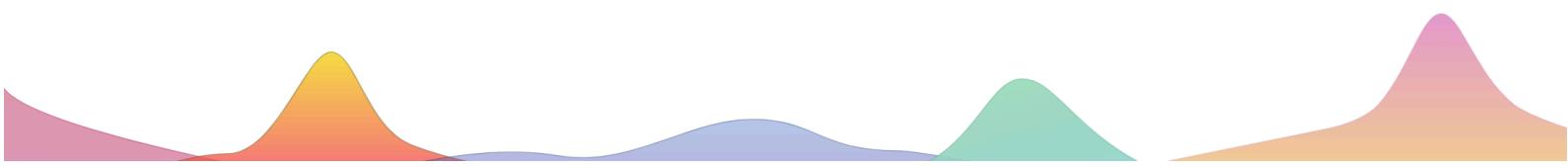




«Отчетность по налогу на доходы физических лиц по муниципалитетам и регионам с 2016 года»

0. О наборе данных

- В датасет вошли 38 показателей из отчетов Федеральной налоговой службы по формам 5-НДФЛ и 7-НДФЛ за 2016–2025 годы: количество поданных деклараций, число налогоплательщиков, налоговая база, удержанный налог и другие. Данные представлены на уровне регионов и отдельных муниципальных образований (3,6 млн наблюдений по форме 5-НДФЛ и 34,4 млн по форме 7-НДФЛ).
- Данные собраны из отчетов, которые публикуются на официальном сайте ФНС в разделе «Статистика и аналитика» → [«Данные по формам статистической налоговой отчетности»](#).
- Данные помогут исследовать пространственное неравенство доходов физических лиц и налоговых поступлений, а также динамику фискального потенциала регионов и муниципальных образований.
- Для некоторых регионов за отдельные годы отсутствовали отчеты на сайте ФНС, поэтому их нет в публикуемом наборе данных.
- Полные таблицы по формам 5-НДФЛ и 7-НДФЛ размещены в формате PARQUET. Данные в формате CSV или XLSX можно скачать по отдельным показателям или регионам [по ссылкам в таблице](#).



1. История изменений

Дата	Версия	Описание изменений
22.12.2025	1.0	Документ создан

2. Основные сведения

Общие сведения	
Наименование набора данных	Отчетность по налогу на доходы физических лиц по муниципалитетам и регионам с 2016 года
Краткое описание набора данных	<p>В наборе представлены данные по формам статистической налоговой отчетности 5-НДФЛ¹ и 7-НДФЛ², которые публикует Федеральная налоговая служба (ФНС). В текущей версии доступны 38 показателей за 2016–2025 годы.</p> <p>Единица наблюдения в наборе данных – муниципальное образование верхнего уровня, субъект РФ или страна. Данные по каждому наблюдению приведены по таким атрибутам, как регион, МО, ОКТМО, название показателя, разрезы для этого показателя, год и период наблюдения, значение показателя и пр. Подробнее в разделе 3. Структура набора данных.</p> <p>Для удобства набор разбит на таблицы по формам, регионам и показателям. Файлы по отдельным регионам, показателям, и сочетанию регион и показатель можно скачать по ссылкам в таблице.</p> <p>Набор доступен для работы в разных форматах: PARQUET (кодировка: «UTF-8», разделитель: «;»), XLSX и CSV (для таблиц в разрезе регионов и показателей). Набор данных охватывает 2016–2025 годы и содержит 38 миллионов наблюдений по 25 атрибутам.</p>
Краткое описание источника набора данных	Данные Федеральной налоговой службы по формам статистической налоговой отчетности 5-НДФЛ и 7-НДФЛ
Тематика набора данных	Разное
Единица наблюдения	Муниципальное образование верхнего уровня, регион, страна в целом
Доступные форматы	<ul style="list-style-type: none">PARQUET – полные таблицы

¹ Отчет о налоговой базе и структуре начислений по налогу на доходы физических лиц по сведениям, представленным налоговыми агентами.

² Отчет о налоговой базе и структуре начислений по расчету сумм налога на доходы физических лиц, исчисленных и удержаных налоговым агентом.

	<ul style="list-style-type: none"> • CSV (кодировка: «UTF-8», разделитель: «;») – таблицы по показателям и регионам • XLSX – таблицы по показателям и регионам
Периодичность публикации и обновления в каталоге «Если быть точным»	
Дата размещения в каталоге	22.12.2025
Покрываемый временной период	2016–2025
Дата последнего обновления набора данных	22.12.2025
Дополнительные сведения	
Цитирование набора данных на русском языке	«Отчетность по налогу на доходы физических лиц по муниципалитетам и регионам с 2016 года» // Федеральная налоговая служба; обработка: « Если быть точным », 2025. URL: https://tochno.st/ndfl
For references (English)	«Personal income tax reporting by Russian municipalities and regions since 2016» // Federal Tax Service (Russia); data-processing: « To Be Precise », 2025. URL: https://tochno.st/ndfl
Ссылка на открытый репозиторий	Открытого репозитория нет
Геоданные (пространственная привязка набора данных)	OKTMO регионов и муниципальных образований
Лицензия, под которой публикуется набор данных	<div style="display: flex; align-items: center;">  Creative Commons BY </div>



3. Структура набора данных

№	Атрибут	Описание	Число пропусков	Тип
1	indicator_name	Название показателя	0	String
2	indicator_unit	Единица измерения	0	String
3	indicator_code	Код показателя	0	String
4	individual_type	Категория физического лица. Пропуски заменены на код CD.	14 893 505	String
5	income	Код дохода. Пропуски заменены на код CD.	14 875 435	String
6	deduction	Код или группа кодов вычета. Пропуски заменены на код CD.	28 184 046	String
7	tax_rate	Налоговая ставка. Пропуски заменены на код CD.	33 078 567	String
8	object_name	Название муниципального образования, региона или страны в целом. Та единица наблюдения, к которой относятся показатели (может не соответствовать региону, предоставившему отчетность report_region_name). Пропуски заменены на код ND.	61 070	String
9	object_level	Уровень (Муниципальное образование верхнего уровня, Субъект РФ, Страна). Пропуски заменены на код ND.	61 070	String
10	object_oktmo	Код ОКТМО ³ (первые 8 цифр)	0	String
11	oktmo_stable	Последний действующий код ОКТМО для муниципального образования. Пропуски заменены на коды ND или CD	3 298 127	String
12	oktmo_history	Тип изменений, произошедших с муниципальным	3 298 127	String

³ Общероссийский классификатор территорий муниципальных образований / Russian Classification of Territories of Municipal Formations

№	Атрибут	Описание	Число пропусков	Тип
		образованием. Пропуски заменены на коды ND или CD		
13	oktmo_year_from	Год введения кода ОКТМО из колонки oktmo. Для всех значений до 2010 года указано значение 2010. Пропуски заменены на коды ND или CD	3 298 127	String
14	oktmo_year_to	Год завершения использования кода ОКТМО из колонки oktmo. В случае актуальности кода на сегодняшний день стоит 2025. Пропуски заменены на коды ND или CD	3 298 127	String
15	year	Год, за который приведено значение показателя	0	Integer
16	indicator_period	Период, за который приводится значение показателя (например, на начало года, в среднем за год и др.)	0	String
17	indicator_value	Значение показателя. Пропуски оставлены без изменений, однако их причины вынесена в отдельный атрибут	126 908	Float
18	null_value_reason	Причина пропуска в значении показателя (ND, CD, HD и NN – not null). Подробнее в разделе 4. Известные ограничения	0	String
19	report_region_id	ФНС код региона, который опубликовал отчетность	0	String
20	report_region_name	Название региона, который опубликовал отчетность	0	String
21	report_date	Дата формирования отчета.	288 418	Date

№	Атрибут	Описание	Число пропусков	Тип
		Пропуски заменены на код 1900-01-01		
22	report_type	Тип отчета («Свод» – сводный по всему региону в целом или «МО» – по муниципальным образованиям). В отчете по муниципальным образованиям публикуются сведения и по региону в целом, однако, для получения общей цифры по региону рекомендуется использовать данные сводного отчета.	0	String
23	tax_authority	Название налогового органа – источника информации. В некоторых регионах информацию публикуют несколько разных инспекций, из-за чего при группировке по регионам есть дубли.	0	String
24	index_form	Форма отчетности (5-НДФЛ или 7-НДФЛ)	0	String
25	source	Источник данных (Форма 2-НДФЛ с признаками «1», «2», «3» или «4», Приложение № 2 к налоговой декларации по налогу на прибыль организаций или Форма 6-НДФЛ)	0	String



4. Известные ограничения

Полнота данных

В текущую версию набора данных вошли следующие наблюдения:

1. За период с 2016 по 2025 годы.
2. По административно-территориальным единицам от муниципального образования верхнего уровня и выше.
3. По определенным муниципальным образованиям и сводам на уровне региона или страны (исключались отчеты, содержащие в шапке «МО: не определен»).

Например, на сайте ФНС отсутствовали отчеты по форме 5-НДФЛ для Калининградской, Нижегородской и Курской областей за 2024 год.

В обозначенных рамках данные представлены только за те годы, периоды, ОКТМО и другие разрезы, которые были доступны для скачивания с официального сайта ФНС на конец ноября 2025 года. То есть данные были размещены, ссылка на скачивание отчета открывалась, архив не был поврежден, содержимое отчета соответствовало заявленному на сайте и пр. Также пропускались отчеты с нестандартной структурой и в pdf файлах.

В наборе данных имеются следующие типы пропущенных значений:

1. **«Нет информации»** — значение отсутствует в источнике, либо является однозначно ошибочным: **ND (no data)** для строкового типа данных и **1900-01-01** для дат. Использовалось для атрибутов `object_name`, `object_level`, `oktmo_stable`, `oktmo_history`, `oktmo_year_from` и `oktmo_year_to` в случае отсутствия расшифровки для кода ОКТМО. А также для атрибута `null_value_reason`, когда значение показателя полностью отсутствовало или было однозначно ошибочным. Код 1900-01-01 проставлен у атрибута `report_date` в случаях, когда внутри отчета не было информации о дате его формирования.
2. **«Неприменимо»** — значение отсутствует, т.к. структура данных не предполагает его наличия в данной колонке в данной строке: **CD (conflict data)** для строкового типа данных. Использовалось в разрезах `individual_type`, `tax_rate`, `income`, `deduction`, когда наличие разреза не предполагалось самим показателем, в `oktmo_stable`, `oktmo_history`, `oktmo_year_from`, `oktmo_year_to` — в наблюдениях, относящихся к субъектам РФ и стране в целом, а также в `null_value_reason`, когда в источнике вместо значения показателя стоял знак «Х».
3. **«Значение скрыто»** — значение скрыто в источнике: **HD (hidden data)** для строкового типа данных. Использовалось для атрибута `null_value_reason`, в случаях, когда в источнике вместо значения показателя стоял знак «*».

5. Источники данных

Наименование источника данных	Данные по формам статистической налоговой отчетности
Владелец(ы) источника данных	Федеральная налоговая служба
Краткое описание источника данных	Данные Федеральной налоговой службы по формам статистической налоговой отчетности 5-НДФЛ и 7-НДФЛ, публикуемые на официальном сайте.
Ссылка на источник данных	<ul style="list-style-type: none">• https://www.nalog.gov.ru/rn77/related_activities/statistics_and_analytics/forms/