ОТЧЕТ О ПРОВЕДЕННОМ АНАЛИЗЕ ДАННЫХ

Тема работы:
«Анализ налоговой платежеспособности малого и среднего бизнеса Приморского края»

Выполнил:
Натаров Иван Петрович – консультант
отдела развития предпринимательства
министерства экономического развития
Приморского края

Владивосток
2021

СОДЕРЖАНИЕ

ПОСТАНОВКА ЦЕЛИ И ЗАДАЧ АНАЛИЗА

В настоящее время в Приморском крае одной из основных задач по развитию малого и среднего предпринимательства является создание благоприятных условий для ведения бизнеса в регионе, способствующих его ускоренному развитию.

Для реализации данной задачи на краевом уровне утверждены региональные проекты и дорожные карты, направленные на улучшение регионального бизнес климата, создана инфраструктура поддержки предпринимательства, действует льготное налоговое законодательство.

Данный инструментарий безусловно важен для региона, но не способен в полной мере обеспечить конкурентное преимущество региону, позволяющее совершить прорыв в улучшении региональной предпринимательской среды.

В эпоху развития четвертой промышленной революции важное значение в принятии взвешенных управленческих решений, будь то частный бизнес, государственный сектор, научные или образовательные организации, играют данные.

Сегодня в них можно найти много полезных инсайтов, изучить поведение пользователей, оценить качество продукта, а также эффективность принятия решений.

Целью данной работы является получение обобщающих показателей и выявление закономерностей налоговой платежеспособности малого и среднего бизнеса на территориях муниципальных образований Приморского края.

Для реализации данной цели были поставлены и решены следующие основные задачи:

1. Осуществлена загрузка и предобработка данных для последующего анализа;
2. Сформирован датасет для последующего анализа данных;
3. Проведен статистический анализ данных;
4. Визуализированы полученные результаты.

1. ЗАГРУЗКА И ПРЕДОБРАБОТКА ДАННЫХ

1.1. СБОР ДАННЫХ ДЛЯ АНАЛИЗА

В целях формирования датасета для проведения статистического анализа данных, с официального сайта Федеральной налоговой службы Российской Федерации (далее - ФНС) [1] была осуществлена загрузка следующих данных, характеризующих развитие малого и среднего предпринимательства Приморского края:

1. Единый реестр субъектов малого и среднего предпринимательства (файл, сформированный по состоянию на 10.08.2020). [2] Каждый август текущего года ФНС формирует реестр за предыдущий год.
2. Сведения о суммах доходов и расходов по данным бухгалтерской (финансовой) отчетности организации за год, предшествующий году размещения таких сведений на сайте ФНС России. [3]
3. Сведения о специальных налоговых режимах, применяемых налогоплательщиками. [4]
4. Сведения о суммах недоимки и задолженности по пеням и штрафам. [5]
5. Сведения об уплаченных организацией в календарном году, предшествующем году размещения указанных сведений в информационно-телекоммуникационной сети «Интернет» в соответствии с пунктом 1.1 статьи 102 Налогового кодекса Российской Федерации, суммах налогов и сборов (по каждому налогу и сбору) без учета сумм налогов (сборов), уплаченных в связи с ввозом товаров на таможенную территорию Евразийского экономического союза, сумм налогов, уплаченных налоговым агентом, о суммах страховых взносов. [6]

Все данные получены по состоянию на 31.12.2019. Отсутствие данных за 2020 год обусловлено особенностями налогового законодательства, такими как отчетные периоды, формирование статистической отчетности ФНС и т.д. Формирование данных за 2020 год происходит в 2021 году.

Всего было выгружено 54 481 файл:

- реестр – 6 211 файлов;
- сведения о доходах и расходах – 12 063 файла;
- сведения о специальных налоговых режимах – 12 067 файлов;
- сведения о суммах недоимки – 12 056 файлов;
- сведения об уплаченных налогах – 12 084 файла.

Все данные были выгружены в формате XML, что делает невозможным их последующий анализ и требует дальнейшей предобработки.

С помощью языка программирования Python были написаны скрипты по обработке полученных данных и последующее приведение их к формату CSV (Comma-Separated Values), удобного для обработки библиотеками по работе с данными Python (Pandas). [7]

В результате были получены следующие датасеты:

- reestr_msp_2019.csv
- company_revenue_2019.csv

- company_tax_regime_2019.csv
- msp_arrears_2019.csv
- msp_tax_2019.csv.

После (средствами Python [8]) данные были объединены в итоговый датасет (data_msp.csv) для последующего анализа.

1.2. ОПИСАНИЕ ПОЛУЧЕННЫХ ДАННЫХ.

Итоговый датасет (data_msp.csv) состоит из 38 366 строк и 53 признаков. Подробное описание признаков представлено в ПРИЛОЖЕНИИ А.

2. СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Техническая реализация статистического анализа была реализована средствами языка программирования Python. [9]

2.1. EDA (exploratory data analysis).

В целях изучения полученных данных был осуществлен исследовательский анализ, описаны типы данных, пропуски, дубликаты, меры центральной тенденции, а также взаимосвязи между переменными.

2.1.1 Описание данных.

В ходе исследовательского анализа установлено, что исходный набор данных содержит 38 366 записей, имеет 53 столбца (признака), из которых 4 столбца содержат числовые дискретные значения, 42 столбца содержат числовые непрерывные значения, 2 столбца содержат категориальные переменные и 5 столбцов содержат текстовые значения, содержит нулевые значения, пропуски, а также содержит не соответствующие типы данных для переменных.

2.1.2 Изменение типов данных.

В исходном наборе данных переменная ИНН приведена к строковому значению так как является не числовым значением, а идентификационным номером. Столбцы «Вид» и «Категория», которые являются категориальными признаками (т.е. признаками, которые принимают ограниченное и обычно фиксированное количество возможных значений) тип данных изменен на категориальный.

2.1.3 Описание пропусков в данных.

В ходе анализа данных на наличие пропусков установлено, что наибольшее количество пропусков (79.4%) присутствует в колонках «Акцизы н», «Водный налог н», «ЕНВД н», «ЕСХН н», «Задолженность и перерасчеты по отмененным налогам н», «Земельный налог н», «Неналоговые доходы, администрируемые налоговыми органами н», «НДС н», «НДПИ н», «НДФЛ н», «Налог на имущество организаций н», «Налог на прибыль н», «УСН н», «Сборы за пользование объектами животного мира и за пользование объектами», «ВБР н», Страховые взносы мед н», «Страховые взносы соц н», «Страховые и другие взносы пенс н», «Торговый сбор н», «Транспортный налог н» и почти половина данных (42.7%) пропущена в колонках «ЕСХН», «УСН», «ЕНВД», «СРП». По остальным колонкам пропуски составляют небольшую долю в данных.

Таким образом, в текущем датасете 6 колонок из 52 не имеют пропуски.

Так как проводить дальнейший анализ по данным, содержащим подавляющее число пропусков нецелесообразно (возможно получить неверные выводы), был сформирован датасет с максимально возможной информацией без пропусков для последующего анализа (48 колонок из 52 не имеют пропуски).

В ПРИЛОЖЕНИ Б представлены визуализации данных до и после удаления пропусков.

2.1.4 Описание дубликатов в данных.

В ходе данной работы был проведен анализ и описаны данные на предмет наличия в них дубликатов.

«Данные о наименованиях субъектов МСП».

Содержат дубликаты так как именование субъекта малого и среднего предпринимательства может быть одинаковым для разных субъектов (например, у индивидуальных предпринимателей может быть полное совпадение фамилии, имя и отчества). В рамках данного анализа информация не представляет особого интереса.

«Данные об идентификационных номерах субъектов МСП».

Не содержат дубликаты так как идентификационный номер — это уникальное значение, которое не может быть одно у двух и более юридических лиц или индивидуальных предпринимателей. По данной информации можно строить агрегаты, а также использовать в качестве уникального ключа, идентифицирующего конкретную запись о субъекте малого и среднего предпринимательства.

«Данные о муниципальных образованиях Приморского края, в которых зарегистрированы субъекты МСП».

Содержат дубликаты так как муниципальное образование — это территория, на которой может осуществлять свою деятельность более одного субъекта малого и среднего предпринимательства. По данной категориальной переменной можно строить агрегированную информацию в разрезе муниципальных образований.

«Данные о категориях субъектов МСП».

Содержат дубликаты так как к одной и той же категории может относится более одного субъекта малого и среднего предпринимательства. По данной категориальной переменной можно строить агрегированную информацию в разрезе категорий.

«Данные о кодах ОКВЭД, указанных как основной вид деятельности у субъектов МСП».

Содержат дубликаты так как субъекты малого и среднего предпринимательства могут осуществлять одинаковые виды деятельности. По данной категориальной переменной можно строить агрегированную информацию в разрезе ОКВЭД (в данном анализе это не представляет особого интереса).

«Данные о наименованиях видов деятельности».

Содержат дубликаты так как субъекты МСП могут осуществлять одинаковые виды деятельности. По данной категориальной переменной можно строить агрегированную информацию в разрезе ОКВЭД (в данном анализе это не представляет особого интереса).

2.1.5 Категориальные данные.

В рамках данной работы были построены распределения количества субъектов малого и среднего предпринимательства в разрезе муниципальных образований и категорий. Визуализация распределений представлена в ПРИЛОЖЕНИИ В.

Из полученных распределений видно, что в Приморском крае подавляющее число субъектов малого и среднего предпринимательства сосредоточено в городе Владивостоке, далее идут города Находка, Артем и Уссурийск. В остальных муниципалитетах сосредоточенно примерно одинаковое количество субъектов малого и среднего предпринимательства региона что вполне соответствует реальности.

Подавляющее число субъектов малого и среднего предпринимательства относится к категории микропредприятий, остальное количество субъектов малого и среднего предпринимательства приходится на малые предприятия и средний бизнес, доля которого крайне незначительна в общем количестве субъектов малого и среднего предпринимательства.

2.1.6 Числовые данные.

В ходе проведенного анализа:

- рассчитаны меры центральной тенденции (результаты расчета представлены в ПРИЛОЖЕНИИ Г), характеризующие распределение числовых переменных;
- для каждой числовой переменной (в целях наглядного отображения распределения данных) построены «боксплот» и частотный график (графики представлены в ПРИЛОЖЕНИИ Д);
- построена корреляционная матрица линейных зависимостей между числовыми переменными (представлена в ПРИЛОЖЕНИИ Е).

Получены следующие выводы:

- данные имеют большой размах;
- в данных присутствуют нулевые значения (естественные, не отсутствие значения);
- данные имеют выбросы (естественные);
- присутствуют отрицательные значения;
- в большинстве признаков (уплата (у) / неуплата (н)) до 75% перцентиля лежат нули;
- среднее значение сильно отличается от медианы;
- признак «Налог на игорный бизнес у» содержит только нулевые значения;
- в числовых данных наиболее часто встречаемое значение равно нулю, что указывает на высокий уровень налоговой ответственности у регионального бизнеса;
- наиболее часто встречаемый вид деятельности 41.20 «Строительство жилых и нежилых зданий»;
- наиболее часто встречаемая категория МСП «Микропредприятие»;

- наиболее часто встречаемый муниципалитет город Владивосток.

Построенные графики также указывают на большой размах данных, асимметрию, выбросы и гамма распределение.

На основании построенной корреляционной матрицы линейных зависимостей между числовыми переменными установлено:

наличие сильной прямой линейной зависимости между доходом и расходом у субъектов малого и среднего предпринимательства что вполне более чем логично. Как правило, чем выше доходы компании, тем крупнее компания и, тем больше нужно нести расходов, обеспечивающих компании прибыль и эффективное функционирование;

наличие средней прямой линейной зависимости между суммами не уплаченных налогов что более чем логично. Обычно, недобросовестные компании не уплачивают все типы налогов (за исключением компаний, где мог быть рассчитан не корректно какой-то конкретный налог и соответственно произошло недопоступление в бюджет), таким образом, чем выше неуплата по одному налогу, тем выше она может быть и по другим налогам (например, налог на прибыль и НДС);

наличие средней прямой линейной зависимости между суммами уплаченных налогов что более чем логично. Обычно, с ростом одного налога к уплате и растет сумма иных налогов (например, налог на прибыль и НДС);

наличие сильной прямой линейной зависимости между суммами не уплаченных страховых взносов что вполне логично, так как база для расчета единая для всех тарифов страховых взносов;

наличие сильной прямой линейной зависимости между суммами уплаченных страховых взносов что вполне логично, так как база для расчета единая для всех тарифов страховых взносов.

Для последующего анализа данные были объединены в следующие признаки и сформирован новый датасет (data_new_features):

«Налоги долг» - сумма всех неуплаченных налогов;

«Налоги уплата» - сумма всех уплаченных налогов;

«Страховые неуплата» - сумма всех неуплаченных страховых взносов;

«Страховые уплата» - сумма все уплаченных страховых взносов

«Иные платежи долг» - сумма неуплаты иных платежей;

«Иные платежи уплата» - сумма уплаты иных платежей.

2.2. ТЕСТИРОВАНИЕ ГИТОПЕЗ.

В ходе проведенного анализа были протестированы следующие гипотезы.

Гипотеза № 1.

Муниципальные образования отличаются по уровню налоговой платежеспособности у субъектов малого и среднего предпринимательства Приморского края.

Гипотеза № 2.

Существует группа муниципальных образований Приморского края, имеющая наиболее низкий уровень налоговой платежеспособности у субъектов малого и среднего предпринимательства Приморского края.

2.2.1 Тестирование гипотезы № 1.

Тестирование гипотезы №1 было осуществлено по следующему алгоритму:

- Сформулирована нулевая и альтернативная гипотезы.

H0 - Муниципальные образования Приморского края не отличаются по уровню налоговой платежеспособности субъектов малого и среднего предпринимательства Приморского края.

H1 - Муниципальные образования Приморского края отличаются по уровню налоговой платежеспособности субъектов малого и среднего предпринимательства Приморского края.

- Проведена проверка распределения на нормальность с помощью теста Андерсона-Дарлинга (работает с большими выборками).

По итогам проведенной проверки получено значение **P-value равное 0.0000000000**, что говорит о ненормальности распределения признака «Налоги долг». Учитывая данный факт для последующего сравнения выборок был использован непараметрический критерий Краскела-Уоллиса.

- Из признака «Налоги долг» были извлечены выборки и сформированы 34 группы (так как в Приморском крае 34 муниципальных образования) для последующего проведения теста Краскела-Уоллиса.

- С помощью теста Краскела-Уоллиса было проведено тестирование с целью установить различия между исследуемыми группами.

По итогам проведенного тестирования было получено **з**начение **P-value равное 0.0000000074**, что указывает на значительные различия между медианными значениями (данный статистический тест проверяет равенство медианных значений в выборках, что так же является плюсом так как средние значения сильно завышены за счет естественных выбросов) по уплате налогов бизнесом в муниципальных образованиях Приморского края. Вероятность совершения ошибки первого рода равна 0.0000000074.

Кроме того, учитывая факт наличия выбросов в тестируемых группах, полученное по итогам тестирование значение P-value могло получится искаженным и не отражать действительности.

В целях проверки указанного предположения было произведено частичное удаление выбросов (остались данные, лежащие в диапазоне + - 3 среднеквадратичных отклонения (правило 3-х сигм)). Таким образом общий объем данных сократился на 17 записей.

Также, после корректировки выбросов в данных было проведено сравнение основных статистик распределения.

Статистика	С выбросами	После удаления выбросов
Среднее	173 309.48	39 963.47
Медиана	18.0	17.5
Мода	0	0
СКО	4 141 243.95	364 253.81
Дисперсия	17 149 901 474 082.15	132 680 839 672.82

После удаления выбросов значительно изменилось среднее значение. Это обусловлено удалением из данных экстремально высоких значений касательно неуплаты налогов, соответственно снизился разброс данных (среднеквадратичное отклонение и дисперсия). Медиана почти не изменилась, мода осталась неизменной. Так как критерий Краскела-Уоллиса проверяет равенство медианных значений в выборках удаление выбросов сильно не сказалось на изменении мер центральной тенденции, на основании которых проводится расчет, поэтому было проведено тестирование (по аналогичному алгоритму) по данным, содержащим значительно меньшее количество выбросов.

По итогам проведенного тестирования было получено значение **P-value равное 0.0000000031**, что указывает на значительные различия между медианными значениями по уплате налогов бизнесом в муниципальных образованиях Приморского края. Вероятность совершения ошибки первого рода равна 0.0000000031.

Таким образом, проведя два теста по данным с выбросами и по данным, взятым в диапазоне +- 3 сигма установлено, что при частичном удалении выбросов вероятность получить статистически значимые различия в группах возрастает, но учитывая низкую вероятности совершения ошибки первого рода проводя тестирование на данных с выбросами, отчистка данных от выбросов в данном анализе не повлияла не результат.

Вывод: на основании проведенного теста можно сделать вывод, что муниципальные образования Приморского края отличаются по уровню налоговой платежеспособности субъектов малого и среднего предпринимательства.

2.2.2 Тестирование гипотезы № 2.

Тестирование гипотезы №2 было осуществлено по следующему алгоритму:

- Построено распределение медианой неуплаты налога субъектом малого и среднего предпринимательства Приморского края в разрезе муниципальных образований Приморского края.
- Из полученного распределения были извлечены выборки и сформированы 3 группы муниципалитетов по уровню неуплаты налога субъектом малого и среднего предпринимательства Приморского края для последующего проведения теста Краскела-Уоллиса:

high - медианная неуплата налога на одного предпринимателя больше 1500 рублей.

medium - медианная неуплата налога на одного предпринимателя больше 500 рублей и меньше или равна 1500 рублей.

low - медианная неуплата налога на одного предпринимателя меньше или равна 500 рублей.

• Визуализированы распределения в группах (ПРИЛОЖЕНИЕ Ж), на основании которых сделано предположение об отличии сравниваемых групп.

• Сформулирована нулевая и альтернативная гипотезы.

H0 - группы не отличаются по уровню медианной неуплаты налога на одного субъекта малого и среднего предпринимательства Приморского края.

H1 - группы отличаются по уровню медианной неуплаты налога на одного субъекта малого и среднего предпринимательства Приморского края.

• По итогам проведенного тестирования было получено значение **P-value равное 0.0000257215**, что указывает на значительные различия между медианными значениями групп, участвующих в тестировании. Вероятность совершения ошибки первого рода равна 0.0000257215.

• Для того, чтобы понять между какими группами существуют отличия (возможно не все группы отличаются между собой) был использован непараметрический апостериорный критерий Данна. Сравнение производилось попарно. Кроме того, в целях минимизации получить ошибку первого рода была применена поправка Бонферрони и Холмса.

Поправка Бонферрони				**Поправка Холмса**			
	high	**low**	**medium**		**high**	**low**	**medium**
high	1.000000	0.002300	1.000000	**high**	1.000000	0.001534	0.466479
low	0.002300	1.000000	0.000862	**low**	0.001534	1.000000	0.000862
medium	1.000000	0.000862	1.000000	**medium**	0.466479	0.000862	1.000000

По результатам проведенного теста Данна получены следующие результаты:

1. Скорректированное значение p-value для разницы между группами high и low составляет:

Поправка Бонферрони - 0.002300

Поправка Холмса - 0.001534

2. Скорректированное значение p-value для разницы между группами high и medium составляет:

Поправка Бонферрони - 1.000000

Поправка Холмса - 0.466479

3. Скорректированное значение p-value для разницы между группами low и medium составляет:

Поправка Бонферрони - 0.000862

Поправка Холмса - 0.000862

Вывод: статистически значимые различия при $\alpha = 0{,}05$ присутствуют между группами high и low и группами medium и low. Таким образом, группы high и medium имеют наиболее низкий уровень налоговой платежеспособности у субъектов малого и среднего предпринимательства Приморского края.

ИТОГОВЫЕ ВЫВОДЫ И РЕКОМЕНДАЦИИ

По результатам проведенного анализа удалось ответить на следующие вопросы:

Отличаются ли муниципальные образования по уровню налоговой платежеспособности у субъектов малого и среднего предпринимательства Приморского края?

В каких муниципальных образованиях наиболее низкий уровень налоговой платежеспособности у субъектов малого и среднего предпринимательства Приморского края?

В ходе проведенного анализа было установлено, что между муниципальными образованиями Приморского края существуют различия по уровню неуплаты налогов субъектами малого и среднего предпринимательства.

В целях установления групп муниципалитетов по уровню неуплаты налогов бизнесом муниципальные образования были объединены в группы по медианному уровню неуплаты налогов на одного предпринимателя.

В результате установлено:

- в Лазовском, Анучинском, Шкотовском, Пожарском, Кавалеровском, Партизанском, Дальнереченском, Спасском, Пограничном районах, а также в городе Спасск-Дальний наблюдается самая высокая недобросовестность бизнеса по уплате налоговых платежей. Данные муниципалитеты входят в группу high и medium. Медианная неуплата налога на одного предпринимателя в группе high составляет 3 686 рублей. Медианная неуплата налога на одного предпринимателя в группе medium составляет 907 рублей;
- в Ольгинском, Хорольском, Михайловском, Кировском, Хасанском, Чугуевском, Надеждинском, Красноармейском, Тернейском, Ханкайском, Октябрьском, Черниговском, Яковлевском, а также в городах Фокино, Лесозаводске, Большом Камне, Партизанске, Находке, Дальнереченске, Артеме, Уссурийске, Владивостоке, Арсеньеве, Дальнегорске, работает самый добросовестный бизнес. Медианная неуплата налога на одного предпринимателя в данной группе муниципалитетов составляет 23 рубля.

Таким образом, в муниципальных образованиях Приморского края входящих в группу high и medium осуществляет деятельность бизнес, либо имеющий низкий уровень знаний налогового законодательства, либо бизнес, целенаправленно не платящий налоги.

По итогам анализа рекомендовано:

- организациям инфраструктуры поддержки субъектов малого и среднего предпринимательства Приморского края, при формировании годового плана обучающих мероприятий бизнеса налоговой грамотности обратить внимание на муниципалитеты, входящие в группу high и medium.

• контрольно-надзорным органам, наделенными полномочиями в области налогового контроля, в целях повышения уровня поступления налогов в консолидированный бюджет Приморского края, усилить налоговый контроль на территориях муниципалитетов, входящих в группу high и medium.

Кроме того, в целях последующего мониторинга уровня налоговой платежеспособности субъектов малого и среднего предпринимательства на территориях муниципалитетах Приморского края, в программном продукте Power BI (ПРИЛОЖЕНИЕ З) был сформирован интерактивный отчет, содержащий следующие метрики:

1. Распределение недоимки на одного субъекта малого и среднего предпринимательства Приморского края в разрезе 34-х муниципальных образований Приморского края;
2. Сумма недоимки на одного субъекта малого и среднего предпринимательства Приморского края в категории high;
3. Сумма недоимки на одного субъекта малого и среднего предпринимательства Приморского края в категории medium;
4. Сумма недоимки на одного субъекта малого и среднего предпринимательства Приморского края в категории low;

В данном отчете предусмотрена возможность фильтрации данных по группам муниципалитетов.

Данные отчета содержат ретроспективную информацию и обновляются по мере выгрузки ФНС России новых наборов данных, отражающих развитие малого и среднего предпринимательства.

Вместе с тем, в целях последующего совершенствования текущего аналитического решения были предложены следующие пути улучшения:

1. Поиск дополнительных источников данных;
2. Восстановление пропущенных значений в данных;
3. Проверка дополнительных гипотез.

Более подробное описание указанных предложений представлено в ПРИЛОЖЕНИИ И.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Официальный сайт Федеральной налоговой службы Российской Федерации [Электронный ресурс] – URL: https://www.nalog.ru (дата обращения: 11.01.2021)
2. Единый реестр субъектов малого и среднего предпринимательства [Электронный ресурс] – URL: https://www.nalog.ru/opendata/7707329152-rsmp (дата обращения: 11.01.2021)
3. Сведения о суммах доходов и расходов по данным бухгалтерской (финансовой) отчетности организации за год, предшествующий году размещения таких сведений на сайте ФНС России [Электронный ресурс] – URL: https://www.nalog.ru/opendata/7707329152-revexp (дата обращения: 11.01.2021).
4. Сведения о специальных налоговых режимах, применяемых налогоплательщиками [Электронный ресурс] – URL: https://www.nalog.ru/opendata/7707329152-snr (дата обращения: 11.01.2021).
5. Сведения о суммах недоимки и задолженности по пеням и штрафам [Электронный ресурс] – URL: https://www.nalog.ru/opendata/7707329152-debtam (дата обращения: 11.01.2021).
6. Сведения об уплаченных организацией в календарном году, предшествующем году размещения указанных сведений в информационно-телекоммуникационной сети «Интернет» в соответствии с пунктом 1.1 статьи 102 Налогового кодекса Российской Федерации, суммах налогов и сборов (по каждому налогу и сбору) без учета сумм налогов (сборов), уплаченных в связи с ввозом товаров на таможенную территорию Евразийского экономического союза, сумм налогов, уплаченных налоговым агентом, о суммах страховых взносов [Электронный ресурс] – URL: https://www.nalog.ru/opendata/7707329152-paytax/ (дата обращения: 11.01.2021).
7. Код по обработке XML файлов.
8. Код по формированию итогового датасета.
9. Код по проведению статистического анализа датасета.

ПРИЛОЖЕНИЕ А

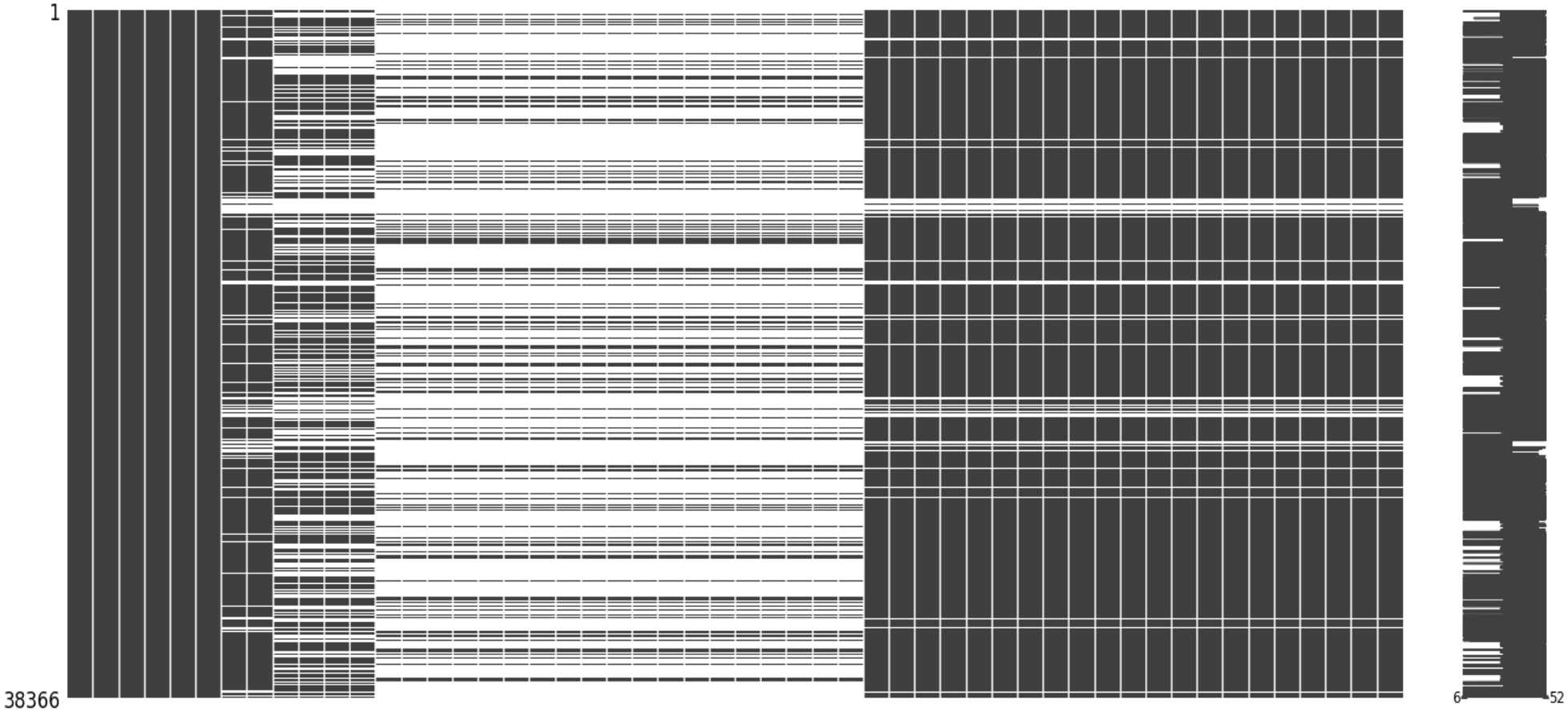
ОПИСАНИЕ ПРИЗНАКОВ ДАТАСЕТА (DATA_MSP.CSV)

№	Название признака	Описание признака
1	Наименование МСП	Содержит данные о полном наименовании субъекта МСП
2	ИНН	Содержит уникальный номер налогоплательщика. Для юридических лиц это номер из 10 чисел, а для индивидуальных предпринимателей номер из 12 чисел
3	Муниципальное образование	Содержит информацию о муниципальном образовании, к которому относится данный субъект МСП
4	Вид	Содержит информацию о виде субъекта МСП (юридическое лицо или индивидуальный предприниматель)
5	Категория	Содержит информацию о категории МСП: - микропредприятие; - малое предприятие; - среднее предприятие. Разделение на категории производится в соответствии со статьей 4 Федерального закона от 24.07.2007 № 209-ФЗ «О развитии малого и среднего предпринимательства в Российской Федерации»
6	Код ОКВЭД	Номер кода вида экономической деятельности в соответствии с Общероссийским классификатором видов экономической деятельности (утв. Приказом Росстандарта от 31.01.2014 № 14-ст)
7	Вид деятельности	Содержит информацию об основном виде экономической деятельности
8	Доход	Информация о полученном доходе субъектом МСП по данным бухгалтерской отчетности, предоставленной в ФНС
9	Расход	Информация о полученном расходе субъектом МСП по данным бухгалтерской отчетности, предоставленной в ФНС
10	ЕСХН	Признак применения единого сельскохозяйственного налога 0 - нет 1 - да
11	УСН	Признак применения упрощённой системы налогообложения 0 - нет 1 - да
12	ЕНВД	Признак применения единого налога на вмененный доход 0 - нет 1 - да

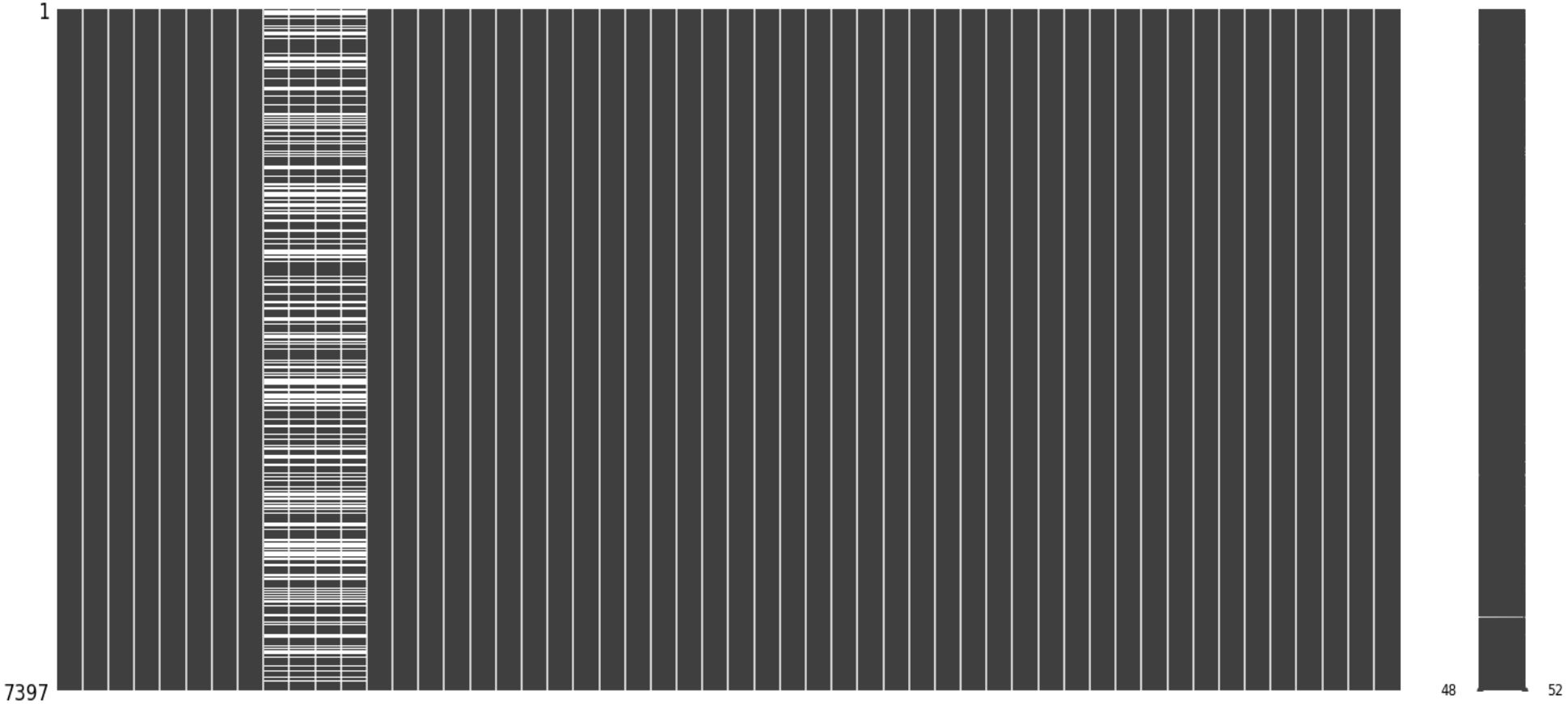
13	СРП	Признак заключения соглашения о разделе продукции 0 - нет 1 - да
14	Акцизы н	Сумма неуплаты акцизов (пени, штрафы, налог)
15	Водный налог н	Сумма неуплаты водного налога (пени, штрафы, налог)
16	ЕНВД н	Сумма неуплаты единого налога на вмененный доход (пени, штрафы, налог)
17	ЕСХН н	Сумма неуплаты единого сельскохозяйственного налога (пени, штрафы, налог)
18	Задолженность и перерасчеты по отмененным налогам н	Задолженность и перерасчеты по отмененным налогам
19	Земельный налог н	Сумма неуплаты земельного налога (пени, штрафы, налог)
20	Неналоговые доходы, администрируемые налоговыми органами н	Денежные взыскания (штрафы) за нарушение законодательства о налогах и сборах и др.
21	НДС н	Сумма неуплаты налога на добавленную стоимость (пени, штрафы, налог)
22	НДПИ н	Сумма неуплаты налога на добычу полезных ископаемых (пени, штрафы, налог)
23	НДФЛ н	Сумма неуплаты налога на доходы физических лиц (пени, штрафы, налог)
24	Налог на имущество организаций н	Сумма неуплаты налога на имущество организаций (пени, штрафы, налог)
25	Налог на прибыль н	Сумма неуплаты налога на прибыль (пени, штрафы, налог)
26	УСН н	Сумма неуплаты налога, уплачиваемого в связи с применением упрощенной системы налогообложения (пени, штрафы, налог)
27	Сборы за пользование объектами животного мира и за пользование объектами ВБР н	Сумма неуплаты сборов за пользование объектами животного мира и за пользование объектами водных биологических ресурсов (пени, штрафы, налог)
28	Страховые взносы мед н	Сумма неуплаты страховых взносов (медицина)
29	Страховые взносы соц н	Сумма неуплаты страховых взносов (социалка)
30	Страховые и другие взносы пенс н	Сумма неуплаты страховых взносов (пенсионка)
31	Торговый сбор н	Сумма неуплаты торгового сбора (пени, штрафы, налог)
32	Транспортный налог н	Сумма неуплаты транспортного налога (пени, штрафы, налог)
33	Акцизы у	Сумма уплаты акцизов (пени, штрафы, налог)
34	Водный налог у	Сумма уплаты водного налога (пени, штрафы, налог)
35	ЕНВД у	Сумма уплаты единого налога на вмененный доход (пени, штрафы, налог)
36	ЕСХН у	Сумма уплаты единого сельскохозяйственного налога (пени, штрафы, налог)

37	Задолженность и перерасчеты по отмененным налогам у	Сумма уплаты задолженности и перерасчетов по отмененным налогам
38	Земельный налог у	Сумма уплаты земельного налога (пени, штрафы, налог)
39	Неналоговые доходы, администрируемые налоговыми органами у	Сумма уплаты денежных взысканий (штрафов) за нарушение законодательства о налогах и сборах и др.
40	НДС у	Сумма уплаты налога на добавленную стоимость (пени, штрафы, налог)
41	НДПИ у	Сумма уплаты налога на добычу полезных ископаемых (пени, штрафы, налог)
42	НДФЛ у	Сумма уплаты налога на доходы физических лиц (пени, штрафы, налог)
43	Налог на игорный бизнес у	Сумма уплаты налога на имущество организаций (пени, штрафы, налог)
44	Налог на имущество организаций у	Сумма уплаты налога на прибыль (пени, штрафы, налог)
45	Налог на прибыль у	Сумма уплаты налога, уплачиваемого в связи с применением упрощенной системы налогообложения (пени, штрафы, налог)
46	УСН у	Сумма уплаты сборов за пользование объектами животного мира и за пользование объектами водных биологических ресурсов (пени, штрафы, налог)
47	Сборы за пользование объектами животного мира и за пользование объектами ВБР у	Сумма уплаты страховых взносов (медицина)
48	Страховые взносы мед у	Сумма уплаты страховых взносов (социалка)
49	Страховые взносы соц у	Сумма уплаты страховых взносов (пенсионка)
50	Страховые и другие взносы пенс у	Сумма уплаты торгового сбора (пени, штрафы, налог)
51	Торговый сбор у	Сумма уплаты транспортного налога (пени, штрафы, налог)
52	Транспортный налог у	Сумма уплаты акцизов (пени, штрафы, налог)
53	Утилизационный сбор у	Сумма уплаты утилизационного сбора

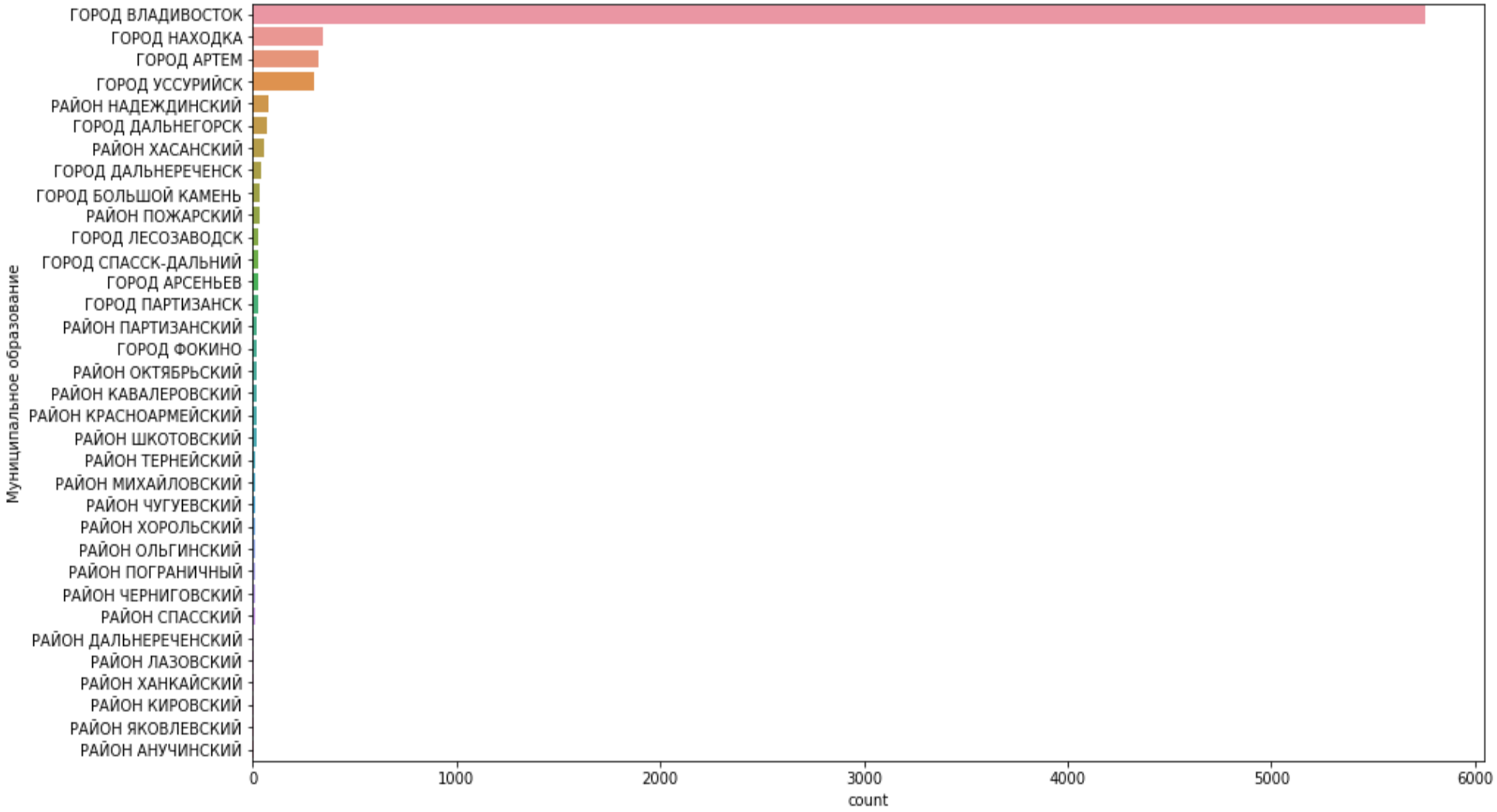
ВИЗУАЛИЗАЦИЯ ПРОПУСКОВ В ИСХОДНОМ ДАТАСЕТЕ



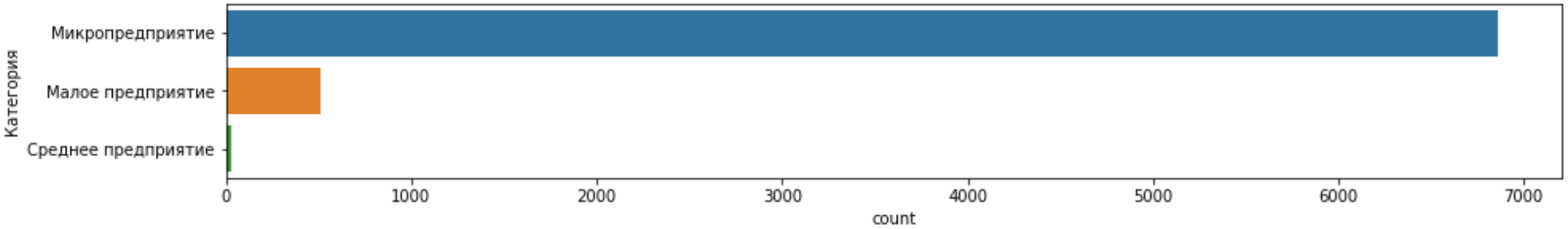
ВИЗУАЛИЗАЦИЯ ПРОПУСКОВ ПОСЛЕ УДАЛЕНИЯ



РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВА СУБЪЕКТОВ МСП ПО МУНИЦИПАЛЬНЫМ ОБРАЗОВАНИЯМ



РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВА СУБЪЕКТОВ МСП ПО КАТЕГОРИЯМ



ПРИЛОЖЕНИЕ Г

МЕРЫ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ

Наименование числовой переменной	count	mean	std	min	25%	50%	75%	max	mode
1. Доход	7397.00	43374076.03	1843882471.72	-173000.00	0.00	711000.00	8653000.00	158305175000.00	0.00
2. Расход	7397.00	42458360.09	1843900718.30	0.00	0.00	767000.00	7906000.00	158305175000.00	0.00
3. ЕСХН	4503.00	0.00	0.06	0.00	0.00	0.00	0.00	1.00	0.00
4. УСН	4503.00	0.95	0.23	0.00	1.00	1.00	1.00	1.00	1.00
5. ЕНВД	4503.00	0.13	0.34	0.00	0.00	0.00	0.00	1.00	0.00
6. СРП	4503.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7. Акцизы н	7397.00	2.72	164.85	0.00	0.00	0.00	0.00	13629.00	0.00
8. Водный налог н	7397.00	182.83	14609.82	0.00	0.00	0.00	0.00	1255666.22	0.00
9. ЕНВД н	7397.00	615.90	10508.81	0.00	0.00	0.00	0.00	431879.57	0.00
10. ЕСХН н	7397.00	29.74	2526.31	0.00	0.00	0.00	0.00	217271.33	0.00
11. Задолженность и перерасчеты по отмененным налогам н	7397.00	61.19	2088.25	0.00	0.00	0.00	0.00	157474.34	0.00
12. Земельный налог н	7397.00	4721.40	301574.71	0.00	0.00	0.00	0.00	25824655.82	0.00
13. Неналоговые доходы, администрируемые налоговыми органами н	7397.00	801.03	19891.68	0.00	0.00	0.00	0.00	1543050.00	0.00
14. НДС н	7397.00	102702.42	2372022.00	0.00	0.00	0.00	0.00	117242849.65	0.00
15. НДПИ н	7397.00	39.49	2957.19	0.00	0.00	0.00	0.00	252380.83	0.00
16. НДФЛ н	7397.00	17603.05	518681.96	0.00	0.00	0.00	0.00	39585371.85	0.00
17. Налог на имущество организаций н	7397.00	26588.82	1247140.17	0.00	0.00	0.00	0.00	80281759.62	0.00
18. Налог на прибыль н	7397.00	15042.21	521480.18	0.00	0.00	0.00	0.00	38338171.78	0.00
19. УСН н	7397.00	3662.49	44706.53	0.00	0.00	0.00	0.00	2191959.24	0.00
20. Сборы за пользование объектами животного мира и за пользование объектами ВБР н	7397.00	2.13	142.03	0.00	0.00	0.00	0.00	11999.53	0.00
21. Страховые взносы мед н	7397.00	7022.70	203966.81	0.00	0.00	0.00	0.60	15372111.02	0.00
22. Страховые взносы соц н	7397.00	2457.02	52521.74	0.00	0.00	0.00	0.41	2551248.71	0.00
23. Страховые и другие взносы пенс н	7397.00	29957.21	844876.04	0.00	0.00	0.00	6.51	61508654.23	0.00
24. Торговый сбор н	7397.00	12.37	1031.27	0.00	0.00	0.00	0.00	88656.40	0.00
25. Транспортный налог н	7397.00	2106.30	118731.67	0.00	0.00	0.00	0.00	9955895.29	0.00
26. Акцизы у	7397.00	5702.88	289392.53	0.00	0.00	0.00	0.00	22223936.00	0.00
27. Водный налог у	7397.00	61.19	2531.79	0.00	0.00	0.00	0.00	196115.00	0.00
28. ЕНВД у	7397.00	9040.63	73817.52	0.00	0.00	0.00	0.00	3225074.00	0.00
29. ЕСХН у	7397.00	81.65	5267.52	0.00	0.00	0.00	0.00	448552.00	0.00
30. Задолженность и перерасчеты по отмененным налогам у	7397.00	0.38	16.55	0.00	0.00	0.00	0.00	872.56	0.00
31. Земельный налог у	7397.00	16863.84	143667.65	0.00	0.00	0.00	0.00	4051650.00	0.00
32. Неналоговые доходы, администрируемые налоговыми органами у	7397.00	118.00	5421.83	0.00	0.00	0.00	0.00	394728.00	0.00
33. НДС у	7397.00	273657.71	2300684.66	0.00	0.00	0.00	0.00	100595330.00	0.00
34. НДПИ у	7397.00	2784.71	116179.49	0.00	0.00	0.00	0.00	9010000.00	0.00

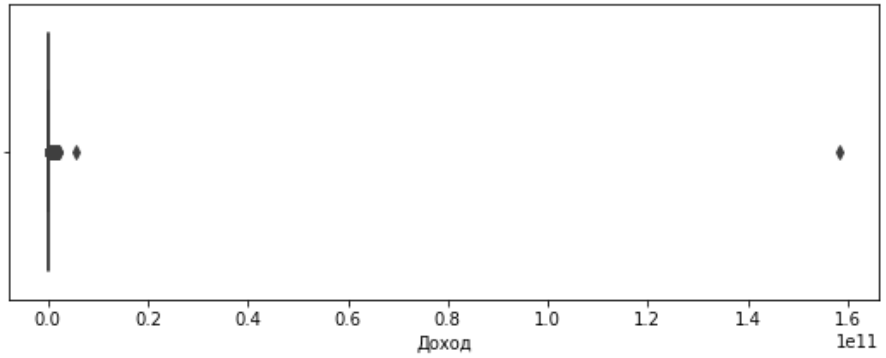
Наименование числовой переменной	count	mean	std	min	25%	50%	75%	max	mode
35. НДФЛ у	7397.00	267.42	5203.04	0.00	0.00	0.00	0.00	326056.00	0.00
36. Налог на игорный бизнес у	7397.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
37. Налог на имущество организаций у	7397.00	16092.43	209374.10	0.00	0.00	0.00	0.00	11005118.00	0.00
38. Налог на прибыль у	7397.00	105148.86	1151649.12	0.00	0.00	0.00	0.00	55659310.00	0.00
39. УСН у	7397.00	92981.14	381700.79	0.00	0.00	0.00	28535.00	9823589.88	0.00
40. Сборы за пользование объектами животного мира и за пользование объектами ВБР у	7397.00	1314.68	77312.87	0.00	0.00	0.00	0.00	6516869.00	0.00
41. Страховые взносы мед у	7397.00	52468.28	249257.35	0.00	0.00	4541.70	26968.81	7465047.00	0.00
42. Страховые взносы соц у	7397.00	28470.63	125833.68	0.00	0.00	2610.00	15307.65	3081898.21	0.00
43. Страховые и другие взносы пенс у	7397.00	230091.76	1041065.91	0.00	0.00	20268.86	118428.00	26984659.00	0.00
44. Торговый сбор у	7397.00	47.65	2929.90	0.00	0.00	0.00	0.00	202500.00	0.00
45. Транспортный налог у	7397.00	3209.13	28869.77	0.00	0.00	0.00	0.00	983910.00	0.00
46. Утилизационный сбор у	7397.00	2285.39	196556.48	0.00	0.00	0.00	0.00	16905000.00	0.00

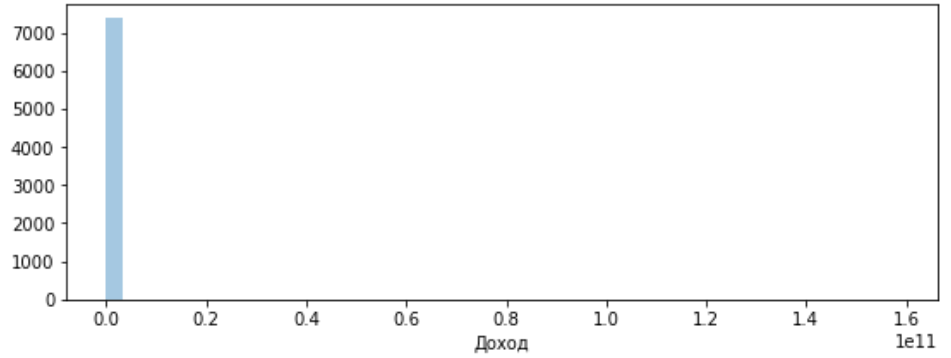
Наименование качественной переменной	mode
1. Наименование МСП	Общество с ограниченной ответственностью «Восток»
2. ИНН	1831187536
3. Муниципальное образование	Город Владивосток
4. Категория	Микропредприятие
5. Код ОКВЭД	41.20
6. Вид деятельности	Строительство жилых и нежилых зданий

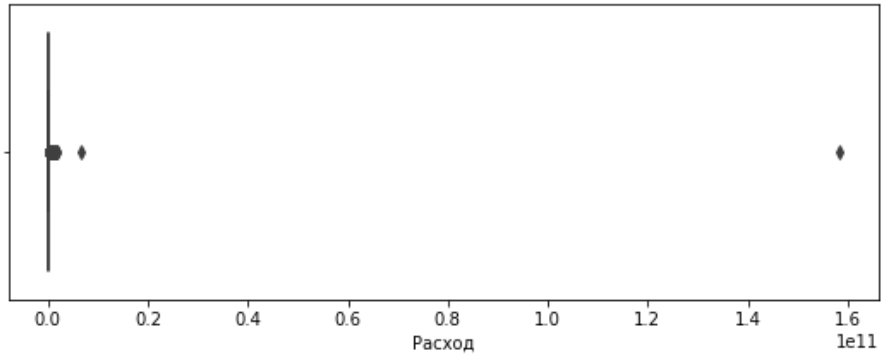
Условные обозначения:

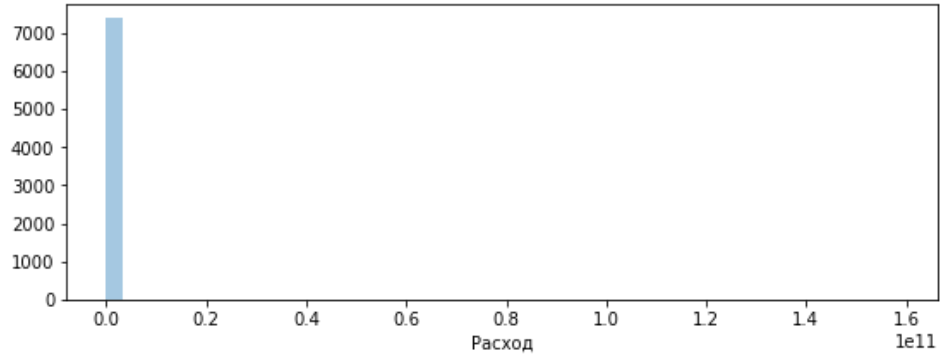
count - количество записей
mean - среднеарифметическое
std - среднеквадратичное отклонение
min - минимальное значение
25% - первый квартиль
50% - второй квартиль / медиана
75% - третий квартиль
max - максимальное значение
mode - мода

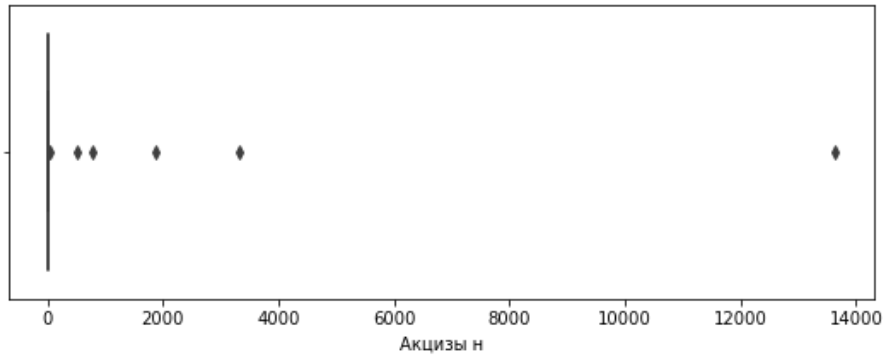
РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВЕННЫХ ПЕРЕМЕННЫХ

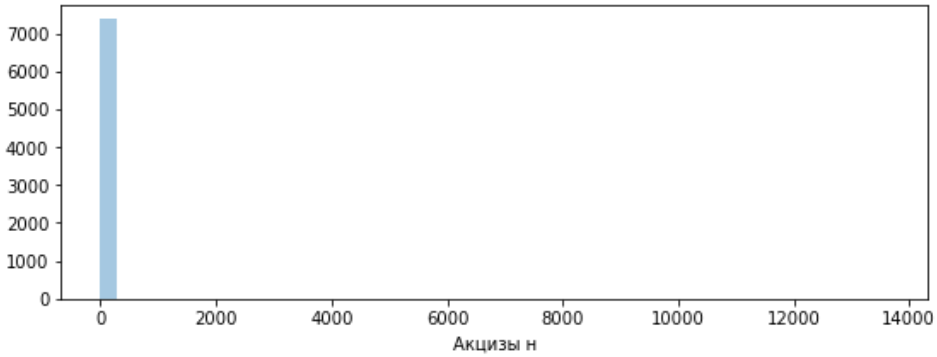


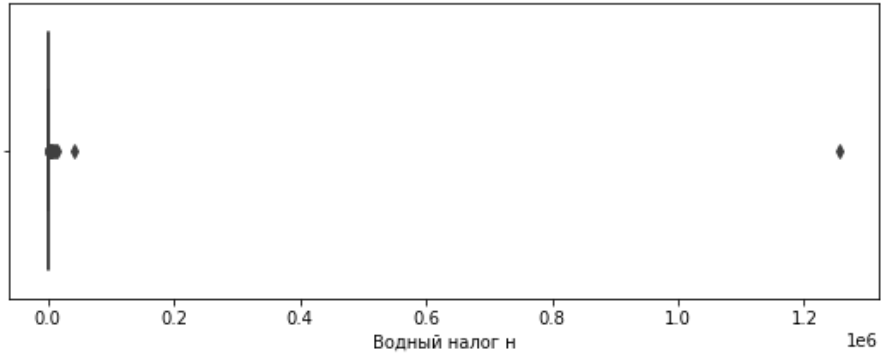


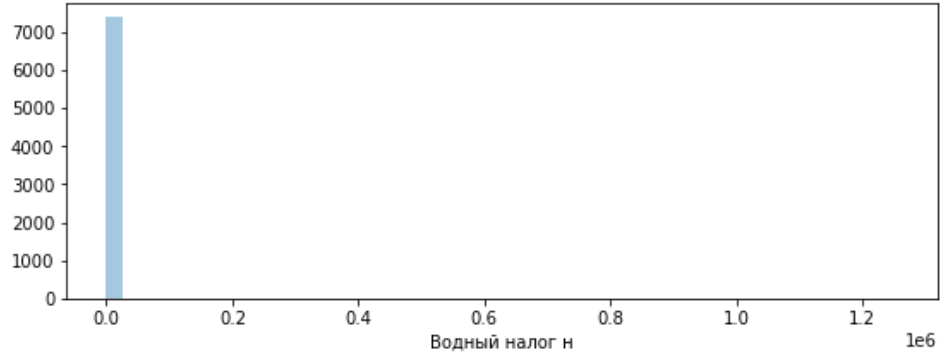


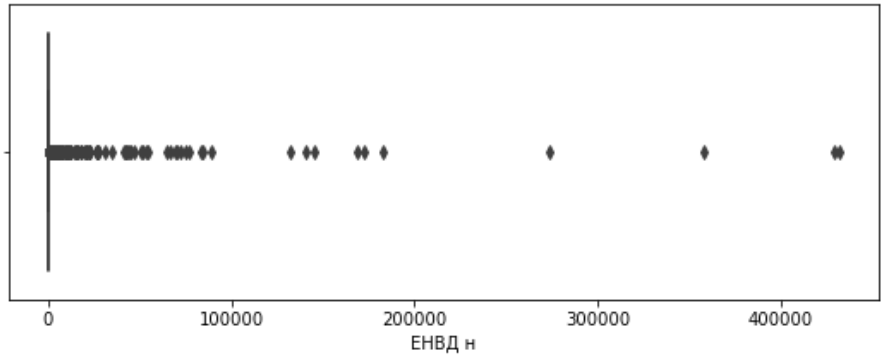


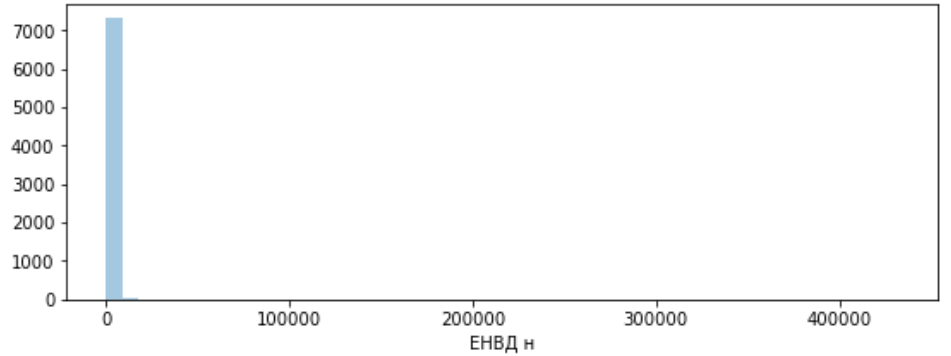


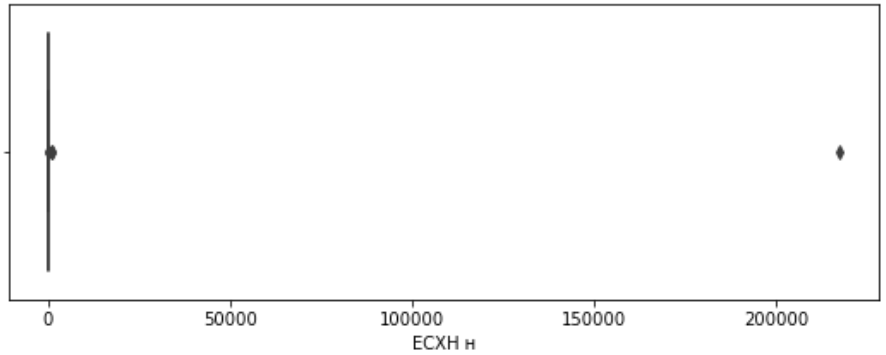


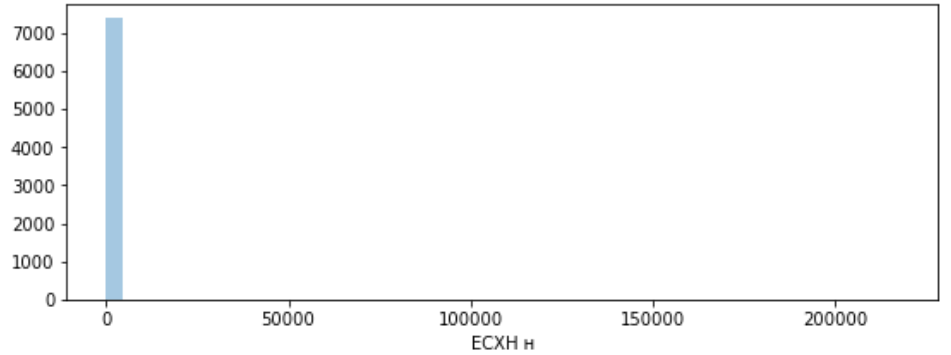

0.0
0.2
0.4
0.6
0.8
1.0
1.2
Водный налог н
1e6

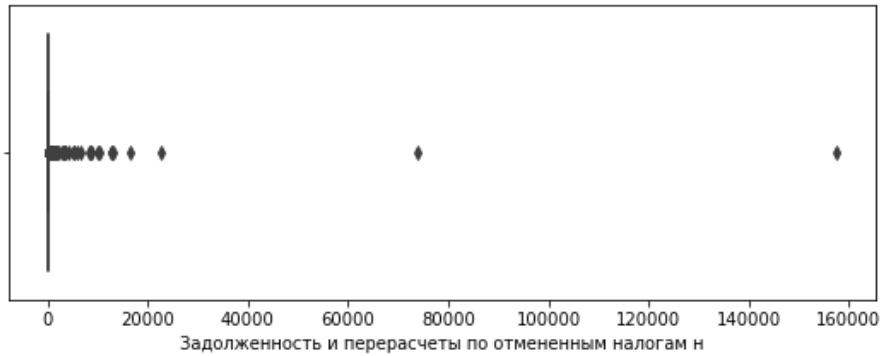

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Водный налог н
1e6

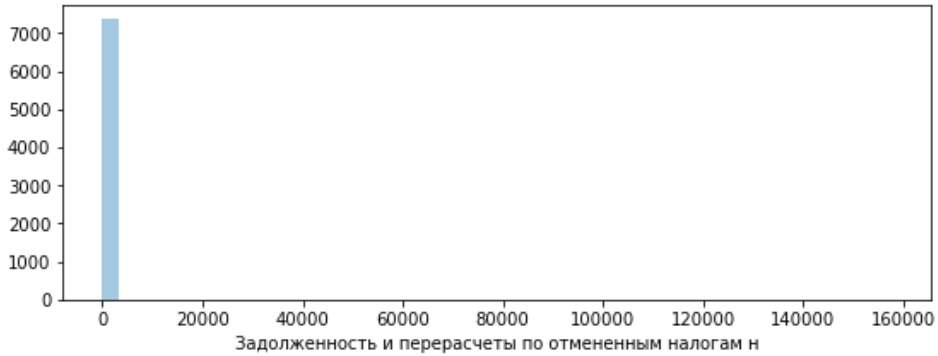

0
100000
200000
300000
400000
ЕНВД н

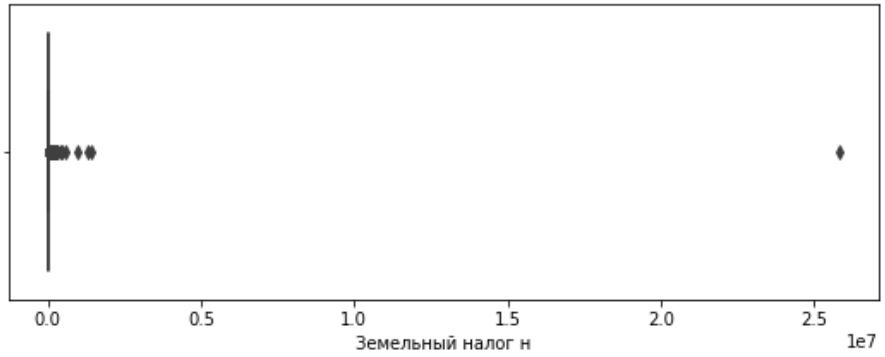

7000
6000
5000
4000
3000
2000
1000
0
0
100000
200000
300000
400000
ЕНВД н

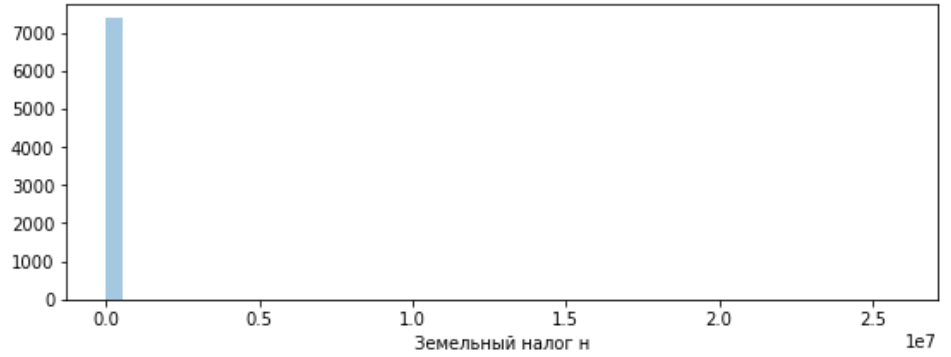

0
50000
100000
150000
200000
ЕСХН н

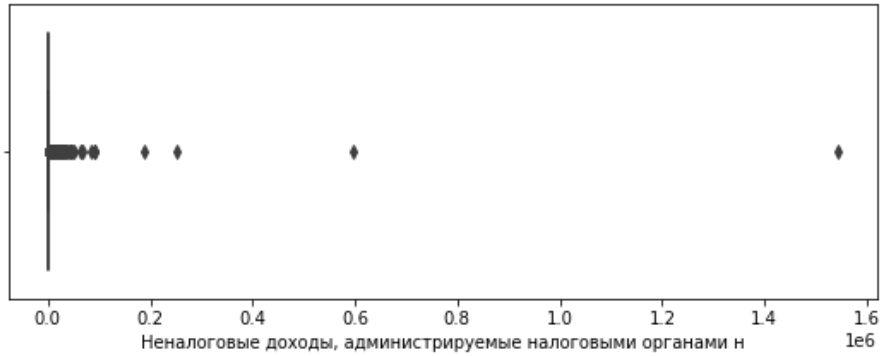

7000
6000
5000
4000
3000
2000
1000
0
0
50000
100000
150000
200000
ЕСХН н

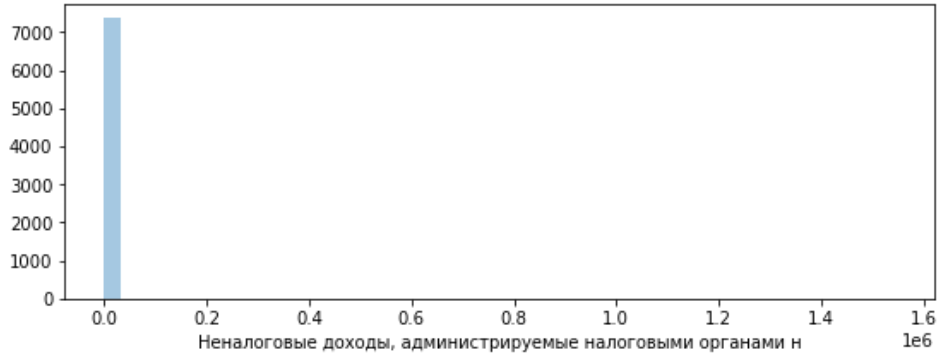

0
20000
40000
60000
80000
100000
120000
140000
160000
Задолженность и перерасчеты по отмененным налогам н

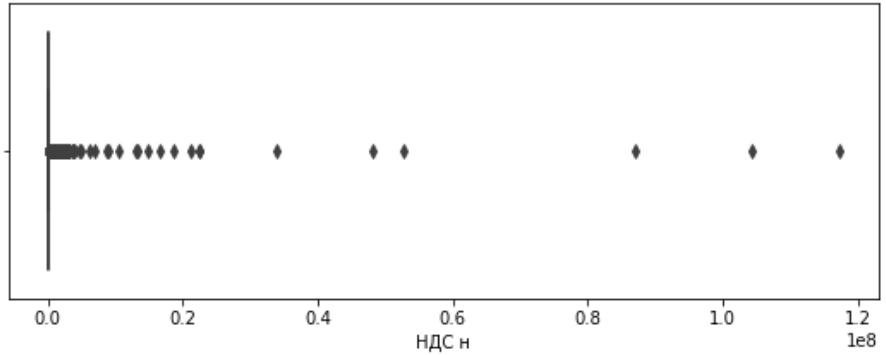

7000
6000
5000
4000
3000
2000
1000
0
0
20000
40000
60000
80000
100000
120000
140000
160000
Задолженность и перерасчеты по отмененным налогам н

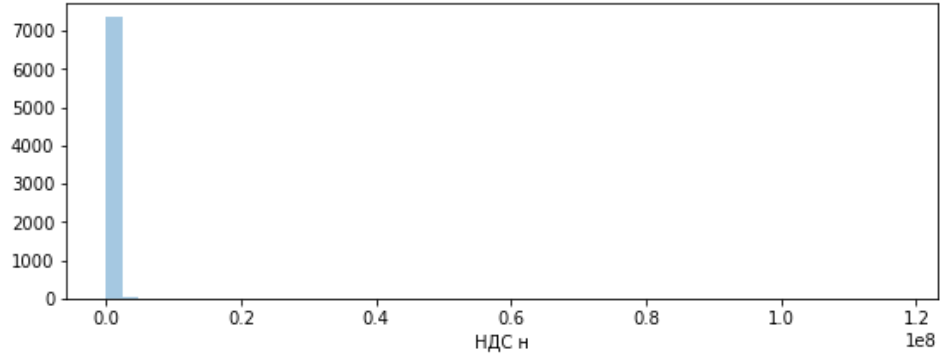

0.0
0.5
1.0
1.5
2.0
2.5
Земельный налог н
1e7

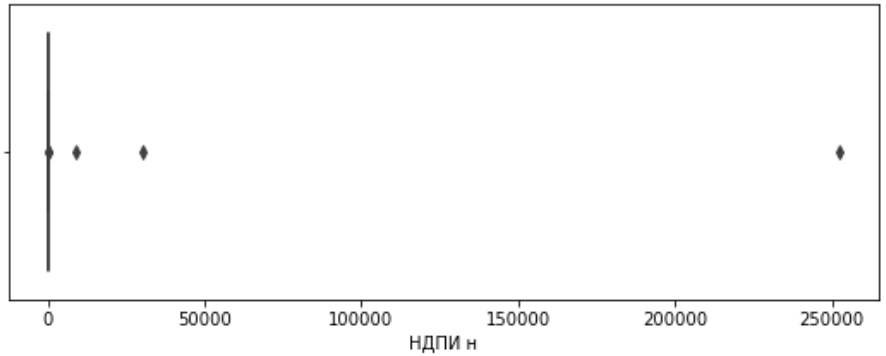

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
Земельный налог н
1e7

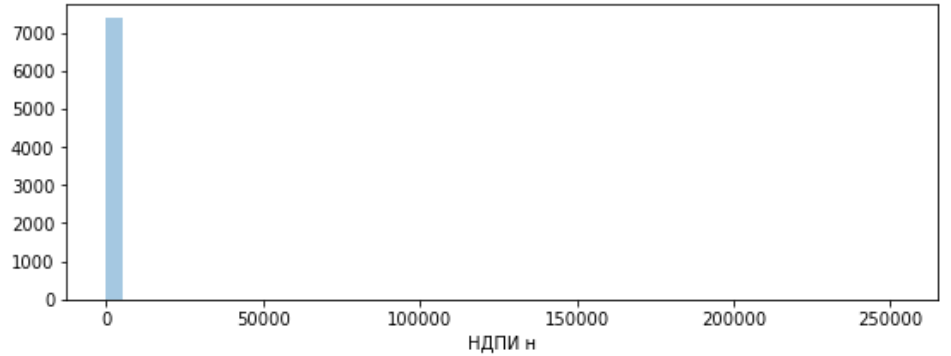

0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Неналоговые доходы, администрируемые налоговыми органами н
1e6

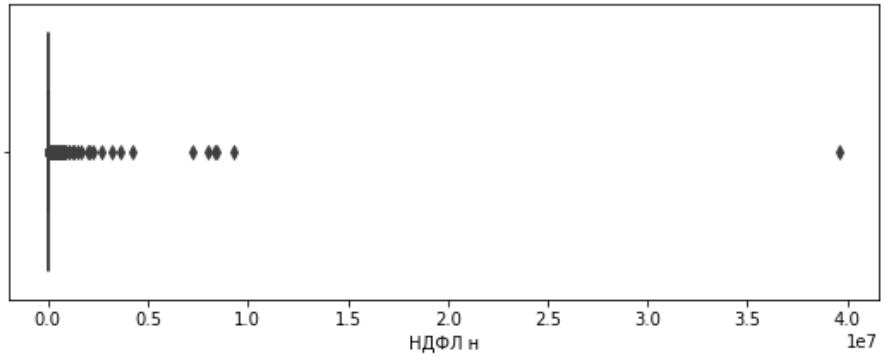

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Неналоговые доходы, администрируемые налоговыми органами н
1e6

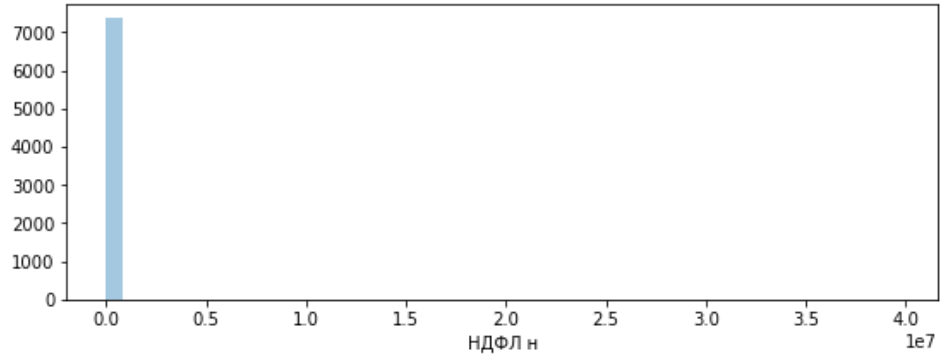

0.0
0.2
0.4
0.6
0.8
1.0
1.2
НДС н
1e8

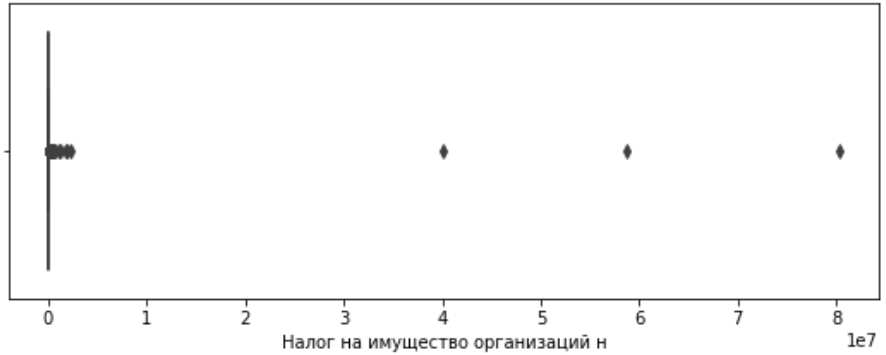

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
НДС н
1e8

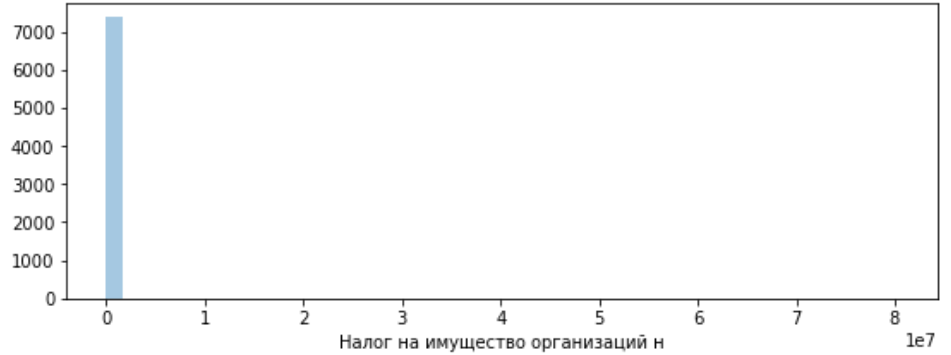

0
50000
100000
150000
200000
250000
НДПИ н

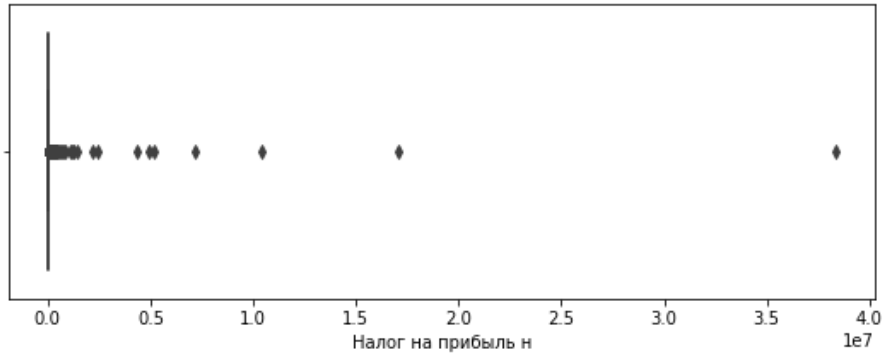

7000
6000
5000
4000
3000
2000
1000
0
0
50000
100000
150000
200000
250000
НДПИ н

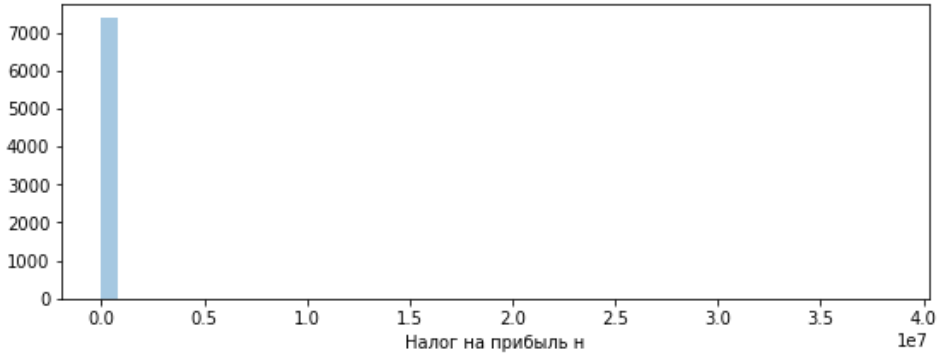

0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
НДФЛ н
1e7

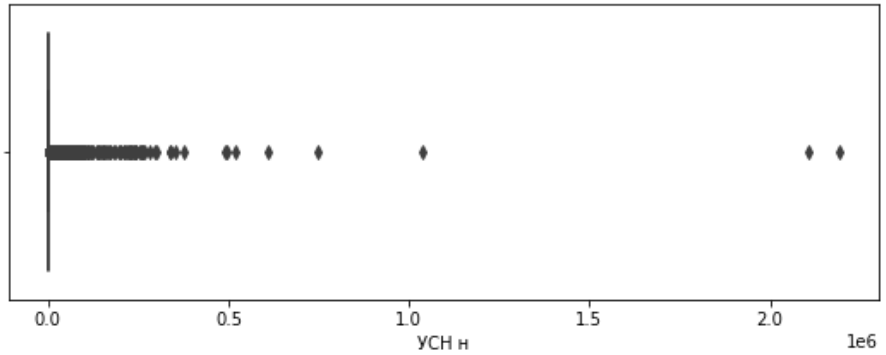

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
НДФЛ н
1e7

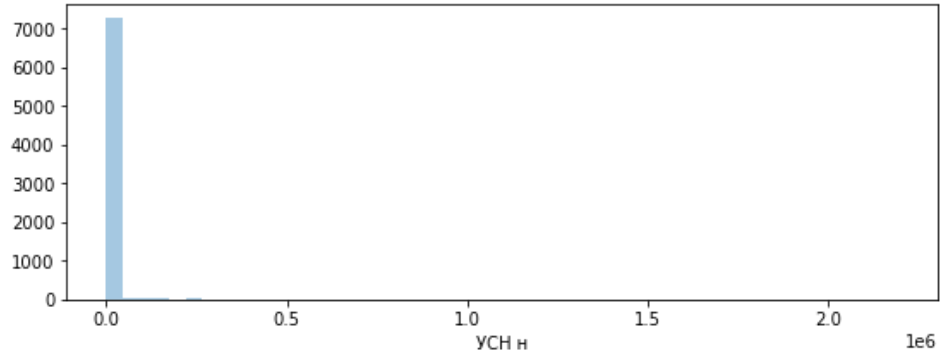

0
1
2
3
4
5
6
7
8
Налог на имущество организаций н
1e7

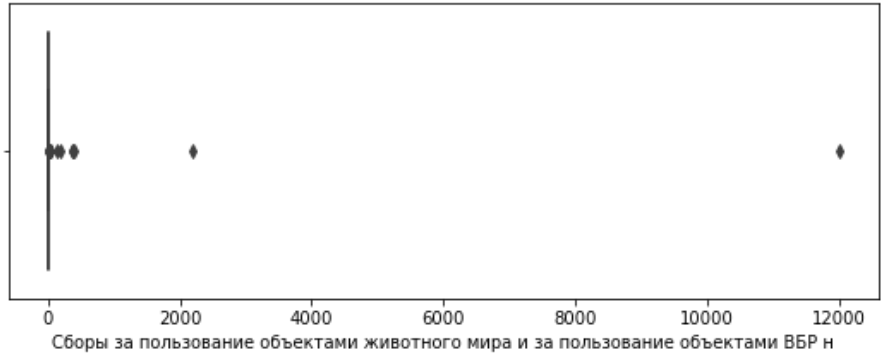

7000
6000
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
7
8
Налог на имущество организаций н
1e7

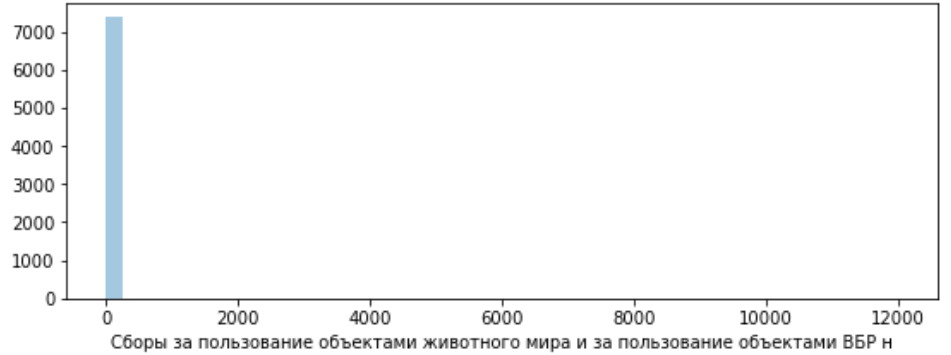

0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Налог на прибыль н
1e7

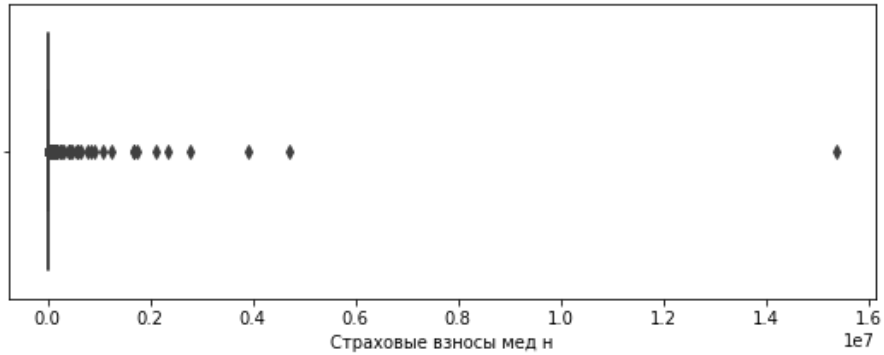

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Налог на прибыль н
1e7

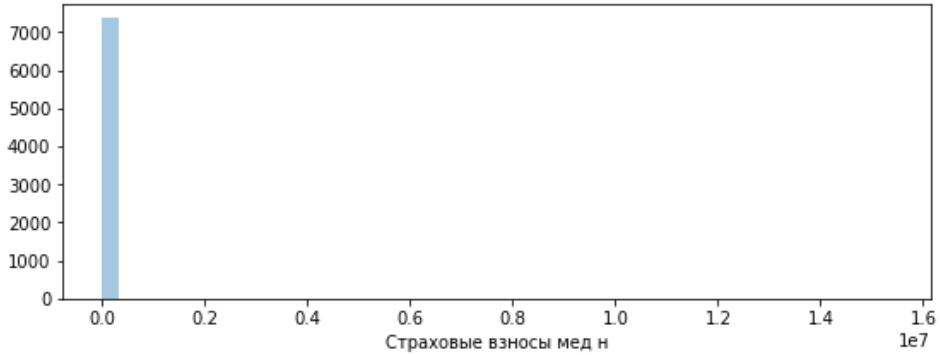

0.0
0.5
1.0
1.5
2.0
УСН н
1e6

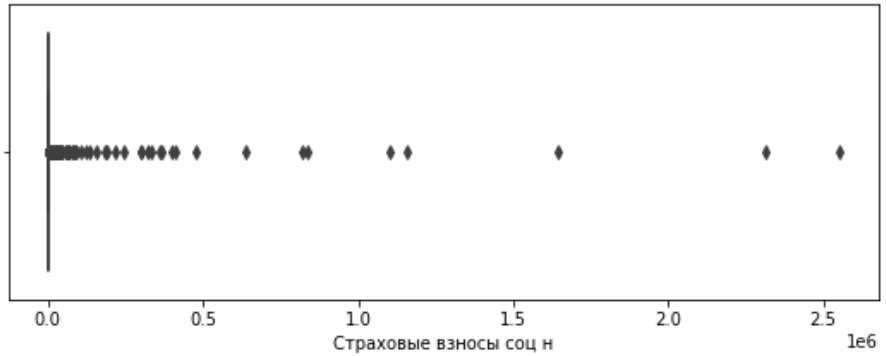

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
УСН н
1e6

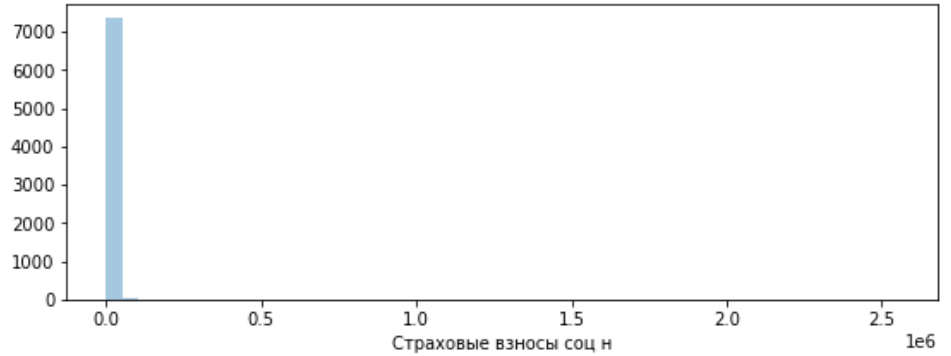

0
2000
4000
6000
8000
10000
12000
Сборы за пользование объектами животного мира и за пользование объектами ВБР н

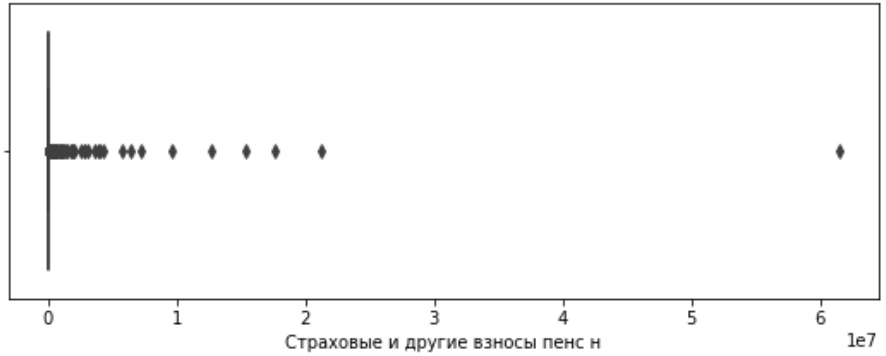

7000
6000
5000
4000
3000
2000
1000
0
0
2000
4000
6000
8000
10000
12000
Сборы за пользование объектами животного мира и за пользование объектами ВБР н

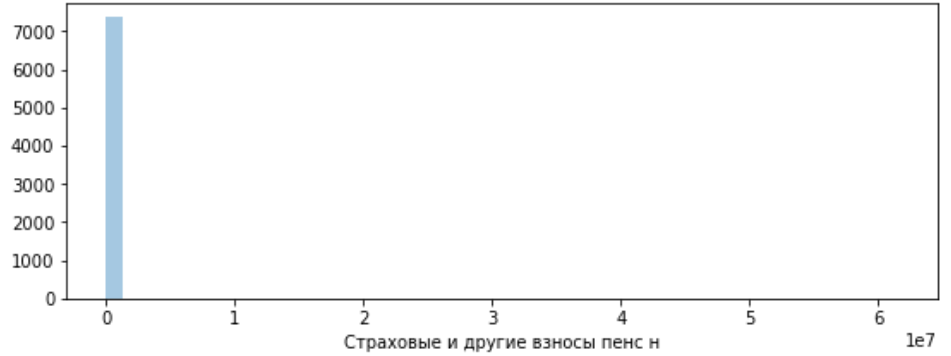

0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1e7
Страховые взносы мед н

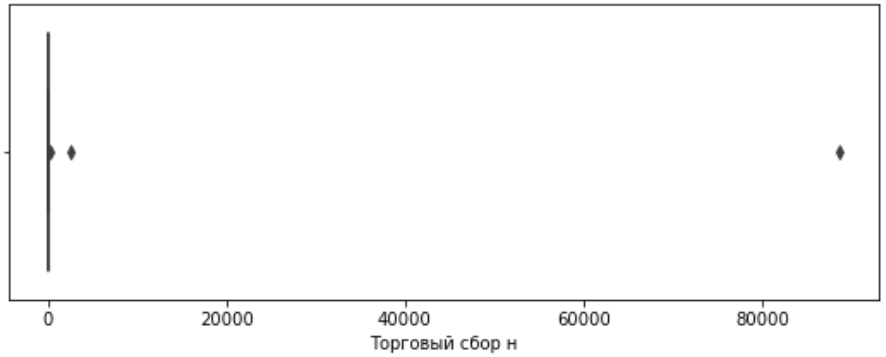

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1e7
Страховые взносы мед н

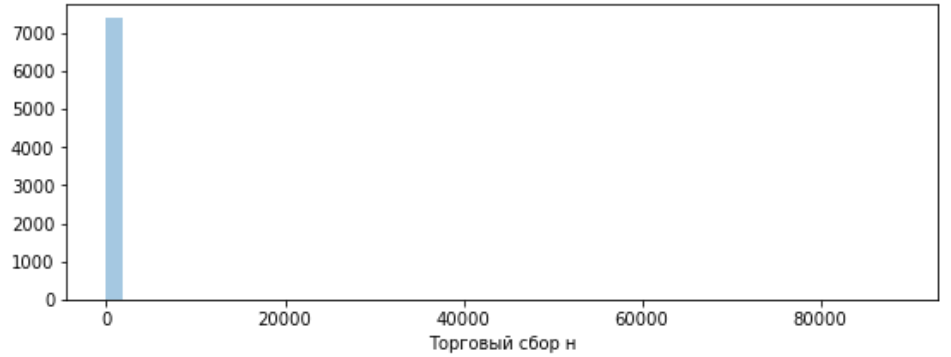

0.0
0.5
1.0
1.5
2.0
2.5
1e6
Страховые взносы соц н

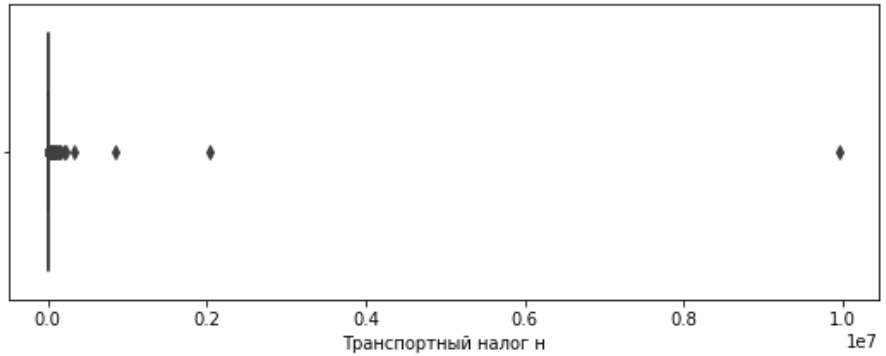

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
1e6
Страховые взносы соц н

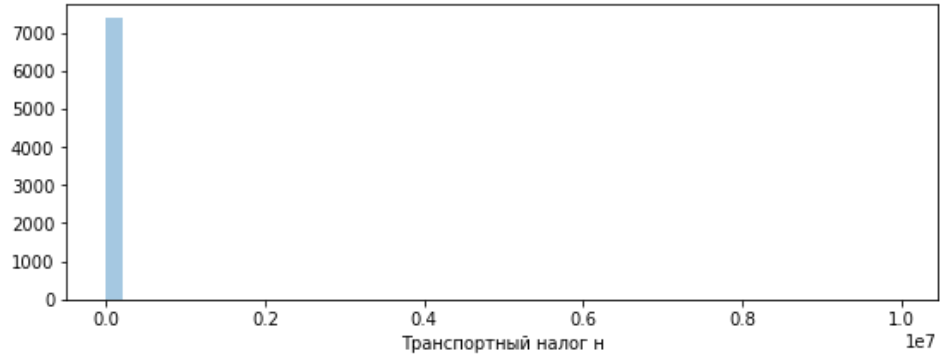

0
1
2
3
4
5
6
Страховые и другие взносы пенс н
1e7

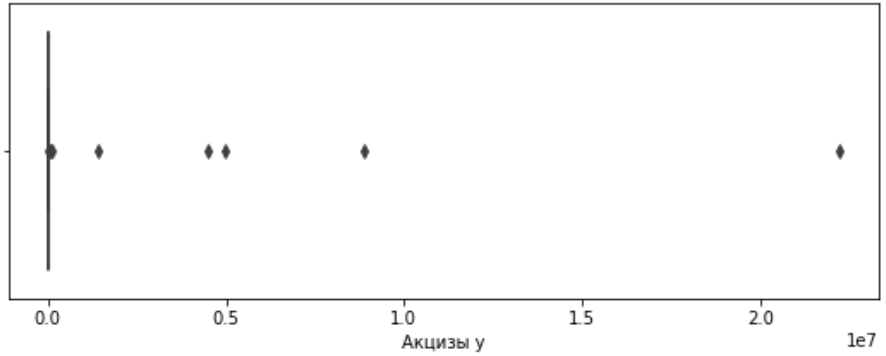

7000
6000
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
Страховые и другие взносы пенс н
1e7

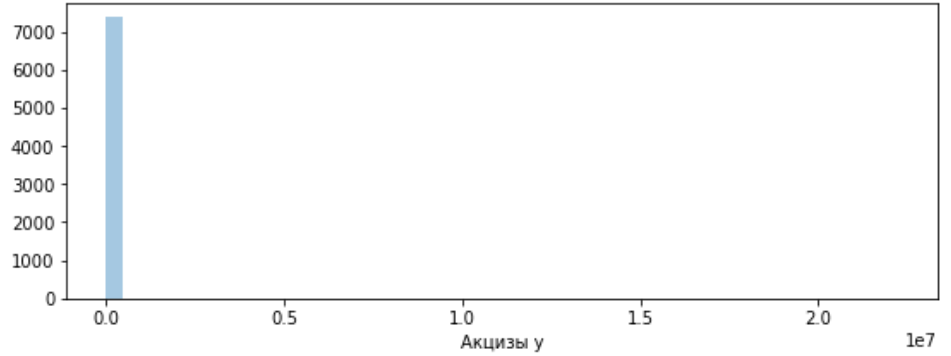

0
20000
40000
60000
80000
Торговый сбор н

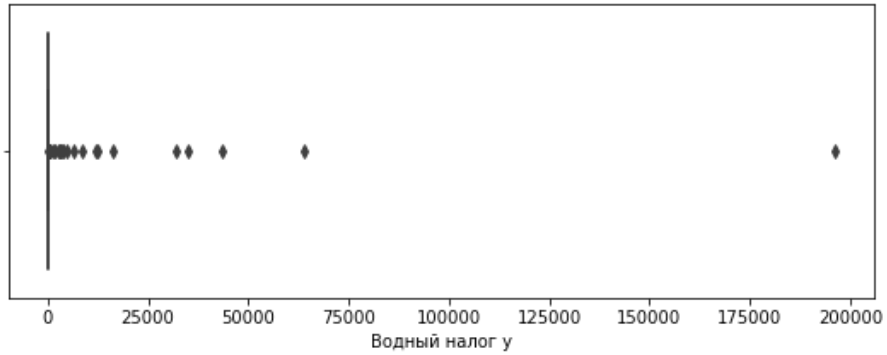

7000
6000
5000
4000
3000
2000
1000
0
0
20000
40000
60000
80000
Торговый сбор н

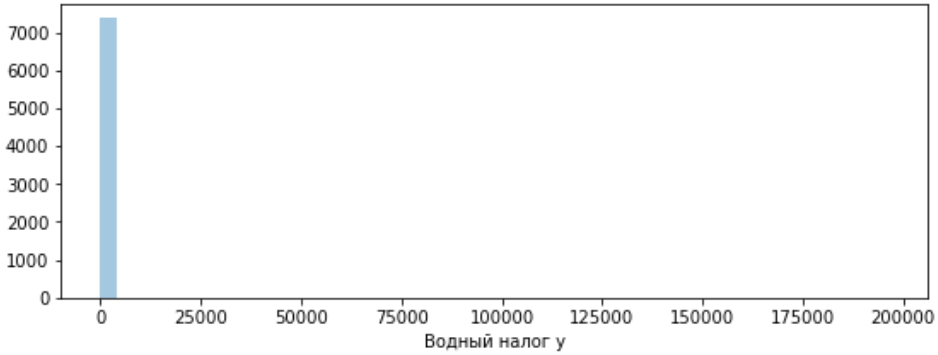

0.0
0.2
0.4
0.6
0.8
1.0
Транспортный налог н
1e7

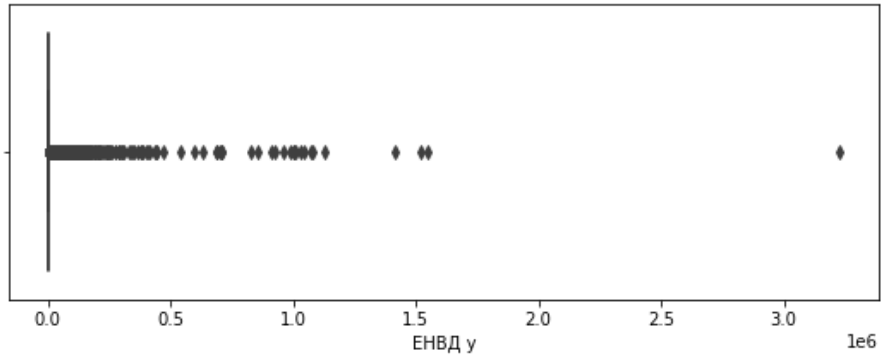

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
Транспортный налог н
1e7

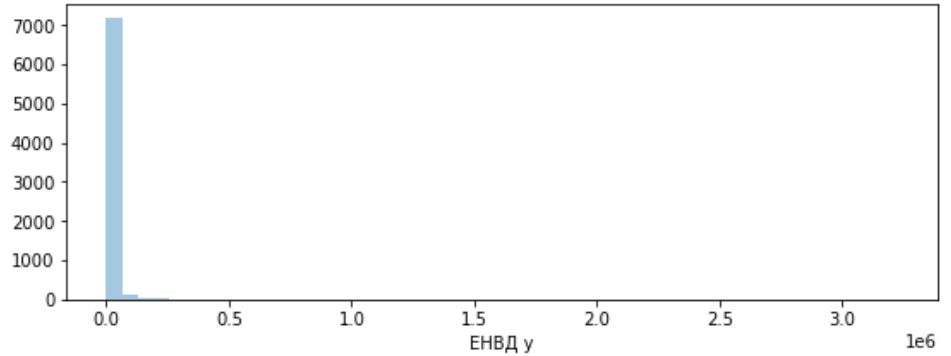

0.0
0.5
1.0
1.5
2.0
Акцизы у
1e7

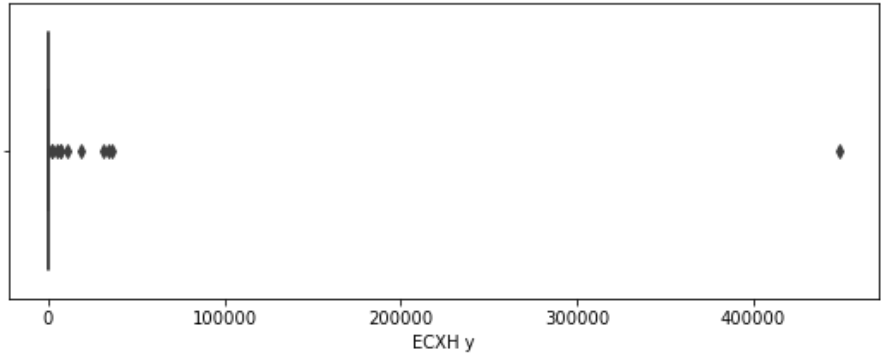

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
Акцизы у
1e7

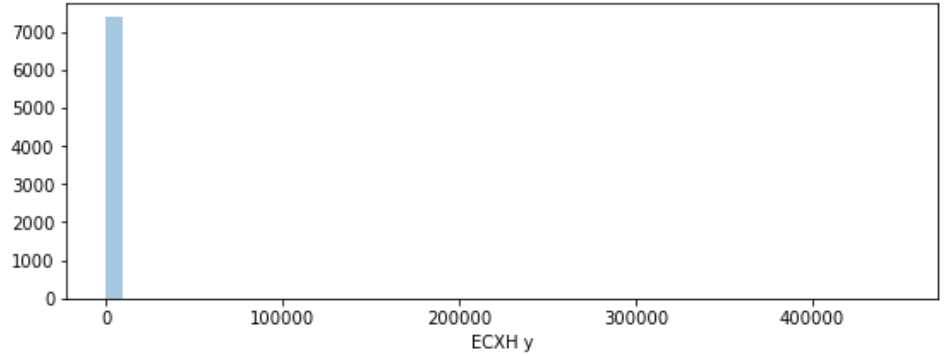

0
25000
50000
75000
100000
125000
150000
175000
200000
Водный налог у

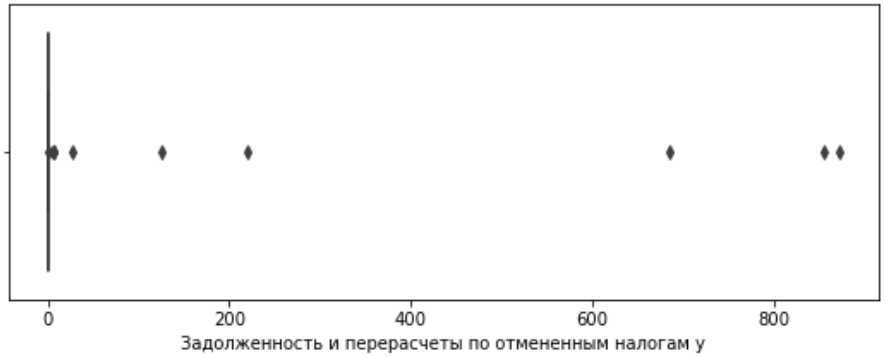

7000
6000
5000
4000
3000
2000
1000
0
0
25000
50000
75000
100000
125000
150000
175000
200000
Водный налог у

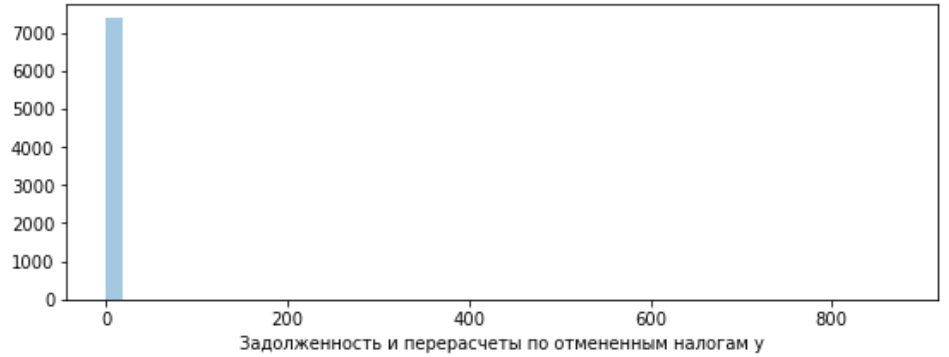

0.0
0.5
1.0
1.5
2.0
2.5
3.0
ЕНВД у
1e6

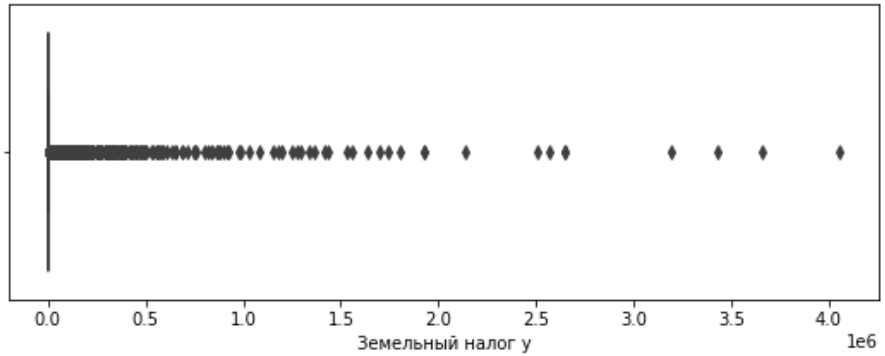

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
ЕНВД у
1e6

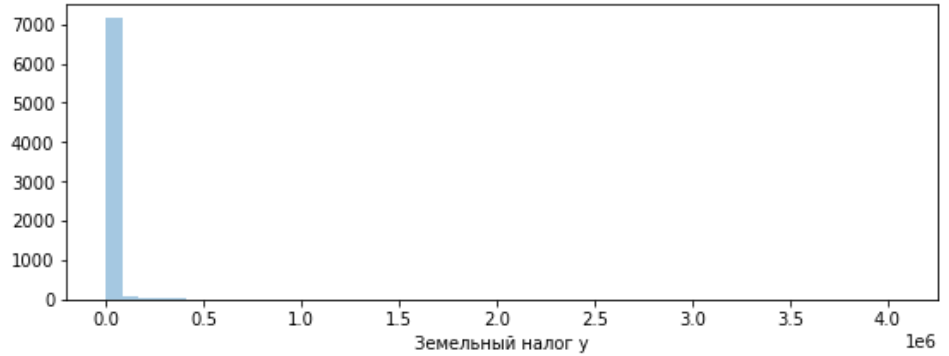

0
100000
200000
300000
400000
ЕСХН у

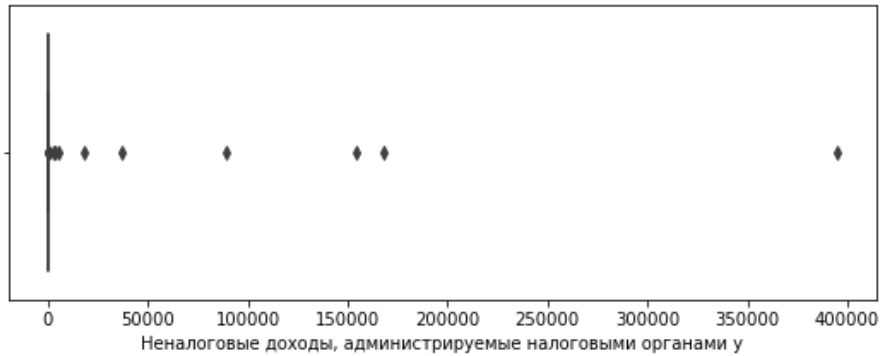

7000
6000
5000
4000
3000
2000
1000
0
0
100000
200000
300000
400000
ЕСХН у

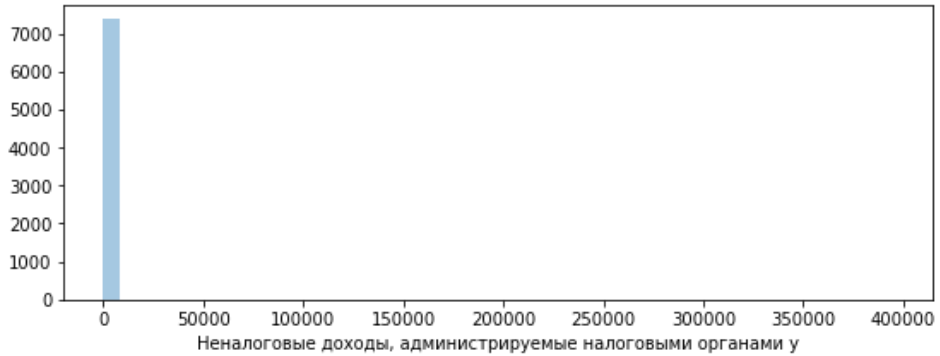

0
200
400
600
800
Задолженность и перерасчеты по отмененным налогам у

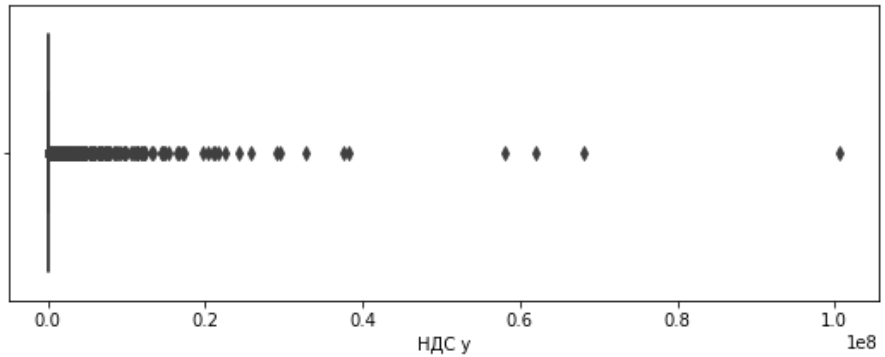

7000
6000
5000
4000
3000
2000
1000
0
0
200
400
600
800
Задолженность и перерасчеты по отмененным налогам у

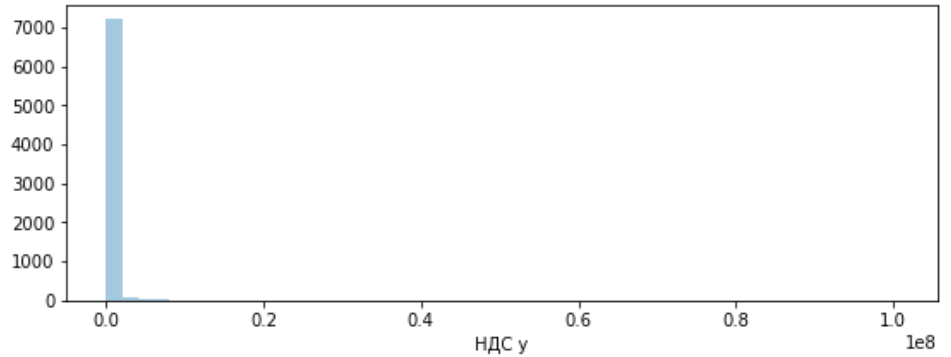

0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e6
Земельный налог у

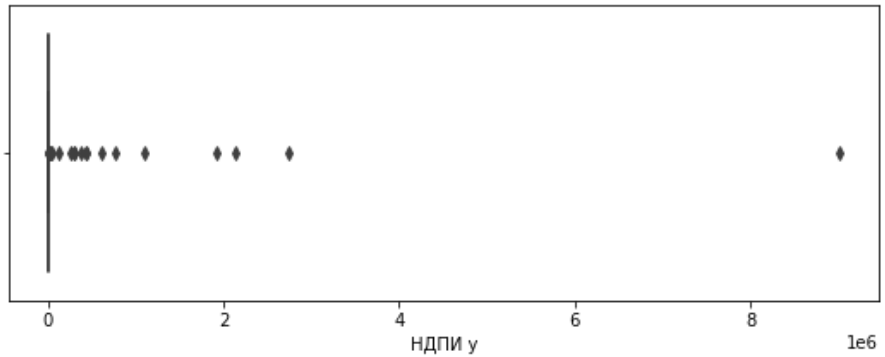

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e6
Земельный налог у

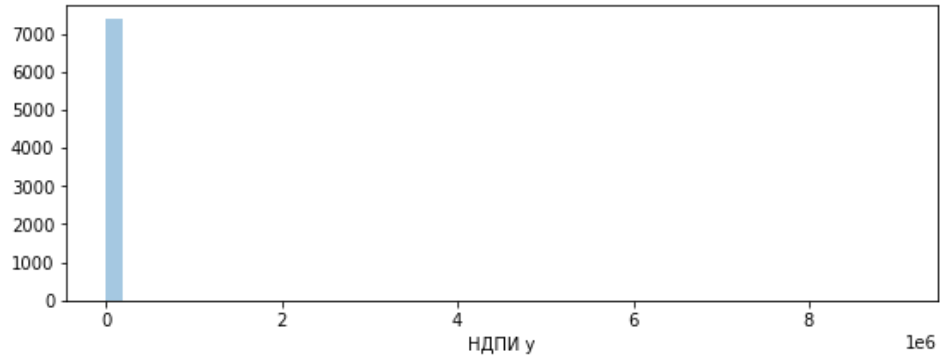

0
50000
100000
150000
200000
250000
300000
350000
400000
Неналоговые доходы, администрируемые налоговыми органами y

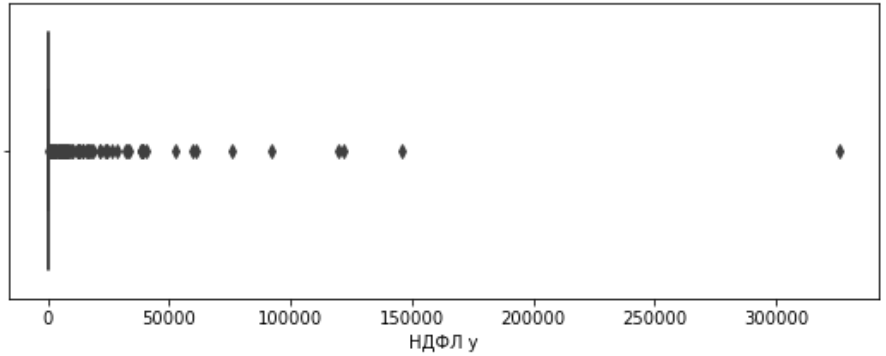

7000
6000
5000
4000
3000
2000
1000
0
0
50000
100000
150000
200000
250000
300000
350000
400000
Неналоговые доходы, администрируемые налоговыми органами y

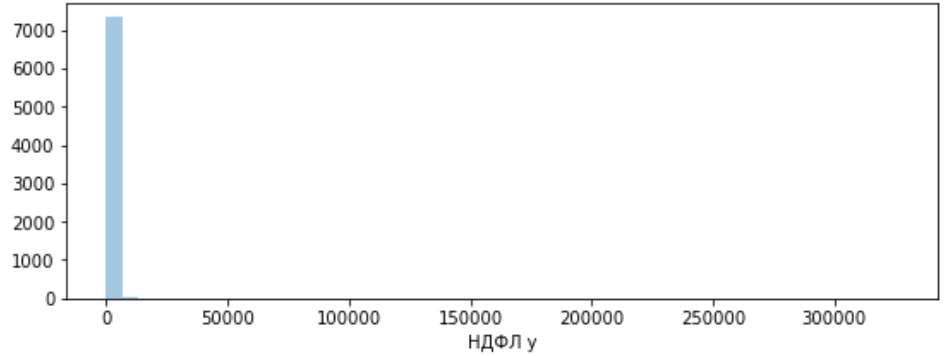

0.0
0.2
0.4
0.6
0.8
1.0
1e8
НДС y

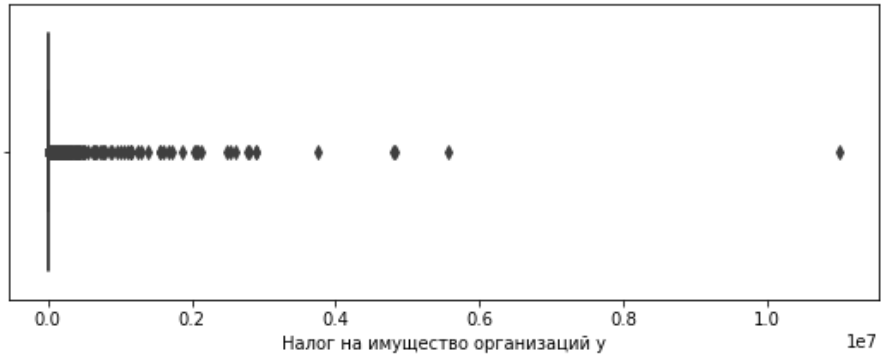

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
1e8
НДС y

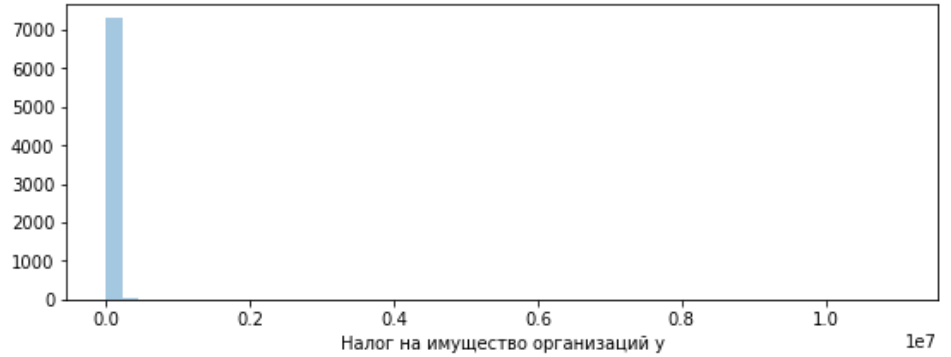

0
2
4
6
8
1e6
НДПИ y

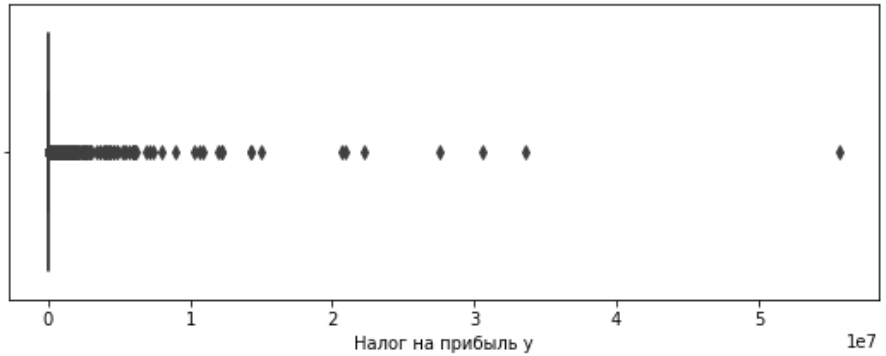

7000
6000
5000
4000
3000
2000
1000
0
0
2
4
6
8
1e6
НДПИ y

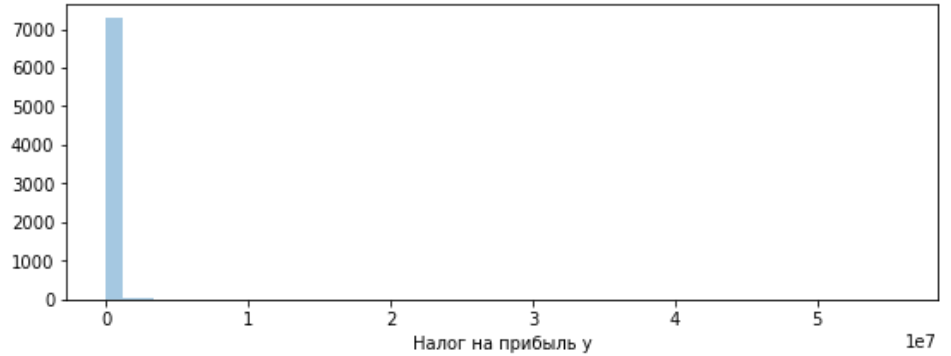

0
50000
100000
150000
200000
250000
300000
НДФЛ у

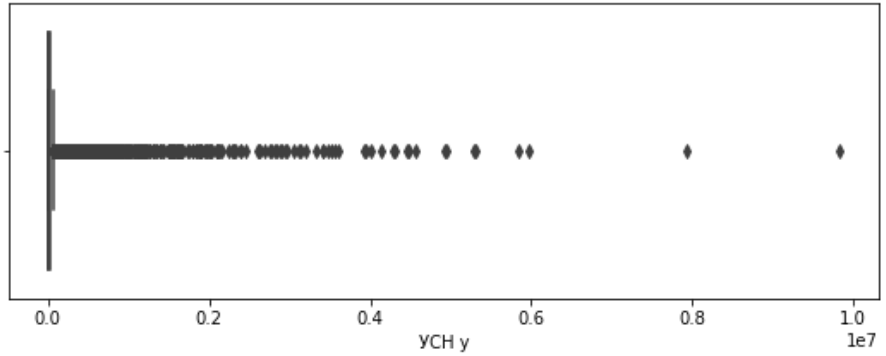

7000
6000
5000
4000
3000
2000
1000
0
0
50000
100000
150000
200000
250000
300000
НДФЛ у

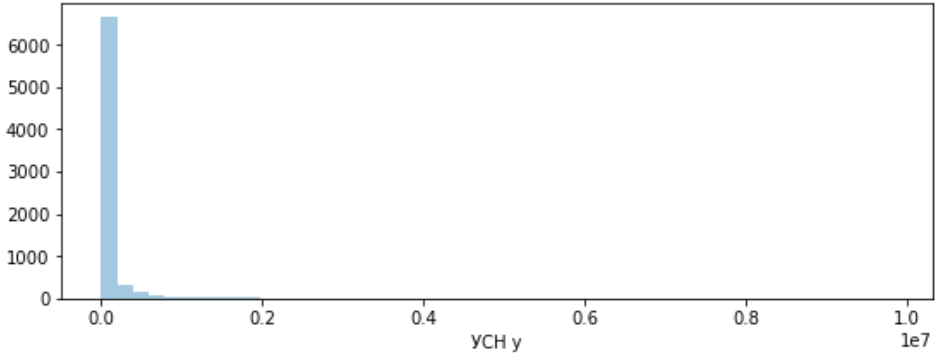

0.0
0.2
0.4
0.6
0.8
1.0
Налог на имущество организаций у
1e7

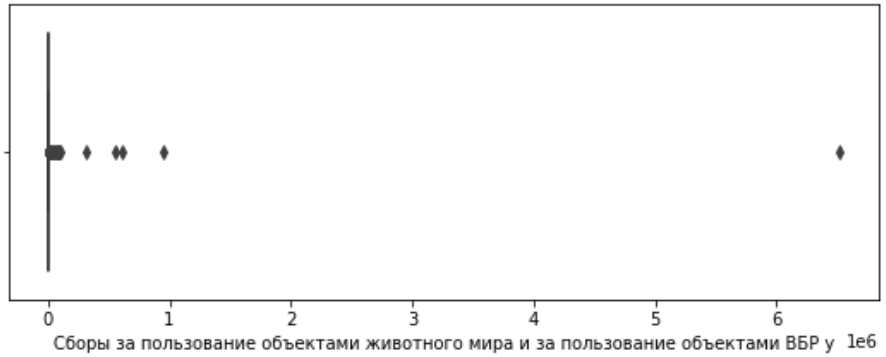

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
Налог на имущество организаций у
1e7

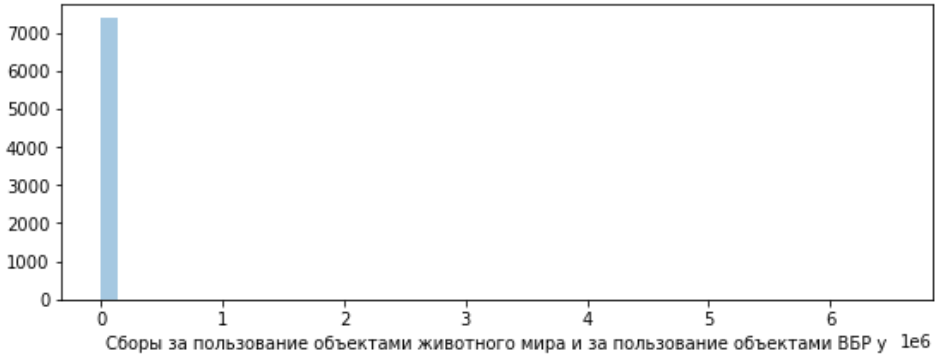

0
1
2
3
4
5
Налог на прибыль у
1e7

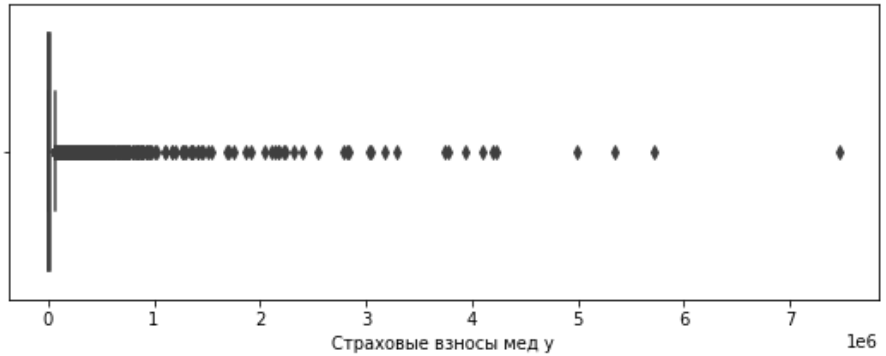

7000
6000
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
Налог на прибыль у
1e7

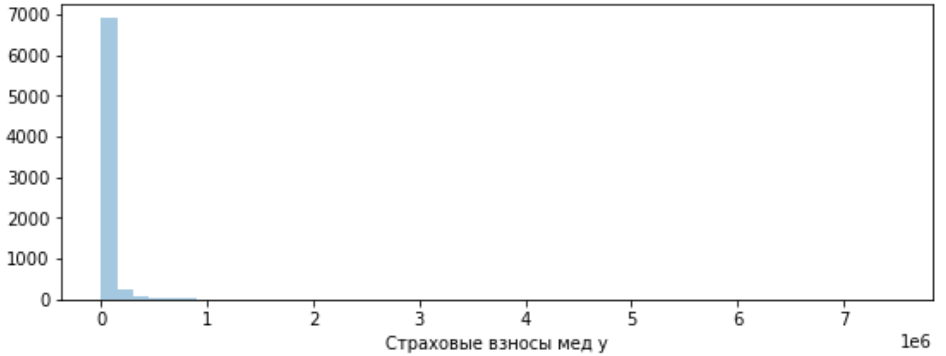

0.0
0.2
0.4
0.6
0.8
1.0
УСН у
1e7

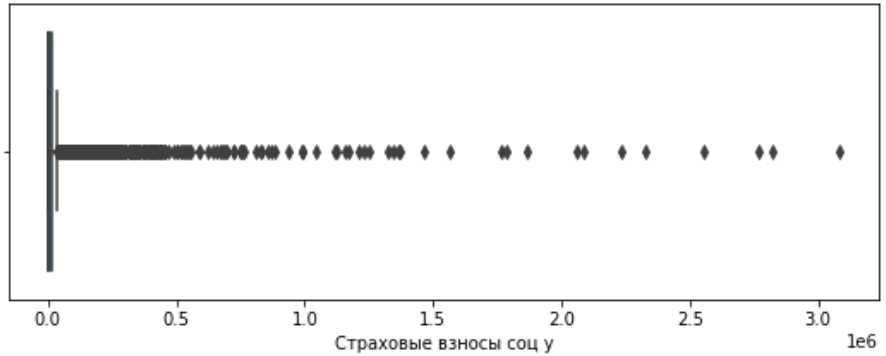

6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
УСН у
1e7

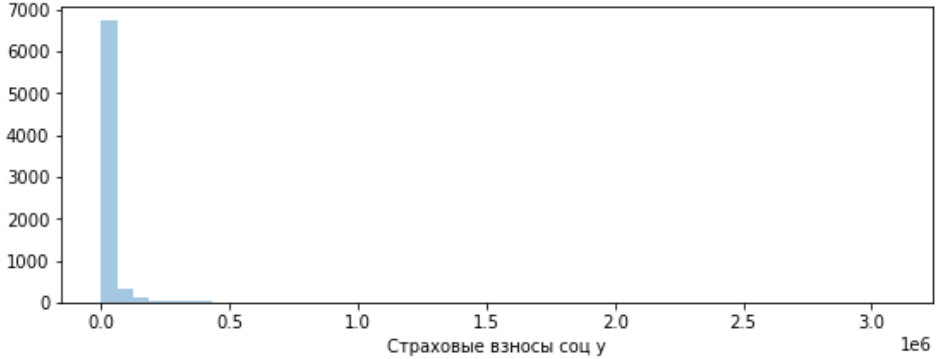

0
1
2
3
4
5
6
Сборы за пользование объектами животного мира и за пользование объектами ВБР у
1e6

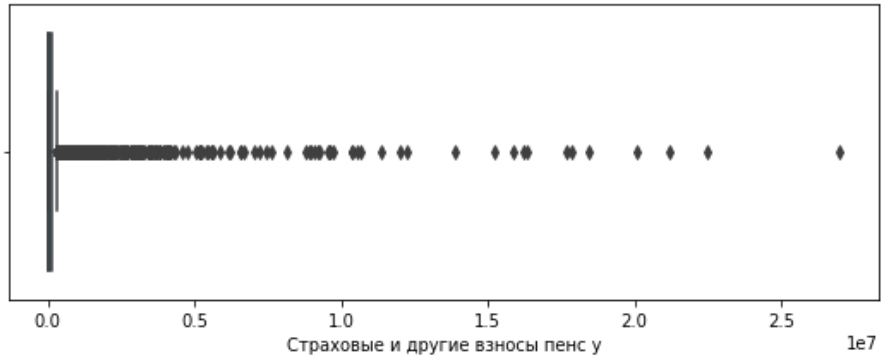

7000
6000
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
Сборы за пользование объектами животного мира и за пользование объектами ВБР у
1e6

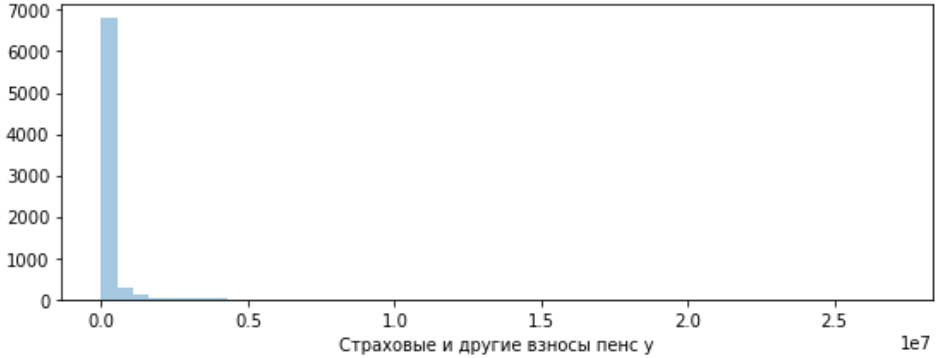

0
1
2
3
4
5
6
7
Страховые взносы мед у
1e6

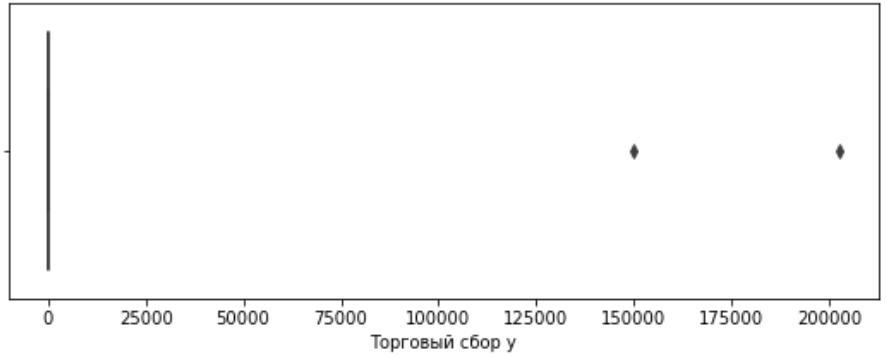

7000
6000
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
7
Страховые взносы мед у
1e6

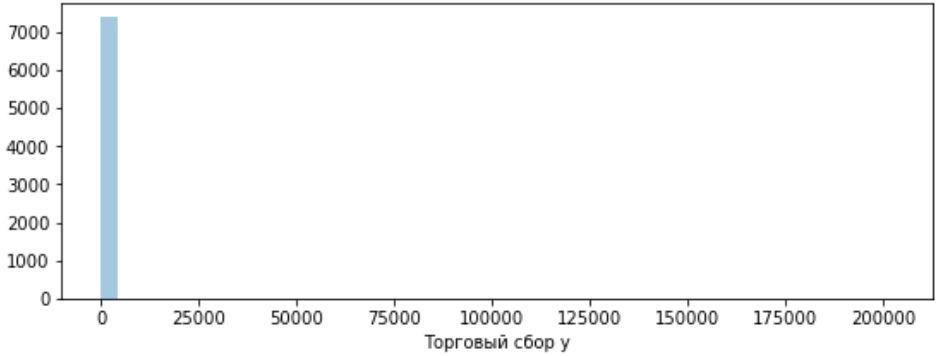

0.0
0.5
1.0
1.5
2.0
2.5
3.0
Страховые взносы соц у
1e6

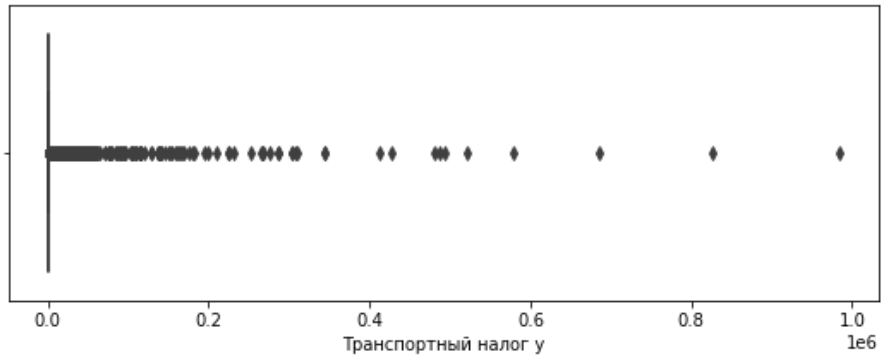

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Страховые взносы соц у
1e6

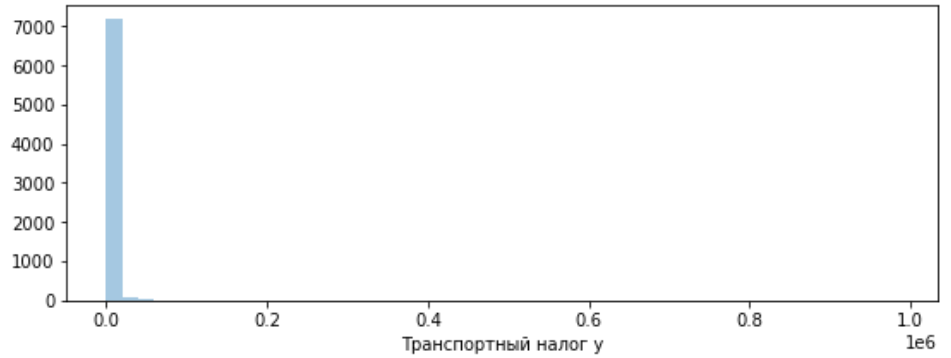

0.0
0.5
1.0
1.5
2.0
2.5
Страховые и другие взносы пенс у
1e7

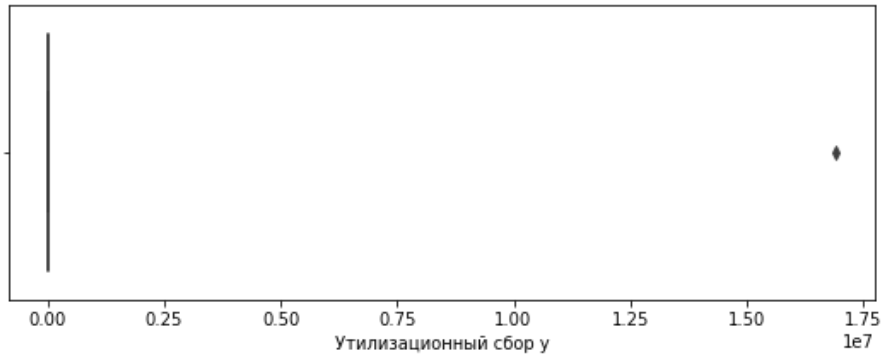

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
Страховые и другие взносы пенс у
1e7

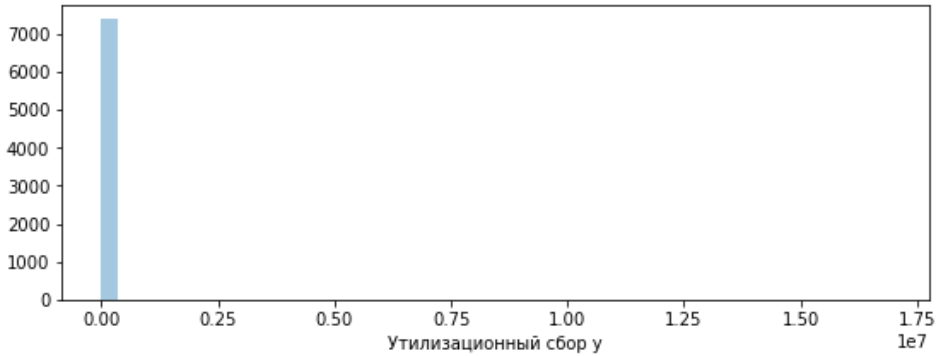

0
25000
50000
75000
100000
125000
150000
175000
200000
Торговый сбор у


7000
6000
5000
4000
3000
2000
1000
0
0
25000
50000
75000
100000
125000
150000
175000
200000
Торговый сбор у


0.0
0.2
0.4
0.6
0.8
1.0
Транспортный налог у
1e6

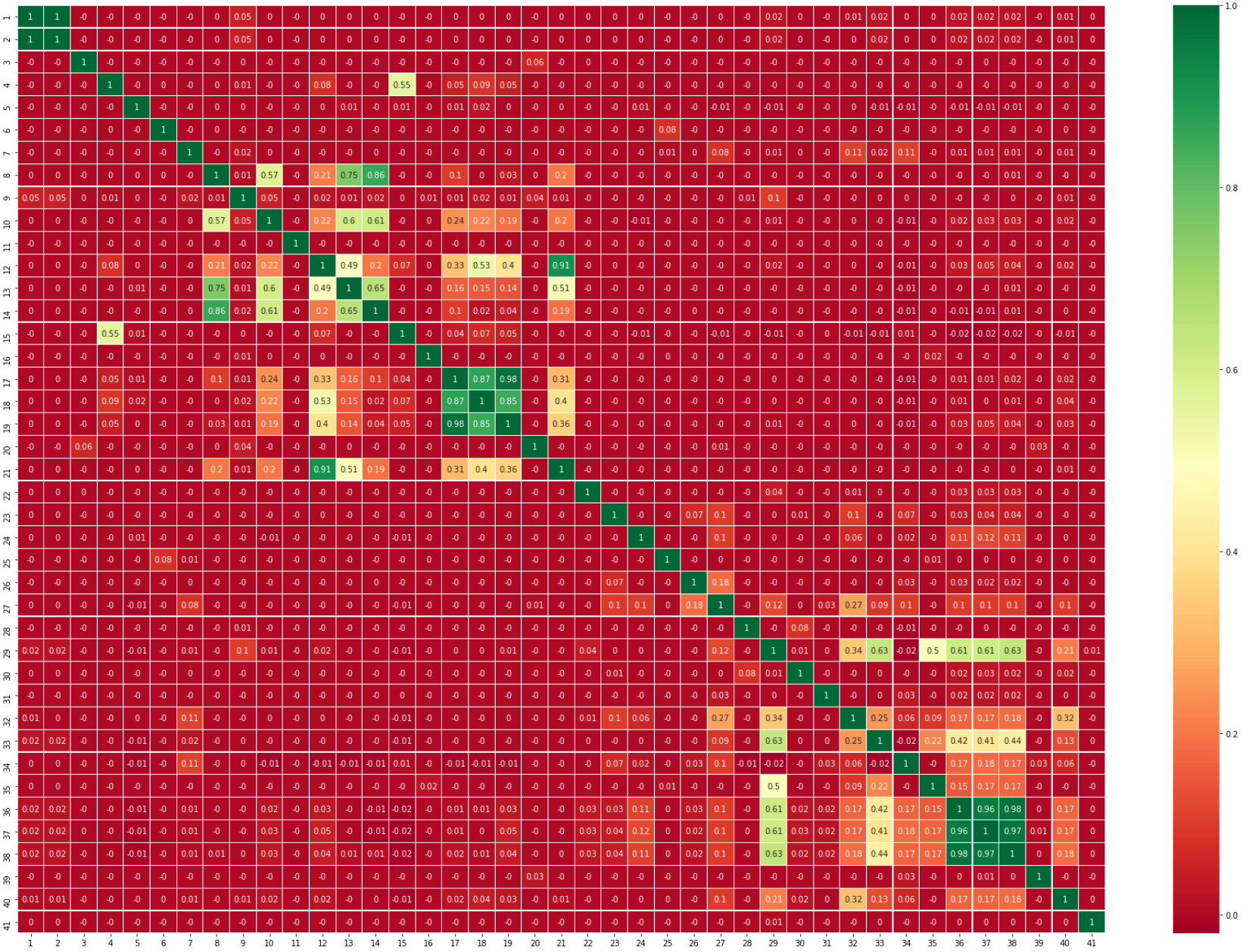

7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
Транспортный налог у
1e6


0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Утилизационный сбор у
1e7


7000
6000
5000
4000
3000
2000
1000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Утилизационный сбор у
1e7

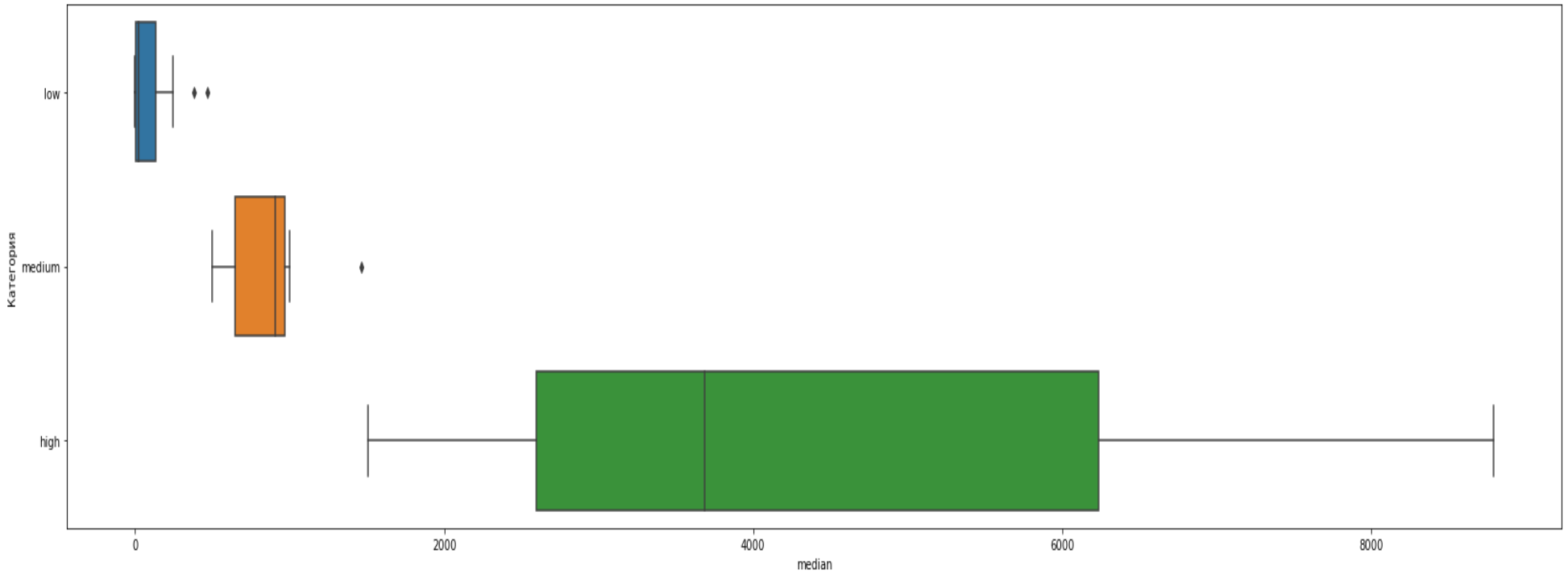
ПРИЛОЖЕНИЕ Е

КОРРЕЛЯЦИОННАЯ МАТРИЦА ЛИНЕЙНЫХ ЗАВИСИМОСТЕЙ МЕЖДУ ЧИСЛОВЫМИ ПЕРЕМЕННЫМИ



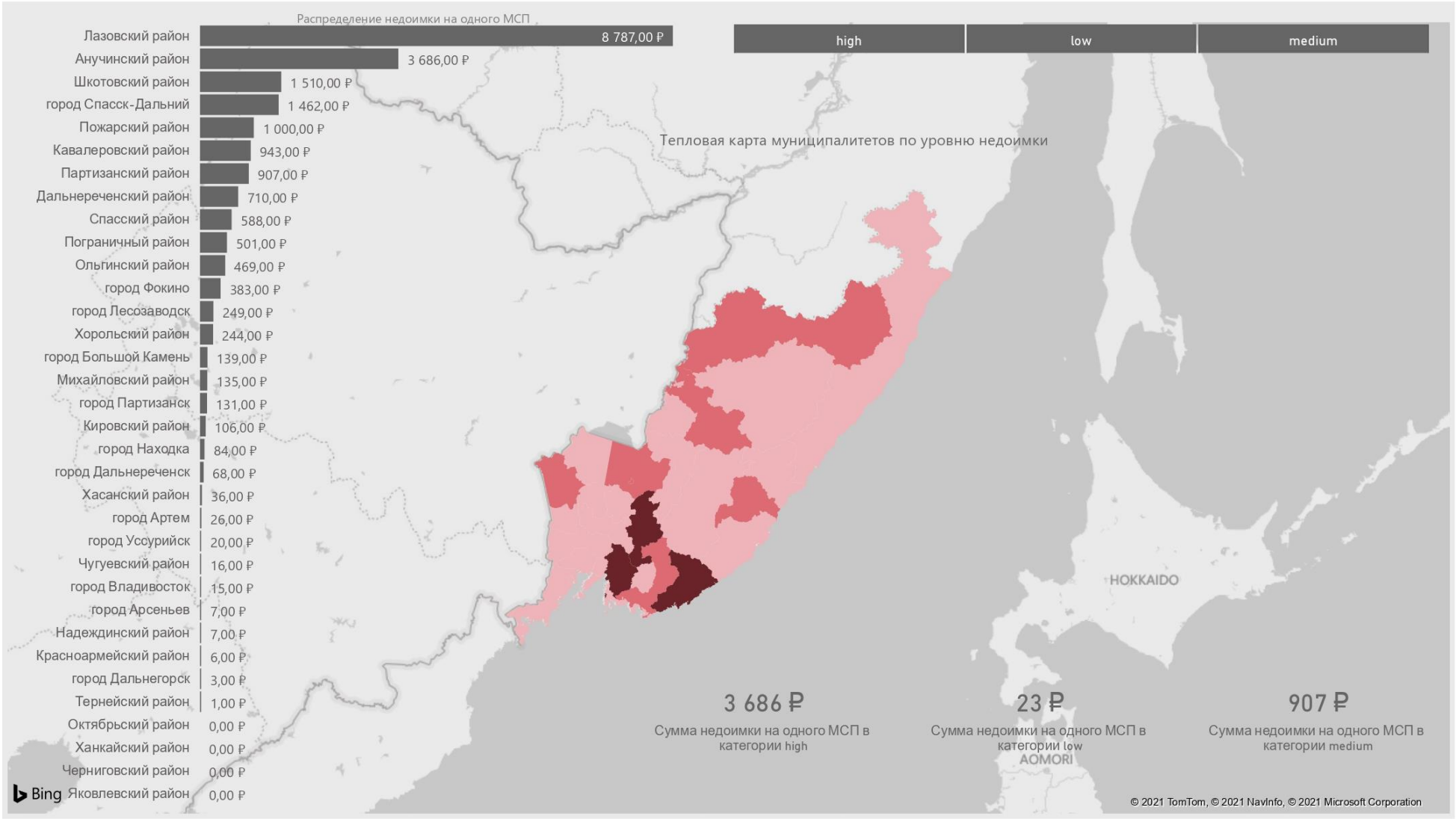
1. Доход
2. Расход
3. Акцизы н
4. Водный налог н
5. ЕНВД н
6. ЕСХН н
7. Задолженность и перерасчеты по отмененным налогам н
8. Земельный налог н
9. Неналоговые доходы, администрируемые налоговыми органами н
10. НДС н
11. НДПИ н
12. НДФЛ н
13. Налог на имущество организаций н
14. Налог на прибыль н
15. УСН н
16. Сборы за пользование объектами животного мира и за пользование объектами ВБР н
17. Страховые взносы мед н
18. Страховые взносы соц н
19. Страховые и другие взносы пенс н
20. Торговый сбор н
21. Транспортный налог н
22. Акцизы у
23. Водный налог у
24. ЕНВД у
25. ЕСХН у
26. Задолженность и перерасчеты по отмененным налогам у
27. Земельный налог у
28. Неналоговые доходы, администрируемые налоговыми органами у
29. НДС у
30. НДПИ у
31. НДФЛ у
32. Налог на имущество организаций
33. Налог на прибыль у
34. УСН у
35. Сборы за пользование объектами животного мира и за пользование объектами ВБР у
36. Страховые взносы мед у
37. Страховые взносы соц у
38. Страховые и другие взносы пенс у
39. Торговый сбор у
40. Транспортный налог у
41. Утилизационный сбор у

РАСПРЕДЕЛЕНИЯ НЕДОИМКИ НА ОДНОГО СУБЪЕКТА МСП В РАЗРЕЗЕ КАТЕГОРИЙ LOW, MEDIUM, HIGH



LOW – неуплата налогов на одного субъекта МСП не превышает 500 рублей
MEDIUM - неуплата налогов на одного субъекта от 501 до 1500 рублей
HIGH – неуплата налогов на одного субъекта МСП более 1500 рублей

ИТОГОВЫЙ ОТЧЕТ, ПОСТРОЕНЫЙ В POWER BI



РЕКОМЕНДАЦИИ ПО СОФЕРШЕНСТВОВАНИЮ ПРЕДЛОЖЕННОГО АНАЛИТИЧЕСКОГО РЕШЕНИЯ

№	Описание проблемы	Мероприятие по устранению	Ожидаемый результат
1.	Отсутствие данных по индивидуальным предпринимателям. В настоящее время в открытых данных ФНС России отсутствует информация по неуплаченным суммам налога индивидуальными предпринимателями.	Налаживание взаимодействия с региональными налоговыми органами по вопросу предоставления данных о суммах неуплаченных налогов индивидуальными предпринимателями.	Получена информация о суммах неуплаченных налогов индивидуальными предпринимателями. Возможно проводить более глубокий анализ налоговой платежеспособности субъектами малого и среднего предпринимательства.
2.	Наличие пропусков в данных по неуплате налогов юридическими лицами.	1.Разработка модели классификации, направленной на прогнозировании неуплаты налога субъектом малого и среднего предпринимательства (необходимо предварительно разметить имеющиеся данные, проставив признак 0 – уплачивает налоги, 1 – имеет неуплату налогов). 2.Разработка модели регрессии, направленной на прогнозирование сумм неуплаченных налогов у субъектов малого и среднего предпринимательства имеющих признак 1.	Увеличения объема данных для последующего статистического анализа.
3.	Тестирование гипотез об отличиях по неуплате налогов между категориями, выручкой у субъектов малого и среднего предпринимательства, а также видами экономической деятельности, осуществляемыми субъектами малого и среднего предпринимательства.	Получить ответы на следующие вопросы: 1.Отличается ли уровень неуплаченных налогов у субъектов малого и среднего предпринимательства в разрезе категорий (микро, малое, среднее)? 2. Отличается ли уровень неуплаченных налогов у субъектов малого и среднего предпринимательства в зависимости от объемов выручки? 3. Отличается ли уровень неуплаченных налогов у субъектов малого и среднего предпринимательства в зависимости от вида осуществляемой деятельности?	Полученные ответы на вопросы позволят установить: - как категория бизнеса влияет на уровень неуплаты налогов; - как размер выручки у бизнеса влияет на уровень неуплаты налогов; - влияет ли осуществляемый субъектом малого и среднего предпринимательства вид деятельности на объем неуплаченных налогов.