Practical 32

Write a python program to find missing data and and perform encoding of it.

[6]:
```
import numpy as np
import pandas as pd

import random
a, b, c = [random.randint(1, 10) for _ in range(10)], [random.randint(1, 10)␣
 ↪for _ in range(10)], [random.randint(1, 10) for _ in range(10)]
df = pd.DataFrame({'A': a,'B': b,'C': c})
print(df)
df = df.sort_values(by=['B'],ascending=[True])
df = df.reset_index(drop=True)
print(df)
```

```
   A  B  C
0  7  6  6
1  1  5  9
2  1  4  3
3  2  2  7
4  8  3  8
5  3  9  7
6  4  5  9
7  4  4  1
8  8  4  9
9  4  5  4
   A  B  C
0  2  2  7
1  8  3  8
2  4  4  1
3  1  4  3
4  8  4  9
5  4  5  9
6  4  5  4
7  1  5  9
8  7  6  6
9  3  9  7
```

[7]:
```
age_list = [['CountryA', 2000, 10000000, 'Region1'],
            ['CountryB', 2005, 15000000, 'Region2'],
            ['CountryC', 2010, 20000000, 'Region3'],
            ['CountryD', 2015, 25000000, 'Region4'],
            ['CountryE', 2020, 30000000, 'Region5']]

# You can add more entries if needed by following the same pattern.
```

```
df = pd.DataFrame(age_list, columns=['Country', 'Year', 'Population',␣
↪'Continent'])
df
```

[7]:

	Country	Year	Population	Continent
0	CountryA	2000	10000000	Region1
1	CountryB	2005	15000000	Region2
2	CountryC	2010	20000000	Region3
3	CountryD	2015	25000000	Region4
4	CountryE	2020	30000000	Region5

[8]:
```
df.sort_values(by=['Year'])
```

[8]:

	Country	Year	Population	Continent
0	CountryA	2000	10000000	Region1
1	CountryB	2005	15000000	Region2
2	CountryC	2010	20000000	Region3
3	CountryD	2015	25000000	Region4
4	CountryE	2020	30000000	Region5

[9]:
```
df.sort_values(by=['Population'], ascending=False)
```

[9]:

	Country	Year	Population	Continent
4	CountryE	2020	30000000	Region5
3	CountryD	2015	25000000	Region4
2	CountryC	2010	20000000	Region3
1	CountryB	2005	15000000	Region2
0	CountryA	2000	10000000	Region1

[10]:
```
df.sort_values(by=['Country', 'Continent'])
```

[10]:

	Country	Year	Population	Continent
0	CountryA	2000	10000000	Region1
1	CountryB	2005	15000000	Region2
2	CountryC	2010	20000000	Region3
3	CountryD	2015	25000000	Region4
4	CountryE	2020	30000000	Region5

[11]:
```
df.sort_values(by=['Country', 'Continent'],ascending=[False, True])
```

[11]:

	Country	Year	Population	Continent
4	CountryE	2020	30000000	Region5
3	CountryD	2015	25000000	Region4
2	CountryC	2010	20000000	Region3
1	CountryB	2005	15000000	Region2
0	CountryA	2000	10000000	Region1

[12]:
```
index = df.index.to_list()
print("index:",index)
np.random.shuffle(index)
print("shuffle index: ",index)
df = df.loc[index]
df
```

```
index: [0, 1, 2, 3, 4]
shuffle index:  [4, 1, 0, 2, 3]
```

[12]:

	Country	Year	Population	Continent
4	CountryE	2020	30000000	Region5
1	CountryB	2005	15000000	Region2
0	CountryA	2000	10000000	Region1
2	CountryC	2010	20000000	Region3
3	CountryD	2015	25000000	Region4

[ ]: