R to Python

Prithviraj Lakkakula

7/25/2022

Contents

Basics of Programming

Creating a dataframe

R code

```
# construct a dataframe
df <- data.frame(a = c(1, 2, 3, 4, 5),
                 b = c(3, 4, 5, 6, 7))
df
```

```
##   a b
## 1 1 3
## 2 2 4
## 3 3 5
## 4 4 6
## 5 5 7
```

Python code

```
import pandas as pd
df = pd.DataFrame({'a': [1, 2, 3, 4, 5],
                           'b': [3, 4, 5, 6, 7]})
print(df)
```

```
##    a  b
## 0  1  3
## 1  2  4
## 2  3  5
## 3  4  6
## 4  5  7
```

Imports

R code

```
sqrt(36)
```

```
## [1] 6
```

Python code

```
import math
math.sqrt(36)

#importing a function
```

```
## 6.0
```

```
from math import sqrt
sqrt(36)

# show all functions in math model
```

```
## 6.0
```

```
print(dir(math))
```

```
## ['__doc__', '__file__', '__loader__', '__name__', '__package__', '__spec__', 'acos', 'acosh', 'asin'
```

Data Types

R code

```
class(5)
```

```
## [1] "numeric"
```

```
class(5.0)
```

```
## [1] "numeric"
```

```
class('Five')
```

```
## [1] "character"
```

```
class(FALSE)
```

```
## [1] "logical"
```

```
## check if an object is of a given type
is.numeric(5)
```

```
## [1] TRUE
```

```
is.character('Five')
```

```
## [1] TRUE
```

```
is.logical(FALSE)
```

```
## [1] TRUE
```

```
## Convert an object to a given type
as.character(5.5)
```

```
## [1] "5.5"
```

```
#as.numeric, as.factor etc to convert to numeric and factor types
```

Python code

```
type(5)
```

```
## <class 'int'>
```

```
type(5.0)
```

```
## <class 'float'>
```

```
type('five')
```

```
## <class 'str'>
```

```
type(False)

## check if an object is of a given type
```

```
## <class 'bool'>
```

```
isinstance(5.0, int)
```

```
## False
```

```
isinstance(5.0, (int, float))
```

```
## True
```

```
isinstance('Five', str)
```

```
## True
```

```
isinstance('Five', int)
```

```
## False
```

```
isinstance(False, bool)

## Convert an object to a given type
```

```
## True
```

```
str(5.5)
```

```
## '5.5'
```

Importing data from a variety of data formats

How to call heads and tails?

How to know the shape/length of the data?

How to access single and multiple columns?

Get summary stats of the data variables

How to graph plots (scatter plot, histogram, boxplot, barplot etc.)?

Handling missing values?

Handling outliers?

Checking class imbalance?

Splitting the dataset into training and test sets for both cross section and time series data cases?

Run a linear regression model along with predictions on test set, model evaluation, performance metrics?

Run a logistic regression model along with predictions on test set, model evaluation, performance metrics?