

# 1.4 категоризация

---

- 1 Выделять из данных словарь категорий;
  - 2 Категоризация числового признака
  - 3 Функции для обработки нескольких ячеек в строке
- 

## переименование столбцов

`.set_axis(['old_name', 'old_name', 'new_name'], axis = 'columns', inplace=True)`

vs

`.rename`

---

## зачем отдельный словарь номер-название категории?

когда много повторов строк

проблемы длинных назв-ий категорий

-визуально хуже

-дольше обработка

-сложнее срезы делать

-дольше редачить категории

---

## Классификация по возрастным группам

если группа мало представлена по ней не сделать вывод

---

## Используем функцию для категоризации - `f(value)`

`df.column.apply(func)`

vs

`pd.cut (karpov)`

---

## Функция для катег-ции по неск столбцам - `f(row)`

(можно 2 столбца как в прош пункте - 1 столбец в индекс)

`row = pd.Series(data=row_values, index=row_columns)`

`row(name_col) = value`

case - конструкция

`df.apply`

`apply(func, axis = 1)`

---

## case:

оптимизацией работы службы поддержки

по нашей **сегментации клиентов по возрасту**  
**и занятости**

может быть создана

система приоритетов поступающих проблем