

Open data and the labor market: a study of vacancies in the forestry industry

*Ilshat Z. Garafiev*², *Farida T. Nezhmetdinova*^{1,*}, *Raushaniya I. Zinurova*², and *A. R. Tuzikov*²

¹ Kazan State Agrarian University, 65, K. Marx St., Kazan, 420015, Republic of Tatarstan, Russian Federation

² Kazan National Research Technological University, 68, K. Marx St., Kazan, 420015, Republic of Tatarstan, Russian Federation

Abstract. The article deals with a study of vacancies in the forestry industry, based on data from the “Work of Russia” portal. The study analyzes various job parameters such as company size, type of employment, work schedule and salary offered. Most of the vacancies were published in December 2023 and relate to the profession “Forester (senior state forest protection inspector)”. Offered salary is about 20 000-35 000 rubles a month. The majority of vacancies posted on the portal are posted by small companies (85.7%) and do not require previous work experience, but require full-time employment (95.1%) and a full-time work schedule (64.7%). The typology of vacancies based on the level of minimum and maximum wages showed that the optimal choice is to identify two clusters. The first cluster included vacancies with a salary of 40,000 rubles or more, the second cluster - with a salary less. This work is an example of the use of open data and can be used by employers and job seekers to make more informed decisions in the labor market.

Key words: forestry industry, rural youth, socialization, rural area, society, development, education, skills, infrastructure, social networks.

1 Introduction

In the modern labor market, vacancies posted publicly on online recruitment platforms are the main source of information about the characteristics of labor demand. The presence of a standardized template for posting a vacancy requires the employer to indicate the key parameters of his offer: salary level, type of employment, work schedule. Information about vacancies accumulated on portals is not only of purely practical value for job seekers, but is also significant as a source of scientific research. Employees are constantly faced with the need to make decisions regarding whether they need to stay in their current job or change it, how to find out about available vacancies, and what criteria to choose from employers' offers.

In this regard, the study of the entire set of parameters of vacancies posted by employers becomes especially relevant.

The forestry labor market is one of the most important and sought-after sectors of the economy. Due to the active development of the industry, growing environmental responsibility and the need to preserve forest resources, the demand for specialists in this field continues to increase.

The main professions in forestry include foresters, arborists, rangers, forest conservation and protection specialists, reforestation engineers.

Foresters are engaged in organizing work on the protection, protection and reproduction of forests, monitoring compliance with forest legislation, carrying out measures to combat forest fires, pests and forest diseases.

Foresters are engaged in the development and implementation of forestry systems, conduct scientific research and develop new methods of forestry, assess the state of the forest fund, and monitor the use of forest resources.

2 Materials and methods

The work provides an analysis of vacancies posted on the portal of the Federal Service for Labor and Employment “Russian Work”. Information is published in the “Open Data” section in the format of multi-level XML structures. The data set is called “Vacancies of all regions of Russia from the EDS “Work in Russia”. It contains data on vacancies from the “Work in Russia” portal.

Vacancies are created from employers' personal accounts on the portal. The data is updated daily (the data set for January 10, 2024 is used for analysis in this

* Corresponding author: nadgmi@mail.ru

article). The following attributes are available for vacancies: information about the employer, position, proposed salary, required work experience, etc. Data collection, processing and analysis were carried out using the pandas, numpy, matplotlib, seaborn, sklearn libraries for the general-purpose Python programming language.

The typology of vacancies according to the values of the maximum and minimum offered wages was made using the clustering method. The Elbow method (“bent knee method”, also known as the “scree method”) was used to determine the optimal number of clusters; the clusters themselves were determined by the K-means method.

3 Results and discussion

The initial data set contains 77 variables, two of these variables allow the selection of only those variables related to the forestry industry. The variable “code profession” is the profession code in accordance with the all-Russian classifier of worker professions, employee positions and tariff categories OK 016-94.

In total, in this classifier we selected 29 professions related to the forestry industry; by selecting only them we received a data set of 289 vacancies/

The variable “code_professional_sphere” contains information about the industry of the vacancy; based on the value of this variable, we selected only vacancies related to the “Forestry, woodworking, pulp and paper industries.” Our final data set includes 224 vacancies.

Our dataset represents 18 of the 29 forestry occupations. Only 7 professions appear in 20 or more vacancies. Thus, the profession “Forester (senior state inspector for forest protection)” appears in 49 vacancies, which is 21.9% of the total number of vacancies, the profession “Forest Master (local state inspector for forest protection)” appears 28 times, which is 12.5% of the total number of vacancies. The profession “State forest protection inspector” occurs 26 times, which is 11.6% of the total number of vacancies, and the profession “Forester (state forest protection inspector)” appears 25 times, which is 11.2% of the total number of vacancies. The profession “Forester” appears 23 times, which is 10.3% of the total number of vacancies, the profession “Forester” appears 20 times, which is 8.9% of the total number of vacancies, and the profession “Forest protection engineer” appears 20 times. which is 8.9% of the total number of vacancies.

The dataset contained a variable published_date - the date of publication of the vacancy; it was translated into a time format - year, month, day of publication of the vacancy. After that, based on the published_date variable, we formed the year, month, day of week variables indicating the year, month and day of the week of publication of the vacancy. All vacancies were published on the website in 2023; the distribution of vacancies posted on the website by day of week and month is presented in Table 2.

Table 1. Distribution of job titles in vacancies

Job title	Count	Frequency (%)
Forester (senior state forest protection inspector)	49	21.9%
Master of Forest (local state forest protection inspector)	28	12.5%
State Forest Protection Inspector	26	11.6%
Forester (state forest protection inspector)	25	11.2%
Arborist	23	10.3%
Forest feller	20	8.9%
Forest protection engineer	20	8.9%
Lumberjack	9	4.0%
Helper in logging	5	2.2%
“Navalshchik” - timber piler	4	1.8%
Other values (8)	15	6.7%

Table 2. Distribution of job postings on the website by day of week and month

	Mon	Tue	Wed	Thu	Fri	Total
July					1	1
August		2	1			3
September	4	4		4	4	16
October	1	1	2	3	3	10
November	2	7	28	11	13	61
December	29	29	16	22	37	133
Total	36	43	47	40	58	224

Our data set includes only vacancies that were current as of January 10, i.e. These are the vacancies for which the employer is still recruiting, therefore most of the vacancies were published in previous months, for example, in December - 133 vacancies and in November - 61 vacancies.

The distribution of posted vacancies by day of the week in December and November was different, so in December more vacancies were posted on Fridays - 37 vacancies, and in November on Wednesdays - 28.

Information on vacancies includes data on the proposed salary; initially, employers fill it out in text form in the format: “from ...”, “from ... to ...”, “to ...”. Filling out the field is not necessary to post a vacancy on the portal.

In the dataset “Vacancies of all regions of Russia from the UDS “Work in Russia”, two variables “salary_max” - the size of the maximum offered salary and “salary_min” - the size of the minimum offered salary were created.

Some employers indicated only the minimum wage and did not indicate the maximum wage. We took all unfilled values of the “salary_max” variable equal to the values of the “salary_min” variable. Also, data on the required work experience in years in the dataset is represented by the “required_experience” variable.

Descriptive statistics of the maximum and minimum wages offered, and previous work experience for all regions are presented in Table 3.

Table 3. Descriptive statistics, required experience, maximum and minimum salary offered

	salary_ min	salary_ max	required_ experience
count	224	224	224
nunique	78	77	5
mean	27812,26	30334,59	0,53
min	0,00	0,00	0,00
25%	20838,75	22741,25	0,00
50%	25000,00	27125,00	0,00
75%	30000,00	35000,00	0,25
max	96006,00	140000,00	5,00
mode	20000(20)	35000(19)	0(168)

The unique number of values for the maximum salary offered is 77 and for the minimum salary offered is 78. The minimum value listed in the vacancies for the maximum salary offered and for the minimum salary offered is 0. The 25th percentile for the minimum salary offered salary is 20838.75 rubles, and for the maximum value of the proposed salary it is equal to 22741.25 rubles

The median level for the minimum value of the proposed salary is RUB 25,000.00, and for the maximum value of the proposed salary it is RUB 35,000.00. The 75th percentile for the minimum value of the proposed salary is RUB 30,000.00, and for the maximum value of the proposed salary it is equal to RUB 35,000.00

The maximum value of the minimum proposed salary is RUB 96,006.00, and for the maximum proposed salary is RUB 140,000.00.

The average value of the minimum proposed salary is 27,812.26 rubles, and for the maximum proposed salary is 30,334.59 rubles.

The most common value of the minimum proposed salary is 20,000 rubles. it occurs 20 times, and for the maximum proposed salary it is equal to 35,000 rubles. It occurs 19 times.

The unique number of values for previous experience is 5, the minimum value, 25th percentile and median are 0, the 75th percentile is 0.25 years. The maximum value is 5 years. The most common value is 0 years, occurring 168 times.

The distribution of vacancies by company size is presented in Table 4.

Table 4. The distribution of vacancies by company size

Value	Count	Frequency (%)
SMALL	192	85.7%
MICRO	19	8.5%
LARGE	7	3.1%
MIDDLE	5	2.2%
BIG	1	0.4%

Based on the table data, we see that most of the vacancies (192 or 85.7%) were posted by small companies, 1 vacancy or 0.4% was posted by a large

company. As for medium-sized companies, they posted only 5 or 2.2% of vacancies. Large companies posted 7 vacancies or 3.1%, and micro companies posted 19 vacancies or 8.5%.

The data set contains a variable called “busy_type”, which reflects the type of employment within the vacancy. The distribution of vacancies by employment type for all regions is presented in Table 5.

Table 5. Distribution of vacancies by employment type

Value	Count	Frequency (%)
Full employment	213	95.1%
Temporary	6	2.7%
Seasonal	4	1.8%
Internship	1	0.4%

Based on the data in Table 5, we see that the majority of vacancies 21 or 95.1% involve full-time employment, 6 vacancies or 2.7% involve temporary employment, 4 or 1.8% of vacancies involve seasonal work, one vacancy involves an internship.

The data set also contains a variable “schedule_type”, reflecting the work schedule within the vacancy. The distribution of vacancies by work schedule for all regions is presented in Table 6.

Table 6. Distribution of vacancies according to work schedule

Value	Count	Frequency (%)
Full time work	145	64.7%
Irregular working hours	65	29.0%
Rotational work	7	3.1%
Shift work	6	2.7%
Flexible schedule	1	0.4%

Based on the data in Table 6, we see that the majority of vacancies (145 or 64.7%) require full-time work, 6 vacancies or 2.7% require a shift schedule, and 65 or 29.0% of vacancies involve irregular working hours. 7 vacancies also offer rotational work.

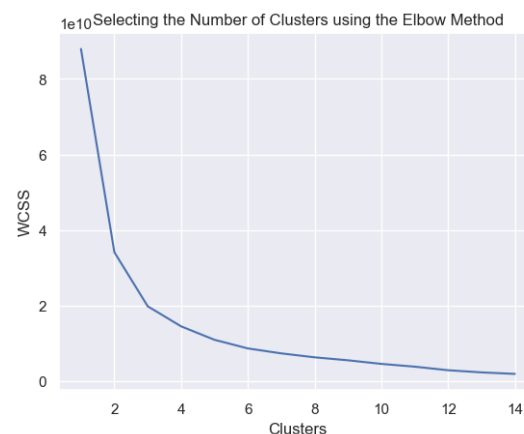


Fig. 1. Determining the optimal number of clusters for maximum and minimum wages offered.

To identify the typology of vacancies based on the values of the maximum and minimum wages offered, we used the clustering method. To determine the optimal number of clusters, we used the “elbow method” (“bent knee method”, also known as the “scree method”).

This method is based on constructing a graph where the number of clusters is marked on the abscissa axis, and the values of the WCSS function of the sum of squares within a cluster are marked on the ordinate axis, which determines the intragroup scatter depending on the number of clusters. From this graph, the optimal number of clusters is determined at the point where the “bend” of the curve appears.

Based on the graph, we see that two is the optimal number of clusters, since this is where the “bend” of the curve appears. The silhouette coefficient is a parameter defined to evaluate the quality of the clusters generated by the algorithm. Scores range from -1 to +1. The higher the silhouette score, the better the model. The symptom score measures the difference between all data points in one cluster.

The smaller this distance, the better the silhouette estimate. The higher this distance, the better.

We calculated the structure of the just built model for clusters 2, 3, 4, 5, 6 and 7. The result was 0.78 for 2 clusters, 0.49 for 3 clusters, 0.5 for 4 clusters, 0.5 for 5 clusters, 0.5 for 6 clusters, 0.44 for 7 clusters.

Next, we visualized the newly built clusters to get an idea of how well our model performs with two given clusters (Fig. 2).

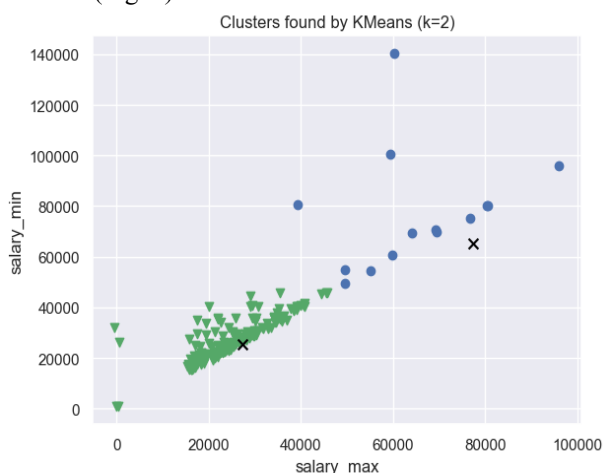


Fig. 2. Distribution of clusters for the maximum and minimum offered wages using the method K-means.

In order to assess the accuracy of the distribution by clusters, a graph of the proximity of values to the centroids of the clusters was plotted using the K-means method.

Based on the data presented in Fig. 2. we can describe the location of the centroids.

The centroid of the first cluster is at a level of more than 60,000 rubles. for the minimum wage and just under 80,000 rubles. for maximum wages. The centroid of the second cluster is at a level of more than 20,000 rubles, but less than 40,000 rubles, according to the values of the maximum and minimum wages offered.

In total, 14 vacancies were included in the first cluster and further we will dwell in detail on their analysis. The distribution of vacancies in the 1st cluster by type of company and region is presented in Table 7.

Table 7. Distribution of vacancies of the 1st cluster by type of company and region

Job position	Type of company	Region
Forest feller	LTD	Irkutsk region
	LTD	Primorsky Krai
	LTD	Irkutsk region
	Public joint stock company	Primorsky Krai
	LTD	Transbaikal region
	Production & construction cooperative	Primorsky Krai
Chief forester	Government agency	Komi Republic
Forest protection engineer	Government agency	Krasnoyarsk region
Forester	Government agency	Magadan region
Forester (senior state security inspector)	Government agency	Murmansk region
	Government agency	Khabarovsk region
Arborist	LTD	Irkutsk region
Lumberjack	LTD	Altai region
	LTD	Irkutsk region

Based on the data in Table 7, we see that the first cluster included 14 vacancies, which were posted for 7 positions, 6 vacancies were posted for the position of “forest feller”, two positions were posted for the position of “logger”, two vacancies were posted for the position of “forester” indicating that this is the position of senior state forest protection inspector. One vacancy is posted for the position of “forester”, indicating that this position is the chief state inspector for forest protection.

There is one vacancy each for the positions of “chief forester”, “forest conservation and protection engineer”, and “forester”. Six vacancies are posted by companies that are limited liability companies, of which 4 vacancies are for the position of “forest feller”, two vacancies for the position of “logger”, one vacancy for the position of “forester”.

5 positions are posted by government agencies, two vacancies for the position of “forest ranger”, one vacancy each for the positions of “chief forester”, “forest conservation and protection engineer” and “forester”.

For the position of “forest feller”, one vacancy was posted by a public joint-stock company (PJSC), one vacancy was posted by an industrial construction cooperative (PSC). Four vacancies were posted by companies from the Irkutsk region, three vacancies were posted by companies from the Primorsky Territory, one vacancy was posted by companies from the Altai Territory, the Trans-Baikal Territory, the Krasnoyarsk Territory, the Magadan Region, the Murmansk Region, the Komi Republic, and the Khabarovsk Territory.

Descriptive statistics of the maximum and minimum offered wages, previous work experience for vacancies from cluster 1 are presented in Table 8.

Table 8. Descriptive statistics of required experience, maximum and minimum offered wages for vacancies from the 1st cluster

	salary_ min	salary_ max	required_ experience
count	14	14	14
nunique	10	9	4
mean	65155,71	77298,57	1,43
min	40000	50346	0
25%	56250	62500	0
50%	62500	72914	1
75%	74371	80000	3
max	96006	140000	5
moda	60000(3)	70000(3)	0(6)

The unique number of values for the maximum proposed salary is 9, and for the minimum proposed salary is 10. The minimum value indicated in the vacancies for the maximum value of the proposed salary is 50,346 rubles, and for the minimum value of the proposed salary is 40,000 rubles.

The 25th percentile for the minimum value of the proposed salary is 56,250 rubles, and for the maximum value of the proposed salary it is equal to 62,500 rubles. The median level for the minimum value of the proposed salary is RUB 62,500.00, and for the maximum value of the proposed salary it is RUB 72,914. The 75th percentile for the minimum value of the proposed salary is 74,371 rubles, and for the maximum value of the proposed salary it is equal to 80,000 rubles.

The maximum value of the minimum proposed salary is RUB 96,006.00, and for the maximum proposed salary is RUB 140,000.00.

The average value of the minimum proposed salary is 65,155.71 rubles, and for the maximum proposed salary is 77,298.57 rubles.

The most common value of the minimum proposed salary is 60,000 rubles. it occurs 3 times, and for the maximum proposed salary it is equal to 70,000 rubles. it occurs 4 times.

The unique number of values for previous experience is 4, the minimum value and 25th percentile are 0, the median is 1 year, the 75th percentile is 3 years. the maximum value is 4 years. The most common value is 0 years, occurring 6 times.

Descriptive statistics of the maximum and minimum offered wages, previous work experience for the forest feller vacancies from cluster 1 are presented in Table 9.

Table 9. Descriptive statistics of required experience, maximum and minimum wages offered for forest feller vacancies from cluster 1

	salary_ min	salary_ max	required_ experience
count	6	6	6
nunique	4	4	2
mean	69334,33	89334,33	0,5
min	40000	60000	0
25%	60000	80000	0
50%	70000	80000	0,5
75%	80000	92004,5	1
max	96006	140000	1
moda	60000(2) 80000(2)	80000(3)	0(3) 1(3)

The unique number of values for the maximum offered salary is 4, and for the minimum proposed salary is 4. The minimum value indicated in the vacancies for the maximum value of the proposed salary is 60,000 rubles, and for the minimum value of the proposed salary is 40,000 rubles.

The 25th percentile for the minimum value of the proposed salary is 60,000 rubles, and for the maximum value of the proposed salary it is equal to 8,000 rubles. The median level for the minimum value of the proposed salary is 70,000 rubles, and for the maximum value of the proposed salary it is 80,000 rubles. The 75th percentile for the minimum value of the proposed salary is 80,000 rubles, and for the maximum value of the proposed salary it is equal to 92,004.5 rubles.

The maximum value of the minimum proposed salary is RUB 96,006.00, and for the maximum proposed salary is RUB 140,000.00.

The average value of the minimum proposed wage is 69,334.33 rubles, and for the maximum proposed wage is 89,334.33 rubles.

The most common value of the minimum proposed salary is 60,000 rubles. and 80,000 rub. each of them occurs 2 times, and for the maximum proposed salary it is equal to 80,000 rubles. it occurs 3 times.

The unique number of values for previous experience is 2, the minimum value and 25th percentile are 0, the median is 0.5 years, the 75th percentile is 1 year. the maximum value is 1 year. The most common values are 0 years and 1 year, each of these values appearing 3 times.

Our analysis of data on vacancies in the forestry industry showed that of the 29 professions associated with this industry, only 18 are represented in the data set. Most of the vacancies were published in December 2023 and relate to the profession “Forester (senior state forest protection inspector)”. Most vacancies offer salaries from 20,000 to 35,000 rubles and do not require previous work experience.

The majority of vacancies (85.7%) are posted by small companies, require full-time employment (95.1%) and a full-time work schedule (64.7%).

The typology of vacancies based on the level of minimum and maximum wages showed that the optimal choice is to identify two clusters. The first cluster included vacancies offering salaries starting from 40,000 rubles; the most in-demand profession was forest feller, which was represented in 6 out of 14 vacancies. The results obtained confirm that employers, by posting their vacancies in the public domain, are engaged in mass recruiting, primarily recruiting working personnel for the company. For the forestry industry, the main workforce is the forest feller, who performs heavy physical work, which is why the turnover in this position is high.

4 Conclusion

The analysis of vacancies in the forestry industry according to such parameters as company size, type of employment, work schedule and offered wages is an example of monitoring the state of the labor market. The possibility of such monitoring is available to any user of the “Work of Russia” portal that is a federal state information system, and a project of the Federal Service for Labor and Employment. Our work is an example of the use of open data; the methods of analyzing vacancies described in it can be used by employers and job seekers to make more informed decisions in the labor market.

References

1. I. Garafiev, G. Garafieva, Territorial features of value orientations and satisfaction with the studies of student youth in the Republic of Tatarstan. E3S Web of Conferences, 210, 18134 (2020).
2. M. Cibik, R. Stepankova Exploring university campuses as urban development boosters and design flagships in urban landscapes. AD ALTA, Journal of Interdisciplinary Research, **10(02)**, 37-45 (2020).
3. F.T. Nezhmetdinova, G.R. Fassakhova, L.R. Shagivaliev, N.Kh. Sharypova, R.I. Zinurova, Digital economy and transformation of personnel training for AIC, Bio Web of Conferences, **17**, 00228 (2020)
4. A.R. Tuzikov, R.I. Zinurova Sociology of students: theoretical status and research practices. Higher education in Russia, **28(6)**, 40-51 (2019).
5. A. Tuzikov, R. Zinurova, Features of the social order for two-level vocational training. Higher Education in Russia, **5**, 113-121 (2008).
6. S. Dyakonov, A. Tuzikov, R. Zinurova, L. Grebnev, Y. Krupnov, Y. Tsarik, Corporate university based on project-based education as a tool for innovative development, Higher Education in Russia, **11**, 3-15 (2006).
7. F.T. Nezhmetdinova, M.E. Guryleva, N.Kh. Sharypova, R.I. Zinurova, A.R. Tuzikov Risks of modern biotechnologies and legal aspects of their implementation in agriculture, Bio Web of Conferences, **17**, 00227 (2020)
8. A.R. Tuzikov, R.I. Zinurova, E.Y. Turner, E.V. Bardasova, E.Y. Girfanova, I.G. Morozova, E.D. Emelina, Global challenges of the 21st century and possible university's answer, Ekoloji, **28(107)**, 33-38 (2019).
9. A.R. Tuzikov, R.I. Zinurova, Ecological imperative in multicultural city agglomeration: ecosphere vs techno sphere, Ekoloji., **28(107)**, 5059-5063 (2019).
10. F. T. Nezhmetdinova, N. S. Barabash, Innovation and expertise, Scientific works, **2 (23)**, 120–131 (2018).