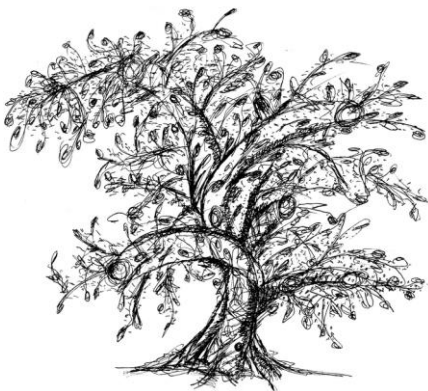


Дерево Глубиной Три

В рамках недельного онлайн-хакатона VSG GAMMA 2021

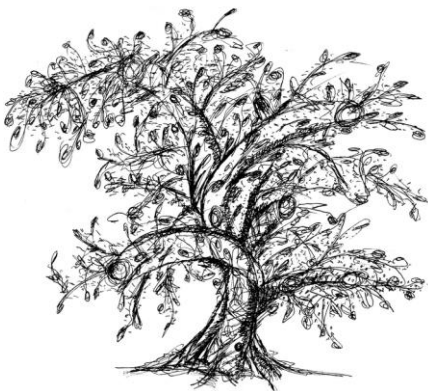
В задаче без рамок от НИИ Гриппа

представляет ...

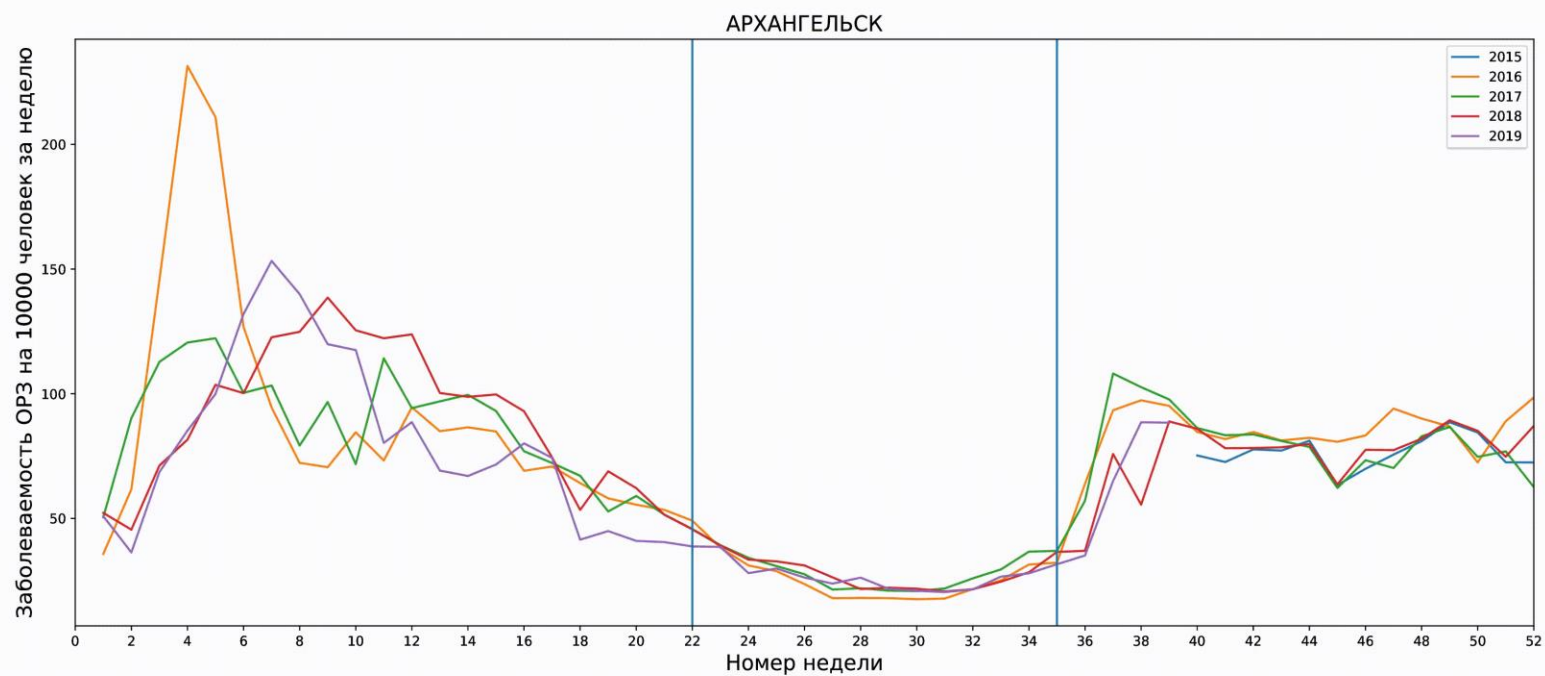


Задача

Предложить гипотезы различной неэпидемической заболеваемости гриппом и ОРВИ в городах РФ

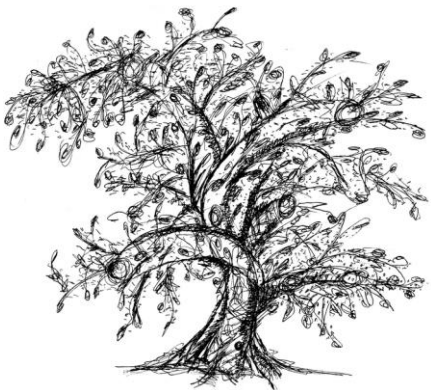


Уровень неэпидемической заболеваемости



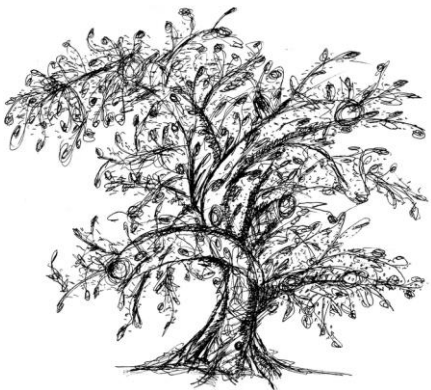
Интерпретация факторов с помощью ML

- ☐ Выделить **факторы**, которые могли бы влиять на неэпидемическую ситуацию
- ☐ **Собрать** необходимые внешние данные
- ☐ Обучение **модели машинного обучения**
- ☐ Оценить важность признаков с помощью **shap-values**
- ☐ Сформулировать **гипотезы**



Собранные внешние данные по городам

- ☐ Погодные
 - ☐ Температура
 - ☐ Давление
 - ☐ Влажность
 - ☐ Скорость ветра



Собранные внешние данные по городам

☐ Погодные

☐ Температура

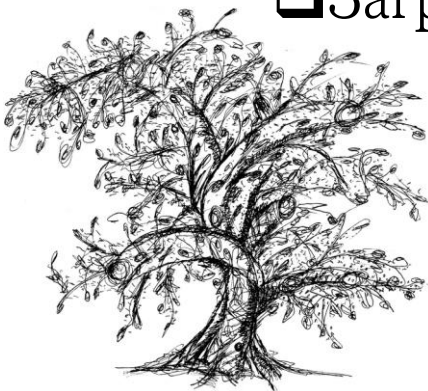
☐ Давление

☐ Влажность

☐ Скорость ветра

☐ Экологические

☐ Загрязненность воздуха



Собранные внешние данные по городам

☐ Погодные

☐ Температура

☐ Давление

☐ Влажность

☐ Скорость ветра

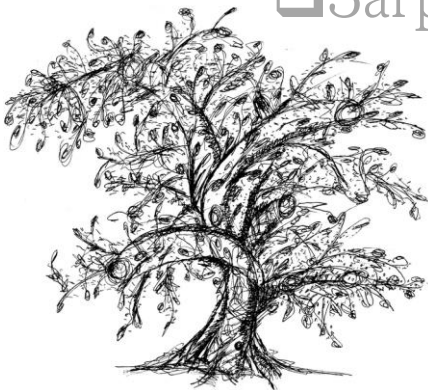
☐ Демографические

☐ Плотность населения

☐ Индекс счастья

☐ Экологические

☐ Загрязненность воздуха



Собранные внешние данные по городам

☐ Погодные

☐ Температура

☐ Давление

☐ Влажность

☐ Скорость ветра

☐ Экологические

☐ Загрязненность воздуха

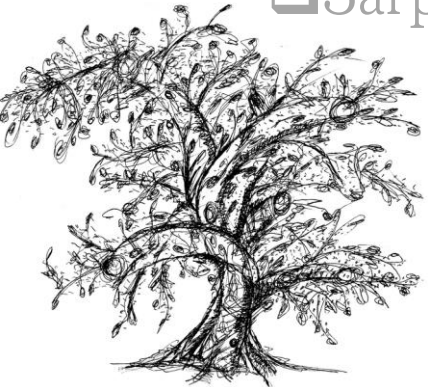
☐ Демографические

☐ Плотность населения

☐ Индекс счастья

☐ Географические

☐ Наличие выхода к морю



Собранные внешние данные по городам

☐ Погодные

☐ Температура

☐ Давление

☐ Влажность

☐ Скорость ветра

☐ Экологические

☐ Загрязненность воздуха

☐ Демографические

☐ Плотность населения

☐ Индекс счастья

☐ Географические

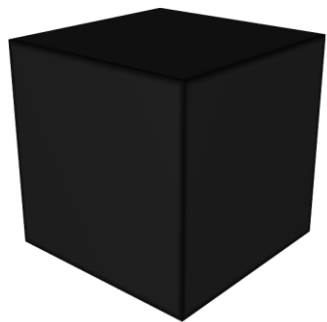
☐ Наличие выхода к морю

☐ Инфраструктура

☐ Наличие выхода к метро

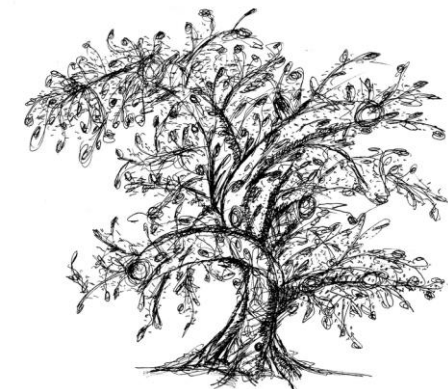


SHAP-values?

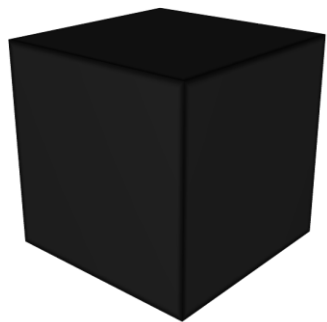


+

Ничего

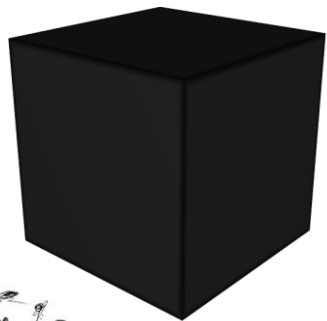


SHAP-values?



+

Ничего



+

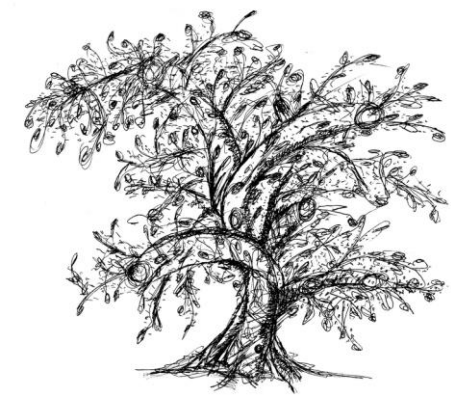
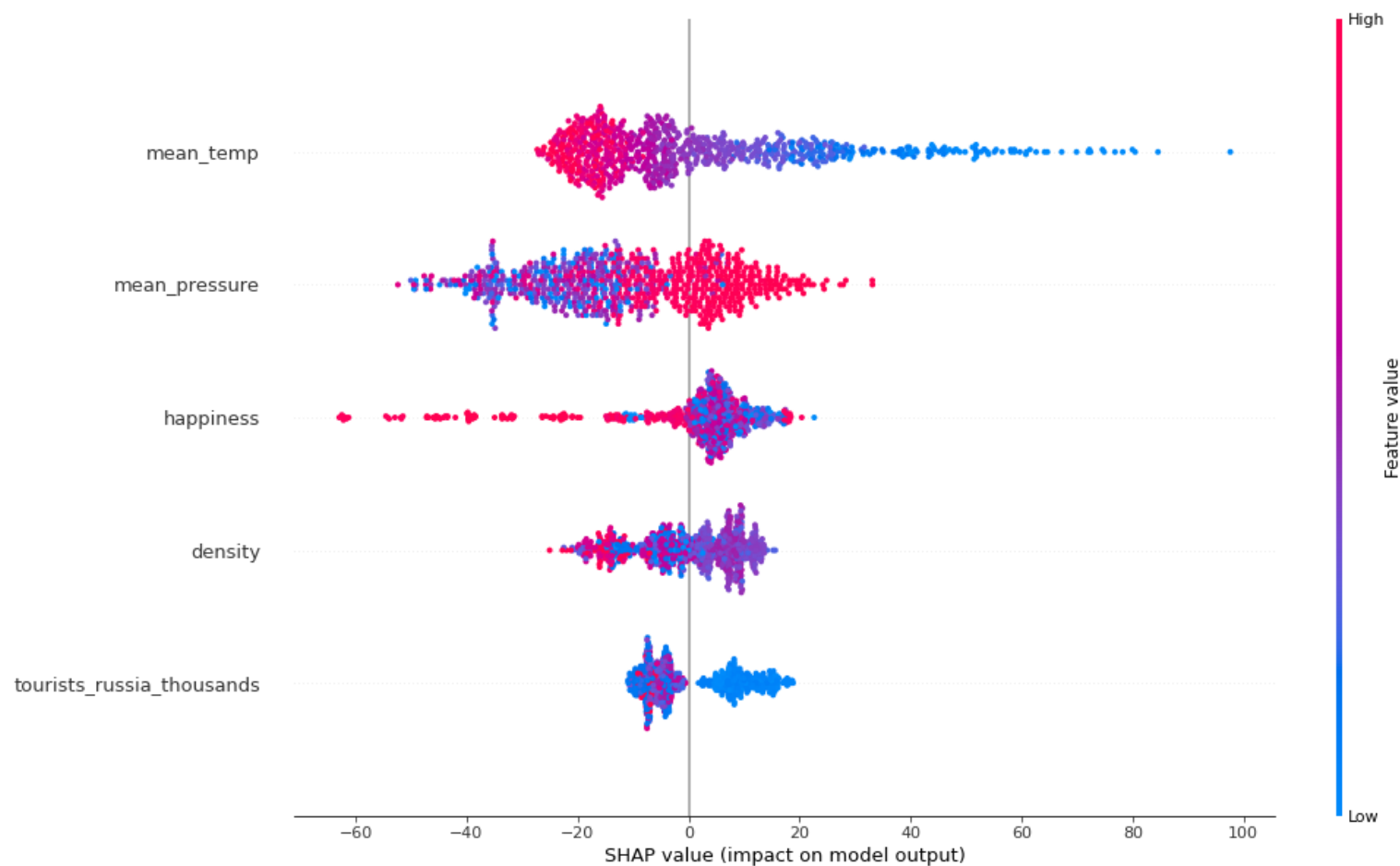
SHAP-values

Интерпретация работы модели
Интерпретация отдельного
предсказания



SHAP-values!

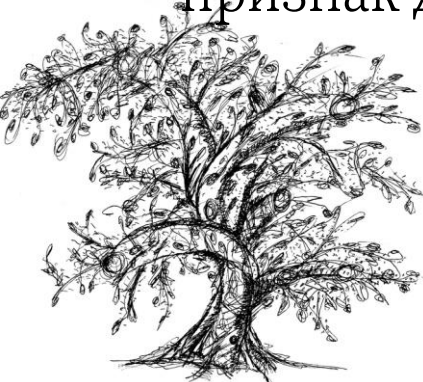
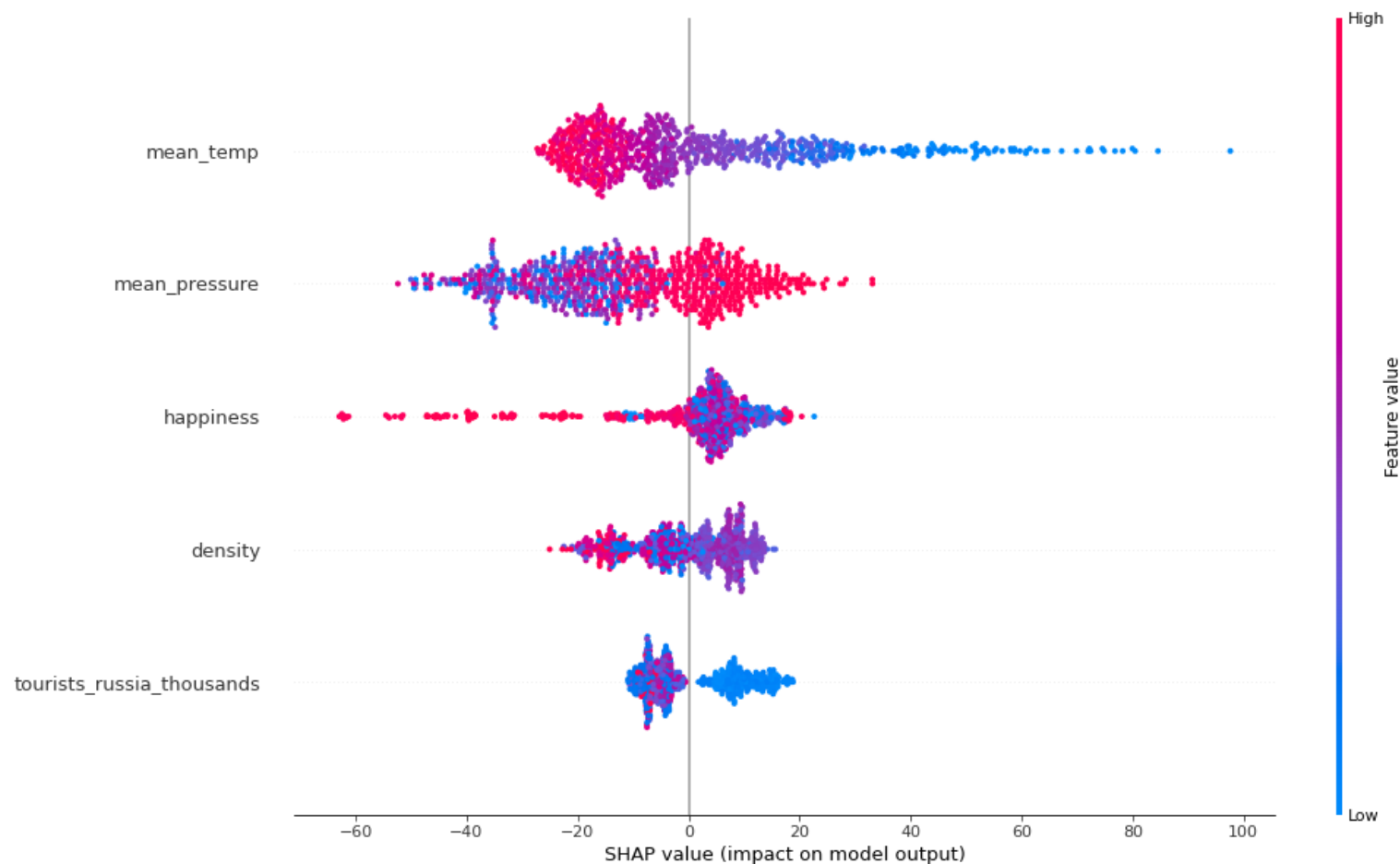
Объяснение модели



SHAP-values!

Объяснение модели

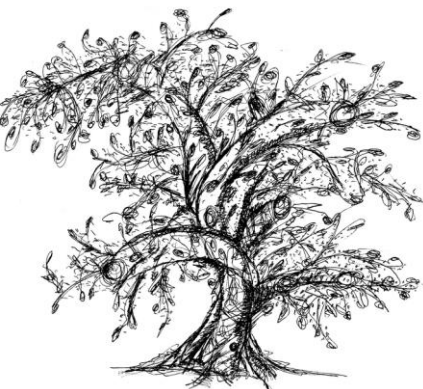
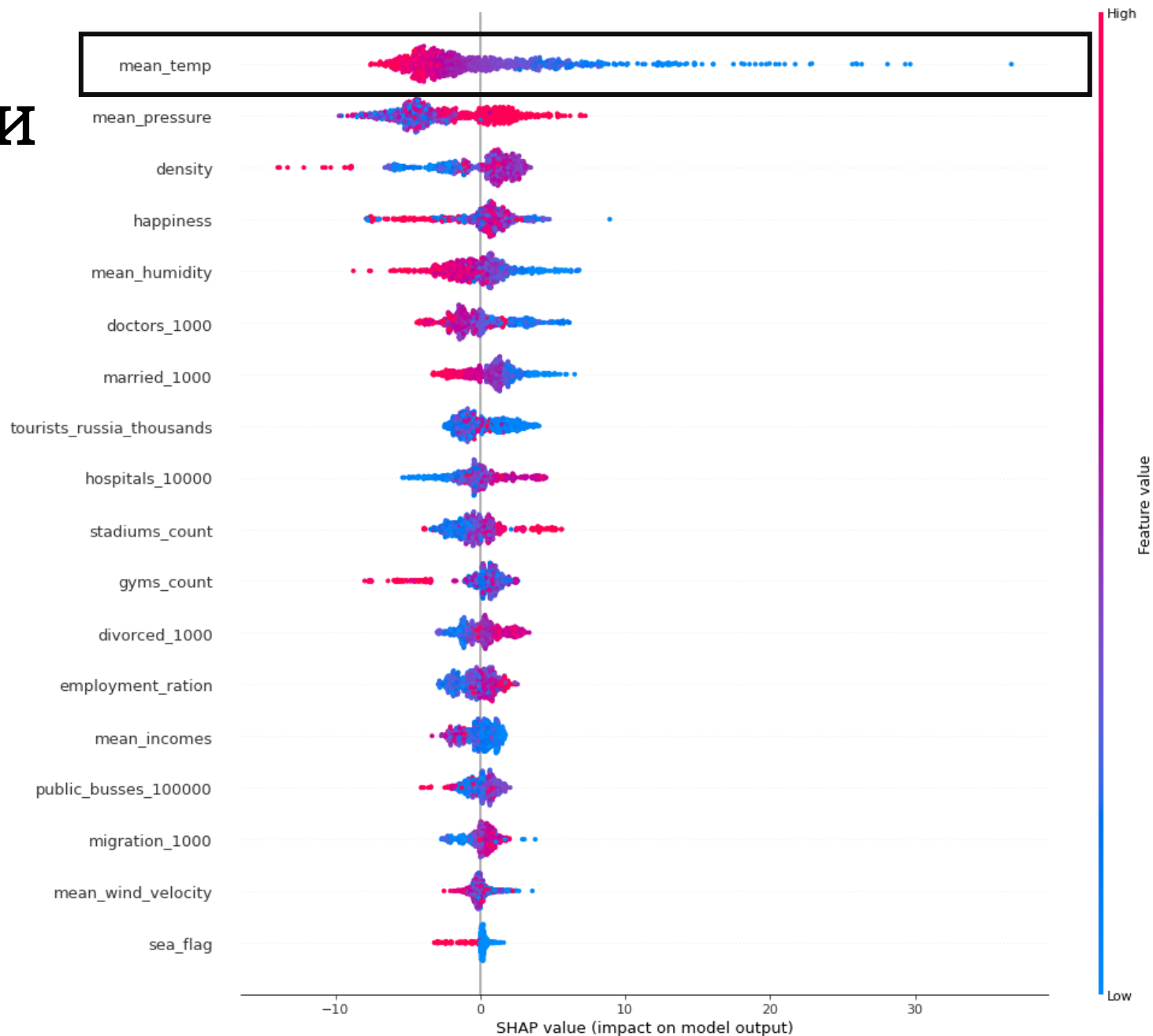
- ❑ Точка - объект
- ❑ Цвет точки отражает **значение** признака. У красной точки значение признака больше.
- ❑ Чем **больше** значение SHAP value для точки, тем **большее влияние** оказывает данный признак для этого объекта



SHAP-values!

Объяснение модели

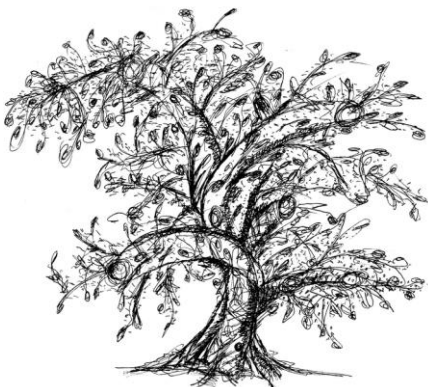
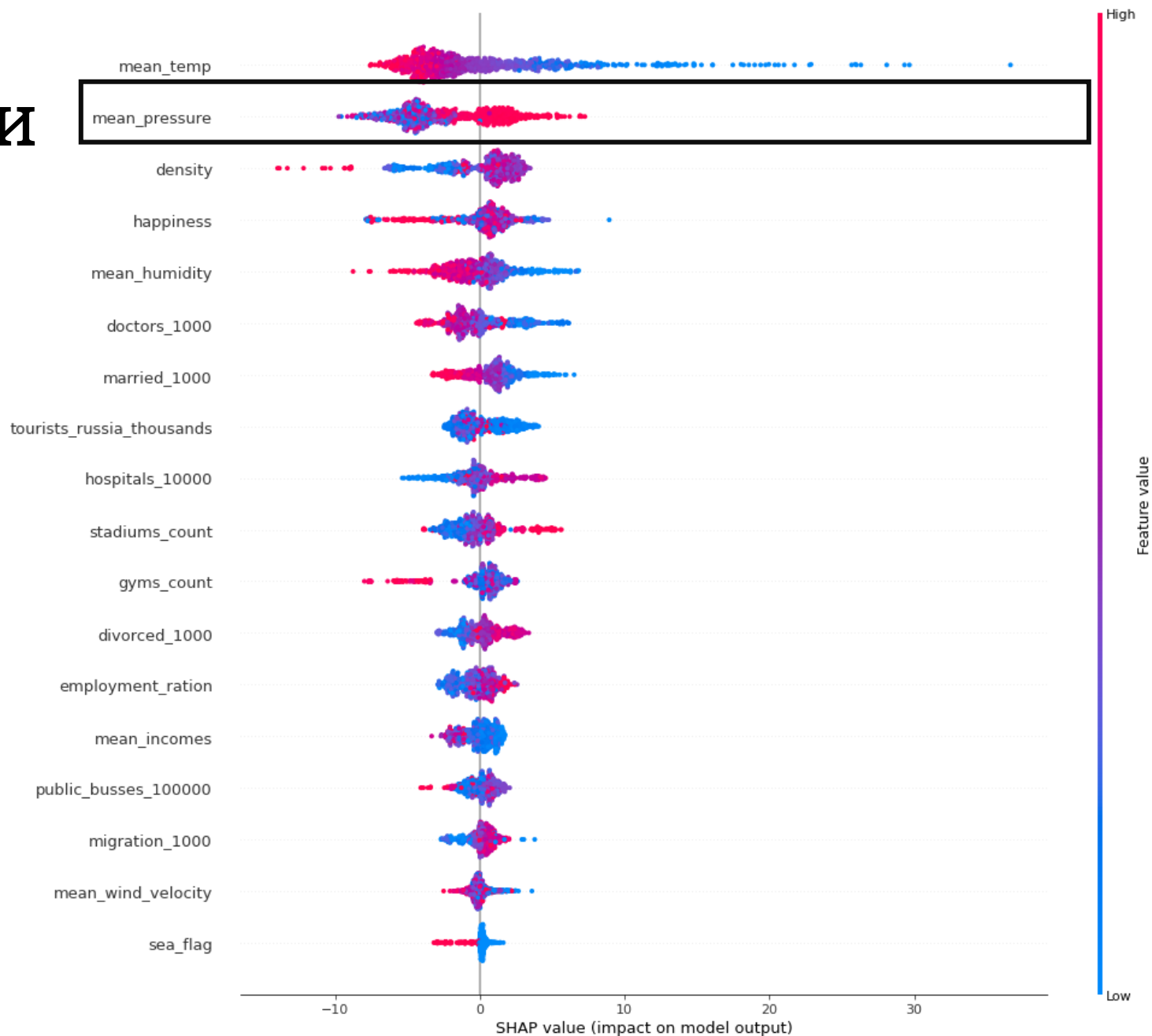
☐ Средняя температура



SHAP-values!

Объяснение модели

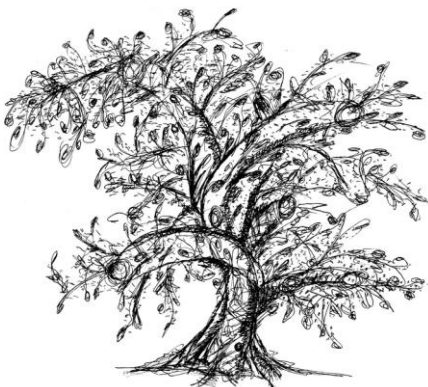
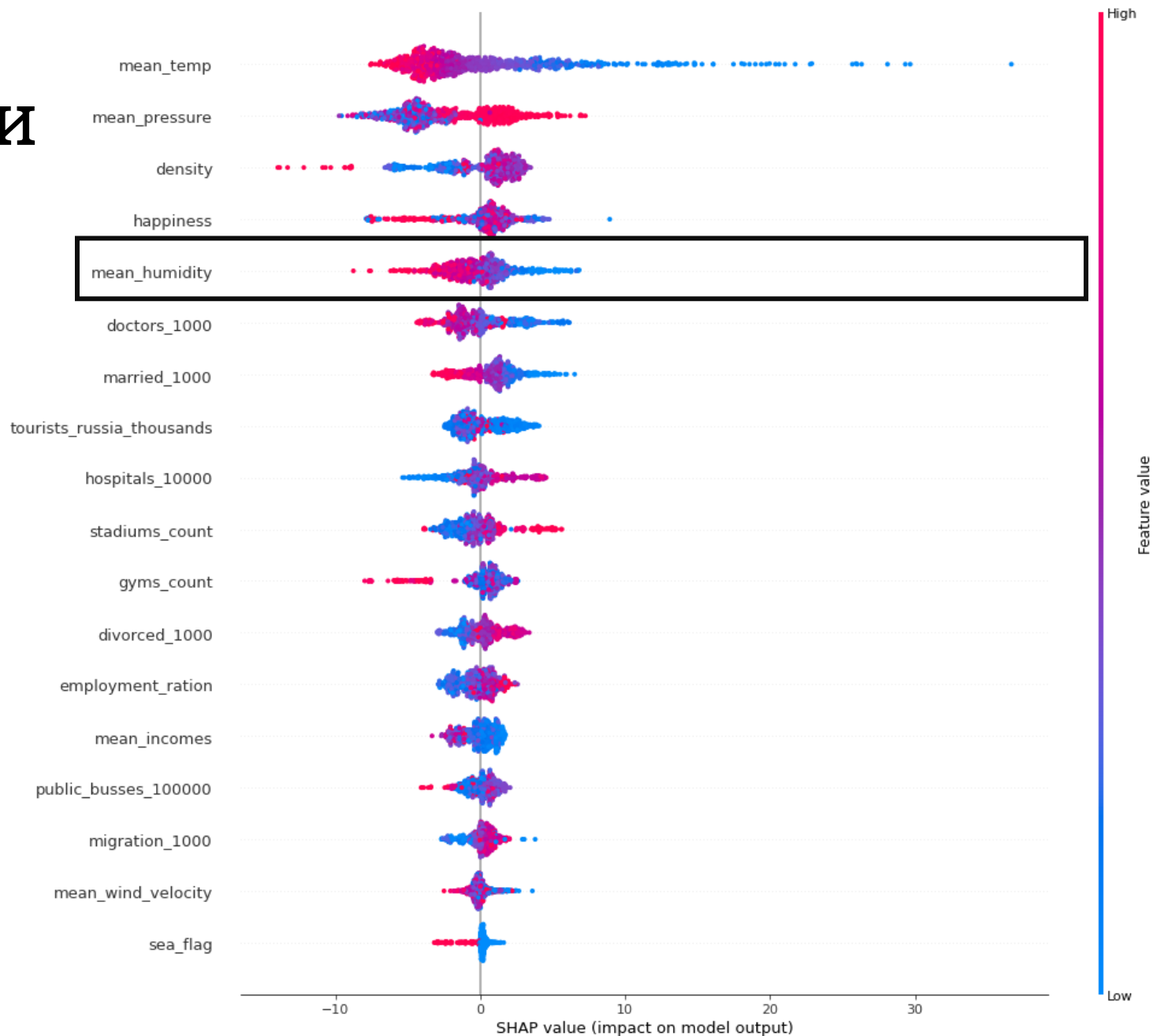
☐ Среднее давление
окружающей среды



SHAP-values!

Объяснение модели

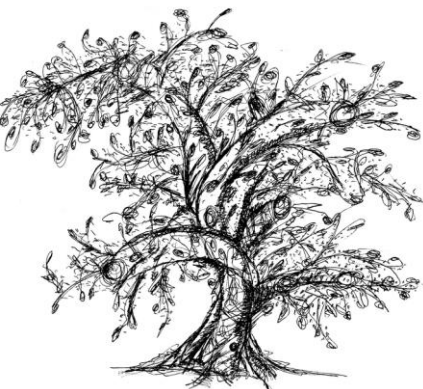
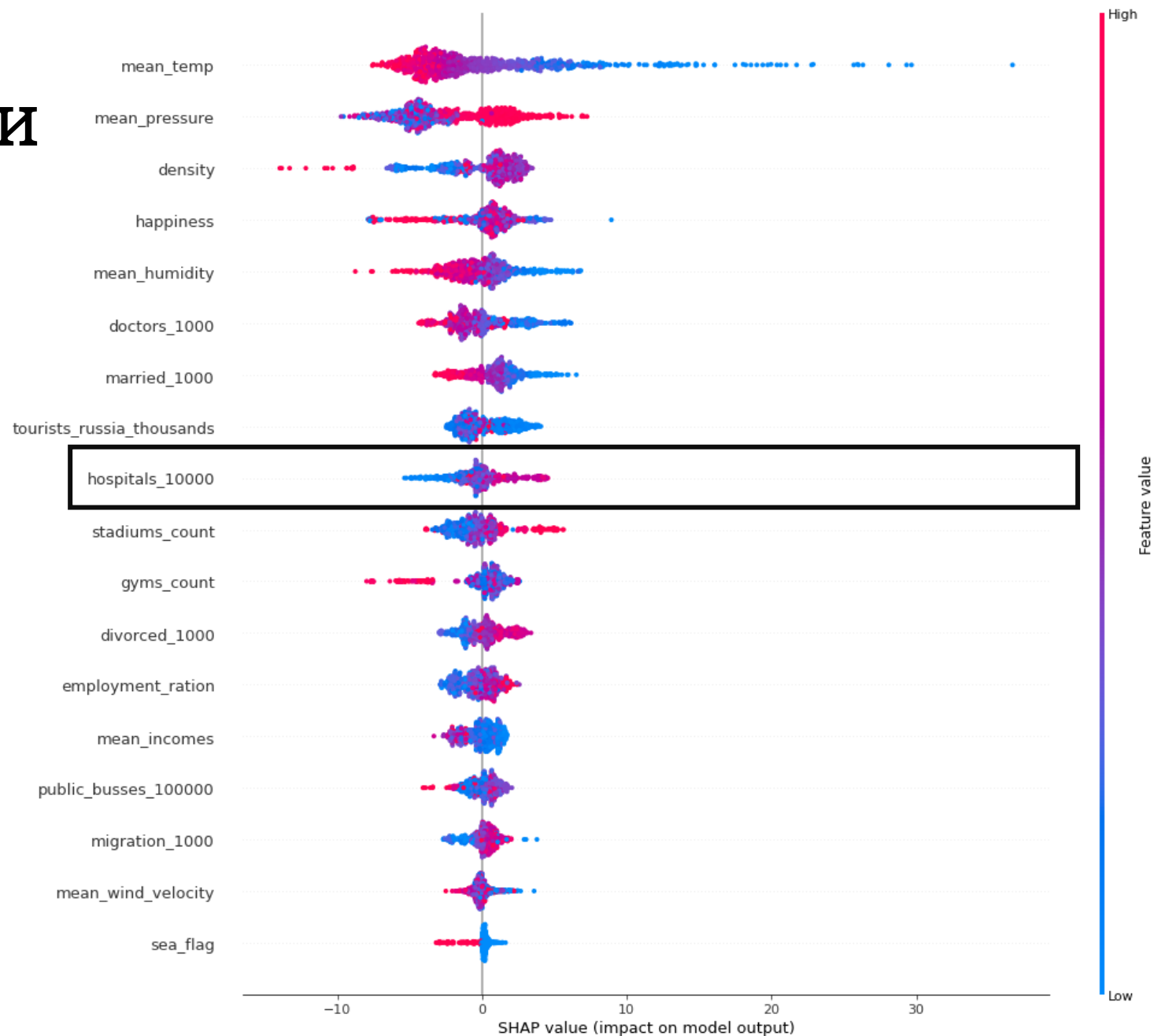
☐ Средняя влажность



SHAP-values!

Объяснение модели

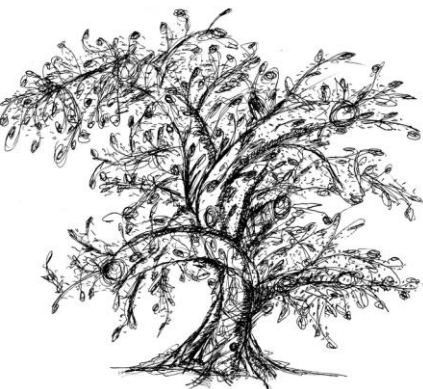
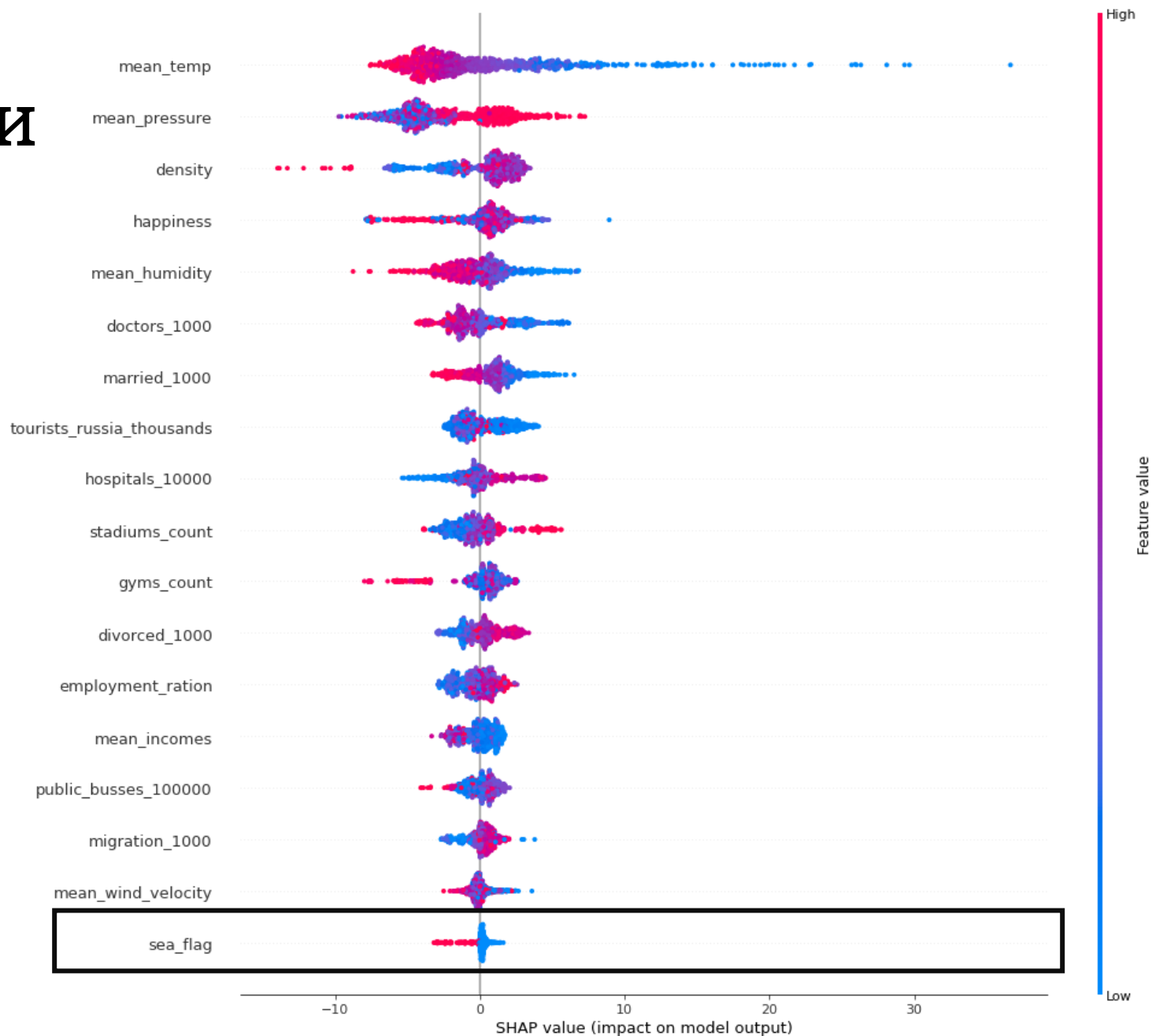
□ Количество больниц на 1000
людей



SHAP-values!

Объяснение модели

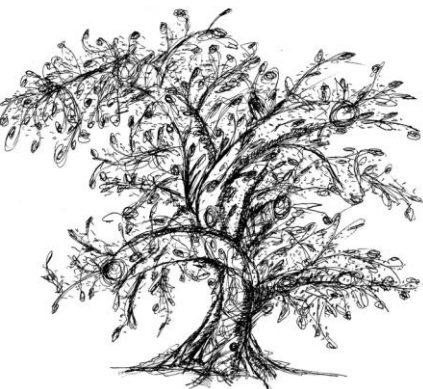
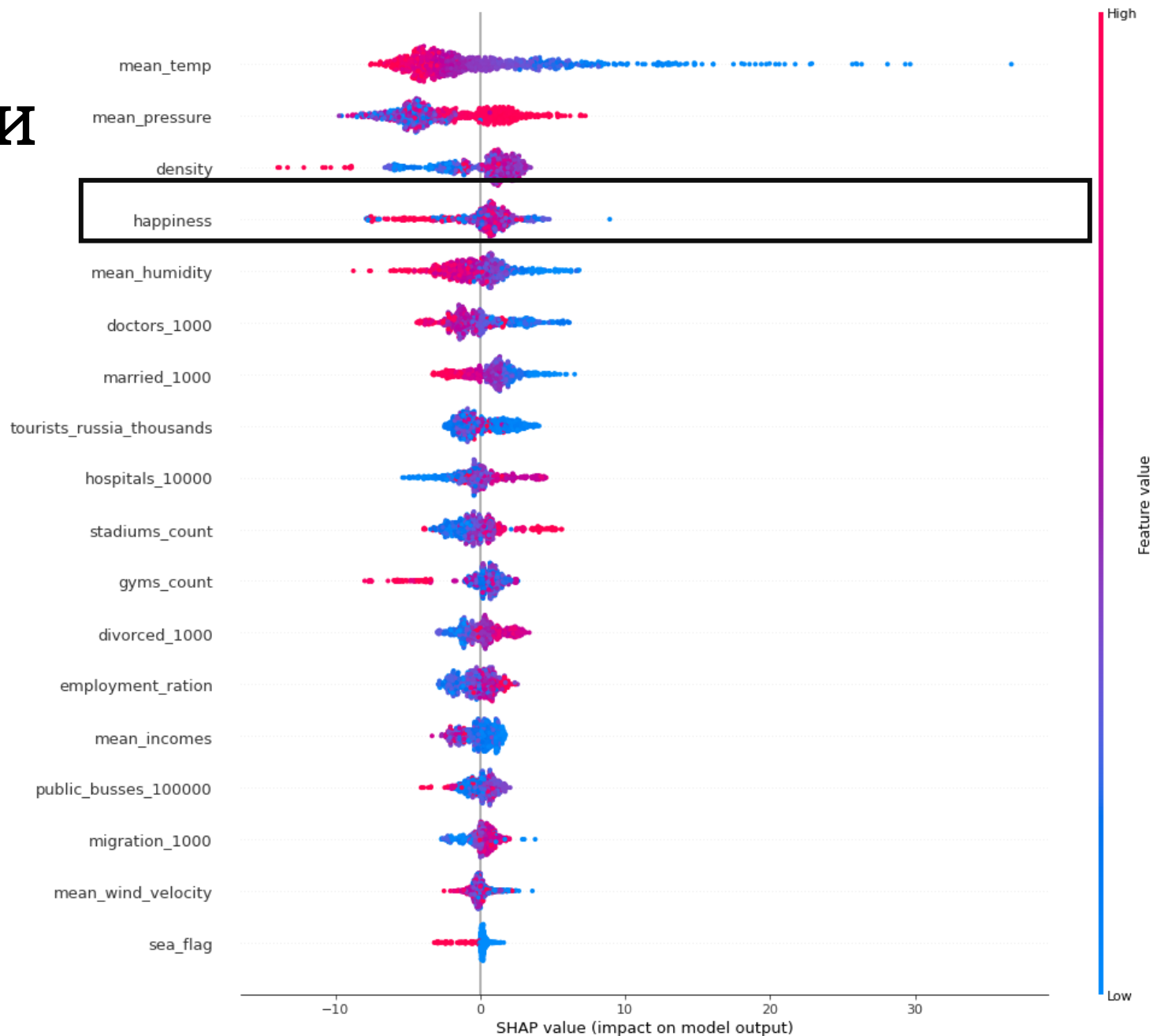
- ❑ Бинарный признак.
Наличие выхода в море



SHAP-values!

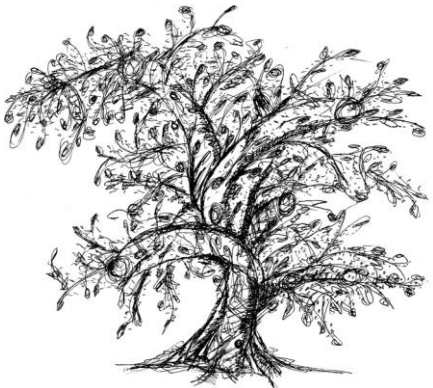
Объяснение модели

☐ Индекс счастья



Предлагаемые гипотезы

- ☐ В тех городах, где живут более **счастливые** люди, заболеваемость **ниже**
- ☐ В тех городах, где **температура** летом выше, заболеваемость ниже
- ☐ В тех городах, где **атмосферное давление** выше, заболеваемость **выше**
- ☐ В тех городах, где есть **море**, заболеваемость ниже



Факторы для дальнейших исследований

☐ Географические

- ☐ Удаленность от Китая

☐ Инфраструктурные

- ☐ Количество заводов

☐ Экологические

- ☐ Прошлые катастрофы

☐ Демографические

- ☐ Доход

- ☐ Уровень образования

- ☐ Уровень безработицы

- ☐ Уровень разводов

- ☐ Уровень алкоголизма

- ☐ Индекс полноты семьи

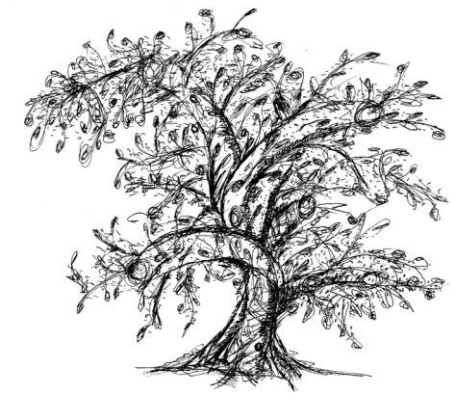
- ☐ Распределение по гендеру



Послесловие

Петропавловск в Казахстане

Петропавловск-Камчатский в России ...



СПАСИБО ЗА ВНИМАНИЕ!

