Pandas

Muchang Bahng

Fall 2024

Contents

1   Introduction

There really isn't anything hard about pandas, but it does have some nice properties that connect with those of relational databases. It also has some subtle features.

Definition 1.1 (Initialization)

A pandas dataframe can be initialized in three ways generally.
1. With a dictionary, where the keys represent the columns.

```
>>> data = {
...      "name" : ["Bob", "Jon", "Mary"],
...      "age" : [14, 19, 21]
... }
>>> df = pd.DataFrame(data)
>>> print(df)
   name  age
0   Bob   14
1   Jon   19
2  Mary   21
```

2. With a list of lists, which are like a stack of rows. Note that this does not provide column names, so it should be added through the `columns` keyword argument.

```
>>> data = [["Bob", 14], ["Jon", 19], ["Mary", 21]]
>>> df = pd.DataFrame(data, columns=["name", "age"])
>>> print(df)
   name  age
0   Bob   14
1   Jon   19
2  Mary   21
```

3. With a list of dictionaries, which is also like a stack of rows but now the columns are provided.

```
>>> data = [
...      {"name" : "Bob", "age" : 14},
...      {"name" : "Jon", "age" : 19},
...      {"name" : "Mary", "age" : 21}
... ]
>>> df = pd.DataFrame(data, columns=["name", "age"])
>>> print(df)
   name  age
0   Bob   14
1   Jon   19
2  Mary   21
```