## **4. Переименование столбцов**

Метод info() возвращает названия столбцов таблицы и сведения о типах данных в ней. Итак, наши проблемы:

* В начале названия одного столбца явно есть несколько лишних пробелов. От них нужно избавиться.
* Название из двух слов содержит пробел, который необходимо заменить на символ нижнего подчёркивания.
* Названия должны быть на одном языке и набраны в одном регистре, чтобы не заучивать уникальный формат для каждого столбца.
* Каждый столбец содержит определенный признак — крайне желательно, чтобы название столбца отражало в краткой форме, какая информация в нём содержится.

Вернёмся к таблице с расстояниями (в миллионах километров) от Земли до небесных тел.

measurements = [['Солнце',146,152], *# Измерения хранятся в списке списков* ['Луна',0.36, 0.41], *# measurements (англ. measurement, «измерение»).* ['Меркурий',82, 217], ['Венера',38, 261], ['Марс',56,401], ['Юпитер',588, 968], ['Сатурн',1195, 1660], ['Уран',2750, 3150], ['Нептун', 4300, 4700], ['Комета Галлея', 6, 5400]] *# Названия столбцов хранятся в переменной header.* header = ['Небесные тела ','MIN', 'MAX'] *# Сохраним структуру данных в переменной celestial (англ. celestial, «небесный»).* celestial = pd.DataFrame(data = measurements, columns = header)

Чтобы как следует разглядеть названия столбцов без отвлекающей информации, удобно запросить значения атрибута columns:

print(celestial.columns)

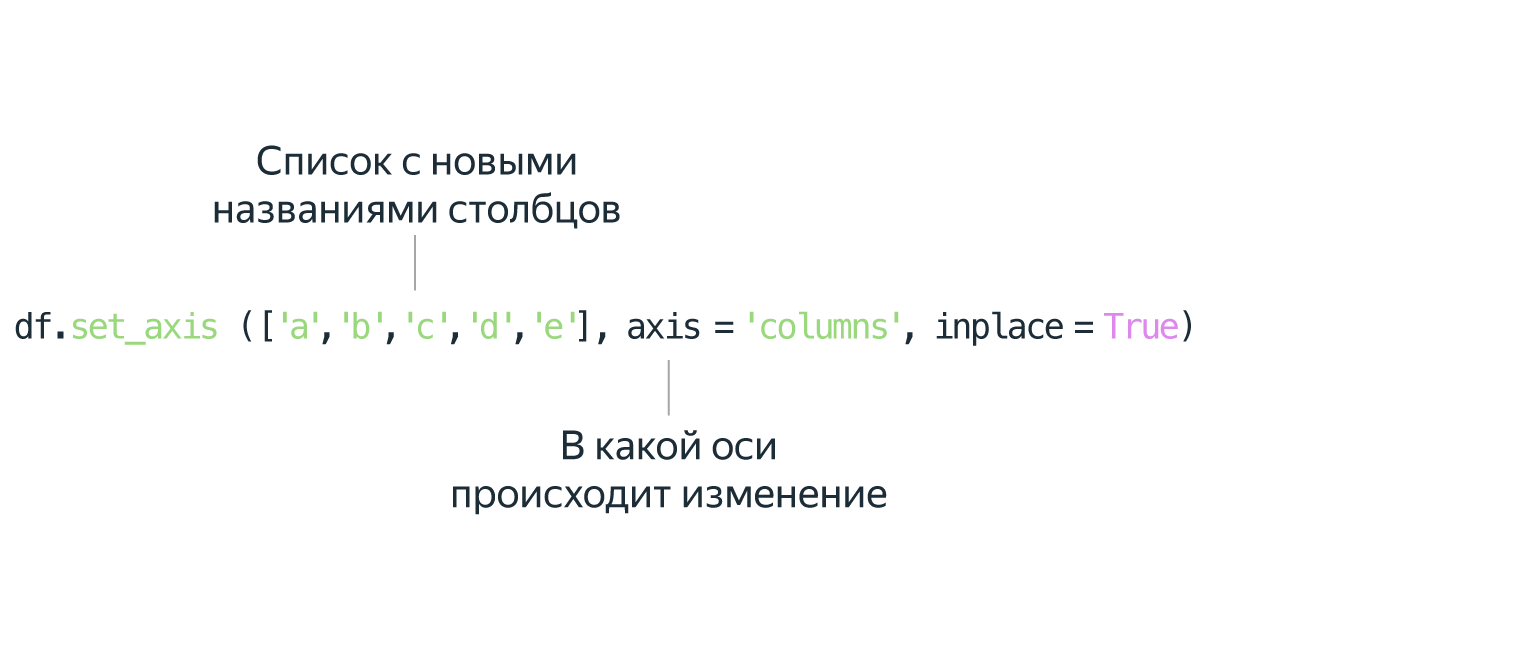
Index(['Небесные тела ', 'MIN', 'MAX'], dtype='object')

Заметно, что столбцы называются на разных языках. Название *Небесные тела* содержит опасный пробел в конце, и слова этого названия разделены пробелом.

Символы за пределами базовой латиницы — недруги аналитика; от них надо избавляться в первую очередь. Правильно будет переименовать «Небесные тела» в *celestial\_bodies*. *MIN* и *MAX* уже написаны латиницей, но эти сокращения мало говорят о том, что за цифры записаны в соотвествующих столбцах. Хорошее решение — назвать их *min\_distance* и *max\_distance* (англ. distance, «расстояние»). Так сразу понятно, что это минимальные и максимальные расстояния.

Чтобы изменить названия столбцов, воспользуйтесь методом **set\_axis()** (англ. set axis, «указать ось»). Он принимает три аргумента:

* cписок с новыми названиями столбцов;
* **axis** (англ. *axis*, «ось») — ось, которой новые названия присваиваются: 'index', если они даются строкам, и 'columns', если это список новых названий столбцов;
* **inplace** (от англ. *in place*, «на месте»). Принимает значения True либо False. В первом случае метод set\_axis() перестраивает структуру данных так, что она замещает прежнюю в переменной с тем же именем.



Заменим названия столбцов таблицы с небесными телами.

celestial.set\_axis(['celestial\_bodies','min\_distance','max\_distance'], axis = 'columns', inplace = True)

Проверим результат:

print(celestial.columns)

Index(['celestial\_bodies', 'min\_distance', 'max\_distance'], dtype='object')

print(celestial)

celestial\_bodies min\_distance max\_distance

0 Солнце 146.00 152.00

1 Луна 0.36 0.41

2 Меркурий 82.00 217.00

3 Венера 38.00 261.00

4 Марс 56.00 401.00

5 Юпитер 588.00 968.00

6 Сатурн 1195.00 1660.00

7 Уран 2750.00 3150.00

8 Нептун 4300.00 4700.00

9 Комета Галлея 6.00 5400.00

Готово! Теперь эта таблица годится для анализа. Давайте проверим названия столбцов в данных Яндекс.Музыки.