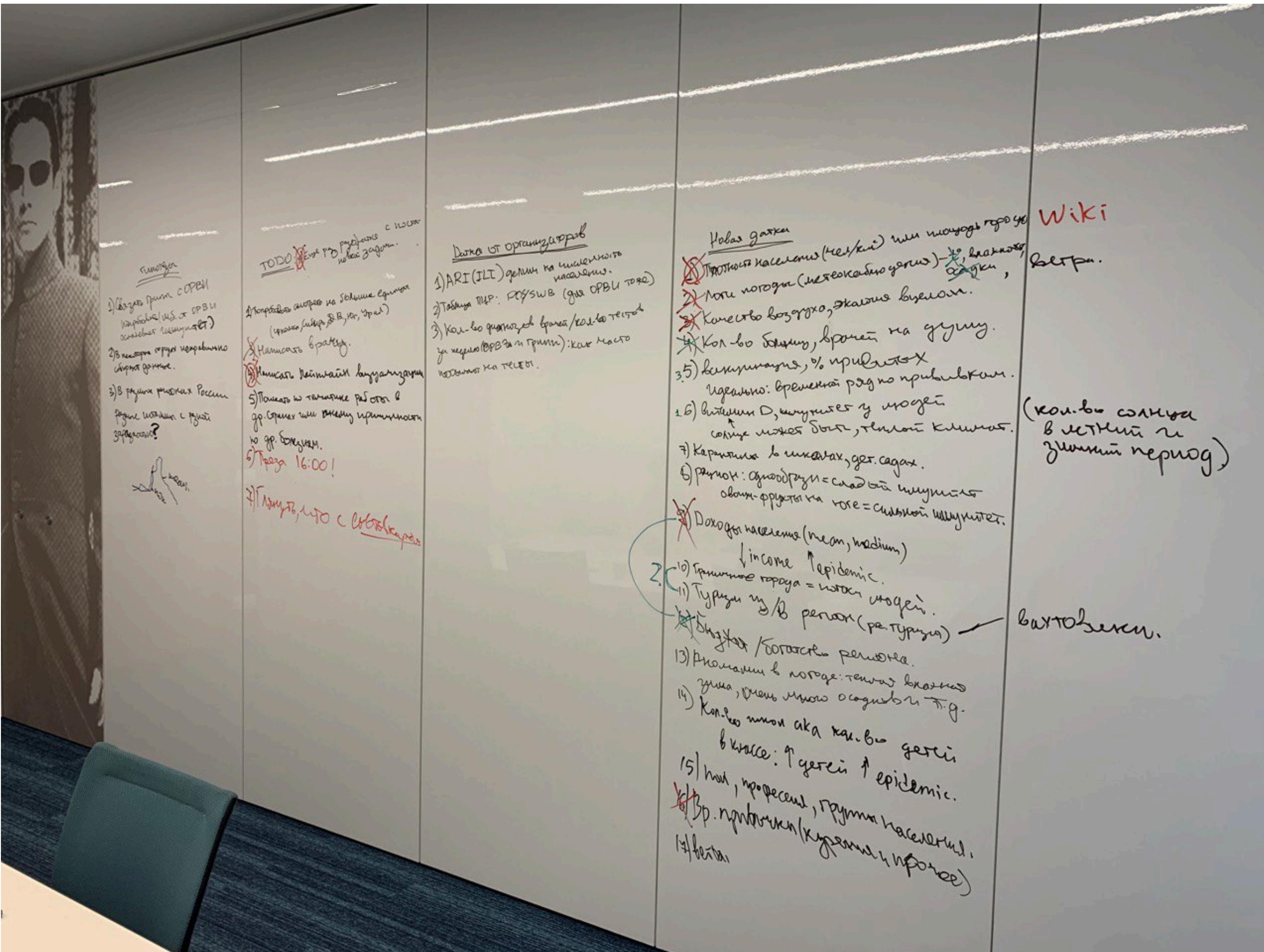


# **Поиск причин различия неэпидемиологического уровня ОРЗ по регионам**

## **НИИ Гриппа**

**Команда Koweschka**



У нас было очень много гипотез,  
но что-то пошло не по плану...

# План

- Данные
- Визуальный анализ
- Интерпретация признаков ML моделей
- Дальнейший исследования

# Данные

- Предоставленные организаторами
- Из открытых источников

# Данные из открытых источников

- Качество воздуха
- Плотность населения на квадратный километр
- Кол-во терапевтов и педиатров на душу населения
- Рейтинг вредных привычек регионов
- Температура и давление атмосферы для разных сезонов
- Доходы населения

# Изначальные гипотезы и данные



# Целевой признак – базовый уровень заболеваемости

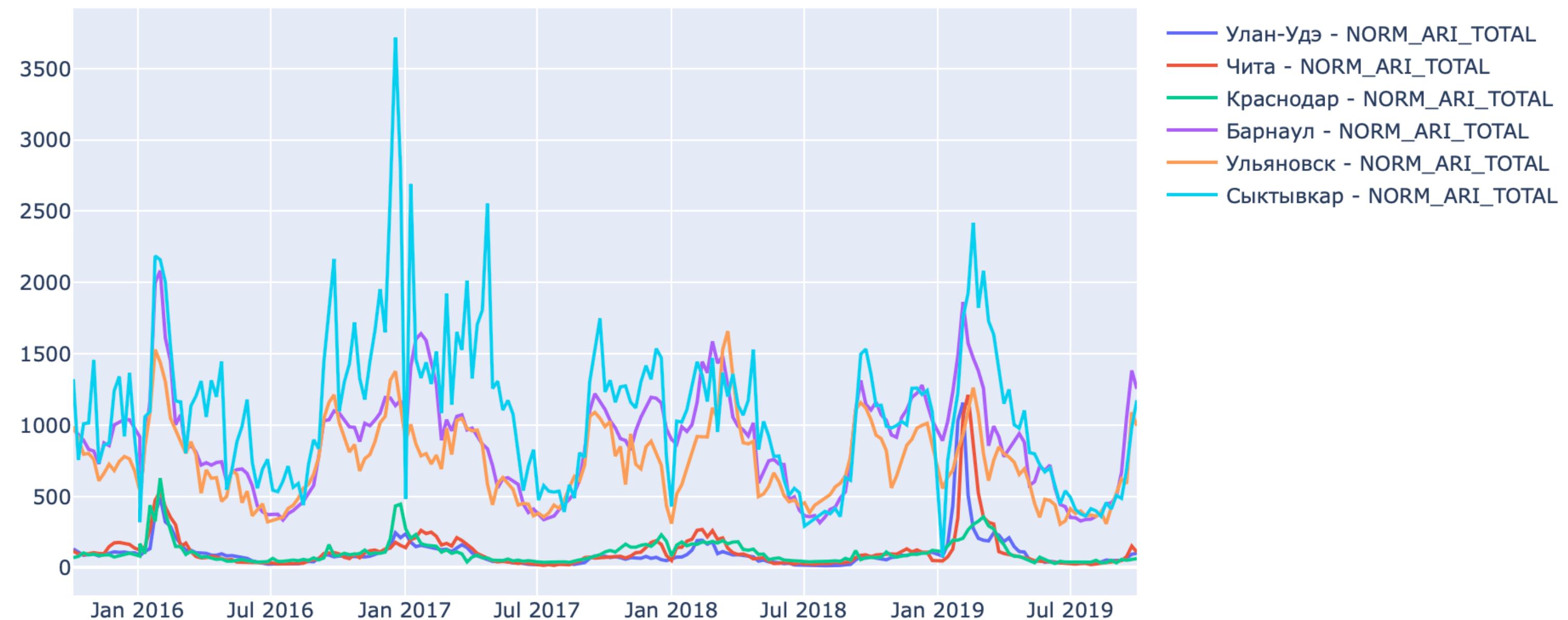
- Среднее значение количества заболевших на 100к человек за все внеэпидемиологические периоды для каждого города
- Позволяет обучать модели и ранжировать города



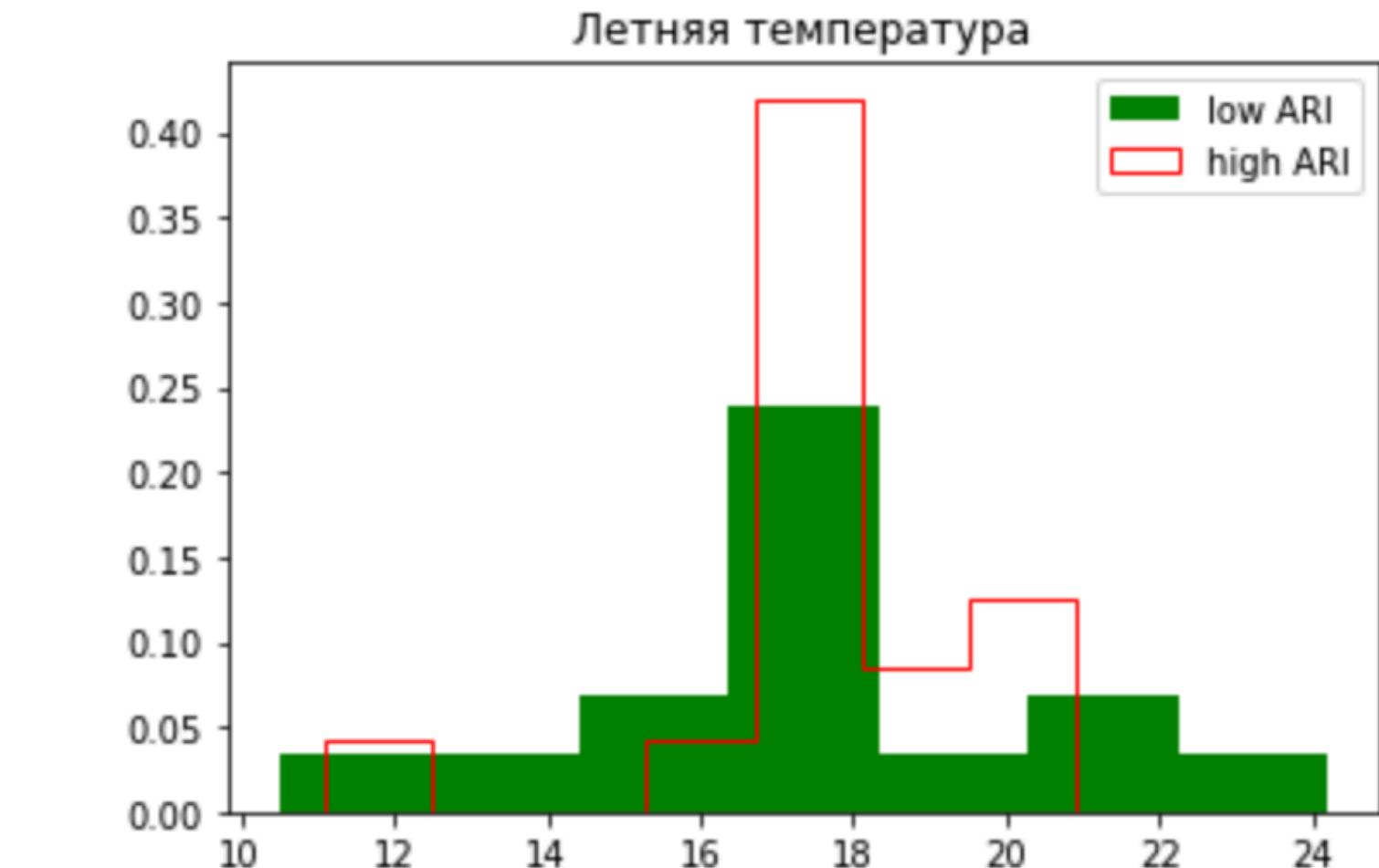
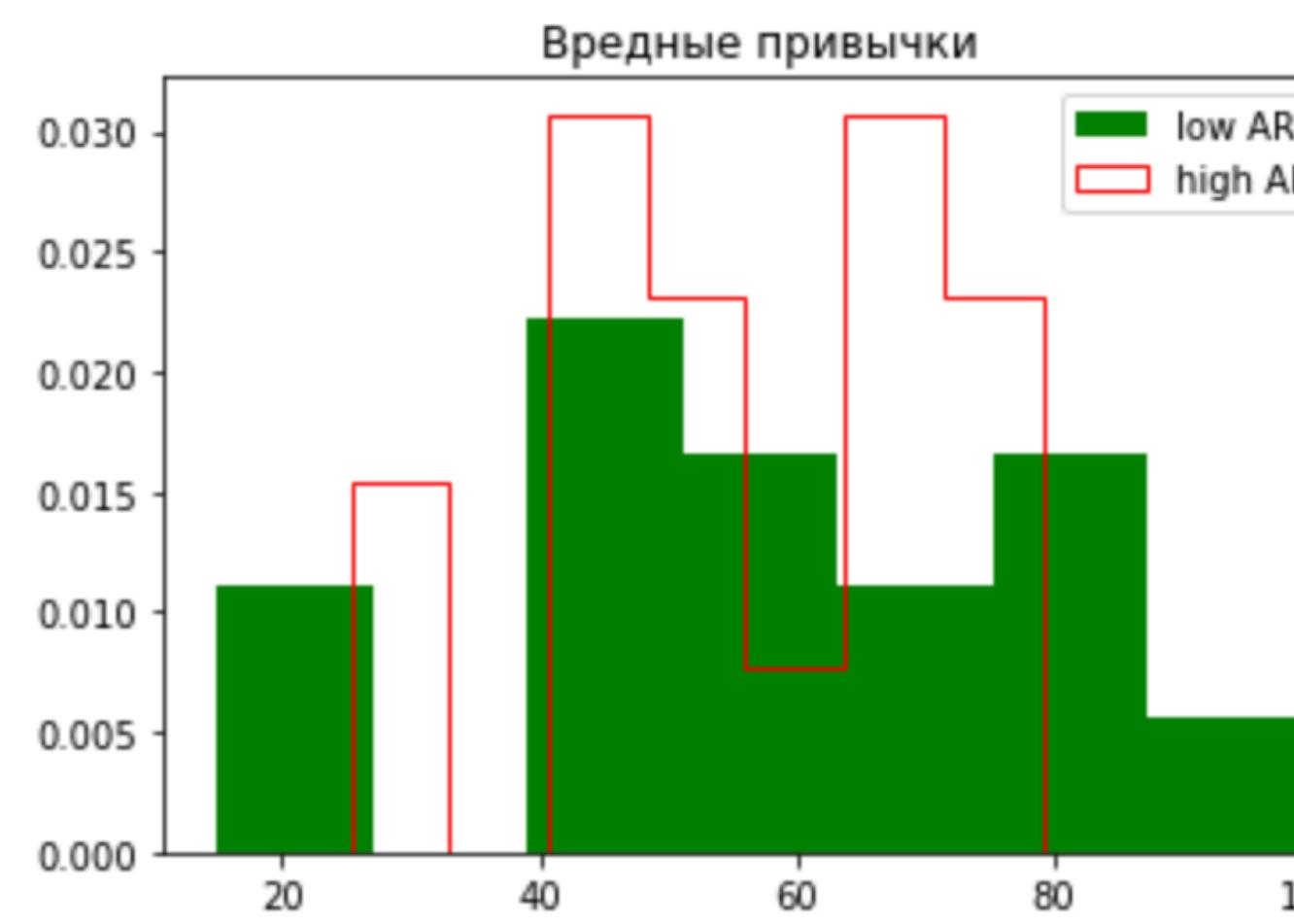
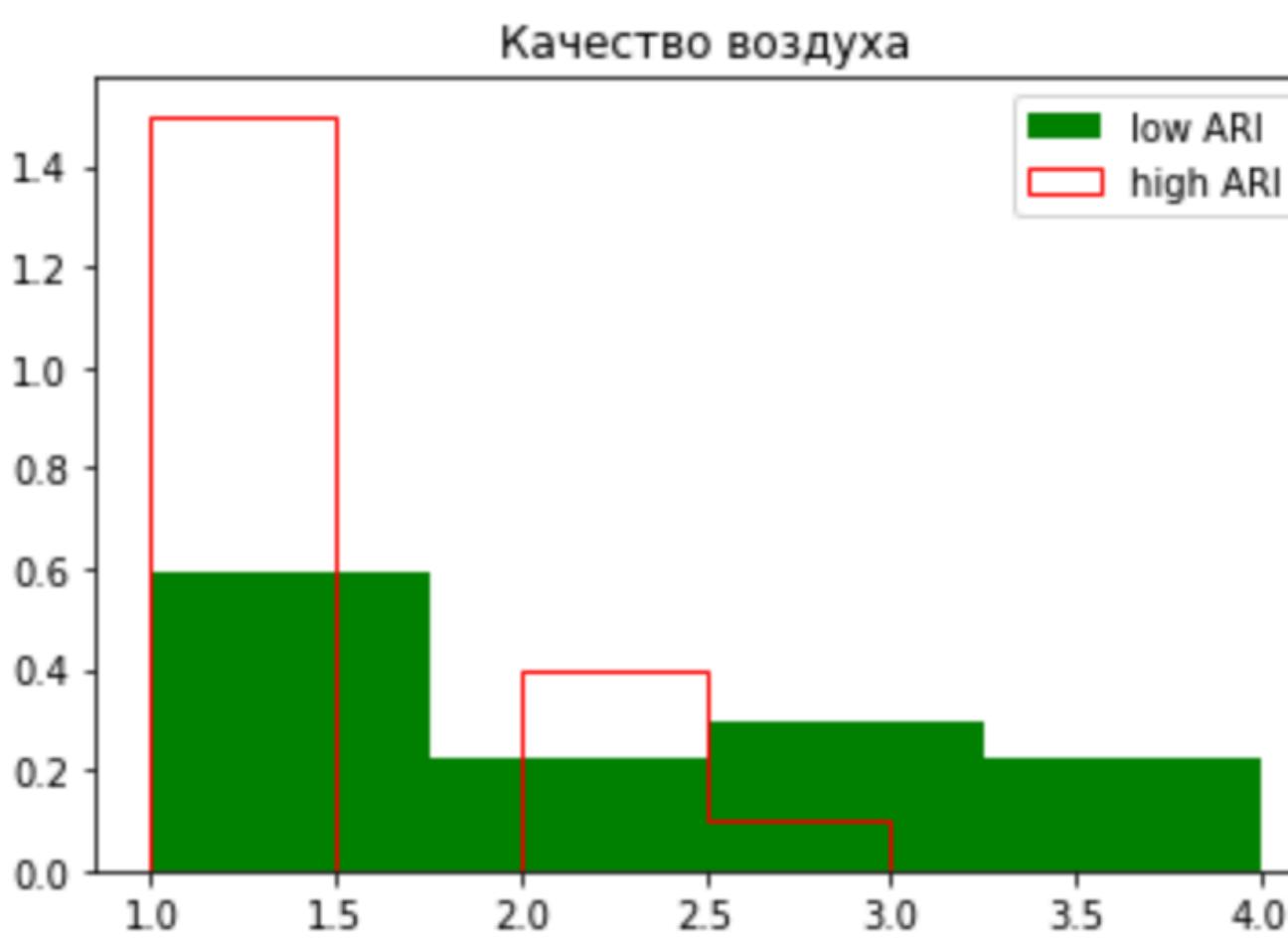
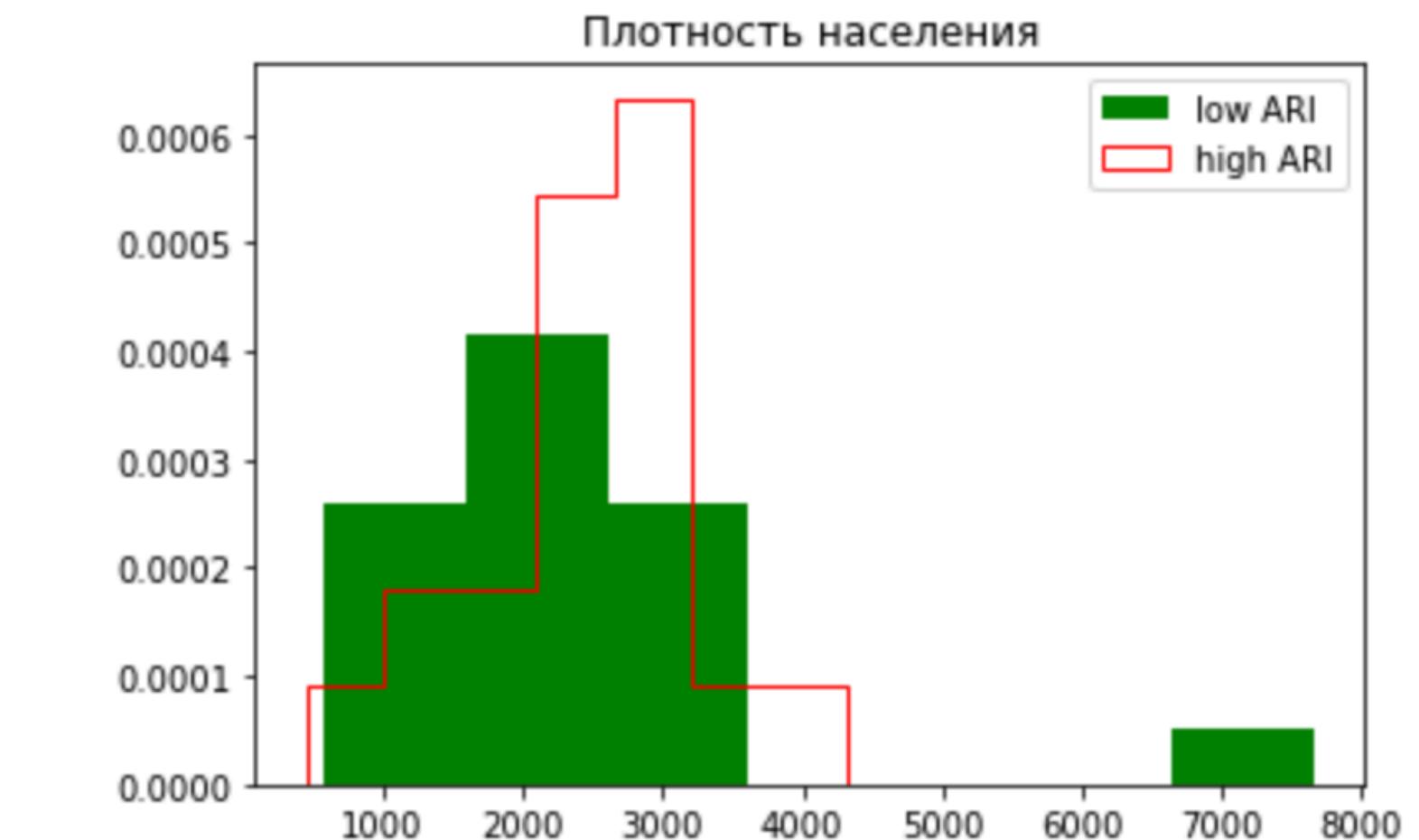
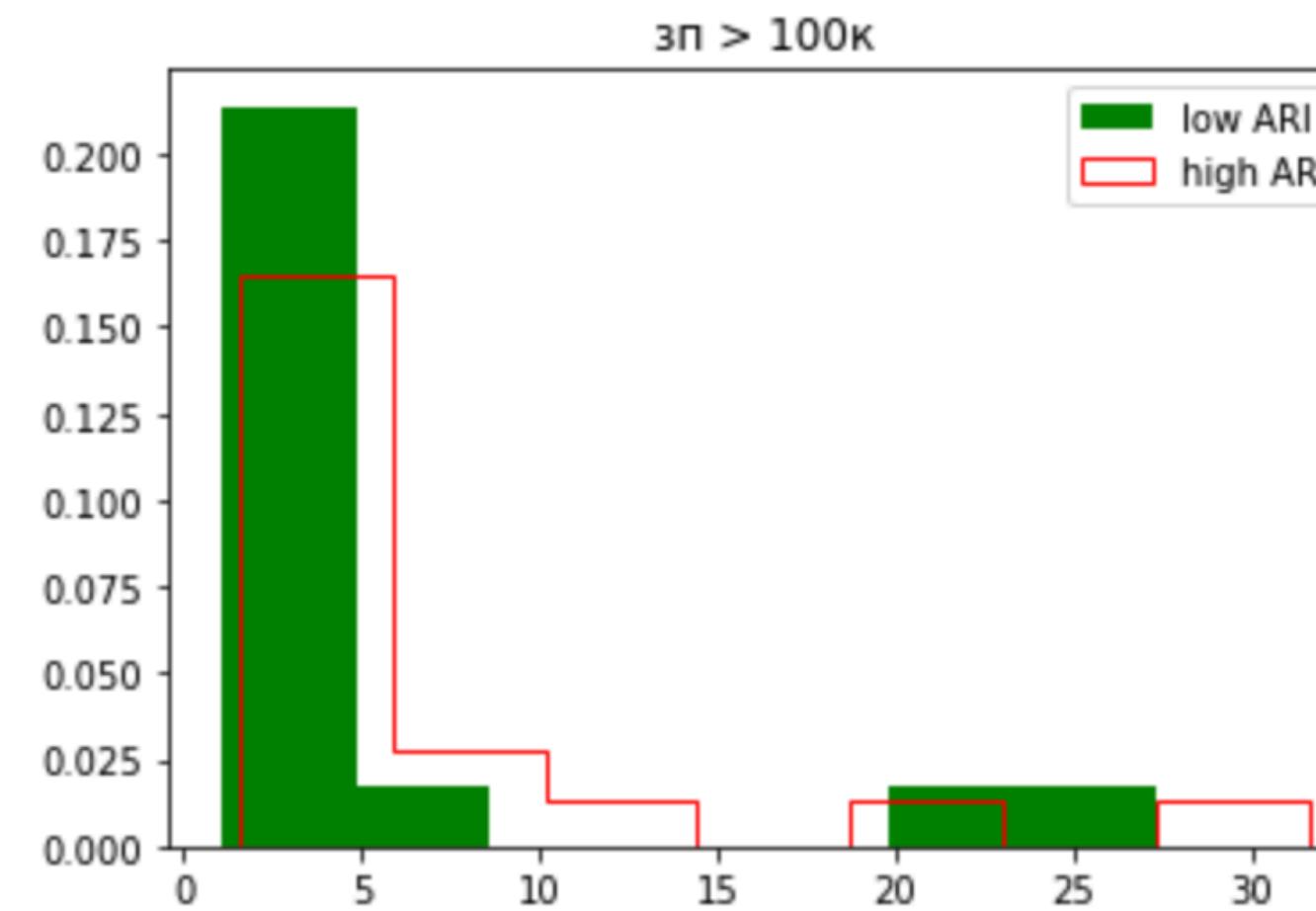
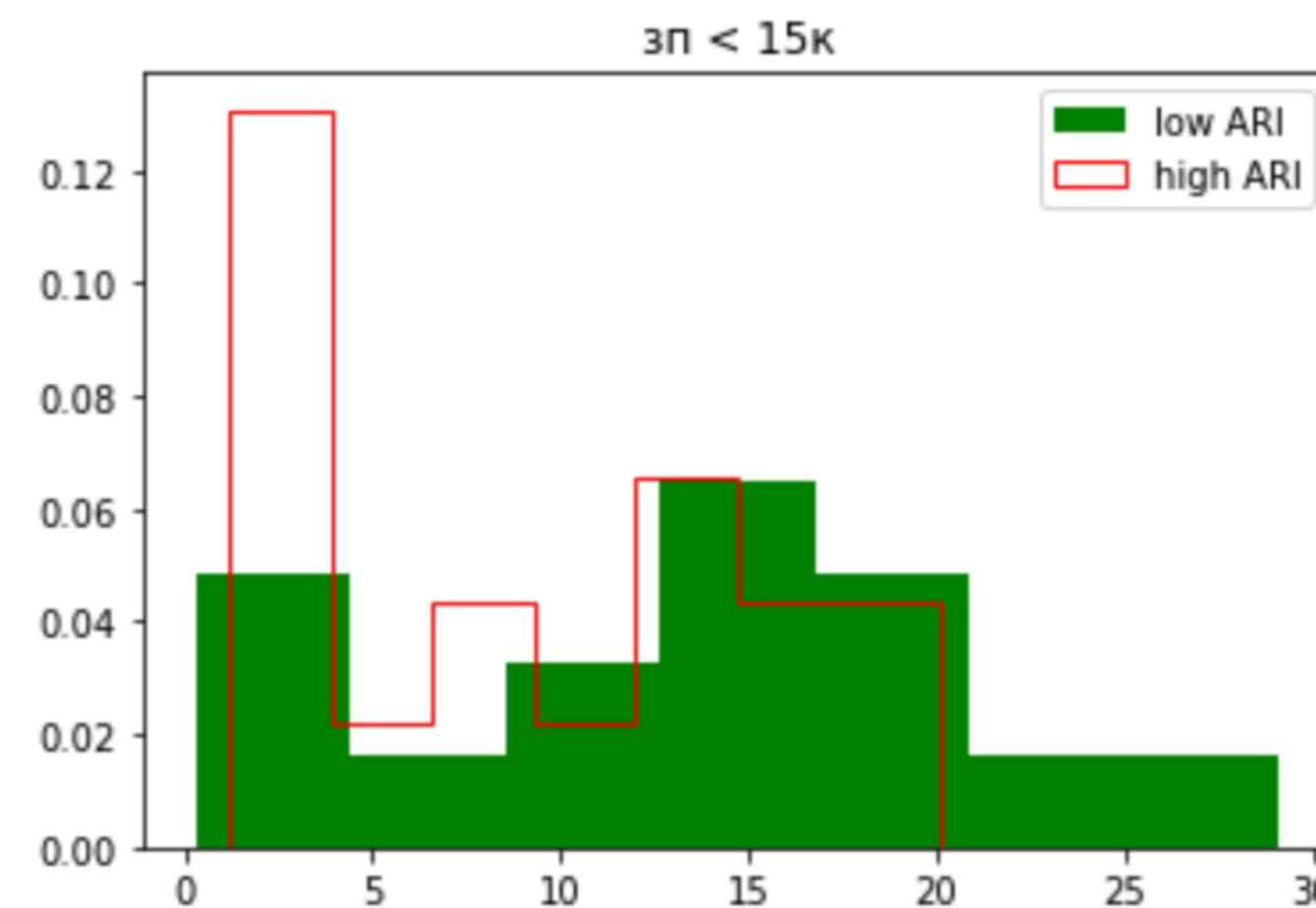
# Визуализация признаков

- Ранжируем города по базовому уровню заболеваемости
- Берем топ-10 городов с самым низким уровнем и топ-10 с высоким
- Для каждого признака строим распределения для выделенных групп городов

	DISTRICT_NAME	NORM_ARI_TOTAL
4	Улан-Удэ	29.440609
6	Чита	31.369864
57	Краснодар	45.523619
14	Пенза	70.763996
3	Петропавловск	147.374441
...	...	...
9	Ижевск	375.871702
44	Владимир	390.558045
32	Барнаул	447.204632
18	Ульяновск	450.761895
29	Сыктывкар	559.384086



# Визуализация признаков





# Интерпретация весов ML моделей

- Модель: линейная регрессия Lasso и Random Forest
- Веса модели позволяют определять значимость признаков
- Целевой признак: базовый уровень заболеваемости (ARI) для разных возрастных групп

# Интерпретация весов Lasso регрессии

0-2 года	3-6 лет	7-14 лет	От 15 лет	Все жители
Процент людей с зп <15k	Процент людей с зп <15k	Вредные привычки	Загрязненность воздуха	Процент людей с зп <15k
Загрязненность воздуха	Загрязненность воздуха	Процент людей с зп <15k	Процент людей с зп <15k	Вредные привычки
Процент людей с зп >100k	Вредные привычки	Температура зимой	Процент людей с зп >100k	Загрязненность воздуха
Температура зимой	Процент людей с зп >100k	Загрязненность воздуха	Вредные привычки	Процент людей с зп >100k
Вредные привычки	Температура зимой	Среднегодовое кол-во осадков	Плотность населения	Температура зимой
Кол-во терапевтов	Кол-во терапевтов	Кол-во педиатров	Среднегодовая температура	Среднегодовое кол-во осадков
Среднегодовая температура	Плотность населения	Кол-во терапевтов	Кол-во терапевтов	Кол-во терапевтов
Кол-во педиатров	Среднегодовое кол-во осадков	Процент людей с зп >100k	Среднегодовое кол-во осадков	Кол-во педиатров
Среднегодовое кол-во осадков	Кол-во педиатров	Температура летом	Процент людей с зп >100k	Температура летом
Плотность населения	Температура летом	Плотность населения	Температура летом	Среднегодовая температура

# Интерпретация Random Forest

<b>0-2 года</b>	<b>3-6 лет</b>	<b>7-14 лет</b>	<b>От 15 лет</b>	<b>Все жители</b>
Температура зимой	Кол-во терапевтов на 100к населения	Качество воздуха	Плотность населения	Качество воздуха
Плотность населения	Кол-во педиатров на 100к населения	Среднегодовое кол-во осадков	Среднегодовое кол-во осадков	Среднегодовое кол-во осадков
Кол-во врачей	Плотность населения	Кол-во врачей	Температура летом	Процент людей с зп < 15к
Процент людей с зп < 15к	Среднегодовая температура	Среднегодовая температура	Качество воздуха	Погодные условия
Температура летом	Качество воздуха	Плотность населения	Температура зимой	Плотность населения
Вредные привычки	Вредные привычки	Температура зимой	Вредные привычки	Кол-во педиатров на 100к населения
Качество воздуха	Среднегодовая температура	Вредные привычки	Кол-во терапевтов	Кол-во терапевтов на 100к населения
Среднегодовое кол-во осадков	Среднегодовое кол-во осадков	Кол-во педиатров	Среднегодовая температура	Вредные привычки
Среднегодовая температура	Процент людей с зп < 15к	Процент людей с зп < 15к	Процент людей с зп >100k	Процент людей с зп >100k
Процент людей с зп >100k	Температура летом	Процент людей с зп >100k	Процент людей с зп < 15к	

# Вывод: Интерпретация весов ML моделей

- Часто в топе фичей видим загрязненность воздуха
- Среднегодовое кол-во осадков. Кажется, что косвенно связанно с влажностью => влияет на распространение
- Плотность населения. Косвенно может влиять на кол-во контактов.
- Также видим количество людей с низкой зп. Косвенно может говорить о размере города либо о невозможности людей уйти на больничный
- Также в топе важный фичей стабильно видим вредные привычки

# Интерпретация результатов

- Качество воздуха влияет на здоровье лёгких граждан и на иммунитет в целом. С другой стороны есть тенденция к тому, что вирус умирает в плохом воздухе быстрее.
- Количество осадков влияет на влажность и на быстроту распространения вируса.
- Плотность населения влияет на количество контактов людей
- Низкая з/п заставляет людей не брать больничные, чтобы побольше работать. Следовательно, болезнь не регистрируется.
- Вредные привычки влияют на иммунитет человека.

# Дальнейшие исследования

- Данные о вакцинации населения
- Разные штаммы гриппа в разных регионах с разной заразностью
- Исследование литературы и научных публикаций с подобными исследованиями за рубежом
- Национальные особенности
- Рацион питания населения
- Роза ветров на местности
- Проверка данных

# Ссылки на данные

- <http://www.statdata.ru/samyi-teplyi-holodnyi-gorod-rossii>
- [http://www.statdata.ru/nasel\\_regions](http://www.statdata.ru/nasel_regions)
- <https://www.data-in.ru/data-catalog/datasets/160/>
- <https://ria.ru/20201123/zarplata-1585826589.html?in=t>
- <https://ria.ru/20201207/vrednye-privychki-1587939088.html>
- [http://voeikovmgo.ru/?option=com\\_content&view=article&id=681&Itemid=236&lang=ru](http://voeikovmgo.ru/?option=com_content&view=article&id=681&Itemid=236&lang=ru)
- <https://minzdrav.gov.ru/ministry/61/22/stranitsa-979/statisticheskie-i-informatsionnye-materialy/statisticheskiy-sbornik-2018-god>



ВСЁ!