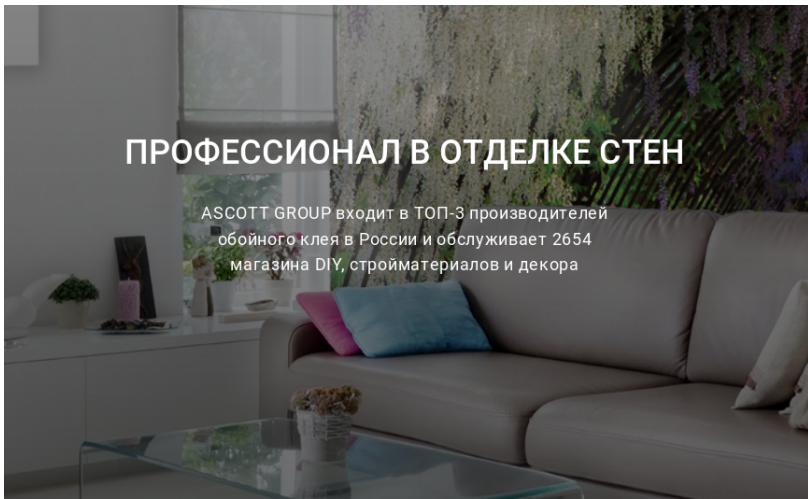


EUGEN BOBROV

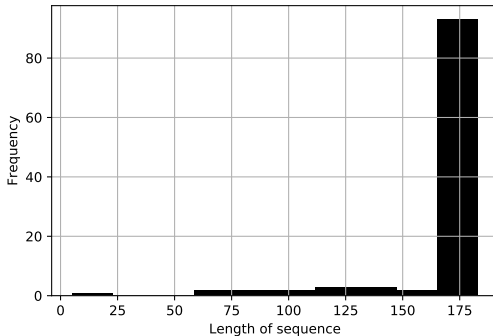
ПРОФЕССИОНАЛ В ОТДЕЛКЕ СТЕН

ASCOTT GROUP входит в ТОП-3 производителей
обойного клея в России и обслуживает 2654
магазина DIY, стройматериалов и декора



В обучающей выборке есть один магазин, которого нет в тестовой. Его последовательность крайне мала – всего 5 записей. Тогда как медиана длин последовательностей равна 180. Необходимо исключить его из обучающей выборки:

- ▶ `idFilial = 15`
- ▶ `idSubGrp = 1`
- ▶ `KanalDB = 1`



Длины последовательностей обучающей выборки:

```
1 data.size().values
```

```
array([167, 179, 180, 180, 180, 179, 114, 171, 180, 180, 182, 180, 144,  
       181, 181, 181, 181, 181, 168, 179, 182, 182, 182, 181, 170, 176,  
       180, 180, 180, 180, 179, 181, 181, 181, 181, 181, 118, 170, 179,  
       181, 180, 176, 177, 180, 183, 183, 182, 182, 180, 181, 181, 181,  
       181, 181, 175, 180, 180, 180, 181, 180, 76, 180, 180, 180, 181,  
       181, 71, 169, 175, 180, 181, 173, 79, 181, 181, 181, 181, 181,  
       146, 179, 180, 180, 180, 181, 5, 150, 157, 127, 168, 105, 137,  
       177, 180, 178, 178, 178, 95, 175, 178, 180, 181, 173, 81, 177,  
       179, 180, 181, 180])
```

```
1 grouped.size().median()
```

180.0

```
1 grouped.size().min()
```

5

```
1 grouped.size().idxmin()
```

(15, 1, 1)

```
1 test[(test['idFilial'] == 15) &  
2      (test['idSubGrp'] == 1) & (test['KanalDB'] == 1)]
```

	wk	N wk	idFilial	KanalDB	idSubGrp
id					

Достаточно вставить следующий код:

```
1 data = pd.read_csv('sample/train_set_weeks.csv', index_col='id')
2 data.shape
```

(18254, 6)

```
1 data = data[~((data['idFilial'] == 15) &
2             (data['idSubGrp'] == 1) & (data['KanalDB'] == 1))]
3 data.shape
```

(18249, 6)