```
C:\Users\sanjeet\Documents\assignment1-big-data\heart disease>python hdp.py
Library Loaded
# rows in dataset 303
--------------------------------------------
# rows in age with ZERO value: 0
# rows in sex with ZERO value: 97
# rows in cp with ZERO value: 0
# rows in trestbps with ZERO value: 0
# rows in chol with ZERO value: 0
# rows in fbs with ZERO value: 258
# rows in restecg with ZERO value: 151
# rows in thalach with ZERO value: 0
# rows in exang with ZERO value: 204
# rows in oldpeak with ZERO value: 99
# rows in slope with ZERO value: 0
# rows in ca with ZERO value: 4
# rows in thal with ZERO value: 2
# rows in target with ZERO value: 164
Building bar plot (works only on jupyter notebook) Press Ctrl + C if it stucks for more than 2-3 minutes.

Removing the columns sex, target, age to ease the computation

Using TSNE

t-SNE done! Time elapsed: 2.13904070854187 seconds
Doing Feature Engineering

# rows in dataset 303
--------------------------------------------
# rows in cp with ZERO value: 0
# rows in trestbps with ZERO value: 0
# rows in chol with ZERO value: 0
# rows in fbs with ZERO value: 258
# rows in restecg with ZERO value: 151
# rows in thalach with ZERO value: 0
# rows in exang with ZERO value: 204
# rows in oldpeak with ZERO value: 99
# rows in slope with ZERO value: 0
# rows in ca with ZERO value: 4
# rows in thal with ZERO value: 2
Training Set : 272
Test Set : 31
Training labels : 272
Test Labels : 31
```

Logistic Regression Result:

```
Running LogisticRegression
Fitting 5 folds for each of 10 candidates, totalling 50 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent workers.
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:947:
  "of iterations.", ConvergenceWarning)
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:947:
  "of iterations.", ConvergenceWarning)
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:947:
  "of iterations.", ConvergenceWarning)
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:947:
  "of iterations.", ConvergenceWarning)
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:947:
  "of iterations.", ConvergenceWarning)
[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:    6.5s
[Parallel(n_jobs=-1)]: Done  50 out of  50 | elapsed:    6.8s finished
[0 1 0 1 0 1 0 0 0 1 0 1 1 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 1 1 1]
Best Params : {'C': 2.7825594022071245, 'penalty': 'l2'}
Classification Report :
              precision    recall  f1-score   support

           0       0.78      0.88      0.82        16
           1       0.85      0.73      0.79        15

    accuracy                           0.81        31
   macro avg       0.81      0.80      0.80        31
weighted avg       0.81      0.81      0.81        31

Accuracy Score : 0.8064516129032258
Confusion Matrix :
 [[14  2]
 [ 4 11]]
Running XGBoost
```

XGBoost Result

```
Running XGBoost
Fitting 5 folds for each of 270 candidates, totalling 1350 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent workers.
[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:    6.3s
[Parallel(n_jobs=-1)]: Done 192 tasks      | elapsed:   39.4s
[Parallel(n_jobs=-1)]: Done 442 tasks      | elapsed:  1.6min
[Parallel(n_jobs=-1)]: Done 792 tasks      | elapsed:  2.8min
[Parallel(n_jobs=-1)]: Done 1242 tasks      | elapsed:  4.3min
[Parallel(n_jobs=-1)]: Done 1350 out of 1350 | elapsed:  4.6min finished
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\model_selection\_search.py:814: DeprecationWarr
ric results when test-set sizes are unequal.
  DeprecationWarning)
[0 1 0 1 1 1 0 1 0 1 0 1 1 0 1 1 0 0 1 1 1 1 1 1 1 0 1 1 1 1 1]
Best Params : {'learning_rate': 0.06999999999999999, 'max_depth': 2, 'n_estimators': 100}
Classification Report :
              precision    recall  f1-score   support

           0       0.67      0.38      0.48        16
           1       0.55      0.80      0.65        15

    accuracy                           0.58        31
   macro avg       0.61      0.59      0.56        31
weighted avg       0.61      0.58      0.56        31

Accuracy Score : 0.5806451612903226
Confusion Matrix :
 [[ 6 10]
 [ 3 12]]
```

Random Forest Result:

```
Running Random Forest
Fitting 5 folds for each of 30 candidates, totalling 150 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent workers.
[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:   18.8s
[Parallel(n_jobs=-1)]: Done 150 out of 150 | elapsed:  1.4min finished
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\model_selection\_search.py:814
ric results when test-set sizes are unequal.
  DeprecationWarning)
[0 1 0 1 1 1 0 0 0 1 0 1 1 0 1 1 1 0 1 0 1 0 1 1 0 0 1 0 1 1 1]
Best Params : {'max_depth': 3, 'n_estimators': 100}
Classification Report :
              precision    recall  f1-score   support

           0       0.85      0.69      0.76        16
           1       0.72      0.87      0.79        15

    accuracy                           0.77        31
   macro avg       0.78      0.78      0.77        31
weighted avg       0.79      0.77      0.77        31

Accuracy Score : 0.7741935483870968
Confusion Matrix :
 [[11  5]
 [ 2 13]]
Running SVC
```

SVC Results:

```
Running SVC
Fitting 5 folds for each of 28 candidates, totalling 140 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent workers.
[Parallel(n_jobs=-1)]: Done 140 out of 140 | elapsed:    0.7s finished
C:\Users\sanjeet\Anaconda3\lib\site-packages\sklearn\model_selection\_search.py:814:
ric results when test-set sizes are unequal.
  DeprecationWarning)
[0 1 0 1 0 1 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 1 1 0 1 1 1 1 1]
Best Params : {'C': 100, 'gamma': 0.0001}
Classification Report :
              precision    recall  f1-score   support

           0       0.69      0.69      0.69        16
           1       0.67      0.67      0.67        15

    accuracy                           0.68        31
   macro avg       0.68      0.68      0.68        31
weighted avg       0.68      0.68      0.68        31

Accuracy Score : 0.6774193548387096
Confusion Matrix :
 [[11  5]
 [ 5 10]]
```