

## Категоризация данных

**Категоризация** - объединение данных в категории. Категоризация может проводиться по данным:

- Из одного столбца. В таком случае метод категоризации принимает только одно значение — значение соответствующего столбца.
- Из нескольких столбцов. В таком случае метод категоризации принимает строку из датафрейма целиком.

### Функция для категоризации по числовым диапазонам

```
# функция age, в зависимости от переданного возраста, показывает категорию граждан
def age_group(age):
    if age <= 18:
        return 'дети'
    if age <= 64:
        return 'взрослые'
    return 'пенсионеры'

print(age_group(14)) # выдаст значение 'дети'
```

### Категоризация на основе одного столбца

```
# создаём новый столбец, в который попадут категории по возрастам
# в зависимости от числового значения в столбце 'age'
clients['age_group'] = clients['age'].apply(age_group)
```

## Категоризация данных

### Категоризация на основе нескольких столбцов датафрейма

```
def age_group_unemployed(row):  
    age = row['age']  
    unemployed = row['unemployed']  
  
    if age <= 18:  
        return 'дети'  
  
    if age <= 64:  
        if unemployed == 1:  
            return 'безработные'  
  
        return 'занятые'  
  
    return 'пенсионеры'  
  
# создаём новый столбец, в который попадёт одна из четырёх категорий  
# в зависимости от значений столбцов 'age' и 'unemployed'  
clients['age_group'] = clients.apply(age_group_unemployed, axis=1)
```