Descriptive analysis of the vacancy database

Jeisson Cárdenas

Documento de Trabajo
Alianza EFI - Colombia Científica
Marzo 2020

Número de serie: WP2-2020-004

Descriptive analysis of the vacancy database[1] [2]

Jeisson Cárdenas Rubio
Institute for Employment Research
University of Warwick
Coventry, United Kingdom

Abstract

Given the high cost of collecting labour demand for skills information through surveys, the composition and dynamics of Colombian labour demand are relatively unknown. However, information regarding unmet labour demand can be collected from job portals with the implementation of relatively novel data mining techniques. This paper provides a descriptive analysis to start evaluating the data from job portals. Among the main results of my analysis of the vacancy database show that 1) job vacancies are concentrated in Bogotá, Antioquia and Bolivar; 2) most of the job positions require a person with at least a high school certificate; 3) most occupations in Colombia correspond to middle- ("Sales demonstrators") and low-skilled occupations ("Kitchen helpers"); 4) the skills most demanded include "Customer service" (knowledge), "Communication" (knowledge) and "Work in teams" (competence). Thus, the vacancy database provided detailed, real-time and valuable information about the Colombian labour demand that, previously, it was not possible to obtain from other sources (e.g. household surveys). Moreover, these initial results suggest that the vacancy database is consistent, or at least it does not contradict itself or external data, such as regional GDP, population, etc.

Key words: Labour demand, vacancy database, online job portals.

JEL classification: J23, J24, J31

[1] This working paper is part of the author's PhD thesis at the University of Warwick.

[2] E-mail: j.cardenas-rubio@warwick.ac.uk (J.Cardenas).

1. Introduction

From a theoretical point of view, Cárdenas (2020a) discussed what can be understood as skill mismatches and how this phenomenon might arise in a particular economy (e.g. due to imperfect information). Cárdenas (2020a) showed that this problem has specific relevance in countries such as Colombia. Indeed, evidence from the labour market in this country suggests that skill mismatches might explain a substantial portion of high unemployment and informality rates. One factor that hinders the design of well-orientated public policies to tackle skill mismatches is the absence or scarcity of detailed labour market information. More specifically, given the high cost of collecting labour demand for skills information through surveys, the composition and dynamics of Colombian labour demand are relatively unknown.

However, information regarding unmet labour demand can be collected from job portals with the implementation of relatively novel data mining techniques. These online sources might provide valuable information in real-time and at a low-cost for the analysis of labour demand, and thus the early identification of the labour demand for skills as well as possible skill shortages. Better understanding this information can provide proper information to training providers and policymakers, and in this way might improve education and public policy designs to tackle issues of unemployment and informality.

This paper describes the main characteristics of vacancy data collected and organised in Cárdenas (2020c). Section 2 shows the number of vacancies and job positions demanded by job portals. Then Section 3 displays the geographical coverage of the Colombian vacancy database. The fourth section provides a descriptive analysis of the labour demand for skills in Colombia, and analyses labour demand composition by education, occupation (at a four-digit level), new job titles, skills and experience requirements. Section 5 shows labour demand organised by sectors. The sixth section analyses the most notable trends in Colombian labour demand by occupation: occupations with higher demand, occupations with a significant increase and occupations for which demand has decreased over time. Section 7 describes the distribution of wages offered by employers, and the last section describes other (secondary) characteristics of the vacancy database, such as the type of contract and the duration of vacancies.

2. Vacancy database composition

Cárdenas (2020c) described the methods and the challenges involved in obtaining and organising vacancy information from job portals. As a result of those methods, a Colombian vacancy database has been generated to be tested and analysed for public policy recommendations. The sample period runs from 1st January 2016 to 31st December 2018. Each observation in the database is a vacancy. By the definition that has been applied to this thesis, a vacancy can require one or more people (the total number of jobs or job placements available) (see Cárdenas, 2020c). Following the above definition, the total number of observations (vacancies) in the database are 2,247,959, while the numbers of jobs are 5,720,513 (Table 1). Consequently, a vacancy advertisement on average contains 2.5 job placements.

As shown in Table 1, by volume most of the vacancies (55.7%) and jobs advertised (total vacancies) come from Computrabajo, followed by Elempleo (33.4%) and Servicio de Empleo (10.8%). Likewise, 65.2% of the total number of jobs orginate from Computrabajo, followed by Elempleo (23.7%) and Serviciodeempleo (10.9%)[3] (Section 4 will discuss the types of job titles posted on each job portal).

Table 1: Total number of vacancies and job positions

Source	Total vacancies		Total jobs	
	Number	Percentage	Number	Percentage
Computrabajo	1,252,366	55.7%	3,734,835	65.2%
Elempleo	752,032	33.4%	1,358,911	23.7%
Servicio de Empleo	243,561	10.8%	626,767	10.9%
Total	2,247,959		5,720,513	

Source: Vacancy information 2016–2018. Own calculations.

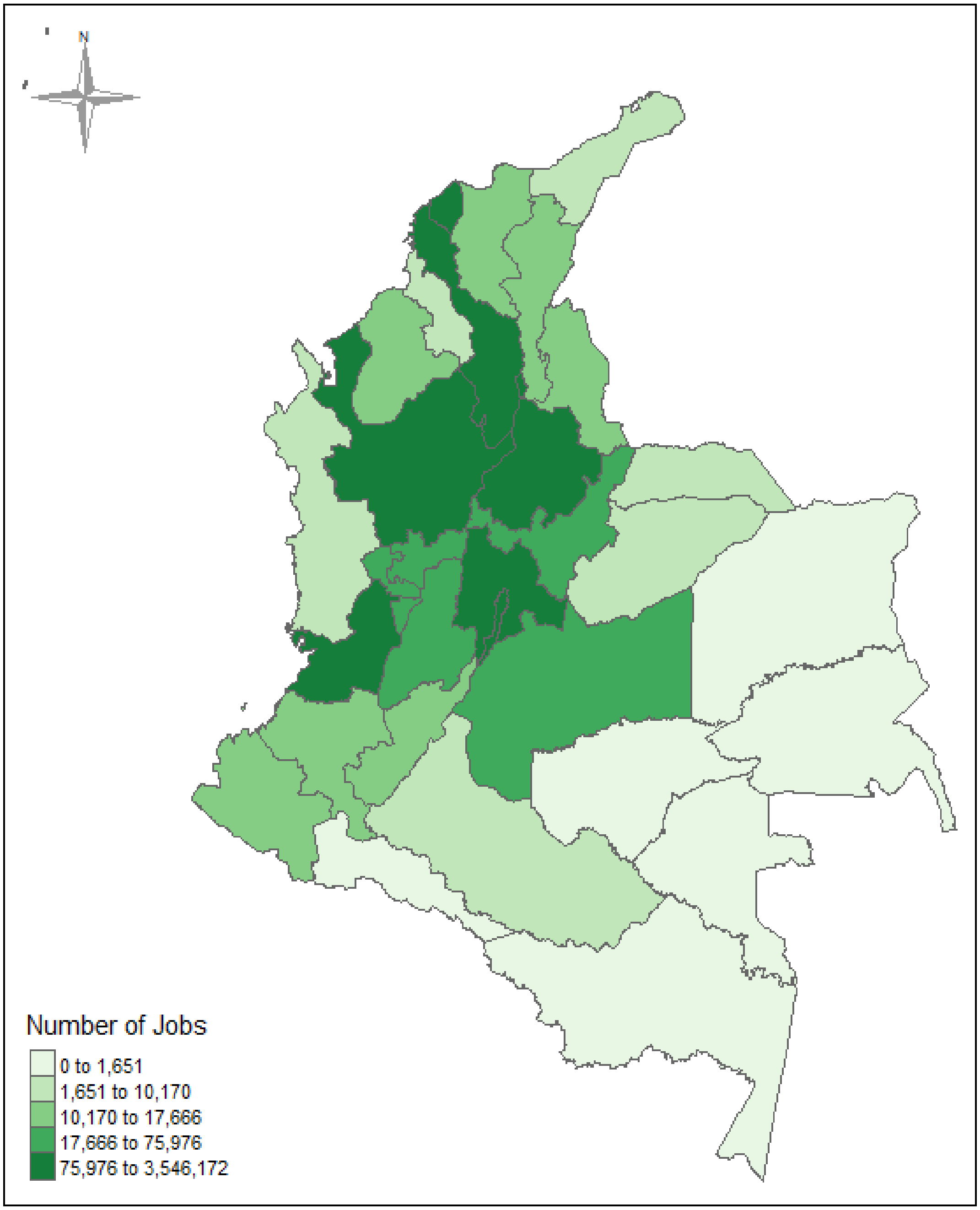
[3] This result reaffirms that websites such as Serviciodeempleo do not necessarily contain the majority of job advertisements, even when the website said that it had 263,621 job vacancies on 30th October 2017 (see Cárdenas, 2020c). As mentioned in Cárdenas (2020c), when clicking on some vacancy announcements on Serviciodeempleo, a new window redirected the search to open another website where the vacancy was posted (e.g. Elempleo).

3. Geographical distribution of vacancies and number of jobs

Figure 1 shows the distribution of vacancies in counties across the Colombian national territory from 2016 to 2018. Colombia is divided into 32 counties[4]. As can be observed, most vacancies and jobs are concentrated in the capital of the country (Bogotá). Indeed, 56.7% (1,276,410) of the total number of vacancies and 61.9% (3,546,172) of the total number of jobs were offered in Bogotá, while 7.9% of vacancies and 9.3% jobs were available in Antioquia, and 15.2% of vacancies and 7.9% of jobs were offered in Bolivar. In contrast, the counties with fewer job placements are Vichada (228 job placements) Guainía (274 job placements) and Vaupes (75 job placements).

[4] Amazonas, Antioquia, Arauca, Atlántico, Bogotá, Bolívar, Boyacá, Caldas, Caquetá, Casanare, Cauca, Cesar, Chocó, Córdoba, Cundinamarca, Guainía, Guaviare, Huila, La Guajira, Magdalena, Meta, Nariño, Norte de Santander, Putumayo, Quindío, Risaralda, San Andrés y Providencia, Santander, Sucre, Tolima, Valle del Cauca, Vaupés and Vichada.

Figure 1: Distribution of job placements by counties 2016-2018



Source: Vacancy information and GEIH 2016-2018. Own calculations. Note: The ranges were chosen according to quintile distribution of job placements in the vacancy database

It is unsurprising more than half of the Colombian job placements are concentrated in Bogotá, and counties such as Vichada possess significantly fewer job placements. First, regarding the population, Bogotá is the biggest city in Colombia. According to the most recent figures published by DANE[5], Bogotá has 8,281,030 inhabitants. This population represents approximately 16.4% of the total Colombian population and 21.3% of the urban Colombian population in 2019. Additionally, Bogotá has 4,609,000 individuals from the economically active population (EAP). This number of people represents 18.6% of the total Colombian EAP and 23.6% of the urban Colombian EAP in 2017[6]. Moreover, the above estimations do not consider that Bogotá attracts workers from its smaller surrounding cities. For instance, it is well-known that people from towns such as Soacha or Chía commute to Bogotá. Thus, by considering the surrounding cities[7], the metropolitan Bogotá population rises to 9,732,848, which represents 19.3% of the Colombian (total) population and 25.1% of the urban Colombian population.

Given the economic concentration in Bogotá, this city produces 24.8% of the Colombian gross domestic product (GDP) (Valencia et al. 2016). Thus, it is logical to expect that the number of available vacancies is higher in Bogotá than elsewhere in the country. Likewise, the second largest county in terms of population and economic activity is Antioquia, followed by Cundinamarca, Atlántico, Bolívar, Valle del Cauca and Santander. Therefore, it is also expected that these counties have a higher number of vacancies when compared to other Colombian counties. In line with this assumption, the counties of Vichada, Guainía and Vaupes contributed only 0.3% of Colombia's GDP in 2017 (DANE, 2017b, p.4) and contained 0.33% of the total Colombian population in 2019[8]. Hence, it is unsurprising that those counties have a lower rate of job placements[9].

Figure 2 shows Colombia's job distribution divided by the EAP in each county from 2016 to 2017. The map does not include information from 2018 due to household data (GEIH) not being currently available (but this information will be added when available)[10]. The first aspect to observe in the map in Figure 2 is the presence of missing values; specifically, in the south-east

---

[5] See: http://www.dane.gov.co/files/investigaciones/poblacion/proyepobla06_20/Municipal_area_1985-2020.xls

[6] See: https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/anexo_empleo_dic_17.xlsx

[7] Soacha, Facatativá, Chía, Zipaquirá, Mosquera, Madrid, Funza, Cajicá, Sibaté, Tocancipá, Tabio, La Calera, Sopó, Cota, Tenjo, El Rosal, Gachancipá and Bojacá.

[8] See: http://www.dane.gov.co/files/investigaciones/poblacion/proyepobla06_20/Municipal_area_1985-2020.xls

[9] Cárdenas (2020d) provides more detailed evidence about the external validity of the vacancy information.

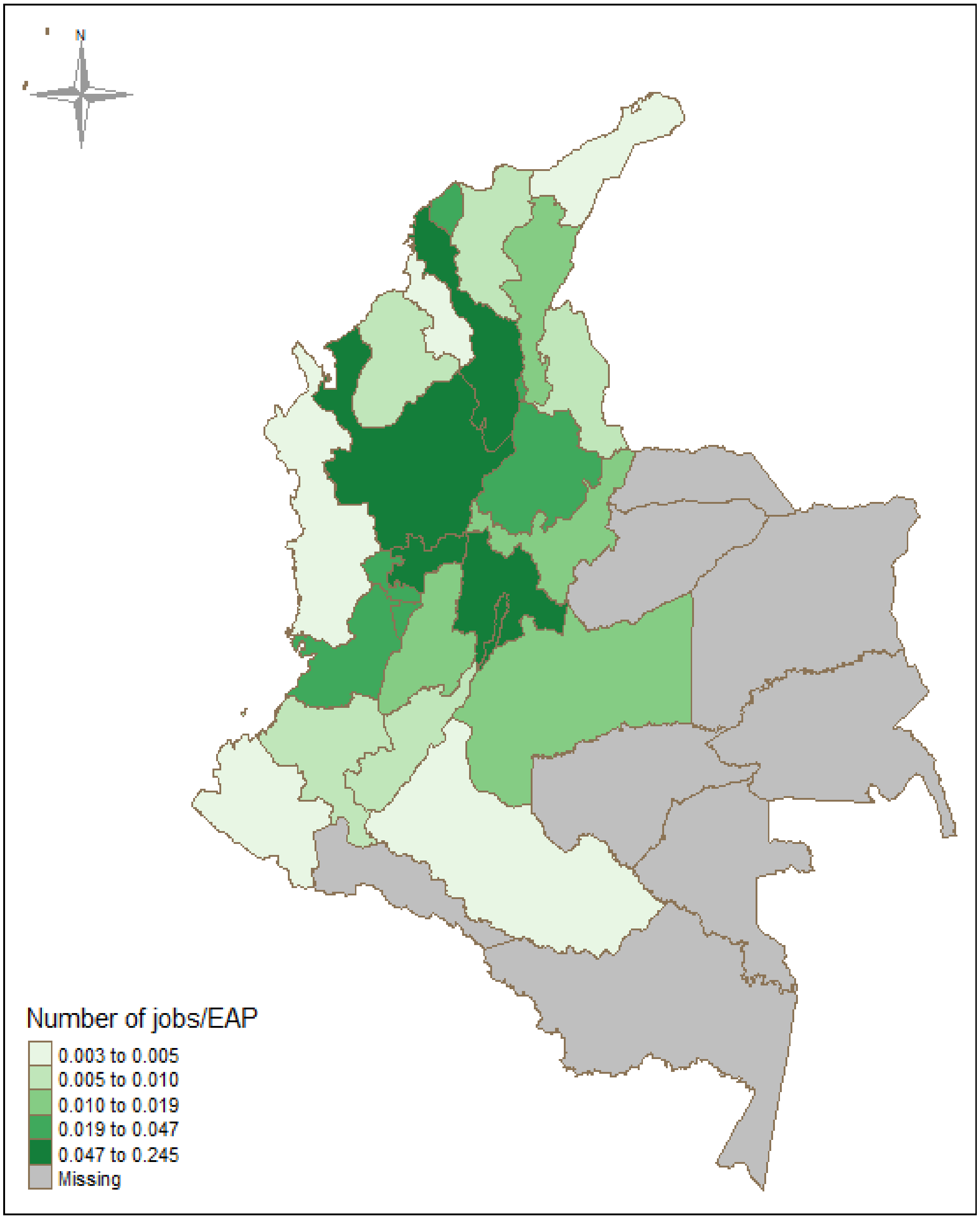
[10] This issue illustrates that there is a degree of delay between the release of household survey results and the problem that researchers or policymakers want to analyse (Cárdenas, 2020a)

zones of Colombian territory. These missing values exist because there is no official information about the labour market (such as EAP and unemployed, among others) in those counties[11]. Consequently, sources such as job portals might facilitate the provision of labour market information where it is difficult to carry out traditional methods (surveys).

According to Figure 2, in Bogotá the ratio between job placements and the EAP is 0.245, which means that for one job placement there are four employed or unemployed workers. For counties such as Antioquia, Cundinamarca, and Caldas y Valle del Cauca, the ratios are around 0.05 (for each job offer there are 20 workers) while for Bolívar the rate is 0.147 (for each job offer there are 6.7 workers).

[11] Due to problems of public order and the difficulty in accessing these areas of the country, the Colombian Bureau of Statistics (since 2012) collects information from the county's capital but no other cities in those south-easterly zones.

Figure 2: Ratio of job placements to the EAP by counties 2016–2017



Source: Vacancy information and GEIH 2016-2017. Own calculations.

Figures 1 and 2 show that the vacancy information is unevenly distributed across the national territory. Online job placements tend to be concentrated in specific zones such as Bogotá, Antioquia, Bolívar, etc. This concentration of data correlates with the relative economic importance of each county. Counties with a larger proportion of the EAP and GDP also tend to have a relatively higher number of job placements. Thus, the geographical results of the vacancy information appear to reflect Colombia's economic and population dynamics[12].

4. Labour demand for skills

As discussed in Cárdenas (2020a), skill is a multi-dimensional concept. However, most of the skill definitions associate this concept to the task complexity attached to each job and the characteristics that each worker needs to successfully carry out the tasks required in a certain job position. Reflecting on the definitions of skill from Cárdenas (2020a), skill is considered as any measurable quality that increases workers' productivity, and can be improved by training or development. Consequently, given the current sources of information available to analyse the labour market (job portals and household surveys), and the information provided by these sources, it is possible to analyse Colombian labour demand by the education, skill and experience demanded (workers' skills), and occupation (skills as job attributes) (see Cárdenas, 2020a).

4.1 Educational requirements

Table 2 provides a general overview of the structure of the Colombian educational system (OEI, 1993)[13]:

[12] Potentially, the labour market analysis in this thesis can be disaggregated at the regional level. However, due to space limitations, (hereinafter) this thesis will present its results aggregated at the national level.

[13] Pre-school education is for children under 6 years old, and basic (and compulsory) education is composed of the elementary and middle school (6th–9th). To have access to higher educational programs it is necessary to have finished high school. People with high school educations can choose between lower, higher vocational or undergraduate programs. Frequently, it is not compulsory to have a lower vocational education qualification to access higher vocational programs. When people finish their undergraduate studies, they can continue studying in a specialisation or a master's program. On the one hand, specialisations are programs that usually involve one year of part-time study, in which people can develop and deepen specific qualifications for a particular occupation, discipline, etc. (MEN, 2016). On the other hand, master's programs usually involve two years of full-time study. To take a PhD (in most cases), it is necessary to first obtain a master's certificate.

Table 2: Structure of the Colombian educational system

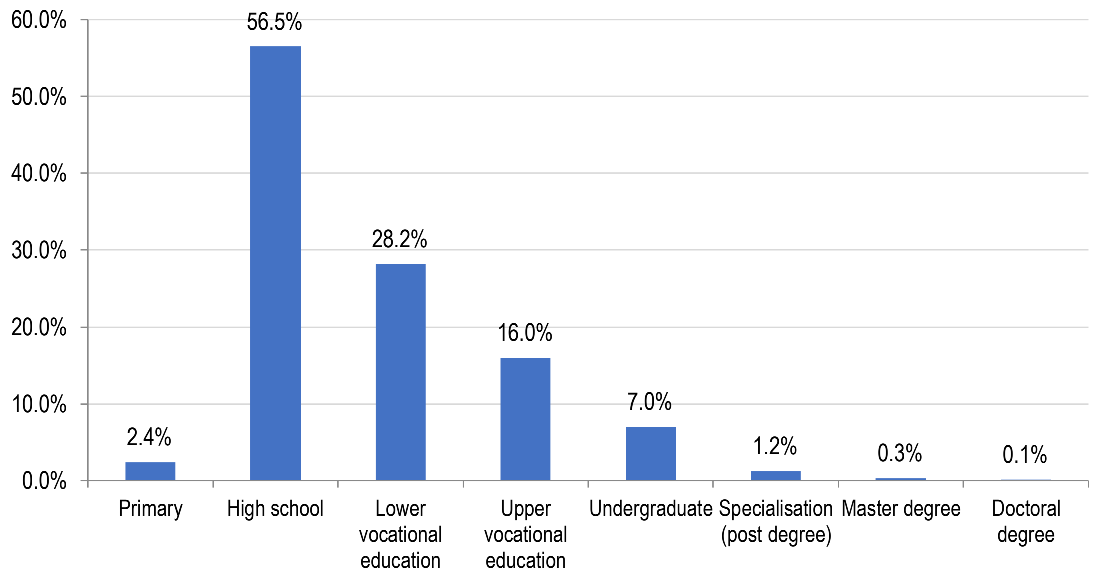
Level of study	Grade
Pre-school	Pre-school
Basic	Elementary /Primary school (1st–5th)
	Middle school (6th–9th)
Intermediate	High school (10th–11th)
Higher education	Lower vocational education
	Higher vocational education
	Undergraduate
	Specialisation
	Master degree
	PhD

Source: OEI, 1993

Figure 3 shows the distribution of jobs available by educational requirements[14]. According to this figure, 56.5% of the job placements ask for a person with (at minimum) a high school degree, followed by lower (28.2%) and upper vocational educational requirements (16.0%). Despite using online sources (job portals), there are a significant number of jobs that require people with a primary school and high school level of education who tend to carry out low- or middle-skilled jobs. This evidence suggests (at least for the Colombian case) that companies do not only search for high-skilled workers when using job portals. As will be seen in more detail in Section 4, job placements posted on job portals cover a variety of low-, middle- and high-skilled jobs.

[14] As pointed out in Appendix B:, employers might be indifferent about educational levels. For instance, a vacancy might require a person with a high school and lower vocational education level. In these cases, the educational dummy variables ("high school" and "lower_vocational _degree") take the value of 1 at the same time. For this reason, the sum of percentages in Figure 3 is more than 100%.

Figure 3: Job placements by minimum educational requirements



Source: Vacancy information. Own calculations.

4.2 Occupational structure

With the occupational variable it is possible to understand labour utilisation and the composition of an economy (high-, middle- and low-skilled jobs), it also allows examining changes (such as job polarisation) in the labour force, and it serves as a guide for training providers and policymakers, among others.

Job portals provide job titles when a vacancy is posted online. As discussed in Cárdenas (2020c), there are techniques available that might help to classify job titles from job portals into occupational groups. However, two concerns arise when using job title information from different job portals for the analysis of labour demand. First, job portals might be biased towards specific groups of occupations. Moreover, given that the vacancy database is composed of a group of main job portals in Colombia (see Cárdenas, 2020c), the results might be biased due to one or more job portals only publishing vacancies for specific occupational groups.

Second, the techniques carried out in Cárdenas (2020c) might misclassify some job titles, and thus the results regarding occupations might be affected[15]. This subsection provides evidence that the concerns mentioned above are not the case for the Colombian vacancy database.

Figure 4 shows a word cloud with the most common job titles for each job portal selected in Cárdenas (2020c). As can be observed, the most frequent job titles are Call centre employees", Customer service" ("cliente"), "Assistants" ("auxilar"), "Salespeople" ("venta"), "Promoter", etc. There are two aspects to highlight from Figure 4. First, the most demanded job titles correspond to low- or middle-skilled occupations. Second, the three job portals offer similar job positions. For instance, in all three job portals one of the most common job titles is "call centre employees". This result suggests that the job portals selected in Cárdenas (2020c) are not biased to a specific market (i.e. high-skilled jobs such as managers or professionals)[16].

[15] For instance, according to job portal information and the techniques carried out in Cárdenas (2020c), one of the most demanded occupations might be "Actors". It is not expected that an occupation such as "Actors" (or other occupations which usually do not have a big market) constitute a significant share of the labour demand.

[16] Cárdenas (2020d) provides more evidence regarding the occupations demanded by job portals.

Figure 4: World cloud. Most frequent job titles by job portals[17]

Elempleo

Computrabajo

Servicio de empleo

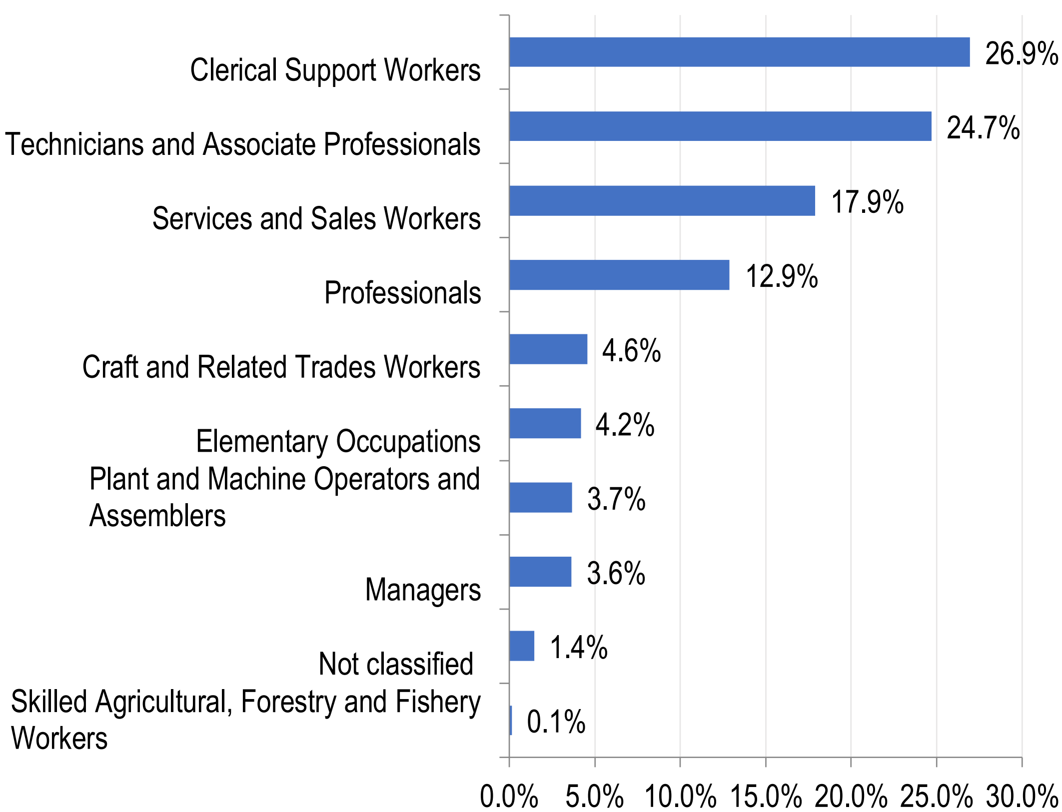
Source: Vacancy information 2016 - 2018. Own calculations.

[17] The text mining figures are presented in the Spanish language because it is the original language used on the Colombian job boards.

However, as mentioned in Cárdenas (2020c), the world cloud figures only acknowledge the most demanded job titles. For a further analysis of the job titles it is necessary to use the occupational classification (e.g. ISCO-08).

Thus, Figure 5 shows the distribution of job placements by their major occupational groups (aggregated at a 1-digit level ISCO-08). Around 26.9% of jobs demand "Clerical support workers", 24.7% demand "Technicians and associate professionals" and 17.9% required "Services and sales workers". Only 1.4% of job placements do not receive an occupational code ("not classified"). These missing values correspond to observations with not enough or a lack of useful information in the job title (for instance, "students", "part-time job", etc.).

Figure 5: Distribution of job placements by major occupational group ISCO-08



Source: Vacancy information 2016 - 2018. Own calculations.

Figure 5 is useful because it shows the general structure of the vacancy database and, potentially, the Colombian labour demand structure. Moreover, these results reaffirm what was stated previously: companies do not only search for high-skilled workers when using job portals.

However, as mentioned in Cárdenas (2020a), Big Data information unlocks the opportunity to monitor job requirements at the disaggregated level (e.g. 4-digits occupation level).

Thus, Table 3 shows the occupational structure (at a four-digit level) of the labour demand[18]. According to the vacancy data, the occupation most required in Colombia during 2016 to 2018 is "Commercial sales representatives" (15.4% of job placements), followed by "Telephone switchboard operators" (8.3% job placements) and "Stock clerks" (8.3% of job placements). These three occupations constitute around 32% of all job placements. Moreover, the most demanded Top 50 occupations form 78.2% of the Colombian labour demand. Consequently, according to the information from job portals, the occupations most required are related to sales, customer services, guards, and food preparation.

Another aspect to highlight is the presence of occupations related to technology and software development, such as "Information and communications technology user support technicians", "Information and communications technology operations technicians" and "Web and multimedia developers". This result confirms what it was mentioned in Cárdenas (2020a), the labour demand for those occupations has dramatically increased during the last years (Section 6 provides detailed evidence about labour demand trends).

Importantly, despite the potential theoretical bias of the information mentioned in Cárdenas (2020a), the results from Table 3 suggest (at least for the Colombian case) that job portals are not entirely focused on high-skilled occupations. Indeed, most of the categories listed in Table 3 are middle (such as "Sales demonstrators") or low-skilled occupations ("Kitchen helpers") which are expected results of a developing economy such as Colombia's.

Additionally, the Top 20 occupations most demanded in Colombia do not express any unusual results. Occupations which usually do not have a big market (such as "Actors") do not constitute a significant share of the Colombian labour demand. All the above results suggest that vacancy information from job portals might provide relevant information for a wide range of low-, middle- and high-skilled occupations.

[18] Given the number of occupational groups, the full list of occupations can be found in the Appendix A: Table A.1.

Table 3: Top 20 occupations most demanded in Colombia

Position	ISCO-08 code	Occupation	Number of jobs	Percentage
1	3322	Commercial sales representatives	878,503	15.4%
2	4223	Telephone switchboard operators	473,021	8.3%
3	4321	Stock clerks	472,076	8.3%
4	5223	Shop sales assistants	269,756	4.7%
5	5242	Sales demonstrators	235,481	4.1%
6	5230	Cashiers and ticket clerks	201,939	3.5%
7	4412	Mail carriers and sorting clerks	123,381	2.2%
8	5414	Security guards	111,717	2.0%
9	2411	Accountants	110,560	1.9%
10	1221	Sales and marketing managers	109,265	1.9%
11	4214	Debt-collectors and related workers	91,483	1.6%
12	9412	Kitchen helpers	75,535	1.3%
13	3343	Administrative and executive secretaries	73,364	1.3%
14	4110	General office clerks	69,875	1.2%
15	4322	Production clerks	67,997	1.2%
16	4311	Accounting and bookkeeping clerks	58,822	1.0%
17	8153	Sewing machine operators	54,628	1.0%
18	4222	Contact centre information clerks	50,337	0.9%
19	3312	Credit and loan officers	48,063	0.8%
20	5321	Health care assistants	45,279	0.8%

Source: Vacancy information 2016 - 2018. Own calculations.

As mentioned above, with the job vacancies categorised into occupations it is possible to identify the share of high-, middle- and low-skilled occupations demanded in Colombia. For instance, the OECD (2017c) defines the following as a high-skilled occupations (classified under the ISCO's major groups): 1) legislators, senior officials and managers, 2) professionals, and, 3) technicians and associate professionals; while middle-skilled jobs include: 4) clerks, 5) craft and related trade workers, and, 6) plant and machine operators and assemblers; and low-skilled include jobs: 7) service workers and shop and market sales workers, 8) agricultural and fishery workers, and, 9) elementary occupations.

Table 4 shows the distribution of jobs according to the above definitions: 22.5% (2,356,979) and 35.7% (2,011,352) of job placements correspond to low-skilled and middle-skilled occupations, respectively. While 41.8% of job placements are high-skilled occupations. It is important to notice that around 878,503 of job placements in the high-skilled group correspond to "Commercial sales representatives". Consequently, the high-skilled group is the most frequent due to the high demand for "Commercial sales representatives". Importantly, the results of Table 4 confirm that the vacancy information from job portals provide a high volume of information for low-, middle- and high-skilled occupations.

Table 4: Distribution of job placements by high-, middle- and low-skilled occupations

Classification	Number of jobs	Percentage
High Skill	2,356,979	41.8%
Middle Skill	2,011,352	35.7%
Low Skill	1,269,604	22.5%

Source: Vacancy information 2016 - 2018. Own calculations.

4.3 New or specific job titles

As pointed out in Cárdenas (2020a), the labour market changes rapidly and new occupations (or job titles) emerge or disappear over time. This thesis defines as "new or specific job titles" those titles that are not in the ISCO Colombian list of occupational titles. Consequently, new or specific job titles can correspond to new job titles or job titles that the ISCO Colombian list of occupational titles has not yet itemised.

As mentioned in Cárdenas (2020a), the early identification of these new labour demand has at least two economic benefits. On the one hand, it allows the curricula of training providers to adapt and, therefore, also adjusts people's skills to suit labour market changes. On the other hand, the identification of emerging patterns in labour demand might provide occupational classifications with real-time information. Consequently, statistics and public policy designs based on an adapted occupational classification might provide more precise results according to different regional and sectorial contexts[19].

Given that job portals generate detailed information on a daily basis, the systematic collection of data from these sources allows the identification of new job titles, and thus provides key information to identify new or emerging occupations. Table 5 presents the most recurrent new job titles in Colombia. It is worth noting the number of new job titles related to social networks and data management, such as "Cloud infrastructure engineer", "Professional SQA" (Software Quality Assurance/Advisor), "Influencer" (which is an industry expert who can influence other's behaviour through social networks, such as Twitter, Facebook, etc.), "Customer service social networks", "Big data specialist" and "Professor in Big Data", among others. However, it is not only in the IT sector that new job titles have emerged. Other job titles related to different activities have emerged: "Supervisor or specialist HSEQ" (Quality, Health, Safety & Environment), "Baristas" (a person specialised in high quality coffee, who creates new and different drinks based on their knowledge), and "Sellers TAT" (Store to store—people who are considered as brand managers, and promote and sell products to local mini-markets).

Interesting occupational titles involve CNC or bobcat operators. In the ISCO-08 occupational titles provided by DANE these job titles are not listed, neither in Spanish nor in the English language; however, these job titles are listed in the ISCO-08 UK version. This result shows that some countries might faster identify emerging occupations compared to other countries, or that the arrival of some technology occurs with certain delays for some developing countries such as Colombia.

In general, new job titles involve new tasks or the use of new technologies. For instance, CNC operators programme and operate manufacturing machines. One difference with other operators is that CNC operators need to programme CNC machines to produce elaborate pieces of work. In contrast, certain kinds of job might be of particular interest for Colombia. For instance, this country is well-known as a producer of high standard coffee, and a significant share of the Colombian economy depends on the performance of this product. Consequently, Baristas jobs

[19] Provided the relevance of this topic for policy and education, institutions such as O*NET have developed a methodology to identify, evaluate, and incorporate new and emerging occupations which have not yet been properly covered in the O*NET-SOC classification system (Dierdorff et al. 2009)

might be essential job opportunities for Colombian workers, especially for informal and unemployed people. Baristas differ from other barman and similar occupations because a Barista job requires a profound knowledge of high-quality coffee.

Thus, job portals are a rich source of changing information which requires the constant updating and adjusting of occupational classifications according to changes in the domestic labour market. To maintain an updated occupational classification requires the continuous monitoring of occupations and new job titles, and might improve labour market matching and, hence, tackle informality and unemployment rates (see, for instance, Cárdenas, 2020b).

Table 5: New job titles

Job titles	Number of jobs
Sellers TAT	52,849
Picking and Packing assistants	8,652
CNC operators	2,840
Supervisor or specialist HSEQ	2,349
Baristas	1,715
Community manager	1,550
NIIF Assistants, manager, or coordinator	1,532
Customer service social networks (Facebook, Twitter, etc.)	368
Cloud infrastructure engineer	169
SEO specialists	167
Maqueteador web (web layout designer)	142
Datacentre operator	125
SSTA inspector	49
Professional SQA	36
Influencer	23
Big data specialist	14
Professor in Big Data	12
Bobcat operators	11

Source: Vacancy information 2016 - 2018. Own calculations.

4.4 Skills most in demand (ESCO classifications)

As mentioned in Cárdenas (2020c), one of the most important characteristics of the vacancy database is that it might provide real-time and low-cost information about the skills most demanded in a particular economy. With the help of text mining techniques, it is possible to identify the skills explicitly demanded by employers according to the ESCO's categories (see Cárdenas, 2020c, Section 6.2). Of the 13,485 skills listed in the ESCO, 4,051 were found in the vacancy database. Around 84.6% of the job advertisements mentioned at least one word related to skill information. For illustrative purposes, Table 6 shows the Top 20 skills most in demand in Colombia. As can be seen, the skill most in demanded is "Customer service" (14.5% of job advertisements), followed by "Communication" (8.4%) and "Work in teams" (5.6%). Most of the skills in Table 6 are cross-sector skills (14 out of 20 skills), followed by sector-specific and transversal skills (e.g. "Work in teams" and "English").

Importantly, the results of Table 6 are consistent with the occupational structure of Colombia (Table 3), where the occupations most demanded are "Commercial sales representatives", "Telephone switchboard operators" and "Stock clerks". Consequently, it is to be expected that the most frequent skills required are related to customer services, communication, and customer insight, among other skills.

Table 6: Top 20 skills most demanded in Colombia

Skills	Level	Skill type	Number of jobs	Percentage
Customer service	Sector-Specific	Knowledge	827,705	14.5%
Communication	Cross-Sector	Knowledge	480,653	8.4%
Work in teams	Transversal	Skill/Competence	322,457	5.6%
Work in shifts	Cross-Sector	Skill/Competence	308,740	5.4%
Logistics	Cross-Sector	Knowledge	208,013	3.6%
Blueprints	Cross-Sector	Knowledge	169,579	3.0%
Telecommunication industry	Cross-Sector	Knowledge	114,998	2.0%
Mechanics	Cross-Sector	Knowledge	106,655	1.9%

English	Transversal	Knowledge	102,874	1.8%
Industrial engineering	Cross-Sector	Knowledge	99,976	1.7%
Manage personnel	Cross-Sector	Knowledge	96,579	1.7%
Customer insight	Sector-Specific	Knowledge	94,318	1.6%
Electronics	Cross-Sector	Knowledge	92,614	1.6%
Financial products	Cross-Sector	Knowledge	66,990	1.2%
Accounting	Cross-Sector	Knowledge	56,240	1.0%
Electricity	Cross-Sector	Knowledge	42,391	0.7%
Telecommunications engineering	Cross-Sector	Knowledge	38,967	0.7%
Sales activities	Sector-Specific	Knowledge	37,411	0.7%
Sales strategies	Sector-Specific	Knowledge	36,383	0.6%
Personal development	Cross-Sector	Knowledge	35,160	0.6%

Source: Vacancy information 2016 - 2018. Own calculations.

Although the ESCO's skills do not contain a full hierarchical structure, it is possible to group most of the 13,485 skills into broader categories: values, ICT safety, application of knowledge, digital communication and collaboration, language, digital data processing, health and safety, problem-solving with digital tools, transversal skills/competences, attitudes and values, social interaction, thinking, attitudes, digital competencies, numeracy and mathematics, working environment and digital content creation. This aggregation provides an overview of the general structure of demanded skills. Table 7 contains the aggregated results of the skills demanded in Colombia. Social interaction skills (such as work in teams, manage personnel, assist customers, etc.) are the most demanded group, followed by language (mainly English) and thinking skills (develop working procedures, plan teamwork, perform market research, among others).

Table 7: Skill groups demanded in Colombia

Broader skill categories	Number of jobs	Percentage
Social interaction	895,530	15.7%
Language	109,708	1.9%
Thinking	46,865	0.8%
Numeracy and mathematics	25,340	0.4%
Health and safety	24,640	0.4%
Attitudes and values	23,881	0.4%
Problem-solving with digital tools	20,088	0.4%
Working environment	9,070	0.2%

Source: Vacancy information 2016 - 2018. Own calculations.

4.5 New or specific skills demanded in the Colombia labour market

As mentioned in Cárdenas (2020c), the ESCO is a useful dictionary to identify the skills required in the labour market in Europe. However, this dictionary might not fully identify the skills demanded in the Colombian labour market because Colombian employers might demand different skills compared to Europe, and updating dictionaries to keep pace with changes in the labour market is challenging. Consequently, this thesis defines new or specific skills to address those skills that are not listed in the ESCO dictionary but are demanded in the Colombian labour market.

The identification of specific or new skills required in Colombia is relevant for tackling skill mismatch issues. With the implementation of new technologies, for instance, early identification of new skills required in the labour market might help people to adapt to Colombian-specific requirements and changes in the labour market and, hence, to reduce unemployment and to divert people from joining the informal sector.

For illustration purposes, Table 8 shows twenty new or specific skills demanded in Colombia. "Packing or Picking" is an Anglicism that describes the process of gathering and placing individual components of an order into a box or envelope addressed to a recipient. These words are an example of terms that the Colombian labour market uses, but are not incorporated into the ESCO dictionary. To identify these words in the vacancy database quantifies the real relevance of these logistic skills for the Colombian labour market. SAP (systems, applications and products) is an integrated business management system designed to model and automate different areas of a company. According to the vacancy database, the use of this technology is necessary for 21,378 jobs between 2016 and 2018. Importantly, employers ask for knowledge in Siigo and Helisa (accounting and administrative software for enterprises). Clearly, this Colombian-specific software is not in the ESCO dictionary because the technology was developed and in demand in Colombia.

It is important to highlight here, that some employer's requirements—such as knowledge in Cloudera, Fintech, Mailings (email marketing)—have recently increased. In 2016, this knowledge was not demanded; however, by 2017 and 2018 these requirements began to be required in the Colombian labour market. This example shows that the analysis of job portal information identifies changes in the labour demand for skills. Yet, it is also worth mentioning, that not only skill changes related to new technologies were found in the Colombian job market. For example, "Cosmetologia" (cosmetology), "*Perifoneos*" (to promote a product or business on the street with the help of a microphone), "*Brandeo*" (an Anglicism of branding) are skills mentioned in the vacancy database that are not listed in the ESCO dictionary. Thus, job portals are a rich source of information to identify new or specific skills which helps to update skills dictionaries such as the ESCO and improve educational and training systems to meet specific requirements and changes in the domestic labour market.

Table 8: Twenty new or specific skills demanded in Colombia

Skills	Number of jobs
Packing or Picking	67,493
Sap	21,378
Siigo	11,784
Pdv	7,360
Helisa	6,024
Scrum	4,219
Cosmetologia	3,201
Apm	878
Perifoneos	858
Mailings	536
Staad	336
Otdr	228
Rph	195
Kaizen	177
Fintech	176
Brandeo	149
Cloudera	138
Bigip	130
Rpgii	110
Ssst	98

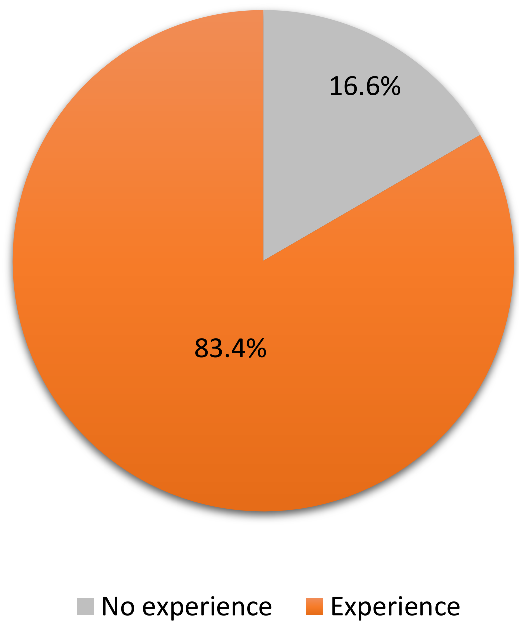
Source: Vacancy information 2016 - 2018. Own calculations.

4.6 Experience requirements

Regarding the experience requirements that employers' are looking for in Colombia, 83.4% of the job placements explicitly require people with some work experience (Figure 6). This result indicates that labour experience is an essential characteristic that workers need to apply for most Colombian vacancies. While the median year for required experience is one year, it is important to note that this variable contains a significant portion of missing values in the database. Indeed, 44.4% of the job placements that require some job experience, but do not report the specific years of work experience required[20].

[20] For instance, an employer might post a job advertisement in the following way: "a person with experience in photography is required to…". The variable "years of experience" can be imputed with the techniques explained in Cárdenas (2020c), and this imputation process will be a part of future research.

Figure 6: Job placements by experience requirements



Source: Vacancy information 2016 - 2018. Own calculations.

5. Demand by sector

As mentioned in Cárdenas (2020c), the vacancy database might provide information about job placements by company name and, hence, by different sectors as company names in job portals refer to the company who posted the job advertisement. An analysis of the vacancies by sector might serve to identify which skills or occupations are sector-specific or generic, which helps to address labour supply according to the needs of each industry. Thus, Table 9 shows the distribution of job placements by sector. More than half of the job placements (around 55%) were coded according to ISIC revision 4 (division groups).

On the one hand, companies related to "Administrative and support service activities" posted around 36.2% of the job positions, followed by "Wholesale and retail trade; repair of motor vehicles and motorcycles" (5.9%), and "Professional, scientific and technical activities" (2.4%). The "Administrative and support service activities" category contains most of the job placements because this group includes companies related to "Temporary employment agency activities" and the "Activities of call centres". Temporary employment agencies act as a third party (intermediary) between companies and employees. They collect CVs and make their client's (employers) vacancies public. Consequently, if a vacancy is posted on job portals and the company's name refers to a temporary employment agency this does not mean that potential employees will work in the "Administrative and support service activities". People who apply for those kinds of vacancies might end up working in other sectors (e.g. manufacturing) (Perhaps, information about the company's name is in the description rather than the company's name variable. Thus, processing and identifying specific patterns in the job description might increase the number of observations of where people will work. However, this further development will be part of future work).

On the other hand, it is expected that companies related to "Activities of call centres" have a considerable share of all job placements. As shown in Table 3, there is a high demand for "Telephone switchboard operators" among other related workers. Indeed, the results of Table 9 correlate with the results of Table 3 as the sectors with the highest number of job placements are related to the most demanded occupations. For instance, in Table 3 the occupations most required are related to "Sales", "Customer services", "Accountants", and "Production clerks", while the sectors with more job placements (apart from administrative and support service activities) are the "Wholesale and retail trade", "Manufacturing", "Financial and insurance activities".

Another aspect to highlight is the relatively low frequency of job placements from sectors such as "Agriculture, forestry and fishing", "Public administration" and "Defence companies", etc. It was expected that these sectors would not have a high participation in the vacancy database because job portals (at least in Colombia) do not adequately cover rural zones where most of agriculture, forestry and fishing companies operates, and, in addition, job portals are not a usual channel for posting vacancies related to public administration and defence, water supply, sewerage, and waste management, among other activities. Instead, these vacancies are advertised on the website of individual companies, and the scraping of that information will be a part of future work.

The issue of missing values in Table 9, and the high participation of temporary employment agency activities might make it difficult to estimate the current level of labour demand by sector. Instead, job portal information might be more useful for the identification of skills and possible skill shortages by industry. As can be observed in Table 9, there are a considerable number of observations for most sectors. This information might provide valuable insights regarding the most demanded generic and sector-specific skills and trends in the labour market by industry (see Cárdenas, 2020d).

Table 9: Job placements by sector

ISCO rev4	Number of jobs	Percentage
Administrative and support service activities	2,070,156	36.2%
Wholesale and retail trade; repair of motor vehicles and motorcycles	338,387	5.9%
Professional, scientific and technical activities	136,955	2.4%
Manufacturing	98,359	1.7%
Financial and insurance activities	97,351	1.7%
Construction	84,935	1.5%
Information and communication	74,502	1.3%
Transportation and storage	67,038	1.2%
Accommodation and food service activities	48,192	0.8%
Human health and social work activities	22,831	0.4%
Other service activities	14,661	0.3%
Arts, entertainment and recreation	13,099	0.2%
Education	11,552	0.2%
Real estate activities	6,205	0.1%
Agriculture, forestry and fishing	4,101	0.1%
Water supply; sewerage, waste management and remediation activities	3,861	0.1%
Public administration and defence; compulsory social security	425	0.0%
Electricity, gas, steam and air conditioning supply	308	0.0%
Activities of households as employers	6	0.0%
Not coded	2,627,589	45.9%
Total	5,720,513	

Source: Vacancy information 2016 - 2018. Own calculations.

6. Trends in the labour demand

Although analysing the structure of labour demand is vital to know the kind of human resources required by employers, this analysis might not be sufficient to improve skills matching in the labour market if trends, seasonal changes, and business cycles, are overlooked. The labour demand for certain occupations might increase over specific periods (i.e. quarters). For instance, in holiday periods the need for "Hotel receptionist" might increase due to an increase in tourism. Moreover, the labour market is dynamic, and the labour demand for certain occupations or skills might increase/decrease over time. The analysis of labour demand cycles, seasons and trends is of paramount importance because it enables the curricula of training providers to adapt, and train people in the required skills for technological change, business cycles, etc.

Given that the Colombian vacancy information was collected with the same techniques (Cárdenas, 2020c) across the same time period, it is possible to analyse trends and seasons in the labour demand to demonstrate the potential of this analysis and the consistency of the vacancy database results. Table 10 shows the distribution of vacancies and job positions across the period of analysis (2016-2018). In 2016, the total number of vacancies and job positions was 688,477 and 1,746,762, respectively. In 2018, the total number of vacancies and job position was 818,160 and 2,073,726, respectively. Consequently, the number of vacancies and the total number of jobs increased from 2016 to 2018, by about 15.8% and 15.7%, respectively. This increase in the number of job advertisements might correspond to Colombian economic growth and the extended use of job portals to advertise job positions[21].

Table 10: Yearly distribution of vacancies and job positions

Year	Total vacancies		Total jobs	
	Number	Percentage	Number	Percentage
2016	688,477	30.63%	1,746,762	30.54%
2017	741,322	32.98%	1,900,025	33.21%
2018	818,160	36.40%	2,073,726	36.25%
Total	2,247,959		5,720,513	

Source: Vacancy information 2016 - 2018. Own calculations.

Figure 7 shows the Colombian labour demand during the period of analysis for major occupational groups (one-digit level ISCO-08)[22]. Most of the major groups demonstrate an

[21] Cárdenas (2020d) provides more detailed evidence regarding this discussion.

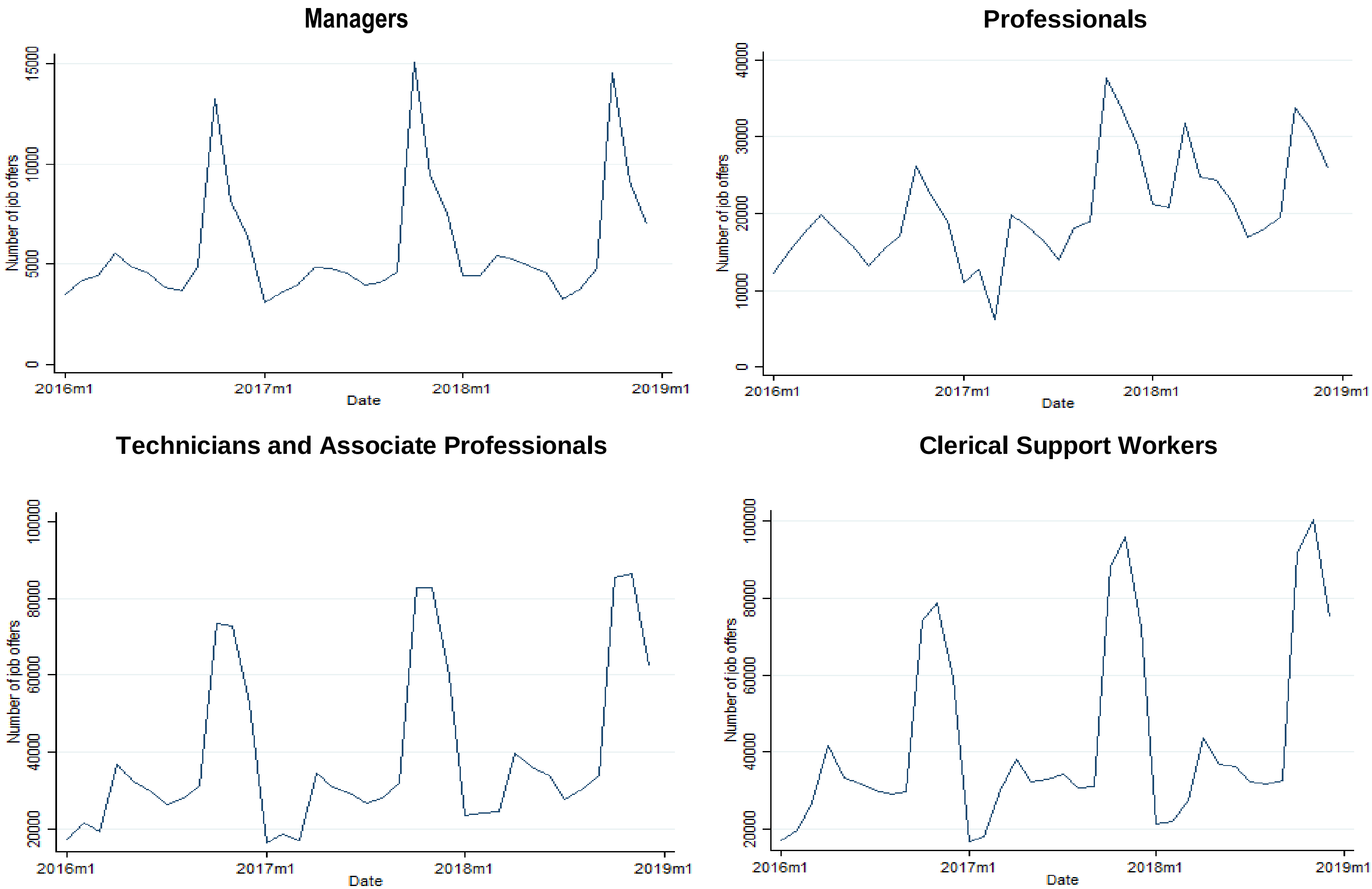
[22] The labels "2016m1", "2017m1", etc., on the x-axis correspond to January 2016, January 2017, and so on.

increase in labour demand between October and December and a substantial decrease in demand between January and March[23].

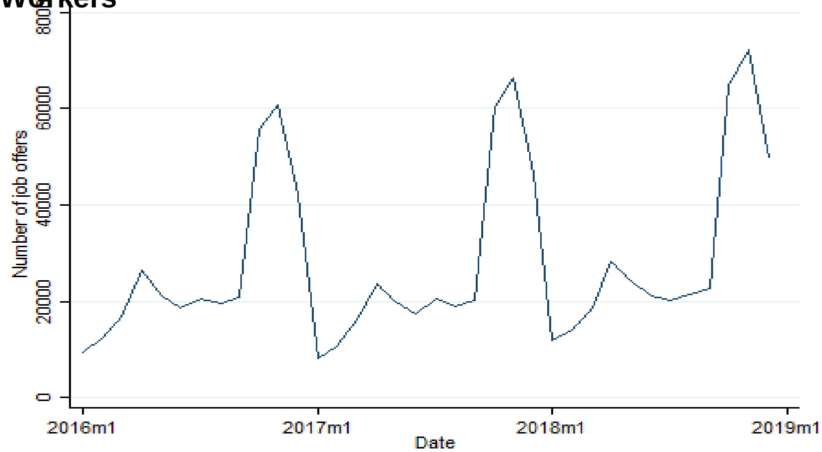
In contrast, the labour demand for “Professionals” grew during the period of analysis. As mentioned in Subsection 4.2, aggregated results are useful because they provide an idea of global labour demand behaviour. However, analysing the results in a disaggregated way over time (for instance at the four-digit level ISCO-08) produces summary measures of trends and amplifies labour demand behaviour.

[23] As will be seen in Cárdenas (2020d) in more detail, this cyclical behaviour correlates with the official unemployment statistics provided by DANE: demonstrating that unemployment rates are relatively low between October and December, and higher between January and March. This result is due to companies hiring people for the December season (when formal workers usually receive a Christmas bonus) and tourism, among other economic activities, significantly increases.

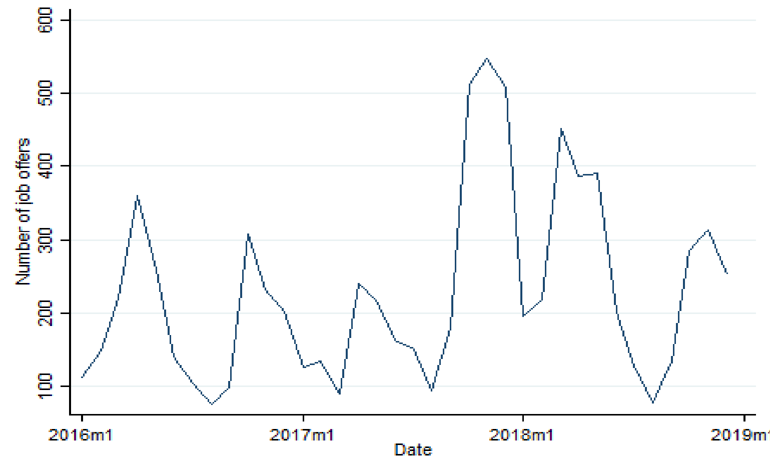
Figure 7: Trends of the labour demand by major occupational ISCO groups



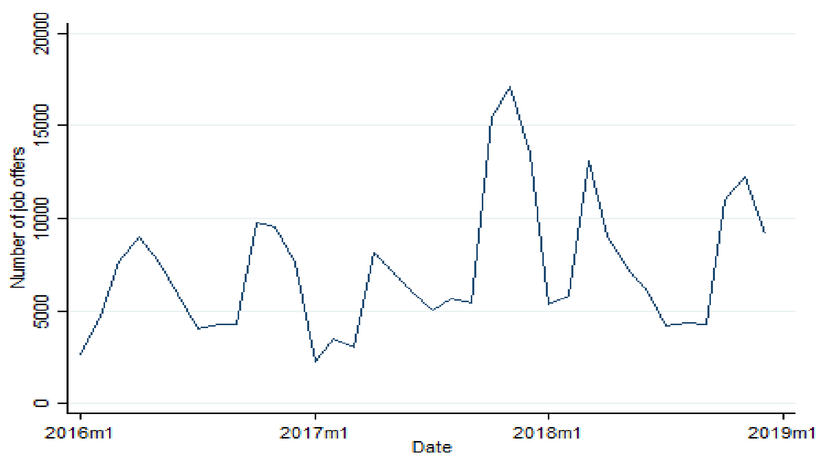
Services and Sales Workers



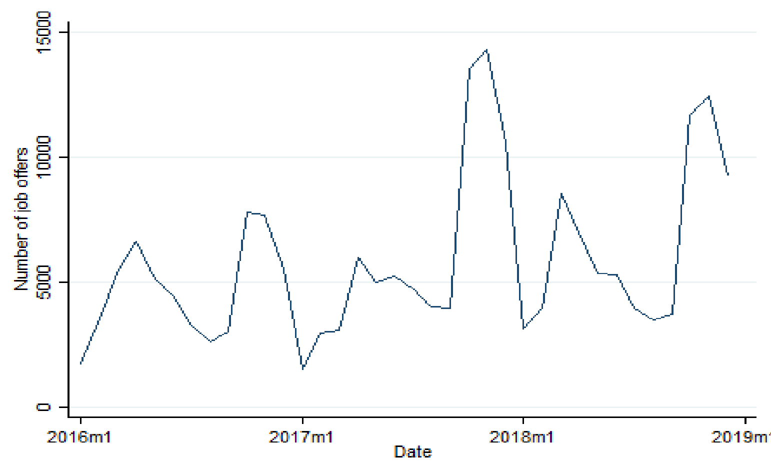
Skilled Agricultural, Forestry and Fishery Workers



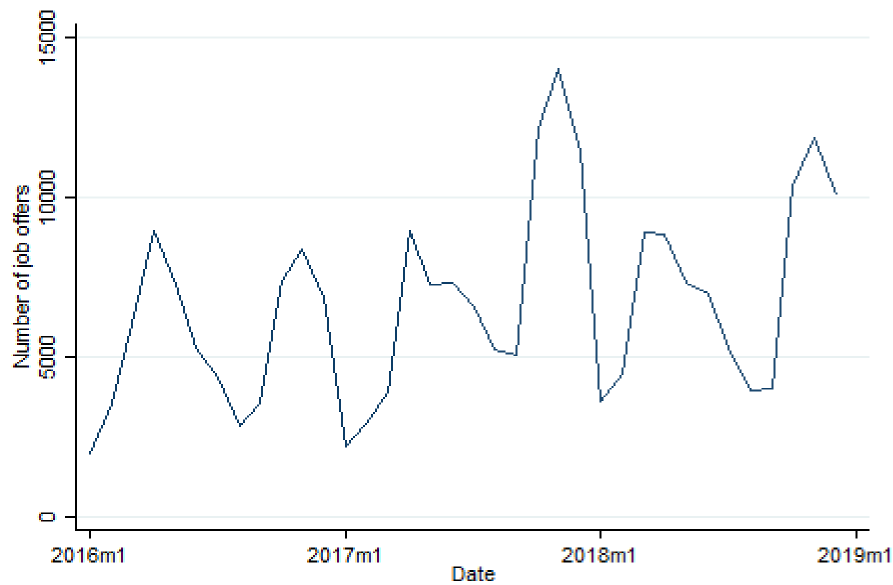
Craft and Related Trades Workers



Plant and Machine Operators and Assemblers



Elementary Occupations

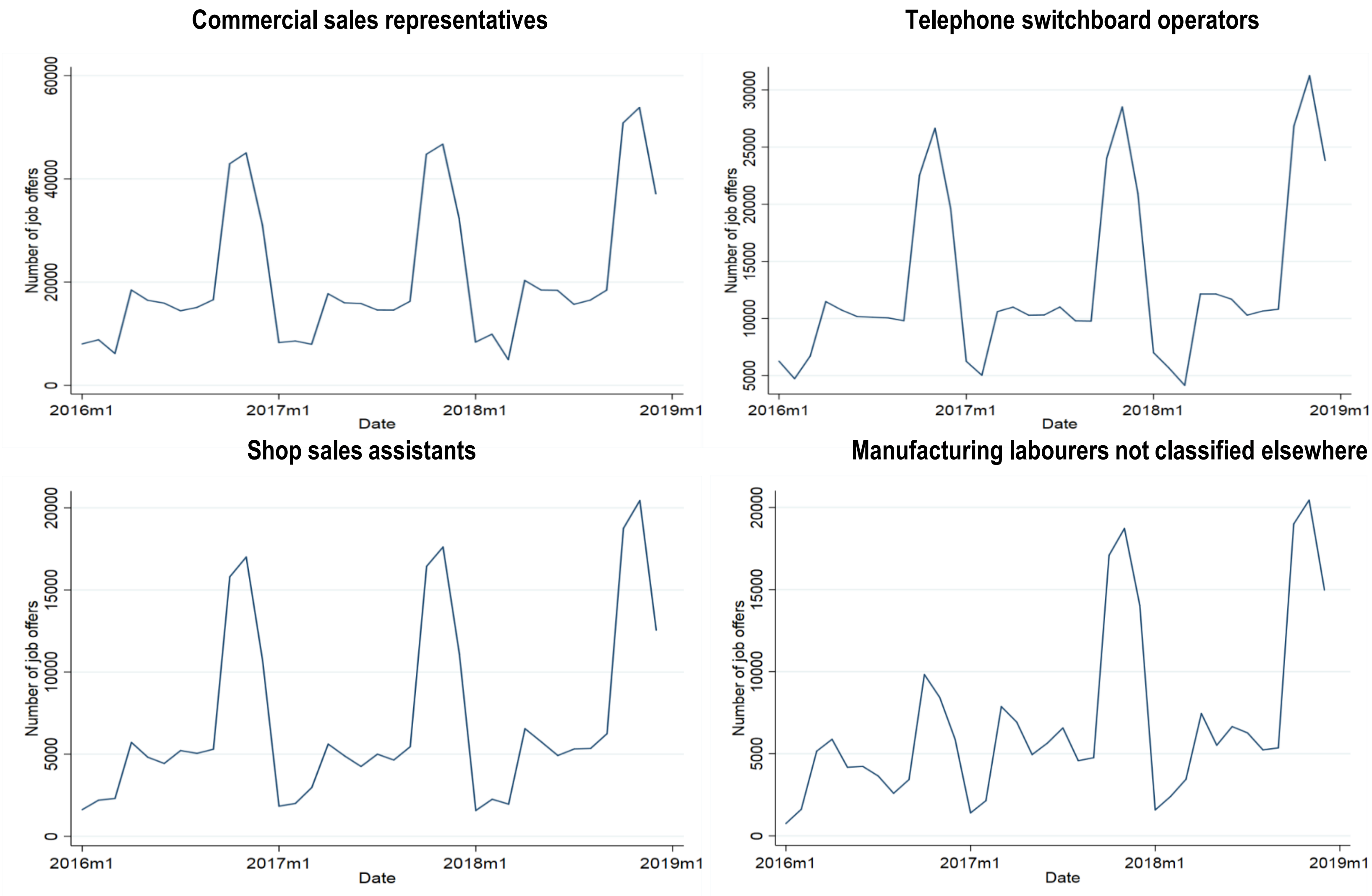


Source: Vacancy information. Own calculations.

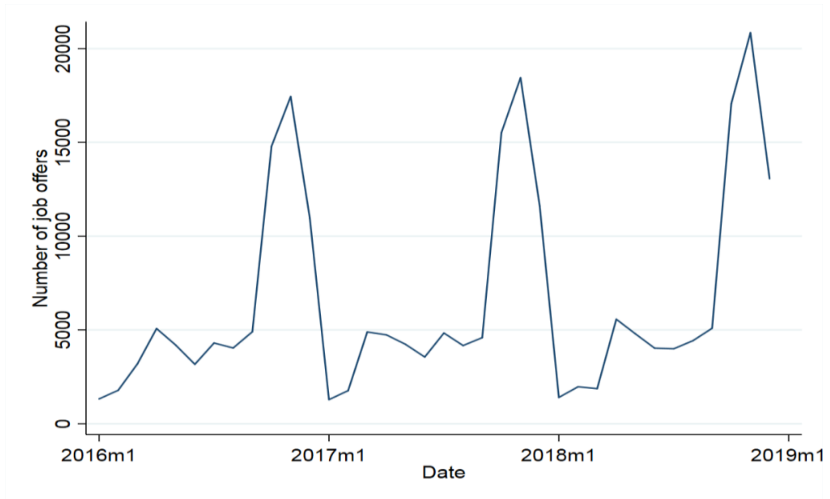
Figure 8, Figure 9 and Figure 10 show the most notable trends by occupational groups (the graphs for each occupational group [304] are available upon request). The charts are divided into three groups: Figure 8 shows the trends of occupations with a higher demand; Figure 9 plots occupations with a significant increase of labour demand during the period of analysis, and Figure 10 displays occupations whose demand has decreased.

As can be observed in Figure 8, occupations with relatively more demand tend to have a similar cyclical pattern over time: a remarkable increase of labour demand between October and December, and a sharp decrease of labour demand between January and March. Alternatively, the labour demand for occupations in Figure 8 slightly increased during the period of analysis. In addition, there are also occupations that always have a high demand and, thus, do not exhibit a significant increase in the last quarter of the year and a decrease in the first quarter of the year. For instance, the labour demand for occupations such as “Accounting and bookkeeping clerks”, “Credit and loan officers”, “General office clerks” and “Contact centre information clerks”, generally increase in the first and sometimes in the last quarter of the year.

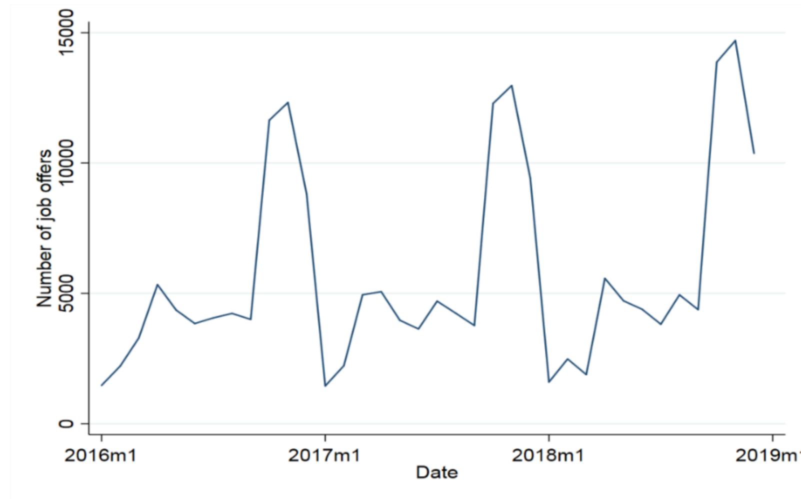
Figure 8: Trends of the most demanded occupations at a four-digit level



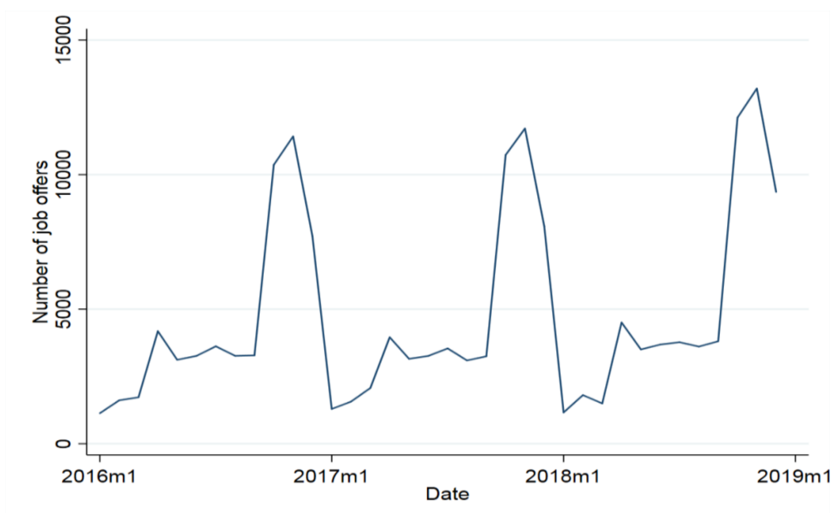
Sales demonstrators



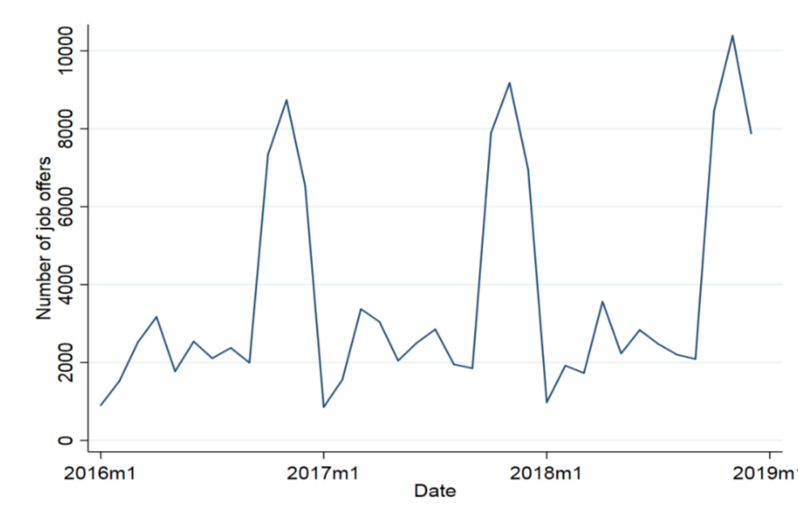
Cashiers and ticket clerks

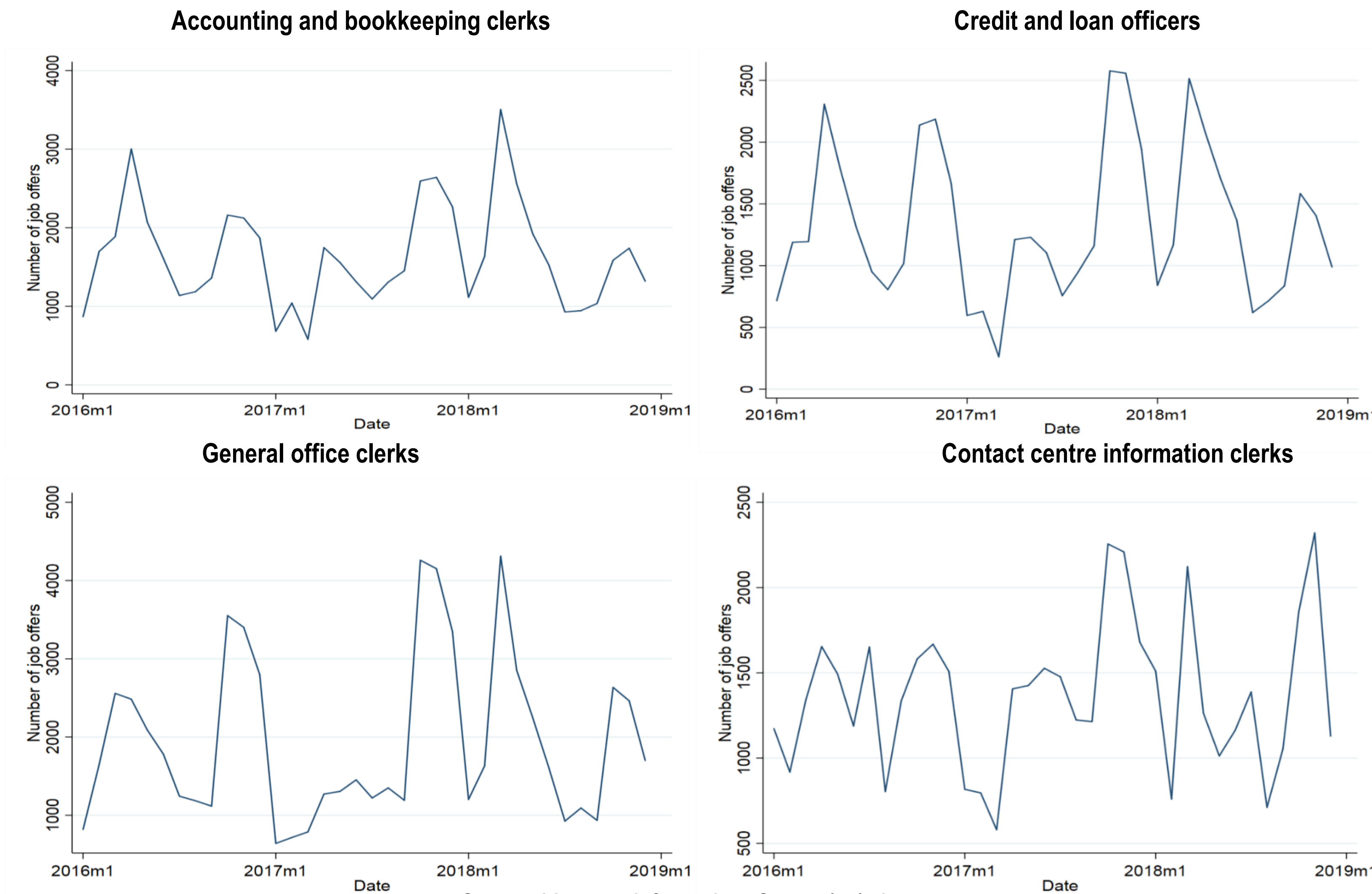


Stock clerks



Management and organisation analysts

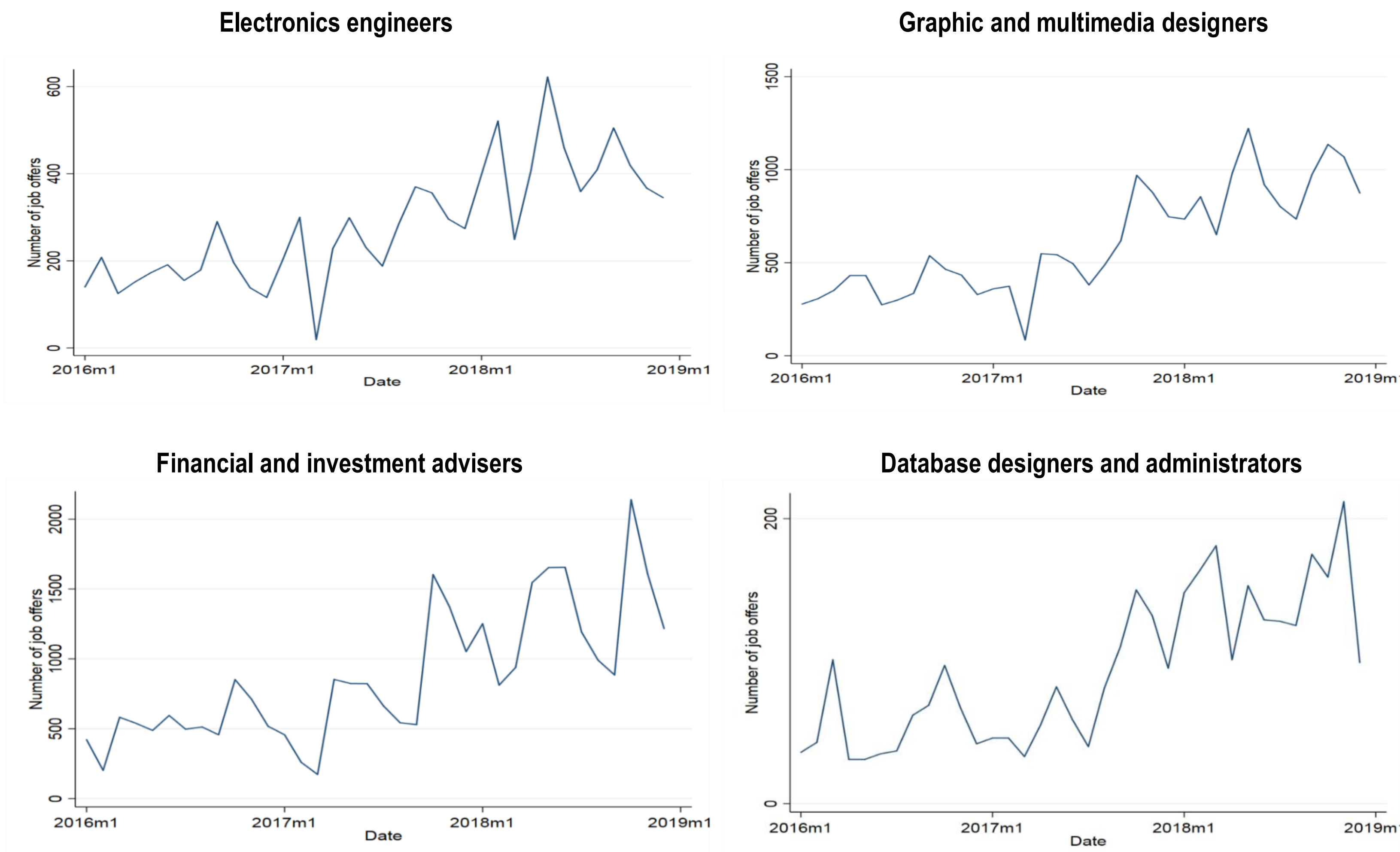




Source: Vacancy information. Own calculations.

Figure 9 plots occupations with a significant increase in labour demand. Despite the relatively short period of analysis (three years), it is possible to observe a growing trend of demand for "Electronics engineers", "Graphic and multimedia designers", "Financial and investment advisers", "Database designers and administrators", "Computer network professionals", "Electronics engineering technicians", "Real estate agents and property managers", and "Information and communications technology user support technicians", among others. These results suggest that in Colombia the labour demand for occupations related to technology, finance and the real estate market is either rapidly growing, or the companies that demand those occupations have increased their use of job portals.

Figure 9: Occupations at a four-digit level with a positive trend

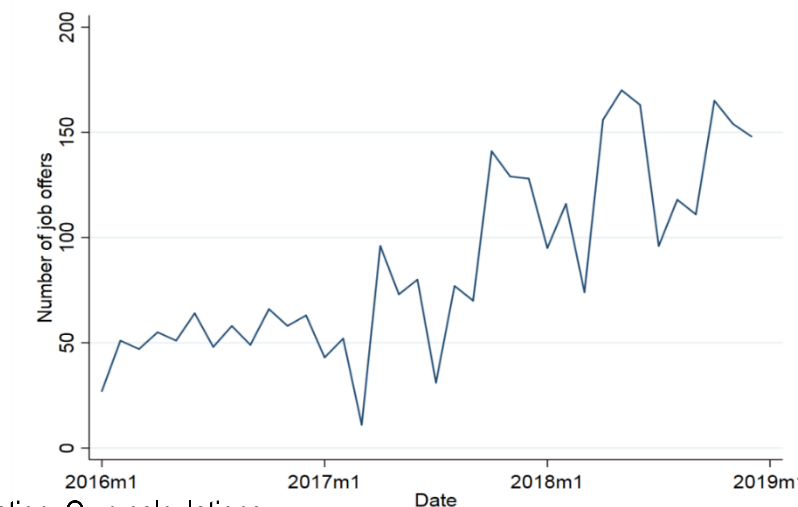


Computer network professionals

Electronics engineering technicians

Real estate agents and property managers

Information and communications technology user support technicians



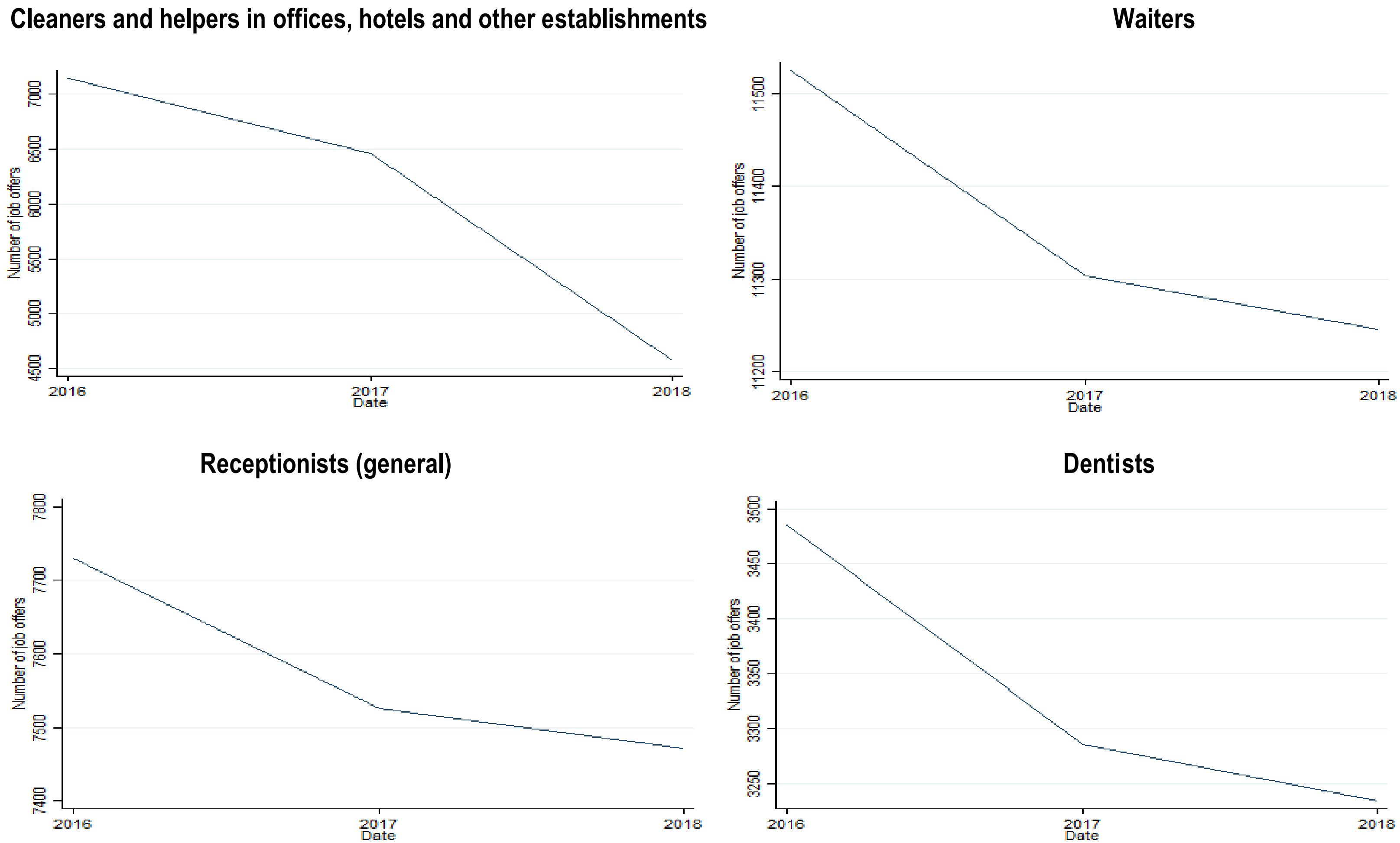
Source: Vacancy information. Own calculations.

Conversely, the labour demand for some occupations has decreased over time. As can be observed in Figure 10, the demand for occupations such as “Cleaners and helpers in offices, hotels and other establishments”, “Waiters”, “Receptionists (general)”, “Dentists”, among others, has decreased from 2016 to 2018. For instance, the labour demand for “Cleaners and helpers in offices, hotels and other establishments” decreased from 7,546 job placements in 2016 to 4,622 job placements in 2018. However, overall there are relatively few occupations for which demand has decreased over time. Additionally, the figures in Figure 10 show that there is not a dramatic decrease in labour demand for specific occupational groups (the results were grouped yearly in order to observe a clearer pattern). These results contrast with Figure 9 where the labour demand for certain occupational groups has dramatically increased.

Two factors might explain the relatively high increase and the slight decrease of labour demand for particular groups of occupations during the period of analysis: First, as mentioned in Cárdenas (2020a), the use of job portals (and Internet use, in general) has increased over the last decades. Consequently, as the number of job portal users increases, so will the number of vacancies posted on the Internet. However, it is interesting to note that the labour demand for certain occupations has decreased despite the increase in job portal usage over time. Thus, the increase of Internet usage might soften the fall of job placements for particular occupations, while this phenomenon intensifies the rise in job placements for other occupations.

Second, it has been widely reported that, over the last decades there has been a skill-based technological change which has increased labour demand and wages for skilled labour (Autor et al. 1998). Thus, the remarkable labour demand increases for occupations such as “Graphic and multimedia designers”, and “Computer network professionals”, among others (Figure 9), is a product of this technological change (i.e. structural change). Nevertheless, the “destruction” of labour demand for certain occupations is a process that might require a relatively long period. For instance, companies might adopt technologies that replace some human labour; however, to make this transition requires a considerable interval. Companies need time to adapt their production process to the new technologies, and legislation exists that protects job positions against (massive) layoffs or other abrupt changes. Thus, falls in the labour demand by occupation might not occur abruptly.

Figure 10: Occupations at four-digit level with a negative trend



Source: Vacancy information. Own calculations.

7. Wages

The analysis of the number of jobs posted in an economy is necessary, but not enough to determine if skill mismatches can be reduced. Jobs can be available; however, the wages of those jobs might not be high enough to create a labour supply to satisfy labour demand. This variable helps to investigate whether the vacancies posted on job portals can offer wages to attract informal workers and the unemployed into formal jobs and, at the same time, helps to determine possible skill mismatches (see Cárdenas, 2020b).

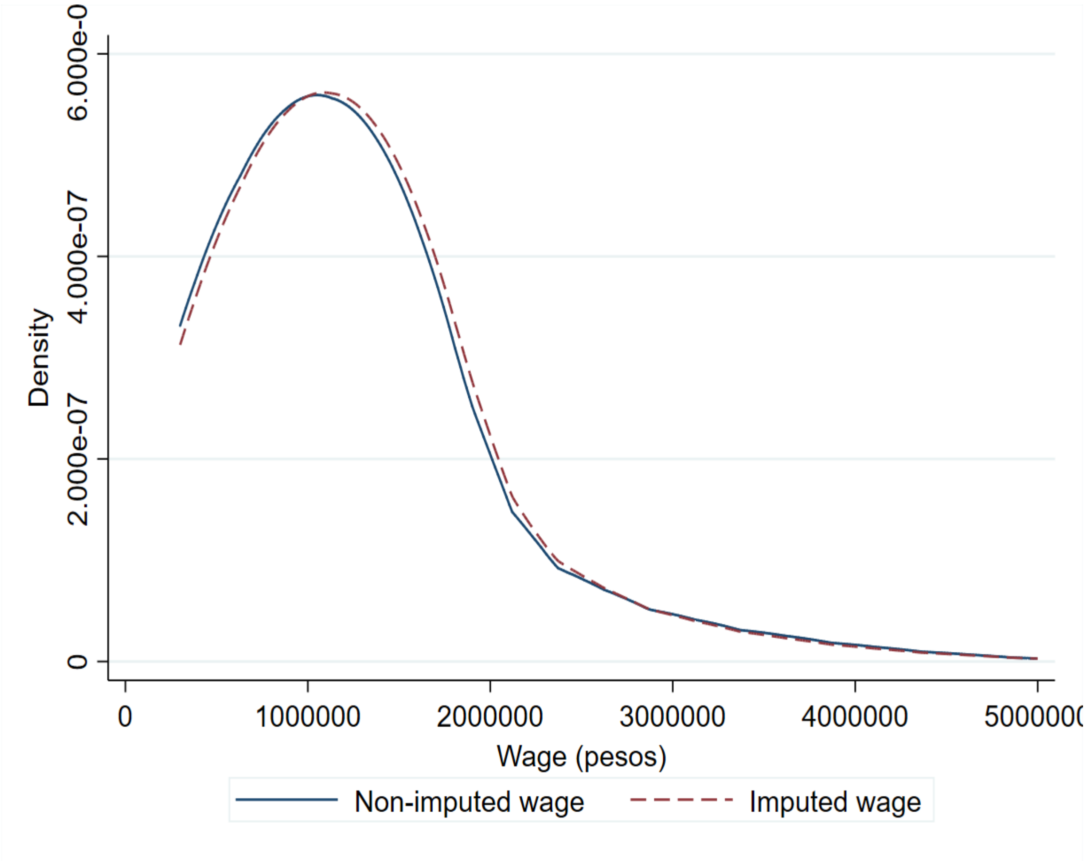
Figure 11 shows the distribution of monthly wages from the vacancy database. The solid blue line represents the "wage" variable without any imputation process, while the dashed red line represents the "imputed wage" variable (see Cárdenas, 2020c). Both the imputed and non-imputed wage variable have a similar distribution because most jobs pay a salary between the minimum wage[24] and 1,500,000 pesos (around £375). Indeed, the average figures for the non-imputed wages and imputed wages are 1,059,667 pesos (around £265) and 1,102,200 pesos (around £275). These results reveal two facts. First, differences between the non-imputed wage and the imputed wage are minimal. Consequently, to use imputed wages in Cárdenas (2020b, 2020d) does not add significant noise or bias to the statistical analysis and, on the contrary, it enables an analysis of all vacancy observations. Second, the distribution of wages is consistent with the results from the previous sections: a high proportion of jobs correspond to low- and middle-skilled occupations. Hence, the data are expected to have a right-skewed distribution (a high concentration of low wages) as in Figure 11[25].

---

[24] In 2016, 2017 and 2018 the Colombian minimum wage was 689,454 Colombian pesos (around £170), $737,717 (around £184) and $781242 (around £195), respectively.

[25] Cárdenas (2020d) provides more evidence about the consistency of the wage variable.

Figure 11: Wage density



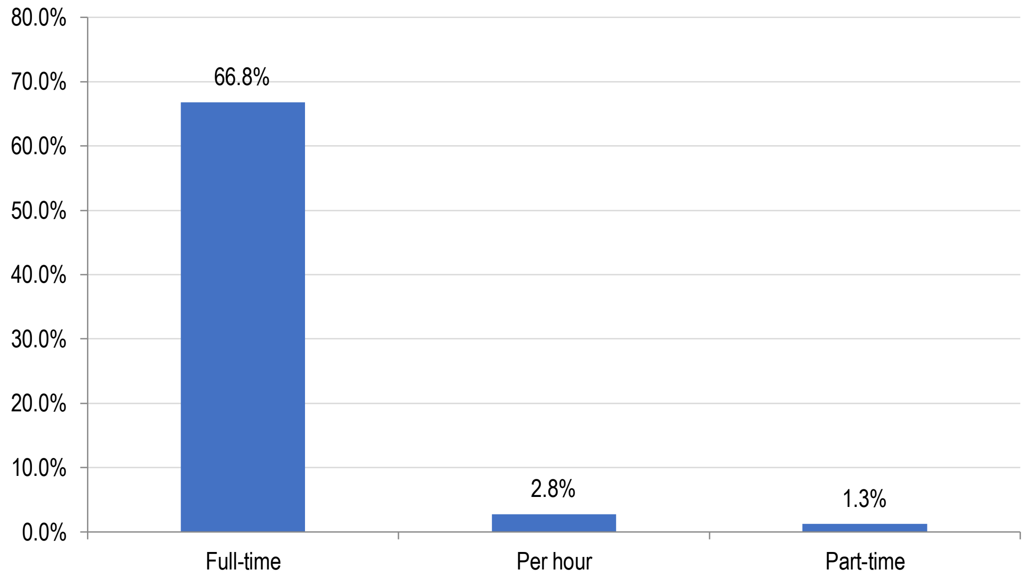
Source: Vacancy information 2016 - 2018. Own calculations.

8. Other characteristics of the vacancy database

As mentioned in Cárdenas (2020c), job portal information is a rich source for identifying different characteristics of labour demand. Some of those characteristics are not directly related to the labour demand for skills. However, this non-related skill information might provide more evidence regarding the consistency of the vacancy database, and it might be useful to tackle skill mismatches for a specific population or type of jobs. For illustration purposes, this section presents some of the most relevant characteristics of the vacancy database that are not directly related to skills information, such as type of contract offered and vacancy duration.

Figure 12 shows the distribution of jobs by type of contract. It is worth mentioning that some employers do not mention the kind of contract being offered. Moreover, and unlike the "education" variable, the variable "type of contract" is not imputed. Consequently, the sum of the percentage in Figure 12 is less than 100%: around 68.8% of jobs available offer a full-time contract, while 2.8% and 1.3% of jobs available offer a per-hour and part-time contract, respectively. This result suggests that job portal information is not biased towards “irregular” jobs such as part-time work or per-hour jobs.

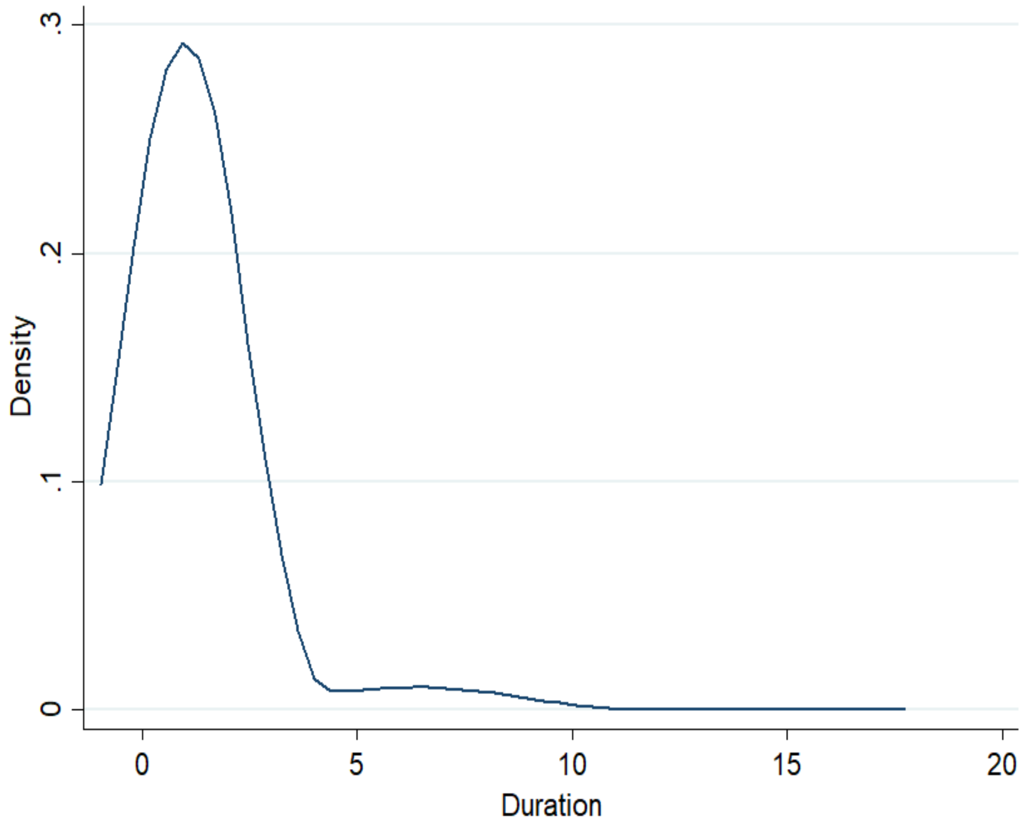
Figure 12: Jobs by type of contract



Source: Vacancy information 2016 - 2018. Own calculations.

Figure 13 shows the duration of job vacancy advertisements. This variable is the difference between the publication and the expiration date provided by the employers in the job advertisement. The median advertising duration is 1.2 months. However, it is important to mention that 73% of observations do not have information regarding their publication or expiration date. Despite the missing values, the results do not have atypical values. This result reaffirms that information provided by employers is consistent, and the problem of atypical or wrong values is minimum.

Figure 13: Duration density (monthly)



Source: Vacancy information 2016 - 2018. Own calculations.

9. Conclusion

The information from job portal information has attracted the attention of researchers and policymakers, since Big Data seem to provide quick and relatively inexpensive access to analyse information about employers' requirements. Currently, for countries such as Colombia job portals are a unique source of labour demand information.

Much has been said about the advantages and limitations of using this information for labour demand analysis (see Cárdenas, 2020a). For instance, given the online nature of these sources of information, job portals data might be biased towards high-skilled occupational groups. Nevertheless, most of the studies which have used job advertisements (printed or online) do not discuss the reliability of this information for labour demand analysis and public policy design (see Cárdenas, 2020a). This paper provides a descriptive analysis to start evaluating the results from the vacancy database, and its usefulness for tackling informality and unemployment problems in Colombia.

The sample period runs from 1st January 2016 to 31st December 2018. The main results of my analysis of the vacancy database show that 1) job vacancies are concentrated in Bogotá, Antioquia and Bolivar. These results are in agreement with other macroeconomic outputs. For instance, the capital (Bogotá) and its surrounding counties have the highest population and GDP rates; 2) most of the job positions require a person with at least a high school certificate; 3) in concordance with the previous result, most occupations in Colombia correspond to middle- ("Sales demonstrators") and low-skilled occupations ("Kitchen helpers") which are expected results from a developing economy such as Colombia; 4) this result also suggests that the job portals selected in Cárdenas (2020c) are not biased to a specific market (e.g. high-skilled jobs, such as managers or professionals); thus, 5) job portals are a rich source of information to continuously update occupational classifications according to changes in the domestic labour market. For instance, among the most relevant new job titles found in the vacancy database are "Sellers TAT", "CNC operators" and "Baristas".

In addition, regarding skill information in the vacancy database: 6) the analysis shows that the skills most demanded in the Colombian labour market include "Customer service" (knowledge), "Communication" (knowledge) and "Work in teams" (competence), which is consistent with the occupation demanded; 7) it is possible to identify new or specific skills such as Fintech, Mailings, and "*perifoneos*" among others. Thus, it is possible to monitor the changes and the specific requirements of the domestic labour market at a low cost by using job portal information, as with a single database (vacancy) it is possible to analyse job attributes (of occupations in demand) and workers' skills requirements.

Moreover, the issue of missing values in the sector variable and the high participation of "Temporary employment agency activities" 8) might make it difficult to estimate the current level of labour demand by sector. Nevertheless, job portal information might be more useful for the identification of skills and possible skill shortages by sector.

Despite increased job portal usage, 9) it is possible to observe clear trends and seasons in labour demand: for instance, labour demand for certain occupations peaks in the last quarter of the year, and the labour demand for occupations related to IT and other technologies is growing. The labour demand for other occupations (such as "cleaners and helpers in offices, hotels and other establishments", "Waiters", and "Receptionists") decreased during the period of analysis.

The results regarding wages show two facts: 10) the differences between non-imputed wage and imputed wage distributions are minimal. Consequently, to use imputed wages in Cárdenas (2020b, 2020d) does not add significant noise or bias for statistical analysis and, on the contrary, it allows analysing all of the vacancy observations; 11) the distribution of wages is consistent with occupations in demand. A high proportion of jobs correspond to low- and middle-skilled occupations and the distribution of wages is right-skewed (a high concentration of low wages).

The analysis of other characteristics of the vacancy database that are not directly related to labour demand for skills shows two facts: 12) information provided by employers is consistent, meaning that issues such as outliers in the wage or vacancy duration variables are minimal; and, 13) the vacancy database can provide different information such as what skill is most demanded by occupation, trends and seasonal changes in labour demand, which might serve as an input to tackle skill shortages for a certain population sample or certain type of jobs (see Cárdenas, 2020b).

In general, the vacancy database provided detailed, real-time and valuable information about the Colombian labour demand that, previously, it was not possible to obtain from other sources (e.g. household surveys). Moreover, these initial results suggest that the vacancy database is consistent, or at least it does not contradict itself or external data, such as regional GDP, population, etc. However, a more detailed examination is necessary to draw conclusions about the reliability and the representativeness of this vacancies data.

10. References

Autor, D. H., Katz, L. F., & Krueger, A. B. (1998). Computing inequality: have computers changed the labor market? The Quarterly Journal of Economics, 113(4), 1169-1213.

Cárdenas R., Jeisson. (2020a). Information Problem in Labour Market and Big Data: Colombian Case. Universidad del Rosario. Working, paper No. WP2-2020-001.

Cárdenas R., Jeisson. (2020b). Possible uses of labour demand and supply information to reduce skill mismatches. Universidad del Rosario. Working, paper No. WP2-2020-002.

Cárdenas R., Jeisson. (2020c). Extracting value from job vacancy information. Universidad del Rosario. Working, paper No. WP2-2020-003.

Cárdenas R., Jeisson. (2020d) Internal and external validity of the vacancy database. Universidad del Rosario. Working Paper No. WP2-2020-005.

DANE (2017b). Producto Interno Bruto (PIB) Departamental. Available at: https://www.dane.gov.co/files/investigaciones/pib/departamentales/B_2015/Bol_dptal_2017preliminar.pdf [Accessed 28 Dic. 2018].

Dierdorff, E. C., Norton, J. J., Drewes, D. W., Kroustalis, C. M., Rivkin, D., & Lewis, P. (2009). *Greening of the world of work: Implications for O* NET®-SOC and new and emerging occupations*. O* NET, February.

MEN (2016). *Ministerio de Educación Nacional.* Decreto No. 1001. Available at:https://www.mineducacion.gov.co/1621/articles-96961_archivo_pdf.pdf [Accessed 10 Jan. 2019].

OECD (2017c). *OECD Employment Outlook 2017.* OECD Publishing, Paris. http://dx.doi.org/10.1787/empl_outlook-2017-en

OEI (1993). *Sistemas Educativos Nacionales. Principios y estructura del sistema educativo.* Available at: https://www.oei.es/historico/quipu/colombia/ [Accessed 10 Jan. 2019].

Valencia, F., Suarez, C., Rocha, C., & Mora, D. (2016). *Composición de la economía de Bogotá. Revista del Banco de la República*, 89(1069), 11-36. Recuperado de lahttps://publicaciones.banrepcultural.org/index.php/banrep/article/view/8010

Appendix A: Additional tables

Table A.1: Occupations demanded in Colombia

Position	ISCO 08 code	Occupation	Number of jobs	Percentage
1	3322	Commercial sales representatives	878,503	15.4%
2	4223	Telephone switchboard operators	473,021	8.3%
3	4321	Stock clerks	472,076	8.3%
4	5223	Shop sales assistants	269,756	4.7%
5	5242	Sales demonstrators	235,481	4.1%
6	5230	Cashiers and ticket clerks	201,939	3.5%
7	4412	Mail carriers and sorting clerks	123,381	2.2%
8	5414	Security guards	111,717	2.0%
9	2411	Accountants	110,560	1.9%
10	1221	Sales and marketing managers	109,265	1.9%
11	4214	Debt-collectors and related workers	91,483	1.6%
12	9412	Kitchen helpers	75,535	1.3%
13	3343	Administrative and executive secretaries	73,364	1.3%
14	4110	General office clerks	69,875	1.2%
15	4322	Production clerks	67,997	1.2%
16	4311	Accounting and bookkeeping clerks	58,822	1.0%
17	8153	Sewing machine operators	54,628	1.0%
18	4222	Contact centre information clerks	50,337	0.9%
19	3312	Credit and loans officers	48,063	0.8%
20	5321	Health care assistants	45,279	0.8%
21	3115	Mechanical engineering technicians	40,808	0.7%
22	9333	Freight handlers	38,009	0.7%
23	4323	Transport clerks	37,532	0.7%
24	2635	Social work and counselling professionals	35,148	0.6%
25	3341	Office supervisors	34,921	0.6%
26	5131	Waiters	34,873	0.6%
27	3512	Information and communications technology user support technicians	33,977	0.6%
28	2221	Nursing professionals	33,337	0.6%
29	8322	Car, taxi and van drivers	31,782	0.6%
30	2412	Financial and investment advisers	31,424	0.5%
31	3511	Information and communications technology operations technicians	31,320	0.5%
32	1120	Managing directors and chief executives	31,212	0.5%
33	5221	Shopkeepers	30,940	0.5%
34	2634	Psychologists	30,926	0.5%
35	8343	Crane, hoist and related plant operators	27,656	0.5%

36	2141	Industrial and production engineers	27,393	0.5%
37	7412	Electrical mechanics and fitters	26,181	0.5%
38	2211	Generalist medical practitioners	25,237	0.4%
39	7512	Bakers, pastry-cooks and confectionery makers	24,275	0.4%
40	7119	Building frame and related trades workers not elsewhere classified	24,019	0.4%
41	4226	Receptionists (general)	22,728	0.4%
42	4415	Filing and copying clerks	22,465	0.4%
43	2166	Graphic and multimedia designers	21,913	0.4%
44	1324	Supply, distribution and related managers	21,376	0.4%
45	3114	Electronics engineering technicians	21,346	0.4%
46	2619	Legal professionals not elsewhere classified	21,295	0.4%
47	3213	Pharmaceutical technicians and assistants	20,709	0.4%
48	9621	Messengers, package deliverers and luggage porters	20,600	0.4%
49	2513	Web and multimedia developers	20,333	0.4%
50	3257	Environmental and occupational health inspectors and associates	19,814	0.3%
51	2151	Electrical engineers	21,238	0.4%
52	3435	Other artistic and cultural associate professionals	20,676	0.4%
53	2149	Engineering professionals not elsewhere classified	19,993	0.4%
54	9112	Cleaners and helpers in offices, hotels and other establishments	18,385	0.3%
55	3323	Buyers	18,069	0.3%
56	5120	Cooks	17,489	0.3%
57	2511	Systems analysts	17,485	0.3%
58	8332	Heavy truck and lorry drivers	17,401	0.3%
59	2431	Advertising and marketing professionals	17,179	0.3%
60	9622	Odd job persons	16,951	0.3%
61	7411	Building and related electricians	16,881	0.3%
62	9111	Domestic cleaners and helpers	16,751	0.3%
63	2523	Computer network professionals	16,659	0.3%
64	9329	Manufacturing labourers not elsewhere classified	15,209	0.3%
65	4227	Survey and market research interviewers	14,790	0.3%
66	7422	Information and communications technology installers and servicers	13,754	0.2%
67	2131	Biologists, botanists, zoologists and related professionals	13,728	0.2%

68	7233	Agricultural and industrial machinery mechanics and repairers	13,724	0.2%
69	9313	Building construction labourers	13,449	0.2%
70	7322	Printers	13,409	0.2%
71	2161	Building architects	13,303	0.2%
72	3113	Electrical engineering technicians	13,285	0.2%
73	2142	Civil engineers	13,256	0.2%
74	2163	Product and garment designers	12,774	0.2%
75	2113	Chemists	12,428	0.2%
76	3112	Civil engineering technicians	12,180	0.2%
77	2212	Specialist medical practitioners	11,948	0.2%
78	8321	Motorcycle drivers	11,695	0.2%
79	2144	Mechanical engineers	11,678	0.2%
80	3122	Manufacturing supervisors	11,336	0.2%
81	2310	University and higher education teachers	11,291	0.2%
82	8142	Plastic products machine operators	10,969	0.2%
83	4313	Payroll clerks	10,835	0.2%
84	3334	Real estate agents and property managers	10,618	0.2%
85	7212	Welders and flamecutters	10,592	0.2%
86	2611	Lawyers	10,564	0.2%
87	2351	Education methods specialists	10,378	0.2%
88	7115	Carpenters and joiners	10,311	0.2%
89	8160	Food and related products machine operators	10,204	0.2%
90	2152	Electronics engineers	10,175	0.2%
91	2261	Dentists	10,005	0.2%
92	3123	Construction supervisors	9,896	0.2%
93	3434	Chefs	9,781	0.2%
94	7231	Motor vehicle mechanics and repairers	9,514	0.2%
95	5142	Beauticians and related workers	8,901	0.2%
96	7221	Blacksmiths, hammersmiths and forging press workers	8,861	0.2%
97	2262	Pharmacists	8,778	0.2%
98	2146	Mining engineers, metallurgists and related professionals	8,507	0.2%
99	3339	Business services agents not elsewhere classified	8,491	0.2%
100	5222	Shop supervisors	8,277	0.1%
101	2622	Librarians and related information professionals	8,213	0.1%
102	2269	Health professionals not elsewhere classified	8,188	0.1%
103	2434	Information and communications technology sales professionals	8,116	0.1%

104	3522	Telecommunications engineering technicians	8,026	0.1%
105	9613	Sweepers and related labourers	7,888	0.1%
106	4416	Personnel clerks	7,795	0.1%
107	1439	Services managers not elsewhere classified	7,630	0.1%
108	6113	Gardeners, horticultural and nursery growers	7,524	0.1%
109	8143	Paper products machine operators	7,420	0.1%
110	2433	Technical and medical sales professionals (excluding ICT)	7,394	0.1%
111	3111	Chemical and physical science technicians	7,223	0.1%
112	2330	Secondary education teachers	7,144	0.1%
113	7131	Painters and related workers	7,048	0.1%
114	2264	Physiotherapists	7,036	0.1%
115	9629	Elementary workers not elsewhere classified	6,792	0.1%
116	3321	Insurance representatives	6,724	0.1%
117	7421	Electronics mechanics and servicers	6,627	0.1%
118	1212	Human resource managers	6,586	0.1%
119	3118	Draughtspersons	6,519	0.1%
120	7223	Metal working machine tool setters and operators	6,404	0.1%
121	1346	Financial and insurance services branch managers	6,269	0.1%
122	8183	Packing, bottling and labelling machine operators	6,232	0.1%
123	5243	Door to door salespersons	6,228	0.1%
124	3411	Police inspectors and detectives	6,214	0.1%
125	7533	Sewing, embroidery and related workers	6,174	0.1%
126	3521	Broadcasting and audio-visual technicians	5,974	0.1%
127	2265	Dieticians and nutritionists	5,734	0.1%
128	3117	Mining and metallurgical technicians	5,567	0.1%
129	7511	Butchers, fishmongers and related food preparers	5,563	0.1%
130	5132	Bartenders	5,363	0.1%
131	5249	Sales workers not elsewhere classified	5,229	0.1%
132	4419	Clerical support workers not elsewhere classified	5,196	0.1%
133	2659	Creative and performing artists not elsewhere classified	5,068	0.1%
134	7323	Print finishing and binding workers	5,041	0.1%
135	2120	Mathematicians, actuaries and statisticians	4,981	0.1%

136	8131	Chemical products plant and machine operators	4,960	0.1%
137	1323	Construction managers	4,918	0.1%
138	1321	Manufacturing managers	4,895	0.1%
139	5322	Home-based personal care workers	4,837	0.1%
140	2512	Software developers	4,740	0.1%
141	2353	Other language teachers	4,710	0.1%
142	4212	Bookmakers, croupiers and related gaming workers	4,577	0.1%
143	7127	Air conditioning and refrigeration mechanics	4,537	0.1%
144	5329	Personal care workers in health services not elsewhere classified	4,451	0.1%
145	2153	Telecommunications engineers	4,385	0.1%
146	8342	Earthmoving and related plant operators	4,365	0.1%
147	5211	Stall and market salespersons	4,320	0.1%
148	1420	Retail and wholesale trade managers	4,300	0.1%
149	5244	Contact centre salespersons	4,041	0.1%
150	3422	Sports coaches, instructors and officials	3,806	0.1%
151	2250	Veterinarians	3,647	0.1%
152	7318	Handicraft workers in textile, leather and related materials	3,627	0.1%
153	2521	Database designers and administrators	3,588	0.1%
154	7321	Pre-press technicians	3,558	0.1%
155	3412	Social work associate professionals	3,488	0.1%
156	8156	Shoemaking and related machine operators	3,463	0.1%
157	7311	Precision-instrument makers and repairers	3,371	0.1%
158	5311	Child care workers	3,364	0.1%
159	5153	Building caretakers	3,360	0.1%
160	3211	Medical imaging and therapeutic equipment technicians	3,337	0.1%
161	5164	Pet groomers and animal care workers	3,174	0.1%
162	3333	Employment agents and contractors	3,169	0.1%
163	2145	Chemical engineers	3,144	0.1%
164	2143	Environmental engineers	3,104	0.1%
165	3514	Web technicians	3,104	0.1%
166	2132	Farming, forestry and fisheries advisers	2,957	0.1%
167	7111	House builders	2,937	0.1%
168	2164	Town and traffic planners	2,755	0.05%
169	5112	Transport conductors	2,728	0.05%
170	1330	Information and communications technology service managers	2,652	0.05%

171	2359	Teaching professionals not elsewhere classified	2,611	0.05%
172	8341	Mobile farm and forestry plant operators	2,587	0.05%
173	7536	Shoemakers and related workers	2,587	0.05%
174	7132	Spray painters and varnishers	2,571	0.05%
175	4221	Travel consultants and clerks	2,563	0.05%
176	3313	Accounting associate professionals	2,488	0.04%
177	9520	Street vendors (excluding food)	2,471	0.04%
178	1411	Hotel managers	2,414	0.04%
179	3212	Medical and pathology laboratory technicians	2,225	0.04%
180	3142	Agricultural technicians	2,186	0.04%
181	2352	Special needs teachers	2,127	0.04%
182	7222	Toolmakers and related workers	2,042	0.04%
183	8181	Glass and ceramics plant operators	1,995	0.04%
184	5163	Undertakers and embalmers	1,973	0.04%
185	2643	Translators, interpreters and other linguists	1,946	0.03%
186	7514	Fruit, vegetable and related preservers	1,913	0.03%
187	7531	Tailors, dressmakers, furriers and hatters	1,875	0.03%
188	5152	Domestic housekeepers	1,855	0.03%
189	7534	Upholsterers and related workers	1,836	0.03%
190	5111	Travel attendants and travel stewards	1,832	0.03%
191	4229	Client information workers not elsewhere classified	1,812	0.03%
192	7211	Metal moulders and coremakers	1,799	0.03%
193	3513	Computer network and systems technicians	1,696	0.03%
194	7114	Concrete placers, concrete finishers and related workers	1,667	0.03%
195	3153	Aircraft pilots and related associate professionals	1,657	0.03%
196	2413	Financial analysts	1,647	0.03%
197	2633	Philosophers, historians and political scientists	1,623	0.03%
198	2133	Environmental protection professionals	1,615	0.03%
199	7112	Bricklayers and related workers	1,597	0.03%
200	5419	Protective services workers not elsewhere classified	1,574	0.03%
201	3352	Government tax and excise officials	1,573	0.03%
202	8159	Textile, fur and leather products machine operators not elsewhere classified	1,567	0.03%
203	2267	Optometrists and ophthalmic opticians	1,557	0.03%
204	7213	Sheet-metal workers	1,551	0.03%
205	9211	Crop farm labourers	1,514	0.03%

206	5141	Hairdressers	1,507	0.03%
207	9411	Fast food preparers	1,485	0.03%
208	8219	Assemblers not elsewhere classified	1,484	0.03%
209	8121	Metal processing plant operators	1,455	0.03%
210	2165	Cartographers and surveyors	1,434	0.03%
211	1211	Finance managers	1,421	0.03%
212	3354	Government licensing officials	1,408	0.03%
213	3133	Chemical processing plant controllers	1,405	0.02%
214	7315	Glass makers, cutters, grinders and finishers	1,399	0.02%
215	2641	Authors and related writers	1,398	0.02%
216	2421	Management and organization analysts	1,395	0.02%
217	7532	Garment and related pattern-makers and cutters	1,380	0.02%
218	2631	Economists	1,317	0.02%
219	3432	Interior designers and decorators	1,312	0.02%
220	7544	Fumigators and other pest and weed controllers	1,245	0.02%
221	3154	Air traffic controllers	1,234	0.02%
222	2642	Journalists	1,234	0.02%
223	3152	Ships' deck officers and pilots	1,223	0.02%
224	3431	Photographers	1,200	0.02%
225	2423	Personnel and careers professionals	1,198	0.02%
226	8157	Laundry machine operators	1,157	0.02%
227	7314	Potters and related workers	1,112	0.02%
228	1349	Professional services managers not elsewhere classified	1,103	0.02%
229	7522	Cabinet-makers and related workers	1,099	0.02%
230	9612	Refuse sorters	1,065	0.02%
231	2632	Sociologists, anthropologists and related professionals	1,062	0.02%
232	4132	Data entry clerks	1,036	0.02%
233	2656	Announcers on radio, television and other media	1,016	0.02%
234	3314	Statistical, mathematical and related associate professionals	1,011	0.02%
235	5245	Service station attendants	1,002	0.02%
236	3331	Clearing and forwarding agents	998	0.02%
237	7122	Floor layers and tile setters	969	0.02%
238	2432	Public relations professionals	966	0.02%
239	3132	Incinerator and water treatment plant operators	948	0.02%
240	1114	Senior officials of special-interest organizations	945	0.02%
241	2114	Geologists and geophysicists	944	0.02%

242	2266	Audiologists and speech therapists	939	0.02%
243	3423	Fitness and recreation instructors and program leaders	939	0.02%
244	8152	Weaving and knitting machine operators	936	0.02%
245	8114	Cement, stone and other mineral products machine operators	925	0.02%
246	8331	Bus and tram drivers	901	0.02%
247	7125	Glaziers	895	0.02%
248	1345	Education managers	876	0.02%
249	3332	Conference and event planners	864	0.02%
250	7224	Metal polishers, wheel grinders and tool sharpeners	829	0.01%
251	2342	Early childhood educators	820	0.01%
252	9623	Meter readers and vending-machine collectors	811	0.01%
253	2529	Database and network professionals not elsewhere classified	803	0.01%
254	7513	Dairy-products makers	797	0.01%
255	3255	Physiotherapy technicians and assistants	791	0.01%
256	1412	Restaurant managers	787	0.01%
257	9510	Street and related service workers	783	0.01%
258	2654	Film, stage and related directors and producers	760	0.01%
259	3155	Air traffic safety electronics technicians	741	0.01%
260	3214	Medical and dental prosthetic technicians	736	0.01%
261	5411	Fire-fighters	733	0.01%
262	7214	Structural-metal preparers and erectors	731	0.01%
263	7313	Jewellery and precious-metal workers	727	0.01%
264	8312	Railway brake, signal and switch operators	720	0.01%
265	1341	Child care services managers	700	0.01%
266	7413	Electrical line installers and repairers	684	0.01%
267	2519	Software and applications developers and analysts not elsewhere classified	683	0.01%
268	8182	Steam engine and boiler operators	675	0.01%
269	8344	Lifting truck operators	671	0.01%
270	3251	Dental assistants and therapists	660	0.01%
271	4213	Pawnbrokers and money-lenders	652	0.01%
272	3254	Dispensing opticians	640	0.01%
273	4225	Enquiry clerks	630	0.01%
274	8151	Fibre preparing, spinning and winding machine operators	583	0.01%
275	5169	Personal services workers not elsewhere classified	558	0.01%

276	2424	Training and staff development professionals	547	0.01%
277	2653	Dancers and choreographers	546	0.01%
278	3342	Legal secretaries	523	0.01%
279	9121	Hand launderers and pressers	504	0.01%
280	7126	Plumbers and pipe fitters	500	0.01%
281	7124	Insulation workers	495	0.01%
282	1222	Advertising and public relations managers	493	0.01%
283	5165	Driving instructors	474	0.01%
284	3139	Process control technicians not elsewhere classified	468	0.01%
285	2621	Archivists and curators	465	0.01%
286	1431	Sports, recreation and cultural centre managers	463	0.01%
287	8350	Ships' deck crews and related workers	461	0.01%
288	9214	Garden and horticultural labourers	442	0.01%
289	3259	Health associate professionals not elsewhere classified	441	0.01%
290	8211	Mechanical machinery assemblers	434	0.01%
291	8154	Bleaching, dyeing and fabric cleaning machine operators	426	0.01%
292	4411	Library clerks	393	0.01%
293	7215	Riggers and cable splicers	381	0.01%
294	5246	Food service counter attendants	364	0.01%
295	3143	Forestry technicians	362	0.01%
296	8155	Fur and leather preparing machine operators	361	0.01%
297	2522	Systems administrators	360	0.01%
298	7317	Handicraft workers in wood, basketry and related materials	352	0.01%
299	1219	Business services and administration managers not elsewhere classified	352	0.01%
300	3151	Ships' engineers	344	0.01%
301	7316	Sign writers, decorative painters, engravers and etchers	337	0.01%
302	7516	Tobacco preparers and tobacco products makers	330	0.01%
303	6121	Livestock and dairy producers	325	0.01%
304	5312	Teachers' aides	308	0.01%
305	2612	Judges	297	0.01%
306	1322	Mining managers	285	0.01%
307	2320	Vocational education teachers	282	0.01%
308	2636	Religious professionals	258	0.005%
309	3240	Veterinary technicians and assistants	257	0.005%
310	4312	Statistical, finance and insurance clerks	256	0.005%

311	7121	Roofers	251	0.004%
312	4211	Bank tellers and related clerks	249	0.004%
313	7541	Underwater divers	242	0.004%
314	3121	Mining supervisors	241	0.004%
315	8111	Miners and quarriers	237	0.004%
316	3141	Life science technicians (excluding medical)	236	0.004%
317	4131	Typists and word processing operators	218	0.004%
318	7535	Pelt dressers, tanners and fellmongers	216	0.004%
319	7232	Aircraft engine mechanics and repairers	208	0.004%
320	3353	Government social benefits officials	206	0.004%
321	6112	Tree and shrub crop growers	196	0.004%
322	8171	Pulp and papermaking plant operators	196	0.004%
323	3134	Petroleum and natural gas refining plant operators	189	0.003%
324	2356	Information technology trainers	188	0.003%
325	3355	Police inspectors and detectives	180	0.003%
326	2222	Midwifery professionals	180	0.003%
327	8141	Rubber products machine operators	179	0.003%
328	2514	Applications programmers	171	0.003%
329	7113	Stonemasons, stone cutters, splitters and carvers	168	0.003%
330	8122	Metal finishing, plating and coating machine operators	167	0.003%
331	8132	Photographic products machine operators	159	0.003%
332	9611	Garbage and recycling collectors	125	0.002%
333	6210	Forestry and related workers	124	0.002%
334	5151	Cleaning and housekeeping supervisors in offices, hotels and other establishments	122	0.002%
335	8112	Mineral and stone processing plant operators	109	0.002%
336	4413	Coding, proof-reading and related clerks	109	0.002%
337	8172	Wood processing plant operators	98	0.002%
338	2655	Actors	94	0.002%
339	7234	Bicycle and related repairers	93	0.002%
340	3258	Ambulance workers	88	0.002%
341	9321	Hand packers	86	0.002%
342	4224	Hotel receptionists	81	0.001%
343	3135	Metal production process controllers	75	0.001%
344	8311	Locomotive engine drivers	75	0.001%
345	3324	Trade brokers	71	0.001%
346	9334	Shelf fillers	67	0.001%
347	9212	Livestock farm labourers	65	0.001%

348	4120	Secretaries (general)	63	0.001%
349	1343	Aged care services managers	60	0.001%
350	3359	Regulatory government associate professionals not elsewhere classified	60	0.001%
351	2341	Primary school teachers	59	0.001%
352	1311	Agricultural and forestry production managers	58	0.001%
353	2355	Other arts teachers	56	0.001%
354	9312	Civil engineering labourers	56	0.001%
355	5113	Travel guides	48	0.001%
356	1342	Health services managers	47	0.001%
357	1223	Research and development managers	43	0.001%
358	6221	Aquaculture workers	41	0.001%
359	6111	Field crop and vegetable growers	34	0.001%
360	3344	Medical secretaries	34	0.001%
361	2354	Other music teachers	31	0.001%
362	9216	Fishery and aquaculture labourers	26	0.0005%
363	7521	Wood treaters	21	0.0004%
364	2651	Visual artists	17	0.0003%
365	6123	Apiarists and sericulturists	14	0.0003%
366	2112	Meteorologists	13	0.0002%
367	8113	Well drillers and borers and related workers	11	0.0002%
368	8189	Stationary plant and machine operators not elsewhere classified	10	0.0002%
369	9129	Other cleaning workers	10	0.0002%
370	7319	Handicraft workers not elsewhere classified	10	0.0002%
371	5412	Police officers	8	0.0001%
372	2162	Landscape architects	8	0.0001%
373	1213	Policy and planning managers	7	0.0001%
374	9311	Mining and quarrying labourers	6	0.0001%
375	5212	Street food salespersons	5	0.0001%

Source: Vacancy information. Own calculations.

Appendix B: Text mining

Column 1 and 2 of Table B.1 shows part of the information provided by employers for a pair of job vacancies. In the job description (Column 1) for the first vacancy, the employers mention that a person is required with an undergraduate certificate; however in the website column, where the employer was supposed to provide specific information regarding educational requirements (Column 2), there is a missing value. In contrast, for the second vacancy (see the second row of Table B.1), in the job description (Column 1) the employer does not mention any educational requirements. Indeed, for this vacancy the information regarding education is available in the "Educational requirements" column (Column 2).

Thus, the algorithm needs to "read" the different columns of the scraped data (not only the educational requirements column) to identify qualification requirements. Some of the relevant information might be only mentioned in the job description or in other specific columns; additionally, information might be repeated in different columns. In this example, the algorithm creates "Dummy_Undergraduate_certificate" and "Dummy_PhD_certificate" columns (the third and fourth columns of Table B.1). Based on the information provided in the job description, the "Dummy_Undergraduate_certificate" column identifies (takes a value of 1) that the first vacancy required a person with an undergraduate certificate while the second vacancy does not require a person with an undergraduate certificate. On the other hand, based on the "Educational requirements" column, the "Dummy_PhD_certificate" column identifies that the second vacancy (second row of Table B.1) requests a person with a PhD. It is important to note that employers might be indifferent about educational levels or another job characteristic. For instance, a vacancy might require a person with a high school or undergraduate level. In these cases, the dummy variables ("high school" and "undergraduate_certificate") take the value of 1 at the same time.

Table B.1: Example of the content of a scraped database

Job description	Educational requirements	Dummy_ Undergraduate certificate	Dummy_ PhD certificate
• Must have some previous production experience in a book publishing or printing environment • Excellent communication and interpersonal skills and consultative customer care approach • Undergraduate certificate • Strong IT skills with experience of using a CRM system advantageous	No information provided by the employer (missing value)	1	0
• Candidates with expertise in cancer genomics and related disciplines, and candidates with experience of working with clinical data, are particularly encouraged to apply. • Editorial experience is not required, although applicants with	A PhD (or equivalent) in a field related to cancer and/or genomics and significant research experience.	0	1

significant editorial experience are encouraged to apply and will potentially be considered for a Senior Editor position.			

Additionally, there is an issue when looking for patterns in the Spanish language. In this language, nouns have a gender; for instance, an undergraduate might be called "*universitario*" (for men) or "*universitaria*" (for women). Given this fact, and the usage of synonyms to express a job requirement, the algorithm looks for patterns in the root of the words[26]. The selection of the root of the words is a critical process where it is necessary to carefully select the proper roots to correctly classify job requirements. In this way, the dummy variables created are guaranteed to correctly identify employers' requirements, even with the presence of synonyms, nouns with genders, etc.

[26] For instance, in the undergraduate case, the algorithm looks for "universi" among other word roots.

Agradecimientos

Esta serie de documentos de trabajo es financiada por el programa "Inclusión productiva y social: programas y políticas para la promoción de una economía formal", código 60185, que conforma Colombia Científica-Alianza EFI, bajo el Contrato de Recuperación Contingente No.FP44842-220-2018.

Acknowledgments

This working paper series is funded by the Colombia Cientfíca-Alianza EFI Research Program, with code 60185 and contract number FP44842-220-2018, funded by The World Bank through the call Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology and Innovation.