

qcut 等分cdf, q=4是quartile: 0~25percentile, 25~50 percentile....

```
data['newCol'] = pd.qcut(data['column'], 4, labels=False)
data['newCol'] = pd.qcut(data['column'], [0, .25, .5, .75, 1.], labels=False)
# Number of quantiles. 10 for deciles, 4 for quartiles, etc.
# Alternately array of quantiles, e.g. [0, .25, .5, .75, 1.] for quartiles.
```

cut 等分距离, 每个interval大小相同

```
data['newCol'] = pd.cut(data['column'], bins=4, labels=False)
```

replace

```
data['newCol'] = data['column'].replace(['NA'], 'NotFound')
```

replace with a dict

```
bottom = data['column'].value_counts()[50:].index
bottom_dict = dict.fromkeys(bottom, 'Other')
data['newCol'] = data['col'].replace(bottom_dict)
```

str

```
names = names.apply(lambda x: x.astype(str).str.lower())
```

```
def id_to_item(df):
    """ function to convert id into counts """
    # 'sum' here is adding two lists into one big list
    ids = df['id'].str.split(',').sum()
    id_list = [0 for i in range(1, 49)]
    for i in ids:
        id_list[int(i) - 1] += 1

    return pd.Series(id_list, index=list(range(1, 49)))
```

pivot

```
p = data_sex.pivot(index='sex', columns='Page', values='ConversionRate')
```