

```
In [1]: %matplotlib inline

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_style("whitegrid")
```

```
In [2]: import types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
client_ca664f4bd18b4ee7be1a254407d5a9cb = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='LnDsssRcQkVEuzSKFi8xiJ_iPDfljTBahv39piIi9d8c',
    ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.eu-geo.objectstorage.service.networklayer.com')

body = client_ca664f4bd18b4ee7be1a254407d5a9cb.get_object(Bucket='oneyearlifexpectancypostthoracic-donotdelete-pr-ydkg5pb794dfn6',Key='Dataset.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

Dataset = pd.read_csv(body)
Dataset.head()
```

Out[2]:

	Diagnosis	FVC	FEV1	Performance	Pain	Haemoptysis	Dyspnoea	Cough	Weakness	Tumour
0	2	2.88	2.16	1	0	0	0	1	1	
1	3	3.40	1.88	0	0	0	0	0	0	
2	3	2.76	2.08	1	0	0	0	1	0	
3	3	3.68	3.04	0	0	0	0	0	0	
4	3	2.44	0.96	2	0	1	0	1	1	

```
In [3]: Dataset.describe()
```

Out[3]:

	Diagnosis	FVC	FEV1	Performance	Pain	Haemoptysis	Dyspnoea	
count	454.000000	454.000000	454.00000	454.000000	454.000000	454.000000	454.000000	4
mean	3.092511	3.287952	2.51685	0.795154	0.059471	0.136564	0.055066	
std	0.715817	0.872347	0.77189	0.531459	0.236766	0.343765	0.228361	
min	1.000000	1.440000	0.96000	0.000000	0.000000	0.000000	0.000000	
25%	3.000000	2.600000	1.96000	0.000000	0.000000	0.000000	0.000000	
50%	3.000000	3.160000	2.36000	1.000000	0.000000	0.000000	0.000000	
75%	3.000000	3.840000	2.97750	1.000000	0.000000	0.000000	0.000000	
max	8.000000	6.300000	5.48000	2.000000	1.000000	1.000000	1.000000	

```
In [4]: # Stats for live and death after 1 yr patients
live = Dataset[Dataset['Death_1yr'] == 0]
death = Dataset[Dataset['Death_1yr'] == 1]

cond = ['FVC', 'FEV1', 'Performance', 'Pain', 'Haemoptysis', 'Dyspnoea', 'Cough', 'Weakness', \
        'Tumor_Size', 'Diabetes_Mellitus', 'MI_6mo', 'PAD', 'Smoking', 'Asthma', 'Age']

l = [np.mean(live[c]) for c in cond]
d = [np.mean(death[c]) for c in cond]

ld = pd.DataFrame(data={'Attribute': cond, 'Live 1yr Mean': l, 'Death 1yr Mean': d})
ld = ld.set_index('Attribute')

print('Death: {:d}, Live: {:d}'.format(len(death), len(live)))
print("1 year death: {:.2f}% out of 454 patients".format(np.mean(Dataset.Death_1yr)*100))
ld
```

Death: 69, Live: 385

1 year death: 15.20% out of 454 patients

Out[4]:

	Live 1yr Mean	Death 1yr Mean
Attribute		
FVC	3.304597	3.195072
FEV1	2.540805	2.383188
Performance	0.774026	0.913043
Pain	0.051948	0.101449
Haemoptysis	0.124675	0.202899
Dyspnoea	0.044156	0.115942
Cough	0.677922	0.797101
Weakness	0.158442	0.246377
Tumor_Size	1.683117	2.014493
Diabetes_Mellitus	0.062338	0.144928
MI_6mo	0.005195	0.000000
PAD	0.015584	0.028986
Smoking	0.815584	0.898551
Asthma	0.005195	0.000000
Age	62.677922	63.333333

```
In [5]: # Percentage difference in means of Live vs death patients
d = np.array(d)
l = np.array(l)

p_diff = (d-l)/l*100

fig, axes = plt.subplots(2,1,figsize=(12,18))

axes[0].bar(cond, p_diff)
axes[0].set_title('Mean Difference % between Dead and Live 1yr', fontsize=18)
axes[0].set_xticks(cond)
axes[0].set_xticklabels(cond, rotation=90)
axes[0].set_ylabel('Percent', fontsize=13)

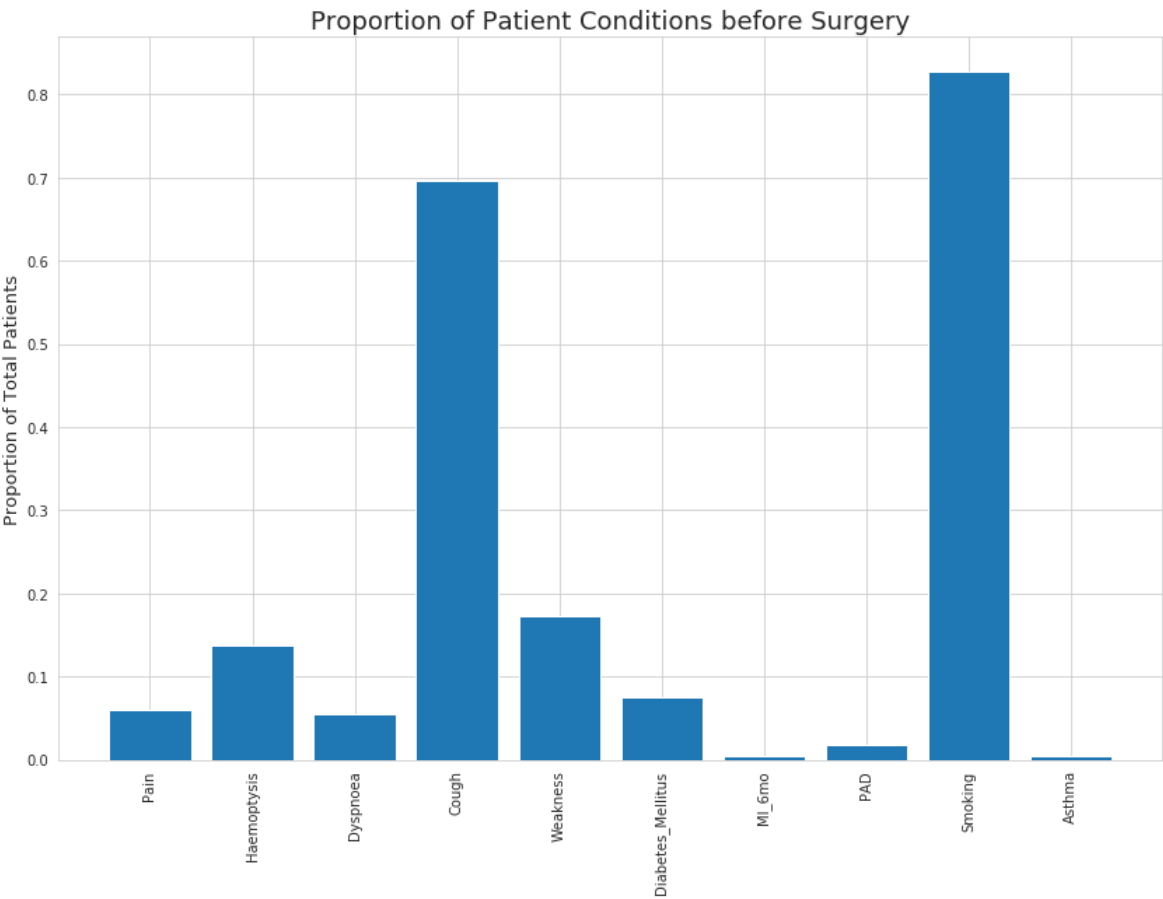
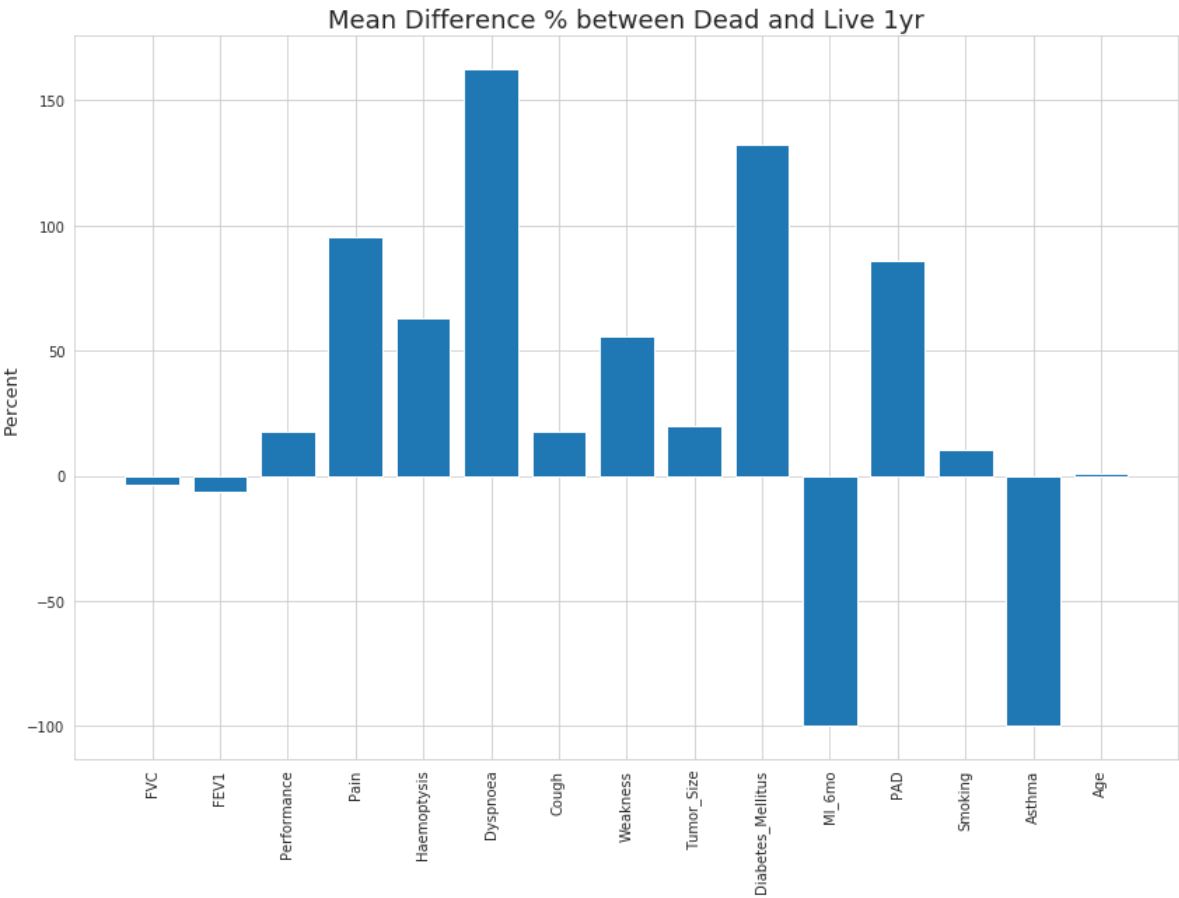
# Count plot of true/false condition columns

tf_col = ['Pain', 'Haemoptysis', 'Dyspnoea', 'Cough', 'Weakness', 'Diabetes_Me
lлитus', 'MI_6mo', 'PAD', 'Smoking', 'Asthma']
tf_sum = [Dataset[col].sum()/454 for col in tf_col]

axes[1].bar(tf_col, tf_sum)
axes[1].set_xticks(tf_col)
axes[1].set_xticklabels(tf_col, rotation=90)
axes[1].set_ylabel('Proportion of Total Patients', fontsize=13)
axes[1].set_title('Proportion of Patient Conditions before Surgery', fontsize=
18)

plt.tight_layout()

plt.show()
```

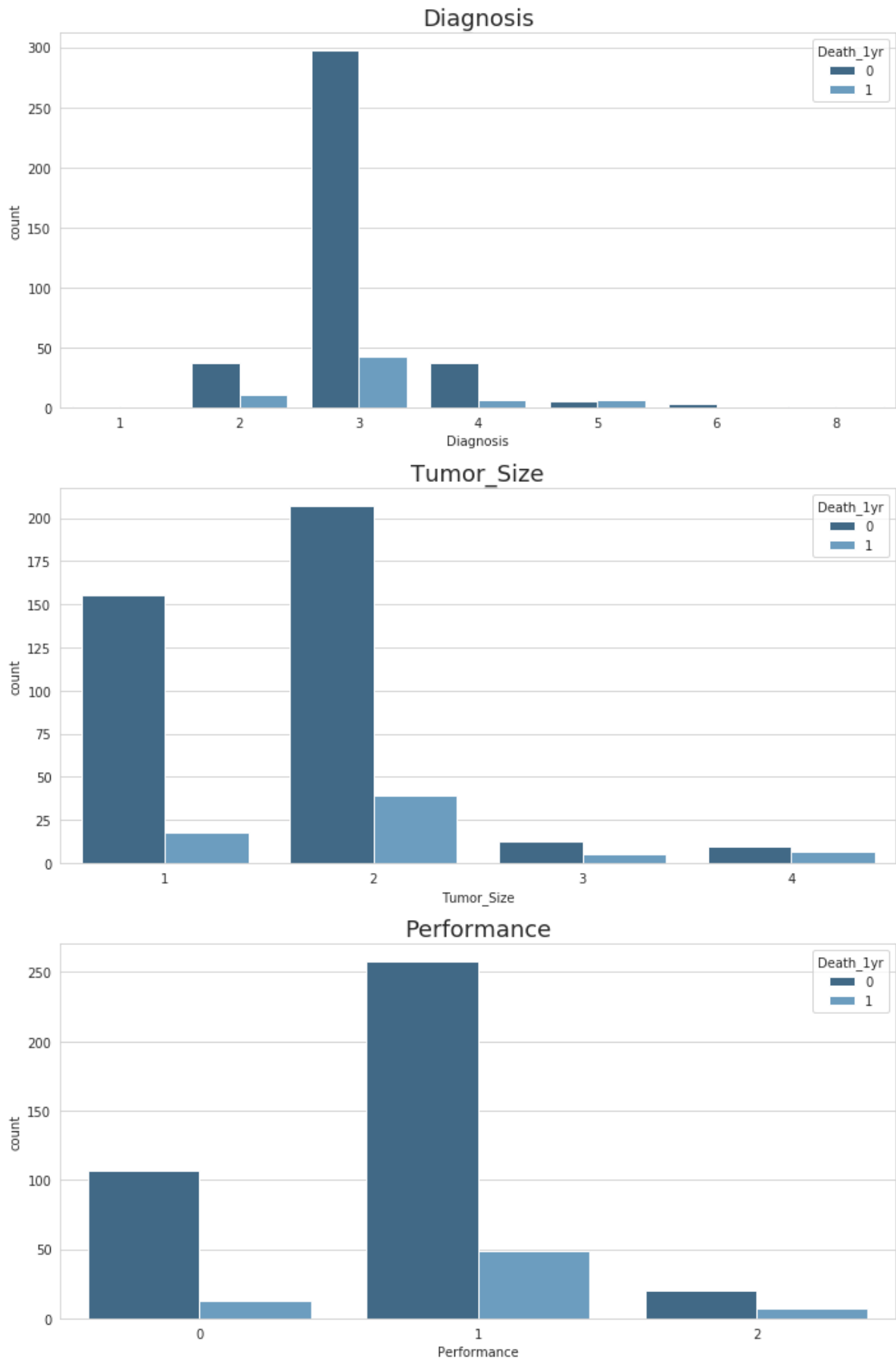


```
In [6]: # Count plots of Diagnosis, Tumor_Size, Performance with difference of Live and death data

fig, axes = plt.subplots(3,1,figsize=(10,15))

sns.countplot(x='Diagnosis', hue='Death_1yr', data=Dataset, palette='Blues_d',
ax=axes[0]).set_title('Diagnosis', fontsize=18)
sns.countplot(x='Tumor_Size', hue='Death_1yr', data=Dataset, palette='Blues_d',
, ax=axes[1]).set_title('Tumor_Size', fontsize=18)
sns.countplot(x='Performance', hue='Death_1yr', data=Dataset, palette='Blues_d',
ax=axes[2]).set_title('Performance', fontsize=18)

plt.tight_layout()
```



```
In [7]: def permutation_sample(data1, data2):
        """Generate a permutation sample from two data sets."""
        data = np.concatenate((data1, data2))
        permuted_data = np.random.permutation(data)

        perm_sample_1 = permuted_data[:len(data1)]
        perm_sample_2 = permuted_data[len(data1):]

        return perm_sample_1, perm_sample_2

def draw_perm_reps(data_1, data_2, func, size=1):
    """Generate multiple permutation replicates."""
    perm_replicates = np.empty(size)

    for i in range(size):
        perm_sample_1, perm_sample_2 = permutation_sample(data_1, data_2)
        perm_replicates[i] = func(perm_sample_1, perm_sample_2)

    return perm_replicates

def diff_of_means(data_1, data_2):
    """Difference in means of two arrays."""
    diff = np.mean(data_1) - np.mean(data_2)
    return diff
```

```
In [8]: # Hypothesis testing with Permutations of data
condition = ['FVC', 'FEV1', 'Performance', 'Pain', 'Haemoptysis', 'Dyspnoea',
             'Cough', 'Weakness',\
             'Tumor_Size', 'Diabetes_Mellitus', 'MI_6mo', 'PAD', 'Smoking', 'Asthma', 'Age']
p_val = []

for c in condition:
    empirical_diff_means = diff_of_means(death[c], live[c])
    perm_replicates = draw_perm_reps(death[c], live[c], diff_of_means, size=1000)
    if empirical_diff_means > 0:
        p = np.sum(perm_replicates >= empirical_diff_means) / len(perm_replicates)
        p_val.append(p)
    else:
        p = np.sum(perm_replicates <= empirical_diff_means) / len(perm_replicates)
        p_val.append(p)

print(list(zip(condition, p_val)))

[('FVC', 0.1666), ('FEV1', 0.0567), ('Performance', 0.03), ('Pain', 0.0994),
 ('Haemoptysis', 0.0617), ('Dyspnoea', 0.023), ('Cough', 0.0259), ('Weakness',
 0.0601), ('Tumor_Size', 0.0004), ('Diabetes_Mellitus', 0.0208), ('MI_6mo', 0.
 7135), ('PAD', 0.3571), ('Smoking', 0.0607), ('Asthma', 0.7213), ('Age', 0.27
 25)]
```



```

In [10]: # Scatter plot for FVC, FEV1, Age columns

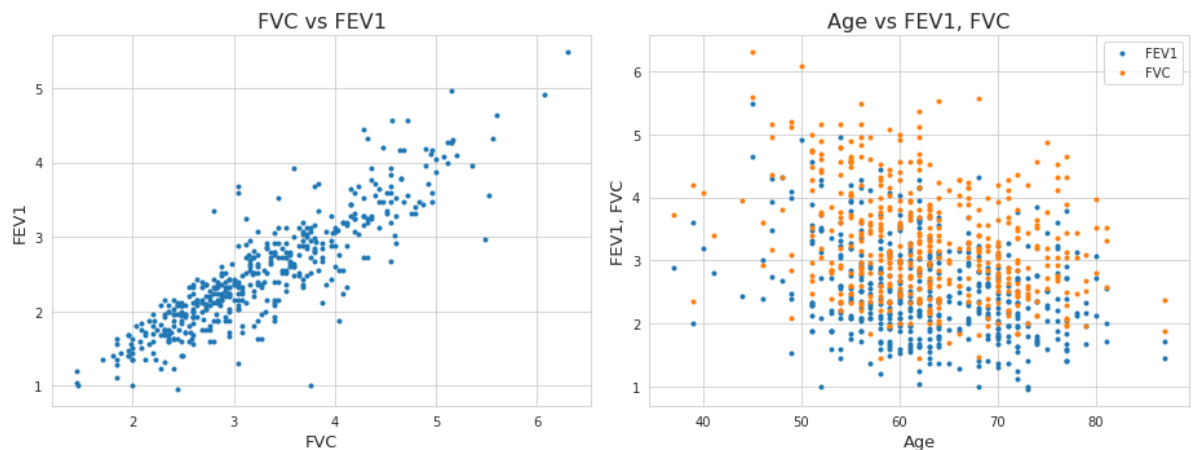
fig, axes = plt.subplots(1,2,figsize=(13,5))
axes[0].plot(Dataset.FVC, Dataset.FEV1, linestyle='none', marker='.')

axes[0].set_xlabel('FVC', fontsize=13)
axes[0].set_ylabel('FEV1', fontsize=13)
axes[0].set_title('FVC vs FEV1', fontsize=16)

axes[1].plot(Dataset.Age, Dataset.FEV1, linestyle='none', marker='.', label='F
EV1')
axes[1].plot(Dataset.Age, Dataset.FVC, linestyle='none', marker='.', label='FV
C')
axes[1].set_xlabel('Age', fontsize=13)
axes[1].set_ylabel('FEV1, FVC', fontsize=13)
axes[1].legend()
axes[1].set_title('Age vs FEV1, FVC', fontsize=16)

plt.tight_layout()

```



```

In [13]: # Correlation coefficients for FVC and FEV1
np.corrcoef(Dataset.FVC, Dataset.FEV1)[0,1]

```

Out[13]: 0.8875452733829

```

In [14]: # Correlation coefficients for Age and FVC
np.corrcoef(Dataset.Age, Dataset.FVC)[0,1]

```

Out[14]: -0.299429919660491

```

In [15]: # Correlation coefficients for Age and FEV1
np.corrcoef(Dataset.Age, Dataset.FEV1)[0,1]

```

Out[15]: -0.3096166273079891

```
In [16]: def ecdf(data):
        """Compute ECDF for a one-dimensional array of measurements."""
        n = len(data)
        x = np.sort(data)
        y = np.arange(1, n+1) / n
        return x, y
```

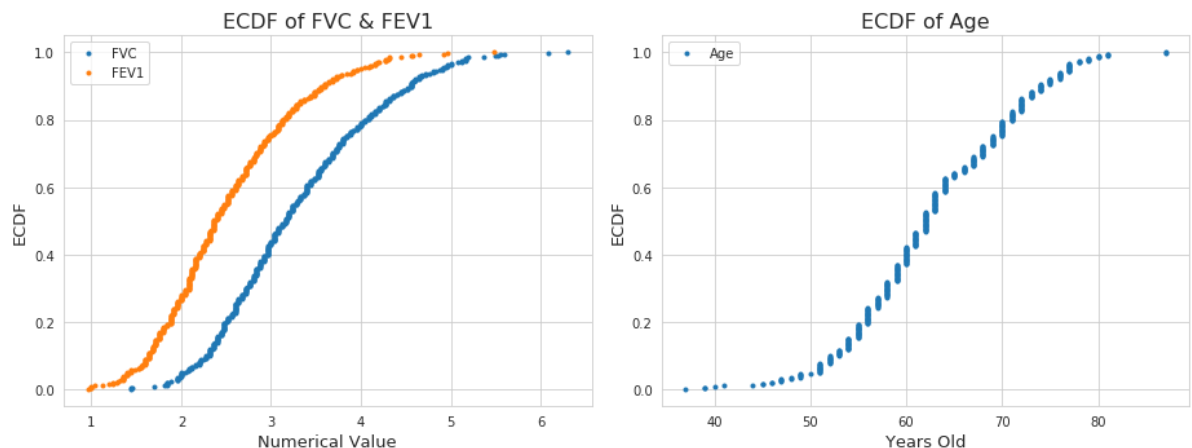
```
In [17]: # ECDF of FVC, FEV1, Age
x_fvc, y_fvc = ecdf(Dataset.FVC)
x_fev1, y_fev1 = ecdf(Dataset.FEV1)
x_age, y_age = ecdf(Dataset.Age)

fig, axes = plt.subplots(1,2,figsize=(13,5))
axes[0].plot(x_fvc, y_fvc, marker='.', linestyle='none', label='FVC')
axes[0].plot(x_fev1, y_fev1, marker='.', linestyle='none', label='FEV1')

axes[0].set_xlabel('Numerical Value', fontsize=13)
axes[0].set_ylabel('ECDF', fontsize=13)
axes[0].legend(loc='upper left')
axes[0].set_title('ECDF of FVC & FEV1', fontsize=16)

axes[1].plot(x_age, y_age, marker='.', linestyle='none', label='Age')
axes[1].set_xlabel('Years Old', fontsize=13)
axes[1].set_ylabel('ECDF', fontsize=13)
axes[1].legend(loc='upper left')
axes[1].set_title('ECDF of Age', fontsize=16)

plt.tight_layout()
```



```
In [18]: from sklearn.linear_model import LogisticRegression
        from sklearn.model_selection import cross_val_score, cross_val_predict, train_
        test_split, GridSearchCV
        from sklearn.metrics import confusion_matrix, classification_report, accuracy_
        score, average_precision_score
        from sklearn.ensemble import RandomForestClassifier
        import itertools
```

```
In [20]: # All attributes excluding target variable, Asthma, and MI_6mo
X = Dataset.drop(['Death_1yr', 'MI_6mo', 'Asthma'], axis=1)

# Attributes of Significance from Hypothesis Testing
X2 = Dataset[['Performance', 'Dyspnoea', 'Cough', 'Tumor_Size', 'Diabetes_Mellitus']]
y = Dataset['Death_1yr']
```

```
In [21]: def plot_confusion_matrix(cm, classes, normalize=False, title='Confusion matrix',
x', cmap=plt.cm.Blues):
    """
    This function prints and plots the confusion matrix.
    Normalization can be applied by setting `normalize=True`.
    """
    if normalize:
        cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
        print("Normalized confusion matrix")
    else:
        print('Confusion matrix')

    print(cm)

    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
    plt.yticks(tick_marks, classes)

    fmt = '.2f' if normalize else 'd'
    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
        plt.text(j, i, format(cm[i, j], fmt),
                 horizontalalignment="center",
                 color="white" if cm[i, j] > thresh else "black")

    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')
```

```
In [22]: def model_report(model, X, y, title, weight=None):  
        """Takes in classifier model with X data and class weight to display score  
        s and confusion matrix."""  
  
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.3, ra  
ndom_state=1111, stratify=y)  
  
        clf = model(class_weight=weight, random_state=1111)  
  
        clf.fit(X_train, y_train)  
        y_pred = clf.predict(X_test)  
  
        class_names = ['Live', 'Death']  
  
        print('Accuracy: {:.2f}'.format(accuracy_score(y_test, y_pred)))  
        print('Average Precision: {:.2f}'.format(average_precision_score(y_test, y  
_pred)))  
        print(classification_report(y_test, y_pred, target_names=class_names))  
  
        cnf_matrix = confusion_matrix(y_test, y_pred)  
        np.set_printoptions(precision=2)  
  
        # Plot confusion matrix  
        plt.figure()  
        plot_confusion_matrix(cnf_matrix, classes=class_names, normalize=True, tit  
le=title)  
        plt.show()
```

```
In [23]: def class_weights_plot(model, title):
    """Plots accuracy, average precision, and correct death predictions of different class weights for X2 data."""
    class_weights = [1,2,3,4,5,6,7,8,9,10]
    acc_scores = []
    prec_scores = []
    correct_deaths = []

    for cw in class_weights:
        clf = model(class_weight={0: 1, 1: cw}, random_state=1111)
        predicted = cross_val_predict(clf, X2, y, cv=5)

        acc_scores.append(accuracy_score(y, predicted))
        prec_scores.append(average_precision_score(y, predicted))
        correct_deaths.append(confusion_matrix(y, predicted)[1,1]/68)

    plt.figure(figsize=(8,6))
    plt.plot(class_weights, acc_scores, marker='.', label='Accuracy Score')
    plt.plot(class_weights, prec_scores, marker='.', label='Average Precision')
    plt.plot(class_weights, correct_deaths, marker='.', label='Correct Death Predictions')
    plt.xticks(class_weights)
    plt.xlabel('Class Weights', fontsize=13)
    plt.ylabel('Scores', fontsize=13)
    plt.legend()
    plt.title(title, fontsize=16)

    plt.show()
```

```
In [24]: def class_weights_tf_plot(model, title):  
        """Plots confusion matrix values of different class weights for X2 dat  
a."""  
        class_weights = [1,2,3,4,5,6,7,8,9,10]  
        true_live = []  
        false_live = []  
        true_death = []  
        false_death = []  
  
        for cw in class_weights:  
            clf = model(class_weight={0: 1, 1: cw}, random_state=1111)  
            predicted = cross_val_predict(clf, X2, y, cv=5)  
  
            true_live.append(confusion_matrix(y, predicted)[0,0]/385)  
            false_live.append(confusion_matrix(y, predicted)[1,0]/68)  
            true_death.append(confusion_matrix(y, predicted)[1,1]/68)  
            false_death.append(confusion_matrix(y, predicted)[0,1]/385)  
  
            plt.figure(figsize=(8,6))  
            plt.plot(class_weights, true_live, marker='.', label='Correct Live Predict  
ions')  
            plt.plot(class_weights, false_live, marker='.', label='False Live Predicti  
ons')  
            plt.plot(class_weights, true_death, marker='.', label='Correct Death Predi  
ctions')  
            plt.plot(class_weights, false_death, marker='.', label='False Death Predic  
tions')  
            plt.xticks(class_weights)  
            plt.xlabel('Class Weights', fontsize=13)  
            plt.ylabel('Scores', fontsize=13)  
            plt.legend()  
            plt.title(title, fontsize=16)  
  
            plt.show()
```

```
In [25]: # Log Reg on X with no class weights
model_report(LogisticRegression, X, y, 'LogReg: X')
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1143: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
```

```
'precision', 'predicted', average, warn_for)
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1143: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
```

```
'precision', 'predicted', average, warn_for)
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1143: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
```

```
'precision', 'predicted', average, warn_for)
```

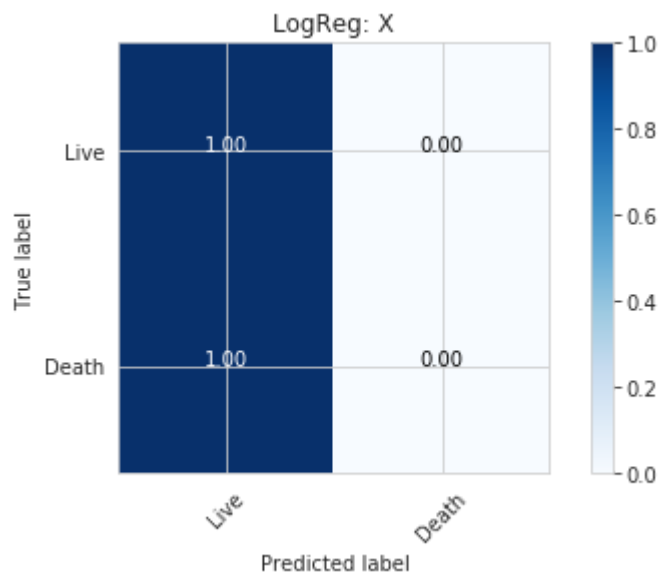
```
Accuracy: 0.85
```

```
Average Precision: 0.15
```

	precision	recall	f1-score	support
Live	0.85	1.00	0.92	116
Death	0.00	0.00	0.00	21
micro avg	0.85	0.85	0.85	137
macro avg	0.42	0.50	0.46	137
weighted avg	0.72	0.85	0.78	137

```
Normalized confusion matrix
```

```
[[1. 0.]
 [1. 0.]]
```



```
In [26]: # Log Reg on X with class weight balanced since imbalanced death numbers (15%)
model_report(LogisticRegression, X, y, 'LogReg: X with class_weight', 'balanced'
)
```

/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)

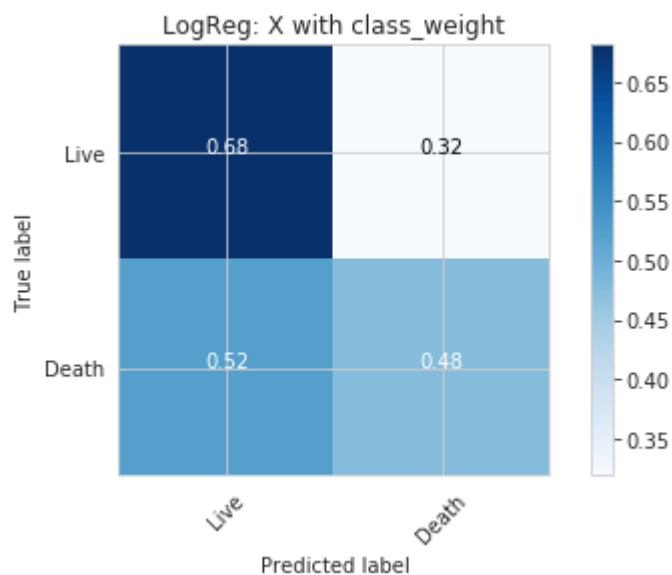
Accuracy: 0.65

Average Precision: 0.18

	precision	recall	f1-score	support
Live	0.88	0.68	0.77	116
Death	0.21	0.48	0.29	21
micro avg	0.65	0.65	0.65	137
macro avg	0.55	0.58	0.53	137
weighted avg	0.78	0.65	0.69	137

Normalized confusion matrix

```
[[0.68 0.32]
 [0.52 0.48]]
```




```
In [27]: # X2 Log reg with no class weights
model_report(LogisticRegression, X2, y, 'LogReg: X2')
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
```

```
FutureWarning)
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1143: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
```

```
'precision', 'predicted', average, warn_for)
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1143: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
```

```
'precision', 'predicted', average, warn_for)
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1143: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
```

```
'precision', 'predicted', average, warn_for)
```

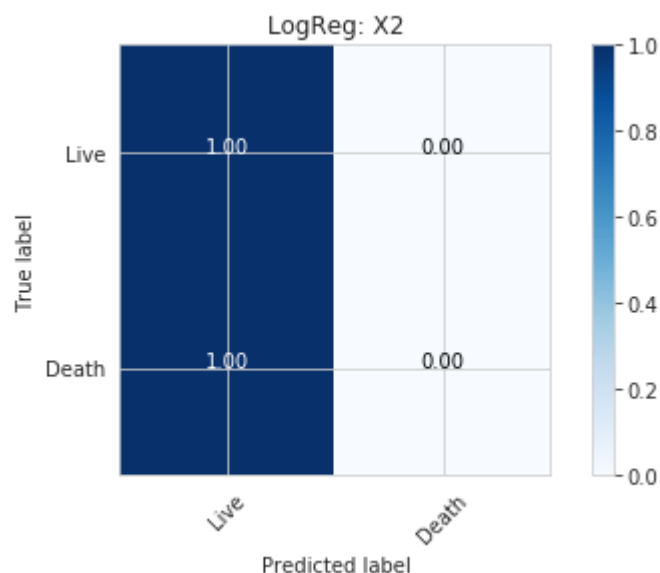
```
Accuracy: 0.85
```

```
Average Precision: 0.15
```

	precision	recall	f1-score	support
Live	0.85	1.00	0.92	116
Death	0.00	0.00	0.00	21
micro avg	0.85	0.85	0.85	137
macro avg	0.42	0.50	0.46	137
weighted avg	0.72	0.85	0.78	137

```
Normalized confusion matrix
```

```
[[1. 0.]
 [1. 0.]]
```



```
In [28]: # X2 Log reg with class weight balanced
model_report(LogisticRegression, X2, y, 'LogReg: X2 with class_weight', 'balanced')
```

/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)

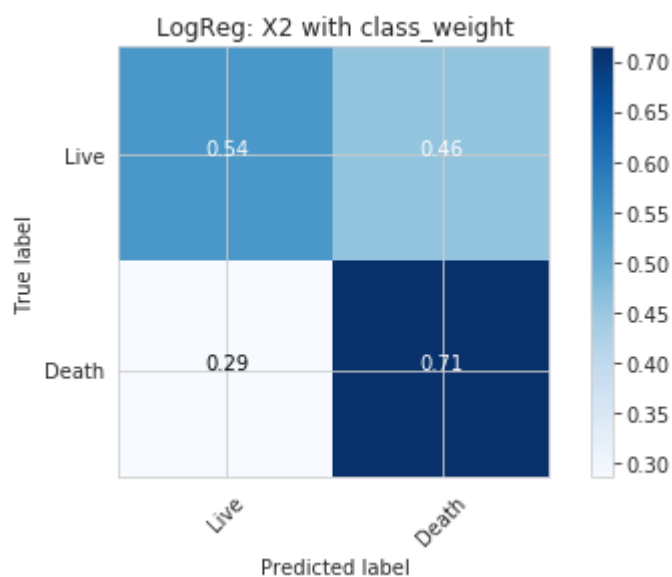
Accuracy: 0.57

Average Precision: 0.20

	precision	recall	f1-score	support
Live	0.91	0.54	0.68	116
Death	0.22	0.71	0.34	21
micro avg	0.57	0.57	0.57	137
macro avg	0.57	0.63	0.51	137
weighted avg	0.81	0.57	0.63	137

Normalized confusion matrix

```
[[0.54 0.46]
 [0.29 0.71]]
```



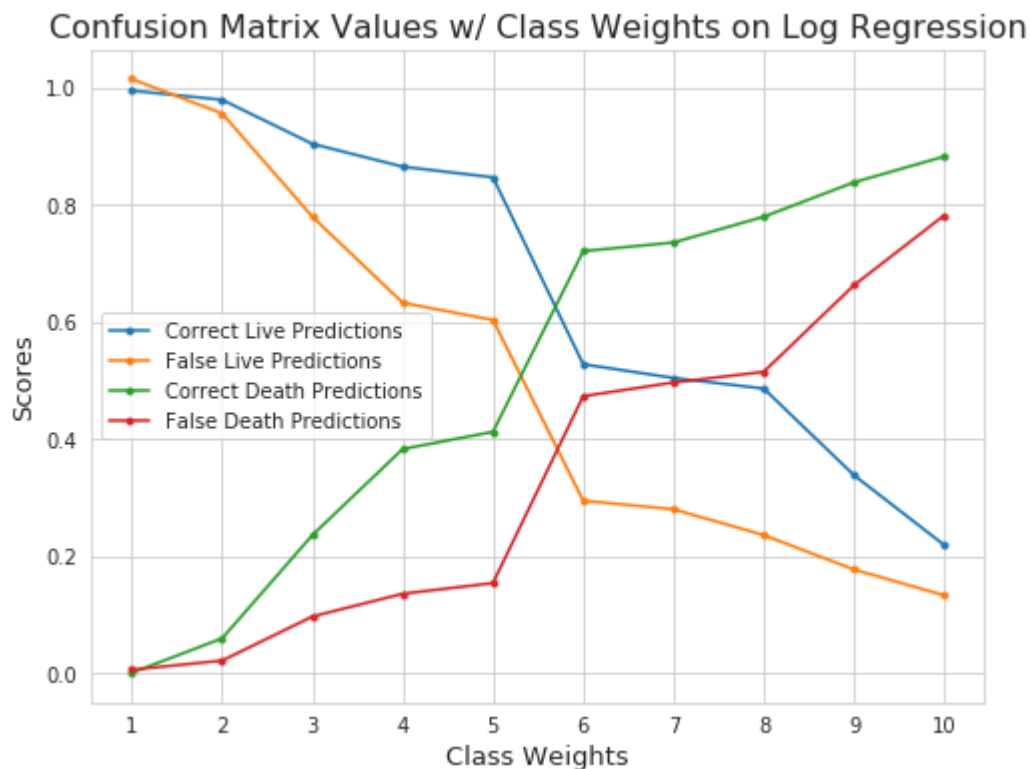
```
In [29]: # Plot different class weights influence on Log Reg X2  
class_weights_tf_plot(LogisticRegression, 'Confusion Matrix Values w/ Class We  
ights on Log Regression')
```



```

FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)

```



```
In [30]: # Plot different class weights influence on LogReg X2  
class_weights_plot(LogisticRegression, 'Class Weights Influence on Log Regression of X2')
```

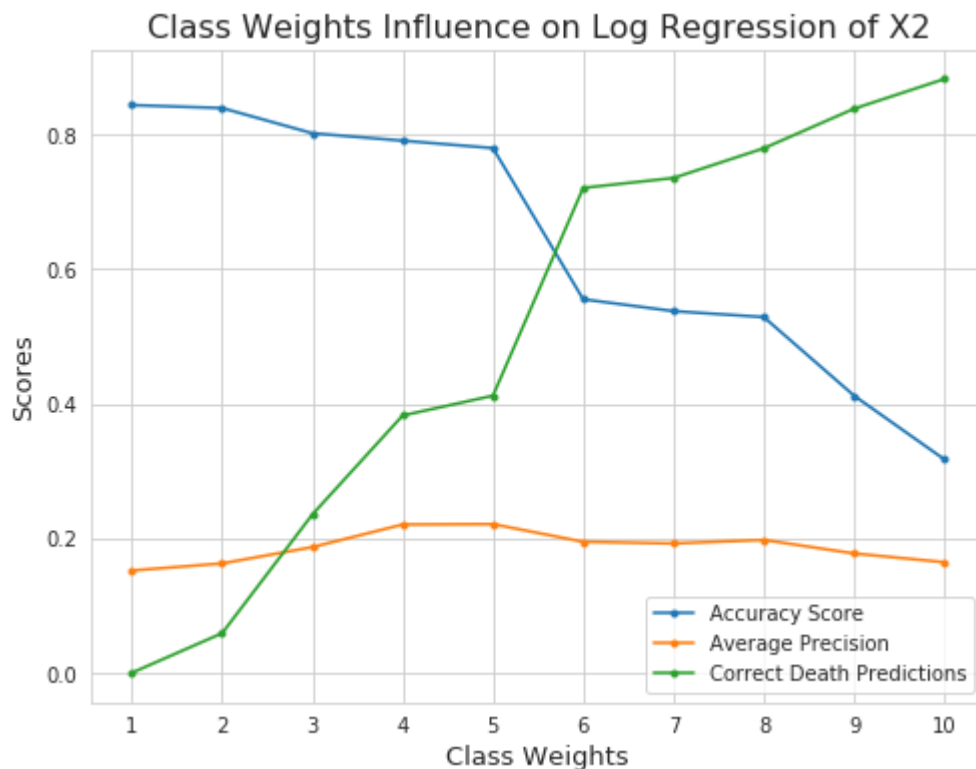

[illegible]

[illegible]


```

FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/linear_model/log
istic.py:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
FutureWarning)

```



```
In [ ]: #Random forest classifier
```

```
In [31]: # X2 Random Forest Classifier with equal class weights 1:1
model_report(RandomForestClassifier, X2, y, 'RF: X2')
```

/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)

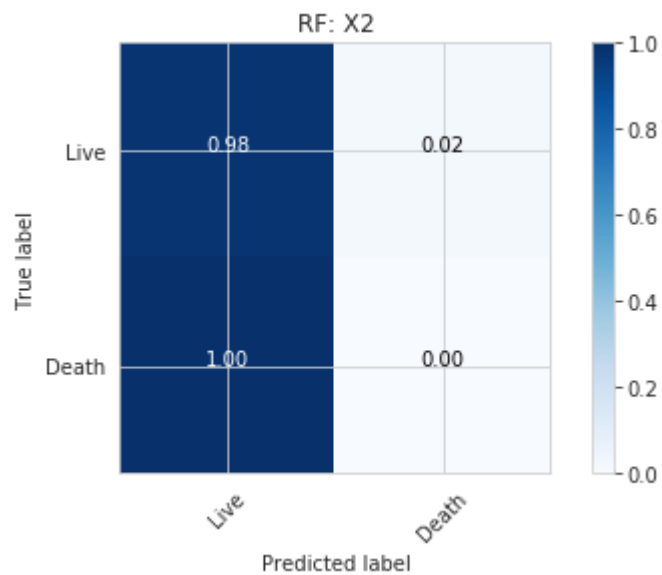
Accuracy: 0.83

Average Precision: 0.15

	precision	recall	f1-score	support
Live	0.84	0.98	0.91	116
Death	0.00	0.00	0.00	21
micro avg	0.83	0.83	0.83	137
macro avg	0.42	0.49	0.45	137
weighted avg	0.72	0.83	0.77	137

Normalized confusion matrix

```
[[0.98 0.02]
 [1.   0.  ]]
```



```
In [32]: # X2 Random Forest Classifier with class weight 5.67
model_report(RandomForestClassifier, X2, y, 'RF: X2 with class_weight', 'balanced')
```

/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)

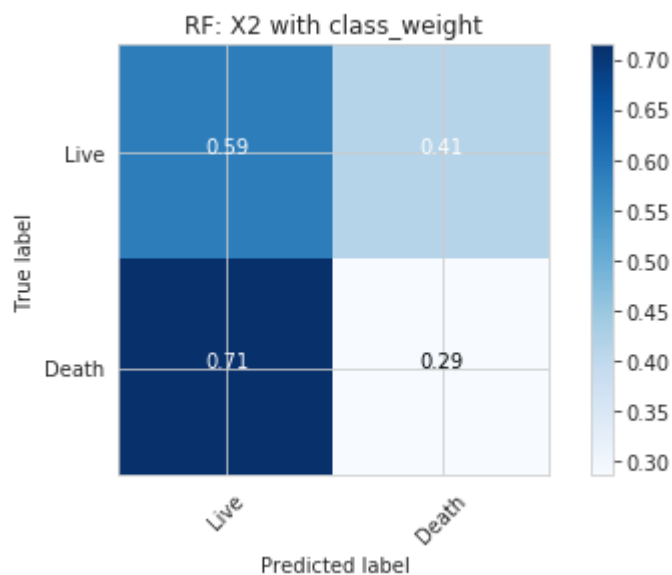
Accuracy: 0.54

Average Precision: 0.14

	precision	recall	f1-score	support
Live	0.82	0.59	0.68	116
Death	0.11	0.29	0.16	21
micro avg	0.54	0.54	0.54	137
macro avg	0.47	0.44	0.42	137
weighted avg	0.71	0.54	0.60	137

Normalized confusion matrix

```
[[0.59 0.41]
 [0.71 0.29]]
```



```
In [33]: # Plot different class weights influence on RF classifier  
class_weights_tf_plot(RandomForestClassifier, 'Confusion Matrix Values w/ Class Weights on Random Forest')
```

[illegible]

in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

```

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)

```



```
In [34]: # Plot different class weights influence on RF classifier  
class_weights_plot(RandomForestClassifier, 'Class Weights Influence on Random  
Forest of X2')
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)
```

```
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)  
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.  
py:246: FutureWarning: The default value of n_estimators will change from 10  
in version 0.20 to 100 in 0.22.  
    "10 in version 0.20 to 100 in 0.22.", FutureWarning)
```

in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

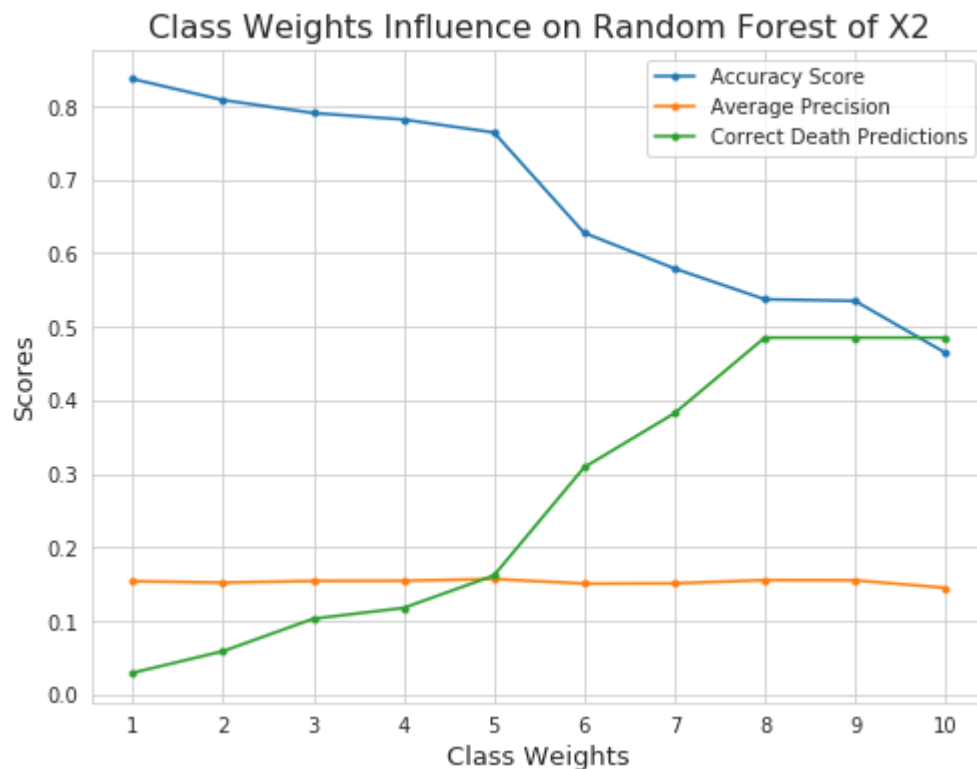
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.py:246: FutureWarning: The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

```

"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/ensemble/forest.
py:246: FutureWarning: The default value of n_estimators will change from 10
in version 0.20 to 100 in 0.22.
"10 in version 0.20 to 100 in 0.22.", FutureWarning)

```




```
In [35]: # Grid Search Hyperparameters for Random Forest
rfc = RandomForestClassifier(random_state=1111)

param_grid = {
    "n_estimators" : [5, 6, 7, 8, 9],
    "max_depth" : [7, 8, 9, 10],
    "min_samples_leaf" : [7, 8, 9, 10],
    "class_weight": [None, 'balanced', {0: 1, 1: 5}]}

CV_rfc = GridSearchCV(estimator=rfc, param_grid=param_grid, cv= 5, scoring='average_precision')
CV_rfc.fit(X2, y)
print(CV_rfc.best_params_)
print('Best average precision score: {:.4f}'.format(CV_rfc.best_score_))

{'class_weight': None, 'max_depth': 7, 'min_samples_leaf': 8, 'n_estimators': 5}
Best average precision score: 0.2609
```

In []: