

3. Прогнозирование временных рядов на карте

На этой неделе вам предстоит освоить работу с геоданными, научиться рисовать объекты на статистических и динамических картах, а также понять, для каких прямоугольных районов прогнозировать спрос на такси на самом деле нужно.

Задание, оцениваемое сокурсниками: Работа с геоданными

Срок сдачи прошел июнь 4, 11:59 вечера PDT

Отправить сейчас

Выполненное задание необходимо отправить как можно раньше, чтобы сокурсники могли оценить вашу работу. В противном случае может не оказаться достаточного количества сокурсников для его оценки. Сдавайте работы как можно раньше!

1. [Инструкции](#)
2. [Моя работа](#)
3. [Обсуждения](#)

Инструкции

В этом задании вам предстоит научиться работать с географическими данными. В Питоне есть огромное количество библиотек, позволяющих эти данные обрабатывать и визуализировать; выберите ту, которая вам больше понравится. Вам может помочь руководство kaggle по работе с картами: "[Семнадцать способов рисовать данные на картах](#)". Смотрите также:

- модуль [basemap](#) ([пример](#), [ещё пример](#))
- плагин [gmaps](#)
- виджет [ipyleaflet](#) (крайне минималистичный [пример](#))
- модуль [folium](#)

Поскольку мы агрегировали данные по равномерной сетке, можно ожидать, что существенное количество ячеек окажется в районах города, где вызов такси невозможен (реки, заливы, острова, парки). Теоретически ряды, соответствующие таким ячейкам, должны состоять из одних нулей, но на практике это может быть не так: устройства, определяющие координаты автомобилей в момент начала поездки, несовершенны и могут ошибаться, показывая, например, что такси находится в середине реки. Чтобы правильно отобрать ячейки, вам понадобится визуализировать их границы на карте и отфильтровать те, из которых поездки на такси невозможны.

Чтобы сдать задание, выполните следующую последовательность действий.

1. Загрузите агрегированные данные о поездках в мае 2016. Просуммируйте общее количество поездок такси из каждой географической зоны и посчитайте количество ячеек, из которых в мае не было совершено ни одной поездки.
2. Нарисуйте статическую карту Нью-Йорка. Поставьте на карте точку там, где находится Эмпайр-Стейт-Билдинг.
3. Поверх статической карты Нью-Йорка визуализируйте данные о поездках из каждой ячейки так, чтобы цветовая шкала, в которую вы окрашиваете каждую ячейку, показывала суммарное количество поездок такси из неё.

4. Вставьте интерактивную карту Нью-Йорка — такую, которую можно прокручивать и увеличивать. Поставьте метку там, где находится статуя свободы.
5. Нарисуйте на интерактивной карте Нью-Йорка ячейки так, чтобы их цвет показывал среднее за месяц количество поездок такси в час из этой зоны.
6. Чтобы не выбирать из всех 2500 ячеек вручную, отфильтруйте ячейки, из которых в мае совершается в среднем меньше 5 поездок в час. Посчитайте количество оставшихся. Проверьте на карте, что среди этих ячеек нет таких, из которых поездки на самом деле невозможны.
7. Сохраните ноутбук в формате html, запакуйте ipynb и html в один архив и загрузите его в форму (html облегчит задачу проверки вашим рецензентам — в ipynb динамические карты часто не сохраняются, а для выполнения кода рецензентам придётся устанавливать все библиотеки).

Review criteriaменьше

Задание будет оцениваться по факту выполнения каждого из его пунктов в приложенном ноутбуке.