

Влад, добрый день

После перерыва возвращаюсь с новым дз к 18 модулю (отправил через Git).

Пришлось повозиться с 8 задачами.

1. С приведением к типу `datetime64` через 2 параметра `parse_dates` и `date_parser` (требующим лямбду в аргументе) я кое-как разобрался. Здесь только хотелось посетовать, что в документации `pandas` очень не хватает явных примеров, чтобы не тыкать синтаксис наугад и не шерстить `StackOverflow`. После чтения документации всегда один эффект: «так, все понятно, осталось узнать, как это пишется в реальном коде». Может, это пока по неопытности. Вот есть же статья «`pandas.read_csv`» с 52 параметрами. Вот почему не дать 1 пример, где они использованы все? Это так, риторически)

2. Как решается вопрос с приведение колонки «year» к типу `int`, допускающему пропуски, я так и не понял. То, что в коде сработало `dtype = {'year': 'Int64'}` – это результат чтения того же `StackOverflow` и антинаучного тыка. В задании была рекомендация использовать `arrays.IntegerArray`. Однако, в документации четко прописано, что при создании `pd.array` требуется задать конечное число возможных значений (e.g.: `[1, 2, None]`), тогда как мне под задачу нужна целочисленная переменная (которая будет принимать подтягиваемые из csv. значения) и плюс возможность для `NaN`. Может, здесь тоже выручает лямбда? Меня постигла удача с `Int64`, так что дальше ломать код не стал. Будет здорово, если поделитесь академичным решением, или советом.

Спасибо!

С уважением,

Иван Лыдкин

11 ноября, 23:30



Иван, здравствуйте!

С возвращением)

По заданиям:

✅ Задание #1: датасет загружен верно;

- ✓ Задание #2: пропуски в годе выпуска автомобиля обработаны;
- ✓ Задание #3: обработка пропусков в типе топлива проведена успешно;
- ✓ Задание #4: пропуски в модели автомобиля заполнены;
- ✓ Задание #5: пропуски в статусе автомобиля исправлены;
- ✓ Задание #6: данные в годе выпуска преобразованы;
- ✓ Задание #7: преобразование типа данных в категориальных переменных проведено успешно;
- ✓ Задание #8: задание принято;

Дополнительные задачи:

- ✓ Задание #9: задание выполнено

Иван, поздравляю с успешно сделанным домашним заданием!)

По вопросам:

- 1) Недоработка документации, такое бывает)
- 2) Тут можно было сделать так:

```
df1 = pd.read_csv('data/vehicles_dataset.csv',  
  
                  usecols=['price', 'year', 'posting_date'],  
  
                  parse_dates=['posting_date'],  
  
                  dtype = {'price': float, 'year': 'Int64'})  
  
df1.head()
```