

Educational and Career Trajectories in Russia: Open Datasets

Datasets general description

The datasets contain data retrieved from the “Jobs in Russia portal”. The datasets contain information about the education, work experience, desired salary and some demographic characteristics of candidates who posted their CVs on the portal between 2015 and 2023. The datasets cover 83 regions of Russia. Dataset 1 contains information on the education and career trajectories (N=6221439), while Dataset 2 contains the information about the general activity of unemployed and job seeking candidates (N=7662089). Each candidate has a unique ID and there are characteristics such as age, education type, region, working experience. The datasets are available as .csv files.

Dataset 1 is designed to study educational and career trajectories. It was obtained by aggregating all CVs for each candidate into a single record with a unique *id_candidate*, then normalizing text information about the candidates' work experience and education, and finally excluding records that do not contain required information.

Dataset 2 includes all records for the available period, but only covers general information and the activity metric. It might be useful to study the general portal dynamics and, therefore, the activity on the Russian labor market.

Dataset 1

General information	
Dataset full name	Educational and career trajectories in Russia
Dataset short name	Educational and career trajectories
Dataset brief description	The dataset contains data on education and career trajectories of candidates who posted their CVs on the Jobs in Russia portal. Each column corresponds to a candidate with a unique <i>id_candidate</i> , their age, their desired position, region, salary, education, working experience. Dataset 1 is also divided into three tables with general information, and detailed information on education and work experience. The tables can be joined by <i>id_candidate</i> . N=6221439
Dataset sources description	The data was gathered from the Jobs in Russia portal
Time frame	2015-2023
Available formats	CSV (“UTF-8”, “[”])

Dataset size	800 mb (.zip); 3.6 gb (.csv)
Dataset topic	Education, Career
Mode of access	Open access

Dataset 1 (general information)

dataset1.cand.csv 6221439 entries

Column	Description	Missing values	Data type
id_candidate	Candidate's unique ID in the system	0	String
date_modify_inner_info	Date when the information was modified by candidate	6246 (0.100%)	Date, format: "YYYY-MM-DD"
position_name	Desired job/position title	4905 (0.079%)	String
salary	Salary requested (monthly)	0	Int
gender	Gender (0=female;1=male)	842334 (13.539%)	Int
birthyear	Year of birth	864655 (13.898%)	Int
region_code	Code for the region of residence in Russia. In most cases, it is a numeric value from 1 to 99, which represents the official code of the regions of Russia. If different regions are specified in the candidate's CV, it is a string value with digits separated by a comma.	3	Int/String

Dataset 1 (work experience)

dataset1.workexp.csv 10254450 entries

Column	Description	Missing values	Data type
id_candidate	Candidate's unique ID in the system	0	String
company_name	The name of the company	5929 (0.058%)	String

date_from	The date when the candidate started working at the job	3374 (0.033%)	Date, format: “YYYY-MM-DD”
date_to	The date when the candidate finished working at the job	624393 (6.089%)	Date, format: “YYYY-MM-DD”
job_title	Name of the job	189075 (1.844%)	String

Dataset 1 (education)

dataset1.edu.csv 9124525 entries

Column	Description	Missing values	Data type
id_candidate	Candidate’s unique ID in the system	0	String
legal_name	Legal name of the educational organization	1149092 (12.593%)	String
graduate_year	Year of graduation	118298 (1.296%)	Int
faculty	Faculty	5223708 (57.249%)	String
qualification	Qualification/degree	4979664 (54.575%)	String
speciality	Speciality/major	5688027 (62.338%)	String
course_name	Name of the additional profession or educational course	7824212 (85.749%)	String

Dataset 2

General information	
Dataset full name	Activity of unemployed and job seeking candidates in Russia
Dataset short name	Activity of unemployed
Dataset brief description	All records for the available period without detailed descriptions of work experience and education. N=7662089
Time frame	2015-2023
Available formats	CSV (“UTF-8”, “ ”)
Dataset size	156 mb (.zip); 710 mb (.cvs)
Dataset topic	Activity of unemployed
Mode of access	Open access

Dataset 2

7662089 entries

Column	Description	Missing values	Data type
id_candidate	Candidate's unique ID in the system	0	String
region_code	Code for the region of residence in Russia. In most cases, it is a numeric value from 1 to 99, which represents the official code of the regions of Russia. If different regions are specified in the candidate's CV, it is a string value with digits separated by a comma.	3	Int/String
birthyear	Year of birth	971842 (12.684%)	Int
gender	Gender (0=female;1=male)	926118 (12.087%)	Int
experience	Years of work experience	44888 (0.586%)	Int
education_type	Level of education (1=higher;2=unfinished higher;3=secondary school;4=vocational education)	2839214 (37.055%)	Int
busy_type	Type of the desired work/employment (1=temporary;2=joint;3=full-time;4=permanent;5=seasonal;6=internship;7=remote;8=part-time)	623580 (8.139%)	Int
salary	Maximum salary requested	0	Int
responses	Total number of responses by the candidate	6025200 (78.637%)	Int

date_creation	Latest date of creation of the CV	0	Date, format: "YYYY-MM-DD"
date_publish	Latest date of publication of the CV	3978 (0.052%)	Date, format: "YYYY-MM-DD"
date_modify_inner_info	Latest date when the information was modified by the candidate	0	Date, format: "YYYY-MM-DD"
is_generated	CV was created automatically or imported from other job seeking sources (0=no;1=yes)	0	Int
cv_count	Number of CVs created on the portal	0	Int
cluster	Cluster ID of candidates' activity (0=passive;1=active)	0	Int