

Python Pandas Working with XML Tutorials

1. Read XML and get Dataframe
2. convert Dataframe to XML

What is XML?

1. XML stands for eXtensible Markup Language
2. XML is a markup language much like HTML
3. XML was designed to store and transport data
4. XML was designed to be self-descriptive
5. XML is a W3C Recommendation

```
In [1]: import pandas as pd
```

```
In [2]: xml='''<?xml version='1.0' encoding='utf-8'?>
<data xmlns="http://example.com">
  <row>
    <shape>square</shape>
    <degrees>360</degrees>
    <sides>4.0</sides>
    <firstname>Krish</firstname>
  </row>
  <row>
    <shape>circle</shape>
    <degrees>360</degrees>
    <sides/>
    <firstname/>
  </row>
  <row>
    <shape>triangle</shape>
    <degrees>180</degrees>
    <sides>3.0</sides>
    <firstname/>
  </row>
</data>'''
```

```
In [3]: pd.read_xml(xml)
```

```
Out[3]:
```

	shape	degrees	sides	firstname
0	square	360	4.0	Krish
1	circle	360	NaN	None
2	triangle	180	3.0	None

```
In [4]: xml='''<?xml version='1.0' encoding='utf-8'?>
<data>
  <row shape="square" degrees="360" sides="4.0" firstname="Krish"/>
  <row shape="circle" degrees="360"/>
  <row shape="triangle" degrees="180" sides="3.0" lastname="Naik"/>
</data>'''
```

```
In [6]: pd.read_xml(xml,xpath="//row")
```

```
Out[6]:
```

	shape	degrees	sides	firstname	lastname
0	square	360	4.0	Krish	None
1	circle	360	NaN	None	None
2	triangle	180	3.0	None	Naik

```
In [7]: xml='''<?xml version='1.0' encoding='utf-8'?>
<doc:data xmlns:doc="https://example.com">
  <doc:row>
    <doc:shape>square</doc:shape>
    <doc:degrees>360</doc:degrees>
    <doc:sides>4.0</doc:sides>
  </doc:row>
  <doc:row>
    <doc:shape>circle</doc:shape>
    <doc:degrees>360</doc:degrees>
    <doc:sides/>
  </doc:row>
  <doc:row>
    <doc:shape>triangle</doc:shape>
    <doc:degrees>180</doc:degrees>
    <doc:sides>3.0</doc:sides>
  </doc:row>
</doc:data>'''
```

```
In [15]: df=pd.read_xml(xml,xpath="//doc:row",namespaces={"doc":"https://example.com"})
```

```
In [16]: df
```

```
Out[16]:
```

	shape	degrees	sides
0	square	360	4.0
1	circle	360	NaN
2	triangle	180	3.0

```
In [17]: df.to_xml('test1.xml')
```

```
In [ ]:
```

