Pandas Fundamentals

GETTING STARTED

Paweł Kordek
SOFTWARE ENGINEER

@pawel_kordek   https://kordek.github.io

Overview

Why is Pandas for you

The dataset

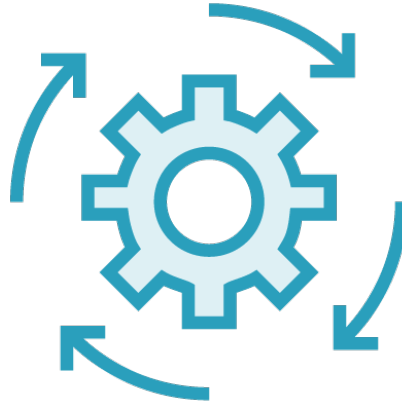Basic objects in Pandas

Why Use Pandas?

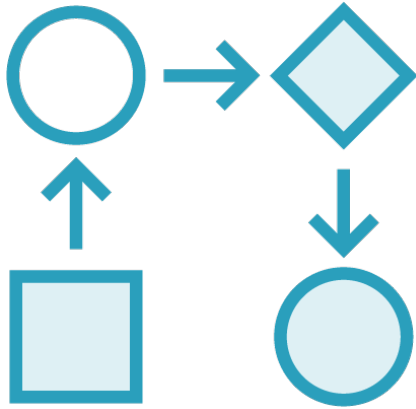# My Story

Large input

Numerous custom transformations

Output

Time saved!

Wide Range of Use Cases

From
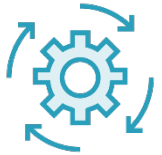Simple data
workflows automation

To
Professional
data analysis

Pandas Features

Intuitive data format

Easy data transformations

Data visualization

The Dataset

Tate Collection metadata

Around 70,000 artworks

Available as JSON or CSV

- https://github.com/ tategallery/collection

Basic Objects

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

Don't Get Scared

These are typical imports.

Types to Be Aware Of

| np.array / np.ndarray | pd.Series | pd.DataFrame |

```
import numpy as np


my_numpy_array = np.random.rand(3)
```

Numpy Arrays – Example

3-element ndarray (random).

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

my_first_series = pd.Series(np.random.rand(3))
```

Pandas Series

Series always has one column.

```
import numpy as np
import pandas as pd

my_first_df = pd.DataFrame(np.random.rand(3, 2))
```

Pandas DataFrame

Two columns with random numbers.

How Does It Look Like

	Col 1	Col 2
Row 1	0.23	0.44
Row 2	0.78	0.99
Row 3	0.55	0.02

Regular/NumPy array

Series

DataFrame

Demo

We will:

- Create a Series object
- Create a DataFrame object
- Inspect their properties