2.5_solutions

January 26, 2023

1 Solutions

1.0.1 pandas dataframes

1. Import the pandas library with the standard namespace

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

2. Create a pandas dataframe with 1 column named 'country' and 4 rows with different country names

[2]:
```
datalist=['USA', 'Monaco', 'France', 'UK']
df=pd.DataFrame(datalist, columns=['country'])
df
```

[2]:
```
  country
0     USA
1  Monaco
2  France
3      UK
```

3. Add a new column named 'Sales' and include some numeric values (imagine its sales for regional brands)

[3]:
```
df['Sales']=[50, 35,80,90]
df
```

[3]:
```
  country  Sales
0     USA     50
1  Monaco     35
2  France     80
3      UK     90
```

4. Use the column country to create an index.

[4]:
```
df.set_index(['country'], inplace=True)
df
```

[4]:
```
         Sales
country
```

```
USA         50
Monaco      35
France      80
UK          90
```

5. Create another column called Cost. again, add some numeric values.

[5]:
```
df['Cost']=(26,23,55,60)  # notice i can also pass tuple to create column
df
```

[5]:
```
         Sales  Cost
country
USA         50    26
Monaco      35    23
France      80    55
UK          90    60
```

6. Compute a new column `Revenue` as `Sales - Cost`.

[6]:
```
df['Revenue']=df['Sales']-df['Cost']
df
```

[6]:
```
         Sales  Cost  Revenue
country
USA         50    26       24
Monaco      35    23       12
France      80    55       25
UK          90    60       30
```

7. What is the shape of the final array?

[8]:
```
print(df.shape)
```

```
(4, 3)
```

8. What is the revenue value for the second country in your dataframe? (write some code to display only this value as output)

[10]:
```
df.iloc[1]['Revenue']
```

[10]:
```
12
```

[ ]:
```

```