

Artificial Intelligence

AI-03: 小テストの解答例



ai-03-quizの正解例

- 1.必要なライブラリをimportし、running.csvのデータをデータフレームdfに読み込む。

```
import numpy as np
import pandas as pd

csv_in = 'running.csv'

df = pd.read_csv(csv_in, delimiter=',', skiprows=0, header=0)
```

2. 行数と列数、各列のデータ型と欠損値でないデータの数、dfの先頭付近を表示して確認。

```
print(df.shape)
print(df.info())
display(df.head())
```

```
(50, 4)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 4 columns):
ID          50 non-null int64
Univ        48 non-null object
Grade       50 non-null int64
Time        49 non-null float64
dtypes: float64(1), int64(2), object(1)
memory usage: 1.6+ KB
None
```

50行, 4列

非欠損値が50未満の行には欠損値がある

np.nan (欠損値) がある数値行はfloat型になる

	ID	Univ	Grade	Time
0	1001	D	3	1870.0
1	1002	<u>NaN</u>	2	1937.0
2	1003	<u>NaN</u>	1	2036.0
3	1004	C	1	1957.0
4	1005	D	1	<u>NaN</u>

np.nan (欠損値)がある数値行はfloat型になる

3. dfから欠損値を含む行を削除し、データフレームdf2に代入。df2の先頭10行を表示して確認。

```
df2 = df.dropna(axis=0).reset_index(drop=True)
print(df2.shape)
print(df2.info())
display(df2.head(10))
```

	ID	Univ	Grade	Time
0	1001	D	3	1870.0
1	1004	C	1	1957.0
2	1006	D	2	2063.0
3	1007	D	3	1839.0
4	1008	C	2	1740.0
5	1009	B	3	2082.0
6	1010	C	4	2045.0
7	1011	A	3	1818.0
8	1012	A	3	1783.0
9	1013	A	2	1965.0

AI-0102の「欠損値の削除」
「indexのリセット」参照

欠損値が削除できている

4. df2からUniv列が 'C' 以外かつGrade列が 3 以下の行を抽出し、データフレームdf3に代入。df3の先頭10行を表示して確認。

```
df3 = df2[ (df2['Univ'] != 'C') & (df2['Grade'] <= 3) ]  
print(df3.shape)  
display(df3.head(10))
```

	ID	Univ	Grade	Time
(31, 4)				
0	1001	D	3	1870.0
2	1006	D	2	2063.0
3	1007	D	3	1839.0
5	1009	B	3	2082.0
7	1011	A	3	1818.0
8	1012	A	3	1783.0
9	1013	A	2	1965.0
10	1014	D	1	1803.0
11	1015	A	2	2123.0
14	1018	D	1	1926.0

C大学以外、かつ1-3年生だけの行が抽出できている。

5. df3のうち、Time列の値が最小の行を表示。

```
display(df3.loc[df3['Time'].idxmin()])
```

```
ID          1030
Univ         D
Grade        3
Time        1761
Name: 26, dtype: object
```

df3から、Time列の値が最小の
列を抽出できた。

5. df2からUniv列が 'C' 以外かつGrade列が 3 以下の行を抽出し、データフレームdf3に代入。df3の先頭10行を表示して確認。

```
df3 = df2[ (df2['Univ'] != 'C') & (df2['Grade'] <= 3) ]  
print(df3.shape)  
display(df3.head(10))
```

	ID	Univ	Grade	Time
(31, 4)				
0	1001	D	3	1870.0
2	1006	D	2	2063.0
3	1007	D	3	1839.0
5	1009	B	3	2082.0
7	1011	A	3	1818.0
8	1012	A	3	1783.0
9	1013	A	2	1965.0
10	1014	D	1	1803.0
11	1015	A	2	2123.0
14	1018	D	1	1926.0

C大学と4年生の行が
削除できている。