```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from sklearn.linear_model import LinearRegression,Ridge
from sklearn.ensemble import GradientBoostingRegressor

path=r"C:\Users\1911s\Downloads\
Healthcare_Insurance_Analysis_Datasets\Hospitalisation details.csv"
df1=pd.read_csv(path)
print('data loaded')
```

```
data loaded
```

```
df1.head()
```

```
  Customer ID  year month  date  children  charges Hospital tier City
tier  \
0      Id2335  1992   Jul     9         0   563.84      tier - 2  tier
- 3
1      Id2334  1992   Nov    30         0   570.62      tier - 2  tier
- 1
2      Id2333  1993   Jun    30         0   600.00      tier - 2  tier
- 1
3      Id2332  1992   Sep    13         0   604.54      tier - 3  tier
- 3
4      Id2331  1998   Jul    27         0   637.26      tier - 3  tier
- 3

  State ID
0    R1013
1    R1013
2    R1013
3    R1013
4    R1013
```

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2343 entries, 0 to 2342
Data columns (total 9 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Customer ID    2343 non-null   object
 1   year           2343 non-null   object
 2   month          2343 non-null   object
 3   date           2343 non-null   int64
 4   children       2343 non-null   int64
 5   charges        2343 non-null   float64
 6   Hospital tier  2343 non-null   object
 7   City tier      2343 non-null   object
 8   State ID       2343 non-null   object
```

```
dtypes: float64(1), int64(2), object(6)
memory usage: 164.9+ KB
```

```
dup=df1[df1.duplicated()]
dup
```

```
Empty DataFrame
Columns: [Customer ID, year, month, date, children, charges, Hospital
tier, City tier, State ID]
Index: []
```

```
path2=r"C:\Users\1911s\Downloads\
Healthcare_Insurance_Analysis_Datasets\Medical Examinations.csv"
df2=pd.read_csv(path2)
print('data loaded')
```

```
data loaded
```

```
df2.head()
```

```
  Customer ID     BMI  HBA1C Heart Issues Any Transplants Cancer
history  \
0         Id1  47.410   7.47           No              No
No
1         Id2  30.360   5.77           No              No
No
2         Id3  34.485  11.87          yes              No
No
3         Id4  38.095   6.05           No              No
No
4         Id5  35.530   5.45           No              No
No

  NumberOfMajorSurgeries smoker
0       No major surgery    yes
1       No major surgery    yes
2                      2    yes
3       No major surgery    yes
4       No major surgery    yes
```

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2335 entries, 0 to 2334
Data columns (total 8 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             2335 non-null   object
 1   BMI                     2335 non-null   float64
 2   HBA1C                   2335 non-null   float64
 3   Heart Issues            2335 non-null   object
```

```
 4   Any Transplants         2335 non-null   object
 5   Cancer history          2335 non-null   object
 6   NumberOfMajorSurgeries  2335 non-null   object
 7   smoker                  2335 non-null   object
dtypes: float64(2), object(6)
memory usage: 146.1+ KB
```

```
pip install pandas openpyxl
```

```
Requirement already satisfied: pandas in c:\users\1911s\appdata\local\
programs\python\python312\lib\site-packages (2.2.2)Note: you may need
to restart the kernel to use updated packages.

Requirement already satisfied: openpyxl in c:\users\1911s\appdata\
local\programs\python\python312\lib\site-packages (3.1.5)
Requirement already satisfied: numpy>=1.26.0 in c:\users\1911s\
appdata\local\programs\python\python312\lib\site-packages (from
pandas) (2.1.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\
1911s\appdata\local\programs\python\python312\lib\site-packages (from
pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\1911s\appdata\
local\programs\python\python312\lib\site-packages (from pandas)
(2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\1911s\
appdata\local\programs\python\python312\lib\site-packages (from
pandas) (2024.1)
Requirement already satisfied: et-xmlfile in c:\users\1911s\appdata\
local\programs\python\python312\lib\site-packages (from openpyxl)
(2.0.0)
Requirement already satisfied: six>=1.5 in c:\users\1911s\appdata\
local\programs\python\python312\lib\site-packages (from python-
dateutil>=2.8.2->pandas) (1.16.0)


[notice] A new release of pip is available: 24.2 -> 25.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

```
difference=df1[~df1['Customer ID'].isin(df2['Customer ID'])]
difference
```

```
     Customer ID  year month  date  children   charges Hospital
tier  \
45             ?  2004   Nov     6         0   1137.01      tier - 3

294            ?  1999   Jun     9         1   2775.19      tier - 2

731            ?  1985   Dec    20         2   6203.90      tier - 1

1863      id2444  1987   Nov    27         2  20984.09      tier - 2
```

```
2098      id3444  2004   Nov     1         2  34303.17      tier - 1

2129           ?  2000   Oct    13         0  35585.58      tier - 1

2158           ?  1992   Oct     6         0  36837.47      tier - 1

2202           ?  1991   Nov    22         2  38711.00      tier - 1


     City tier State ID
45    tier - 1    R1013
294   tier - 1    R1012
731   tier - 2    R1012
1863  tier - 2    R1015
2098  tier - 3    R1013
2129  tier - 2    R1011
2158  tier - 2    R1011
2202  tier - 3    R1011
```

```
df1=df1[df1['Customer ID'].isin(df2['Customer ID'])]
df1.head()
```

```
  Customer ID  year month  date  children  charges Hospital tier City
tier  \
0      Id2335  1992   Jul     9         0   563.84      tier - 2  tier
- 3
1      Id2334  1992   Nov    30         0   570.62      tier - 2  tier
- 1
2      Id2333  1993   Jun    30         0   600.00      tier - 2  tier
- 1
3      Id2332  1992   Sep    13         0   604.54      tier - 3  tier
- 3
4      Id2331  1998   Jul    27         0   637.26      tier - 3  tier
- 3

  State ID
0    R1013
1    R1013
2    R1013
3    R1013
4    R1013
```

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2335 entries, 0 to 2342
Data columns (total 9 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Customer ID    2335 non-null   object
 1   year           2335 non-null   object
```

```
 2   month          2335 non-null   object
 3   date           2335 non-null   int64
 4   children       2335 non-null   int64
 5   charges        2335 non-null   float64
 6   Hospital tier  2335 non-null   object
 7   City tier      2335 non-null   object
 8   State ID       2335 non-null   object
dtypes: float64(1), int64(2), object(6)
memory usage: 182.4+ KB
```

```
path3=r"C:\Users\1911s\Downloads\
Healthcare_Insurance_Analysis_Datasets\Names.xlsx"
df3=pd.read_excel(path3)
print('data loaded')
```

```
data loaded
```

```
df3.head()
```

```
  Customer ID                      name
0         Id1       Hawks, Ms.  Kelly
1         Id2  Lehner, Mr.  Matthew D
2         Id3           Lu, Mr.  Phil
3         Id4    Osborne, Ms.  Kelsey
4         Id5    Kadala, Ms.  Kristyn
```

```
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2335 entries, 0 to 2334
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Customer ID  2335 non-null   object
 1   name         2335 non-null   object
dtypes: object(2)
memory usage: 36.6+ KB
```

```
mergedata=pd.merge(df1,df2,on='Customer ID',how='left')
mergedata.head()
```

```
  Customer ID  year month  date  children  charges Hospital tier City
tier  \
0      Id2335  1992   Jul     9         0   563.84      tier - 2  tier
- 3
1      Id2334  1992   Nov    30         0   570.62      tier - 2  tier
- 1
2      Id2333  1993   Jun    30         0   600.00      tier - 2  tier
- 1
3      Id2332  1992   Sep    13         0   604.54      tier - 3  tier
- 3
```

```
4      Id2331  1998   Jul    27         0   637.26      tier - 3  tier
- 3

  State ID    BMI  HBA1C Heart Issues Any Transplants Cancer
history  \
0    R1013  17.58   4.51           No               No             No

1    R1013  17.60   4.39           No               No             No

2    R1013  16.47   6.35           No               No            Yes

3    R1013  17.70   6.28           No               No             No

4    R1013  22.34   5.57           No               No             No


  NumberOfMajorSurgeries smoker
0                      1     No
1                      1     No
2                      1     No
3                      1     No
4                      1     No

mergedata.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2335 entries, 0 to 2334
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             2335 non-null   object
 1   year                    2335 non-null   object
 2   month                   2335 non-null   object
 3   date                    2335 non-null   int64
 4   children                2335 non-null   int64
 5   charges                 2335 non-null   float64
 6   Hospital tier           2335 non-null   object
 7   City tier               2335 non-null   object
 8   State ID                2335 non-null   object
 9   BMI                     2335 non-null   float64
 10  HBA1C                   2335 non-null   float64
 11  Heart Issues            2335 non-null   object
 12  Any Transplants         2335 non-null   object
 13  Cancer history          2335 non-null   object
 14  NumberOfMajorSurgeries  2335 non-null   object
 15  smoker                  2335 non-null   object
dtypes: float64(3), int64(2), object(11)
memory usage: 292.0+ KB

mergedata.tail()
```

```
     Customer ID  year month  date  children   charges Hospital
tier  \
2330         Id5  1989   Jun    19         0  55135.40      tier - 1

2331         Id4  1991   Jun     6         1  58571.07      tier - 1

2332         Id3  1970     ?    11         3  60021.40      tier - 1

2333         Id2  1977   Jun     8         0  62592.87      tier - 2

2334         Id1  1968   Oct    12         0  63770.43      tier - 1


     City tier State ID     BMI  HBA1C Heart Issues Any Transplants  \
2330  tier - 2    R1012  35.530   5.45           No               No
2331  tier - 3    R1024  38.095   6.05           No               No
2332  tier - 1    R1012  34.485  11.87          yes               No
2333  tier - 3    R1013  30.360   5.77           No               No
2334  tier - 3    R1013  47.410   7.47           No               No

     Cancer history NumberOfMajorSurgeries smoker
2330             No       No major surgery    yes
2331             No       No major surgery    yes
2332             No                      2    yes
2333             No       No major surgery    yes
2334             No       No major surgery    yes
```

```
data=pd.merge(df3,mergedata,on='Customer ID',how='right')
data.head()
```

```
  Customer ID                                name  year month  date
children  \
0      Id2335               German, Mr.  Aaron K  1992   Jul     9
0
1      Id2334             Rosendahl, Mr.  Evan P  1992   Nov    30
0
2      Id2333                 Albano, Ms.  Julie  1993   Jun    30
0
3      Id2332  Riveros Gonzalez, Mr.  Juan D. Sr.  1992   Sep    13
0
4      Id2331              Brietzke, Mr.  Jordan  1998   Jul    27
0


   charges Hospital tier City tier State ID    BMI  HBA1C Heart Issues
\
0   563.84      tier - 2  tier - 3    R1013  17.58   4.51           No

1   570.62      tier - 2  tier - 1    R1013  17.60   4.39           No

2   600.00      tier - 2  tier - 1    R1013  16.47   6.35           No
```

```
3   604.54      tier - 3  tier - 3    R1013  17.70   6.28           No

4   637.26      tier - 3  tier - 3    R1013  22.34   5.57           No


  Any Transplants Cancer history NumberOfMajorSurgeries smoker
0              No             No                       1     No
1              No             No                       1     No
2              No            Yes                       1     No
3              No             No                       1     No
4              No             No                       1     No

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2335 entries, 0 to 2334
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             2335 non-null   object
 1   name                    2335 non-null   object
 2   year                    2335 non-null   object
 3   month                   2335 non-null   object
 4   date                    2335 non-null   int64
 5   children                2335 non-null   int64
 6   charges                 2335 non-null   float64
 7   Hospital tier           2335 non-null   object
 8   City tier               2335 non-null   object
 9   State ID                2335 non-null   object
 10  BMI                     2335 non-null   float64
 11  HBA1C                   2335 non-null   float64
 12  Heart Issues            2335 non-null   object
 13  Any Transplants         2335 non-null   object
 14  Cancer history          2335 non-null   object
 15  NumberOfMajorSurgeries  2335 non-null   object
 16  smoker                  2335 non-null   object
dtypes: float64(3), int64(2), object(12)
memory usage: 310.2+ KB

duplicate=data[data.duplicated()]
duplicate

Empty DataFrame
Columns: [Customer ID, name, year, month, date, children, charges,
Hospital tier, City tier, State ID, BMI, HBA1C, Heart Issues, Any
Transplants, Cancer history, NumberOfMajorSurgeries, smoker]
Index: []

data.isna().sum()
```

```
Customer ID                0
name                       0
year                       0
month                      0
date                       0
children                   0
charges                    0
Hospital tier              0
City tier                  0
State ID                   0
BMI                        0
HBA1C                      0
Heart Issues               0
Any Transplants            0
Cancer history             0
NumberOfMajorSurgeries     0
smoker                     0
dtype: int64
```

```
for col in data.columns:
    ##print(col)
    dataset=data[data.apply(lambda row:'?' not in row.values,axis=1)]

dataset
```

```
     Customer ID                                name  year month  date
\
0         Id2335                German, Mr.  Aaron K  1992   Jul     9

1         Id2334               Rosendahl, Mr.  Evan P  1992   Nov    30

2         Id2333                  Albano, Ms.  Julie  1993   Jun    30

3         Id2332  Riveros Gonzalez, Mr.  Juan D. Sr.  1992   Sep    13

4         Id2331                Brietzke, Mr.  Jordan  1998   Jul    27

...          ...                                 ...   ...   ...   ...

2329         Id6               Baker, Mr.  Russell B.  1962   Aug     4

2330         Id5                 Kadala, Ms.  Kristyn  1989   Jun    19

2331         Id4                 Osborne, Ms.  Kelsey  1991   Jun     6

2333         Id2               Lehner, Mr.  Matthew D  1977   Jun     8

2334         Id1                    Hawks, Ms.  Kelly  1968   Oct    12


      children   charges Hospital tier City tier State ID     BMI
```

```
HBA1C  \
0            0    563.84      tier - 2  tier - 3    R1013  17.580
4.51
1            0    570.62      tier - 2  tier - 1    R1013  17.600
4.39
2            0    600.00      tier - 2  tier - 1    R1013  16.470
6.35
3            0    604.54      tier - 3  tier - 3    R1013  17.700
6.28
4            0    637.26      tier - 3  tier - 3    R1013  22.340
5.57
...        ...       ...           ...       ...      ...     ...    .
..
2329         0  52590.83      tier - 1  tier - 3    R1011  32.800
6.59
2330         0  55135.40      tier - 1  tier - 2    R1012  35.530
5.45
2331         1  58571.07      tier - 1  tier - 3    R1024  38.095
6.05
2333         0  62592.87      tier - 2  tier - 3    R1013  30.360
5.77
2334         0  63770.43      tier - 1  tier - 3    R1013  47.410
7.47

     Heart Issues Any Transplants Cancer history
NumberOfMajorSurgeries smoker
0              No               No             No
1     No
1              No               No             No
1     No
2              No               No            Yes
1     No
3              No               No             No
1     No
4              No               No             No
1     No
...           ...              ...            ...                    ..
.    ...
2329           No               No             No       No major
surgery    yes
2330           No               No             No       No major
surgery    yes
2331           No               No             No       No major
surgery    yes
2333           No               No             No       No major
surgery    yes
2334           No               No             No       No major
surgery    yes
```

```
[2325 rows x 17 columns]
```

```
for col in dataset.columns:
    print(col)
    print(dataset[col].unique())
```

```
Customer ID
['Id2335' 'Id2334' 'Id2333' ... 'Id4' 'Id2' 'Id1']
name
['German, Mr.  Aaron K' 'Rosendahl, Mr.  Evan P' 'Albano, Ms.
Julie' ...
 'Osborne, Ms.  Kelsey' 'Lehner, Mr.  Matthew D' 'Hawks, Ms.  Kelly']
year
['1992' '1993' '1998' '2001' '1995' '2002' '1997' '1999' '1996' '2004'
 '2000' '1994' '2003' '1989' '1991' '1988' '1990' '1987' '1986' '1981'
 '1985' '1984' '1983' '1982' '1980' '1979' '1977' '1978' '1974' '1973'
 '1972' '1975' '1971' '1970' '1969' '1976' '1968' '1966' '1965' '1967'
 '1962' '1963' '1964' '1961' '1960' '1959' '1958']
month
['Jul' 'Nov' 'Jun' 'Sep' 'Dec' 'Aug' 'Oct']
date
[ 9 30 13 27 20  1  4 29 12 14 22 28  7 18 19  5 17 24 26  6 11 15 25
10
  2 16  8 21  3 23]
children
[0 1 3 2 4 5]
charges
[  563.84   570.62   600.   ... 58571.07 62592.87 63770.43]
Hospital tier
['tier - 2' 'tier - 3' 'tier - 1']
City tier
['tier - 3' 'tier - 1' 'tier - 2']
State ID
['R1013' 'R1012' 'R1011' 'R1015' 'R1019' 'R1016' 'R1018' 'R1025'
'R1024'
 'R1023' 'R1014' 'R1021' 'R1017' 'R1020' 'R1026' 'R1022']
BMI
[17.58 17.6  16.47 ... 36.96 36.4  47.41]
HBA1C
[ 4.51  4.39  6.35  6.28  5.57  4.29  5.22  5.26 10.67  5.6   4.54
5.94
  5.43  5.53  5.73  5.81  5.91  5.29  6.11  5.42  4.2   5.64  5.27
5.8
  4.36  4.45  4.86  8.48 11.46  6.05  5.4   5.5   5.33  5.28  8.46
4.37
  5.49  9.13  5.71  6.08 10.36  4.07  4.27  4.08  5.03  7.73  6.02
4.15
  5.44  4.02  4.44 10.95  4.96  5.13  6.15  9.44  4.42  5.45  4.52
6.21
```

```
  4.7   4.76  5.51  7.03  5.87  8.59  5.12  6.37  4.49  4.01  4.99
4.77
  4.13  5.7   9.46  4.92  5.85  4.1   5.56  4.03 11.11  7.4   4.83
6.24
  4.12  5.65  4.32 10.01  6.22 11.84  9.79  4.38  4.11  4.19  6.29
5.59
  4.5   4.3   5.96  5.24  4.55  5.84  5.08  4.59  6.01  4.28  6.09
5.68
  5.39  5.74  5.52  4.17  4.41  5.99  6.    5.05  4.69  5.06  5.62
5.18
  5.54  4.04  5.97  5.    5.14  5.32  4.26  4.78 11.92  4.58  5.36
6.19
  4.87  6.14  4.43 11.03 11.19 10.16  9.49  4.34  4.47  4.97  9.5
9.88
  4.24  9.27  6.45 11.15  4.33  4.48  4.89  4.23  5.93  4.95  5.47
6.41
  4.56  6.06  6.1   6.2   4.66  5.09  4.21  4.79  5.04  4.6   4.46
5.11
  7.62 11.05  6.12  4.64  5.92  6.66  6.16  8.82  6.04 11.62  6.43
4.81
  5.88  5.61  6.03  6.3   6.17  4.68  4.75  5.25 11.02  8.93  4.73
4.57
  4.65  6.32  5.3   5.83  5.55  4.8   4.91  4.84  9.9   4.4   4.18
9.39
  7.09  5.34  4.    5.63  4.25  5.72  6.26  4.14  4.62  5.16  4.31
4.09
  5.2   7.66  5.95  6.25  4.67  4.72  6.23  5.21  8.4   6.8   5.46
5.98
  8.75  6.38  6.91  4.98 10.27 11.47  8.63  6.49  5.01  9.07  6.18
5.17
 10.97  5.79  5.77  8.94  4.82  5.82  4.05  5.78  4.9   5.69  6.77
9.62
  5.19  5.86  5.37  6.07  4.06  6.46  5.1   4.16  6.6  11.69 11.48
9.25
 11.66  9.99  4.93 10.93  5.07 11.07 11.61 11.72  5.15 10.34  4.74
5.31
  6.72  6.44  8.43  4.63  4.71  5.38  8.55  8.66  7.63  9.6   5.76
11.56
  9.54  4.35  7.81 10.45  6.64 11.73  7.32  9.12  9.2   8.62  7.41
7.94
 10.66 10.9   9.92  9.96 10.14  5.75  6.27  6.9  11.68 11.96  5.48
11.75
  5.67  6.89  9.61 11.85  7.59 11.63 10.85  5.41  6.94 10.42  9.31
11.41
  9.72  5.58 11.09  4.85  9.71  4.88  5.66  5.89 11.18  8.18 11.04
6.98
  6.57  4.61  8.53  8.5   9.21  6.93 10.84  7.02  8.41 11.93 11.17
8.6
 10.83  6.59 11.98 10.59 11.58 10.74 10.09  5.35  7.46 11.81  8.7
```

```
7.5
  8.27  9.17  7.48 11.37 11.44 10.81 10.56  9.55  6.69 10.53  9.14
9.16
  8.12  5.9  11.4   8.87  9.28  9.91 10.82 11.71  8.    8.22  8.19
7.04
  7.8  10.75 11.2   7.85 10.68 11.06  8.84  6.5   7.95 11.08  9.59
7.98
  7.57  6.54  6.74  8.06  6.81 11.35  6.36 11.39 10.08  7.67  9.45
7.93
  8.8   8.57 12.    9.89  9.23 11.24 10.1   7.61  8.17  9.85 11.82
8.38
 11.94  6.53  7.71 10.24 10.98  9.47 10.54  7.99  8.76  8.33 10.46
8.29
  9.04 11.78 10.18  8.51  8.45  8.68 11.49  7.96 11.51  6.68  8.13
7.34
  5.23  8.96  7.58 11.8  10.37  9.63  8.85  6.13 11.3   7.42  8.05
6.79
  9.94 11.52  8.58  8.92  6.96 10.17  7.83  7.45  9.48  6.4   9.58
10.72
  7.19  9.8   7.72  6.7   7.33 11.42  7.56  7.53 11.36 10.79 10.19
11.76
  9.08  9.05  9.77  7.69  8.26 11.31  8.34  8.21  7.37  6.99  9.81
4.94
  8.3   9.03  7.2   8.02  9.18 11.22  7.6   7.74  7.16 11.83  7.91
10.12
  7.92 10.57  7.23  7.05  9.29  6.31  8.31  7.26 11.1  10.78  9.65
10.52
 11.5   9.1   8.83  8.69  8.24 10.02 10.7  10.62 11.21  7.82 10.58
8.47
  8.81 11.    8.25 11.86  8.95  8.07  7.51  7.68  7.21 10.8   8.77
11.33
 10.64  9.19  9.32  9.24 10.22  9.4   9.56  6.65  9.22  7.49  9.06
10.48
 11.38  7.86  7.06  7.12  7.79 11.14 10.3  11.87  9.02  8.16  7.
6.55
  7.65 11.01  9.83  8.11 10.49  8.08 10.33  8.23  8.74  5.02 11.25
10.4
  7.39  9.53 11.88  6.71 11.97 10.2   7.11  7.76  7.18  4.22  9.66
10.99
  8.28  7.3  11.89  6.83 11.43 10.87 10.44  9.95  9.36  9.51  8.64
4.53
 11.34 10.43  6.47  6.87  7.31  8.37  6.84 10.55  8.9  11.32  8.1
6.76
  6.39 10.5   8.56 11.64  6.85  9.52  6.33 10.47  6.97  8.71 11.95
10.96
  8.86  6.48 10.21  9.86 11.74 10.86 10.25  9.7  11.99  7.84 11.55
10.6
 10.38  9.42  9.68  8.72 11.12 10.73  7.22  9.26  6.75 10.04  8.01
7.54
```

```
  6.92  9.78 10.51  6.67 11.57 10.92  9.33 10.61  9.34  8.49  6.52
11.91
  6.86  8.44  7.01  7.89  7.29 11.9   7.47]
Heart Issues
['No' 'yes']
Any Transplants
['No' 'yes']
Cancer history
['No' 'Yes']
NumberOfMajorSurgeries
['1' 'No major surgery' '2' '3']
smoker
['No' 'yes']
```

```
pd.options.display.float_format='{:,.4f}'.format
variance=dataset.var(numeric_only=True).reset_index()
variance.columns=['Feature','Variance']
variance
```

```
    Feature         Variance
0      date          76.0434
1  children           1.5239
2   charges 140,742,458.9775
3       BMI          76.4739
4     HBA1C           4.9590
```

```
categorical=dataset.select_dtypes(exclude='number')
categorical=categorical.drop(columns=['year','Hospital tier','City
tier','month','State ID','Customer ID','name'],axis=1)
categorical
```

```
     Heart Issues Any Transplants Cancer history
NumberOfMajorSurgeries smoker
0              No              No             No
1     No
1              No              No             No
1     No
2              No              No            Yes
1     No
3              No              No             No
1     No
4              No              No             No
1     No
...           ...             ...            ...                      ..
.    ...
2329           No              No             No       No major
surgery    yes
2330           No              No             No       No major
surgery    yes
2331           No              No             No       No major
```

```
surgery    yes
2333           No               No              No       No major
surgery    yes
2334           No               No              No       No major
surgery    yes

[2325 rows x 5 columns]
```

```
categorical['NumberOfMajorSurgeries'].unique()
```

```
array(['1', 'No major surgery', '2', '3'], dtype=object)
```

```
for col in categorical.columns:
    print(col)
    print(categorical[col].unique())
```

```
Heart Issues
['No' 'yes']
Any Transplants
['No' 'yes']
Cancer history
['No' 'Yes']
NumberOfMajorSurgeries
['1' 'No major surgery' '2' '3']
smoker
['No' 'yes']
```

```
numeric=dataset.select_dtypes(include='number')
numeric
```

```
      date  children     charges     BMI  HBA1C
0        9         0    563.8400 17.5800 4.5100
1       30         0    570.6200 17.6000 4.3900
2       30         0    600.0000 16.4700 6.3500
3       13         0    604.5400 17.7000 6.2800
4       27         0    637.2600 22.3400 5.5700
...    ...       ...         ...     ...    ...
2329     4         0 52,590.8300 32.8000 6.5900
2330    19         0 55,135.4000 35.5300 5.4500
2331     6         1 58,571.0700 38.0950 6.0500
2333     8         0 62,592.8700 30.3600 5.7700
2334    12         0 63,770.4300 47.4100 7.4700

[2325 rows x 5 columns]
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()

for col in categorical.columns:
    dataset[col]=le.fit_transform(dataset[col])
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1595813537.py:2: 
SettingWithCopyWarning: 
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: 
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  dataset[col]=le.fit_transform(dataset[col])
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1595813537.py:2: 
SettingWithCopyWarning: 
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: 
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  dataset[col]=le.fit_transform(dataset[col])
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1595813537.py:2: 
SettingWithCopyWarning: 
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: 
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  dataset[col]=le.fit_transform(dataset[col])
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1595813537.py:2: 
SettingWithCopyWarning: 
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: 
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  dataset[col]=le.fit_transform(dataset[col])
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1595813537.py:2: 
SettingWithCopyWarning: 
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: 
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  dataset[col]=le.fit_transform(dataset[col])

dataset.head()

  Customer ID                                name  year month  date  
children  \
```

```
0      Id2335                German, Mr.  Aaron K  1992   Jul     9
0
1      Id2334              Rosendahl, Mr.  Evan P  1992   Nov    30
0
2      Id2333                  Albano, Ms.  Julie  1993   Jun    30
0
3      Id2332  Riveros Gonzalez, Mr.  Juan D. Sr.  1992   Sep    13
0
4      Id2331               Brietzke, Mr.  Jordan  1998   Jul    27
0

   charges Hospital tier City tier State ID     BMI  HBA1C  Heart
Issues  \
0 563.8400      tier - 2  tier - 3    R1013 17.5800 4.5100
0
1 570.6200      tier - 2  tier - 1    R1013 17.6000 4.3900
0
2 600.0000      tier - 2  tier - 1    R1013 16.4700 6.3500
0
3 604.5400      tier - 3  tier - 3    R1013 17.7000 6.2800
0
4 637.2600      tier - 3  tier - 3    R1013 22.3400 5.5700
0

   Any Transplants  Cancer history  NumberOfMajorSurgeries  smoker
0                0               0                       0       0
1                0               0                       0       0
2                0               1                       0       0
3                0               0                       0       0
4                0               0                       0       0

dataset.info()

<class 'pandas.core.frame.DataFrame'>
Index: 2325 entries, 0 to 2334
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             2325 non-null   object
 1   name                    2325 non-null   object
 2   year                    2325 non-null   object
 3   month                   2325 non-null   object
 4   date                    2325 non-null   int64
 5   children                2325 non-null   int64
 6   charges                 2325 non-null   float64
 7   Hospital tier           2325 non-null   object
 8   City tier               2325 non-null   object
 9   State ID                2325 non-null   object
 10  BMI                     2325 non-null   float64
 11  HBA1C                   2325 non-null   float64
```

```
 12  Heart Issues             2325 non-null   int64
 13  Any Transplants          2325 non-null   int64
 14  Cancer history           2325 non-null   int64
 15  NumberOfMajorSurgeries   2325 non-null   int64
 16  smoker                   2325 non-null   int64
dtypes: float64(3), int64(7), object(7)
memory usage: 327.0+ KB
```

```
dataset['State ID'].unique()
```

```
array(['R1013', 'R1012', 'R1011', 'R1015', 'R1019', 'R1016', 'R1018',
       'R1025', 'R1024', 'R1023', 'R1014', 'R1021', 'R1017', 'R1020',
       'R1026', 'R1022'], dtype=object)
```

```
for col in dataset.columns:
    print(col)
    print(dataset[col].unique())
```

```
Customer ID
['Id2335' 'Id2334' 'Id2333' ... 'Id4' 'Id2' 'Id1']
name
['German, Mr.  Aaron K' 'Rosendahl, Mr.  Evan P' 'Albano, Ms.
Julie' ...
 'Osborne, Ms.  Kelsey' 'Lehner, Mr.  Matthew D' 'Hawks, Ms.  Kelly']
year
['1992' '1993' '1998' '2001' '1995' '2002' '1997' '1999' '1996' '2004'
 '2000' '1994' '2003' '1989' '1991' '1988' '1990' '1987' '1986' '1981'
 '1985' '1984' '1983' '1982' '1980' '1979' '1977' '1978' '1974' '1973'
 '1972' '1975' '1971' '1970' '1969' '1976' '1968' '1966' '1965' '1967'
 '1962' '1963' '1964' '1961' '1960' '1959' '1958']
month
['Jul' 'Nov' 'Jun' 'Sep' 'Dec' 'Aug' 'Oct']
date
[ 9 30 13 27 20  1  4 29 12 14 22 28  7 18 19  5 17 24 26  6 11 15 25
10
  2 16  8 21  3 23]
children
[0 1 3 2 4 5]
charges
[  563.84   570.62   600.   ... 58571.07 62592.87 63770.43]
Hospital tier
['tier - 2' 'tier - 3' 'tier - 1']
City tier
['tier - 3' 'tier - 1' 'tier - 2']
State ID
['R1013' 'R1012' 'R1011' 'R1015' 'R1019' 'R1016' 'R1018' 'R1025'
'R1024'
 'R1023' 'R1014' 'R1021' 'R1017' 'R1020' 'R1026' 'R1022']
BMI
[17.58 17.6  16.47 ... 36.96 36.4  47.41]
```

```
HBA1C
[ 4.51  4.39  6.35  6.28  5.57  4.29  5.22  5.26 10.67  5.6   4.54
5.94
  5.43  5.53  5.73  5.81  5.91  5.29  6.11  5.42  4.2   5.64  5.27
5.8
  4.36  4.45  4.86  8.48 11.46  6.05  5.4   5.5   5.33  5.28  8.46
4.37
  5.49  9.13  5.71  6.08 10.36  4.07  4.27  4.08  5.03  7.73  6.02
4.15
  5.44  4.02  4.44 10.95  4.96  5.13  6.15  9.44  4.42  5.45  4.52
6.21
  4.7   4.76  5.51  7.03  5.87  8.59  5.12  6.37  4.49  4.01  4.99
4.77
  4.13  5.7   9.46  4.92  5.85  4.1   5.56  4.03 11.11  7.4   4.83
6.24
  4.12  5.65  4.32 10.01  6.22 11.84  9.79  4.38  4.11  4.19  6.29
5.59
  4.5   4.3   5.96  5.24  4.55  5.84  5.08  4.59  6.01  4.28  6.09
5.68
  5.39  5.74  5.52  4.17  4.41  5.99  6.    5.05  4.69  5.06  5.62
5.18
  5.54  4.04  5.97  5.    5.14  5.32  4.26  4.78 11.92  4.58  5.36
6.19
  4.87  6.14  4.43 11.03 11.19 10.16  9.49  4.34  4.47  4.97  9.5
9.88
  4.24  9.27  6.45 11.15  4.33  4.48  4.89  4.23  5.93  4.95  5.47
6.41
  4.56  6.06  6.1   6.2   4.66  5.09  4.21  4.79  5.04  4.6   4.46
5.11
  7.62 11.05  6.12  4.64  5.92  6.66  6.16  8.82  6.04 11.62  6.43
4.81
  5.88  5.61  6.03  6.3   6.17  4.68  4.75  5.25 11.02  8.93  4.73
4.57
  4.65  6.32  5.3   5.83  5.55  4.8   4.91  4.84  9.9   4.4   4.18
9.39
  7.09  5.34  4.    5.63  4.25  5.72  6.26  4.14  4.62  5.16  4.31
4.09
  5.2   7.66  5.95  6.25  4.67  4.72  6.23  5.21  8.4   6.8   5.46
5.98
  8.75  6.38  6.91  4.98 10.27 11.47  8.63  6.49  5.01  9.07  6.18
5.17
 10.97  5.79  5.77  8.94  4.82  5.82  4.05  5.78  4.9   5.69  6.77
9.62
  5.19  5.86  5.37  6.07  4.06  6.46  5.1   4.16  6.6  11.69 11.48
9.25
 11.66  9.99  4.93 10.93  5.07 11.07 11.61 11.72  5.15 10.34  4.74
5.31
  6.72  6.44  8.43  4.63  4.71  5.38  8.55  8.66  7.63  9.6   5.76
11.56
```

```
  9.54  4.35  7.81 10.45  6.64 11.73  7.32  9.12  9.2   8.62  7.41
7.94
 10.66 10.9   9.92  9.96 10.14  5.75  6.27  6.9  11.68 11.96  5.48
11.75
  5.67  6.89  9.61 11.85  7.59 11.63 10.85  5.41  6.94 10.42  9.31
11.41
  9.72  5.58 11.09  4.85  9.71  4.88  5.66  5.89 11.18  8.18 11.04
6.98
  6.57  4.61  8.53  8.5   9.21  6.93 10.84  7.02  8.41 11.93 11.17
8.6
 10.83  6.59 11.98 10.59 11.58 10.74 10.09  5.35  7.46 11.81  8.7
7.5
  8.27  9.17  7.48 11.37 11.44 10.81 10.56  9.55  6.69 10.53  9.14
9.16
  8.12  5.9  11.4   8.87  9.28  9.91 10.82 11.71  8.    8.22  8.19
7.04
  7.8  10.75 11.2   7.85 10.68 11.06  8.84  6.5   7.95 11.08  9.59
7.98
  7.57  6.54  6.74  8.06  6.81 11.35  6.36 11.39 10.08  7.67  9.45
7.93
  8.8   8.57 12.    9.89  9.23 11.24 10.1   7.61  8.17  9.85 11.82
8.38
 11.94  6.53  7.71 10.24 10.98  9.47 10.54  7.99  8.76  8.33 10.46
8.29
  9.04 11.78 10.18  8.51  8.45  8.68 11.49  7.96 11.51  6.68  8.13
7.34
  5.23  8.96  7.58 11.8  10.37  9.63  8.85  6.13 11.3   7.42  8.05
6.79
  9.94 11.52  8.58  8.92  6.96 10.17  7.83  7.45  9.48  6.4   9.58
10.72
  7.19  9.8   7.72  6.7   7.33 11.42  7.56  7.53 11.36 10.79 10.19
11.76
  9.08  9.05  9.77  7.69  8.26 11.31  8.34  8.21  7.37  6.99  9.81
4.94
  8.3   9.03  7.2   8.02  9.18 11.22  7.6   7.74  7.16 11.83  7.91
10.12
  7.92 10.57  7.23  7.05  9.29  6.31  8.31  7.26 11.1  10.78  9.65
10.52
 11.5   9.1   8.83  8.69  8.24 10.02 10.7  10.62 11.21  7.82 10.58
8.47
  8.81 11.    8.25 11.86  8.95  8.07  7.51  7.68  7.21 10.8   8.77
11.33
 10.64  9.19  9.32  9.24 10.22  9.4   9.56  6.65  9.22  7.49  9.06
10.48
 11.38  7.86  7.06  7.12  7.79 11.14 10.3  11.87  9.02  8.16  7.
6.55
  7.65 11.01  9.83  8.11 10.49  8.08 10.33  8.23  8.74  5.02 11.25
10.4
  7.39  9.53 11.88  6.71 11.97 10.2   7.11  7.76  7.18  4.22  9.66
```

```
10.99
  8.28  7.3  11.89  6.83 11.43 10.87 10.44  9.95  9.36  9.51  8.64
4.53
 11.34 10.43  6.47  6.87  7.31  8.37  6.84 10.55  8.9  11.32  8.1
6.76
  6.39 10.5   8.56 11.64  6.85  9.52  6.33 10.47  6.97  8.71 11.95
10.96
  8.86  6.48 10.21  9.86 11.74 10.86 10.25  9.7  11.99  7.84 11.55
10.6
 10.38  9.42  9.68  8.72 11.12 10.73  7.22  9.26  6.75 10.04  8.01
7.54
  6.92  9.78 10.51  6.67 11.57 10.92  9.33 10.61  9.34  8.49  6.52
11.91
  6.86  8.44  7.01  7.89  7.29 11.9   7.47]
Heart Issues
[0 1]
Any Transplants
[0 1]
Cancer history
[0 1]
NumberOfMajorSurgeries
[0 3 1 2]
smoker
[0 1]

filterdata=dataset[dataset['State
ID'].isin(['R1011','R1012','R1013'])]
filterdata

     Customer ID                                  name  year month  date
\
0         Id2335                 German, Mr.  Aaron K  1992   Jul     9

1         Id2334               Rosendahl, Mr.  Evan P  1992   Nov    30

2         Id2333                  Albano, Ms.  Julie  1993   Jun    30

3         Id2332  Riveros Gonzalez, Mr.  Juan D. Sr.  1992   Sep    13

4         Id2331                Brietzke, Mr.  Jordan  1998   Jul    27

...          ...                                  ...   ...   ...   ...

2328         Id7               Macpherson, Mr.  Scott  1994   Oct    27

2329         Id6               Baker, Mr.  Russell B.  1962   Aug     4

2330         Id5                 Kadala, Ms.  Kristyn  1989   Jun    19

2333         Id2               Lehner, Mr.  Matthew D  1977   Jun     8
```

```
2334         Id1                   Hawks, Ms.  Kelly  1968   Oct    12


      children     charges Hospital tier City tier State ID     BMI
HBA1C  \
0             0    563.8400      tier - 2  tier - 3    R1013 17.5800
4.5100
1             0    570.6200      tier - 2  tier - 1    R1013 17.6000
4.3900
2             0    600.0000      tier - 2  tier - 1    R1013 16.4700
6.3500
3             0    604.5400      tier - 3  tier - 3    R1013 17.7000
6.2800
4             0    637.2600      tier - 3  tier - 3    R1013 22.3400
5.5700
...         ...          ...           ...       ...      ...     ...
...
2328          1 51,194.5600      tier - 1  tier - 3    R1011 36.4000
6.0700
2329          0 52,590.8300      tier - 1  tier - 3    R1011 32.8000
6.5900
2330          0 55,135.4000      tier - 1  tier - 2    R1012 35.5300
5.4500
2333          0 62,592.8700      tier - 2  tier - 3    R1013 30.3600
5.7700
2334          0 63,770.4300      tier - 1  tier - 3    R1013 47.4100
7.4700

      Heart Issues  Any Transplants  Cancer history
NumberOfMajorSurgeries  \
0                0                 0               0
0
1                0                 0               0
0
2                0                 0               1
0
3                0                 0               0
0
4                0                 0               0
0
...            ...               ...             ...
...
2328             0                 0               0
3
2329             0                 0               0
3
2330             0                 0               0
3
2333             0                 0               0
3
```

```
2334             0                 0               0
3

      smoker
0          0
1          0
2          0
3          0
4          0
...      ...
2328       1
2329       1
2330       1
2333       1
2334       1

[1755 rows x 17 columns]
```

```
filterdata['State ID'].unique()
```

```
array(['R1013', 'R1012', 'R1011'], dtype=object)
```

```
filterdata['State ID']=le.fit_transform(filterdata['State ID'])
filterdata
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\2542322041.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata['State ID']=le.fit_transform(filterdata['State ID'])

     Customer ID                                  name  year month  date
\
0         Id2335                German, Mr.  Aaron K  1992   Jul     9

1         Id2334              Rosendahl, Mr.  Evan P  1992   Nov    30

2         Id2333                 Albano, Ms.  Julie  1993   Jun    30

3         Id2332  Riveros Gonzalez, Mr.  Juan D. Sr.  1992   Sep    13

4         Id2331               Brietzke, Mr.  Jordan  1998   Jul    27

...          ...                                 ...   ...   ...   ...

2328         Id7             Macpherson, Mr.  Scott  1994   Oct    27
```

```
2329         Id6              Baker, Mr.  Russell B.  1962   Aug     4

2330         Id5                Kadala, Ms.  Kristyn  1989   Jun    19

2333         Id2              Lehner, Mr.  Matthew D  1977   Jun     8

2334         Id1                  Hawks, Ms.  Kelly  1968   Oct    12


      children     charges Hospital tier City tier  State ID     BMI
HBA1C  \
0            0    563.8400      tier - 2  tier - 3         2 17.5800
4.5100
1            0    570.6200      tier - 2  tier - 1         2 17.6000
4.3900
2            0    600.0000      tier - 2  tier - 1         2 16.4700
6.3500
3            0    604.5400      tier - 3  tier - 3         2 17.7000
6.2800
4            0    637.2600      tier - 3  tier - 3         2 22.3400
5.5700
...        ...          ...           ...       ...       ...     ...
...
2328         1 51,194.5600      tier - 1  tier - 3         0 36.4000
6.0700
2329         0 52,590.8300      tier - 1  tier - 3         0 32.8000
6.5900
2330         0 55,135.4000      tier - 1  tier - 2         1 35.5300
5.4500
2333         0 62,592.8700      tier - 2  tier - 3         2 30.3600
5.7700
2334         0 63,770.4300      tier - 1  tier - 3         2 47.4100
7.4700

      Heart Issues  Any Transplants  Cancer history
NumberOfMajorSurgeries  \
0                0                 0               0
0
1                0                 0               0
0
2                0                 0               1
0
3                0                 0               0
0
4                0                 0               0
0
...            ...               ...             ...
...
2328             0                 0               0
3
```

```
2329              0                  0                 0
3
2330              0                  0                 0
3
2333              0                  0                 0
3
2334              0                  0                 0
3
```

```
      smoker
0          0
1          0
2          0
3          0
4          0
...      ...
2328       1
2329       1
2330       1
2333       1
2334       1

[1755 rows x 17 columns]
```

```
filterdata.rename(columns={'date':'day'},inplace=True)
filterdata
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\2268384193.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata.rename(columns={'date':'day'},inplace=True)
```

```
     Customer ID                                  name  year month  day
\
0         Id2335                German, Mr.  Aaron K  1992   Jul    9

1         Id2334              Rosendahl, Mr.  Evan P  1992   Nov   30

2         Id2333                  Albano, Ms.  Julie  1993   Jun   30

3         Id2332  Riveros Gonzalez, Mr.  Juan D. Sr.  1992   Sep   13

4         Id2331               Brietzke, Mr.  Jordan  1998   Jul   27

...          ...                                 ...   ...   ...  ...

2328         Id7              Macpherson, Mr.  Scott  1994   Oct   27
```

```
2329         Id6              Baker, Mr.  Russell B.  1962   Aug    4

2330         Id5                Kadala, Ms.  Kristyn  1989   Jun   19

2333         Id2              Lehner, Mr.  Matthew D  1977   Jun    8

2334         Id1                   Hawks, Ms.  Kelly  1968   Oct   12


      children     charges Hospital tier City tier  State ID     BMI
HBA1C  \
0            0    563.8400      tier - 2  tier - 3         2 17.5800
4.5100
1            0    570.6200      tier - 2  tier - 1         2 17.6000
4.3900
2            0    600.0000      tier - 2  tier - 1         2 16.4700
6.3500
3            0    604.5400      tier - 3  tier - 3         2 17.7000
6.2800
4            0    637.2600      tier - 3  tier - 3         2 22.3400
5.5700
...        ...          ...           ...       ...       ...     ...
...
2328         1 51,194.5600      tier - 1  tier - 3         0 36.4000
6.0700
2329         0 52,590.8300      tier - 1  tier - 3         0 32.8000
6.5900
2330         0 55,135.4000      tier - 1  tier - 2         1 35.5300
5.4500
2333         0 62,592.8700      tier - 2  tier - 3         2 30.3600
5.7700
2334         0 63,770.4300      tier - 1  tier - 3         2 47.4100
7.4700

      Heart Issues  Any Transplants  Cancer history
NumberOfMajorSurgeries  \
0                0                 0               0
0
1                0                 0               0
0
2                0                 0               1
0
3                0                 0               0
0
4                0                 0               0
0
...            ...               ...             ...
...
2328             0                 0               0
```

```
3
2329              0                 0                 0
3
2330              0                 0                 0
3
2333              0                 0                 0
3
2334              0                 0                 0
3

      smoker
0          0
1          0
2          0
3          0
4          0
...      ...
2328       1
2329       1
2330       1
2333       1
2334       1

[1755 rows x 17 columns]
```

```
for col in filterdata.columns:
    print(col)
    print(filterdata[col].unique())
```

```
Customer ID
['Id2335' 'Id2334' 'Id2333' ... 'Id5' 'Id2' 'Id1']
name
['German, Mr.  Aaron K' 'Rosendahl, Mr.  Evan P' 'Albano, Ms.
Julie' ...
 'Kadala, Ms.  Kristyn' 'Lehner, Mr.  Matthew D' 'Hawks, Ms.  Kelly']
year
['1992' '1993' '1998' '2001' '1995' '2002' '1997' '1999' '1996' '2004'
 '2000' '1994' '2003' '1989' '1991' '1988' '1990' '1987' '1986' '1981'
 '1985' '1984' '1983' '1982' '1980' '1979' '1977' '1978' '1974' '1973'
 '1972' '1975' '1971' '1970' '1969' '1976' '1968' '1966' '1965' '1967'
 '1962' '1963' '1964' '1961' '1960' '1959' '1958']
month
['Jul' 'Nov' 'Jun' 'Sep' 'Dec' 'Aug' 'Oct']
day
[ 9 30 13 27 20  1  4 29 12 14 22 28  7 18 19  5 17 24 26  6 11 15 25
10
  2 16  8 21  3 23]
children
[0 1 3 2 4 5]
charges
```

```
[  563.84   570.62   600.   ... 55135.4  62592.87 63770.43]
Hospital tier
['tier - 2' 'tier - 3' 'tier - 1']
City tier
['tier - 3' 'tier - 1' 'tier - 2']
State ID
[2 1 0]
BMI
[17.58 17.6  16.47 ... 36.96 36.4  47.41]
HBA1C
[ 4.51  4.39  6.35  6.28  5.57  4.29  5.22  5.26 10.67  5.6   4.54
5.94
  5.43  5.53  5.73  5.81  5.91  5.29  6.11  5.42  4.2   5.64  5.27
5.8
  4.36  4.45  4.86  8.48 11.46  6.05  5.4   5.5   5.33  5.28  8.46
4.37
  5.49  9.13  5.71  6.08 10.36  4.07  4.27  4.08  5.03  7.73  6.02
4.15
  5.44  4.02  4.44 10.95  4.96  5.13  6.15  9.44  4.42  5.45  4.52
6.21
  4.7   4.76  5.51  7.03  5.87  8.59  5.12  6.37  4.49  4.01  4.99
4.77
  4.13  5.7   9.46  4.92  5.85  4.1   5.56  4.03 11.11  7.4   4.83
6.24
  4.12  5.65  4.32 10.01  6.22 11.84  9.79  4.38  4.11  4.19  6.29
5.59
  4.5   4.3   5.96  5.24  4.55  5.84  5.08  4.59  6.01  4.28  6.09
5.68
  5.39  5.74  5.52  4.17  4.41  5.99  6.    5.05  5.62  5.54  4.04
5.97
  5.    5.14  5.32  4.26  4.78 11.92  4.58  5.36  6.19  4.87  4.43
11.03
 11.19 10.16  9.49  4.34  4.47  4.97  9.5   4.24  9.27  6.45  4.33
4.48
  4.89  4.23  5.93  4.95  5.47  6.41  4.56  6.06  6.1   6.2   4.66
4.79
  6.14  4.46  5.11  7.62 11.05  4.6   6.12  4.64  5.92  6.66  6.16
6.04
  6.43  4.81  5.88  5.61  6.03  6.3   6.17  4.68  4.75  5.25 11.02
4.73
  4.57  6.32  5.3   5.83  5.55  4.91  4.84  9.9   4.4   4.18  9.39
7.09
  4.    5.63  4.25  5.72  4.09  5.2   7.66  6.25  4.67  4.72  6.23
8.4
  6.8   5.46  5.98  8.75  6.91 10.27 11.47  8.63  9.07  6.18  5.17
10.97
  5.79  5.77  8.94  4.82  5.82  4.05  5.78  4.9   5.69  6.77  9.62
5.16
  5.19  5.37  4.65  6.07  6.26  5.86  4.06  5.21  4.16  6.6  11.69
```

```
11.48
  9.25 11.66  9.99  4.21 10.93  5.07 11.07 11.72  5.15 10.34  4.74
5.31
  6.72  6.44  8.43  5.38  8.55  8.66  7.63  9.6   5.76 11.56  9.54
4.35
  8.93  7.81 10.45 11.73  9.12  4.62  9.2   8.62  7.41  7.94  5.01
10.66
 10.9   9.92  9.96  6.38  8.82  5.75  6.9  11.68 11.96 11.75  5.67
6.89
  9.61 11.85  5.09  7.59 11.63 10.85  5.41  6.94  4.69 10.42  9.31
11.41
  9.72 11.09  4.85  9.71  4.88  5.66  5.89 11.18  8.18  4.71 11.04
6.98
  6.57  4.61  8.53  8.5   9.21  6.93 10.84  7.02  8.41 11.93 11.17
8.6
 11.98 10.59 11.61  4.14 11.58 10.09  7.46 11.81  8.7   7.5   8.27
9.17
 10.74  7.48 11.37 11.44 10.56  9.55  6.69 10.53  9.14  9.16  8.12
7.32
  4.93 11.4   8.87  9.28  9.91 10.82 11.71  8.    8.22  8.19  7.04
7.8
 10.75 11.2   7.85 10.68  8.84  6.5   7.95 11.08  9.59  6.27  7.98
7.57
  5.58  6.54  6.74  8.06  6.81 11.35 11.39 10.08  7.67  9.45  7.93
8.8
  8.57 12.    9.89  9.23 11.24 10.1   7.61  9.85 11.82  5.18  5.04
11.94
  6.53 10.81 10.24 10.98  9.47 10.54  7.99 10.46  8.29  9.04 10.18
8.45
 11.06 11.51  8.13  7.34  5.23  7.58 10.37  9.63 11.3   9.88  8.05
4.8
  6.79  9.94  8.58  8.68  5.1   8.96 10.17  7.83  7.45  9.48  9.58
10.72
  7.19  9.8   7.72 11.42  7.53 11.36 10.79 10.19 11.76  9.08  9.05
9.77
  8.26 11.31 10.83  4.31  8.34  8.21  5.06  7.37  9.81  8.02  7.33
9.18
  7.6  11.49  7.74  7.16  7.42  7.91 10.12  7.92  7.23  9.29  8.31
7.26
 10.78 10.52 11.52  9.1   8.83  7.69  5.9   8.69  8.24 10.62 11.21
7.82
 10.58  8.47  8.81 11.    8.25  7.21 11.33  9.03 11.86 10.64  9.32
9.24
 10.22  6.65  9.22  7.49  9.06 10.48 11.38  7.86 11.14 10.3   6.7
11.87
  9.02  8.16  6.55 11.01  9.83 10.14  8.08  6.13  5.02  6.68  7.39
9.53
 11.88  5.35 11.97 10.4   7.76 11.78  4.22 10.99  7.3  11.89 11.43
7.71
```

```
  7.   10.87  8.38  8.64  8.51  4.53 10.43  6.47  6.36 10.55  5.34
7.79
  9.51  8.1   6.76 10.5  11.64  7.05  6.85  9.52  6.33  6.97 11.95
10.96
  8.86  8.92  6.48  6.84 10.21  9.86 11.74 10.86 10.25  4.98 11.8
11.99
  7.84  5.48 11.55 10.6  10.38  9.42 11.83  9.68  8.72 11.12 10.73
11.34
  7.22  9.26  6.75 10.04  8.9   8.01  7.54  9.4   8.77  6.92  9.78
10.51
  6.67 11.57 10.92  7.96  9.33 10.61  9.34  8.49 11.91  5.95  6.86
8.71
  7.01  7.89 11.5  11.32 11.9   6.59  7.47]
Heart Issues
[0 1]
Any Transplants
[0 1]
Cancer history
[0 1]
NumberOfMajorSurgeries
[0 3 1 2]
smoker
[0 1]

filterdata.info()

<class 'pandas.core.frame.DataFrame'>
Index: 1755 entries, 0 to 2334
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             1755 non-null   object
 1   name                    1755 non-null   object
 2   year                    1755 non-null   object
 3   month                   1755 non-null   object
 4   day                     1755 non-null   int64
 5   children                1755 non-null   int64
 6   charges                 1755 non-null   float64
 7   Hospital tier           1755 non-null   object
 8   City tier               1755 non-null   object
 9   State ID                1755 non-null   int64
 10  BMI                     1755 non-null   float64
 11  HBA1C                   1755 non-null   float64
 12  Heart Issues            1755 non-null   int64
 13  Any Transplants         1755 non-null   int64
 14  Cancer history          1755 non-null   int64
 15  NumberOfMajorSurgeries  1755 non-null   int64
 16  smoker                  1755 non-null   int64
dtypes: float64(3), int64(8), object(6)
memory usage: 246.8+ KB
```

```
month_map = {'Jan': 1, 'Feb': 2, 'Mar': 3, 'Apr': 4,
             'May': 5, 'Jun': 6, 'Jul': 7, 'Aug': 8,
             'Sep': 9, 'Oct': 10, 'Nov': 11, 'Dec': 12}

filterdata['month'] = filterdata['month'].map(month_map)
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1336257953.py:5:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata['month'] = filterdata['month'].map(month_map)
```

```
filterdata['DOB'] = pd.to_datetime(filterdata['year'].astype(str) +
'-' +
                                   filterdata['month'].astype(str) +
'-' +
                                   filterdata['day'].astype(str),
format='%Y-%m-%d')
filterdata
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1283776996.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata['DOB'] = pd.to_datetime(filterdata['year'].astype(str) +
'-' +
```

```
     Customer ID                                name  year  month  day
\
0         Id2335                German, Mr.  Aaron K  1992      7    9

1         Id2334               Rosendahl, Mr.  Evan P  1992     11   30

2         Id2333                  Albano, Ms.  Julie  1993      6   30

3         Id2332  Riveros Gonzalez, Mr.  Juan D. Sr.  1992      9   13

4         Id2331                Brietzke, Mr.  Jordan  1998      7   27

...          ...                                  ...   ...    ...  ...

2328         Id7               Macpherson, Mr.  Scott  1994     10   27
```

```
2329         Id6              Baker, Mr.  Russell B.  1962      8    4

2330         Id5                Kadala, Ms.  Kristyn  1989      6   19

2333         Id2              Lehner, Mr.  Matthew D  1977      6    8

2334         Id1                   Hawks, Ms.  Kelly  1968     10   12


      children     charges Hospital tier City tier  State ID     BMI
HBA1C  \
0            0    563.8400      tier - 2  tier - 3         2 17.5800
4.5100
1            0    570.6200      tier - 2  tier - 1         2 17.6000
4.3900
2            0    600.0000      tier - 2  tier - 1         2 16.4700
6.3500
3            0    604.5400      tier - 3  tier - 3         2 17.7000
6.2800
4            0    637.2600      tier - 3  tier - 3         2 22.3400
5.5700
...        ...          ...           ...       ...       ...     ...
...
2328         1 51,194.5600      tier - 1  tier - 3         0 36.4000
6.0700
2329         0 52,590.8300      tier - 1  tier - 3         0 32.8000
6.5900
2330         0 55,135.4000      tier - 1  tier - 2         1 35.5300
5.4500
2333         0 62,592.8700      tier - 2  tier - 3         2 30.3600
5.7700
2334         0 63,770.4300      tier - 1  tier - 3         2 47.4100
7.4700

      Heart Issues  Any Transplants  Cancer history
NumberOfMajorSurgeries  \
0                0                 0               0
0
1                0                 0               0
0
2                0                 0               1
0
3                0                 0               0
0
4                0                 0               0
0
...            ...               ...             ...
...
2328             0                 0               0
3
```

```
2329              0                  0                 0
3
2330              0                  0                 0
3
2333              0                  0                 0
3
2334              0                  0                 0
3
```

```
      smoker        DOB
0          0 1992-07-09
1          0 1992-11-30
2          0 1993-06-30
3          0 1992-09-13
4          0 1998-07-27
...      ...        ...
2328       1 1994-10-27
2329       1 1962-08-04
2330       1 1989-06-19
2333       1 1977-06-08
2334       1 1968-10-12

[1755 rows x 18 columns]
```

```
from datetime import datetime

current_date = pd.to_datetime(datetime.now().date())
age_years = current_date.year - filterdata['DOB'].dt.year
age_adjusted = age_years - ((current_date.month <
filterdata['DOB'].dt.month) | ((current_date.month ==
filterdata['DOB'].dt.month) & (current_date.day <
filterdata['DOB'].dt.day)))
filterdata['Age'] = age_adjusted
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\4149288236.py:4:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata['Age'] = age_adjusted
```

```
tier_map = {'tier - 1': '1', 'tier - 2': '2', 'tier - 3': '3'}
filterdata['Hospital tier'] = filterdata['Hospital
tier'].map(tier_map)
filterdata['City tier'] = filterdata['City tier'].map(tier_map)
filterdata
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1562771818.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata['Hospital tier'] = filterdata['Hospital
tier'].map(tier_map)
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\1562771818.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata['City tier'] = filterdata['City tier'].map(tier_map)
```

	Customer ID	name	year	month	day
0	Id2335	German, Mr. Aaron K	1992	7	9
1	Id2334	Rosendahl, Mr. Evan P	1992	11	30
2	Id2333	Albano, Ms. Julie	1993	6	30
3	Id2332	Riveros Gonzalez, Mr. Juan D. Sr.	1992	9	13
4	Id2331	Brietzke, Mr. Jordan	1998	7	27
...	...	...	...	...	...
2328	Id7	Macpherson, Mr. Scott	1994	10	27
2329	Id6	Baker, Mr. Russell B.	1962	8	4
2330	Id5	Kadala, Ms. Kristyn	1989	6	19
2333	Id2	Lehner, Mr. Matthew D	1977	6	8
2334	Id1	Hawks, Ms. Kelly	1968	10	12

	children	charges	Hospital tier	City tier	State ID	BMI	HBA1C
0	0	563.8400	2	3	2	17.5800	4.5100
1	0	570.6200	2	1	2	17.6000	4.3900

```
2             0    600.0000              2          1          2 16.4700
6.3500
3             0    604.5400              3          3          2 17.7000
6.2800
4             0    637.2600              3          3          2 22.3400
5.5700
...         ...         ...            ...        ...        ...     ...
...
2328          1 51,194.5600              1          3          0 36.4000
6.0700
2329          0 52,590.8300              1          3          0 32.8000
6.5900
2330          0 55,135.4000              1          2          1 35.5300
5.4500
2333          0 62,592.8700              2          3          2 30.3600
5.7700
2334          0 63,770.4300              1          3          2 47.4100
7.4700

      Heart Issues  Any Transplants  Cancer history
NumberOfMajorSurgeries  \
0                0                 0                0
0
1                0                 0                0
0
2                0                 0                1
0
3                0                 0                0
0
4                0                 0                0
0
...            ...               ...              ...
...
2328             0                 0                0
3
2329             0                 0                0
3
2330             0                 0                0
3
2333             0                 0                0
3
2334             0                 0                0
3

      smoker        DOB  Age
0          0 1992-07-09   32
1          0 1992-11-30   32
2          0 1993-06-30   31
3          0 1992-09-13   32
```

```
4          0 1998-07-27   26
...      ...        ...  ...
2328       1 1994-10-27   30
2329       1 1962-08-04   62
2330       1 1989-06-19   35
2333       1 1977-06-08   47
2334       1 1968-10-12   56

[1755 rows x 19 columns]
```

```
title =
filterdata['name'].str.split(',').str[1].str.split('.').str[0].str.str
ip()
gender_map = {'Mr': 0, 'Ms': 1, 'Mrs': 1}
filterdata['Gender'] = title.map(gender_map)
filterdata
```

```
C:\Users\1911s\AppData\Local\Temp\ipykernel_17092\2925448966.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filterdata['Gender'] = title.map(gender_map)
```

	Customer ID	name	year	month	day
0	Id2335	German, Mr. Aaron K	1992	7	9
1	Id2334	Rosendahl, Mr. Evan P	1992	11	30
2	Id2333	Albano, Ms. Julie	1993	6	30
3	Id2332	Riveros Gonzalez, Mr. Juan D. Sr.	1992	9	13
4	Id2331	Brietzke, Mr. Jordan	1998	7	27
...	...	...	...	...	...
2328	Id7	Macpherson, Mr. Scott	1994	10	27
2329	Id6	Baker, Mr. Russell B.	1962	8	4
2330	Id5	Kadala, Ms. Kristyn	1989	6	19
2333	Id2	Lehner, Mr. Matthew D	1977	6	8
2334	Id1	Hawks, Ms. Kelly	1968	10	12

```
      children     charges Hospital tier City tier  State ID     BMI
HBA1C  \
0            0    563.8400             2         3         2 17.5800
4.5100
1            0    570.6200             2         1         2 17.6000
4.3900
2            0    600.0000             2         1         2 16.4700
6.3500
3            0    604.5400             3         3         2 17.7000
6.2800
4            0    637.2600             3         3         2 22.3400
5.5700
...        ...         ...           ...       ...       ...     ...
...
2328         1 51,194.5600             1         3         0 36.4000
6.0700
2329         0 52,590.8300             1         3         0 32.8000
6.5900
2330         0 55,135.4000             1         2         1 35.5300
5.4500
2333         0 62,592.8700             2         3         2 30.3600
5.7700
2334         0 63,770.4300             1         3         2 47.4100
7.4700

      Heart Issues  Any Transplants  Cancer history
NumberOfMajorSurgeries  \
0                0                0               0
0
1                0                0               0
0
2                0                0               1
0
3                0                0               0
0
4                0                0               0
0
...            ...              ...             ...
...
2328             0                0               0
3
2329             0                0               0
3
2330             0                0               0
3
2333             0                0               0
3
2334             0                0               0
```

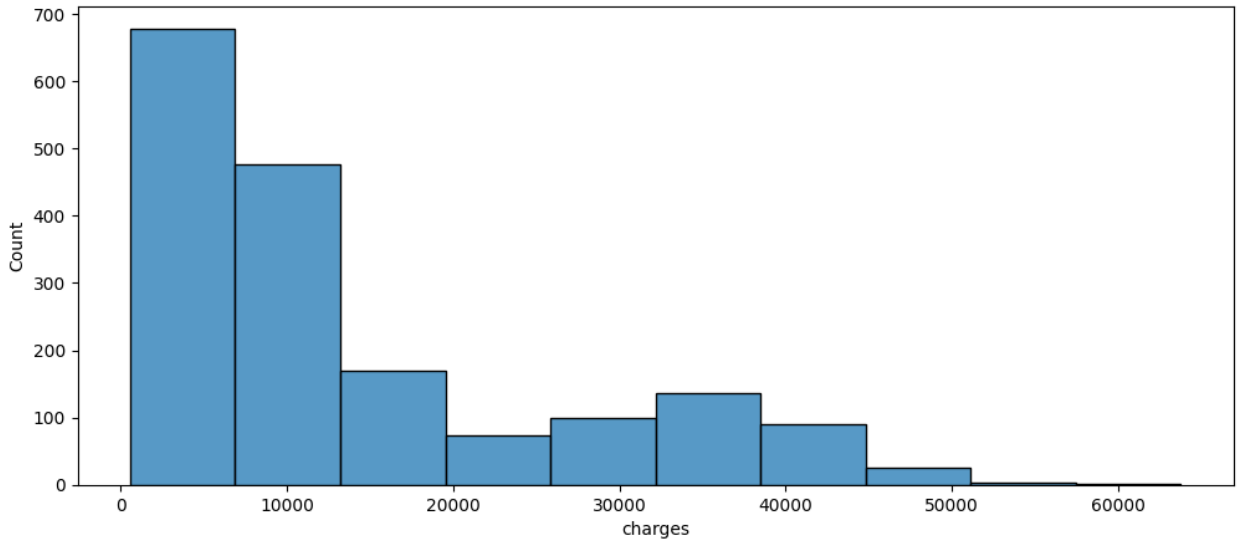
```
3
```

	smoker	DOB	Age	Gender
0	0	1992-07-09	32	0
1	0	1992-11-30	32	0
2	0	1993-06-30	31	1
3	0	1992-09-13	32	0
4	0	1998-07-27	26	0
...	...	...	...	...
2328	1	1994-10-27	30	0
2329	1	1962-08-04	62	0
2330	1	1989-06-19	35	1
2333	1	1977-06-08	47	0
2334	1	1968-10-12	56	1

```
[1755 rows x 20 columns]
```

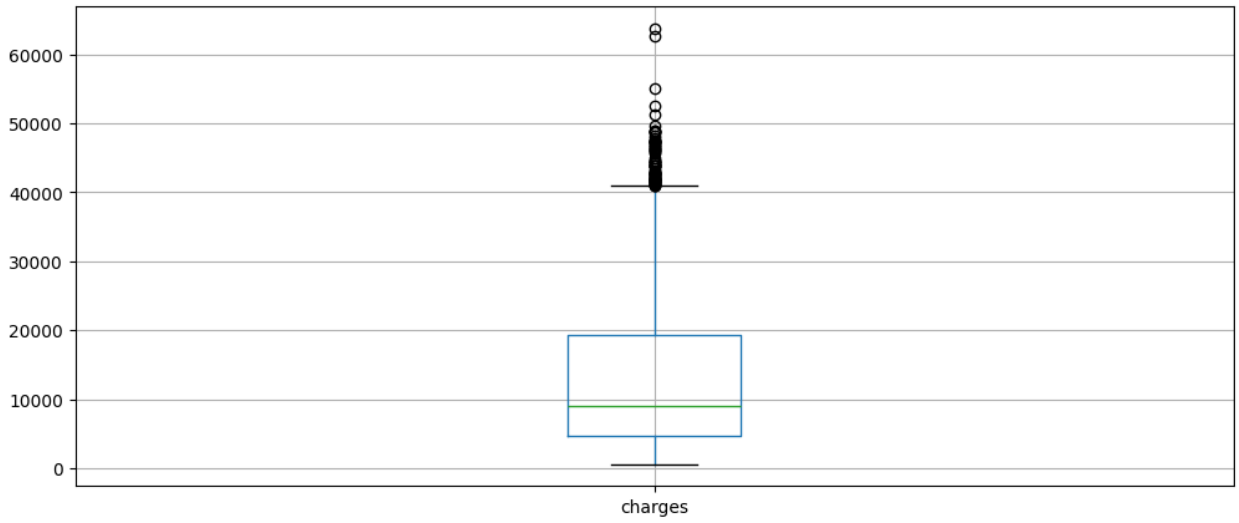
```
import seaborn as sb
plt.figure(figsize=(12,5))
sb.histplot(filterdata['charges'],kde=False,bins=10)
```

```
<Axes: xlabel='charges', ylabel='Count'>
```




```
plt.figure(figsize=(12,5))
filterdata['charges'].plot(kind='box',title='Cost of hospitalisation')
plt.show()
```


```
plt.figure(figsize=(12,5))
filterdata.boxplot(column='charges')
plt.show()
```




```
plt.figure(figsize=(25,5))
sb.swarmplot(x='year',y='charges',hue='Gender', data=filterdata)
plt.title('Swarm Plot')
plt.show()
```

```
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 11.6% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
```

```
packages\seaborn\categorical.py:3399: UserWarning: 22.2% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 40.5% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 36.8% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 15.4% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 50.0% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 27.3% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 20.0% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 38.7% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 57.9% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 33.3% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
```

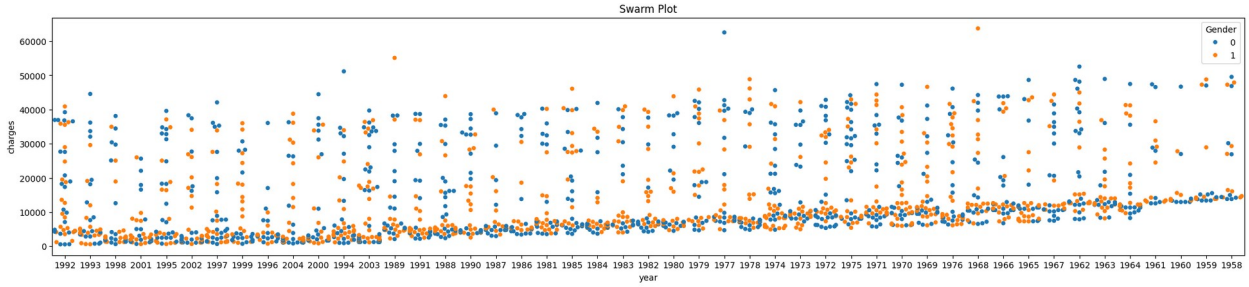
```
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 21.4% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 52.6% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 27.0% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 25.6% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 28.9% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 34.4% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 30.0% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 35.7% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 20.5% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 37.1% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
```

```
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 18.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 27.8% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 34.3% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 6.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 21.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 38.1% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 17.4% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 23.5% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 20.9% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:3399: UserWarning: 11.9% of the points cannot be placed; you may want to decrease the size of the markers or
```

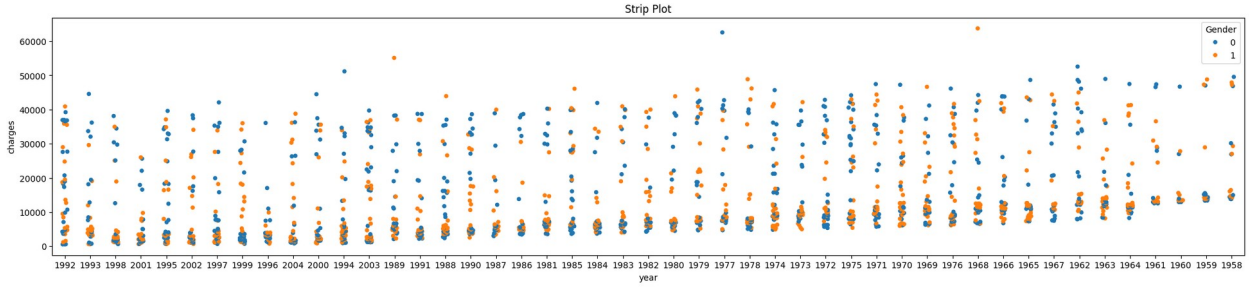
```
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 27.9% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 21.7% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 22.0% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 19.4% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 24.3% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 32.4% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 16.1% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 13.9% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 21.9% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 36.1% of the points
```

```
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 11.8% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 41.2% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 11.1% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 37.8% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 45.7% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 30.6% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 16.7% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 31.0% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 15.2% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
```

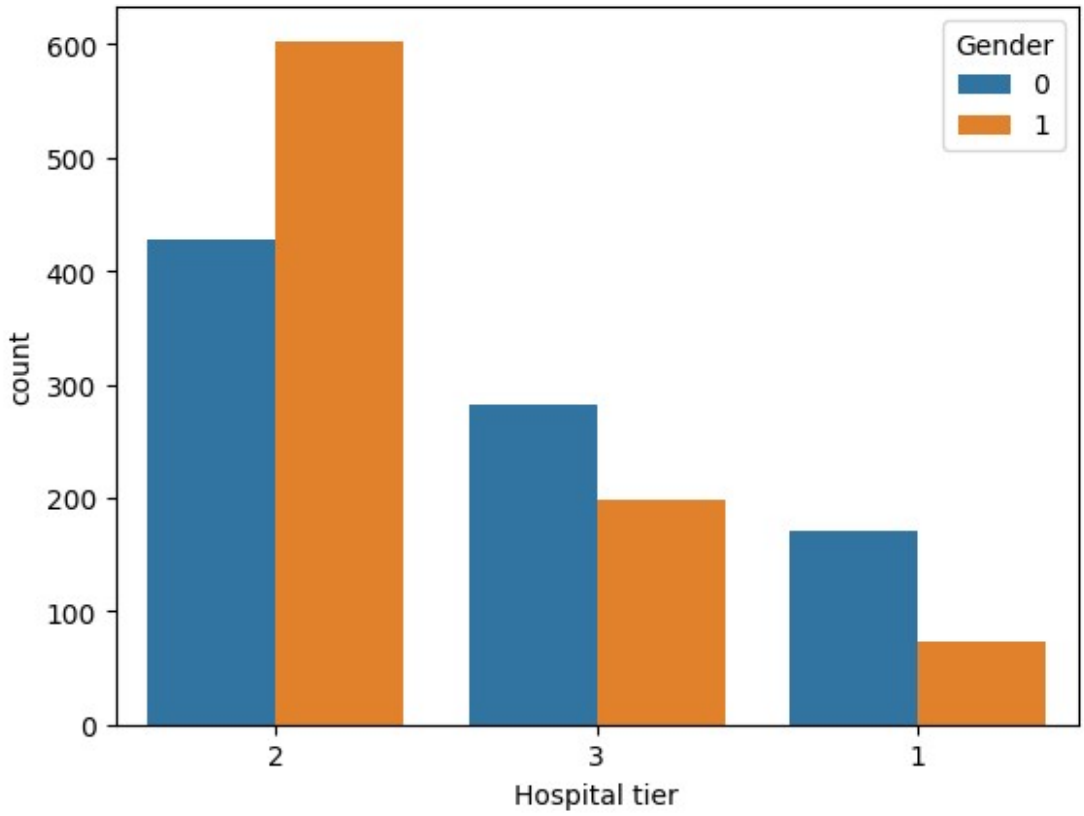
```
packages\seaborn\categorical.py:3399: UserWarning: 25.0% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\1911s\AppData\Local\Programs\Python\Python312\Lib\site-
packages\seaborn\categorical.py:3399: UserWarning: 19.5% of the points
cannot be placed; you may want to decrease the size of the markers or
use stripplot.
  warnings.warn(msg, UserWarning)
```




```
plt.figure(figsize=(25,5))
sb.stripplot(x='year',y='charges',hue='Gender', data=filterdata)
plt.title('Strip Plot')
plt.show()
```




```
sb.countplot(x=filterdata['Hospital
tier'],hue='Gender',data=filterdata)
plt.show()
```




```
filterdata.groupby(['Hospital tier'])
['charges'].median().reset_index()
```

```
  Hospital tier     charges
0             1 32,694.1550
1             2  6,846.1000
2             3 10,231.5000
```

```
df=pd.DataFrame(dict(r=['32,694.1550','6,846.1000','10,231.5000'],thet
a=['Tier-1 hospital','Tier-2 hospital','Tier-3 hospital']))

import plotly.express as px

fig=px.line_polar(df,r='r',theta='theta',line_close=True)
fig.update_traces(fill='toself')
fig.show()
```


```
{"config":{"plotlyServerURL":"https://plot.ly"},"data":
[{"fill":"toself","hovertemplate":"r=%{r}<br>theta=%{theta}<extra></
extra>","legendgroup":"","line":
{"color":"#636efa","dash":"solid"},"marker":
{"symbol":"circle"},"mode":"lines","name":"","r":
["32,694.1550","6,846.1000","10,231.5000","32,694.1550"],"showlegend":
false,"subplot":"polar","theta":["Tier-1 hospital","Tier-2
hospital","Tier-3 hospital","Tier-1
```

```
hospital"],"type":"scatterpolar"}],"layout":{"autosize":true,"legend":
{"tracegroupgap":0},"margin":{"t":60},"polar":{"angularaxis":
{"direction":"clockwise","rotation":90,"type":"category"},"domain":
{"x":[0,1],"y":[0,1]},"radialaxis":{"autorange":true,"range":
[0,2],"type":"category"}},"template":{"data":{"bar":[{"error_x":
{"color":"#2a3f5f"},"error_y":{"color":"#2a3f5f"},"marker":{"line":
{"color":"#E5ECF6","width":0.5},"pattern":
{"fillmode":"overlay","size":10,"solidity":0.2}},"type":"bar"}],"barpo
lar":[{"marker":{"line":{"color":"#E5ECF6","width":0.5},"pattern":
{"fillmode":"overlay","size":10,"solidity":0.2}},"type":"barpolar"}],"
carpet":[{"aaxis":
{"endlinecolor":"#2a3f5f","gridcolor":"white","linecolor":"white","min
orgridcolor":"white","startlinecolor":"#2a3f5f"},"baxis":
{"endlinecolor":"#2a3f5f","gridcolor":"white","linecolor":"white","min
orgridcolor":"white","startlinecolor":"#2a3f5f"},"type":"carpet"}],"ch
oropleth":[{"colorbar":
{"outlinewidth":0,"ticks":""},"type":"choropleth"}],"contour":
[{"colorbar":{"outlinewidth":0,"ticks":""},"colorscale":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]],"type":"contour"}],"contourcarpet":[{"colorbar":
{"outlinewidth":0,"ticks":""},"type":"contourcarpet"}],"heatmap":
[{"colorbar":{"outlinewidth":0,"ticks":""},"colorscale":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]],"type":"heatmap"}],"heatmapgl":[{"colorbar":
{"outlinewidth":0,"ticks":""},"colorscale":[[0,"#0d0887"],
[0.1111111111111111,"#46039f"],[0.2222222222222222,"#7201a8"],
[0.3333333333333333,"#9c179e"],[0.4444444444444444,"#bd3786"],
[0.5555555555555556,"#d8576b"],[0.6666666666666666,"#ed7953"],
[0.7777777777777778,"#fb9f3a"],[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]],"type":"heatmapgl"}],"histogram":[{"marker":{"pattern":
{"fillmode":"overlay","size":10,"solidity":0.2}},"type":"histogram"}],
"histogram2d":[{"colorbar":{"outlinewidth":0,"ticks":""},"colorscale":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]],"type":"histogram2d"}],"histogram2dcontour":
[{"colorbar":{"outlinewidth":0,"ticks":""},"colorscale":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
```

```
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]],"type":"histogram2dcontour"}],"mesh3d":[{"colorbar":
{"outlinewidth":0,"ticks":""},"type":"mesh3d"}],"parcoords":[{"line":
{"colorbar":{"outlinewidth":0,"ticks":""}},"type":"parcoords"}],"pie":
[{"automargin":true,"type":"pie"}],"scatter":[{"fillpattern":
{"fillmode":"overlay","size":10,"solidity":0.2},"type":"scatter"}],"sc
atter3d":[{"line":{"colorbar":{"outlinewidth":0,"ticks":""}},"marker":
{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scatter3d"}],"scattercarpet":
[{"marker":{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scattercarpet"}],"scattergeo":
[{"marker":{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scattergeo"}],"scattergl":
[{"marker":{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scattergl"}],"scattermapbox":
[{"marker":{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scattermapbox"}],"scatterpolar"
:[{"marker":{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scatterpolar"}],"scatterpolargl
":[{"marker":{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scatterpolargl"}],"scatterterna
ry":[{"marker":{"colorbar":
{"outlinewidth":0,"ticks":""}},"type":"scatterternary"}],"surface":
[{"colorbar":{"outlinewidth":0,"ticks":""},"colorscale":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]],"type":"surface"}],"table":[{"cells":{"fill":
{"color":"#EBF0F8"},"line":{"color":"white"}},"header":{"fill":
{"color":"#C8D4E3"},"line":
{"color":"white"}},"type":"table"}]},"layout":{"annotationdefaults":
{"arrowcolor":"#2a3f5f","arrowhead":0,"arrowwidth":1},"autotypenumbers
":"strict","coloraxis":{"colorbar":
{"outlinewidth":0,"ticks":""}},"colorscale":{"diverging":
[[0,"#8e0152"],[0.1,"#c51b7d"],[0.2,"#de77ae"],[0.3,"#f1b6da"],
[0.4,"#fde0ef"],[0.5,"#f7f7f7"],[0.6,"#e6f5d0"],[0.7,"#b8e186"],
[0.8,"#7fbc41"],[0.9,"#4d9221"],[1,"#276419"]],"sequential":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],[1,"#f0f921"]],"sequentialminus":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
```

[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],[1,"#f0f921"]]},"colorway":
["#636efa","#EF553B","#00cc96","#ab63fa","#FFA15A","#19d3f3","#FF6692"
,"#B6E880","#FF97FF","#FECB52"],"font":{"color":"#2a3f5f"},"geo":
{"bgcolor":"white","lakecolor":"white","landcolor":"#E5ECF6","showlake
s":true,"showland":true,"subunitcolor":"white"},"hoverlabel":
{"align":"left"},"hovermode":"closest","mapbox":
{"style":"light"},"paper_bgcolor":"white","plot_bgcolor":"#E5ECF6","po
lar":{"angularaxis":
{"gridcolor":"white","linecolor":"white","ticks":""},"bgcolor":"#E5ECF
6","radialaxis":
{"gridcolor":"white","linecolor":"white","ticks":""}},"scene":
{"xaxis":
{"backgroundcolor":"#E5ECF6","gridcolor":"white","gridwidth":2,"lineco
lor":"white","showbackground":true,"ticks":"","zerolinecolor":"white"}
,"yaxis":
{"backgroundcolor":"#E5ECF6","gridcolor":"white","gridwidth":2,"lineco
lor":"white","showbackground":true,"ticks":"","zerolinecolor":"white"}
,"zaxis":
{"backgroundcolor":"#E5ECF6","gridcolor":"white","gridwidth":2,"lineco
lor":"white","showbackground":true,"ticks":"","zerolinecolor":"white"}
},"shapedefaults":{"line":{"color":"#2a3f5f"}},"ternary":{"aaxis":
{"gridcolor":"white","linecolor":"white","ticks":""},"baxis":
{"gridcolor":"white","linecolor":"white","ticks":""},"bgcolor":"#E5ECF
6","caxis":
{"gridcolor":"white","linecolor":"white","ticks":""}},"title":
{"x":5.0e-2},"xaxis":
{"automargin":true,"gridcolor":"white","linecolor":"white","ticks":"",
"title":
{"standoff":15},"zerolinecolor":"white","zerolinewidth":2},"yaxis":
{"automargin":true,"gridcolor":"white","linecolor":"white","ticks":"",
"title":{"standoff":15},"zerolinecolor":"white","zerolinewidth":2}}}}}

```
hos_tier=filterdata.groupby(['Hospital
tier']).size().rename_axis('City&hospital_tier').reset_index(name='Hos
pitalcount')
hos_tier
```

	City&hospital_tier	Hospitalcount
0	1	244
1	2	1030
2	3	481

```
cit_tier=filterdata.groupby(['City
tier']).size().rename_axis('City&hospital_tier').reset_index(name='Cit
ycount')
cit_tier
```

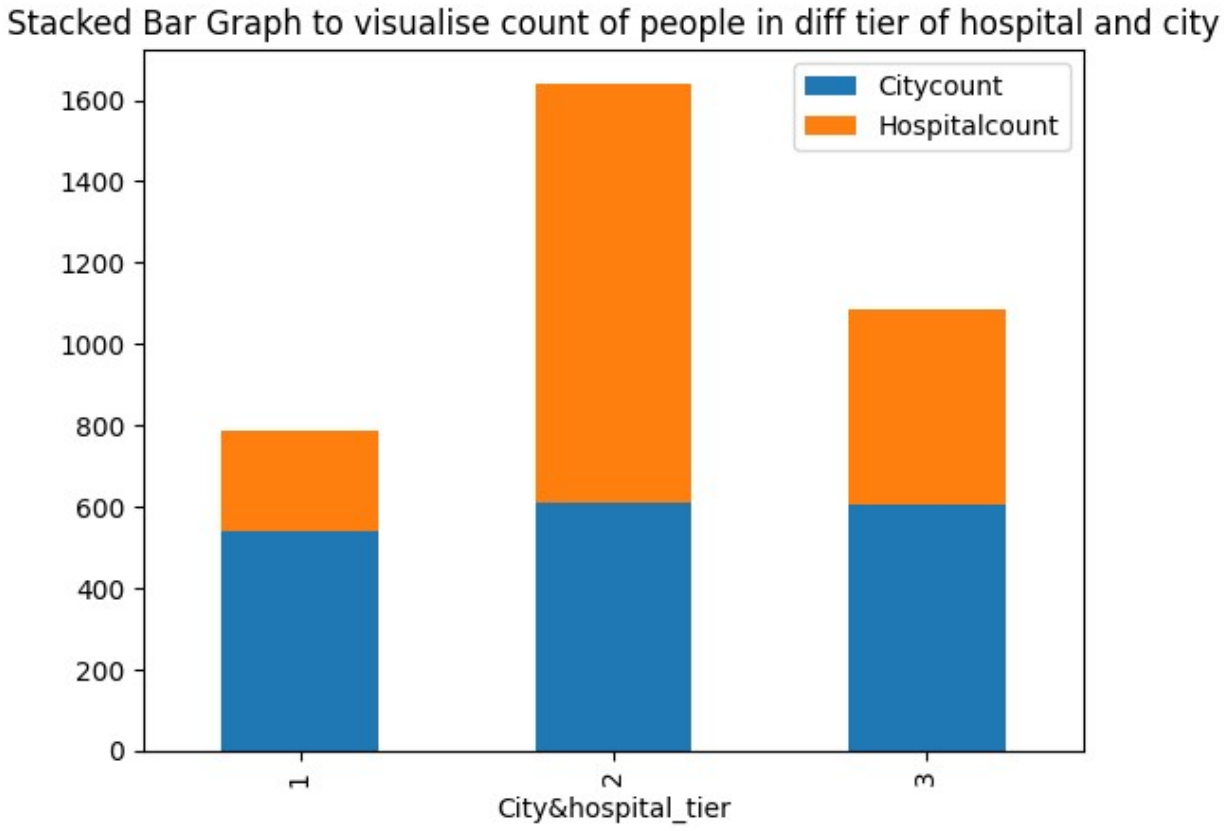
	City&hospital_tier	Citycount
0	1	541

```
1                    2         611
2                    3         603
```

```
df = pd.merge(cit_tier, hos_tier, on ='City&hospital_tier')
df
```

```
  City&hospital_tier  Citycount  Hospitalcount
0                  1        541            244
1                  2        611           1030
2                  3        603            481
```

```
df.plot(x='City&hospital_tier', kind='bar', stacked=True,
        title='Stacked Bar Graph to visualise count of people in diff
tier of hospital and city')
plt.show()
```




```
from scipy.stats import friedmanchisquare

stat,p=friedmanchisquare(32694.1550,6846.1000,10231.5000)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

```
stat=2.000, p=0.368
Probably the same distribution
```

Since p>0.05 we accept null hypothesis for hospital tier.

```
filterdata.groupby(['City tier'])['charges'].median().reset_index()
```

```
  City tier    charges
0         1 9,174.1400
1         2 8,522.0000
2         3 9,304.7000
```

```
stat,p=friedmanchisquare(9174.1400,8522.0000,9304.7000)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different distributions')
```

```
stat=2.000, p=0.368
Probably the same distribution
```

Since p>0.05 we accept null hypothesis for city tier.

```
smokdf=filterdata.groupby(['smoker'])
['charges'].median().reset_index()
smokdf
```

```
   smoker     charges
0       0  6,778.6550
1       1 34,218.0200
```

```
from scipy.stats import chi2_contingency

chi2, p, dof, expected = chi2_contingency(smokdf)
print(f"Chi-Square: {chi2}")
print(f"P-Value: {p}")
print(f"Degrees of Freedom: {dof}")
print(f"Expected Frequencies:{expected}")
```

```
Chi-Square: 0.0
P-Value: 1.0
Degrees of Freedom: 1
Expected Frequencies:[[1.65342425e-01 6.77848966e+03]
 [8.34657575e-01 3.42181853e+04]]
```

P-Value: 1.0 - A p-value of 1.0 means there is no statistical significance. In other words, the null hypothesis cannot be rejected, and the data does not provide evidence of an association between the variables.

```
table = [[filterdata["Heart Issues"].value_counts()],
[filterdata["smoker"].value_counts()]]
stat, p, dof, expected = chi2_contingency(table)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably independent')
else:
    print('Probably dependent')
```

```
stat=97.244, p=0.000
Probably dependent
```

```
modeldata=filterdata.drop(columns=['Customer
ID','name','year','month','day','DOB'],axis=1)
modeldata
```

```
      children     charges Hospital tier City tier  State ID     BMI
HBA1C  \
0            0    563.8400              2         3         2 17.5800
4.5100
1            0    570.6200              2         1         2 17.6000
4.3900
2            0    600.0000              2         1         2 16.4700
6.3500
3            0    604.5400              3         3         2 17.7000
6.2800
4            0    637.2600              3         3         2 22.3400
5.5700
...        ...         ...            ...       ...       ...     ...
...
2328         1 51,194.5600              1         3         0 36.4000
6.0700
2329         0 52,590.8300              1         3         0 32.8000
6.5900
2330         0 55,135.4000              1         2         1 35.5300
5.4500
2333         0 62,592.8700              2         3         2 30.3600
5.7700
2334         0 63,770.4300              1         3         2 47.4100
7.4700

      Heart Issues  Any Transplants  Cancer history
NumberOfMajorSurgeries  \
0                0                0               0
0
1                0                0               0
```

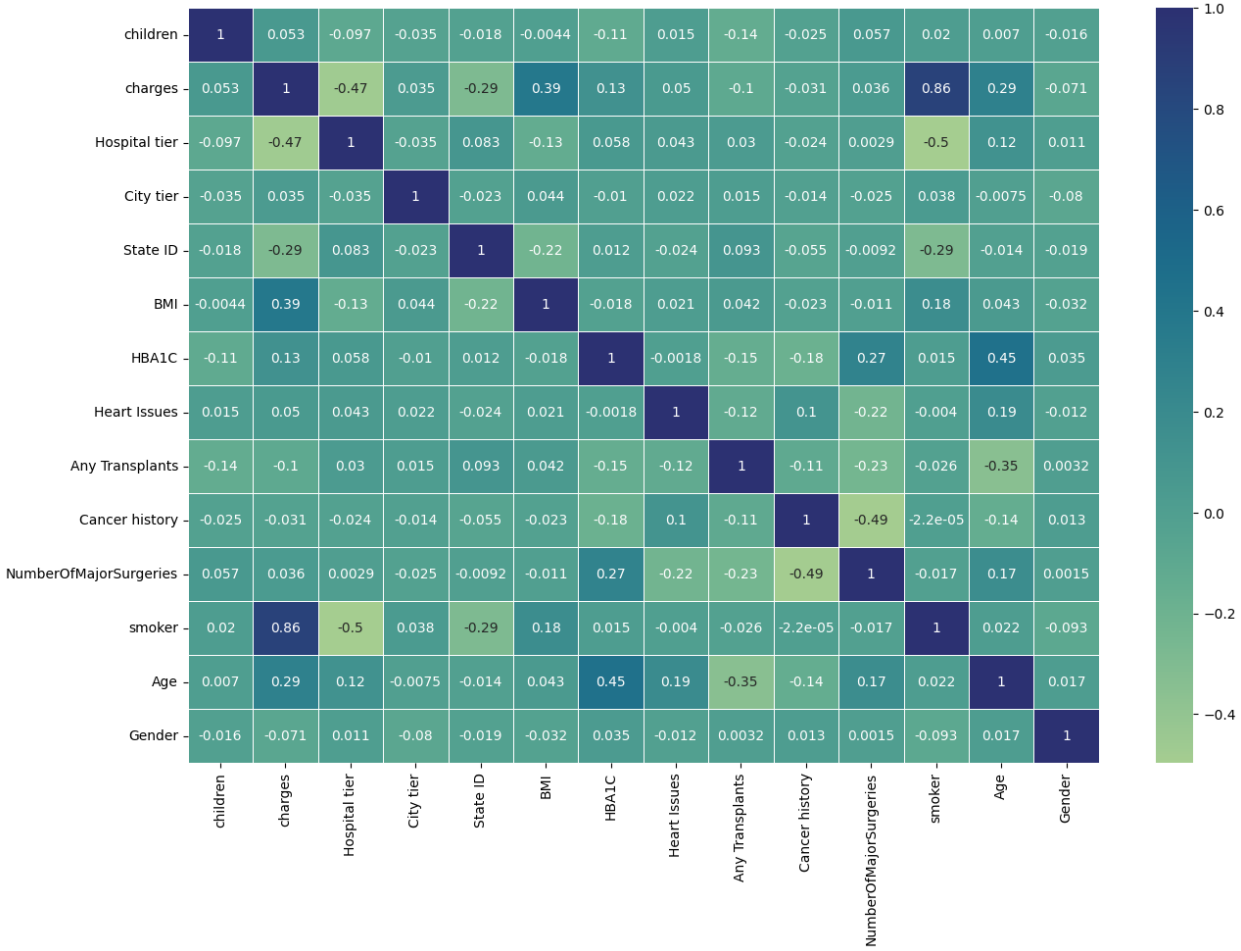
```
0
2                 0                 0                 1
0
3                 0                 0                 0
0
4                 0                 0                 0
0
...             ...               ...               ...
...
2328              0                 0                 0
3
2329              0                 0                 0
3
2330              0                 0                 0
3
2333              0                 0                 0
3
2334              0                 0                 0
3

      smoker  Age  Gender
0          0   32       0
1          0   32       0
2          0   31       1
3          0   32       0
4          0   26       0
...      ...  ...     ...
2328       1   30       0
2329       1   62       0
2330       1   35       1
2333       1   47       0
2334       1   56       1

[1755 rows x 14 columns]

modeldata.corr()

                        children  charges  Hospital tier  City tier
State ID  \
children                  1.0000   0.0533        -0.0966    -0.0353
-0.0180
charges                   0.0533   1.0000        -0.4718     0.0347
-0.2885
Hospital tier            -0.0966  -0.4718         1.0000    -0.0353
0.0832
City tier                -0.0353   0.0347        -0.0353     1.0000
-0.0234
State ID                 -0.0180  -0.2885         0.0832    -0.0234
1.0000
BMI                      -0.0044   0.3857        -0.1342     0.0437
```

```
-0.2170
HBA1C                      -0.1144   0.1342         0.0582    -0.0101
0.0123
Heart Issues                0.0151   0.0504         0.0432     0.0224
-0.0239
Any Transplants            -0.1351  -0.1028         0.0301     0.0149
0.0934
Cancer history             -0.0248  -0.0312        -0.0241    -0.0141
-0.0554
NumberOfMajorSurgeries      0.0574   0.0359         0.0029    -0.0246
-0.0092
smoker                      0.0200   0.8615        -0.4966     0.0383
-0.2884
Age                         0.0070   0.2949         0.1250    -0.0075
-0.0139
Gender                     -0.0156  -0.0714         0.0106    -0.0804
-0.0187

                              BMI   HBA1C  Heart Issues  Any Transplants
\
children                  -0.0044 -0.1144        0.0151           -0.1351

charges                    0.3857  0.1342        0.0504           -0.1028

Hospital tier             -0.1342  0.0582        0.0432            0.0301

City tier                  0.0437 -0.0101        0.0224            0.0149

State ID                  -0.2170  0.0123       -0.0239            0.0934

BMI                        1.0000 -0.0177        0.0207            0.0424

HBA1C                     -0.0177  1.0000       -0.0018           -0.1492

Heart Issues               0.0207 -0.0018        1.0000           -0.1185

Any Transplants            0.0424 -0.1492       -0.1185            1.0000

Cancer history            -0.0231 -0.1812        0.1025           -0.1065

NumberOfMajorSurgeries    -0.0108  0.2728       -0.2226           -0.2337

smoker                     0.1755  0.0153       -0.0040           -0.0257

Age                        0.0427  0.4548        0.1938           -0.3473

Gender                    -0.0316  0.0348       -0.0117            0.0032


                          Cancer history  NumberOfMajorSurgeries  smoker
\
```

```
children                          -0.0248                    0.0574  0.0200

charges                           -0.0312                    0.0359  0.8615

Hospital tier                     -0.0241                    0.0029 -0.4966

City tier                         -0.0141                   -0.0246  0.0383

State ID                          -0.0554                   -0.0092 -0.2884

BMI                               -0.0231                   -0.0108  0.1755

HBA1C                             -0.1812                    0.2728  0.0153

Heart Issues                       0.1025                   -0.2226 -0.0040

Any Transplants                   -0.1065                   -0.2337 -0.0257

Cancer history                     1.0000                   -0.4896 -0.0000

NumberOfMajorSurgeries            -0.4896                    1.0000 -0.0168

smoker                            -0.0000                   -0.0168  1.0000

Age                               -0.1417                    0.1710  0.0222

Gender                             0.0135                    0.0015 -0.0931
```

```
                            Age  Gender
children                 0.0070 -0.0156
charges                  0.2949 -0.0714
Hospital tier            0.1250  0.0106
City tier               -0.0075 -0.0804
State ID                -0.0139 -0.0187
BMI                      0.0427 -0.0316
HBA1C                    0.4548  0.0348
Heart Issues             0.1938 -0.0117
Any Transplants         -0.3473  0.0032
Cancer history          -0.1417  0.0135
NumberOfMajorSurgeries   0.1710  0.0015
smoker                   0.0222 -0.0931
Age                      1.0000  0.0169
Gender                   0.0169  1.0000
```

```
plt.figure(figsize=(15,10))
sb.heatmap(modeldata.corr(), annot=True, linewidth=.5, cmap="crest")
plt.savefig("heatmap.png")
plt.show()
```




```
X=modeldata.drop('charges',axis=1)
y=modeldata['charges']

from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.8,rando
m_state=42)

from sklearn.preprocessing import StandardScaler

from sklearn.pipeline import Pipeline
linear_pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('linear', LinearRegression())
])

ridge_pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('ridge', Ridge())
])

from sklearn.model_selection import cross_val_score,KFold,GridSearchCV
```

```
kf = KFold(n_splits=5, shuffle=True, random_state=42)

linear_scores = cross_val_score(linear_pipeline, X_train, y_train,
cv=kf, scoring='neg_mean_squared_error')

param_grid = {'ridge__alpha': [0.1, 1.0, 10.0, 100.0]}
ridge_cv = GridSearchCV(ridge_pipeline,param_grid ,cv=kf,
scoring='neg_mean_squared_error')

results=ridge_cv.fit(X_train, y_train)

pred=ridge_cv.predict(X_test)

from sklearn.metrics import r2_score
r2=r2_score(y_test,pred)
r2
```

```
0.8660373861456069
```

```
ridge_model_after_gridcv = Ridge(alpha =
results.best_params_.get('ridge__alpha')).fit(X_train, y_train)

ridge_scores = ridge_cv.cv_results_['mean_test_score']

print(f"Linear Regression MSE: {-linear_scores.mean()} (+/-
{linear_scores.std()})")
print(f"Ridge Regression MSE: {-ridge_scores.mean()} (+/-
{ridge_scores.std()})")
print(f"Best Ridge Alpha: {ridge_cv.best_params_}")
```

```
Linear Regression MSE: 19333367.836520676 (+/- 3068234.3295021374)
Ridge Regression MSE: 21647924.431065653 (+/- 3884219.203457555)
Best Ridge Alpha: {'ridge__alpha': 0.1}
```

```
from sklearn.metrics import mean_squared_error, mean_absolute_error
sgd_mae = mean_absolute_error(y_test, pred)
sgd_mse = mean_squared_error(y_test, pred)

print("MAE:", sgd_mae)
print("MSE:", sgd_mse)
```

```
MAE: 2842.307816127824
MSE: 21091170.90446624
```

```
pd.DataFrame(ridge_model_after_gridcv.coef_,index=X_train.columns,colu
mns=['Feature_imp'])
```

	Feature_imp
children	786.4827
Hospital tier	-1,381.3698
City tier	-324.8769
State ID	-177.1128
BMI	334.2270

```
HBA1C                         -7.9492
Heart Issues                   9.2283
Any Transplants            3,142.5018
Cancer history             1,397.4244
NumberOfMajorSurgeries       228.4282
smoker                    22,193.6349
Age                          256.7677
Gender                       -68.4132
```

```
gbr=GradientBoostingRegressor(random_state=42)

gbr.fit(X_train,y_train)
```

```
GradientBoostingRegressor(random_state=42)
```

```
pred_y=gbr.predict(X_test)

score=gbr.score(X_test,y_test)
score
```

```
0.8977969045603712
```

```
error= mean_squared_error(y_test, pred_y)
error2=mean_absolute_error(y_test, pred_y)
print('Root mean Square error: {:.2f}'.format(error))
print('Root mean absolute error: {:.2f}'.format(error2))
```

```
Root mean Square error: 16090929.33
Root mean absolute error: 2405.88
```

```
modeldata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1755 entries, 0 to 2334
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   children                1755 non-null   int64  
 1   charges                 1755 non-null   float64
 2   Hospital tier           1755 non-null   object 
 3   City tier               1755 non-null   object 
 4   State ID                1755 non-null   int64  
 5   BMI                     1755 non-null   float64
 6   HBA1C                   1755 non-null   float64
 7   Heart Issues            1755 non-null   int64  
 8   Any Transplants         1755 non-null   int64  
 9   Cancer history          1755 non-null   int64  
 10  NumberOfMajorSurgeries  1755 non-null   int64  
 11  smoker                  1755 non-null   int64  
 12  Age                     1755 non-null   int32  
 13  Gender                  1755 non-null   int64  
```

```
dtypes: float64(3), int32(1), int64(8), object(2)
memory usage: 198.8+ KB
```

```
h=170/100
BMI=85/(h*h)
BMI
```

```
29.411764705882355
```

```
agec=datetime.now()-datetime(1988,12,28)
agec
years = agec.days // 365
months = (agec.days % 365) // 30
days = (agec.days % 365) % 30
print(f"Age: {years} years, {months} months, {days} days")
```

```
Age: 36 years, 2 months, 6 days
```

```
customdata=[{'children':2,'Hospital tier':1,'City tier':1,'State
ID':0,'BMI':29.411764705882355,'HBA1C':5.8,'Heart Issues':0,
            'Any Transplants':0,'Cancer
history':1,'NumberOfMajorSurgeries':0,'smoker':0,'Age':36,'Gender':1}]
customdf=pd.DataFrame(customdata)
customdf
```

```
   children  Hospital tier  City tier  State ID     BMI  HBA1C  Heart
Issues  \
0         2              1          1         0 29.4118 5.8000
0

   Any Transplants  Cancer history  NumberOfMajorSurgeries  smoker
Age  \
0                0               1                       0       0
36

   Gender
0       1
```

```
ridge_pipeline.fit(X_train,y_train)
```

```
Pipeline(steps=[('scaler', StandardScaler()), ('ridge', Ridge())])
```

```
HospitalCost=[]
predRidge=ridge_pipeline.predict(customdf)
HospitalCost.append(predRidge)

predgbr=gbr.predict(customdf)
HospitalCost.append(predgbr)

HospitalCost
```

```
[array([10195.79110491]), array([11792.93862416])]
```

```
avgCost=np.mean(HospitalCost)
avgCost
```

```
np.float64(10994.364864531472)
```

The avg cost estimation will be 10994.364864531472 Rs