```
import pandas as pd
import numpy as np
```

```
# What are vectorized operations
a = np.array([1,2,3,4])
a * 4
```

```
# problem in vectorized opertions in vanilla python
s = ['cat','mat',None,'rat']

[i.startswith('c') for i in s]
```

```
# How pandas solves this issue?

s = pd.Series(['cat','mat',None,'rat'])
# string accessor
s.str.startswith('c')

# fast and optimized
```

+ Code + Text

```
# import titanic
df = pd.read_csv('/content/titanic.csv')
df['Name']
```

```
# Common Functions
# lower/upper/capitalize/title
df['Name'].str.upper()
df['Name'].str.capitalize()
df['Name'].str.title()
# len
df['Name'][df['Name'].str.len() == 82].values[0]
# strip
"                   nitish                                ".strip()
df['Name'].str.strip()
```

```
# split -> get
df['lastname'] = df['Name'].str.split(',').str.get(0)
df.head()
```

```
df[['title','firstname']] = df['Name'].str.split(',').str.get(1).str.strip().str.split(' ', n=1, expand=True
df.head()

df['title'].value_counts()
```

```
# replace
df['title'] = df['title'].str.replace('Ms.','Miss.')
df['title'] = df['title'].str.replace('Mlle.','Miss.')
```

```
df['title'].value_counts()
```

```
# filtering
# startswith/endswith
df[df['firstname'].str.endswith('A')]
# isdigit/isalpha...
df[df['firstname'].str.isdigit()]
```

```
# applying regex
# contains
# search john -> both case
df[df['firstname'].str.contains('john',case=False)]
# find lastnames with start and end char vowel
df[df['lastname'].str.contains('^[^aeiouAEIOU].+[^aeiouAEIOU]$')]
```

```
# slicing
df['Name'].str[::-1]
```

```

```