hospital-readmission-prediction

November 8, 2023

1 Importing necessary libraries

[225]:

```
import pandas as pd
import numpy as np
import tensorflow as tf
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.metrics import accuracy_score, classification_report,␣
  ↪confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
```

2 Importing dataset

[226]:

```
df = pd.read_csv('/content/readmissions.csv')
df.head()
```

[226]:

```
   time_in_hospital  num_lab_procedures  num_procedures  num_medications  \
0                14                  41               0               11
1                 2                  30               0               12
2                 5                  66               0               22
3                 3                  63               0                8
4                 5                  40               0                6

   number_outpatient  number_emergency  number_inpatient  number_diagnoses  \
0                  0                 0                 0                 6
1                  0                 0                 1                 9
2                  1                 0                 2                 9
3                  0                 0                 0                 8
4                  0                 0                 1                 9

   race_Caucasian  race_AfricanAmerican  ...  citoglipton_No  insulin_No  \
0            True                 False  ...            True        True
1            True                 False  ...            True       False
2            True                 False  ...            True        True
```

```
3            True                 False  ...          True        True
4            True                 False  ...          True        True

   glyburide-metformin_No  glipizide-metformin_No  \
0                    True                    True
1                    True                    True
2                    True                    True
3                    True                    True
4                    True                    True

   glimepiride-pioglitazone_No  metformin-rosiglitazone_No  \
0                         True                        True
1                         True                        True
2                         True                        True
3                         True                        True
4                         True                        True

   metformin-pioglitazone_No  change_No  diabetesMed_Yes  readmitted
0                       True       True             True           0
1                       True      False             True           1
2                       True       True             True           1
3                       True       True             True           1
4                       True       True            False           0

[5 rows x 65 columns]
```

3 Exploring Dataset

3.1 Checking shape of dataset

[227]:

```
df.shape
```

[227]:

```
(25000, 65)
```

3.2 Getting info of dataset

[228]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25000 entries, 0 to 24999
Data columns (total 65 columns):
 #   Column                                   Non-Null Count  Dtype
---  ------                                   --------------  -----
 0   time_in_hospital                         25000 non-null  int64
 1   num_lab_procedures                       25000 non-null  int64
 2   num_procedures                           25000 non-null  int64
```

```
3   num_medications                           25000 non-null  int64
4   number_outpatient                         25000 non-null  int64
5   number_emergency                          25000 non-null  int64
6   number_inpatient                          25000 non-null  int64
7   number_diagnoses                          25000 non-null  int64
8   race_Caucasian                            25000 non-null  bool
9   race_AfricanAmerican                      25000 non-null  bool
10  gender_Female                             25000 non-null  bool
11  age_[70-80)                               25000 non-null  bool
12  age_[60-70)                               25000 non-null  bool
13  age_[50-60)                               25000 non-null  bool
14  age_[80-90)                               25000 non-null  bool
15  age_[40-50)                               25000 non-null  bool
16  payer_code_?                              25000 non-null  bool
17  payer_code_MC                             25000 non-null  bool
18  payer_code_HM                             25000 non-null  bool
19  payer_code_SP                             25000 non-null  bool
20  payer_code_BC                             25000 non-null  bool
21  medical_specialty_?                       25000 non-null  bool
22  medical_specialty_InternalMedicine        25000 non-null  bool
23  medical_specialty_Emergency/Trauma        25000 non-null  bool
24  medical_specialty_Family/GeneralPractice  25000 non-null  bool
25  medical_specialty_Cardiology              25000 non-null  bool
26  diag_1_428                                25000 non-null  bool
27  diag_1_414                                25000 non-null  bool
28  diag_1_786                                25000 non-null  bool
29  diag_2_276                                25000 non-null  bool
30  diag_2_428                                25000 non-null  bool
31  diag_2_250                                25000 non-null  bool
32  diag_2_427                                25000 non-null  bool
33  diag_3_250                                25000 non-null  bool
34  diag_3_401                                25000 non-null  bool
35  diag_3_276                                25000 non-null  bool
36  diag_3_428                                25000 non-null  bool
37  max_glu_serum_None                        25000 non-null  bool
38  A1Cresult_None                            25000 non-null  bool
39  metformin_No                              25000 non-null  bool
40  repaglinide_No                            25000 non-null  bool
41  nateglinide_No                            25000 non-null  bool
42  chlorpropamide_No                         25000 non-null  bool
43  glimepiride_No                            25000 non-null  bool
44  acetohexamide_No                          25000 non-null  bool
45  glipizide_No                              25000 non-null  bool
46  glyburide_No                              25000 non-null  bool
47  tolbutamide_No                            25000 non-null  bool
48  pioglitazone_No                           25000 non-null  bool
49  rosiglitazone_No                          25000 non-null  bool
50  acarbose_No                               25000 non-null  bool
```

```
 51  miglitol_No                          25000 non-null  bool
 52  troglitazone_No                      25000 non-null  bool
 53  tolazamide_No                        25000 non-null  bool
 54  examide_No                           25000 non-null  bool
 55  citoglipton_No                       25000 non-null  bool
 56  insulin_No                           25000 non-null  bool
 57  glyburide-metformin_No               25000 non-null  bool
 58  glipizide-metformin_No               25000 non-null  bool
 59  glimepiride-pioglitazone_No          25000 non-null  bool
 60  metformin-rosiglitazone_No           25000 non-null  bool
 61  metformin-pioglitazone_No            25000 non-null  bool
 62  change_No                            25000 non-null  bool
 63  diabetesMed_Yes                      25000 non-null  bool
 64  readmitted                           25000 non-null  int64
dtypes: bool(56), int64(9)
memory usage: 3.1 MB
```

3.3 Converting all Boolean values to integer values

[229]:

```
df *= 1
df.head()
```

[229]:

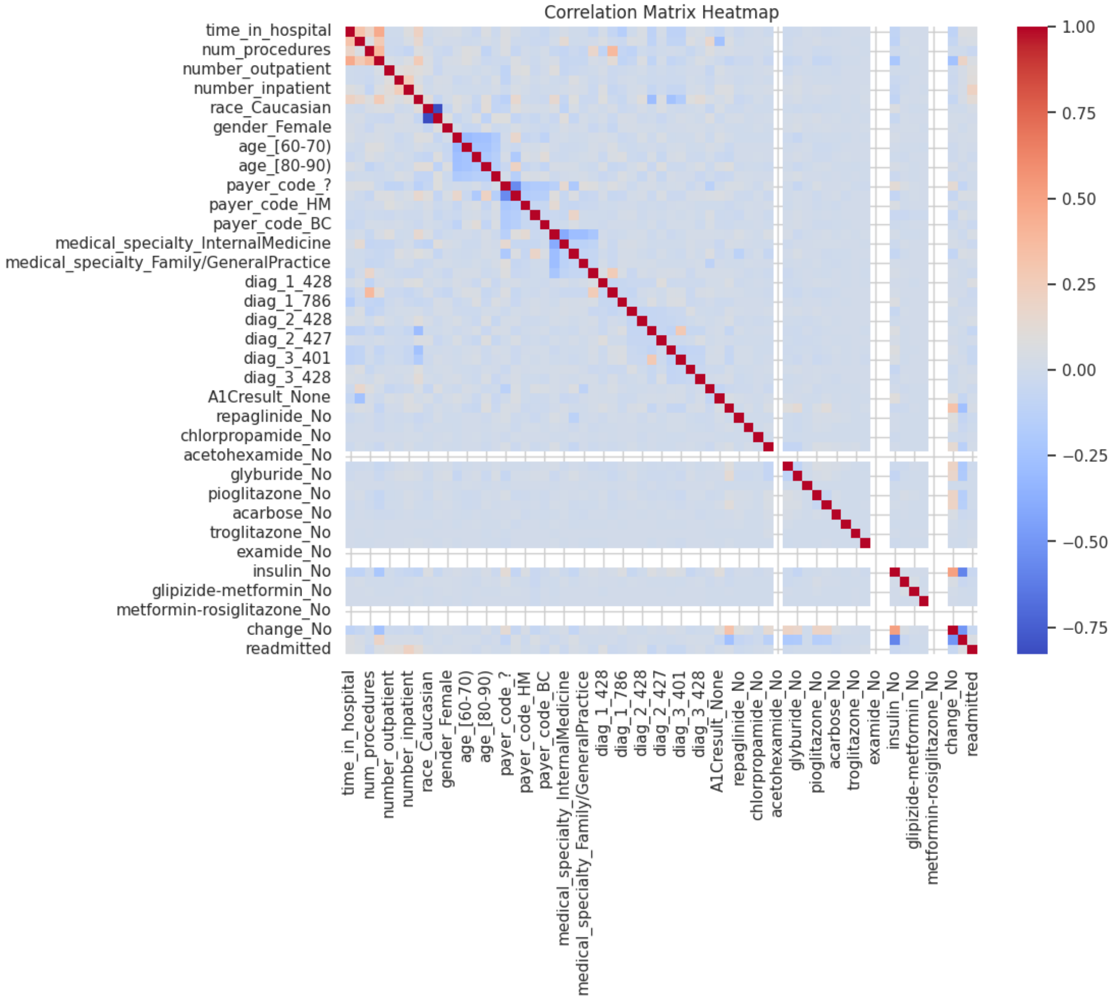
```
   time_in_hospital  num_lab_procedures  num_procedures  num_medications  \
0                14                  41               0               11
1                 2                  30               0               12
2                 5                  66               0               22
3                 3                  63               0                8
4                 5                  40               0                6

   number_outpatient  number_emergency  number_inpatient  number_diagnoses  \
0                  0                 0                 0                 6
1                  0                 0                 1                 9
2                  1                 0                 2                 9
3                  0                 0                 0                 8
4                  0                 0                 1                 9

   race_Caucasian  race_AfricanAmerican  …  citoglipton_No  insulin_No  \
0               1                     0  …               1           1
1               1                     0  …               1           0
2               1                     0  …               1           1
3               1                     0  …               1           1
4               1                     0  …               1           1

   glyburide-metformin_No  glipizide-metformin_No  \
0                       1                       1
1                       1                       1
2                       1                       1
```

```
3                       1                       1
4                       1                       1

   glimepiride-pioglitazone_No  metformin-rosiglitazone_No  \
0                            1                           1
1                            1                           1
2                            1                           1
3                            1                           1
4                            1                           1

   metformin-pioglitazone_No  change_No  diabetesMed_Yes  readmitted
0                          1          1                1           0
1                          1          0                1           1
2                          1          1                1           1
3                          1          1                1           1
4                          1          1                0           0

[5 rows x 65 columns]
```

3.4 Plotting the correlation matrix

[230]:

```python
correlation_matrix = df.corr()

# Create a heatmap of the correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=False, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix Heatmap')
plt.show()
```



3.5 Dropping any possible null values

```
[231]: df = df.dropna()
       df.head()
```

```
[231]:    time_in_hospital  num_lab_procedures  num_procedures  num_medications  \
       0                14                  41               0               11
       1                 2                  30               0               12
       2                 5                  66               0               22
       3                 3                  63               0                8
       4                 5                  40               0                6

          number_outpatient  number_emergency  number_inpatient  number_diagnoses  \
       0                  0                 0                 0                 6
       1                  0                 0                 1                 9
       2                  1                 0                 2                 9
```

```
3                  0                 0                  0                 8
4                  0                 0                  1                 9

   race_Caucasian  race_AfricanAmerican  …   citoglipton_No  insulin_No  \
0               1                     0  …                1           1
1               1                     0  …                1           0
2               1                     0  …                1           1
3               1                     0  …                1           1
4               1                     0  …                1           1

   glyburide-metformin_No  glipizide-metformin_No  \
0                       1                       1
1                       1                       1
2                       1                       1
3                       1                       1
4                       1                       1

   glimepiride-pioglitazone_No  metformin-rosiglitazone_No  \
0                            1                           1
1                            1                           1
2                            1                           1
3                            1                           1
4                            1                           1

   metformin-pioglitazone_No  change_No  diabetesMed_Yes  readmitted
0                          1          1                1           0
1                          1          0                1           1
2                          1          1                1           1
3                          1          1                1           1
4                          1          1                0           0

[5 rows x 65 columns]
```

4 Preparing the dataset

4.1 Extracting features and target from the dataset

[232]:

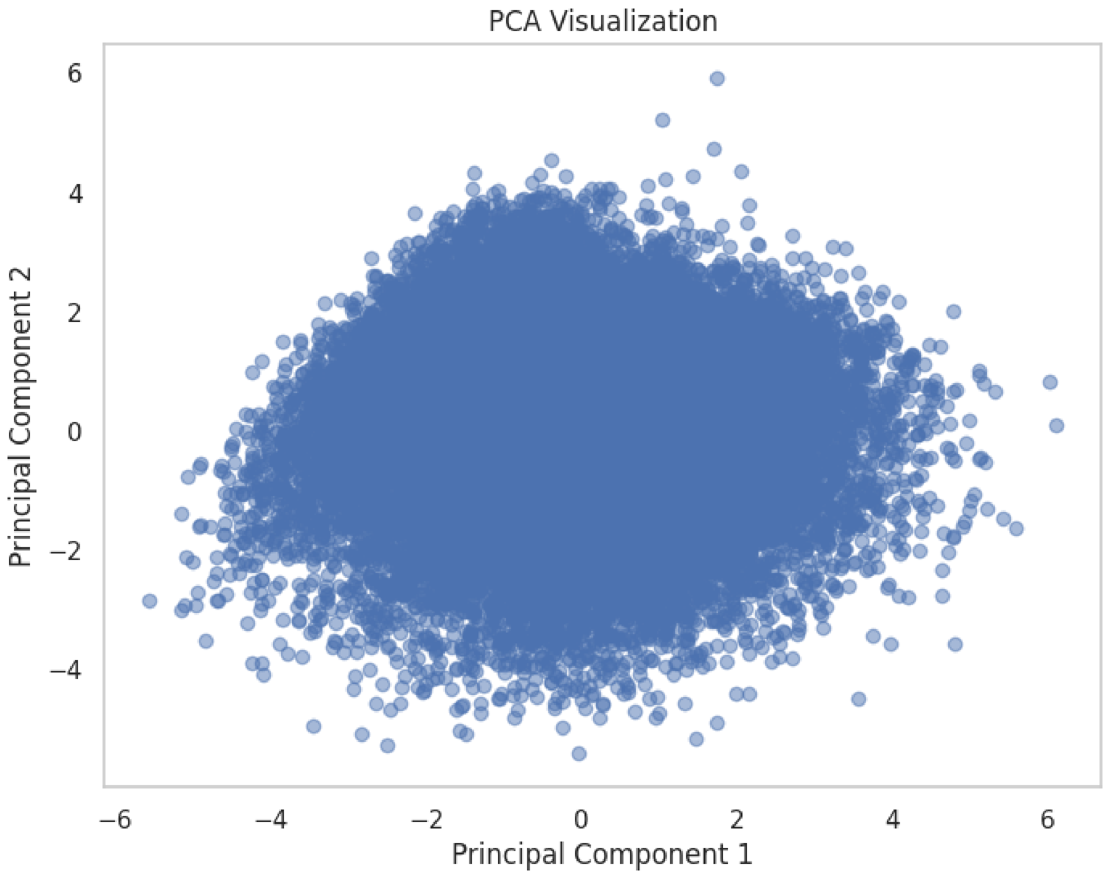
```
X = df.drop('readmitted', axis=1)
y = df['readmitted']
```

4.2 Visualizing the features using PCA

[233]:

```
scaler = StandardScaler()
X_vis = scaler.fit_transform(X)

n_components = 2
```

```
pca = PCA(n_components=n_components)
X_vis = pca.fit_transform(X_vis)

plt.figure(figsize=(8, 6))
plt.scatter(X_vis[:,0], X_vis[:,1], alpha=0.5)
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('PCA Visualization')
plt.grid()
plt.show()
```



4.3 Splitting the dataset into Training, Testing and Validation dataset

[234]:
```
X_train, X_temp, y_train, y_temp = train_test_split(X, y, test_size=0.3,␣
  ↪random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5,␣
  ↪random_state=42)
```

5 Model Training

5.1 Initializing the model

[235]:

```
model = tf.keras.Sequential([
    tf.keras.layers.Input(shape=(X_train.shape[1],)),
    tf.keras.layers.Dense(64, activation='sigmoid'),
    tf.keras.layers.Dense(32, activation='sigmoid'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
```

5.2 Compiling the model

[236]:

```
model.compile(optimizer='adam', loss='binary_crossentropy',␣
 ↪metrics=['accuracy'])
```

5.3 Training

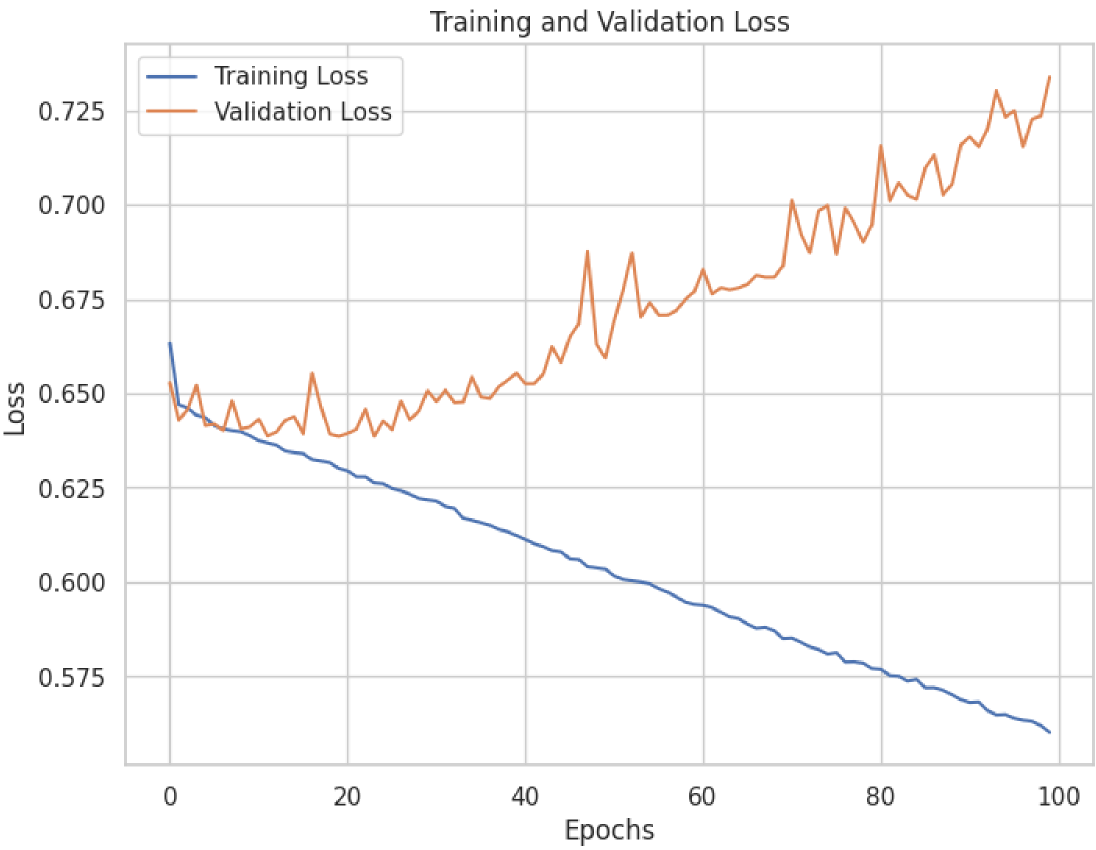
[237]:

```
history = model.fit(X_train, y_train, epochs=100, batch_size=20,␣
 ↪validation_data=(X_val, y_val), verbose=0)
```

5.4 Plotting the Training and Validation losses

[238]:

```
plt.figure(figsize=(8, 6))
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.title('Training and Validation Loss')
plt.legend()
plt.show()
```



6 Model Evaluation

[239]:
```
_, accuracy = model.evaluate(X_test, y_test)
print("Test Accuracy: ", accuracy)
```

```
118/118 [==============================] - 0s 2ms/step - loss: 0.7325 -
accuracy: 0.6067
Test Accuracy:  0.6066666841506958
```

6.1 Predicting binary output values from model output

[240]:
```
y_pred = model.predict(X_test)

# Convert probabilities to binary predictions
y_pred_binary = (y_pred > 0.5).astype(int)
```

```
118/118 [==============================] - 0s 2ms/step
```

6.2 Visualizing model results

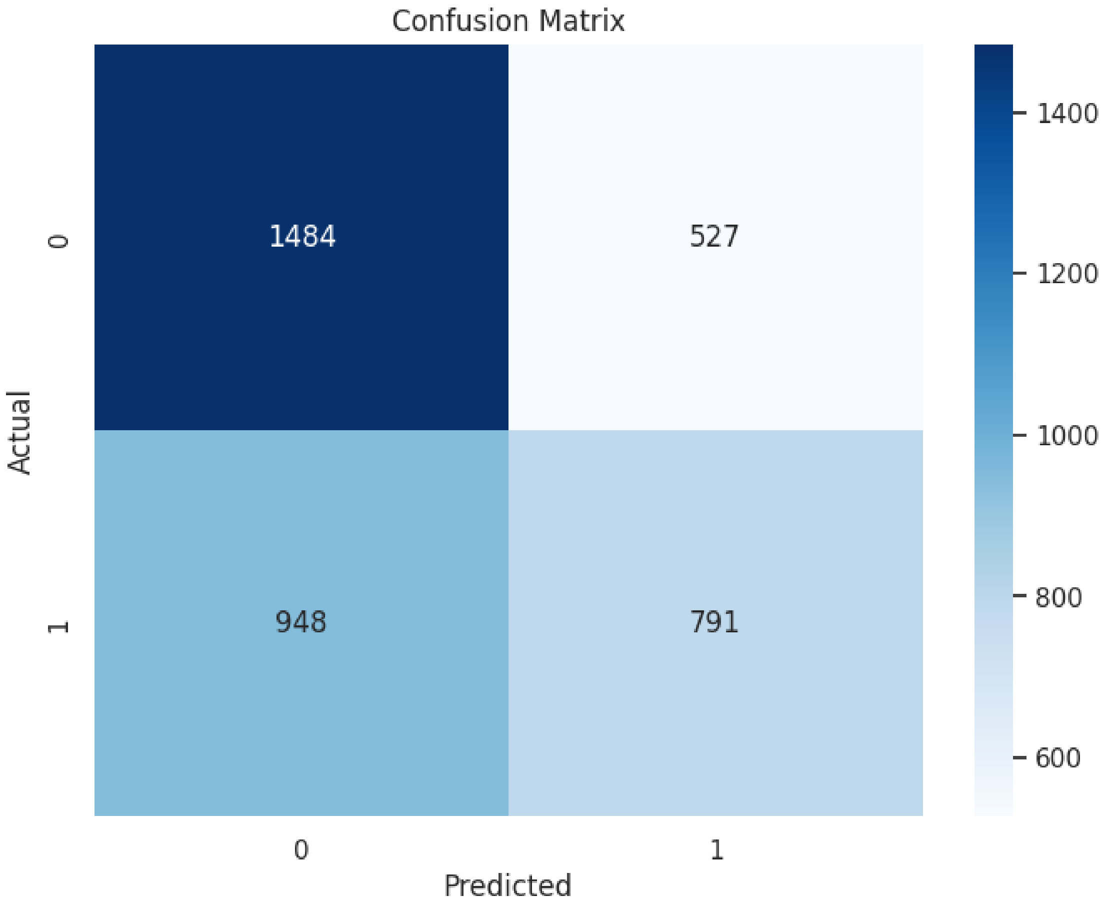
[241]:

```
# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred_binary)
print("Accuracy:", accuracy)

# Generate a classification report
report = classification_report(y_test, y_pred_binary)
print("Classification Report:\n", report)
```

```
Accuracy: 0.6066666666666667
Classification Report:
               precision    recall  f1-score   support

           0       0.61      0.74      0.67      2011
           1       0.60      0.45      0.52      1739

    accuracy                           0.61      3750
   macro avg       0.61      0.60      0.59      3750
weighted avg       0.61      0.61      0.60      3750
```

[242]:

```
conf_matrix = confusion_matrix(y_test, y_pred_binary)
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues")
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```


Confusion Matrix
Actual
0
1
1484
527
948
791
0
1
Predicted
1400
1200
1000
800
600