

# Assign\_3

March 8, 2023

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

```
[2]: df = pd.read_csv('data.csv')
df
```

```
[2]:
```

|        | Number | City   | Gender | Age | Income   | Illness |
|--------|--------|--------|--------|-----|----------|---------|
| 0      | 1      | Dallas | Male   | 41  | 40367.0  | No      |
| 1      | 2      | Dallas | Male   | 54  | 45084.0  | No      |
| 2      | 3      | Dallas | Male   | 42  | 52483.0  | No      |
| 3      | 4      | Dallas | Male   | 40  | 40941.0  | No      |
| 4      | 5      | Dallas | Male   | 46  | 50289.0  | No      |
| ...    | ...    | ...    | ...    | ... | ...      | ...     |
| 149995 | 149996 | Austin | Male   | 48  | 93669.0  | No      |
| 149996 | 149997 | Austin | Male   | 25  | 96748.0  | No      |
| 149997 | 149998 | Austin | Male   | 26  | 111885.0 | No      |
| 149998 | 149999 | Austin | Male   | 25  | 111878.0 | No      |
| 149999 | 150000 | Austin | Female | 37  | 87251.0  | No      |

[150000 rows x 6 columns]

```
[3]: frame = pd.DataFrame(df)
```

```
[4]: frame
```

```
[4]:
```

|        | Number | City   | Gender | Age | Income   | Illness |
|--------|--------|--------|--------|-----|----------|---------|
| 0      | 1      | Dallas | Male   | 41  | 40367.0  | No      |
| 1      | 2      | Dallas | Male   | 54  | 45084.0  | No      |
| 2      | 3      | Dallas | Male   | 42  | 52483.0  | No      |
| 3      | 4      | Dallas | Male   | 40  | 40941.0  | No      |
| 4      | 5      | Dallas | Male   | 46  | 50289.0  | No      |
| ...    | ...    | ...    | ...    | ... | ...      | ...     |
| 149995 | 149996 | Austin | Male   | 48  | 93669.0  | No      |
| 149996 | 149997 | Austin | Male   | 25  | 96748.0  | No      |
| 149997 | 149998 | Austin | Male   | 26  | 111885.0 | No      |
| 149998 | 149999 | Austin | Male   | 25  | 111878.0 | No      |
| 149999 | 150000 | Austin | Female | 37  | 87251.0  | No      |

[150000 rows x 6 columns]

```
[9]: frame.isnull()
```

```
[9]:
```

|        | Number | City  | Gender | Age   | Income | Illness |
|--------|--------|-------|--------|-------|--------|---------|
| 0      | False  | False | False  | False | False  | False   |
| 1      | False  | False | False  | False | False  | False   |
| 2      | False  | False | False  | False | False  | False   |
| 3      | False  | False | False  | False | False  | False   |
| 4      | False  | False | False  | False | False  | False   |
| ...    | ...    | ...   | ...    | ...   | ...    | ...     |
| 149995 | False  | False | False  | False | False  | False   |
| 149996 | False  | False | False  | False | False  | False   |
| 149997 | False  | False | False  | False | False  | False   |
| 149998 | False  | False | False  | False | False  | False   |
| 149999 | False  | False | False  | False | False  | False   |

[150000 rows x 6 columns]

```
[10]: frame.isna().sum()
```

```
[10]:
```

|         |   |
|---------|---|
| Number  | 0 |
| City    | 0 |
| Gender  | 0 |
| Age     | 0 |
| Income  | 0 |
| Illness | 0 |

dtype: int64

```
[11]: frame
```

```
[11]:
```

|        | Number | City   | Gender | Age | Income   | Illness |
|--------|--------|--------|--------|-----|----------|---------|
| 0      | 1      | Dallas | Male   | 41  | 40367.0  | No      |
| 1      | 2      | Dallas | Male   | 54  | 45084.0  | No      |
| 2      | 3      | Dallas | Male   | 42  | 52483.0  | No      |
| 3      | 4      | Dallas | Male   | 40  | 40941.0  | No      |
| 4      | 5      | Dallas | Male   | 46  | 50289.0  | No      |
| ...    | ...    | ...    | ...    | ... | ...      | ...     |
| 149995 | 149996 | Austin | Male   | 48  | 93669.0  | No      |
| 149996 | 149997 | Austin | Male   | 25  | 96748.0  | No      |
| 149997 | 149998 | Austin | Male   | 26  | 111885.0 | No      |
| 149998 | 149999 | Austin | Male   | 25  | 111878.0 | No      |
| 149999 | 150000 | Austin | Female | 37  | 87251.0  | No      |

[150000 rows x 6 columns]

```
[12]: frame.describe()
```

```
[12]:
```

|       | Number        | Age           | Income        |
|-------|---------------|---------------|---------------|
| count | 150000.000000 | 150000.000000 | 150000.000000 |
| mean  | 75000.500000  | 44.950200     | 91252.798273  |
| std   | 43301.414527  | 11.572486     | 24989.500948  |
| min   | 1.000000      | 25.000000     | -654.000000   |
| 25%   | 37500.750000  | 35.000000     | 80867.750000  |
| 50%   | 75000.500000  | 45.000000     | 93655.000000  |
| 75%   | 112500.250000 | 55.000000     | 104519.000000 |
| max   | 150000.000000 | 65.000000     | 177157.000000 |

```
[13]: frame['Age_Group'] = frame.apply(lambda row: 'young' if (row.Age < 41 and row.  
↪ Age > 19) else ('mid' if (row.Age < 51 and row.Age > 39) else 'old'), axis=1)
```

```
[22]: frame
```

```
[22]:
```

|        | Number | City   | Gender | Age | Income   | Illness | Age_Group |
|--------|--------|--------|--------|-----|----------|---------|-----------|
| 0      | 1      | Dallas | Male   | 41  | 40367.0  | No      | mid       |
| 1      | 2      | Dallas | Male   | 54  | 45084.0  | No      | old       |
| 2      | 3      | Dallas | Male   | 42  | 52483.0  | No      | mid       |
| 3      | 4      | Dallas | Male   | 40  | 40941.0  | No      | young     |
| 4      | 5      | Dallas | Male   | 46  | 50289.0  | No      | mid       |
| ...    | ...    | ...    | ...    | ... | ...      | ...     | ...       |
| 149995 | 149996 | Austin | Male   | 48  | 93669.0  | No      | mid       |
| 149996 | 149997 | Austin | Male   | 25  | 96748.0  | No      | young     |
| 149997 | 149998 | Austin | Male   | 26  | 111885.0 | No      | young     |
| 149998 | 149999 | Austin | Male   | 25  | 111878.0 | No      | young     |
| 149999 | 150000 | Austin | Female | 37  | 87251.0  | No      | young     |

[150000 rows x 7 columns]

```
[19]: young_filtered = frame[~frame['Age_Group'].isin(['mid', 'old'])]
```

```
[20]: young_filtered
```

```
[20]:
```

|        | Number | City   | Gender | Age | Income   | Illness | Age_Group |
|--------|--------|--------|--------|-----|----------|---------|-----------|
| 3      | 4      | Dallas | Male   | 40  | 40941.0  | No      | young     |
| 5      | 6      | Dallas | Female | 36  | 50786.0  | No      | young     |
| 6      | 7      | Dallas | Female | 32  | 33155.0  | No      | young     |
| 7      | 8      | Dallas | Male   | 39  | 30914.0  | No      | young     |
| 9      | 10     | Dallas | Female | 30  | 50082.0  | No      | young     |
| ...    | ...    | ...    | ...    | ... | ...      | ...     | ...       |
| 149994 | 149995 | Austin | Female | 37  | 106097.0 | No      | young     |
| 149996 | 149997 | Austin | Male   | 25  | 96748.0  | No      | young     |
| 149997 | 149998 | Austin | Male   | 26  | 111885.0 | No      | young     |
| 149998 | 149999 | Austin | Male   | 25  | 111878.0 | No      | young     |

```
149999 150000 Austin Female 37 87251.0 No young
```

```
[58255 rows x 7 columns]
```

```
[21]: young_filtered.describe()
```

```
[21]:
```

|       | Number        | Age          | Income        |
|-------|---------------|--------------|---------------|
| count | 58255.000000  | 58255.000000 | 58255.000000  |
| mean  | 75237.552571  | 32.700970    | 91332.803502  |
| std   | 43338.282104  | 4.486425     | 24956.481752  |
| min   | 4.000000      | 25.000000    | -654.000000   |
| 25%   | 37904.000000  | 29.000000    | 80909.000000  |
| 50%   | 75153.000000  | 33.000000    | 93708.000000  |
| 75%   | 112799.500000 | 37.000000    | 104589.000000 |
| max   | 150000.000000 | 40.000000    | 170965.000000 |

```
[23]: mid_filtered = frame[~frame['Age_Group'].isin(['young', 'old'])]
```

```
[25]: mid_filtered
```

```
[25]:
```

|        | Number | City   | Gender | Age | Income  | Illness | Age_Group |
|--------|--------|--------|--------|-----|---------|---------|-----------|
| 0      | 1      | Dallas | Male   | 41  | 40367.0 | No      | mid       |
| 2      | 3      | Dallas | Male   | 42  | 52483.0 | No      | mid       |
| 4      | 5      | Dallas | Male   | 46  | 50289.0 | No      | mid       |
| 10     | 11     | Dallas | Female | 48  | 41524.0 | Yes     | mid       |
| 11     | 12     | Dallas | Male   | 47  | 54777.0 | No      | mid       |
| ...    | ...    | ...    | ...    | ... | ...     | ...     | ...       |
| 149976 | 149977 | Austin | Male   | 49  | 91519.0 | No      | mid       |
| 149979 | 149980 | Austin | Male   | 43  | 86737.0 | No      | mid       |
| 149981 | 149982 | Austin | Male   | 50  | 77134.0 | No      | mid       |
| 149982 | 149983 | Austin | Female | 47  | 92157.0 | No      | mid       |
| 149995 | 149996 | Austin | Male   | 48  | 93669.0 | No      | mid       |

```
[37613 rows x 7 columns]
```

```
[26]: mid_filtered.describe()
```

```
[26]:
```

|       | Number        | Age          | Income        |
|-------|---------------|--------------|---------------|
| count | 37613.000000  | 37613.000000 | 37613.000000  |
| mean  | 74850.556855  | 45.472071    | 91166.036716  |
| std   | 43376.260062  | 2.877640     | 24969.662567  |
| min   | 1.000000      | 41.000000    | 2474.000000   |
| 25%   | 37082.000000  | 43.000000    | 80840.000000  |
| 50%   | 74777.000000  | 45.000000    | 93617.000000  |
| 75%   | 112456.000000 | 48.000000    | 104380.000000 |
| max   | 149996.000000 | 50.000000    | 176746.000000 |

```
[27]: old_filtered = frame[~frame['Age_Group'].isin(['young', 'mid'])]
```

```
[28]: old_filtered
```

```
[28]:
```

|        | Number | City   | Gender | Age | Income  | Illness | Age_Group |
|--------|--------|--------|--------|-----|---------|---------|-----------|
| 1      | 2      | Dallas | Male   | 54  | 45084.0 | No      | old       |
| 8      | 9      | Dallas | Male   | 51  | 68667.0 | No      | old       |
| 14     | 15     | Dallas | Female | 61  | 38429.0 | No      | old       |
| 20     | 21     | Dallas | Male   | 57  | 17378.0 | No      | old       |
| 24     | 25     | Dallas | Male   | 58  | 55190.0 | No      | old       |
| ...    | ...    | ...    | ...    | ... | ...     | ...     | ...       |
| 149986 | 149987 | Austin | Female | 63  | 80381.0 | No      | old       |
| 149987 | 149988 | Austin | Female | 55  | 62501.0 | No      | old       |
| 149989 | 149990 | Austin | Male   | 52  | 83688.0 | No      | old       |
| 149991 | 149992 | Austin | Male   | 51  | 97510.0 | No      | old       |
| 149993 | 149994 | Austin | Male   | 64  | 89906.0 | No      | old       |

[54132 rows x 7 columns]

```
[29]: old_filtered.describe()
```

```
[29]:
```

|       | Number        | Age          | Income        |
|-------|---------------|--------------|---------------|
| count | 54132.000000  | 54132.000000 | 54132.000000  |
| mean  | 74849.578438  | 57.769785    | 91226.984667  |
| std   | 43209.231986  | 4.204792     | 25038.947447  |
| min   | 2.000000      | 51.000000    | 4371.000000   |
| 25%   | 37361.750000  | 54.000000    | 80838.000000  |
| 50%   | 74958.500000  | 58.000000    | 93640.000000  |
| 75%   | 112215.500000 | 61.000000    | 104544.250000 |
| max   | 149994.000000 | 65.000000    | 177157.000000 |