

R to Python

Prithviraj Lakkakula

7/25/2022

Contents

Creating a dataframe and performing set operations	1
Importing data from a variety of data formats	2
How to call heads and tails?	2
How to know the shape/length of the data?	2
How to access single and multiple columns?	2
Get summary stats of the data variables	2
How to graph plots (scatter plot, histogram, boxplot, barplot etc.)?	2
Handling missing values?	2
Handling outliers?	2
Checking class imbalance?	2
Splitting the dataset into training and test sets for both cross section and time series data cases? .	2
Run a linear regression model along with predictions on test set, model evaluation, performance metrics?	2
Run a logistic regression model along with predictions on test set, model evaluation, performance metrics?	2

Creating a dataframe and performing set operations

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
```

```
#intersect(df$a, df$b)
```

Python code

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

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

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?