Реестр субъектов малого и среднего предпринимательства за 2016–2024 гг. с доходами, расходами, численностью работников и географическими координатами

Лицензия: CC BY-SA
Доступ: Доступен всем пользователям
Дата обновления: 06.11.2024
Теги:
лёгкая промышленность, малый и средний бизнес, микроданные

Описание

Набор данных содержит информацию из реестра субъектов малого и среднего предпринимательства (МСП), дополненную сведениями о доходах, расходах, среднесписочной численности работников, а также географическими координатами населённых пунктов, в которых зарегистрированы субъекты МСП. Он сформирован на базе выгрузок реестра МСП и других открытых данных Федеральной налоговой службы России. 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 июнь 2024 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 2023 год.

Характеристики датасета

Использование

Лицензия: CC BY-SA
Доступ: Доступен всем пользователям

Предоставление данных

Формат:
CSV

Характеристики набора

Количество наблюдений: 63855163

Обновления

Размещено: 06.11.2024
Обновлено: 06.11.2024
Частота обновления: не указано
Дата следующего обновления: не указано

Происхождение

Источники данных: Единый реестр субъектов малого и среднего предпринимательства

Структура набора данных

panel.csv (panel.csv)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
tin	Идентификационный номер налогоплательщика (ИНН)	0			String
year	Год, для которого приводятся сведения. Может повторяться для каждого ИНН, если в течение года данные о субъекте МСП менялись.	0			Integer
reg_number	Основной государственный регистрационный номер (ОГРН)	14746093			Srting
kind	Вид субъекта МСП: 1 — юридическое лицо, 2 — индивидуальный предприниматель, 3 — самозанятый	0			Enum
category	Категория субъекта малого и среднего предпринимательства: 1 — микропредприятие, 2 — малое предприятие, 3 — среднее предприятие	0			Enum
org_name	Полное наименование организации	37496460			String
org_short_name	Сокращённое наименование организации	37592651			String
activity_code_main	Основной код ОКВЭД (общероссийского классификатора видов деятельности)	416287			String
region_iso_code	Код субъекта федерации, в котором зарегистрирована организация, по ISO 3166	605			String
region_code	Код субъекта федерации, в котором зарегистрирована организация, по классификации ФНС России	605			String
region	Название субъекта федерации, в котором зарегистрирована организация	605			String
area	Часть территории субъекта (район, муниципальное образование), в которой зарегистрирована организация	48280101			String

settlement	Населённый пункт, в котором зарегистрирована организация	724004			String
settlement_type	Сокращённый вид населённого пункта, в котором зарегистрирована организация. Сокращения в соответствии с ФИАС	992521			String
oktmo	Код ОКТМО муниципального образования, в котором зарегистрирована организация	1125880			String
lat	Географическая широта населённого пункта, в котором зарегистрирована организация (EPSG:4326)	992521			Float
lon	Географическая долгота населённого пункта, в котором зарегистрирована организация (EPSG:4326)	992521			Float
start_date	Начальная дата периода, в течение которого актуальны сведения. Период может охватывать несколько лет — в таком случае строки за каждый год дублируются. Формат даты — yyyy-mm-dd	0			String
end_date	Конечная дата периода, в течение которого актуальны сведения. Период может охватывать несколько лет — в таком случае строки за каждый год дублируются. Формат даты — yyyy-mm-dd	0			String
revenue	Доходы организации за год	51161158	Рубль	383	Float
expenditure	Расходы организации за год	51161158	Рубль	383	Float
employees_count	Среднесписочная численность работников за год	50407358	Человек	792	Integer

Постпроцессинг

Исходные наборы открытых данных были преобразованы в CSV-формат. 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/).