



Методы оптимизации pandas

Константин Башевой

Аналитик-разработчик, Яндекс



Константин Башевой
Аналитик-разработчик
Яндекс

Помогаю аналитикам с инфраструктурой
Собираю инструменты обработки данных
Рассказываю как это весело

Последние 10 лет:

Rambler&Co

Ростелеком

Яндекс

Что сегодня будет

Программа на сегодня

4



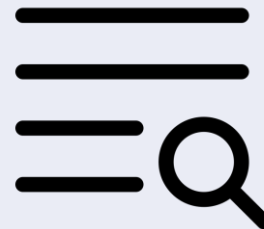
Поиск

Как его ускорять
простыми методами



Pivot и фильтры

Еще про ускорение
и вычисления



Румсистем

Работа с падежами
и склонениями

Основные идеи простой оптимизации

Простая оптимизация

6

Время поиска по списку пропорционально длине списка

```
5 in [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

```
True
```

Время поиска по множеству не зависит от времени

```
5 in {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}
```

```
True
```

Простая оптимизация

7

Время поиска по ключам словаря не зависит от времени.
Кортежи могут быть ключами словаря

```
costs_dict = {  
    ('2020-01-01', 'campaign_1', 'banner_1'): 100,  
    ('2020-01-01', 'campaign_1', 'banner_2'): 200,  
}  
  
costs_dict[('2020-01-01', 'campaign_1', 'banner_1')]  
  
100
```