

employees-data

February 22, 2025

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[3]: df=pd.read_csv(r'C:\Users\kmuba\Downloads\Employee Dataset.csv ')
```

```
[5]: df
```

```
[5]:
```

	id	groups	age	healthy_eating	active_lifestyle	salary
0	0	A	36	5	5	2297
1	1	A	26	3	5	1134
2	2	A	61	8	1	4969
3	3	O	24	3	6	902
4	4	O	39	6	2	3574
5	5	O	32	5	3	2761
6	6	AB	25	4	6	1484
7	7	B	39	8	6	3809
8	8	A	32	5	6	2065
9	9	A	28	4	8	1020
10	10	A	30	6	9	1950
11	11	O	40	7	1	4387
12	12	O	23	3	2	1830
13	13	O	30	6	8	2182
14	14	AB	51	7	5	3460
15	15	B	26	3	8	662
16	16	A	30	3	5	1134
17	17	A	22	4	8	1020
18	18	A	27	2	6	779
19	19	O	32	7	6	3228
20	20	O	39	7	6	3228
21	21	O	26	1	6	1360
22	22	AB	40	7	5	3460
23	23	B	37	6	4	3110
24	24	A	28	5	8	1601
25	25	A	30	6	7	2414
26	26	A	62	9	6	700

27	27	0	37	3	5	1134
28	28	0	30	6	7	2414
29	29	0	37	8	6	3809
30	30	AB	28	4	7	1252
31	31	B	30	7	8	2764
32	32	A	22	3	8	662
33	33	A	25	4	6	1484
34	34	A	29	5	7	1833
35	35	0	27	2	4	785
36	36	0	39	5	5	2297
37	37	0	26	3	5	1134
38	38	AB	26	3	7	670
39	39	B	29	2	6	779
40	40	A	30	6	8	2182
41	41	A	32	5	3	2761
42	42	A	31	6	7	2414
43	43	0	23	4	10	556
44	44	0	35	6	6	2646
45	45	0	25	5	8	1601
46	46	AB	29	4	3	2179
47	47	B	30	4	6	1484
48	48	A	30	5	6	2065
49	49	A	29	6	9	1950

```
[7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id              50 non-null    int64
1   groups          50 non-null    object
2   age             50 non-null    int64
3   healthy_eating  50 non-null    int64
4   active_lifestyle 50 non-null    int64
5   salary          50 non-null    int64
dtypes: int64(5), object(1)
memory usage: 2.5+ KB
```

```
[9]: df.describe()
```

```
[9]:
```

	id	age	healthy_eating	active_lifestyle	salary
count	50.00000	50.000000	50.000000	50.000000	50.000000
mean	24.50000	31.880000	4.920000	5.900000	2027.500000
std	14.57738	8.390082	1.827678	2.022829	1076.861885
min	0.00000	22.000000	1.000000	1.000000	556.000000

25%	12.25000	26.250000	3.250000	5.000000	1134.000000
50%	24.50000	30.000000	5.000000	6.000000	1950.000000
75%	36.75000	35.750000	6.000000	7.000000	2732.250000
max	49.00000	62.000000	9.000000	10.000000	4969.000000

```
[11]: df.shape
```

```
[11]: (50, 6)
```

```
[13]: df.columns
```

```
[13]: Index(['id', 'groups', 'age', 'healthy_eating', 'active_lifestyle', 'salary'],  
        dtype='object')
```

```
[ ]:
```