«Работа в России»: обработанные и объединенные сведения о вакансиях, резюме, откликах и приглашениях портала trudvsem.ru

Лицензия: CC BY-SA

Доступ: Доступен всем пользователям

Дата обновления: 02.12.2021

Теги:

трудовая занятость, рынок труда

Описание

Архивные и актуальные сведения о резюме, вакансиях, откликах соискателей и приглашениях на собеседование, размещенные на портале Роструда «Работа России». Сведения публикуются в разделе «Открытые данные» в формате многоуровневых структур ХМL. В ходе подготовки этого датасета данные были объединены в совокупность связанных плоских таблиц и предобработаны.

Единица наблюдения в датасете (строка): резюме соискателя, вакансия от работодателя или действия с вакансией (отклики, приглашения). Для резюме и вакансий приводится несколько версий в случае, если соискатели или работодатели вносили в них значимые изменения.

Для резюме доступны атрибуты: пол, дата рождения, образование, желаемая должность и зарплата, опыт работы и регион заявителя, а также сведения о различных статусах при публикации и обновлении резюме и т.д. Часть атрибутов сгруппированы в отдельных таблицах по принципу one-to-many (эти атрибуты могут соответствовать нескольким резюме). Для вакансий доступны атрибуты: сведения о работодателе, должность, обязанности, предлагаемая зарплата, требуемое образование, опыт работы, а также сведения о датах и связанных с ними статусах при публикации и обновлении вакансии и т.д.

Датасет состоит из 13 плоских связанных таблиц и охватывает все резюме и вакансии, публикуемые на портале с 01.01.2018 по 30.10.2021. Всего набор содержит около 20 млн наблюдений по более чем 100 атрибутам.

Набор размещен в открытом режиме доступа в формате CSV (кодировка: «UTF-8», разделитель: «;»).

Характеристики датасета

Использование

Лицензия: <u>CC BY-SA</u>

Доступ: Доступен всем пользователям

Предоставление данных

Формат:

CSV

Характеристики набора

Временной период: 01.01.2018 — 30.10.2021

Количество наблюдений: 2000000

Обновления

Размещено: 02.12.2021 Обновлено: 02.12.2021

Частота обновления: Ежеквартально

Дата следующего обновления: не указано

Происхождение

Источники данных: Федеральная служба по труду и занятости

Ответственные за набор данных:,

Структура набора данных

Справочник сфер деятельности (industries)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
date_modify	Дата изменения сущности в наборе	0			date
deleted	Отметка удаления	0			boolean
date_creation	Дата создания сущности в наборе	0			date
date_creation_mistake	Ошибки в датах атрибута date_creation (ошибка unix time)	34			boolean
industry_name	Наименование сферы деятельности (на русском языке)	0			text
industry_code	Наименование сферы деятельности (на английском языке)	0			text
active	Отметка активности	0			boolean
date_modify_mistake	Ошибки в датах атрибута date_modify (ошибка unix time)	34			boolean

Опыт работы, указанный в резюме (workexp)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
date_mistake	Отметка об ошибке в датах	0			boolean
company_name		86614			text
date_last_updated	Дата первой публикации этой версии записи об опыте работы в открытых данных портала «Работа России»	0			date
date_to	Дата окончания работы	810747			date
achievements	Достижения. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля соискателем необязательно	10323883			cat
achievements_modified	Достижения. Атрибут добавлен при обработке данных. Аналогично атрибуту achievements. Дополнительно унифицированы некоторые значения: «Нет».	10323883			cat

date_from	Дата начала работы.	26749		date
demands	Должностные обязанности	216616		text
id_cv	Идентификатор резюме	0		text
job_title	Название должности	330214		cat

Приглашения на собеседование, отправленные работодателями (invitations)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
is_new	Признак нового отклика	0			boolean
response_type	Приглашение	0			date
date_last_updated	Дата первой публикации записи в открытых данных портала «Работа России»	0			date
activity_flag_candidate	Отметка кандидата. В личном кабинете соискателя и работодателя есть архив приглашений и откликов, соответственно, приглашения и отклики могут быть активными либо в архиве	0			boolean
date_creation_mistake	Ошибки в датах атрибута date_creation (ошибка unix time)	6984541			boolean
activity_flag_manager	Отметка менеджера. В личном кабинете соискателя и работодателя есть архив приглашений и откликов, соответственно, приглашения и отклики могут быть активными либо в архиве	0			boolean
date_modify_mistake	Ошибки в датах атрибута date_modify (ошибка unix time)	6984541			boolean
id_candidate	Идентификатор соискателя	415			text
date_creation	Дата создания резюме	0			date
date_modify	Дата изменения отклика	380224			date
id_hiring_organization	Идентификатор нанимающей организации	193			cat
id_cv	Идентификатор резюме	415			text
id_reply	Идентификатор отклика	6569568			cat
id_vacancy	Идентификатор вакансии	193			text
id_invitation	Идентификатор приглашения	0			text
region_code	Код региона	0			integer
	·				

Аналитические данные по гражданам на портале (stat_citizens)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
region_code_mistake	Ошибки в датах атрибута region_code (длина не 13 цифр)	0			boolean
region_code		0			text
region_name	Наименование региона (на русском)	0			text
medium_salary	Средняя заработная плата	0			numeric
cvs_count	Количество резюме	0			integer

Субъекты РФ и основная аналитическая информация о них (regions)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
attraction_region	Регионы, оказывающие поддержку при переезде	1			numeric
kindergarten_accessibil- ity	Доступность детских садов	3			numeric
region_code	Код региона	0			integer
region_name	Наименование региона	0			text
accomodation_accessibil- ity	Доступность жилья	0			numeric
economic_growth	Темп экономического роста	1			numeric
medium_salary_difference	Средняя заработная плата	1			numeric
price_level	Уровень цен	1			numeric
unemployment_level	Уровень безработицы	1			numeric

Сведения об образовании, указанные в резюме (edu)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
id_cv	Идентификатор резюме	4			text
grad_year_mistake	Ошибки в датах атрибута «graduate_year»: ранее 70 лет назад и позже 4 лет вперед от даты обработки атрибута	0			boolean

date_last_updated	Дата первой публикации этой версии записи о полученном образовании в открытых данных портала «Работа России»	0		date
faculty	Название факультета	2394932		cat
graduate_year	Год окончания обучения	54880		numeric
speciality	Специальность	2927646		cat
legal_name	Юридическое название учебного заведения	80362		text
qualification	Квалификация	2214748		cat

Аналитические данные по работодателям на портале (stat_companies)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
small_company	Компаний с численностью сотрудников от 51 до 100 человек	0			integer
large_company	Компаний с численностью сотрудников более 500 человек	0			integer
midle_company	Компаний с численностью сотрудников от 101 до 250 человек	0			integer
company_count	Компаний, разместивших вакансии (всего)	0			integer
big_company	Компаний с численностью сотрудников от 251 до 500 человек	0			integer
micro_company	Компаний с численностью сотрудников менее 50 человек	0			integer
region_code	Код региона	0			text
region_name	Наименование региона (на русском)	0			text

Справочник специальностей (professions)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
active	Отметка активности	0			date
deleted	Отметка удаления	0			boolean
date_creation_mistake	Ошибки в датах атрибута date_creation (ошибка unix time)	0			boolean

date_creation	Дата создания сущности в наборе	0		date
profession_code	Профессия по ОКПДТР	0		integer
profession_name	Наименование специальности (на русском)	0		text
category	Категория специальности по ОКПДТР	574		cat
date_modify	Дата изменения сущности в наборе	0		date
date_modify_mistake	Ошибки в датах атрибута date_modify (ошибка unix time)	0		boolean
etks	Тариф указан в соответствии со справочником ЕТКС	2600		cat

Дополнительное образование, указанное в резюме (addedu)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
course_name	Название курса	0			text
date_last_updated	Дата первой публикации этой версии записи о дополнительном образовании в открытых данных портала «Работа России»	0			date
id_cv	Идентификатор резюме	0			text
graduate_year	Год выпуска	81169			numeric
description	Описание обучения	1294911			cat
grad_year_mistake	Ошибки в датах атрибута «graduate_year»: ранее 70 лет назад и позже 4 лет вперед от даты обработки атрибута	0			boolean
legal_name	Юридическое название учебного заведения	0			text

Отклики соискателей на вакансии (responses)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
date_last_updated	Дата первой публикации этой версии вакансии в открытых данных портала	0			date
date_creation_mistake	Ошибки в датах атрибута date_creation (ошибка unix time)	0			boolean

activity_flag_candidate	Отметка кандидата. В личном кабинете соискателя и работодателя есть архив приглашений и откликов, соответственно, приглашения и отклики могут быть активными либо в архиве	0	boolean
date_modify_mistake	Ошибки в датах атрибута date_modify (ошибка unix time)	1820667	boolean
date_creation	Дата создания отклика	0	date
date_modify	Дата изменения отклика	61748	date
activity_flag_manager	Отметка менеджера. В личном кабинете соискателя и работодателя есть архив приглашений и откликов, соответственно, приглашения и отклики могут быть активными либо в архиве	0	boolean
id_candidate	Идентификатор соискателя	737	text
id_cv	Идентификатор резюме	737	text
id_hiring_organization	Идентификатор нанимающей организации	42	cat
is_new	Признак нового отклика	0	boolean
id_reply	Идентификатор отклика	1820556	cat
id_response	Идентификатор отклика	0	text
id_vacancy	Идентификатор вакансии	42	cat
region_code	Код региона	0	integer
response_type	Тип отклика	0	cat
	·	·	

Резюме по субъектам РФ (curricula_vitae)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
abilympics_status	Подтверждено участие в движении «Абилимпикс»	10580192			date
birthday_mistake	Отметка об ошибке в годе рождения	2437634			boolean
inactive	Статус инактивации	0			boolean
abilympics_participation	Участие в движении «Абилимпикс»	10573585			boolean
nark_certificate	Наличие свидетельства о независимой оценке квалификации	10520847			numeric

nark_inspection_status	Данные о наличии свидетельства о независимой оценке квалификации подтверждены	10580877	numeric
additional_skills	Личные навыки	5301615	text
business_trips	Готовность к командировкам	2751413	boolean
busy_type	Тип занятости	39	cat
country	Гражданство	1	cat
date_creation	Дата создания резюме	13347	date
date_inactivation	Дата, начиная с которой эта версия резюме отсутствует в наборе открытых данных портала	5630018	date
date_last_updated	Дата первой публикации этой версии резюме в открытых данных портала	0	date
date_modify_inner_info	Дата последнего изменения резюме на портале	13348	date
date_publish	Дата публикации резюме на портале	3	date
worldskills_is_internat- ional	Отметка о статусе международной компетенции WorldSkills International (WSI)	10556148	boolean
date_time_publish	Дата и время публикации резюме на портале	3	date
drive_licences	Уровень лицензии водительских прав, представлен в виде списка перечисленных категорий водительских прав	6802359	cat
driver_licence_b	Наличие водительских прав категории В	0	boolean
driver_licence_c	Наличие водительских прав категории С	0	boolean
driver_licence_e	Наличие водительских прав категории Е	0	boolean
education_type	Тип полученного образования соискателем	4700918	cat
experience	Опыт работы соискателя	301651	numeric
experience_mistake	Нереалистичное значение атрибута «Опыт работы соискателя»	0	boolean
id_candidate	Идентификатор кандидата	1	text
id_cv	Идентификатор резюме	0	text
id_user_inner_info	Внутренний идентификатор пользователя сервиса: модератора (сотрудника СЗН/ЦЗН)	31740	cat

industry_code	Сфера деятельности	68	cat
inner_info_deleted	Отметка удаления резюме	1	date
inner_info_fullness_rate	Процент заполненности резюме	946	numeric
inner_info_status	Статус резюме на портале	1	cat
inner_info_visibility	Видимость резюме	1	numeric
locality	Код КЛАДР	2259	numeric
other_info	Иная информация о соискателе	9366641	cat
other_info_modified	Иная информация о соискателе	9366641	cat
position_name	Желаемая должность	22110	cat
profession_code	Профессия по ОКПДТР	5912055	numeric
region_code	Код региона	1496	numeric
relocation	Готовность к переезду	2746774	boolean
retraining_capability	Готовность к переобучению	2288235	boolean
salary	Желаемая зарплата	1	numeric
schedule_type	Желаемый график работы. Представлен в виде списка перечисленных типов графиков работы: вахтовый метод; гибкий график; ненормированный рабочий день; неполный рабочий день; полный рабочий день; сменный график	16	cat
schedule_type_1	Желаемый график работы: вахтовый метод	0	boolean
schedule_type_2	Желаемый график работы: гибкий график	0	boolean
schedule_type_3	Желаемый график работы: ненормированный рабочий день	0	boolean
schedule_type_4	Желаемый график работы: неполный рабочий день	0	boolean
schedule_type_5	Желаемый график работы: полный рабочий день	0	boolean
schedule_type_6	Желаемый график работы: сменный график	0	boolean
skills	Профессиональные качества	6282880	text
time_publish	Время публикации резюме на портале	3	time
worldskills_internation- al_name	Наименование компетенции Worldskills на английском языке	10559620	cat
worldskills_russian_name	Наименование компетенции Worldskills на русском языке	10556148	cat
worldskills_skill_abbre- viation	Код навыка компетенции Worldskills	10556704	cat

worldskills_type	Статус участия	10546498		boolean
worldskills_inspection status	Статус проверки наличия конкретной квалификации Worldskills	792137		cat
worldskills_inspection status_code	Код статуса проверки наличия квалификации Worldskills	10129128		cat
add_certificates	Иные сертификаты, удостоверения, достижения, звание, чин	9699055		cat
add_certificates_modified	Иные сертификаты, удостоверения, достижения, звание, чин	9699055		cat
birthday	Год рождения соискателя	2774941		numeric
driver_licence_a	Наличие водительских прав категории А	0		boolean
driver_licence_d	Наличие водительских прав категории D	0		boolean
gender	Пол	2121367		cat

Организации работодателей, размещающих вакансии на портале (organizations)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
date_creation_mistake	Ошибки в датах атрибута date_creation (ошибка unix time)	914287			boolean
date_moderation_inner_i- nfo_mistake	Ошибки в датах атрибута date_modify_inner_info (ошибка unix time)	593007			boolean
inner_info_code_externa- l_system	Источник контакта	253945			date
business_size	Размер организации в классификации	42799			cat
hr_agency	Кадровое агентство	705660			boolean
inner_info_deleted	Отметка удаления организации с портала	0			boolean
inner_info_disable_join- _manager	Запрет на добавление менеджеров	26356			boolean
description	Описание организации	901306			cat
inn	Код ИНН	7955			numeric
inner_info_id_author	Идентификатор автора, создавшего аккаунт организации	472530			text
inner_info_manager_ids	Идентификатор менеджеров организации	514456			text

inner_info_registration- _status	Статус регистрации	0	cat
inner_info_status	Статус организации	325	cat
inner_info_disable_impo- rt_vacancy	Запрет импорта вакансий из ОБВ	0	boolean
inner_info_disable_join- _company	Запрет на присоединение компаний	26356	boolean
inner_info_id_user	Код пользователя	695408	cat
inner_info_id_small_icon	Идентификатор изображения	902566	cat
inner_info_moderation_c- omment	Комментарий модератора	913315	cat
inner_info_state_program	Идентификатор государственной программы	913044	cat
inner_info_id_logo	Идентификатор изображения	914214	cat
inner_info_external_id	Внешний идентификатор	914280	cat
kpp	Код КПП	284678	cat
legal_form_code	Код ОКОПФ	762767	numeric
legal_form_name	Наименование кода ОКОПФ	762767	cat
company_structure_hidder	Скрывать структуру компании	41030	boolean
date_modify_inner_info mistake	Ошибки в датах атрибута date_modify_inner_info (ошибка unix time)	914287	boolean
first_rate_company	Статус крупнейшей компании	0	boolean
inner_info_is_moderated	Статус модерации организации	0	boolean
inner_info_disable_impo- rt_info	Запрет обновления описания компании при импорте	0	boolean
legal_name	Юридическое наименование организации	0	text
date_creation	Дата создания сущности в наборе	21862	date
name	Наименование организации	0	text
ogrn	Код ОГРН	1437	numeric
region_code	Код региона	122	numeric
site	Сайт организации	832920	cat
source	Источник информации об организации	7209	cat
state_program	Идентификатор государственной программы	913044	date
id_organization	Идентификатор организации	0	text
date_change_inner_info mistake	Ошибки в датах атрибута date_change_inner_info (ошибка unix time)	593007	boolean

address_code	Код КЛАДР	122	numeric
date_change_inner_info	Дата изменения сущности в наборе	463890	date
address	Адрес работодателя (индекс, регион, область, населенный пункт, улица, дом, помещение и т.д.)	549003	cat
code_parent_company	Код родительской организации	911928	cat
time_change_inner_info	Время изменения сущности в наборе.	463890	cat
date_moderation_inner_i- nfo	Дата изменения сущности в наборе	394	date
date_time_change_inner info	Дата и время изменения сущности в наборе	463890	date
date_modify_inner_info	Дата последнего изменения вакансии	0	date
date_last_updated	Дата первой публикации этой версии записи в открытых данных портала	0	date

Вакансии (vacansies)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
date_time_posted	Дата и время создания вакансии. Доступно за 2015—2021 гг.	584441			
date_posted	Дата публикации вакансии. Доступно за 2021 г. Формат YYYY-MM-DD	0			
date_posted_mistake	Ошибки в датах атрибута date_posted (ошибка unix time). Атрибут добавлен при обработке данных. Принимает значения: 1— в случае неадекватного значения; О— ошибки нет	0			
date_time_change_inner info	Дата и время изменения вакансии. Доступно за 2015—2021 гг. Формат YYYY-MM-DD HH-MM-SS	0			
caring_workers	Социальная категория граждан «Работники, осуществляющие уход за больными членами своих семей в соответствии с медицинским заключением». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0			

	I			
disabled	Социальная категория граждан «Инвалиды». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
dms	Преимущества вакансии «ДМС». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
job_location_additional- _address_info	Дополнительные данные адреса работодателя. Заполнение поля необязательно	0		
vouchers_health_institu- tions	Преимущества вакансии «Путевки в оздоровительные учреждения». Принимает значения: 1 — да; 0 — нет	0		
requirements_required_c- ertificates	Требуемые сертификаты. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		
date_creation_mistake	Ошибки в датах атрибута date_creation (ошибка unix time). Атрибут добавлен при обработке данных. Принимает значения: 1—в случае неадекватного значения; 0—ошибки нет	0		
date_inactivation	Дата, начиная с которой эта версия вакансии отсутствует в наборе открытых данных портала. Атрибут добавлен при обработке данных. Формат: YYYY-MM-DD	0		
date_change_inner_info	Дата изменения вакансии. Устаревший к 01.12.2021 атрибут. Может дублировать date_modify_inner_info 2015—2021 гг. Формат YYYY-MM-DD	584441		
date_last_updated	Дата первой публикации этой версии записи в открытых данных портала. Атрибут добавлен при обработке данных. Формат: YYYY-MM-DD	0		
id_hiring_organization	Идентификатор нанимающей организации. Совпадает с атрибутом organization. Атрибут позволяет связать между собой таблицы: vacancies, organizations, responces, invitations	0		
identifier	Идентификатор вакансии. Атрибут позволяет связать между собой таблицы: vacancies, curricula_vitae, responces, invitations, organizations	0		

industry	Сфера деятельности вакансии в соответствии со справочником сфер деятельности. Атрибут позволяет связать между собой таблицы: vacancies, curricula_vitae	0		
organization	Код идентификатор организации. Совпадает с id_hiring_organization. Атрибут позволяет связать между собой таблицы: vacancies, organizations, responces, invitations	0		
profession	Профессия по Общероссийскому классификатору профессий рабочих, должностей служащих и тарифных разрядов (ОКПДТР). Расшифровка кодов содержится во вспомогательной таблице professons. Атрибут позволяет связать между собой таблицы: vacancies, curricula_vitae	0		
region	Код субъекта РФ. Расшифровка кодов содержится во вспомогательной таблице regions. Атрибут позволяет связать между собой таблицы: vacancies, curricula_vitae	0		
accommodation_capability	Жилищные условия. Принимает значения: 1 — жилье / предоставляется; 0 — жилье не предоставляется. Заполнение поля необязательно	0		
accommodation_housing	Тип предлагаемого жилья. Принимает значения: «DORMITORY»; «ROOM»; «FLAT»; «HOUSE». Заполнение поля необязательно	0		
additional_info	Дополнительная информация по требованиям к кандидату. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		
base_salary	Предлагаемая заработная плата. Заполняется текстом в формате: «от», «от до», «до». Заполнение поля необязательно	0		
base_salary_min	Минимальная предлагаемая заработная плата	0		
base_salary_max	Максимальная предлагаемая заработная плата	0		
career_perspective	Перспективы карьерного роста. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		

	Ошибки в датах атрибута date_change_inner_info (ошибка			
date_change_inner_info mistake	unix time). Атрибут добавлен при обработке данных. Принимает значения: 1 — в случае неадекватного значения; 0 — ошибки нет	0		
date_creation	Дата создания вакансии. 2015—2021 гг. Формат YYYY-MM-DD	0		
date_modify_inner_info	Дата последнего изменения вакансии. Доступно за 2021 г. Формат YYYY-MM-DD	0		
date_modify_inner_info mistake	Ошибки в датах атрибута date_modify_inner_info (ошибка unix time). Атрибут добавлен при обработке данных. Принимает значения: 1 — в случае неадекватного значения, 0 — ошибки нет	0		
drive_licences	Уровень лицензии водительских прав, представлен в виде списка перечисляемых категорий водительских прав. Далее представлен one-hot переменными: driver_licence_A; driver_licence_E; driver_licence_E	0		
driver_licence_A	Водительские права категории «А». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
driver_licence_B	Водительские права категории «В». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; О — нет	0		
driver_licence_C	Водительские права категории «С». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
driver_licence_D	Водительские права категории «D». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
driver_licence_E	Водительские права категории «Е». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
education_academic_degr	Требуемая академическая степень. Принимает значения: «Кандидат наук»; «Доктор наук» Заполнение поля необязательно	0		

	Требуемый уровень образования.			
education_requirements education_type	Принимает значения: «Высшее»; «Незаконченное высшее»; «Среднее профессиональное»; «Среднее». Заполнение поля необязательно	0		
education_requirements speciality	Требуемая образовательная специальность. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		
employment_type	Тип занятости. Принимает значения: «Полная занятость»; «Сезонная»; «Удаленная»; «Временная»; «Частичная занятость»; «Стажировка»	0		
experience_requirements	Требуемый опыт работы (количество лет). Количество лет. Например: 0, 1, 5. Заполнение поля необязательно	0		
federal_district	Номер федерального округа. Принимает значения от 1 до 9	0		
inactive	Статус инактивации. Атрибут добавлен при обработке данных. Принимает значения: 0 — запись в последней версии; 1 — запись в архиве или существует более новая версия записи	0		
incentive_compensation transport_compensation	Транспортная компенсация. Принимает значения: «PASSAGE_PAID»; «FUEL_PAID»; «AUTO». Заполнение поля необязательно	0		
inner_info_contact_source	Источник контакта вакансии. Принимает значения: «CZN» — разместил вакансию центр занятости населения; «СОМРАNУ» — разместила вакансию организация. Заполнение поля необязательно	0		
inner_info_deleted	Отметка удаления вакансии. Принимает значения: 0— вакансия не удалена; 1— вакансия удалена	0		
inner_info_metro_ids	Идентификатор станции метро, представлен в виде списка перечисляемых идентификаторов станций метрополитена. Сгенерирован в исходных данных. Пример: [e182b311-81cf-11e4-a6da-9be8de68 e183766c-81cf-11e4-a6da-9be8de68 Заполнение поля необязательно			

inner_info_source_type	Тип источника вакансии. Принимает значения: «Служба занятости»; «Работодатель»; «Аккредитованное частное агентство занятости»; «Кадровое агентство»; «Вакансия работодателя». Заполнение поля необязательно	0		
inner_info_status	Статус размещения вакансии. Принимает значения: «Одобрено»; «Не одобрено»; «Ожидает подтверждения». Заполнение поля необязательно	0		
inner_info_visibility	Видимость вакансии. Принимает значения: «Видно всем». Принимает значения: 1— «Видно всем»	0		
is_uzbekistan_recruitment	Отметка вакансии для граждан Узбекистана. Принимает значения: 1 — да; 0 — нет	0		
job_benefits	Преимущества вакансии. Атрибут принимает значения, передаваемые списком: «Оплата занятий спортом», «Путевки в оздоровительные учреждения», «Оплата питания», «ДМС». Далее представлен one-hot переменными: раумент_sports_activities: «Оплата занятий спортом», vouchers_health_institutions: «Путевки в оздоровительные учреждения», раумент_meals: «Оплата питания», dms: «ДМС». Заполнение поля необязательно	0		
job_benefits_other_bene- fits	Дополнительные преимущества. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		
job_location_address	Адрес работодателя (индекс, регион, область, населенный пункт, улица, дом, помещение и т.д.). Может содержать только часть адреса, например «352900, г. Армавир, д. 125» или «Набережная 1-го Мая»	0		
job_location_geo_longit- ude	Координаты долготы	0		
job_location_geo_latitude	Координаты широты	0		
large_families	Социальная категория граждан «Многодетные семьи». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		

metro_station	Станция метро, атрибут представлен в виде строки перечисляемых названий станций метрополитена. Заполнение поля необязательно	0		
minor_workers	Социальная категория граждан «Несовершеннолетние работники». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
need_medcard	Требование наличия медицинской книжки. Принимает значения: 1 — требуется; 0 — не требуется	0		
okso_code	Код по Общероссийскому классификатору специальностей по образованию: https://classifikators.ru/okso. Заполнение поля необязательно	0		
payment_meals	Преимущества вакансии «Оплата питания». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
payment_sports_activities	Преимущества вакансии «Оплата занятий спортом». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
premium_size	Размер премирования	0		
premium_type	Тип премии. Принимает значения: «Ежемесячная премия»; «Ежеквартальная премия»; «Ежегодная премия». Заполнение поля необязательно	0		
released_persons	Социальная категория граждан «Лица; освобождаемые из мест лишения свободы». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
requirements_id_priorit- y_category	Указатель на приоритетную категорию. Принимает значение: «Опытные кадры»	0		
requirements_qualificat- ions	Требуемая квалификация. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		
responsibilities	Должностные обязанности. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		

			_	_
retraining_capability	Наличие готовности к переобучению. Принимает значения: 1 — готов к переобучению; 0 — не готов к переобучению. Заполнение поля соискателем необязательно	0		
retraining_condition	Предоставляемое обучение. Текст в свободной форме. Удалены html-теги, дефектные пропуски значений и пр. Заполнение поля необязательно	0		
retraining_grant	Наличие стипендии. Принимает значения: 1— есть стипендия; 0— нет стипендии. Заполнение поля соискателем необязательно	0		
retraining_grant_value	Размер гранта на переобучение	0		
single_parent	Социальная категория граждан «Матери и отцы, воспитывающие без супруга (супруги) детей в возрасте до пяти лет». Атрибут добавлен при обработке данных. Принимает значения: 1 — да; 0 — нет	0		
social_protecteds_socia- l_protected	Категории социальной защиты. Арибут принимает значения, передаваемые списком (см. в колонке справа). Далее представлен one-hot переменными: disabled — «Инвалиды»; released_persons — «Лица, освобождаемые из мест лишения свободы»; single_parent — «Матери и отцы, воспитывающие без супруга (супруги) детей в возрасте до пяти лет»; large_families — «Многодетные семьи»; minor_workers — «Несовершеннолетние работники»; workers_with_disabled_children — «Работники, имеющие детей-инвалидов»; caring_workers — «Работники, осуществляющие уход за больными членами своих семей в соответствии с	0		
source	MERUHARE MARAREAM: ТЕБИЯТИРА ТРЕБИЯТИРА ТРЕ	0		
time_change_inner_info	Время изменения сущности в наборе. Формат HH-MM-SS	584441		

time_posted	Время публикации вакансии. Формат HH-MM-SS	0		
title	Наименование вакансии. Текст в свободной форме. Заполнение поля необязательно	0		
vac_url	URL-адрес вакансии на сайте trudvsem.ru	0		
work_hours	График работы. Принимает значения: «Сменный график»; «Полный рабочий день»; «Гибкий график»; «Неполный рабочий день»; «Ненормированный рабочий день»; «Вахтовый метод». Заполнение поля необязательно	0		
work_places	Количество рабочих мест	0		
workers_with_disabled_c- hildren	Социальная категория граждан «Работники, имеющие детей-инвалидов». Принимает значения: 1 — да; 0 — нет	0		

Внесенные в набор данных изменения по сравнению с источниками

- Текстовые поля со свободным вводом исходных данных имели в виде артефактов html-теги, которые при обработке были удалены. Дополнительно были унифицированы очевидно схожие значения текстовых полей, например: обозначение факта отсутствия какого-либо навыка, документа и пр. (варианты написания слова «нет», «отсутствует» и т.д.); обозначения факта наличия военного билета (варианты написания: «военный билет», «ВОЕННЫЙ билет» и т.д.) В некоторых случаях в исходном тексте отсутствуют пробелы между словами, пробелы были восстановлены. Пустые строки, содержащие только один символ либо последовательность знаков препинания, цифр и пробелов, были заменены на NULL.
- В исходных данных атрибутов, связанных с датами («birthday», «experience», «date_to», «date_from», «graduate_year»), встречаются нереалистичные значения. Были созданы атрибуты («birthday_mistake», «experience_mistake», «date_mistake», «grad_year_mistake»), указывающие на вероятную ошибку в значении соответствующего атрибута.
- В aтрибутах «date_change_inner_info», «date_posted», «date_publish», «date_change_inner_info», где указываются дата и время, было произведено разделение на переменные date_ и time_, оригинальные переменные оставлены без изменения с приставкой date time_в названиях.
- Записи из таблиц об откликах, приглашениях и организациях содержат дополнительную переменную «date_last_updated» дата первой публикации записи в открытых данных (ftp). Формат YYYY-MM-DD. Таким образом, можно отслеживать, когда сущность впервые появилась в наборе данных.
- Записи из таблиц о резюме и вакансиях имеют большой список переменных свободного ввода и содержат несколько дополнительных переменных «date_last_updated» дата первой публикации записи в открытых данных (ftp), «inactive» статус инактивациии, принимает значение «1», если запись в архиве или существует более новая версия записи, «date_inactivation» дата изменения переменной inactive. Таким образом, можно не только увидеть дату первого появления вакансии или резюме в наборе данных, но и все архивные версии, начиная с 2018 г.
- Исходный атрибут «social_protecteds_social_protected», представлявший категории социальной защиты в разных сочетаниях, передаваемые списком: «Инвалиды», «Лица, освобождаемые из мест лишения свободы», «Матери и отцы, воспитывающие без супруга (супруги) детей в возрасте до пяти лет», «Многодетные семьи», «Несовершеннолетние работники», «Работники, имеющие детей-инвалидов», «Работники, осуществляющие уход за больными членами своих семей в соответствии с медицинским заключением», был преобразован в отдельные атрибуты

(«disabled», «released_persons», «single_parent», «large_families», «minor_workers», «workers_with_dis

- по каждой категории со значением 1 в случае соответствия и 0 при отсутствии данной категории.
- Исходный атрибут «job_benefits», представлявший преимущества вакансии в разных сочетаниях, передаваемые списком: «Оплата занятий спортом», «Путевки в оздоровительные учреждения», «Оплата питания», «ДМС», был преобразован в отдельные атрибуты («payment_sports_activities», «vouchers_health_institutions», «payment_meals», «dms») по каждой категории со значением 1 в случае соответствия и 0 при отсутствии данной категории.
- Для атрибута «company_name» все случаи заполнения одним символом, знаками препинания, пробелами или цифрами заменены на пустую строку. Три нуля в начале названия «000» были заменены на «000».
- Атрибуты «course_name» и «legal_name». В случаях когда соискатель оставил поля незаполненными, значения были изменены на «Название курса не было указано» и «Название организации не было указано» соответственно.
- Были очищены излишне детализованные данные, которые пользователь оставил в не подходящих для этого полях. В таких случаях они были заменены на REDACTED.
- Были произведены и другие изменения.

Визуализация данных

Портал по поиску работы и сотрудников ИАС ОБВ «Работа в России»

Владелец источника данных	Федеральная служба по труду и занятости (Роструд)
Краткое описание источника данных	Портал «Работа России» — это федеральная государственная информационная система, оператором которой выступает Роструд. На портале размещаются вакансии от российских работодателей и резюме от российских граждан. Основное наполнение базы резюме портала происходит за счет регистрации граждан, которым оказывается государственная услуга по содействию занятости в центрах занятости населения (ЦЗН). Портал регулярно публикует обновляемые датасеты о вакансиях, резюме, откликах и приглашениях в разделе «Открытые данные». Данные публикуются в формате XML и могут содержать лишние символы, ошибки, описки и другие недостатки, что является следствием как ручного ввода сведений со стороны работодателей и соискателей, так и недостаточных процедур очистки и предобработки публикуемых открытых данных. Данные на портале «Работа России» доступны по API.
Ссылка на источник данных	https://trudvsem.ru/opendata/datasets
Перечень используемых классификаторов	Общероссийский классификатор профессий рабочих, должностей служащих и тарифных разрядов ОК 016-94
Методология и изменения методологии для источника данных	Детальный порядок сбора информации описан в Постановлении Правительства РФ от 25.08.2015 № 885 «Об информационно-аналитической системе Общероссийская база вакансий «Работа в России»
Ссылки на методологию	http://government.ru/docs/19388/