepers, payroll managers and wage clerks n.e.c.	9,028	0.37%
Low	6135/02	Domiciliary care workers	19,973	0.81%
Low	9251/00	Elementary storage supervisors	1,531	0.06%
Low	2222/00	Occupational therapists	9,617	0.39%
Low	2131/00	IT project managers	3,975	0.16%
Low	2111/99	Chemical scientists n.e.c.	339	0.01%
Low	3533/01	Accounting technicians	4,619	0.19%
Low	5434/05	Pastry chefs	1,649	0.07%
Low	4159/04	Property administrators	4,615	0.19%
Low	1134/03	Procurement and purchasing managers and directors	1,973	0.08%
Low	8143/00	Routine inspectors and testers	4,512	0.18%
Low	3116/00	Planning, process and production technicians	2,011	0.08%
Low	3574/99	Other vocational and industrial trainers n.e.c.	4,586	0.19%
Low	5223/10	Motor vehicle part fitters	2,084	0.08%
Low	4134/00	Transport and distribution clerks and assistants	5,830	0.24%
Low	1132/03	Sales directors	2,233	0.09%
Low	3534/99	Financial accounts managers n.e.c.	5,422	0.22%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	9261/00	Bar and catering supervisors	879	0.04%
Low	2452/02	Planning officers	1,087	0.04%
Low	2319/01	Adult education tutors	1,393	0.06%
Low	3533/03	Insolvency administrators	3,903	0.16%
Low	2434/01	Researchers in media and entertainment	4,062	0.17%
Low	2423/00	Taxation experts	9,996	0.41%
Low	3543/00	Project support officers	8,564	0.35%
Low	2133/02	Data architects	2,269	0.09%
Low	6135/01	Community support workers	44,070	1.79%
Low	2419/99	Legal professionals n.e.c.	4,544	0.18%
Low	1139/06	Intellectual property managers and directors	1,195	0.05%
Low	4159/01	Business administrators	10,940	0.45%
Low	4131/08	Purchasing administrators	3,882	0.16%
Low	3532/00	Insurance underwriters	4,005	0.16%
Low	6232/00	Caretakers	3,339	0.14%
Low	3222/03	Learning and behaviour mentors	2,931	0.12%
Low	3520/99	Legal associate professionals n.e.c.	2,810	0.11%
Low	2133/06	IT solutions designers	2,774	0.11%
Low	3556/99	Sales accounts and business development managers n.e.c.	7,523	0.31%
Low	1133/02	Public relations managers	1,009	0.04%
Low	4136/02	Training administrators	2,827	0.12%
Low	3222/01	Child protection officers	778	0.03%
Low	2432/01	Commercial managers	2,261	0.09%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	3531/04	Stockbrokers	1,678	0.07%
Low	3131/99	IT operations technicians n.e.c.	3,563	0.15%
Low	1136/02	Employee relations managers	1,466	0.06%
Low	4136/01	Recruitment administrators	2,020	0.08%
Low	5250/02	Maintenance supervisors	479	0.02%
Low	5244/00	Computer system and equipment installers and servicers	4,360	0.18%
Low	2412/99	Solicitors and lawyers n.e.c.	2,199	0.09%
Low	2161/01	Design managers	2,508	0.10%
Low	2316/01	Special educational teachers and coordinators	6,545	0.27%
Low	2132/05	Software development managers	2,427	0.10%
Low	2434/99	Business and related research professionals n.e.c.	1,154	0.05%
Low	3531/02	Insurance and mortgage brokers	3,859	0.16%
Low	3556/04	Export and import managers	2,293	0.09%
Low	2455/02	Transport planners	1,232	0.05%
Low	2135/00	Cyber security professionals	8,159	0.33%
Low	1221/00	Hotel and accommodation managers and proprietors	1,049	0.04%
Low	4111/99	National government administrative occupations n.e.c.	686	0.03%
Low	1251/05	Shopping centre managers	1,402	0.06%
Low	4111/02	National government revenue officers	994	0.04%
Low	2125/03	Industrial and production engineers	456	0.02%
Low	4131/01	Admissions officers	1,189	0.05%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	3119/02	Science technicians	6,135	0.25%
Low	2482/03	Quality assurance professionals	3,047	0.12%
Low	4112/02	Housing assistants	508	0.02%
Low	2455/01	Construction managers	1,365	0.06%
Low	3574/05	Training assessors	5,778	0.24%
Low	3541/01	Claims assessors	743	0.03%
Low	3221/00	Youth and community workers	6,004	0.24%
Low	4131/99	Records clerks and assistants n.e.c.	1,753	0.07%
Low	3549/04	Contract coordinators	884	0.04%
Low	7113/00	Telephone salespersons	5,269	0.21%
Low	8219/00	Road transport drivers n.e.c.	9,212	0.37%
Low	3556/05	Sales managers	9,138	0.37%
Low	5223/03	Catering equipment technicians	1,502	0.06%
Low	4129/01	Account administrators	3,332	0.14%
Low	2422/04	Mortgage advisers	4,105	0.17%
Low	6135/03	Residential care workers	21,381	0.87%
Low	2419/03	Paralegals and legal assistants	2,856	0.12%
Low	2412/01	Criminal solicitors and lawyers	1,988	0.08%
Low	2129/99	Engineering professionals n.e.c.	1,157	0.05%
Low	4129/06	Payments administrators	891	0.04%
Low	4142/00	Office supervisors	5,856	0.24%
Low	4131/07	Medical administrators	1,966	0.08%
Low	4151/00	Sales administrators	8,345	0.34%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	3131/03	Network and systems administrators	2,632	0.11%
Low	1132/02	Marketing directors	2,752	0.11%
Low	5241/99	Electricians and electrical fitters n.e.c.	3,428	0.14%
Low	2161/99	Research and development (R&D) managers n.e.c.	3,653	0.15%
Low	3554/03	Marketing consultants	12,568	0.51%
Low	2129/10	Mechatronic engineers	697	0.03%
Low	9223/99	Cleaners and domestics n.e.c.	9,462	0.39%
Low	3229/06	Mental health workers	6,462	0.26%
Low	3113/01	Aircraft engineering technicians	1,662	0.07%
Low	2451/02	Landscape architects	1,036	0.04%
Low	3229/08	Student support workers	1,073	0.04%
Low	3120/99	CAD, drawing and architectural technicians n.e.c.	2,376	0.10%
Low	7111/00	Sales and retail and assistants	19,645	0.80%
Low	1122/00	Production managers and directors in construction	4,622	0.19%
Low	3552/00	Business sales executives	42,738	1.74%
Low	3534/01	Claims managers	2,405	0.10%
Low	2421/99	Chartered and certified accountants n.e.c.	1,329	0.05%
Low	5223/06	Heavy plant maintenance technicians	1,239	0.05%
Low	2494/01	Advertising account managers	2,246	0.09%
Low	6111/00	Early education and childcare assistants	10,844	0.44%
Low	3557/03	Festival managers and organisers	1,934	0.08%
Low	2421/05	Fund accountants	2,985	0.12%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	2439/99	Business, research and administrative professionals n.e.c.	1,134	0.05%
Low	1132/01	Advertising directors	1,909	0.08%
Low	6136/01	Senior community support workers	2,524	0.10%
Low	8141/00	Assemblers (electrical and electronic products)	1,062	0.04%
Low	2412/04	Property solicitors and lawyers	4,717	0.19%
Low	4129/05	Mortgage administrators	1,016	0.04%
Low	1139/01	Accounts managers and directors	1,530	0.06%
Low	7219/99	Customer service occupations n.e.c.	7,326	0.30%
Low	2123/99	Electrical engineers n.e.c.	1,702	0.07%
Low	2493/99	Public relations professionals n.e.c.	859	0.03%
Low	3574/04	Retail and store trainers	3,472	0.14%
Low	2421/02	Financial accountants	6,200	0.25%
Low	1242/00	Managers in storage and warehousing	3,635	0.15%
Low	2136/00	IT quality and testing professionals	4,529	0.18%
Low	4129/02	Cashiers	511	0.02%
Low	2494/03	Fundraising managers	1,756	0.07%
Low	9223/01	Domestic cleaners	3,926	0.16%
Low	3133/02	Data managers	2,506	0.10%
Low	2431/01	Business analysts	9,234	0.38%
Low	3223/00	Housing officers	2,846	0.12%
Low	3556/03	Communications managers	5,869	0.24%
Low	3229/07	Probation service workers	1,773	0.07%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	2134/99	Programmers and software development professionals n.e.c.	22,659	0.92%
Low	3554/99	Marketing associate professionals n.e.c.	18,415	0.75%
Low	2311/01	Teaching professionals of agriculture and related subjects	3,205	0.13%
Low	5223/01	Automation maintenance technicians	1,254	0.05%
Low	1111/02	Chief executives	3,266	0.13%
Low	4152/00	Data entry administrators	2,565	0.10%
Low	2139/01	DevOps engineers	4,222	0.17%
Low	3556/02	Business development managers	13,418	0.55%
Low	8149/00	Assemblers and routine operatives n.e.c.	2,784	0.11%
Low	2422/01	Credit risk analysts	3,366	0.14%
Low	1222/03	Restaurant managers	1,089	0.04%
Low	4133/03	Supply chain coordinators	2,555	0.10%
Low	6212/00	Travel agents	3,397	0.14%
Low	2412/03	Family solicitors and lawyers	2,257	0.09%
Low	7213/99	Communication operators n.e.c.	2,053	0.08%
Low	2119/99	Natural and social science professionals n.e.c.	766	0.03%
Low	2422/99	Finance and investment analysts and advisers n.e.c.	7,911	0.32%
Low	2421/07	Management accountants	12,014	0.49%
Low	6131/10	Theatre assistants	8,350	0.34%
Low	3554/02	Market researchers	2,525	0.10%
Low	2461/01	Adult social workers	2,263	0.09%
Low	4122/03	Payroll officers	10,447	0.43%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Low	2127/02	Project engineers	1,987	0.08%
Low	2481/03	Quality control engineers	1,765	0.07%
Low	2432/99	Marketing and commercial managers n.e.c.	3,281	0.13%
Low	6113/03	Learning support assistants	16,913	0.69%
Low	1136/99	Human resources managers and directors n.e.c.	9,198	0.37%
Low	2235/00	Mental health nurses	28,549	1.16%
Low	4212/02	Legal secretaries	3,657	0.15%
Low	2412/02	Commercial solicitors and lawyers	5,108	0.21%
Low	2440/01	Business change managers	2,403	0.10%
Medium - Low	3219/07	Health trainers	333	0.01%
Medium - Low	7131/03	Car traders	205	0.01%
Medium - Low	9259/99	Elementary storage occupations n.e.c.	315	0.01%
Medium - Low	1137/02	Operations managers	1,512	0.06%
Medium - Low	5241/02	Electro-mechanical technicians	861	0.04%
Medium - Low	1137/04	Technical directors	1,016	0.04%
Medium - Low	2133/05	IT business analysts	1,297	0.05%
Medium - Low	8159/10	Thermal insulation operatives	528	0.02%
Medium - Low	2236/03	School nurses	1,447	0.06%
Medium - Low	2311/05	Teaching professionals of computer science	856	0.03%
Medium - Low	2311/06	Teaching professionals of creative arts, crafts and design	363	0.01%
Medium - Low	9121/00	Groundworkers	858	0.03%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	9231/02	Court enforcement officers	407	0.02%
Medium - Low	8134/99	Water and sewerage plant operatives n.e.c.	375	0.02%
Medium - Low	4132/01	Claims handlers	1,902	0.08%
Medium - Low	2132/07	Technical support managers	1,396	0.06%
Medium - Low	6131/01	Blood donor carers	962	0.04%
Medium - Low	2319/03	Drama teachers	1,533	0.06%
Medium - Low	9252/00	Warehouse operatives	39,357	1.60%
Medium - Low	2141/01	Applications designers	411	0.02%
Medium - Low	4129/03	Cost administrators	1,100	0.04%
Medium - Low	2322/99	Education managers n.e.c.	907	0.04%
Medium - Low	7129/99	Sales related occupations n.e.c.	1,958	0.08%
Medium - Low	6136/99	Senior care workers n.e.c.	5,015	0.20%
Medium - Low	8211/00	Large goods vehicle drivers	11,454	0.47%
Medium - Low	5222/02	Tool makers	345	0.01%
Medium - Low	6112/02	School teaching assistants	12,762	0.52%
Medium - Low	2132/01	Information managers	870	0.04%
Medium - Low	3432/04	Golf coaches	115	0.00%
Medium - Low	5316/99	Carpenter and joiners n.e.c.	787	0.03%
Medium - Low	5213/99	Welding trades n.e.c.	1,130	0.05%
Medium - Low	1251/99	Property, housing and estate managers n.e.c.	3,242	0.13%
Medium - Low	3111/99	Laboratory technicians n.e.c.	1,234	0.05%
Medium - Low	3560/02	Health service administrators	973	0.04%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	5223/08	Lift technicians	738	0.03%
Medium - Low	1241/99	Managers in transport and distribution n.e.c.	434	0.02%
Medium - Low	9223/02	Commercial cleaners	1,076	0.04%
Medium - Low	1232/99	Residential, day and domiciliary care managers and proprietors n.e.c.	1,172	0.05%
Medium - Low	9264/00	Waiters and waitresses	1,997	0.08%
Medium - Low	3549/01	Aviation consultants	270	0.01%
Medium - Low	1232/01	Day care managers	2,393	0.10%
Medium - Low	5223/04	Door fitters and makers	401	0.02%
Medium - Low	1231/99	Health care practice managers n.e.c.	461	0.02%
Medium - Low	2161/02	Laboratory managers	757	0.03%
Medium - Low	3533/02	Financial controllers	2,488	0.10%
Medium - Low	8139/03	Laser operators	725	0.03%
Medium - Low	3557/99	Conference and exhibition managers and organisers n.e.c.	1,142	0.05%
Medium - Low	2111/01	Analytical chemists	465	0.02%
Medium - Low	3582/02	Fire safety managers and officers	652	0.03%
Medium - Low	5434/01	Chefs in takeaway and fast food establishments	772	0.03%
Medium - Low	3115/03	Quality assurance auditors	864	0.04%
Medium - Low	1137/01	Information security managers	400	0.02%
Medium - Low	2141/99	Web design professionals n.e.c.	390	0.02%
Medium - Low	8115/99	Metal making and treating process operatives n.e.c.	231	0.01%
Medium - Low	5233/00	Vehicle paint technicians	290	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	2433/04	Mathematicians	674	0.03%
Medium - Low	5434/02	Executive chefs	345	0.01%
Medium - Low	1171/03	Hospital managers	299	0.01%
Medium - Low	2129/03	Brewers	88	0.00%
Medium - Low	3541/99	Estimators, valuers and assessors n.e.c.	1,826	0.07%
Medium - Low	3119/99	Science, engineering and production technicians n.e.c.	2,663	0.11%
Medium - Low	8145/99	Tyre, exhaust and windscreen fitters n.e.c.	67	0.00%
Medium - Low	3319/06	Security consultants and managers	1,249	0.05%
Medium - Low	3429/99	Design occupations n.e.c.	938	0.04%
Medium - Low	3542/00	Importers and exporters	1,208	0.05%
Medium - Low	2469/06	Mediators and restorative practice professionals	835	0.03%
Medium - Low	2224/03	Cognitive behavioural therapists	282	0.01%
Medium - Low	1112/99	Elected officers and representatives n.e.c.	160	0.01%
Medium - Low	7129/03	Hire controllers	527	0.02%
Medium - Low	8234/03	Shunters	79	0.00%
Medium - Low	2452/01	Chartered architectural technologists	629	0.03%
Medium - Low	1241/04	Station managers	311	0.01%
Medium - Low	1171/99	Health services and public health managers and directors n.e.c.	1,271	0.05%
Medium - Low	5225/00	Air-conditioning and refrigeration installers and repairers	1,213	0.05%
Medium - Low	2129/14	Test engineers	980	0.04%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	3533/99	Financial and accounting technicians n.e.c.	1,083	0.04%
Medium - Low	2259/11	Sexual health advisers	952	0.04%
Medium - Low	3416/99	Arts officers, producers and directors n.e.c.	918	0.04%
Medium - Low	4111/01	National government benefits officers	878	0.04%
Medium - Low	3229/03	Charity workers	272	0.01%
Medium - Low	3416/01	Art consultants	141	0.01%
Medium - Low	7220/00	Customer service supervisors	2,309	0.09%
Medium - Low	4211/02	Medical secretaries	1,775	0.07%
Medium - Low	3412/06	Script writers	507	0.02%
Medium - Low	9231/99	Security guards and related occupations n.e.c.	695	0.03%
Medium - Low	5435/00	Cooks	1,720	0.07%
Medium - Low	9219/01	Reprographics assistants	478	0.02%
Medium - Low	5412/03	Leather workers	218	0.01%
Medium - Low	2493/01	Communications officers	708	0.03%
Medium - Low	4132/02	Insurance administrators	1,438	0.06%
Medium - Low	2461/03	Children, family and school social workers	1,991	0.08%
Medium - Low	6114/00	Childminders	686	0.03%
Medium - Low	9129/01	Builder's labourers	1,016	0.04%
Medium - Low	5223/12	Textile machine technicians	871	0.04%
Medium - Low	2315/00	Nursery education teaching professionals	14,915	0.61%
Medium - Low	5321/02	Plasterers	441	0.02%
Medium - Low	5423/99	Print finishing and binding workers n.e.c	158	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	3319/03	Fraud investigators	536	0.02%
Medium - Low	4123/99	Bank and post office clerks n.e.c.	911	0.04%
Medium - Low	2492/02	Newspaper journalists and reporters	785	0.03%
Medium - Low	2316/99	Special needs education teaching professionals n.e.c.	3,699	0.15%
Medium - Low	4123/01	Bank and building society clerks	2,173	0.09%
Medium - Low	3412/03	Copywriters	723	0.03%
Medium - Low	2322/01	Higher education managers	643	0.03%
Medium - Low	8111/99	Food, drink and tobacco process operatives n.e.c.	2,111	0.09%
Medium - Low	1243/02	Supply chain managers	973	0.04%
Medium - Low	3229/01	Care service officers	1,210	0.05%
Medium - Low	3229/04	Health advisors	1,044	0.04%
Medium - Low	3412/04	Literary editors	473	0.02%
Medium - Low	2113/99	Biochemists and biomedical scientists n.e.c.	1,321	0.05%
Medium - Low	8120/01	Metal dressers and finishers	740	0.03%
Medium - Low	4131/04	Enumerators	529	0.02%
Medium - Low	1241/02	Dispatch and distribution managers	372	0.02%
Medium - Low	1139/05	Events mangers and directors	386	0.02%
Medium - Low	4133/99	Stock control clerks and assistants n.e.c.	1,449	0.06%
Medium - Low	2124/02	Electronics design engineers	653	0.03%
Medium - Low	2142/01	Animators	687	0.03%
Medium - Low	5330/00	Construction and building trades supervisors	1,898	0.08%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	8133/01	Boiler operatives	486	0.02%
Medium - Low	1139/02	Administration managers and directors	758	0.03%
Medium - Low	8222/00	Fork-lift truck drivers	6,222	0.25%
Medium - Low	9267/05	Stewards	644	0.03%
Medium - Low	3115/02	Quality and test technicians	730	0.03%
Medium - Low	1251/04	Sales and lettings mangers	1,551	0.06%
Medium - Low	1131/04	Investment bankers	927	0.04%
Medium - Low	3113/99	Engineering technicians n.e.c.	1,090	0.04%
Medium - Low	3113/02	Commissioning engineering technicians	745	0.03%
Medium - Low	6113/01	Behaviour support assistants	1,208	0.05%
Medium - Low	2237/00	Other nursing professionals	51,714	2.10%
Medium - Low	9269/03	Porters	85	0.00%
Medium - Low	2494/02	Creative directors	615	0.03%
Medium - Low	2259/99	Other health professionals n.e.c.	1,137	0.05%
Medium - Low	8160/00	Production, factory and assembly supervisors	1,116	0.05%
Medium - Low	8114/99	Plastics process operatives n.e.c.	217	0.01%
Medium - Low	8142/02	Metal goods assemblers (non-vehicle)	363	0.01%
Medium - Low	9222/00	Street cleaners	206	0.01%
Medium - Low	3120/01	BIM and CAD technicians	574	0.02%
Medium - Low	3224/04	Relationship counsellors	226	0.01%
Medium - Low	2433/99	Actuaries, economists and statisticians n.e.c.	2,008	0.08%
Medium - Low	2129/08	Heating and ventilating engineers	752	0.03%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	7122/02	Debt collectors	565	0.02%
Medium - Low	2471/00	Librarians	552	0.02%
Medium - Low	8131/00	Paper and wood machine operatives	1,014	0.04%
Medium - Low	4122/02	Auditors	3,675	0.15%
Medium - Low	2225/00	Clinical psychologists	3,561	0.14%
Medium - Low	1223/00	Publicans and managers of licensed premises	332	0.01%
Medium - Low	9211/02	Mail sorters	1,168	0.05%
Medium - Low	1251/03	Property developers	1,926	0.08%
Medium - Low	3531/99	Brokers n.e.c.	393	0.02%
Medium - Low	3133/03	Website editors	401	0.02%
Medium - Low	2316/03	Teachers of visually impaired	1,396	0.06%
Medium - Low	3417/99	Photographers, audio-visual and broadcasting equipment operators n.e.c.	370	0.02%
Medium - Low	2132/08	Test managers	521	0.02%
Medium - Low	3555/01	Estate agents	4,124	0.17%
Medium - Low	2482/02	Patent and trademark attorneys	968	0.04%
Medium - Low	2313/14	Teachers of physical education	5,575	0.23%
Medium - Low	3229/05	Housing advice and support workers	1,462	0.06%
Medium - Low	3582/01	Asbestos safety managers and officers	1,596	0.06%
Medium - Low	2121/07	Water engineers	607	0.02%
Medium - Low	3560/04	Recycling officers	326	0.01%
Medium - Low	9263/00	Kitchen and catering assistants	7,130	0.29%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	3115/99	Quality assurance technicians n.e.c.	551	0.02%
Medium - Low	8132/99	Mining and quarry workers and related operatives n.e.c.	98	0.00%
Medium - Low	1131/01	Bank, building society and post office managers	1,508	0.06%
Medium - Low	9253/00	Delivery operatives	1,126	0.05%
Medium - Low	3111/02	Food technicians	348	0.01%
Medium - Low	3229/09	Victim support workers	426	0.02%
Medium - Low	2319/04	Forest school practitioners	431	0.02%
Medium - Low	2317/00	Teachers of English as a Foreign Language	6,949	0.28%
Medium - Low	2313/99	Secondary education teaching professionals n.e.c.	10,764	0.44%
Medium - Low	1241/03	Fleet managers	406	0.02%
Medium - Low	2313/08	Teachers of geography	5,392	0.22%
Medium - Low	2313/04	Teachers of design technology	1,229	0.05%
Medium - Low	9231/06	Security officers	2,700	0.11%
Medium - Low	7132/00	Sales supervisors - retail and wholesale	4,363	0.18%
Medium - Low	7115/02	Vehicle sales executives	2,005	0.08%
Medium - Low	1259/04	Library managers	329	0.01%
Medium - Low	3574/01	Business coaches	504	0.02%
Medium - Low	2251/00	Pharmacists	8,087	0.33%
Medium - Low	1259/05	Recruitment agency managers	1,017	0.04%
Medium - Low	9232/00	School midday and crossing patrol occupations	2,367	0.10%
Medium - Low	3131/04	Quality assurance testers	586	0.02%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	2121/02	Environmental engineers	421	0.02%
Medium - Low	2122/01	Automotive engineers	176	0.01%
Medium - Low	2212/06	Haematologists	198	0.01%
Medium - Low	1171/02	Directors of nursing, midwifery and allied heath professionals	2,802	0.11%
Medium - Low	1134/04	Supply chain managers and directors	418	0.02%
Medium - Low	3412/08	Translators and interpreters	737	0.03%
Medium - Low	2433/01	Actuaries and actuarial analysts	1,244	0.05%
Medium - Low	2234/02	Operating department practitioners	2,154	0.09%
Medium - Low	3114/02	Civil engineering technicians	463	0.02%
Medium - Low	2313/11	Teachers of mathematics	11,552	0.47%
Medium - Low	3512/99	Ship and hovercraft officers n.e.c.	324	0.01%
Medium - Low	2229/09	Psychological wellbeing practitioners	1,106	0.05%
Medium - Low	8133/99	Energy plant operatives n.e.c.	266	0.01%
Medium - Low	2141/02	UI and UX designers	2,215	0.09%
Medium - Low	8151/01	Riggers	379	0.02%
Medium - Low	5241/01	Electric vehicle charging point installers	552	0.02%
Medium - Low	5317/02	Window fabricators	85	0.00%
Medium - Low	1134/01	Commissioning managers and directors	787	0.03%
Medium - Low	3222/04	Welfare and attendance officers	448	0.02%
Medium - Low	3549/10	Transport consultants	367	0.01%
Medium - Low	2221/02	Physiotherapists	4,602	0.19%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	2132/99	IT managers n.e.c.	1,230	0.05%
Medium - Low	1133/01	Communications managers	1,206	0.05%
Medium - Low	1222/02	Catering managers	104	0.00%
Medium - Low	6133/01	Dental nurses	1,200	0.05%
Medium - Low	2314/00	Primary education teaching professionals	41,785	1.70%
Medium - Low	1172/00	Social services managers and directors	1,668	0.07%
Medium - Low	9131/00	Industrial cleaning process occupations	1,091	0.04%
Medium - Low	2129/13	Robotics engineers	597	0.02%
Medium - Low	1251/02	Landlords, property and housing managers	594	0.02%
Medium - Low	2421/01	Auditors	1,284	0.05%
Medium - Low	2461/04	Criminal justice social workers	1,040	0.04%
Medium - Low	2329/99	Other educational professionals n.e.c.	503	0.02%
Medium - Low	3112/01	Avionics technicians	182	0.01%
Medium - Low	8142/99	Assemblers (vehicles and metal goods) n.e.c.	854	0.03%
Medium - Low	8239/00	Other drivers and transport operatives n.e.c.	917	0.04%
Medium - Low	3520/03	Conveyancing executives	639	0.03%
Medium - Low	2211/99	Generalist medical practitioners n.e.c.	3,890	0.16%
Medium - Low	6131/08	Speech and language therapy assistants	4,335	0.18%
Medium - Low	5231/99	Vehicle technicians, mechanics and electricians n.e.c.	4,383	0.18%
Medium - Low	1134/99	Purchasing managers and directors n.e.c.	452	0.02%
Medium - Low	7131/99	Shopkeepers and owners - retail and wholesale n.e.c.	481	0.02%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	1171/01	Clinical managers	1,628	0.07%
Medium - Low	4159/03	Proofreaders	748	0.03%
Medium - Low	7213/02	Emergency services call handlers	153	0.01%
Medium - Low	8142/03	Vehicle and vehicle part assemblers	349	0.01%
Medium - Low	7122/05	Vending machine operators	284	0.01%
Medium - Low	2132/03	Network managers	616	0.03%
Medium - Low	4213/02	School secretaries	238	0.01%
Medium - Low	5231/03	Heavy and commercial vehicle technicians	1,262	0.05%
Medium - Low	9225/02	Refuse collectors	283	0.01%
Medium - Low	7131/05	Jewellers	166	0.01%
Medium - Low	9139/00	Elementary process plant occupations n.e.c.	1,541	0.06%
Medium - Low	6231/00	Housekeepers and related occupations	2,931	0.12%
Medium - Low	5250/03	Mechanical supervisors	233	0.01%
Medium - Low	3232/00	Early education and childcare practitioners	4,349	0.18%
Medium - Low	1139/08	Research and development managers and directors	1,926	0.08%
Medium - Low	1232/02	Domiciliary care managers	771	0.03%
Medium - Low	9119/99	Fishing and other elementary agriculture occupations n.e.c.	222	0.01%
Medium - Low	2461/02	Addiction and substance abuse social workers	736	0.03%
Medium - Low	3572/99	Careers advisers and vocational guidance specialists n.e.c.	1,959	0.08%
Medium - Low	2454/03	Land surveyors	1,073	0.04%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	2324/00	Early education and childcare services managers	1,416	0.06%
Medium - Low	2234/99	Nurse practitioners n.e.c.	3,090	0.13%
Medium - Low	9129/07	Plumber's assistants	2,241	0.09%
Medium - Low	3549/99	Business associate professionals n.e.c.	751	0.03%
Medium - Low	2134/02	Computer programmers	1,836	0.07%
Medium - Low	2454/01	Building control surveyors	2,459	0.10%
Medium - Low	2115/06	GIS analysts	637	0.03%
Medium - Low	8159/03	Building maintenance operatives	1,405	0.06%
Medium - Low	3422/99	Clothing, fashion and accessories designers n.e.c.	98	0.00%
Medium - Low	3231/00	Higher level teaching assistants	4,219	0.17%
Medium - Low	5231/04	Motorcycle technicians	1,203	0.05%
Medium - Low	5231/05	Motorsport technicians	2,365	0.10%
Medium - Low	1243/01	Logistics managers	691	0.03%
Medium - Low	6131/03	Maternity care assistants	652	0.03%
Medium - Low	6138/04	Funeral directors and undertakers	70	0.00%
Medium - Low	2115/99	Social and humanities scientists n.e.c.	1,613	0.07%
Medium - Low	9224/99	Launderers, dry cleaners and pressers n.e.c.	237	0.01%
Medium - Low	6131/09	Sterile services assistants	1,488	0.06%
Medium - Low	1225/00	Travel agency managers and proprietors	211	0.01%
Medium - Low	5249/99	Electrical and electronic trades n.e.c.	865	0.04%
Medium - Low	2137/00	IT network professionals	4,326	0.18%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	4132/99	Pensions and insurance clerks and assistants	670	0.03%
Medium - Low	2119/01	Sports scientists	1,159	0.05%
Medium - Low	3549/07	Marine consultants	484	0.02%
Medium - Low	5316/03	Joiners	964	0.04%
Medium - Low	3541/05	Valuers	856	0.03%
Medium - Low	5434/06	Sous chefs	2,885	0.12%
Medium - Low	3432/99	Sports coaches, instructors and officials n.e.c.	541	0.02%
Medium - Low	4159/02	Church administrators	9,116	0.37%
Medium - Low	5319/07	Steel fixers and underpinners	517	0.02%
Medium - Low	1259/06	Security services managers	573	0.02%
Medium - Low	1131/02	Company secretaries, treasurers and company registrars	2,771	0.11%
Medium - Low	3549/06	Clinical trials assistants	580	0.02%
Medium - Low	3412/99	Authors, writers and translators n.e.c.	751	0.03%
Medium - Low	2121/99	Civil engineers n.e.c.	456	0.02%
Medium - Low	3520/05	Law clerks	149	0.01%
Medium - Low	2455/99	Construction project managers and related professionals n.e.c.	640	0.03%
Medium - Low	8145/01	Tyre technicians	327	0.01%
Medium - Low	5434/99	Chefs n.e.c.	4,520	0.18%
Medium - Low	7122/99	Debt, rent and other cash collectors n.e.c.	534	0.02%
Medium - Low	6117/00	Playworkers	807	0.03%
Medium - Low	1111/03	Diplomats and foreign office officials	604	0.02%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium - Low	1111/99	Chief executives and senior officials n.e.c.	683	0.03%
Medium - Low	2439/02	Registrars	542	0.02%
Medium - Low	4133/02	Inventory and stock controllers	1,230	0.05%
Medium - Low	2226/04	Health psychologists	561	0.02%
Medium - Low	5249/02	Field engineers	732	0.03%
Medium - Low	5449/07	Paint sprayers	722	0.03%
Medium - Low	6112/99	Teaching assistants n.e.c.	21,451	0.87%
Medium - Low	6112/01	Cover supervisors	5,887	0.24%
Medium	5231/07	Vehicle testers	763	0.03%
Medium	1140/00	Directors in logistics, warehousing and transport	153	0.01%
Medium	2236/02	Paediatric nurses	3,759	0.15%
Medium	3111/01	Chemical laboratory technicians	454	0.02%
Medium	5323/00	Painters and decorators	1,123	0.05%
Medium	3311/01	Non-commissioned Army officers and others ranks	203	0.01%
Medium	3319/02	Customs and border control officers	271	0.01%
Medium	2321/01	Heads of further education and sixth form colleges	1,246	0.05%
Medium	2226/03	Forensic psychologists	550	0.02%
Medium	8112/01	Colourists	94	0.00%
Medium	8135/00	Printing machine assistants	628	0.03%
Medium	6211/99	Sports and leisure assistants n.e.c.	212	0.01%
Medium	2112/02	Biologists	542	0.02%
Medium	2259/06	Mental health workers	387	0.02%
Medium	2212/10	Paediatricians	349	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	8112/04	Winders	212	0.01%
Medium	3311/02	Non-commissioned Royal Air Force officers and other ranks	148	0.01%
Medium	3582/03	Nuclear safety managers and officers	278	0.01%
Medium	2113/02	Biomedical scientists	1,874	0.08%
Medium	9226/00	Vehicle valeters and cleaners	300	0.01%
Medium	3531/03	Shipbrokers	217	0.01%
Medium	1212/03	Forestry managers	100	0.00%
Medium	5243/00	TV, video and audio servicers and repairers	271	0.01%
Medium	2469/05	Health education professionals	561	0.02%
Medium	2319/05	Music and singing teachers	2,681	0.11%
Medium	2411/01	Barristers and advocates	383	0.02%
Medium	2114/02	Geophysicists	203	0.01%
Medium	5231/02	Car/light vehicle technicians	1,208	0.05%
Medium	5316/01	Carpenters	1,077	0.04%
Medium	7129/04	Print brokers	193	0.01%
Medium	8112/03	Spinners	335	0.01%
Medium	6132/00	Ambulance staff (excluding paramedics)	1,949	0.08%
Medium	2492/01	Broadcast journalists	255	0.01%
Medium	5234/00	Aircraft maintenance and related trades	135	0.01%
Medium	9111/00	Farm labourers	256	0.01%
Medium	3413/02	Broadcasters, podcasters and presenters	67	0.00%
Medium	5112/00	Horticultural trades	250	0.01%
Medium	3414/99	Dancers and choreographers n.e.c.	69	0.00%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	4112/04	Revenue officers	434	0.02%
Medium	9119/01	Cutters and pickers	385	0.02%
Medium	4217/03	Typists	210	0.01%
Medium	5213/02	Pipe welders	586	0.02%
Medium	2123/01	Control and instrumentation engineers	120	0.00%
Medium	2162/00	Other researchers, unspecified discipline	1,717	0.07%
Medium	2321/02	Heads of higher education establishments	527	0.02%
Medium	2322/02	School managers	735	0.03%
Medium	2481/01	Garment technologists	184	0.01%
Medium	5319/01	Builders	175	0.01%
Medium	2129/11	Nuclear engineers	437	0.02%
Medium	3429/04	Set designers	400	0.02%
Medium	9262/00	Hospital porters	439	0.02%
Medium	4113/02	Clerks to governors	158	0.01%
Medium	6134/01	Foster carers	512	0.02%
Medium	3417/01	Camera operators and videographers	370	0.02%
Medium	5241/03	Installation and maintenance electricians	266	0.01%
Medium	3512/02	Deck officers	82	0.00%
Medium	3574/03	Forklift instructors	138	0.01%
Medium	3219/03	Assistant practitioners	258	0.01%
Medium	2151/03	Heritage officers	113	0.00%
Medium	5223/07	Hydraulic technicians	190	0.01%
Medium	4131/02	Booking clerks	275	0.01%
Medium	2111/02	Industrial chemists	109	0.00%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	2122/99	Mechanical engineers n.e.c.	360	0.01%
Medium	8144/03	Weighers	195	0.01%
Medium	1224/03	Gym, fitness and leisure services managers	98	0.00%
Medium	3212/00	Pharmaceutical technicians	2,429	0.10%
Medium	2451/01	Architects (except landscape)	740	0.03%
Medium	8112/02	Cutters	177	0.01%
Medium	2311/07	Teaching professionals of education	253	0.01%
Medium	2125/02	Control engineers	159	0.01%
Medium	8133/02	Compressor operatives	110	0.00%
Medium	3415/01	Composers and musical arrangers	449	0.02%
Medium	2114/99	Physical scientists n.e.c.	2,153	0.09%
Medium	3520/06	Litigation officers	326	0.01%
Medium	5316/04	Model makers	261	0.01%
Medium	8119/04	Rubber workers	85	0.00%
Medium	3581/01	Animal health inspectors	142	0.01%
Medium	3572/01	Careers advisers and coaches	544	0.02%
Medium	3417/04	Sound engineers	129	0.01%
Medium	3415/99	Musicians n.e.c.	188	0.01%
Medium	2121/03	Geotechnical engineers	416	0.02%
Medium	2151/99	Conservation professionals n.e.c.	354	0.01%
Medium	2119/02	University researchers (unspecified discipline)	274	0.01%
Medium	3432/14	Swimming coaches	82	0.00%
Medium	3429/01	Industrial designers	329	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	7114/99	Pharmacy and optical dispensing assistants n.e.c	608	0.02%
Medium	2311/17	Teaching professionals of subjects allied to medicine	230	0.01%
Medium	3319/04	Immigration enforcement officers	273	0.01%
Medium	3224/01	Bereavement counsellors	121	0.00%
Medium	3213/03	Dental technicians	171	0.01%
Medium	5213/03	Plate welders	130	0.01%
Medium	2454/99	Chartered surveyors n.e.c.	1,658	0.07%
Medium	2454/04	Property surveyors	654	0.03%
Medium	3224/02	Debt advisers	156	0.01%
Medium	3222/99	Child and early years officers n.e.c.	318	0.01%
Medium	2313/03	Teachers of chemistry	3,343	0.14%
Medium	2121/06	Transportation engineers	185	0.01%
Medium	8133/03	Nuclear decommissioning operatives	224	0.01%
Medium	3131/01	Configuration managers	521	0.02%
Medium	3319/99	Protective service associate professionals n.e.c.	180	0.01%
Medium	3113/03	Wind turbine technicians	518	0.02%
Medium	3111/04	Medical laboratory technicians	813	0.03%
Medium	2433/03	Economists	980	0.04%
Medium	5213/01	Braziers and solderers	284	0.01%
Medium	3574/02	Cycle instructors	548	0.02%
Medium	1139/99	Functional managers and directors n.e.c.	439	0.02%
Medium	6113/02	Communication support assistants	527	0.02%
Medium	8152/99	Road construction operatives n.e.c.	207	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	2469/03	Childrens and family services professionals	329	0.01%
Medium	5223/13	Turbine technicians	342	0.01%
Medium	2461/06	Mental health and healthcare social workers	743	0.03%
Medium	7129/02	Energy consultants	316	0.01%
Medium	1224/06	Parks and gardens managers	127	0.01%
Medium	2452/03	Urban designers	158	0.01%
Medium	8229/99	Mobile machine drivers and operatives n.e.c.	589	0.02%
Medium	2115/02	Archaeologists	167	0.01%
Medium	6129/07	Animal welfare workers	114	0.00%
Medium	2254/99	Medical radiographers n.e.c.	1,273	0.05%
Medium	2481/99	Quality control and planning engineers n.e.c.	233	0.01%
Medium	3531/01	Commodity brokers	681	0.03%
Medium	2472/99	Archivists and curators n.e.c.	87	0.00%
Medium	7131/01	Antique dealers	201	0.01%
Medium	2211/01	General practitioners	1,658	0.07%
Medium	1255/99	Managers and directors in the creative industries n.e.c.	349	0.01%
Medium	2240/00	Veterinarians	538	0.02%
Medium	2224/02	Child and adolescent psychotherapists	373	0.02%
Medium	4214/02	Company secretaries	405	0.02%
Medium	2129/12	Patent agents	92	0.00%
Medium	2323/01	Educational advisers	910	0.04%
Medium	2319/02	Dance teachers	3,113	0.13%
Medium	5231/01	Auto electricians	688	0.03%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	3224/05	Substance misuse counsellors	311	0.01%
Medium	2462/00	Probation officers	404	0.02%
Medium	2259/08	Occupational health advisers	1,372	0.06%
Medium	2313/06	Teachers of economics	6,771	0.28%
Medium	2313/05	Teachers of drama	317	0.01%
Medium	1252/00	Garage managers and proprietors	167	0.01%
Medium	5223/02	Bicycle mechanics	233	0.01%
Medium	4211/01	Medical administrators	740	0.03%
Medium	2311/99	Higher education teaching professionals n.e.c.	355	0.01%
Medium	3224/99	Counsellors n.e.c.	267	0.01%
Medium	7121/02	Insurance agents	101	0.00%
Medium	2313/19	Teachers of Welsh	1,191	0.05%
Medium	9132/99	Packers, bottlers, canners and fillers n.e.c.	1,257	0.05%
Medium	2259/05	Medical physicists and nuclear medicine scientists	1,473	0.06%
Medium	2483/00	Environmental health professionals	426	0.02%
Medium	2323/02	School inspectors	234	0.01%
Medium	2112/05	Microbiologists and bacteriologists	692	0.03%
Medium	2433/05	Statisticians	463	0.02%
Medium	3319/05	Private investigators	235	0.01%
Medium	5423/01	Binders and print finishers	77	0.00%
Medium	2233/01	Intensive care nurses	1,152	0.05%
Medium	8134/02	Leakage operatives	73	0.00%
Medium	2229/05	Nutritionists	81	0.00%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	2329/01	Educational administrators	656	0.03%
Medium	2461/05	Homelessness and housing social workers	191	0.01%
Medium	2254/01	Diagnostic radiographers and sonographers	1,626	0.07%
Medium	2226/99	Other psychologists n.e.c.	1,957	0.08%
Medium	3520/07	Probate officers	391	0.02%
Medium	2233/99	Specialist nurses n.e.c.	3,330	0.14%
Medium	7131/04	Greengrocers	70	0.00%
Medium	3414/03	Dance teachers	124	0.01%
Medium	3581/99	Inspectors of standards and regulations n.e.c.	111	0.00%
Medium	4135/00	Library clerks and assistants	320	0.01%
Medium	2212/01	Anaesthetists	458	0.02%
Medium	5422/02	Flexographic printers	68	0.00%
Medium	1224/01	Cinema managers	224	0.01%
Medium	5113/04	Landscape gardeners	441	0.02%
Medium	2152/04	Environmental scientists	292	0.01%
Medium	5311/00	Steel erectors	89	0.00%
Medium	5212/04	Platers	143	0.01%
Medium	3560/01	Electoral services officers	1,082	0.04%
Medium	5241/05	Solar panel installers	133	0.01%
Medium	7112/00	Retail cashiers and check-out operators	266	0.01%
Medium	5317/99	Glaziers, window fabricators and fitters n.e.c.	147	0.01%
Medium	2121/01	Building and building services engineers	607	0.02%
Medium	4214/01	Company administrators	213	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	5313/00	Bricklayers	563	0.02%
Medium	9211/99	Postal workers, mail sorters and messengers n.e.c.	1,666	0.07%
Medium	3549/03	Business systems analysts	306	0.01%
Medium	2311/11	Teaching professionals of law	297	0.01%
Medium	2312/22	Physics teaching professionals	346	0.01%
Medium	1255/01	Art gallery managers	104	0.00%
Medium	7219/02	Reservations agents	462	0.02%
Medium	3557/04	Wedding planners and organisers	205	0.01%
Medium	4133/01	Despatch clerks	451	0.02%
Medium	2115/08	Political researchers	999	0.04%
Medium	3111/03	Health physics monitors	489	0.02%
Medium	3114/01	Building engineering technicians	309	0.01%
Medium	2115/03	Behavioural scientists	245	0.01%
Medium	9132/02	Manual packers, bottlers, canners and fillers	445	0.02%
Medium	6131/04	Occupational therapy assistants	1,024	0.04%
Medium	9225/01	Recycling operatives	670	0.03%
Medium	2319/06	Private tutors	2,427	0.10%
Medium	8144/99	Weighers, graders and sorters n.e.c.	570	0.02%
Medium	3429/03	Packaging technologists	256	0.01%
Medium	2122/02	Marine engineers	332	0.01%
Medium	3120/03	Draughtspersons	270	0.01%
Medium	1257/00	Hire services managers and proprietors	170	0.01%
Medium	1254/06	Sewage works and water treatment managers	71	0.00%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	3314/01	Operational support grades	156	0.01%
Medium	3560/03	Job Centre officers	465	0.02%
Medium	2259/10	Physiologists	412	0.02%
Medium	2115/09	Public health analysts	370	0.02%
Medium	4113/99	Officers of non-governmental organisations n.e.c.	800	0.03%
Medium	5249/01	Broadcast and communications technicians	126	0.01%
Medium	7114/01	Optical dispensing assistants	1,237	0.05%
Medium	5317/03	Window fitters	196	0.01%
Medium	2151/01	Conservationists	234	0.01%
Medium	5432/01	Bakers	289	0.01%
Medium	2111/04	Research and development chemists	284	0.01%
Medium	8139/01	Aerial installers	197	0.01%
Medium	2463/00	Clergy	673	0.03%
Medium	2129/01	Acoustic engineers	435	0.02%
Medium	6131/07	Radiography assistants	147	0.01%
Medium	8212/00	Bus and coach drivers	1,664	0.07%
Medium	1137/99	Information technology directors n.e.c.	649	0.03%
Medium	6222/02	Makeup artists	138	0.01%
Medium	8232/00	Marine and waterways transport operatives	82	0.00%
Medium	3411/02	Art technicians	137	0.01%
Medium	8159/02	Blind fitters	196	0.01%
Medium	9129/99	Elementary construction occupations n.e.c.	120	0.00%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	3222/02	Family liaison officers	356	0.01%
Medium	2313/17	Teachers of religious education	2,821	0.11%
Medium	2469/04	Community and social development professionals	441	0.02%
Medium	5250/01	Electrical supervisors	487	0.02%
Medium	6221/99	Hairdressers and barbers n.e.c.	65	0.00%
Medium	8120/02	Press and stamp operators	492	0.02%
Medium	3413/99	Actors, entertainers and presenters n.e.c.	139	0.01%
Medium	2152/02	Environmental and geo-environmental engineers	414	0.02%
Medium	5114/00	Groundsmen and greenkeepers	414	0.02%
Medium	2316/02	Teachers of the deaf	755	0.03%
Medium	5441/07	Potters	161	0.01%
Medium	3549/11	Work measurement officers	341	0.01%
Medium	2212/14	Surgeons	356	0.01%
Medium	8115/02	Galvanisers and platers	64	0.00%
Medium	2313/16	Teachers of psychology	110	0.00%
Medium	5421/00	Pre-press technicians	121	0.00%
Medium	1137/03	Programme managers	486	0.02%
Medium	1259/02	Educational establishment directors	322	0.01%
Medium	9259/01	Dockers, slingers and stevedores	540	0.02%
Medium	5422/03	Lithographic printers	219	0.01%
Medium	2112/99	Biological scientists n.e.c.	631	0.03%
Medium	9132/01	Machine operatives	1,949	0.08%
Medium	5231/06	Roadside assistance technicians	205	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	5214/00	Pipe fitters	206	0.01%
Medium	8115/06	Powder coaters	192	0.01%
Medium	5411/02	Trimmers and upholsterers	229	0.01%
Medium	5211/99	Sheet metal workers n.e.c.	273	0.01%
Medium	1222/99	Restaurant and catering establishment managers and proprietors n.e.c.	223	0.01%
Medium	7131/02	Art dealers	265	0.01%
Medium	3416/02	Directors	324	0.01%
Medium	3555/03	Auctioneers	241	0.01%
Medium	2129/02	Biomedical engineers	385	0.02%
Medium	1134/02	Estimating managers and directors	242	0.01%
Medium	4112/99	Local government administrative occupations n.e.c.	390	0.02%
Medium	8134/03	Water treatment operatives	314	0.01%
Medium	6131/99	Nursing auxiliaries and assistants n.e.c.	739	0.03%
Medium	8159/07	Drainage operatives	282	0.01%
Medium	8146/00	Sewing machinists	437	0.02%
Medium	7122/03	Meter readers	750	0.03%
Medium	8144/01	Graders	216	0.01%
Medium	2129/04	Clinical Engineers	332	0.01%
Medium	9219/99	Elementary administration occupations n.e.c.	919	0.04%
Medium	7213/01	Air traffic control assistants	76	0.00%
Medium	1259/01	Cleaning and hygiene services managers	184	0.01%
Medium	2212/12	Radiologists	252	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	2152/99	Environmental professionals n.e.c.	296	0.01%
Medium	9266/00	Coffee shop workers	3,235	0.13%
Medium	9231/01	CCTV operators	208	0.01%
Medium	9211/01	Postal delivery workers	1,211	0.05%
Medium	3541/02	Energy advisers and assessors	667	0.03%
Medium	3560/99	Public services associate professionals n.e.c.	345	0.01%
Medium	2234/01	Mental health practitioners	5,189	0.21%
Medium	1224/04	Marina managers	86	0.00%
Medium	2123/03	Power systems engineers	165	0.01%
Medium	7129/01	Demonstrators	244	0.01%
Medium	1255/06	Theatre production managers	196	0.01%
Medium	2313/15	Teachers of physics	789	0.03%
Medium	6121/99	Pest control officers n.e.c.	170	0.01%
Medium	6219/02	Ship stewards	145	0.01%
Medium	8139/99	Plant and machine operatives n.e.c.	673	0.03%
Medium	8113/00	Chemical and related process operatives	5,569	0.23%
Medium	1112/02	Councillors	88	0.00%
Medium	2452/99	Chartered architectural technologists, planning officers and consultants n.e.c.	932	0.04%
Medium	1241/05	Traffic managers	525	0.02%
Medium	6134/99	Houseparents and residential wardens n.e.c.	152	0.01%
Medium	3417/05	Theatre technicians	324	0.01%
Medium	3582/04	Occupational health managers and officers	722	0.03%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
Medium	6131/02	First aid trainers	1,110	0.05%
Medium	3413/03	Children's entertainers	148	0.01%
Medium	6129/04	Animal technicians	88	0.00%
Medium	2255/00	Paramedics	1,641	0.07%
Medium	2229/99	Therapy professionals n.e.c.	190	0.01%
Medium	2152/03	Environmental consultants	301	0.01%
High	5431/01	Butchers	240	0.01%
High	5222/99	Tool makers, tool fitters and markers-out n.e.c	191	0.01%
High	8114/03	Plastic fabricators	71	0.00%
High	3520/01	Adjudicators	132	0.01%
High	8153/00	Rail construction and maintenance operatives	71	0.00%
High	2313/09	Teachers of history	1,341	0.05%
High	3213/01	Cardiac physiologists	535	0.02%
High	7214/02	Mystery shoppers	640	0.03%
High	6131/05	Phlebotomists	549	0.02%
High	1135/00	Charitable organisation managers and directors	505	0.02%
High	2259/04	Health promotion officers	517	0.02%
High	6222/99	Beauticians and related occupations n.e.c.	1,480	0.06%
High	2112/07	Pharmacologists	99	0.00%
High	2226/05	Occupational psychologists	196	0.01%
High	3114/99	Building and civil engineering technicians n.e.c.	160	0.01%
High	6211/04	Leisure and recreation assistants	337	0.01%
High	2472/03	Conservators	131	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	5119/99	Agricultural and fishing trades n.e.c.	178	0.01%
High	9112/99	Forestry and related workers n.e.c.	162	0.01%
High	2311/18	Teaching professionals of veterinary science	186	0.01%
High	2313/10	Teachers of ICT and computer science	402	0.02%
High	5111/01	Agricultural contractors	66	0.00%
High	5223/11	Pump technicians	222	0.01%
High	6312/01	Car park attendants	153	0.01%
High	8115/01	Furnace operators	136	0.01%
High	8231/00	Train and tram drivers	152	0.01%
High	8215/01	Car driving instructors	2,367	0.10%
High	3319/01	Coastguard officers	124	0.01%
High	2312/08	Construction, planning and the built environment teaching professionals	190	0.01%
High	3511/05	Navigators	223	0.01%
High	8111/03	Dairy workers	98	0.00%
High	2211/02	Public health doctors	538	0.02%
High	5224/01	Calibration and precision instrument technicians	229	0.01%
High	2112/04	Horticultural scientists	151	0.01%
High	1222/01	Café managers	104	0.00%
High	5113/02	Gardeners	532	0.02%
High	3433/02	Personal trainers	8,642	0.35%
High	3431/99	Sports players n.e.c.	392	0.02%
High	2114/04	Medical physicists	630	0.03%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	2311/04	Teaching professionals of business and administrative studies	189	0.01%
High	2311/02	Teaching professionals of architecture, building and planning	554	0.02%
High	5314/01	Felt and flat roofers	68	0.00%
High	2311/13	Teaching professionals of mathematical sciences	232	0.01%
High	9233/00	Exam invigilators	901	0.04%
High	2141/03	Web designers	470	0.02%
High	2111/03	Nuclear and radiochemists	202	0.01%
High	2252/00	Optometrists	1,758	0.07%
High	3311/03	Royal Marines Commandos	148	0.01%
High	3219/06	Foot health practitioners	280	0.01%
High	3413/08	Singers	77	0.00%
High	9112/01	Forest workers	99	0.00%
High	2231/00	Midwifery nurses	1,329	0.05%
High	8114/02	Moulders	70	0.00%
High	8132/03	Onshore oil and gas workers	66	0.00%
High	7121/03	Sales representatives	213	0.01%
High	2229/01	Art therapists	112	0.00%
High	2226/07	Sports psychologists	170	0.01%
High	2226/06	Research psychologists	157	0.01%
High	1111/01	Chairperson	161	0.01%
High	2224/99	Psychotherapists and cognitive behaviour therapists n.e.c.	423	0.02%
High	2221/03	Veterinary physiotherapists	3,160	0.13%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	5119/01	Arborists and arboriculturalists	115	0.00%
High	1259/03	Funeral services and crematorium managers	184	0.01%
High	1163/01	Senior officers in ambulance services	134	0.01%
High	2212/07	Homeopaths (medically qualified)	541	0.02%
High	5442/03	Furniture fitters and makers	97	0.00%
High	3312/99	Police officers (sergeant and below) n.e.c.	179	0.01%
High	8111/02	Brewery and distillery workers	142	0.01%
High	1163/03	Senior officers in immigration services	61	0.00%
High	9249/01	Order pickers	278	0.01%
High	9231/03	Court ushers	198	0.01%
High	8213/00	Taxi and cab drivers and chauffeurs	96	0.00%
High	1255/04	Publishing managers	143	0.01%
High	5422/01	Digital printers	223	0.01%
High	6129/99	Animal care services occupations n.e.c.	69	0.00%
High	3411/99	Artists n.e.c.	81	0.00%
High	2311/03	Teaching professionals of biological sciences	168	0.01%
High	7214/01	Market researchers	617	0.03%
High	2114/06	Nuclear energy scientists	249	0.01%
High	3219/04	Dietetic assistants	438	0.02%
High	3417/02	Lighting designers	90	0.00%
High	2313/02	Teachers of biology	292	0.01%
High	4217/99	Typists and related keyboard occupations n.e.c.	193	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	6134/02	Residential care officers	114	0.00%
High	4112/03	Parish clerks	122	0.00%
High	2254/02	Therapeutic radiographers	907	0.04%
High	6211/05	Lifeguards	195	0.01%
High	2124/03	Telecommunications engineers	83	0.00%
High	2212/11	Psychiatrists	493	0.02%
High	2313/12	Teachers of modern foreign languages	206	0.01%
High	2313/07	Teachers of English	2,059	0.08%
High	3581/04	Health and safety inspectors	106	0.00%
High	5113/01	Garden designers	80	0.00%
High	5322/01	Carpet and linoleum fitters	156	0.01%
High	5449/09	Sign makers and writers	116	0.00%
High	9267/02	Cinema attendants	138	0.01%
High	5235/00	Boat and ship builders and repairers	150	0.01%
High	2233/02	Theatre nurses	1,199	0.05%
High	6213/02	Passenger services assistants	299	0.01%
High	8233/02	Flight dispatchers	408	0.02%
High	6312/02	Civil enforcement officers	136	0.01%
High	2329/03	Examiners and markers	744	0.03%
High	5224/04	Meter fitters	105	0.00%
High	3213/99	Medical and dental technicians n.e.c.	590	0.02%
High	3573/00	Information technology trainers	1,005	0.04%
High	8144/02	Sorters	178	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	2469/01	Addiction and substance abuse advice professionals	196	0.01%
High	3433/99	Fitness and wellbeing instructors n.e.c.	1,001	0.04%
High	2481/02	Planning engineers	158	0.01%
High	5224/99	Precision instrument makers and repairers n.e.c.	502	0.02%
High	5449/08	Pattern makers	79	0.00%
High	2129/09	Materials scientists	243	0.01%
High	1253/01	Beauty salon managers	88	0.00%
High	5411/99	Upholsterers n.e.c.	178	0.01%
High	3572/02	Placement officers	1,311	0.05%
High	3214/05	Hospital play specialists	115	0.00%
High	9129/02	Carpenter's and joiner's assistants	89	0.00%
High	5422/04	Screen printers	76	0.00%
High	2121/04	Mining engineers	121	0.00%
High	5249/04	Signal workers	160	0.01%
High	6131/06	Physiotherapy assistants	342	0.01%
High	6219/01	Holiday representatives	85	0.00%
High	5119/03	Falconers	101	0.00%
High	2212/05	Gynaecologists and obstetricians	2,500	0.10%
High	2212/99	Specialist medical practitioners n.e.c.	9,766	0.40%
High	1233/00	Early education and childcare services proprietors	79	0.00%
High	8152/03	Road surfacers	82	0.00%
High	1171/04	Public health managers	270	0.01%
High	2113/01	Biochemists	221	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	2232/02	Mental health and psychiatric community nurses	2,641	0.11%
High	9265/00	Bar staff	1,582	0.06%
High	9231/04	Detention officers	116	0.00%
High	5119/09	Tree surgeons	145	0.01%
High	1139/04	Complaints and customer service directors	1,905	0.08%
High	3115/01	Quality analysts	130	0.01%
High	2124/01	Broadcast engineers	76	0.00%
High	2212/02	Cardiologists	806	0.03%
High	2454/02	Hydrographic surveyors	102	0.00%
High	5212/99	Metal plate workers, smiths, moulders and related occupations n.e.c.	132	0.01%
High	5423/03	Photographic technicians	74	0.00%
High	3549/05	Clinical coders	222	0.01%
High	5250/99	Skilled metal, electrical and electronic trades supervisors n.e.c.	128	0.01%
High	3581/09	Trading standards officers	109	0.00%
High	3219/08	Home economists	63	0.00%
High	3416/03	Editors	108	0.00%
High	1255/05	Radio production managers	106	0.00%
High	5314/99	Roofers, roof tilers and slaters n.e.c.	282	0.01%
High	3240/00	Veterinary nurses	337	0.01%
High	2321/03	Heads of nursery schools	270	0.01%
High	3213/04	Hearing aid dispensers	377	0.02%
High	3412/05	Poets	138	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	3213/02	Dental hygienists	83	0.00%
High	3214/99	Complementary health associate professionals n.e.c.	107	0.00%
High	2115/01	Anthropologists	143	0.01%
High	8139/05	Vehicle dismantlers	70	0.00%
High	1212/05	Racehorse trainers	90	0.00%
High	1211/01	Managers and proprietors in agriculture	102	0.00%
High	3432/13	Sports development officers	158	0.01%
High	8233/03	Ramp agents	84	0.00%
High	3229/02	Celebrants	79	0.00%
High	2259/02	Audiologists	219	0.01%
High	8221/00	Crane drivers	205	0.01%
High	9119/02	Hatchery workers	194	0.01%
High	1259/99	Managers and proprietors in other services n.e.c.	140	0.01%
High	5441/06	Optical glass makers	87	0.00%
High	2129/07	Gas engineers	193	0.01%
High	4131/06	Indexers	83	0.00%
High	3415/03	Instrumentalists	151	0.01%
High	5431/02	Slaughterers	84	0.00%
High	2321/05	Heads of secondary schools	267	0.01%
High	2229/03	Drama therapists	65	0.00%
High	4113/01	Charity administrators	493	0.02%
High	8233/99	Air transport operatives n.e.c.	93	0.00%
High	2112/01	Agricultural scientists	138	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	5422/99	Printers n.e.c.	136	0.01%
High	1231/01	Managers of dental practices	102	0.00%
High	2253/99	Dental practitioners n.e.c.	447	0.02%
High	2129/15	Traffic engineers	157	0.01%
High	3581/08	Nuclear inspectors	80	0.00%
High	2115/05	Genealogists	1,398	0.06%
High	2312/25	Retail and commercial enterprise teaching professionals	138	0.01%
High	3581/02	Building control officers	119	0.00%
High	2259/03	Dietitians and nutritionists	1,369	0.06%
High	9241/00	Shelf fillers	729	0.03%
High	2329/02	Educational consultants	72	0.00%
High	2112/06	Pathologists	235	0.01%
High	2127/01	Contract managers	204	0.01%
High	1258/03	Management consultancy directors	77	0.00%
High	9231/05	Door supervisors	72	0.00%
High	2115/04	Epidemiologists	174	0.01%
High	2123/02	Laser engineers	75	0.00%
High	5322/02	Floor layers	104	0.00%
High	2113/03	Biotechnologists	154	0.01%
High	6134/03	School matrons and houseparents	95	0.00%
High	2236/01	Neonatal nurses	973	0.04%
High	2472/02	Collection managers and curators	131	0.01%
High	3415/02	Conductors	186	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	2152/05	Sustainability officers	375	0.02%
High	5321/01	Dry liners	118	0.00%
High	5113/03	Green roof installers	70	0.00%
High	8159/99	Construction operatives n.e.c.	155	0.01%
High	7219/03	Student ambassadors	173	0.01%
High	2464/00	Youth work professionals	123	0.01%
High	2469/99	Welfare professionals n.e.c.	97	0.00%
High	7131/07	Wholesalers	168	0.01%
High	9269/99	Other elementary service occupations n.e.c.	103	0.00%
High	9267/99	Leisure and theme park attendants n.e.c.	108	0.00%
High	7123/00	Roundspersons and van salespersons	104	0.00%
High	3512/03	Marine engineers	157	0.01%
High	7124/00	Market and street traders and assistants	386	0.02%
High	2212/04	Gastroenterologists	739	0.03%
High	2319/99	Teaching professionals n.e.c.	3,046	0.12%
High	2226/01	Counselling psychologists	1,105	0.04%
High	9224/01	Carpet cleaners	317	0.01%
High	5319/05	Insulation technicians	144	0.01%
High	7114/02	Pharmacy dispensing assistants	1,361	0.06%
High	3411/04	Illustrators (non-medical & non-scientific)	79	0.00%
High	2312/99	Further education teaching professionals n.e.c.	321	0.01%
High	2152/01	Energy managers	143	0.01%
High	2312/14	Health, public services and care teaching professionals	187	0.01%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	3413/07	Models	75	0.00%
High	9129/06	Marquee erectors	123	0.01%
High	3413/09	Social media influencers	139	0.01%
High	3432/03	Football coaches	114	0.00%
High	2256/00	Podiatrists	687	0.03%
High	5443/00	Florists	121	0.00%
High	3432/08	Outdoor pursuits instructors	109	0.00%
High	7214/04	Traffic enumerators	471	0.02%
High	2411/99	Barristers and judges n.e.c	119	0.00%
High	3414/04	Dancers	102	0.00%
High	2151/02	Ecologists	146	0.01%
High	8111/01	Bakery workers	2,311	0.09%
High	1161/03	Royal Marines officers	188	0.01%
High	8229/01	Agricultural machinery drivers	212	0.01%
High	2313/18	Teachers of sociology	89	0.00%
High	8214/00	Delivery drivers and couriers	13,132	0.53%
High	3413/01	Actors	72	0.00%
High	3213/06	Orthotic and prosthetic technicians	314	0.01%
High	2435/00	Professional/chartered company secretaries	71	0.00%
High	9119/04	Shearers	62	0.00%
High	2321/04	Heads of primary schools	518	0.02%
High	2311/16	Teaching professionals of social studies	211	0.01%
High	3520/04	Cost lawyers	116	0.00%
High	3313/02	Firefighters	116	0.00%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	6116/02	Nannies	324	0.01%
High	3314/02	Prison officers	199	0.01%
High	3541/04	Loss adjusters	958	0.04%
High	3422/02	Clothing and fashion designers	83	0.00%
High	1255/03	Film and television production managers	115	0.00%
High	6219/99	Leisure and travel service occupations n.e.c.	227	0.01%
High	3417/03	Photographers	530	0.02%
High	8119/02	Concrete operatives	183	0.01%
High	3131/02	Games testers	631	0.03%
High	5449/06	Musical instrument repairers, makers and turners	82	0.00%
High	2112/03	Botanical scientists	118	0.00%
High	5316/02	Fitters	323	0.01%
High	3213/07	Nursing associates	971	0.04%
High	3219/99	Health associate professionals n.e.c.	272	0.01%
High	9269/05	Turnstile operators	178	0.01%
High	8159/09	Sign fitters	77	0.00%
High	2313/13	Teachers of music	145	0.01%
High	2115/07	Historians	424	0.02%
High	8119/01	Ceramic workers	167	0.01%
High	2125/01	Chemical engineers	250	0.01%
High	3582/05	Road traffic and transport safety managers and officers	161	0.01%
High	1224/07	Theatre managers	261	0.01%
High	8119/99	Process operatives n.e.c.	106	0.00%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	5319/04	Industrial climbers	89	0.00%
High	1111/04	Senior public service officials	255	0.01%
High	3211/00	Dispensing opticians	413	0.02%
High	2311/09	Teaching professionals of historical and philosophical studies	120	0.00%
High	6312/99	Parking and civil enforcement occupations n.e.c	170	0.01%
High	3131/05	Software technicians	815	0.03%
High	7121/01	Credit agents	85	0.00%
High	2212/09	Oncologists	218	0.01%
High	2321/06	Heads of special needs education schools	604	0.02%
High	3512/01	Captains	64	0.00%
High	4112/01	Benefits officers	108	0.00%
High	6221/02	Hairdressers	267	0.01%
High	2313/01	Teachers of art	2,545	0.10%
High	2259/07	Neurophysiologists	270	0.01%
High	5322/03	Tilers	97	0.00%
High	3432/15	Tennis coaches	130	0.01%
High	2312/19	Leisure, travel and tourism teaching professionals	92	0.00%
High	8215/99	Driving instructors n.e.c.	1,622	0.07%
High	5113/99	Gardeners and landscape gardeners n.e.c	112	0.00%
High	9225/99	Refuse and salvage occupations n.e.c.	75	0.00%
High	6213/99	Air travel assistants n.e.c.	188	0.01%
High	2232/01	Learning disability community nurses	1,960	0.08%

Coefficient of Variation	Code SOC 2020 6-digit	Group SOC 2020 6-digit	Number of vacancies	Percentage
High	8151/02	Scaffolders and stagers	92	0.00%
High	2419/01	Adjudicators	77	0.00%
High	2232/99	Community nurses n.e.c.	4,521	0.18%
High	2472/01	Archivists	75	0.00%
High	5449/02	Candle makers	72	0.00%

Source: IER-LMI. Own calculations