

@learneverythingai



# 14 Tips for Structuring your CSV files, implement in your next Data Science Project



**SHIVAM MODI**

@learneverythingai

@learneverythingai

1

# Loading Dataset

```
df = pd.read_csv('filename.csv')
```



@learneverythingai

2

# Opening from URL

```
import requests  
from io import StringIO
```

```
url = "your-url-here"
```

```
headers = {"User-Agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X  
10.14; rv:66.0) Gecko/20100101Firefox/66.0"}
```

```
req = request.get(url, headers=headers)
```

```
data = StringIO(req.text)
```

```
pd.read_csv(data)
```



@learneverythingai

3

# Separated by Tabs

```
pd.read_csv('filename.tsv', sep='\t')
```



@learneverythingai

# 4

## Using selected columns

```
pd.read_csv('filename.csv', usecols=['sno', 'name', 'rating'])
```

	sno	name	release_year	rating	votes	genres
0	m0	10 things i hate about you	1999	6.9	62847.0	['comedy' 'romance']
1	m1	1492: conquest of paradise	1992	6.2	10421.0	['adventure' 'biography' 'drama' 'history']
2	m2	15 minutes	2001	6.1	25854.0	['action' 'crime' 'drama' 'thriller']
3	m3	2001: a space odyssey	1968	8.4	163227.0	['adventure' 'mystery' 'sci-fi']
4	m4	48 hrs.	1982	6.9	22289.0	['action' 'comedy' 'crime' 'drama' 'thriller']

@learneverythingai

# 5

## Choosing Index Column

```
pd.read_csv('filename.csv', index_col='col')
```

	city	city_development_index	gender
<b>enrollee_id</b>			
8949	city_103	0.920	Male
29725	city_40	0.776	Male
11561	city_21	0.624	NaN

@learneverythingai

# 6

## Starting Header with specific number

```
pd.read_csv('filename.csv', header=1)
```

0	enrollee_id	city	city_development_index
1	29725	city_40	0.776
2	11561	city_21	0.624
3	33241	city_115	0.789

@learneverythingai

7

# Using n number of rows

```
pd.read_csv('filename.csv', nrows=100)
```

100 rows x 14 columns



@learneverythingai

8

# Skipping Bad Lines

```
pd.read_csv('filename.csv', error_bad_lines=False)
```

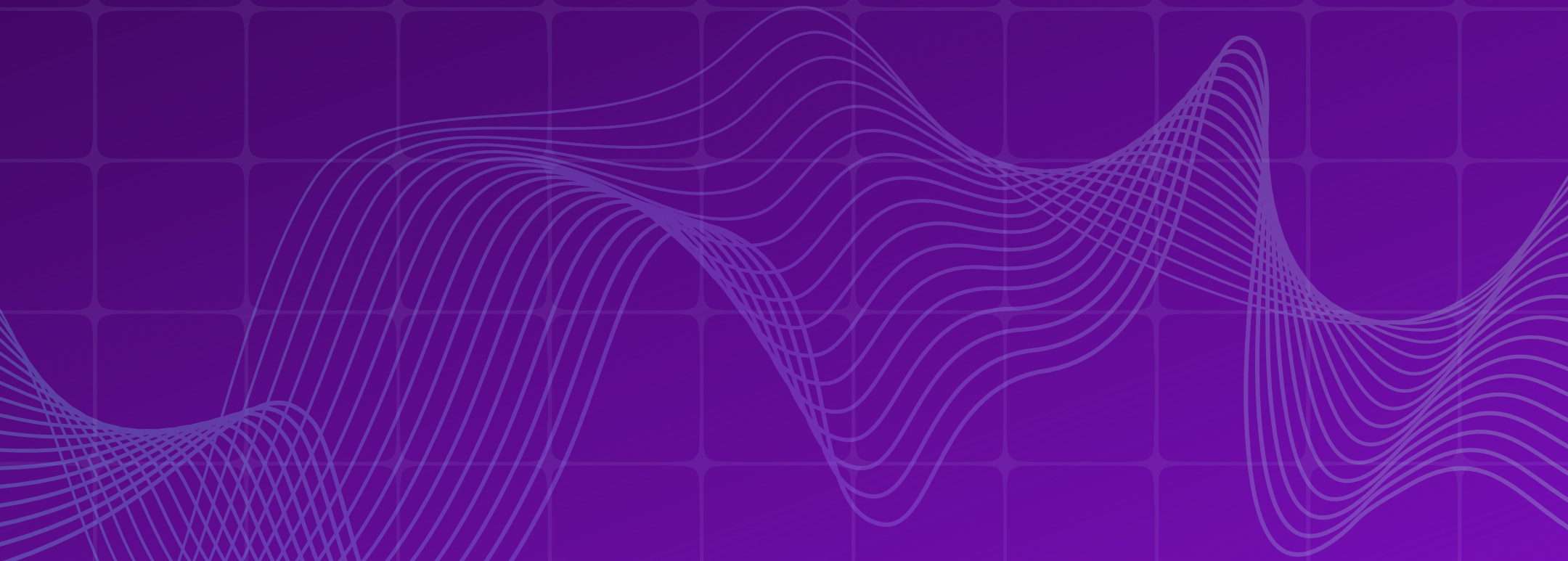
```
b'Skipping line 6452: expected 8 fields, saw 7
b'Skipping line 92038: expected 8 fields, saw 7
b'Skipping line 144058: expected 8 fields, saw 7
b'Skipping line 209388: expected 8 fields, saw 7
```

@learneverythingai

9

# Encoding Parameters

```
pd.read_csv('filename.csv', encoding='latin-1')
```



@learneverythingai

10

# Changing Data type

```
pd.read_csv('filename.csv', dtype={'column':int})
```

```
13 target 19158 non-null int32
```

@learneverythingai

# 11

## Parsing Dates

```
pd.read_csv('filename.csv', parse_dates=['col'])
```

```
2    date    816 non-null    datetime64[ns]
```

@learneverythingai

12

# Renaming Value

*function*



`pd.read_csv('filename.csv', converters={'col': rename})`

```
def rename(name):  
    if name == "Royal Challengers Bangalore":  
        return "RCB"  
    else:  
        return name
```

```
rename("Royal Challengers Bangalore")
```

@learneverythingai

13

# Loading Data into Chunks

```
dfs = pd.read_csv('filename.csv', chunksize=5000)
```

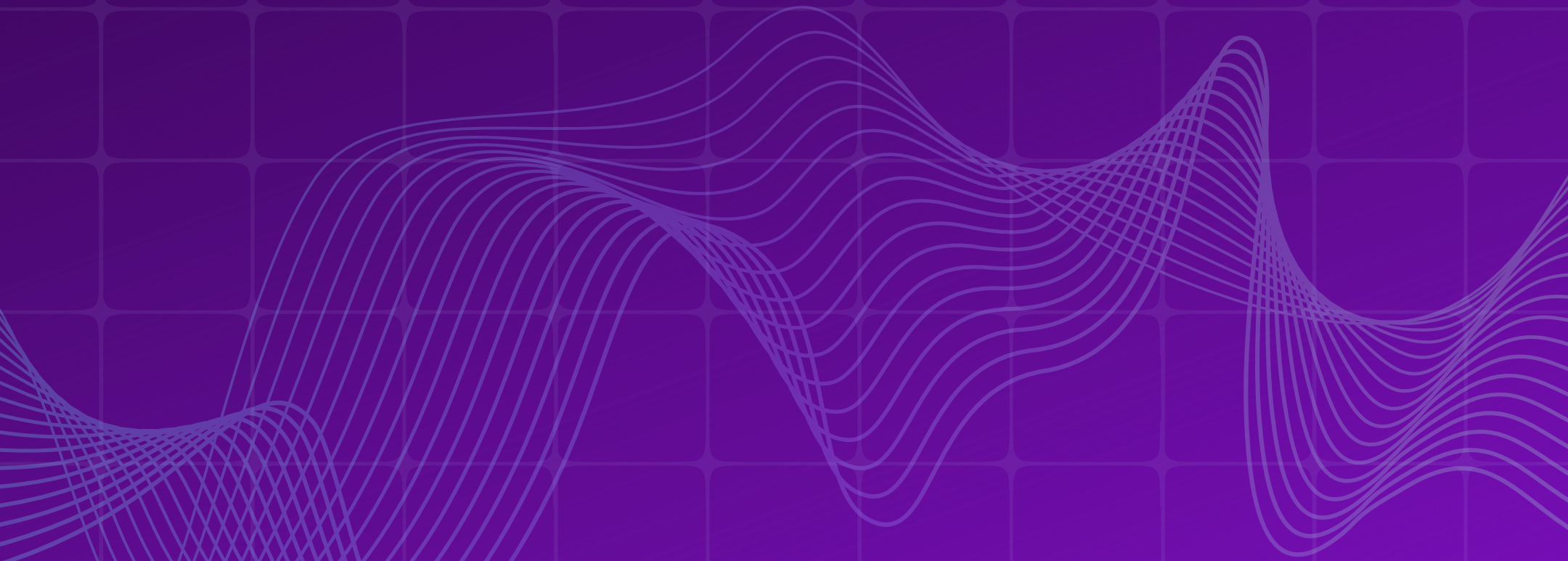


@learneverythingai

14

# Specifying NA Values

```
pd.read_csv('filename.csv', na_values=['val',])
```



# Like this Post?

- Follow Me
- Share with your friends
- Check out my previous posts



SHIVAM MODI

@learneverythingai



Share This

FOLLOW

SAVE THIS

[www.learneverythingai.com](http://www.learneverythingai.com)

