Малые и средние предприятия в сфере лёгкой промышленности за 2016–2023 гг.

Лицензия: CC BY-SA
Доступ: Доступен всем пользователям
Дата обновления: 13.03.2024
Теги:
лёгкая промышленность, малый и средний бизнес, микроданные

Описание

Набор данных содержит некоторую информацию о российских малых и средних организациях, ведущих деятельность в сфере лёгкой промышленности. Он сформирован на базе реестра субъектов малого и среднего предпринимательства и других открытых данных Федеральной налоговой службы России. К организациям в сфере лёгкой промышленности отнесены организации, у которых основной код ОКВЭД начинается с 13, 14 или 15. Набор содержит такие сведения, как ИНН и ОГРН, наименование (полное и сокращённое) юридического лица, его адрес регистрации (с точностью до населённого пункта), категорию в соответствии с классификацией малых и средних предприятий (микропредприятия, малые предприятия, средние предприятия), доходы, расходы и среднесписочная численность работников. Кроме того, он содержит географическую привязку: для каждой организации указаны географические координаты и код ОКТМО населённого пункта, в котором она зарегистрирована. Набор отражает хронологию изменения сведений за период с августа 2016 по ноябрь 2023 года. Данные о доходах, расходах и числе работников даются с детализацией по годам, остальные — с детализацией вплоть до месяцев. Данные о доходах, расходах и числе работников доступны не за все годы: обычно они имеются за период с 2019 по 2022 год.

Характеристики датасета

Использование

Лицензия: CC BY-SA

Доступ: Доступен всем пользователям

Предоставление данных

Формат:

CSV

Характеристики набора

Количество наблюдений: 146561

Обновления

Размещено: 13.03.2024

Обновлено: 13.03.2024

Частота обновления: не указано

Дата следующего обновления: не указано

Происхождение

Источники данных: Единый реестр субъектов малого и среднего предпринимательства

Структура набора данных

smb.csv (smb.csv)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
reg_number	Основной государственный регистрационный номер (ОГРН)	12739			string
tin	Идентификационный номер налогоплательщика (ИНН)	0			string
category	Категория субъекта малого и среднего предпринимательства: 1 — микропредприятие, 2 — малое предприятие, 3 — среднее предприятие	0			string
org_name	Полное наименование организации	0			string
org_short_name	Сокращённое наименование организации	93			string
activity_code_main	Основной код ОКВЭД (общероссийского классификатора видов деятельности)	0			string
region	Субъект федерации, в котором зарегистрирована организация	0			string
area	Часть территории субъекта (район, муниципальное образование), в которой зарегистрирована организация	30581			string
settlement	Населённый пункт, в котором зарегистрирована организация	9850			string
settlement_type	Сокращённый вид населённого пункта, в котором зарегистрирована организация. Сокращения в соответствии с ФИАС	567			string
oktmo	Код ОКТМО муниципального образования, в котором зарегистрирована организация	672			string
lat	Географическая широта населённого пункта, в котором зарегистрирована организация (десятичные в EPSG:4326)	567			Float
lon	Географическая долгота населённого пункта, в котором зарегистрирована организация (десятичные в EPSG:4326)	567			Float

address_raw	Адрес организации, содержащийся в исходных данных. Представляет собой структурированную строку вида «регион / тип региона / район / тип района / город / тип города / населённый пункт / тип населённого пункта», где косая черта — это разделитель, а элементы и структура адреса в основном соответствуют ФИАС	0			string
start_date	Начальная дата периода, в течение которого актуальны сведения	0			date
end_date	Конечная дата периода, в течение которого актуальны сведения.	0			date

revexp.csv (revexp.csv)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
tin	Идентификационный номер налогоплательщика (ИНН)	0			string
year	Год, для которого приводятся сведения	0			Integer
revenue	Доходы организации за год	0			Float
expenditure	Расходы организации за год	0			Float

empl.csv (empl.csv)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
tin	Идентификационный номер налогоплательщика (ИНН)	0			string
year	Год, для которого приводятся сведения	0			Integer
employees_count	Среднесписочная численность работников за год	0			Integer

panel.csv (panel.csv)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат

tin	Идентификационный номер налогоплательщика (ИНН)	0			String
reg_number	Основной государственный регистрационный номер (ОГРН)	39939			Srting
category	Категория субъекта малого и среднего предпринимательства: 1 — микропредприятие, 2 — малое предприятие, 3 — среднее предприятие	0			Enum
org_name	Полное наименование организации	0			String
org_short_name	Сокращённое наименование организации	350			String
activity_code_main	Основной код ОКВЭД (общероссийского классификатора видов деятельности)	0			String
region	Субъект федерации, в котором зарегистрирована организация	0			String
area	Часть территории субъекта (район, муниципальное образование), в которой зарегистрирована организация	124290			String
settlement	Населённый пункт, в котором зарегистрирована организация	37211			String
settlement_type	Сокращённый вид населённого пункта, в котором зарегистрирована организация. Сокращения в соответствии с ФИАС	1608			String
oktmo	Код ОКТМО муниципального образования, в котором зарегистрирована организация	1608			String
lat	Географическая широта населённого пункта, в котором зарегистрирована организация (EPSG:4326)	1608			Float
lon	Географическая долгота населённого пункта, в котором зарегистрирована организация (EPSG:4326)	1608			Float
address_raw	Адрес организации, содержащийся в исходных данных. Представляет собой структурированную строку вида «регион / тип региона / район / тип района / город / тип города / населённый пункт / тип населённого пункта», где косая черта — это разделитель, а элементы и структура адреса в основном соответствуют ФИАС	0			String

year	Год, для которого приводятся сведения	0			Integer
confidence	Условный показатель «надёжности» сведений. Определяется как единица, делённая на количество вариантов тех или иных сведений об организации в течение года. Например, если у организации в течение года поменялось наименование, в результате чего за год у неё есть два варианта сведений, то confidence = 0.5. Если confidence не равна единице (то есть вариантов сведений несколько), то таблица содержит последний вариант из доступных. Все варианты можно посмотреть в таблице smb.csv	0			Float
revenue	Доходы организации за год	0			Float
expenditure	Расходы организации за год	0			Float
employees_count	Среднесписочная численность работников за год	0			Integer

Постпроцессинг

Исходные наборы открытых данных были преобразованы в CSV-формат. Сведения об индивидуальных предпринимателях были исключены, чтобы избежать проблем с персональными данными. Дублирующиеся записи, возникшие из-за особенностей публикации исходных данных, были удалены. Адреса были нормализованы, к ним добавлены коды ОКТМО и географические координаты населённых пунктов в соответствии с наборами данных «Населенные пункты России: численность населения и географические координаты» (https://data.rcsi.science/data-catalog/datasets/160), «Города России» (https://github.com/hflabs/city), а также с ФИАС (https://fias.nalog.ru/).