



Предположим, у нас есть такие данные:

First Name	Last Name	Phone	Salary
Jessica	Higgins	999-9999-99	6604
Owen	null	144-6330-95	2882
Samantha	Brown	654-1875-50	9362
null	Hunt	505-4760-87	4823
Camila	Evans	999-9999-99	7053

Мы хотим:

- Пустые значения для фамилий заменить на "unknown"
- Номера телефонов "999-9999-99" заменить на private
- Выкинуть строки с записями, для которых неизвестно имя

Какие из предложенных вариантов кода дадут верный результат?

a)

```
df.na.drop(subset="First Name") \
.na.fill({"Last Name": "unknown"}) \
.na.replace("999-9999-99", "private")
```

b)

```
df.withColumn(
  "last_name",
  when(col("Last Name").isNull(), "unknown").otherwise(col("Last Name"))
).drop("Last Name") \
.filter(df["First Name"].isNotNull()) \
.withColumn(
  "phone",
  when(col("Phone") == "999-9999-99", "private").otherwise(col("Phone"))
)
```

c)

```
df.rdd.filter(
  lambda r: r[0] is not None
).map(
  lambda r: (
    r[0],
    r[1] or "unknown",
    "private" if r[2] == "999-9999-99" else r[2],
    r[3])
).toDF()
```

Выберите ответ(ы) из списка

- a
- b
- c



▶ Запустить стенд



Дедлайн 07 июля, 23:59 Мск

