

Table of contents

<i>Introduction and Research Area</i>	<i>2</i>
<i>Obtaining the Data</i>	<i>3</i>
<i>'Skill Level' as a Measurement Tool</i>	<i>5</i>
<i>Location and Municipalities</i>	<i>5</i>
<i>Analysis</i>	<i>6</i>
<i>Interactive Folium Map</i>	<i>6</i>
<i>Distribution of Requested Job Skill Levels across Municipalities</i>	<i>8</i>
<i>Comparing Regions on Job Adverts and Skill Level</i>	<i>10</i>
<i>Considering Firm Size</i>	<i>11</i>
<i>Discussion</i>	<i>12</i>
<i>Conclusion</i>	<i>13</i>
<i>Bibliography</i>	<i>14</i>
<i>Appendix 1.</i>	<i>15</i>

Introduction and Research Area

Variance in socioeconomic structures between rural and urban areas is a current topic in Danish politics, exemplified by the government's plans to move public workplaces from the capital to other areas of Denmark¹. Research from the UK shows that one of the challenges facing rural areas is the low skill-level within the workforce, as well as restricted access to skills development for the firms.² This issue can be attributed to the small and sparsely distributed populations, remote locations, and the fact that firms in rural areas are skewed more towards small firms than in urban areas (Owen, Li & Green, 2013: v). If the challenges faced by rural areas in Denmark are similar to those of the UK, we could find it reflected in the profiles of advertised job vacancies; the proportion of adverts for high skilled jobs would be higher in urban areas than in rural areas, as companies settle in areas with a suitably competent workforce. The aim of this paper is to explore the variance between urban and rural areas in Denmark by examining the level of skill requested by job adverts; we wish to investigate our hypothesis that the demand for workers with a high skill level will be higher in urban areas than in rural areas.

Data from Statistics Denmark about job vacancies is available, however only from a national and regional observation - therefore the data is not as explicit in detail as it could be if it was taken from a municipal perspective. Furthermore, it only contains data from firms within Industry Codes B-N (DB07)³, and does not include job vacancies within public administration, teaching, health services, and other occupations that primarily fall within the public sector. To get a more geographically detailed and comprehensive occupational profile, we choose to obtain our data from JobNet (<https://job.jobnet.dk>), using data scraping techniques to capture information from every job advertisement (at the time of scraping) on job titles, employer, postcode, municipality, city, longitude and latitude. To create a measurement of skill level that we could apply to job vacancies, we will refer to the official Danish version of International Standard Classification of Occupations (DISCO-08, V.1.0:2011). Containing 9 categories, the DISCO classification groups occupations according to the similarity of their work requirements

¹ “Regeringen flytter 3.900 statslige arbejdspladser ud af København”, as seen on dr.dk/nyheder on October 1st, 2015.

² “Towards a one nation economy: A 10-point plan for boosting economy in rural areas” (August 2015), *Department for Environmental Food & Rural Affairs*.

³ Dansk Branchekode DB07. Available for download at [/www.dst.dk](http://www.dst.dk) > Statistik > dokumentation > nomenklaturer > dansk-branchekode-db07.

and specialization and ascribes each group a Skill-Level rating.⁴ Using this rating, associated with different occupations, we will construct a dictionary to create a new skill level variable. Rural and urban areas are defined on a municipal scale: urban, intermediate, rural, and peripheral, created by the Ministry of Food, Agriculture and Fisheries as part of the administration of the Danish Rural Development Programme (DRDP)⁵. With the locational information gathered from the job vacancies, we will investigate whether the distribution of high-skill level job advertisements is unequal, and much denser in urban areas of Denmark, than in rural areas. To visualize the results, we will use the longitude and latitude of the job adverts to create an interactive folium map displaying Denmark, and showing the location of every job advertisement as a marker. The markers will have colours corresponding to different Skill-Level classifications scattered in different regions according to job adverts.

For this project, we will use Python. There are many appropriate alternatives available; R, for example, is definitely capable of allowing us to manipulate and visualise our data, due to its large library of statistical packages. However, Python is simpler to master in a shorter time, allowing us to perform website scraping, machine learning and more. Although its library is not as extensive as R, the Pandas, Matplotlib and Folium tools were enough for our investigation.

We will use a range of effective visualization tools following. These are developed in accordance with the theory described in BDSS.⁶

Obtaining the Data

We used JobNet as the basis of our investigation due to its large database of jobs in Denmark; between 12,000 and 19,000 jobs are advertised each day, larger than alternative search engines such as www.jobsincopenhagen.com, and www.workindenmark.dk. The advantage of JobNet is its variety of jobs advertised in many different sectors of the DISCO scale. At the time of collecting the data, there were 13,422 job adverts. In order to scrape all of the required information from the available adverts, we constructed a code which generated the URL of every single page, containing the information of 20 job advertisements per page. This method follows the three-step procedure described in Edelman, B. (2012).

⁴ DISCO-08, *Danmarks statistiks fagklassifikation*, p. 7-9. København, Marts, 2011

⁵ "Official definitions of Nordic rural areas*", available on www.nordregion.se > About Nordregio > Noregion News and the journal of Noregio > Journal of Noregio 2010 > Journal of Noregio No 2, 2010.

⁶ Foster, Ian, et al. "Big Data and Social Science—A Practical Guide to Methods and Tools." (2017). ch 9.3

Since web pages are written in HTML format, and a lot of structural coding is included once downloaded, we had initially planned to use the Web Scraping method to extract information. Using Python's 'BeautifulSoup' library, we wrote a code to parse, search through, and reformat the data into a spreadsheet which could be structured into our dataset. However, we found that we were able to extract the data as a .json file from the JobNet server, using the API. This allowed us to preserve the data with readily assigned headings, allowing us easier interpretation of the data. The magnitude of the data we had collected, paired with the inconsistency of keywords used in job titles meant that we had to create a dictionary. The dictionary let us create 'keys' corresponding to keywords from the official index of occupational titles related to the main DISCO-groups.⁷ The index provides a map between "real world" terms used to describe jobs and the formally named titles of groups defined in the DISCO classification structure. Once the dictionary was made, sorting all occupations into 9 categories, we created yet another dictionary to sort those 9 categories into skill levels, assigning every type of occupation a value from 1 (lowest) to 4 (highest). This method of classification follows Grimmer, J., & Stewart, B. M. (2013). After placing all observations in DISCO categories, we were left with 12,409 observations, due to some job titles which did not fit into our model.

A similar dictionary method was used to sort the 98 municipalities into a urban-periphery continuum according to Danish Rural Development Programme (DRDP), created by the Ministry of Food, Agriculture and Fisheries.⁸ As well as adding a new variables to the dataset for analysis, we also drop the data columns which we deemed to be unnecessary in our investigation.

JobNet is a public website, with the purpose of making it easy for people to find vacancies. When considering the ethics of our data collection, we assessed that our collected data can be accessed by anyone with access to the internet, we therefore have not accessed information not available to the public. In our model we incorporated a limiter, so that we could control the speed at which we were downloading the data. We tested various limits to find the fastest way we could download the data without disrupting JobNet's server.

⁷ The index is available for download on Statistics Denmark's webpage www.dst.dk > Statistik > Dokumentation > Nomenklaturer > DISCO-08-i-loenstatistikken.

⁸ 'Classification of Municipalities', The Danish Rural Development Programme, 2007-2013. <http://www.nordregio.se>

Link:

‘Skill Level’ as a Measurement Tool

Skill, in the context of labour, is defined as the ability to carry out the tasks and duties of a given job with pre-determined results. When arranging the occupations from the DISCO scale into a ranked category, we considered the level of skill required to perform tasks related to the job. Our resulting skill level variable, used to test our hypothesis, was made with the dictionary created using DISCO-08 occupational groups (below).

Table 1 - Disco to ‘Skill Level’ converter

<i>DISCO groups</i>	<i>Skill level</i>
1 Managers	3+4
2 Professionals	4
3 Technicians and Associate Professionals	3
4 Clerical Support Workers 5 Services and Sales Workers 6 Skilled Agricultural, Forestry and Fishery Workers 7 Craft and Related Trades Workers 8 Plant and Machine Operators and Assemblers	2
9 Elementary Occupations	1

Source: “DISCO-08 i lønstatistikken”, 2. udgave. Danmarks Statistik. Dec. 2011. Pdf-udgave kan hentes på www.dst.dk/disco-08lon

Skill level, our resulting measuring tool, is defined as the complexity and range of tasks performed in an occupation, and ranks from 1 to 4; 1 being the lowest ranking, and 4 being the highest ranking. Higher degrees of competence and specialization are required for occupations higher up on the DISCO scale, such as Managers and Professionals, and lower rankings were attributed to Elementary Occupations, including workers in retail. This is shown in the ‘Mapping of DISCO-08 Major Groups to Skill Levels’ table (Table 1) which shows the International Standard Classification of Occupations. Eight of the ten Major groups in the DISCO-08 contain occupations in only one of the four Skill Level categories, with two having occupations which fall into two. Once we created the variable, arranging each of the many thousands of job entries into skill levels, we were able to move onto the next step of our investigation.

Location and Municipalities

The next step of the investigation was to visualise the job entries and their skill levels on a map so that we could inspect the distribution, and see if there was any correlation between specific skill levels and their density in urban and rural regions of Denmark. Our JobNet dataset provided information on the region, municipality, longitude, and latitude, and, initially, we planned to use

the 5 Danish regions as a measurement of geographical locations. However, we believed that regions did not give enough detail, and they are often comprised of both urban and rural areas, for example in Nordjylland. We decided to measure location by the 98 different municipalities in Denmark, for a more explicit perspective into our hypothesis. Using the DRDP classification of the 98 municipalities⁹, we categorized the municipalities on an urban-rural continuum consisting of four categories of municipality type: urban, intermediate, rural, or peripheral. The DRDP classification is based on 14 indicators weighted equally - including population demography, proportion of area in rural zones and mean distance to motorway¹⁰. With this information we were able to find out exactly how the various skill level jobs advertised were distributed in urban areas, intermediate areas, rural areas and remote areas.

Analysis

Interactive Folium Map

For a simple visualisation showing the distribution of various skill levelled jobs around Denmark, we used Python's Folium package with the Leaflet library to create an interactive map. Using the longitude and latitude provided by each job advert, Folium allowed us to create markers for every single job advert placed in Denmark. Due to the magnitude of the data, cluster close markers together for easier interpretation. These clusters are colour-coded based on their density: from green (small cluster) to orange (dense cluster). As you zoom closer to the map, the clusters split into smaller clusters, eventually showing individual entries in their exact location. When clicking on a single marker, one can view the job title. See Example (pictures) in appendix 1. When you zoom in on the map enough to view the street layout, you will notice the individual markers, coloured pink, red, green and black. Pink corresponds to the skill level 1, red to skill level 2, green to skill level 3, and black to skill level 4.

The interactive map showing information obtained from our data set includes areas covering the whole of Denmark, meaning that we are able to draw a stronger conclusion, as our data set contains a great magnitude of entries, covering many of the 98 municipalities. However,

⁹ The list of categorized municipalities is available on www.livogland.dk > lokale-aktionsgrupper > grundlag-lokale-aktionsgrupper > geografisk-grundlag-lokale-aktionsgrupper

¹⁰ "Official definitions of Nordic rural areas*", available on www.nordregion.se > About Nordregio > Noregion News and the journal of Noregio > Journal of Noregio 2010 > Journal of Noregio No 2, 2010.

due to missing information on longitude and latitude by some adverts, we were only able to place 10,231 job adverts from the 12,409 categorised into DISCO categories. Comparing the locations of the different skill levelled job adverts with the DRDP municipality classification scale, we are able to look at the demand for different skill levels in different areas. Thus, we can investigate whether it is true that there is more of a demand for high-skilled employment in urban areas, compared to rural areas. The interactive map can be downloaded using this link: [Interactive Map - Onedrive](#) ¹¹.

Comparing Municipality Categories on Job Adverts and Skill Level

Table 2 - Descriptive statistics on ‘Skill Level’ distributed by municipalities groups

<i>Area</i>	<i>Job adverts (Share)</i>	<i>Skill level (Mean)</i>
<i>Urban</i>	59%	2,83
<i>Intermediate</i>	14%	2,64
<i>Rural</i>	21%	2,62
<i>Peripheral</i>	5%	2,45

Using a table to view our findings and make an analysis, we took the four DRDP categories of municipality, and totalled the number of jobs advertised per category (job adverts (Count)), and compared that with the calculated mean Skill level of each category (skill level (Mean)). The results of the table show us that urban municipalities have the highest share of total job adverts, and also the highest mean skill level of all the different types of municipality, supporting our hypothesis.

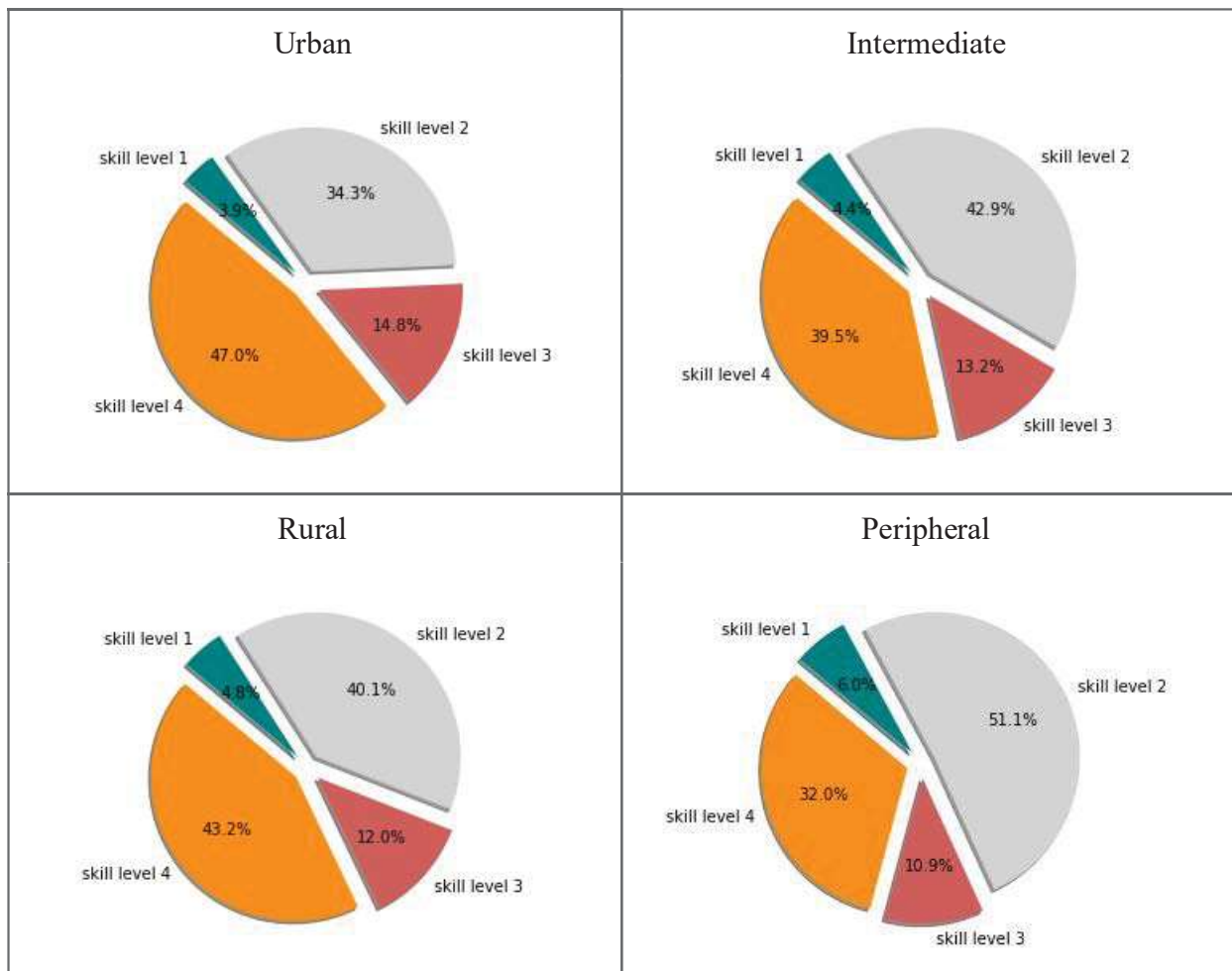
This finding, despite supporting our hypothesis, is not a very strong correlation. This could result from the larger population and wider population demographic in urban areas; they have a wider variety of industries than rural and peripheral areas, and most likely require workers with a diverse skill level set. For example, there are job roles associated with higher skill level requirements which are vital for urban city areas, such as Bankers, Managers, University professors. However, there are also many jobs associated with low skill levels, such as sales assistants, cleaners, and workers in craftsmanship which are also vital for urban city areas due to the tourism industry, and infrastructural needs.

¹¹ You have to press download in the onedrive folder and open it, otherwise it will only show the HTML code. Direct link for onedrive folder - <https://1drv.ms/u/s!Ape6rozO2wamlXjyBaSqlMZKNiTq>

Distribution of Requested Job Skill Levels across Municipalities

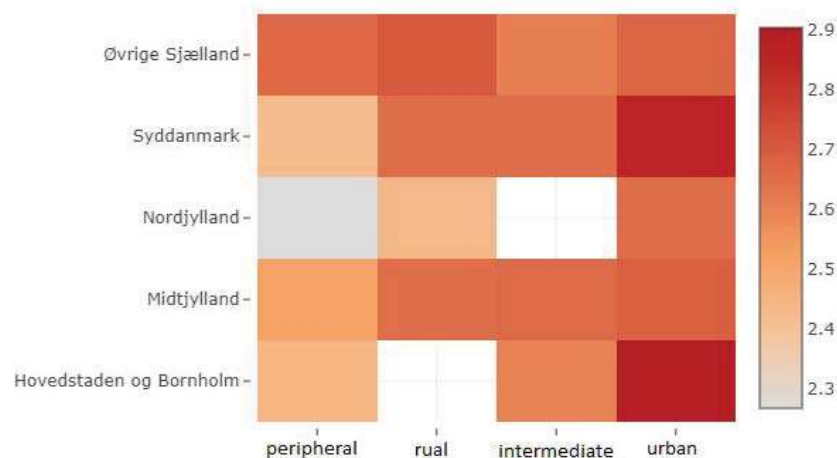
The pie charts (below) show the distribution of jobs belonging to specific skill level categories in the different types of municipalities. In the urban municipalities, we can see that the largest share of vacancies were for occupations in the skill level 4 category, at 47,0%, indicating that urban municipalities are generally searching for high-skilled employees. Rural municipalities also have the majority of their jobs advertising occupations associated with skill Level 4, at 43.2%. However, the intermediate municipalities, although a large part of the advertisements were for skill level 4 jobs, the majority of jobs were for skill level 2, at 42,9%. Peripheral municipalities had over half of their job advertisements for skill level 2 jobs, at 51,1%. Peripheral also had the highest percentage of skill level 1 jobs, at 6,0%, whereas the other categories ranged between 3,9% and 4,8%.

Pie Charts 1 - 'Skill Level' distribution by municipality groups



From the pie charts, a pattern emerges; in highly-populated areas, the quantity of job advertisements requiring employees with a high-skill background is higher than in less populated areas. Similarly, lesser populated areas had higher quantities of job vacancies requiring employees with a background with level 1 and 2 skills, which supports our hypothesis.

Heatmap 1 - Region, area type and mean skill level



Heatmap 1 gives an insight into the mean skill levels of the 5 Danish regions, and further detail into the performance of their different municipality groups types. It shows the mean skill level required by jobs, ranging from 2.3 to 2.9 with the higher means taking a darker colour. The absence of data, shown, for example, in Hovedstaden og Bornholm's rural area, is shown by an empty square. Heatmap 1, created using Plotly, explicitly shows that the demand for high skilled workers is concentrated in urban areas. For example, if we look at Hovedstaden og Bornholm's urban areas, they are at the top of the scale when compared with other municipality types. In general, you can see that throughout the different regions, the urban areas are darker, and show a higher mean of high skilled jobs available, supporting the findings of the pie charts. The intermediate areas on the heatmap show the second highest means of job skill required, after urban. The peripheral column stands out, showing a lower mean for every region, in comparison to urban, intermediate and rural areas. In region Nordjylland the urban areas has the highest mean skill level for the postings, and it is clear to see that the rural and peripheral areas has a lower mean skill level for the job adverts in their areas. Regions Syddanmark and Hovedstaden og Bornholm show similarity with the concentration of high skilled job adverts being concentrated in urban areas. Regions Øvrige Sjælland and Midtjylland, are similar in having

their mean skill-level is more spread out through the areas. The heatmap can be used to support our hypothesis that the concentration of job advertisements requiring high-skilled workers are higher in urban areas, in comparison to other areas.

Comparing Regions on Job Adverts and Skill Level

Table 3 - Descriptive statistics on ‘Skill Level’ distributed by regions

<i>Region</i>	<i>Job adverts (Share)</i>	<i>Skill level (Mean)</i>
<i>Hovedstaden og Bornholm</i>	37%	2,88
<i>Midtjylland</i>	24%	2,65
<i>Nordjylland</i>	8%	2,50
<i>Syddanmark</i>	20%	2,70
<i>Øvrige Sjælland</i>	12%	2,65

In a simpler view of Denmark than Table 2, we can see how the different regions perform in their mean skill levels, and inspect the distribution of their job adverts. Hovedstaden og Bornholm accounts for 36,75% of the total job adverts and also has the highest mean skill level. This supports the argument mentioned in the discussion above the with heat map, where it was clear most of the job adverts in Hovedstaden and Bornholm were high-skilled job adverts in urban areas. The same also occurs for Syddanmark, which also has a relatively high mean skill level, and accounts for 20,03% of the job adverts.

Midtjylland and Øvrige Sjælland have the same mean skill level. As mentioned above in regards to the heatmap, we could see that the job adverts were more diverse, further supported by identical mean skill level. Together they can account for 35,14% of the job adverts. Nordjylland is distinct, by accounting for the smallest amount of job adverts, and having the lowest mean skill level. This could be due to the fact that half of its municipalities are classed as peripheral, which generally advertise jobs which require a low skill level.

Considering Firm Size

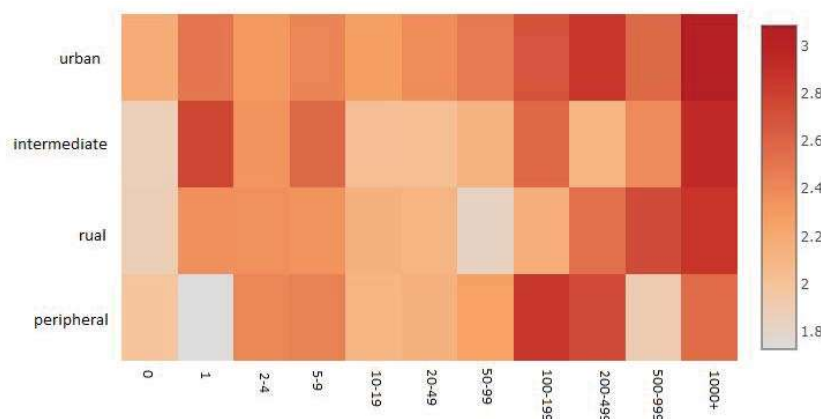
Table 4 - Firm sizes distribution by municipality groups

Area	0	1	2 - 4	5 - 9	10 - 19	20 - 49	50 - 99	100 - 199	200 - 499	500 - 999	1000+
Urban	44%	56%	49%	54%	60%	59%	52%	70%	59%	48%	59%
Intermediate	19%	18%	14%	15%	14%	17%	17%	12%	11%	15%	13%
Rural	30%	23%	28%	25%	21%	20%	23%	14%	25%	33%	21%
Peripheral	6%	3%	9%	6%	5%	4%	9%	4%	6%	4%	6%

The Danish Business Authority has information on Danish companies publicly available on their webpage virk.dk.¹² We used the list of CVR-numbers, unique to the companies, from our JobNet dataset, and extracted information on the number of employees (in full time equivalents) from each company via <https://datacvr.virk.dk>. We then merged the two datasets to create a new variable, showing the size of the company, in terms of employees.

As we can see from Table 4 (above), the majority of bigger firms are located in the urban municipalities. As noted in Table 2, urban areas also account for 59,0% of all job adverts. As firms get bigger, they tend to be located in urban areas.

Heatmap 2 - Area type, firm size and mean skill level



Heatmap 2 plots mean skill level against municipality type and company size, and shows that firms with more than 1.000 employees tend to have a higher mean skill level. This applies for all areas. There is a correlation between the size of the firm, and the level of skill that they require from potential workers; the bigger the firm, the higher mean skill level. This observation,

¹² Erhvervsstyrelsen

combined with the fact that urban areas account for 59% of the job adverts from firms with more than 1.000 employees can partly explain why urban areas tends to require a higher mean skill level from employees.

From the data we collected we can see a correlation between area, skill level and the size of the firms. By combining the three variables we are able to tell whether a certain area demands high or low skilled employment.

Discussion

Although the model we created for the investigation was successful in extracting all the necessary information required when testing our hypothesis, we are aware that it has many flaws. The first concerns our method of data collection; we used only one source for our data extraction: JobNet. Despite its wide database, it does not give information on every single job advertised in the country of Denmark; it cannot give us information on job adverts that are advertised locally in small or closed communities. An alternative source of information for the investigation could have been <https://jobindex.dk>. However due to time constraint on the project, it was a choice of one between the two. Additionally, our findings are only a snapshot of the day in which we gathered the information. Unless we were able to create a model which could update itself daily with JobNet, our findings are arguably only an insight into the distribution of high skill level jobs adverts on 18/08/2017. Secondly, our model was not entirely efficient when categorizing all of the entries from our dataset into DISCO categories; some jobs had no keywords that our model could recognise, and some jobs had keywords that fit into more than one category.

However, despite the drawbacks, our findings provide, generally, an interesting insight into the location of the supply of high and low skilled jobs. It's implications in wider society could be as a tool for people to identify where their skill set would provide better work opportunities. Referring back to the Danish government's plan to lessen the disparity between rural and urban areas, our findings could spark awareness into how the location of big firms and the spread of high skilled jobs could be influencing their disparity.

Conclusion

This project was an inquiry into the distribution of job skill level across urban and rural municipalities in Denmark. We sought to investigate our hypothesis, that the demand for workers with higher skill levels would be greater in the urban areas of Denmark, than in rural areas. From the data we created, manipulated and visualised, our analysis can conclude that there is a correlation between municipality type and the requested mean skill level: in urban areas, there was a higher demand for high skilled workers, than in rural areas. In the data we also found that there was a tendency for firms with a larger workforce to be located in urban areas, insinuating that these larger firms are looking for workers with a highly skilled background. This, combined with their urban location may be used to explain why urban areas have the highest mean skill-level. Overall, we can conclude that there is correlation between the type of municipality and mean skill-level municipality type and the size of the firm. These two correlations can be explained by the correlation of firm size and mean skill-level.

Bibliography

Articles.

- *DISCO-08, Danmarks statistiks fagklassifikation*, 2. udgave,. København, Marts, 2011. Available for download on www.dst.dk/DISCO-08
- Edelman, B. (2012). Using internet data for economic research. *The Journal of Economic Perspectives*, 26(2), 189-206.
- Foster, Ian, et al. "Big Data and Social Science–A Practical Guide to Methods and Tools." (2017). ch 9.3
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.
- Owen, D., Li, Y. and Green A. (2013) “Secondary analysis of employer surveys: Urban and rural differences in jobs, training, and skills”, *UK Commission for Employment and Skills, Evidence Report 75, October 2013*, UK Commission for Employment and Skills, Wath-upon-Deane.
- “Towards a one nation economy: A 10-point plan for boosting economy in rural areas” (August 2015), Department for Environmental Food & Rural Affairs,

Webpages

- “*Regeringen flytter 3.900 statslige arbejdspladser ud af København*”, as seen on dr.dk/nyheder on October 1st, 2015.
- “*DISCO-08, V.1.0: 2011*”. www.dst.dk > Statistik > Dokumentation > Nomenklaturer > DISCO-08-i-loenstatistikken > koder og kategorier > Download > csv. For a list of the official index of occupational titles related to the main DISCO-groups. As seen on aug. 21st.
- “*Official definitions of Nordic rural areas**”, available on www.nordregion.se > About Nordregion > Nordregion News and the journal of Nordregion > Journal of Nordregion 2010 > Journal of Nordregion No 2, 2010. As seen on aug. 21st
- “*Geografisk grundlag for lokale aktionsgrupper*”, www.livogland.dk > lokale-aktionsgrupper > grundlag-lokale-aktionsgrupper > geografisk-grundlag-lokale-aktionsgrupper. as seen on aug. 21st.

Appendix 1.

