THE LINK BETWEEN EMPLOYMENT DEPRIVATION AND PEOPLE WHO HAVE NEVER WORKED

MARTINUS PETRUS KLEIWEG

Contents

The link between employment deprivation and people who have never worked

Summary

The number of people in the United Kingdom who have never had a paid job has increased by almost 15 percent in the last decade (ONS, 2019), even though UK employment was at record levels before the Covid-19 pandemic. People that are 16-24 years old represent most of the population who have never had a paid job.

This report examines what areas of England have the highest percentage of people under 25 who have never worked. I analysed the labour market data from the UK census in 2011 and found the five Middle Layer Super Output Areas (MSOAs) with the highest percentage: County Durham 066, Medway 024, Bradford 045, Weymouth and Portland 008, and Kingston upon Hull 003.

My analysis shows that areas classified as industrial and multi-ethnic (Bradford and Kingston upon Hull), mining/manufacturing legacy (County Durham), or ethnically diverse metropolitan living have on average a relatively high percentage of young people who have never worked, while areas classified as rural areas, farming areas and college towns (full-time students are not included in the 'never worked' category) have on average a low percentage.

By combining the census dataset with the deprivation indices dataset, I examined if there is a correlation between the 'never worked' percentage in a MSOA and the employment deprivation ranking of that area. The correlation strength turned out to be moderate, so the lack of job opportunities in a MSOA is likely a reason why people there have never worked.

Further data analysis is needed to confirm this and to find other reasons before policies can be established to address this problem. For example, the Lower Layer Super Output Areas (LSOAs) that are part of the identified MSOAs with high percentages can be analysed to get more information (e.g., education, sex, ethnicity) about the people who miss out on the job market.

Aims and objectives

This study investigates the link between the level of employment deprivation of an area and the percentage of residents of that area who have never worked.

It combines two datasets: one that contains the labour market statistics of all residents 16 and over for each MSOA and one that ranks the 32.844 LSOAs in England by deprivation level. The first dataset covers England and Wales, the second dataset England.

This study considers two questions: what are the five MSOAs in England with the highest percentage of people between 16 and 24 years old who have never worked, and is there a correlation between the percentage of people who have never worked and the employment deprivation ranking of the area where they live?

The report does not require any specialised knowledge or understanding.

Background

According to the Office for National Statistics (ONS) (2019), around 3.6 million adults in the UK under 65 years old have never been paid for work. This is almost ten percent of the total workforce under 65. From that group, 52 percent are younger than 30 years old (without including full-time students).

The main reasons for never having had a paid job before, according to interviews conducted by the ONS, are studying full-time (55%), actively seeking work but still unemployed (7%), looking after the family, or home (14%) or being short- or long-term sick (12%).

This investigation uses the labour market data of the UK census of 2011 and the English indices of deprivation to analyse people who have never worked. To avoid ambiguity, people who have never worked in this investigation are 16 years and older and have never been in paid employment but would wish to be (ONS, n.d.).

One limitation of both datasets is that they are both (in part) based on surveys filled in by householders. The answers they gave to the questions can be not or only partly true. In addition, people that do not participate are not considered in the same way as participants, even though ONS uses complex statistical techniques to adjust the census for those people missed (ONS, n.d.). A further limitation of both datasets is that the choice of components and the weighting of those components that lead to the rankings (deprivation indices) and the classifications (census) is unavoidably subjective.

Data sources

Full details of data sources and licences are given in Appendix 2.

English indices of deprivation 2015

The deprivation data used in this report was originally obtained from the Gov.uk website (Ministry of Housing, Communities & Local Government, 2015). The original Excel-file was converted into a CSV file and underwent some minor cleaning after it was imported as a data frame into Jupyter Notebook using Python.

The data contains the deprivation rankings (overall and per domain) of the 32844 LSOAs in England, where the LSOA (between 1000 and 3000 residents) with a rank of 1 is the most deprived and a rank of 32844 the least deprived.

The Employment deprivation domain measures the proportion of involuntary unemployment in an area (UK Government, 2015).

UK census report DC6206EW

This dataset that includes labour market statistics of different age groups was originally obtained from Nomis (2013). Nomis is a service provided by ONS that gives free access to UK labour market statistics from official sources.

The dataset consists of four CSV-files, one for each age group: 16-24, 25-49, 50-64, and 65 and over. The files underwent some minor cleaning after they were imported as data frames into Jupyter Notebook using Python. There is also an extra CSV-file, the Output Area Classification Table (ONS, n.d.), which contains classifications for all the LSOAs and Local Authority Districts (LADs) as determined by ONS.

The data contains labour market statistics for all the 7201 MSOAs in England and Wales. For every MSOA (average population is 8288) it has the number of people in 12 labour market categories, including 'Never worked'.

Higher professional occupations	Lower managerial occupations	Intermediate occupations	Never worked	Long-term unemployed	Full-time students
482	1296	824	227	98	516

Table 1: The average of six categories for all age groups and all MSOAs combined.

For the first question, I used the dataset for the age group 16-24. To plot the data on a map, I used a GeoJSON file of all the English LADs (Chlorley, 2017). For the second question, I used all four datasets and the deprivation dataset. I did not encounter any missing values in the datasets.

Analysis pipeline

Finding the five MSOAs in England with the highest percentage of people from 16 to 24 years old who have never worked.

After importing the census dataset for the age group 16-24, I first wanted to know the total population of every MSOA for that age group. I got that number by calculating the sum of each row using Python. To get the percentage of people who have never worked, I divided the number of people who never worked by the total population and multiplied by 100. I added the columns 'Total population' and 'Percentage never worked' to the imported dataset.

Because I only wanted to focus on England here, I then filtered out all the MSOA-codes that start with a 'W', which stands for Wales. That left me with a table of 6791 rows. By sorting that table on 'Percentage never worked', I found the five MSOAs with the highest percentages and the five MSOAs with the lowest percentages.

To give more context to these data, I linked the LAD Classification column from the Output Area Classification Table to the table. I then created a bar chart to see what LAD classifications have the highest percentage of people who have never worked in the age category 16-24. I did this by grouping the data frame by LAD classification using Python and then again calculating the 'never worked' percentage.

To finalise this question, I also grouped the data frame by LAD. I did this by removing the last four characters of each MSOA name and combining the rows with the same name. I created a bar chart using the resulting table. I chose for this bar chart to only consider LADs with more than 50.000 residents, to give an overview of the more populated areas. I also used this data frame to plot the data on a map of England, using a GeoJSON file of all the LADs and the Python library Folium.

Is there a correlation between the percentage of people that have never worked and the employment deprivation ranking of the area where they live?

To find an answer to the second question, I started by importing the four CSV-files of the census dataset. After that, I combined the resulting data frames into one data frame that contained the labour market data of everyone 16 years and older. I did this by adding up all the rows with the same MSOA code using Python. Then, like before, I calculated the total population and the 'Percentage never worked' for each row and added both as additional columns to the data frame.

I imported the deprivation dataset and did some minor data cleaning. Because this dataset contains data for the LSOAs, I needed to find a way to convert it to a dataset for MSOAs so that I would be able to combine it with the census dataset. I did this by removing the last character of each LSOA name (Adur 001A), which gives the MSOA name (Adur 001).

I could not combine the rows by taking the sum of the rows, because a sum of the rankings would not make sense. I, therefore, decided to calculate the mean of the rankings instead.

After that, I was able to combine the census dataset and the deprivation dataset using the MSOA name. By combining, I lost the MSOAs that are part of Wales, because the deprivation dataset has no data about Wales.

	Percentage Never Worked	Average Multiple Deprivation Rank	Average Employment Deprivation Rank
Mean	4.28	16450.32	16469.64
Minimum	0.64	22.75	61.40
Maximum	29.34	32625.60	32822.67

Table 2: The mean, minimum and maximum of the 16 and over table.

To find if there is a correlation between the percentage who have never worked and the employment ranking, I used Python to plot a scatter graph with the two columns. I also used Python and its scipy.stats module to perform Pearson's correlation test.

To finalise this question, I used the k-means clustering algorithm as a data mining tool to divide the data points of the Employment ranking scatter graph into two clusters, to see if there was a difference in correlation between the lower and the higher rankings.

Findings

16-24 years old people who have never worked

I used the census dataset of the age group 16-24 years to find the five MSOAs that have the highest percentage of people in that age group who have never worked and the five MSOAs with the lowest percentage. To give context, I added the classification of the LAD that each MSOA is part of.

MSOA name	Percentage never worked	Total population	LAD Classification
County Durham 066	25.18	1116	Mining Legacy
Medway 024	24.91	1377	City Periphery
Bradford 045	24.16	890	Industrial and Multi-ethnic
Weymouth and Portland 008	22.81	982	Seaside Living
Kingston upon Hull 003	22.63	1109	Industrial and Multi-ethnic

Table 3: The five MSOAs in age group 16-24 years old that have the highest percentage of people who have never worked.

MSOA name	Percentage never worked	Total population	LAD Classification
Bristol 030	0.43	1388	Larger Towns and Cities
Cambridge 007	0.37	7066	University Towns and Cities
Oxford 008	0.36	9876	University Towns and Cities
Newcastle upon Tyne 013	0.36	5853	Larger Towns and Cities
Cambridge 005	0.34	5017	University Towns and Cities

Table 4: The five MSOAs in age group 16-24 years that have the lowest percentage of people who have never worked.

The reasons why the percentages in Cambridge and Oxford are so low is probably because these are university towns and people in the age group 16-24 that live in these towns will most likely be students and are therefore not part of the 'never worked' category. The same applies to Bristol 030 (home of the University of Bristol) and Newcastle upon Tyne 013 (Newcastle University and Northumbria University).

The reasons why the MSOAs in table 5 have such high percentages are less clear. The LADs of Bradford and Kingston upon Hull are classified by ONS as 'Industrial and Multi-ethnic', which is nationally the category with the highest percentage of people who have never worked in this age category (see figure 6). The classifications of County Durham 066 (Mining Legacy) and Weymouth and Portland 008 (Seaside Living) also score high.
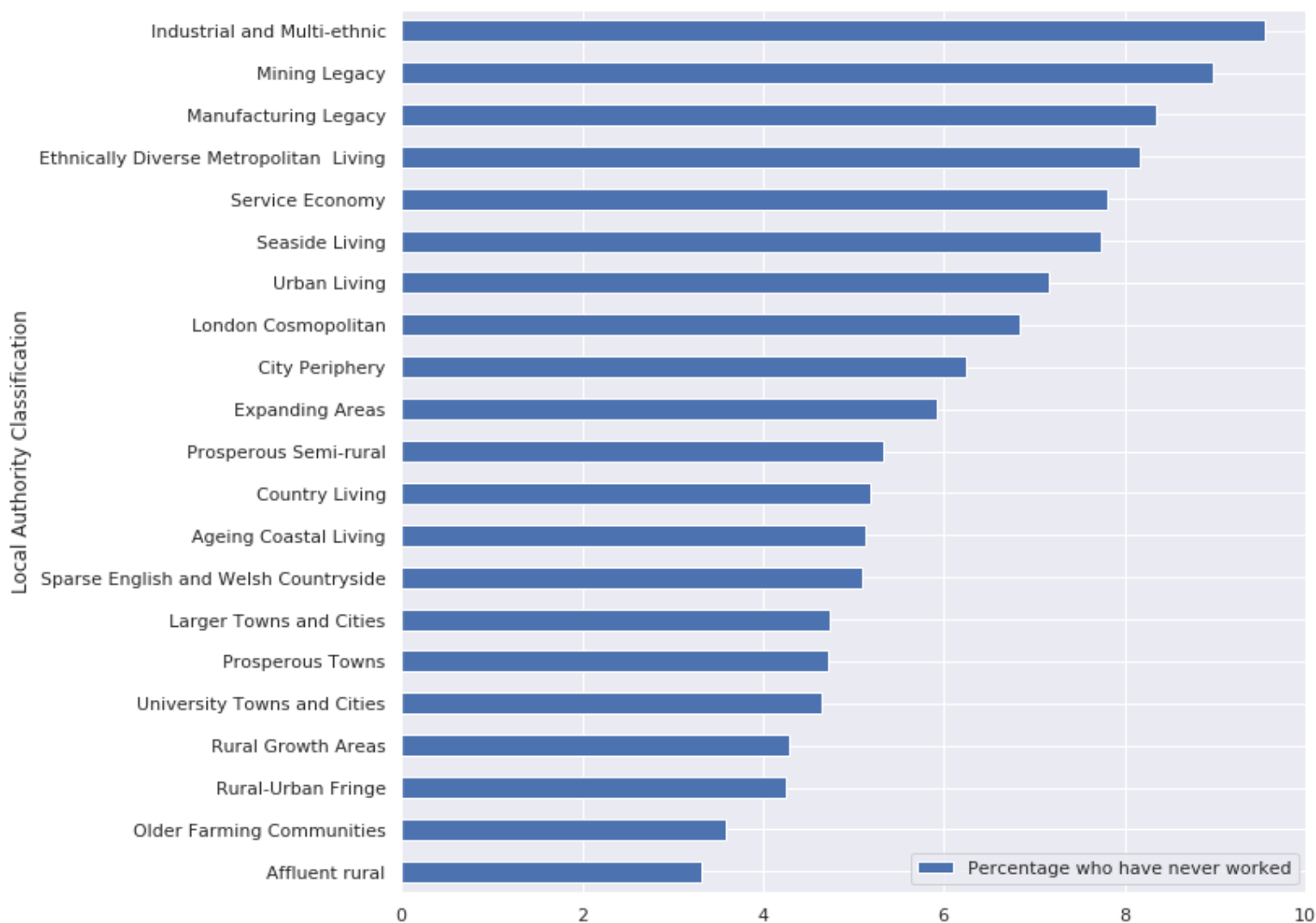
Figure 1: Percentage of people of 16-24 years old who have never worked by LAD Classification.

Figure 1 confirms that young people living in university towns and rural areas are less likely to be part of the 'never worked' category. In contrast, young people living in industrial and ethnically diverse towns are more likely to be part of the 'never worked' category.
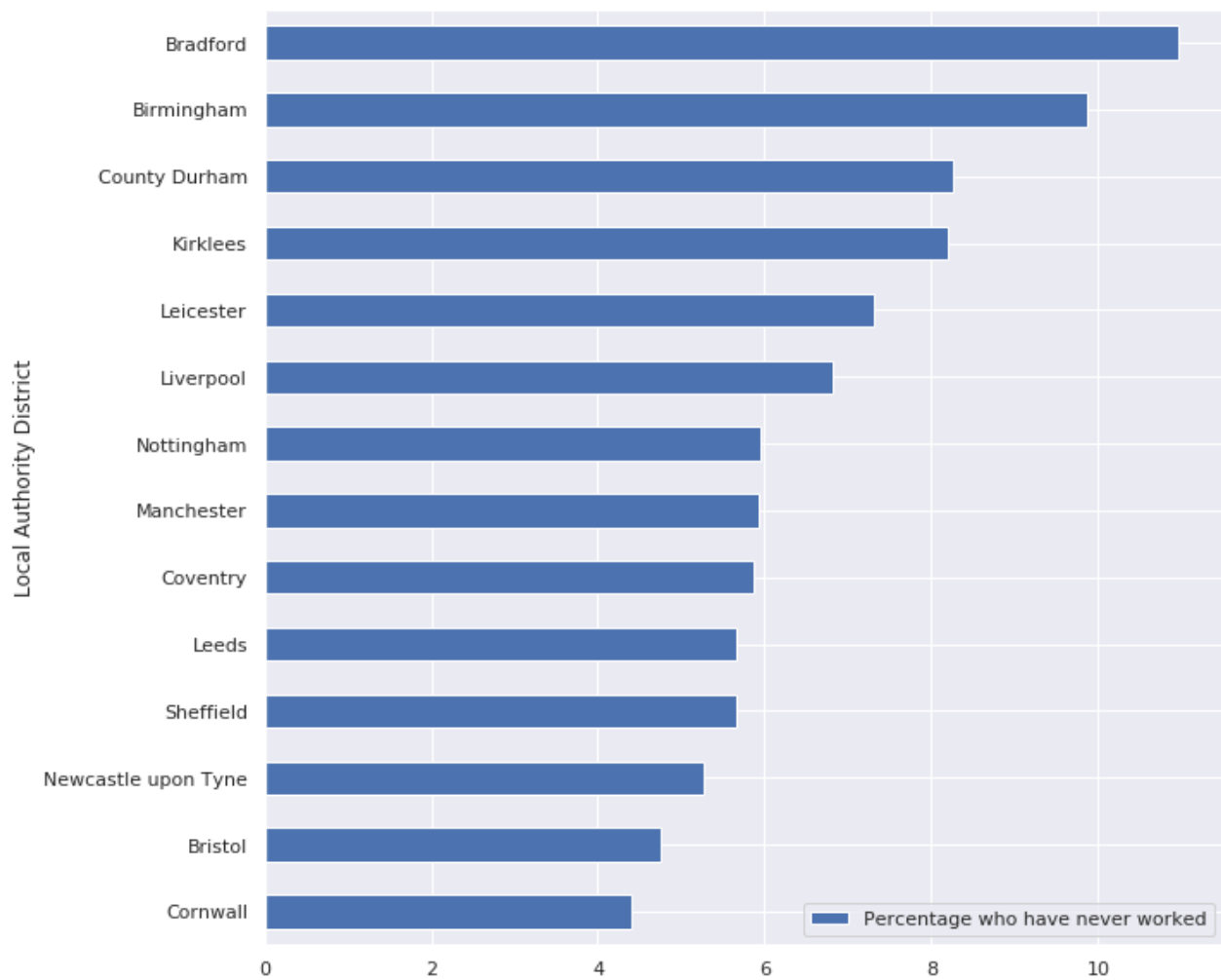
Figure 2: Percentage of people of 16-24 years old who have never worked by LAD of more than 50.000 residents.
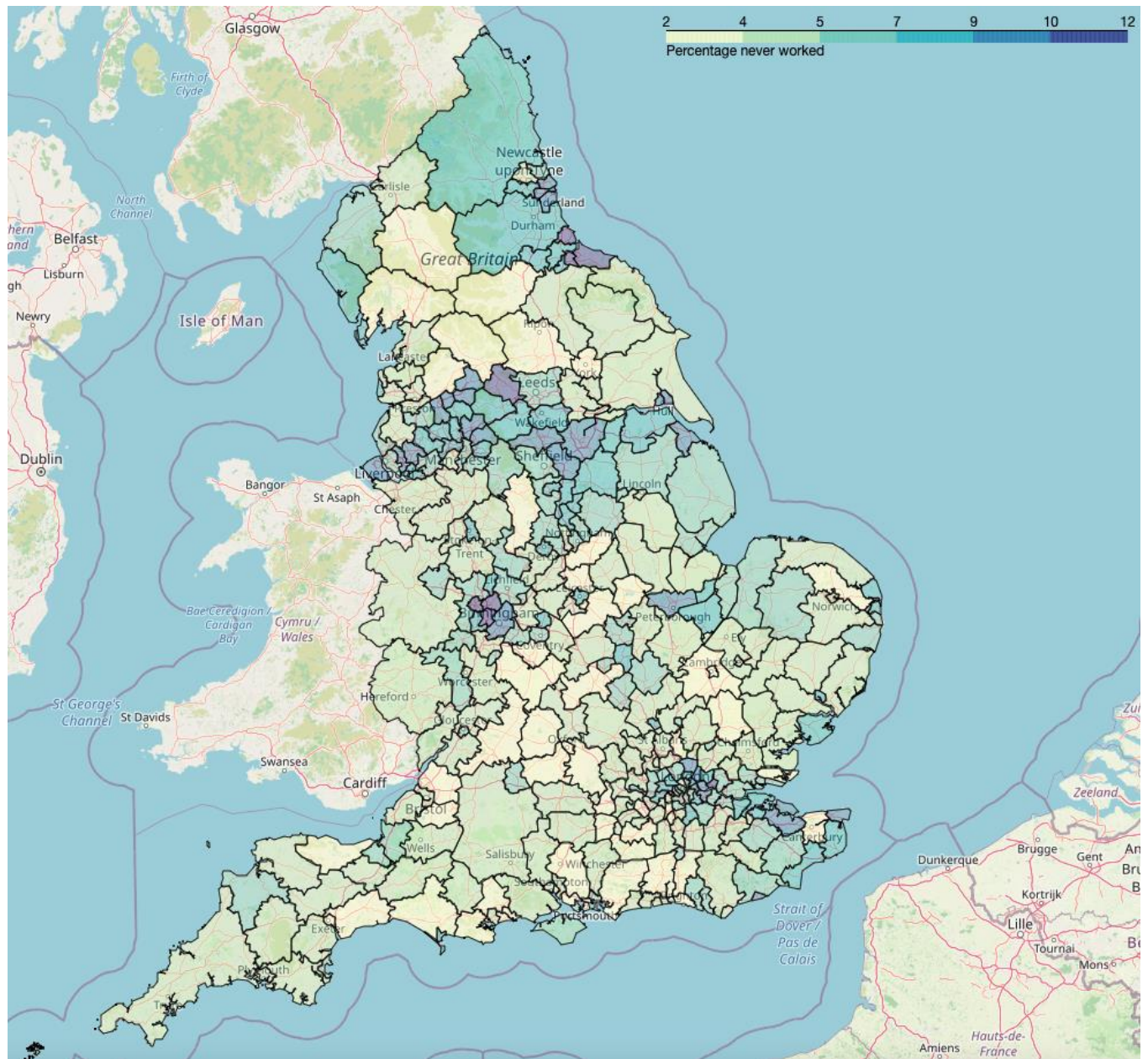
Figure 3: The percentages of people of 16-24 years old in each LAD who have never worked plotted on a map of England.

Figure 2 and especially figure 3 show that in and around some of the bigger cities the percentages are higher, while the less populated and more rural areas have low percentages.

Around Birmingham and the areas around Liverpool, Manchester, Bradford, Hartlepool, Middlesbrough and the eastern districts around London have the highest percentages of people under 25 who have never worked.

The link between the percentage of people who have never worked and the employment deprivation ranking of their area

I used the census dataset of all age groups and combined it with the deprivation dataset to get a table with columns for the 'Percentage never worked', the employment deprivation ranking, and the multiple deprivation ranking for each MSOA in England.

MSOA Name 2011	Percentage never worked	Average Multiple Deprivation Rank	Average Employment Rank
Birmingham 051	29.34	1014.40	1762.40
Birmingham 077	29.23	839.75	1876.00
Birmingham 082	28.60	1520.00	2902.75

Table 5: The census dataset and the deprivation dataset combined into one table.

This data is as I expected given the low deprivation rankings for the MSOAs with the highest percentage of people who have never worked. To get a better idea of the existence of correlation, I plotted a scatter chart of the 'Percentage never worked' against the 'Average employment rank'.
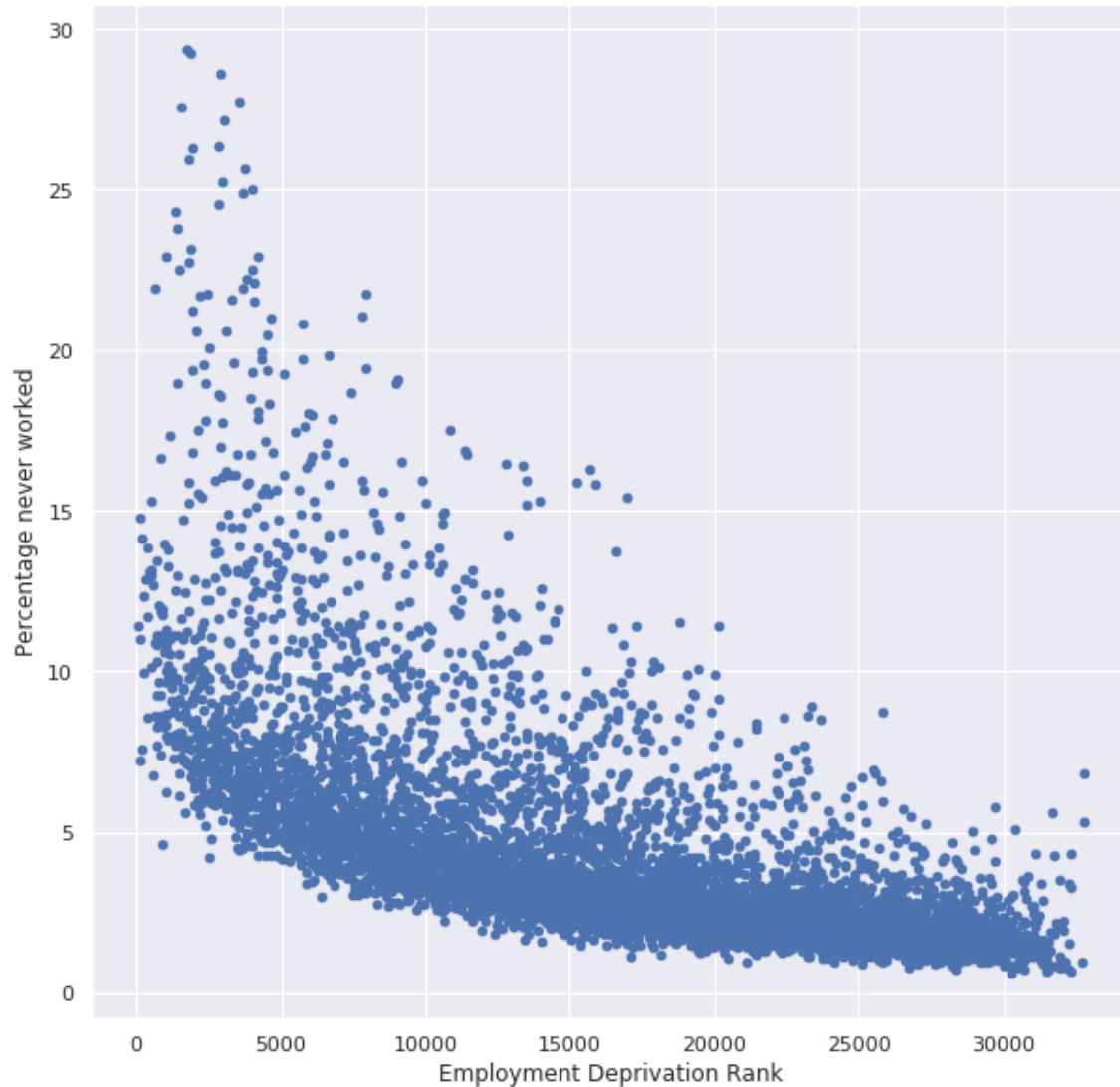
Figure 4: Scatter graph of the percentage who have never worked against the average employment rank.

This graph indicates that there is a correlation, but not very strong. I performed Pearson's test to see how strong the correlation is between the percentage of people who never worked and the average employment rank. It gave an r of -0.66, with p<0.001 (significant). This means a moderate negative correlation.

The scatter graph suggests that there may be a difference in correlation strength between the more deprived and the less deprived MSOAs. To check if that is correct, I used the kmeans data mining algorithm to divide the data points into two clusters (see figure 4).
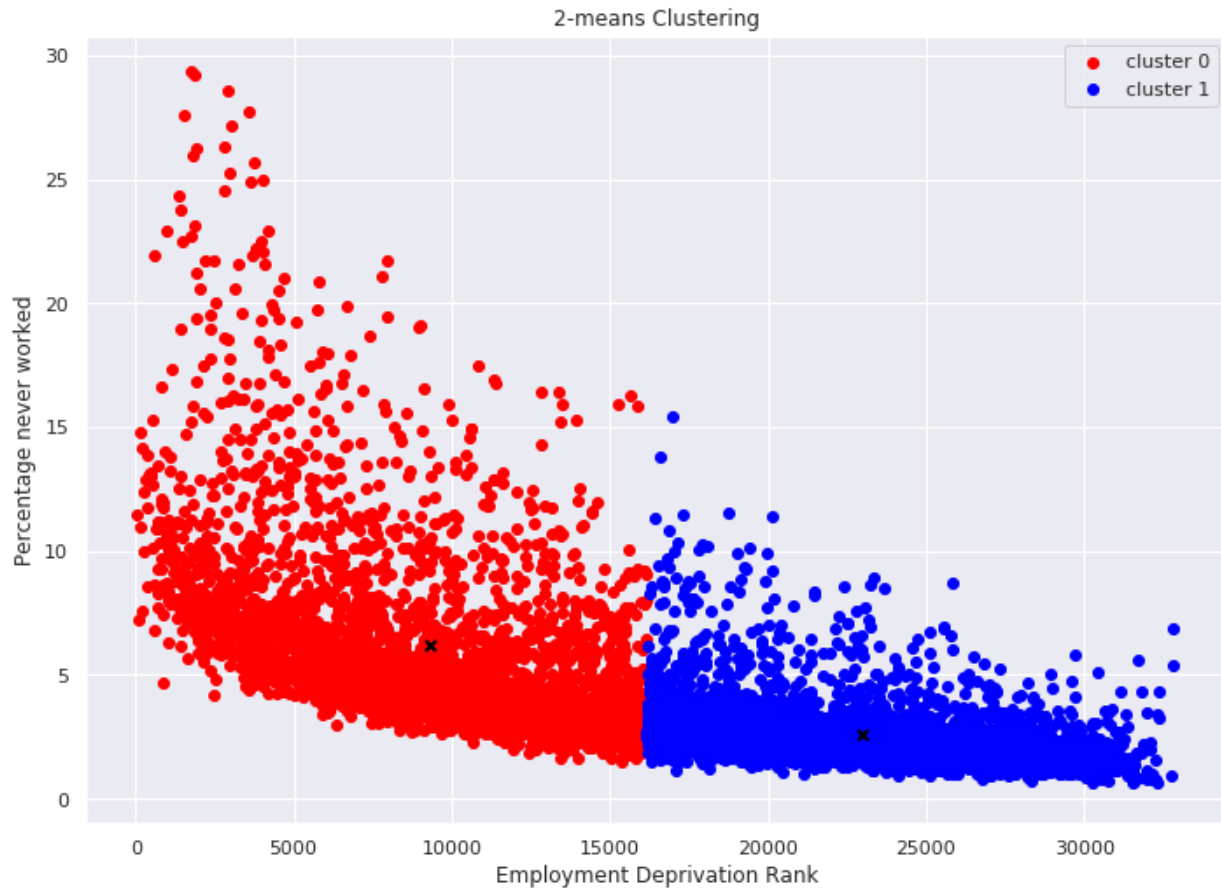
Figure 5: I divided the data points of the scatter plot into two clusters using data mining.

Using Pearson's test, I found that the r for cluster 0, the more deprived MSOAs, is -0.53 (p<0.001). In contrast, the r of the less deprived MSOAs is -0.38 (p<0.001). The correlation between the percentage of people who have never worked, and the employment deprivation rank is therefore stronger for the more deprived MSOAs.

Conclusions

The analysis based on the census dataset shows that the five MSOAs with the highest percentage of people from 16-24 years old who have never worked are County Durham 066, Medway 024, Bradford 045, Weymouth and Portland 008, and Kingston upon Hull 003.

It also shows that people in that age group from English LADs classified as 'Industrial and Multi-ethnic', 'Mining Legacy' or 'Manufacturing Legacy' are more likely to have never had a job than in other English LADs, especially those classified as rural, farming community or university town.

Based on the combination of the census dataset and the deprivation indices dataset, there is a moderate correlation between the percentage of people in a MSOA who have never worked and the employment deprivation ranking of that area.

14

In other words, it is likely that the lack of job opportunities in a MSOA is a reason why people there have never worked. Further data analysis on a lower level is needed to confirm this and to find other reasons before policies can be established to address this problem.

Moreover, we need to take in mind the limitations of the datasets these conclusions are based on, like the subjectivity of the deprivation indices and the census classifications and the fact that both datasets are (in part) based on surveys.

Reflection

Both datasets were easy to work with. There were no missing values and only minor data cleaning was needed. The answer to my first question was therefore not very hard to find.

I decided to use some data from the output area classification CSV-file for context, but unfortunately, there were only classifications for the LSOAs and the LADs, not for the MSOAs from the census table. I therefore settled for the classifications of the LADs.

To plot the data on a map was a bit tough because the kernel of Jupyter Notebooks shut down unexpectedly a few times while generating the map. I decided after that to skip the map for the second question.

Answering the second question was harder, especially because it took some time to figure out how to combine the two datasets. I stumbled on a few dead ends there, but after I figured out that I could use the LSOA name to get to the MSOA name and then take the average of the rankings, it became clear how to combine the datasets and get to the answer.

(2998 words)

References

- Chorley, M. (2017), *GeoJSON versions of UK Boundary Data* [Online]. Available at https://github.com/martinjc/UK-GeoJSON (Accessed 2 June 2021).

- Kleiweg, M. (2021), *TM351 TMA 02*, submitted to The Open University as part of TM351 assessment.

- Ministry of Housing, Communities & Local Government (2015), *The English indices of deprivation 2015* [Online]. Available at https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (Accessed 1 June 2021)

- Nomis (2013), *DC6206EW - NS-SeC by ethnic group by sex by age* [Online]. Available at https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=682 (Accessed 25 May 2021)

- Office for National Statistics (ONS) (2019), *Despite record employment, nearly 10% of adults have never done paid work* [Online]. Available at https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/economicinactivity/articles/despiterecordemploymentnearly10ofadultshaveneverdonepaidwork/2019-02-28 (Accessed 1 June 2021)

- Office for National Statistics (ONS) (n.d.), *The National Statistics Socio-economic classification (NS-SEC)* [Online]. Available at https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 (Accessed 1 June 2021)

- Office for National Statistics (ONS) (n.d.), *Quality and methods* [Online]. Available at https://www.ons.gov.uk/census/2011census/2011censusdata/2011censususerguide/qualityandmethods (Accessed 1 June 2021)

- UK Government (2015), *The English Indices of Deprivation 2015* [Online]. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/465791/English_Indices_of_Deprivation_2015_-_Statistical_Release.pdf (Accessed 1 June 2021)

Appendix 1: Notebooks

Full details of the data processing and analysis are given in the following Notebooks.

Notebook	Contents
mk8978_project_diary	Investigation of census data and deprivation data
LAD map	Creating a folium map for first question

Appendix 2: Data catalogue

The English indices of deprivation 2015 dataset used in this report came from the Ministry of Housing, Communities & Local Government and is hosted on Gov.uk https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467765/File_2_ID_2015_Domains_of_deprivation.xlsx.

The UK census dataset DC6206EW used in this report came from Nomis, a service provided by ONS that gives free access to UK labour market statistics from official sources. The dataset is hosted on Nomisweb https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=682.

The GeoJSON file that I used to plot the map with the 'Percentage never worked' for each Local Authority District came from the Github repository of Dr. Martin Chorley (https://github.com/martinjc/UK-GeoJSON) but was originally obtained from ONS.

These datasets are all public-sector information and licensed under the Open Government Licence v3.0. This licence allows me to adapt and exploit the information.