NeuroFinalProject

December 9, 2023

0.1 Import packages and dataset

```
[4]:               ID M/F Hand  Age  Educ  SES  MMSE  CDR  eTIV   nWBV    ASF  \
     0     OAS1_0001_MR1   F    R   74   2.0  3.0  29.0  0.0  1344  0.743  1.306
     1     OAS1_0002_MR1   F    R   55   4.0  1.0  29.0  0.0  1147  0.810  1.531
     2     OAS1_0003_MR1   F    R   73   4.0  3.0  27.0  0.5  1454  0.708  1.207
     3     OAS1_0004_MR1   M    R   28   NaN  NaN   NaN  NaN  1588  0.803  1.105
     4     OAS1_0005_MR1   M    R   18   NaN  NaN   NaN  NaN  1737  0.848  1.010
     ..              ...  ..  ...  ...   ...  ...   ...  ...   ...    ...    ...
     431   OAS1_0285_MR2   M    R   20   NaN  NaN   NaN  NaN  1469  0.847  1.195
     432   OAS1_0353_MR2   M    R   22   NaN  NaN   NaN  NaN  1684  0.790  1.042
     433   OAS1_0368_MR2   M    R   22   NaN  NaN   NaN  NaN  1580  0.856  1.111
     434   OAS1_0379_MR2   F    R   20   NaN  NaN   NaN  NaN  1262  0.861  1.390
     435   OAS1_0395_MR2   F    R   26   NaN  NaN   NaN  NaN  1283  0.834  1.368

           Delay
     0       NaN
     1       NaN
     2       NaN
     3       NaN
     4       NaN
     ..      ...
     431     2.0
     432    40.0
     433    89.0
     434     2.0
     435    39.0

     [436 rows x 12 columns]
```

1 Data preprocessing

1.0.1 Data cleaning

```
[5]: ID        object
     M/F       object
     Hand      object
```

```
Age        int64
Educ       float64
SES        float64
MMSE       float64
CDR        float64
eTIV       int64
nWBV       float64
ASF        float64
Delay      float64
dtype: object
```

```
[6]:      M/F Hand  Age  Educ  SES  MMSE  CDR  eTIV   nWBV    ASF  Delay
     0     F    R   74   2.0  3.0  29.0  0.0  1344  0.743  1.306    NaN
     1     F    R   55   4.0  1.0  29.0  0.0  1147  0.810  1.531    NaN
     2     F    R   73   4.0  3.0  27.0  0.5  1454  0.708  1.207    NaN
     3     M    R   28   NaN  NaN   NaN  NaN  1588  0.803  1.105    NaN
     4     M    R   18   NaN  NaN   NaN  NaN  1737  0.848  1.010    NaN
     ..   ..   ..   ..   ...  ...   ...  ...   ...    ...    ...    ...
     431   M    R   20   NaN  NaN   NaN  NaN  1469  0.847  1.195    2.0
     432   M    R   22   NaN  NaN   NaN  NaN  1684  0.790  1.042   40.0
     433   M    R   22   NaN  NaN   NaN  NaN  1580  0.856  1.111   89.0
     434   F    R   20   NaN  NaN   NaN  NaN  1262  0.861  1.390    2.0
     435   F    R   26   NaN  NaN   NaN  NaN  1283  0.834  1.368   39.0

     [436 rows x 11 columns]
```

```
[7]: M/F      0
     Hand     0
     Age      0
     Educ   201
     SES    220
     MMSE   201
     CDR    201
     eTIV     0
     nWBV     0
     ASF      0
     Delay  416
     dtype: int64
```

```
[8]:      M/F Hand  Age  Educ  SES  MMSE  CDR  eTIV   nWBV    ASF
     0     F    R   74   2.0  3.0  29.0  0.0  1344  0.743  1.306
     1     F    R   55   4.0  1.0  29.0  0.0  1147  0.810  1.531
     2     F    R   73   4.0  3.0  27.0  0.5  1454  0.708  1.207
     3     M    R   28   NaN  NaN   NaN  NaN  1588  0.803  1.105
     4     M    R   18   NaN  NaN   NaN  NaN  1737  0.848  1.010
     ..   ..   ..   ..   ...  ...   ...  ...   ...    ...    ...
     431   M    R   20   NaN  NaN   NaN  NaN  1469  0.847  1.195
```

```
432   M    R    22    NaN   NaN   NaN   NaN   1684   0.790   1.042
433   M    R    22    NaN   NaN   NaN   NaN   1580   0.856   1.111
434   F    R    20    NaN   NaN   NaN   NaN   1262   0.861   1.390
435   F    R    26    NaN   NaN   NaN   NaN   1283   0.834   1.368

[436 rows x 10 columns]
```

[10]:
```
M/F     0
Hand    0
Age     0
Educ    0
SES     0
MMSE    0
CDR     0
eTIV    0
nWBV    0
ASF     0
dtype: int64
```

[11]:
```
      M/F  Age  Educ  SES  MMSE  CDR  eTIV   nWBV   ASF
0      F   74   2.0  3.0  29.0  0.0  1344  0.743  1.306
1      F   55   4.0  1.0  29.0  0.0  1147  0.810  1.531
2      F   73   4.0  3.0  27.0  0.5  1454  0.708  1.207
8      M   74   5.0  2.0  30.0  0.0  1636  0.689  1.073
9      F   52   3.0  2.0  30.0  0.0  1321  0.827  1.329
..    ..   ..   ...  ...   ...  ...   ...    ...    ...
411    F   70   1.0  4.0  29.0  0.5  1295  0.748  1.355
412    F   73   3.0  2.0  23.0  0.5  1536  0.730  1.142
413    F   61   2.0  4.0  28.0  0.0  1354  0.825  1.297
414    M   61   5.0  2.0  30.0  0.0  1637  0.780  1.072
415    F   62   3.0  3.0  26.0  0.0  1372  0.766  1.279

[216 rows x 9 columns]

<class 'pandas.core.frame.DataFrame'>
Int64Index: 216 entries, 0 to 415
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   M/F     216 non-null    object
 1   Age     216 non-null    int64
 2   Educ    216 non-null    float64
 3   SES     216 non-null    float64
 4   MMSE    216 non-null    float64
 5   CDR     216 non-null    float64
 6   eTIV    216 non-null    int64
 7   nWBV    216 non-null    float64
```

```
 8   ASF     216 non-null     float64
dtypes: float64(6), int64(2), object(1)
memory usage: 16.9+ KB
```

1.0.2 Feature Engineering

```
[13]: 0      0.0
      1      0.0
      2      0.5
      8      0.0
      9      0.0
              ...
      411    0.5
      412    0.5
      413    0.0
      414    0.0
      415    0.0
      Name: CDR, Length: 216, dtype: float64
```

```
[15]: 0      0.0
      1      0.0
      2      1.0
      8      0.0
      9      0.0
              ...
      411    1.0
      412    1.0
      413    0.0
      414    0.0
      415    0.0
      Name: CDR, Length: 216, dtype: float64
```

```
[69]:       M/F  Age  Educ  SES  MMSE  CDR  eTIV   nWBV   ASF
      0       0   74   2.0  3.0  29.0  0.0  1344  0.743  1.306
      1       0   55   4.0  1.0  29.0  0.0  1147  0.810  1.531
      2       0   73   4.0  3.0  27.0  1.0  1454  0.708  1.207
      8       1   74   5.0  2.0  30.0  0.0  1636  0.689  1.073
      9       0   52   3.0  2.0  30.0  0.0  1321  0.827  1.329

      ..     ...  ...   ...  ...   ...  ...   ...    ...    ...
      411     0   70   1.0  4.0  29.0  1.0  1295  0.748  1.355
      412     0   73   3.0  2.0  23.0  1.0  1536  0.730  1.142
      413     0   61   2.0  4.0  28.0  0.0  1354  0.825  1.297
      414     1   61   5.0  2.0  30.0  0.0  1637  0.780  1.072
      415     0   62   3.0  3.0  26.0  0.0  1372  0.766  1.279

      [216 rows x 9 columns]
```
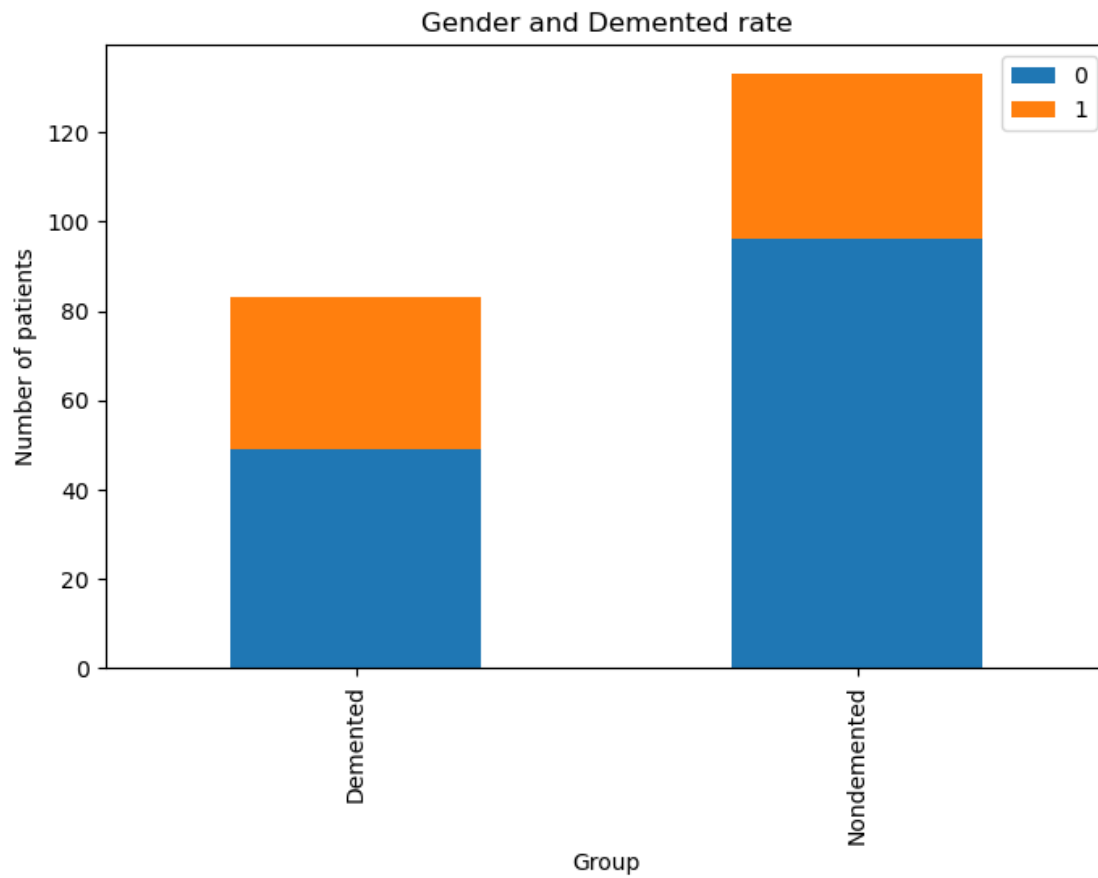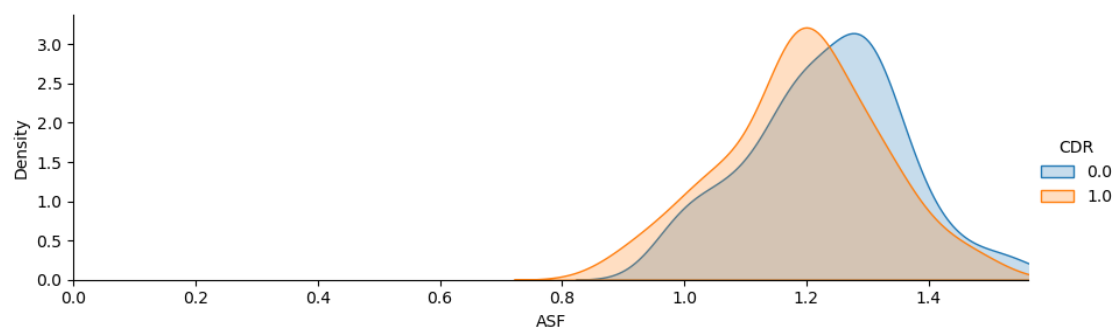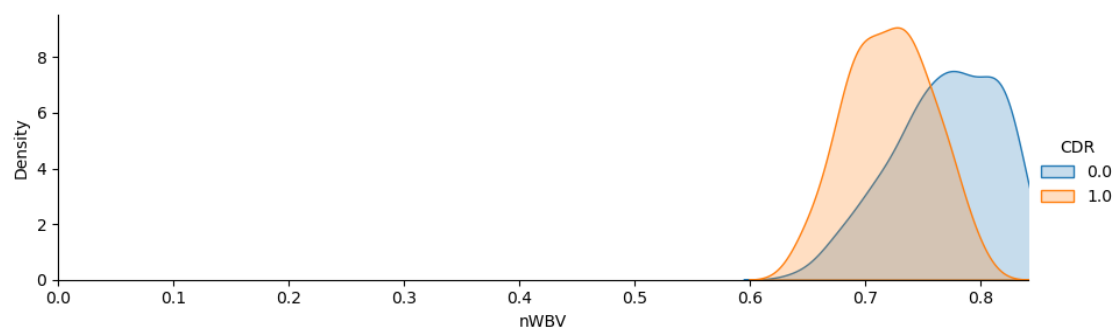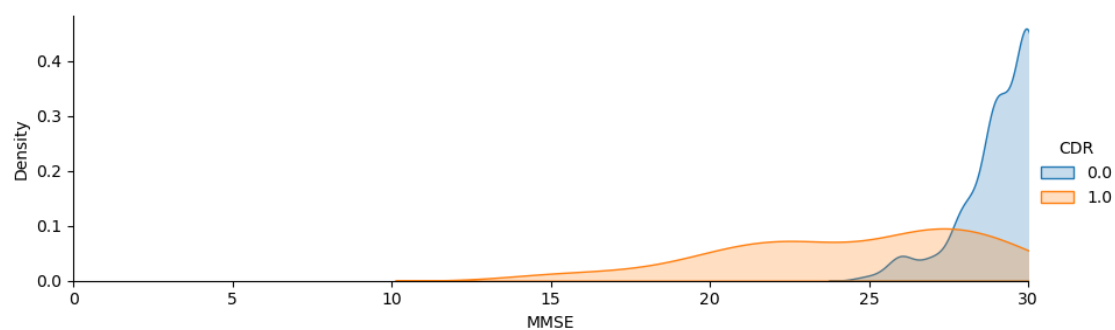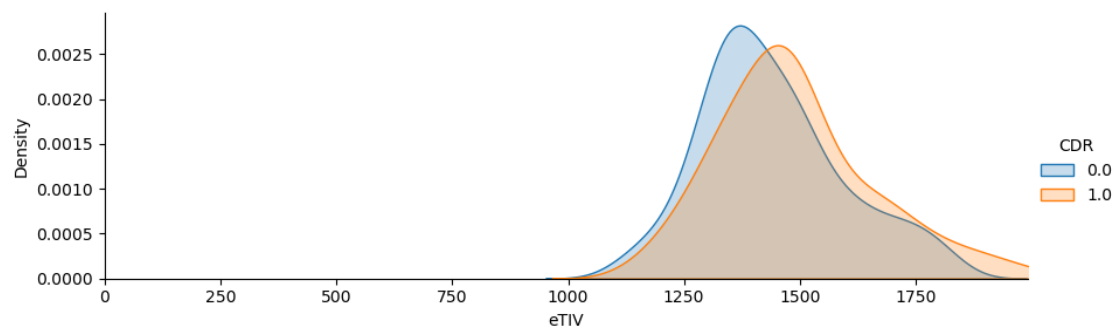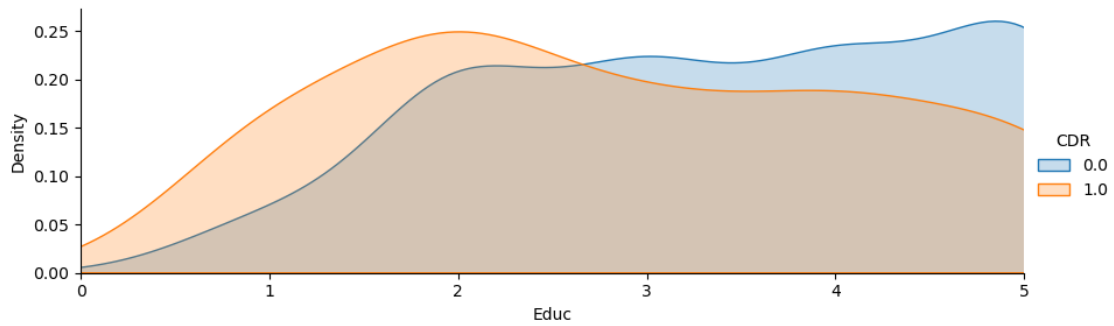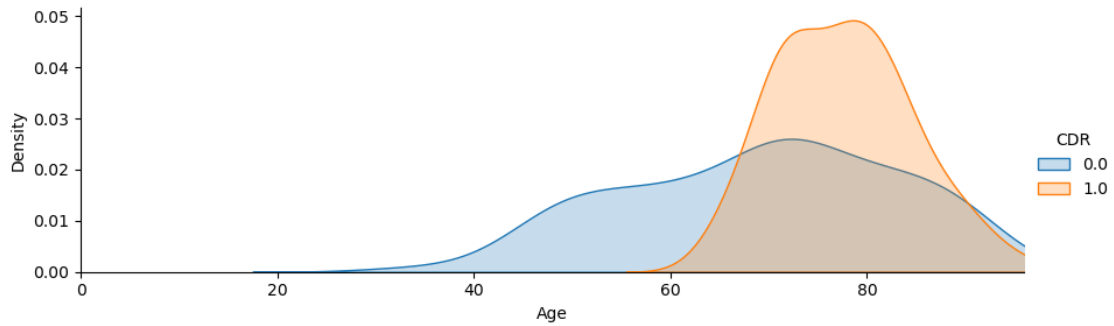
2 Exploratory Data Analysis

- Distribution of variables
- Exploratory PCAs
- Model implementations
  - Linear regression model with automatic feature selection
  - Automatic ML model selection (RandomforestClassifier)
  - Model testing & Results visualization

2.1 Distributions

[18]: Text(0.5, 1.0, 'Gender and Demented rate')

Gender and Demented rate

2.2 Exploratory Data analysis – PCAs

- General PCA
- PCA without age
- PCA with social factors only (no results because the dataset is only 2-D)
- PCA with brain measurement factors

```
Training set shape: (172, 8) (172,)
Test set shape: (44, 8) (44,)
```

```
[25]:            PC1        PC2  CDR
      0    -1.874614   1.613698  0.0
      1     1.736401   1.000079  0.0
      2    -0.479995  -1.221007  1.0
      3    -1.865039   2.591256  NaN
      4     2.742506  -1.813851  NaN
      ..         …          …  …
      411        NaN        NaN  1.0
      412        NaN        NaN  1.0
```
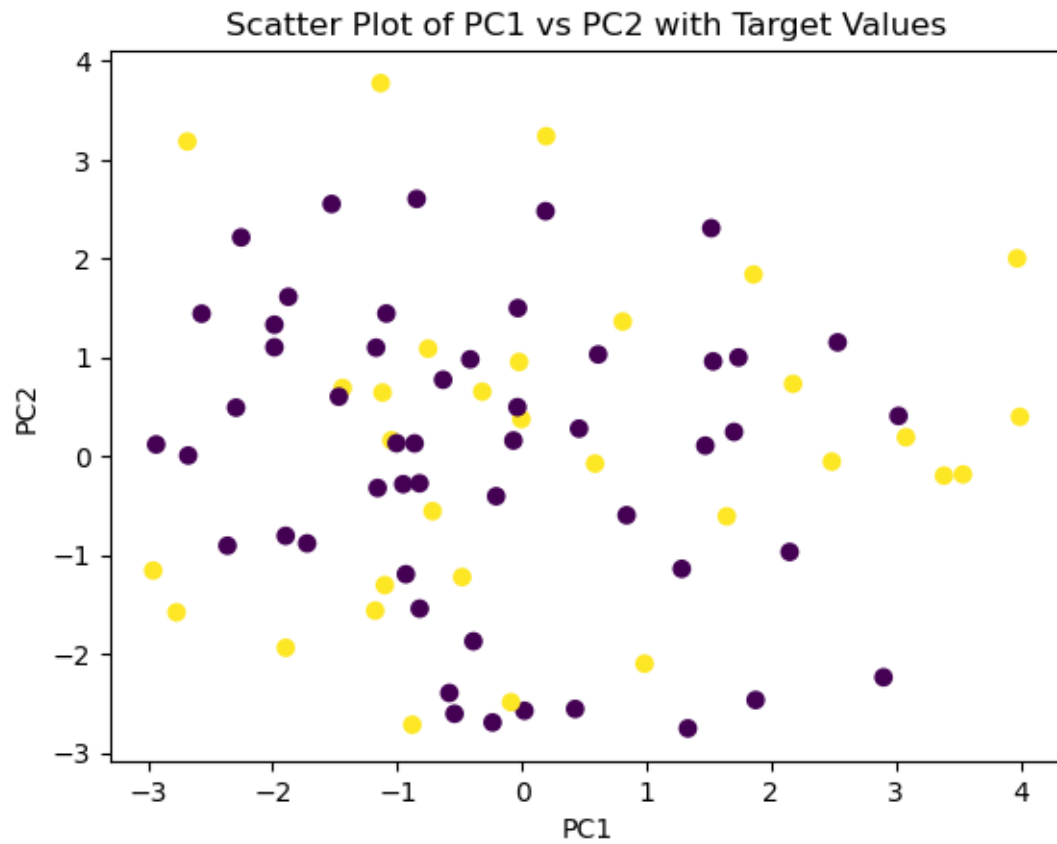
```
413        NaN        NaN  0.0
414        NaN        NaN  0.0
415        NaN        NaN  0.0

[306 rows x 3 columns]
```
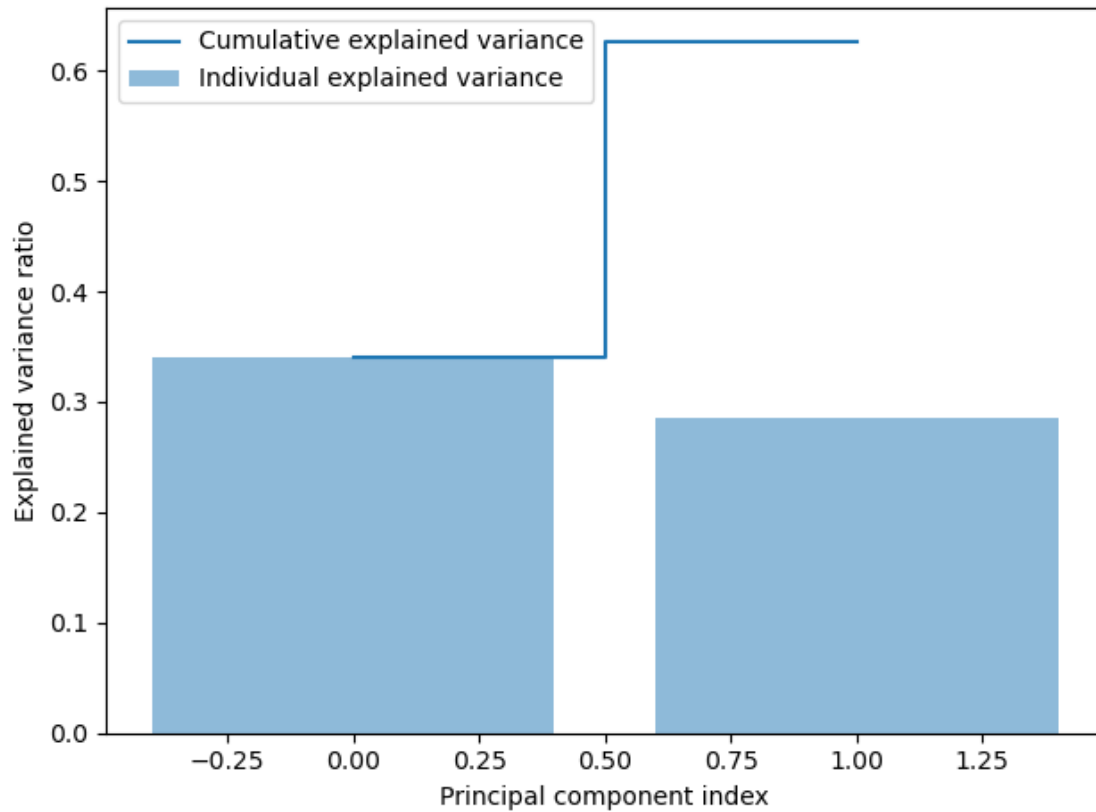
Scatter Plot of PC1 vs PC2 with Target Values


```
The explained variances by PC1 and PC2 is [0.34073802 0.28484481]
The total explained variance by PCA is 0.6255828327408242
```

2.2.1 Inferring the PCA results

```
[28]: array([[ 0.43829675, -0.03454681,  0.29904305, -0.30799529,  0.0638827 ,
               0.55369213, -0.10709964, -0.5479857 ],
             [ 0.09280395,  0.482203  , -0.34705162,  0.28885494, -0.47398471,
               0.12421983, -0.54564549, -0.1310138 ]])
```

```
[29]: array(['M/F', 'Age', 'Educ', 'SES', 'MMSE', 'eTIV', 'nWBV', 'ASF'],
           dtype='<U4')
```

```
[30]:    vars    to_PC1    to_PC2
     0   M/F   0.438297  0.092804
     1   Age  -0.034547  0.482203
     2  Educ   0.299043 -0.347052
     3   SES  -0.307995  0.288855
     4  MMSE   0.063883 -0.473985
     5  eTIV   0.553692  0.124220
     6  nWBV  -0.107100 -0.545645
     7   ASF  -0.547986 -0.131014
```

2.2.2 Data2 without Age

Since Age is a huge predictor of Alzheimer's, it is possible that the PCA results were skewed by the Age feature.

So this section tries the entire process without Age

```
[31]:        M/F  Educ  SES  MMSE  CDR  eTIV   nWBV   ASF
        0      0   2.0  3.0  29.0  0.0  1344  0.743  1.306
        1      0   4.0  1.0  29.0  0.0  1147  0.810  1.531
        2      0   4.0  3.0  27.0  1.0  1454  0.708  1.207
        8      1   5.0  2.0  30.0  0.0  1636  0.689  1.073
        9      0   3.0  2.0  30.0  0.0  1321  0.827  1.329
        ..    ...  ...  ...  ...   ...   ...    ...    ...
        411    0   1.0  4.0  29.0  1.0  1295  0.748  1.355
        412    0   3.0  2.0  23.0  1.0  1536  0.730  1.142
        413    0   2.0  4.0  28.0  0.0  1354  0.825  1.297
        414    1   5.0  2.0  30.0  0.0  1637  0.780  1.072
        415    0   3.0  3.0  26.0  0.0  1372  0.766  1.279

        [216 rows x 8 columns]

        Training set shape: (172, 7) (172,)
        Test set shape: (44, 7) (44,)

[34]:          PC1        PC2  CDR
        0   -1.816204   1.733007  0.0
        1    1.759053   1.535174  0.0
        2   -0.505545  -1.925298  1.0
        3   -1.774946   2.793992  NaN
        4    2.646523  -0.964920  NaN
        ..      ...        ...   ...
        411      NaN        NaN  1.0
        412      NaN        NaN  1.0
        413      NaN        NaN  0.0
        414      NaN        NaN  0.0
        415      NaN        NaN  0.0

        [306 rows x 3 columns]
```
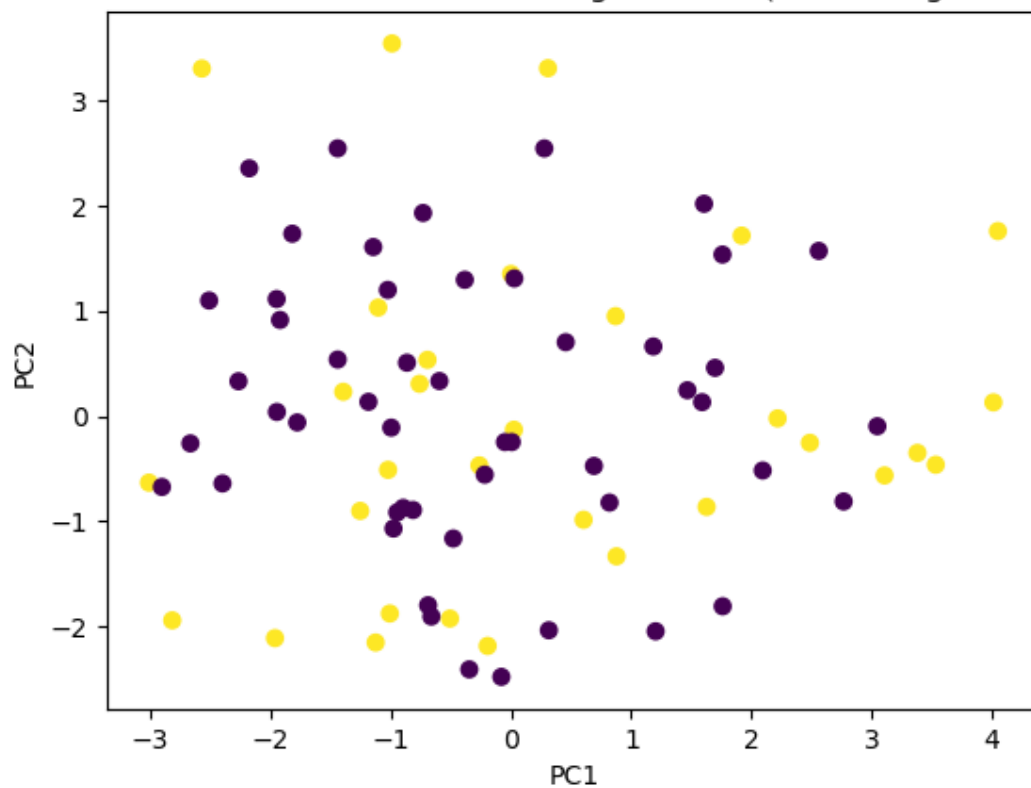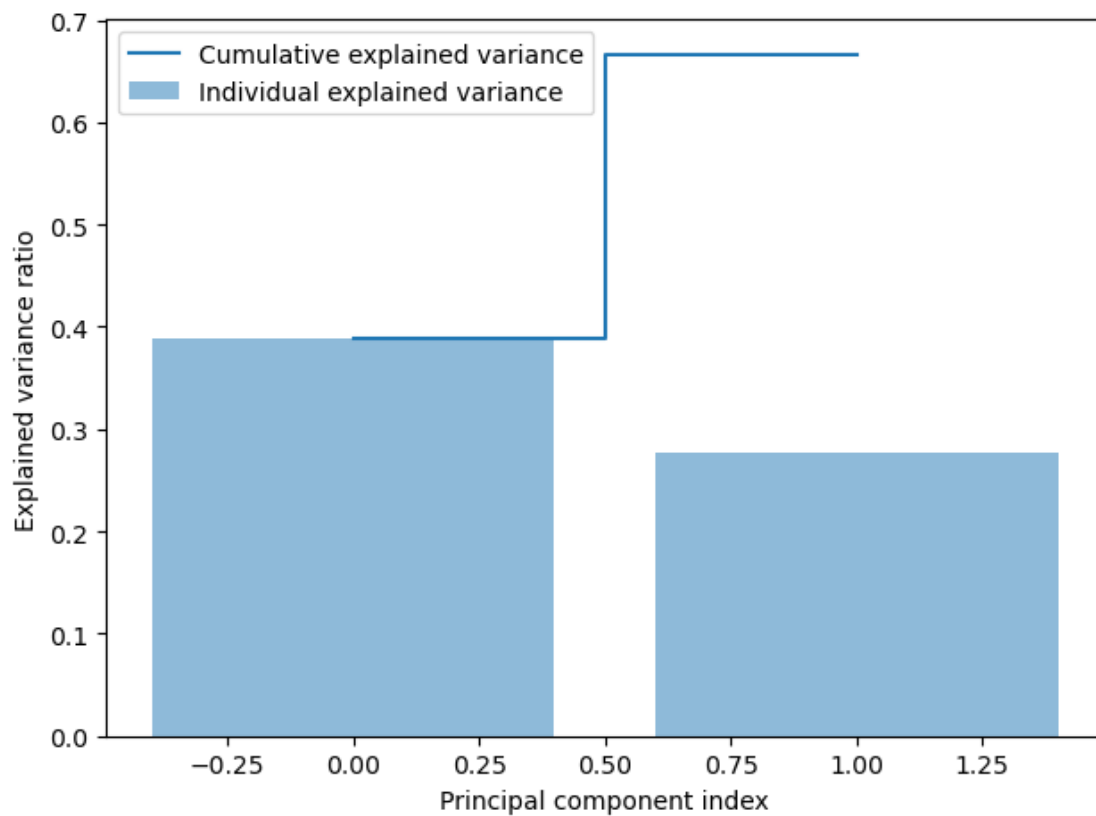
Scatter Plot of PC1 vs PC2 with Target Values (without Age feature)

The explained variances by PC1 and PC2 is [0.38922668 0.27753814]
The total explained variance by PCA is 0.666764812569268

```
[37]:    vars     to_PC1     to_PC2
      0   M/F   0.439875   0.149675
      1  Educ   0.289599  -0.498214
      2   SES  -0.300694   0.450204
      3  MMSE   0.049019  -0.526259
      4  eTIV   0.556647   0.159572
      5  nWBV  -0.127619  -0.441487
      6   ASF  -0.551072  -0.171388

[39]:      M/F  Age  MMSE  CDR  eTIV   nWBV   ASF
      0      0   74  29.0  0.0  1344  0.743  1.306
      1      0   55  29.0  0.0  1147  0.810  1.531
      2      0   73  27.0  1.0  1454  0.708  1.207
      8      1   74  30.0  0.0  1636  0.689  1.073
      9      0   52  30.0  0.0  1321  0.827  1.329
      ..    ..   ..   ...   ..   ...    ...   ...
      411    0   70  29.0  1.0  1295  0.748  1.355
      412    0   73  23.0  1.0  1536  0.730  1.142
      413    0   61  28.0  0.0  1354  0.825  1.297
      414    1   61  30.0  0.0  1637  0.780  1.072
      415    0   62  26.0  0.0  1372  0.766  1.279
```
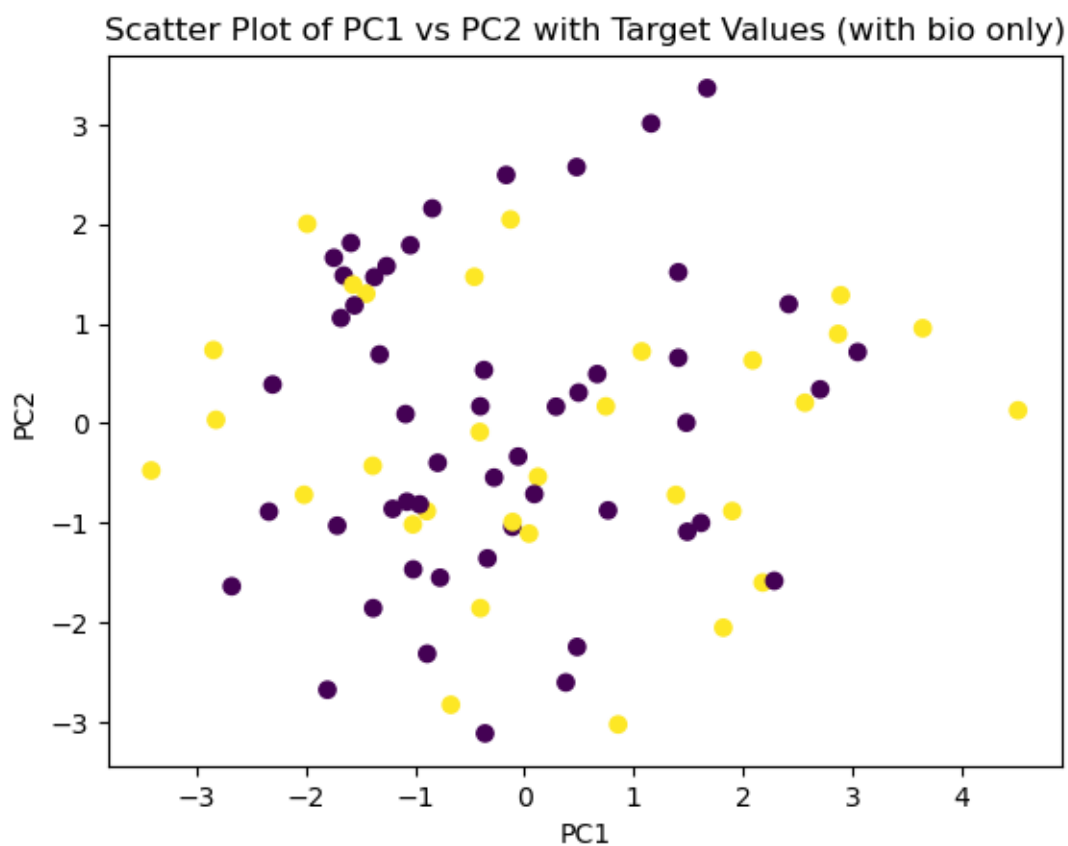
```
[216 rows x 7 columns]

Training set shape: (172, 6) (172, 6)
Test set shape: (44,) (44,)
```

[42]:
```
          PC1        PC2  CDR
0    -0.770960  -1.550299  0.0
1     2.419336   1.196371  0.0
2    -1.386507  -0.426424  1.0
3    -0.276261  -2.160778  NaN
4     1.719195   2.819163  NaN
..         ...        ...  ...
411        NaN        NaN  1.0
412        NaN        NaN  1.0
413        NaN        NaN  0.0
414        NaN        NaN  0.0
415        NaN        NaN  0.0

[306 rows x 3 columns]
```


Scatter Plot of PC1 vs PC2 with Target Values (with bio only)

```
[44]:    vars     to_PC1     to_PC2
    0   M/F   0.461271   0.184471
    1   Age   0.134123  -0.587461
    2  MMSE  -0.131233   0.468814
    3  eTIV   0.577746   0.194324
    4  nWBV  -0.292716   0.571614
    5   ASF  -0.576667  -0.191230
```

2.3   Model implementation – (1) Linear regression Model

With Automatic Feature Selection

```
                                          model        RSS
0  <statsmodels.regression.linear_model.Regressio…   21.324458
1  <statsmodels.regression.linear_model.Regressio…   20.374411
2  <statsmodels.regression.linear_model.Regressio…   22.886511
3  <statsmodels.regression.linear_model.Regressio…   26.465201
4  <statsmodels.regression.linear_model.Regressio…   19.488219
5  <statsmodels.regression.linear_model.Regressio…   19.336666
6  <statsmodels.regression.linear_model.Regressio…   19.296314
7  <statsmodels.regression.linear_model.Regressio…   19.393491
                       OLS Regression Results
================================================================================
=======
Dep. Variable:                    CDR   R-squared (uncentered):
0.708
Model:                            OLS   Adj. R-squared (uncentered):
0.695
Method:                 Least Squares   F-statistic:
57.05
Date:                Sat, 09 Dec 2023   Prob (F-statistic):
7.27e-41
Time:                        15:48:43   Log-Likelihood:
-55.926
No. Observations:                 172   AIC:
125.9
Df Residuals:                     165   BIC:
147.9
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
M/F            0.0651      0.068      0.952      0.342      -0.070       0.200
Educ           0.0195      0.031      0.621      0.535      -0.042       0.081
SES            0.0454      0.035      1.286      0.200      -0.024       0.115
MMSE          -0.0705      0.009     -7.944      0.000      -0.088      -0.053
eTIV           0.0015      0.000      8.978      0.000       0.001       0.002
```

```
nWBV           -2.7252      0.646     -4.222      0.000     -4.000     -1.451
ASF             1.5881      0.240      6.610      0.000      1.114      2.063
==============================================================================
Omnibus:                       10.640   Durbin-Watson:                  1.803
Prob(Omnibus):                  0.005   Jarque-Bera (JB):              11.549
Skew:                           0.629   Prob(JB):                     0.00311
Kurtosis:                       2.836   Cond. No.                    3.78e+04
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[3] The condition number is large, 3.78e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[48]: `<class 'statsmodels.iolib.table.SimpleTable'>`

```
Empty DataFrame
Columns: []
Index: []
```

2.4 Model implementation – (2) Automatic ML Model Selection

```
[73]: M/F        int64
      Age        int64
      Educ     float64
      SES      float64
      MMSE     float64
      CDR      float64
      eTIV       int64
      nWBV     float64
      ASF      float64
      dtype: object
```

```
[78]: M/F       object
      Age      float64
      Educ     float64
      SES      float64
      MMSE     float64
      CDR       object
      eTIV     float64
      nWBV     float64
      ASF      float64
      dtype: object
```

```
Best Model: Pipeline(steps=[('preprocessor',
```

```
            ColumnTransformer(transformers=[('num',
                                            Pipeline(steps=[('scaler',
    StandardScaler())]),
                                            Index(['M/F', 'Age', 'Educ',
    'SES', 'MMSE', 'eTIV', 'nWBV', 'ASF'], dtype='object')),
                                            ('cat',
                                            Pipeline(steps=[('onehot',
    OneHotEncoder(handle_unknown='ignore'))]),
                                            Index([], dtype='object'))])),
                ('classifier', RandomForestClassifier())])
    Cross-Validation Accuracy: 0.8601680672268905
    Test Accuracy: 0.7727272727272727

    Accuracy: 0.7272727272727273
```

[126]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at
       0x7fc2439d2610>
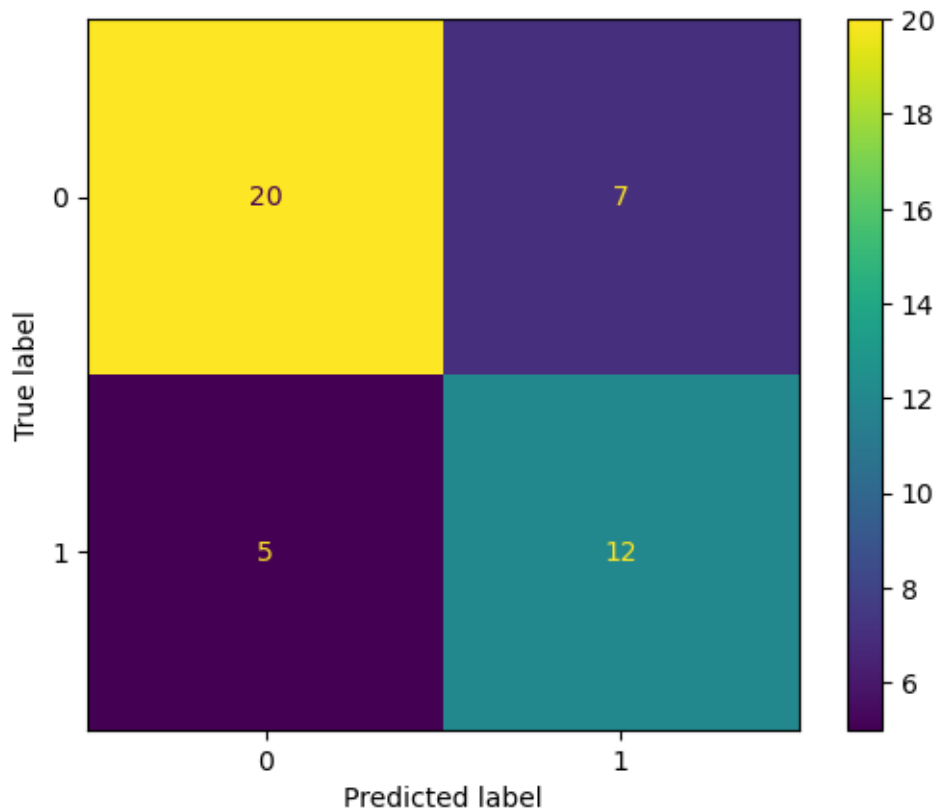
2.5 Codes not working

Requirement already satisfied: graphviz in /Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (0.20.1)

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
Input In [116], in <cell line: 15>()
     12 graph = graphviz.Source(dot_data)
     13 #display(graph)
---> 15 with open("tree") as f:
     16     dot_graph = f.read()
     17     display(graphviz.Source(dot_graph))

FileNotFoundError: [Errno 2] No such file or directory: 'tree'
```

```
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-
packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed
to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
```

[89]: LogisticRegression()

2.5.1 Alternative model selection method: MDR

(Trying to implement)

```
Collecting scikit-mdr
  Downloading scikit_MDR-0.4.5-py3-none-any.whl (15 kB)
Requirement already satisfied: scikit-learn in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from scikit-mdr)
(1.2.2)
Requirement already satisfied: scipy in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from scikit-mdr)
(1.10.1)
Requirement already satisfied: numpy in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from scikit-mdr)
(1.21.5)
Requirement already satisfied: matplotlib in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from scikit-mdr)
```

```
(3.7.1)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (4.25.0)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (2.8.2)
Requirement already satisfied: cycler>=0.10 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (0.11.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (3.0.9)
Requirement already satisfied: kiwisolver>=1.0.1 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (1.4.4)
Requirement already satisfied: packaging>=20.0 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (23.0)
Requirement already satisfied: importlib-resources>=3.2.0 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (5.2.0)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (1.0.5)
Requirement already satisfied: pillow>=6.2.0 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from
matplotlib->scikit-mdr) (9.4.0)
Requirement already satisfied: zipp>=3.1.0 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from importlib-
resources>=3.2.0->matplotlib->scikit-mdr) (3.11.0)
Requirement already satisfied: six>=1.5 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from python-
dateutil>=2.7->matplotlib->scikit-mdr) (1.16.0)
Requirement already satisfied: joblib>=1.1.1 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from scikit-
learn->scikit-mdr) (1.1.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/Users/yufeimeng/opt/anaconda3/lib/python3.9/site-packages (from scikit-
learn->scikit-mdr) (2.2.0)
Installing collected packages: scikit-mdr
Successfully installed scikit-mdr-0.4.5
Note: you may need to restart the kernel to use updated packages.
```

[60]:

	M/F	Age	Educ	SES	MMSE	eTIV	nWBV	ASF
352	0	77	2.0	4.0	22.0	1350	0.736	1.300
300	1	72	1.0	3.0	29.0	1734	0.762	1.012

```
200    0   75    5.0   1.0   30.0   1317   0.742   1.332
212    0   77    1.0   4.0   20.0   1376   0.701   1.275
189    1   51    5.0   2.0   29.0   1714   0.819   1.024

..    ...  ...   ...   ...   ...    ...    ...
153    0   74    2.0   3.0   29.0   1395   0.787   1.258
17     0   89    5.0   1.0   30.0   1536   0.715   1.142
152    0   81    2.0   3.0   28.0   1495   0.687   1.174
262    1   83    3.0   2.0   26.0   1992   0.706   0.881
263    0   73    2.0   2.0   19.0   1274   0.745   1.377

[172 rows x 8 columns]
```

```
---------------------------------------------------------------------------
KeyError                                    Traceback (most recent call last)
File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/indexes/base.py:
 ↪3802, in Index.get_loc(self, key, method, tolerance)
   3801 try:
-> 3802     return self._engine.get_loc(casted_key)
   3803 except KeyError as err:


File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/_libs/index.pyx:138, in
 ↪pandas._libs.index.IndexEngine.get_loc()


File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/_libs/index.pyx:165, in
 ↪pandas._libs.index.IndexEngine.get_loc()


File pandas/_libs/hashtable_class_helper.pxi:5745, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()


File pandas/_libs/hashtable_class_helper.pxi:5753, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()


KeyError: 0


The above exception was the direct cause of the following exception:


KeyError                                    Traceback (most recent call last)
Input In [61], in <cell line: 1>()
----> 1 my_mdr.fit(features, labels)
      2 my_mdr.transform(features)


File ~/opt/anaconda3/lib/python3.9/site-packages/mdr/mdr.py:81, in MDRBase.
 ↪fit(self, features, class_labels)
     79 self.class_count_matrix = defaultdict(lambda: defaultdict(int))
     80 for row_i in range(features.shape[0]):
---> 81     feature_instance = tuple(features[row_i])
     82     self.class_count_matrix[feature_instance][class_labels[row_i]] += 1
```

19

```
      83 self.class_count_matrix = dict(self.class_count_matrix)

File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/frame.py:3807, in␣
 ↪DataFrame.__getitem__(self, key)
   3805 if self.columns.nlevels > 1:
   3806     return self._getitem_multilevel(key)
-> 3807 indexer = self.columns.get_loc(key)
   3808 if is_integer(indexer):
   3809     indexer = [indexer]

File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/indexes/base.py:
 ↪3804, in Index.get_loc(self, key, method, tolerance)
   3802     return self._engine.get_loc(casted_key)
   3803 except KeyError as err:
-> 3804     raise KeyError(key) from err
   3805 except TypeError:
   3806     # If we have a listlike key, _check_indexing_error will raise
   3807     #  InvalidIndexError. Otherwise we fall through and re-raise
   3808     #  the TypeError.
   3809     self._check_indexing_error(key)

KeyError: 0
```