In [131]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

a) load/merge data and visualize logerror

In [132]:

```
# load data into DataFrames
df1=pd.read_csv("train.csv")
df2=pd.read_csv("properties.csv")
df_final = df1.merge(df2, on = 'id', how = 'left')
```

In [133]:

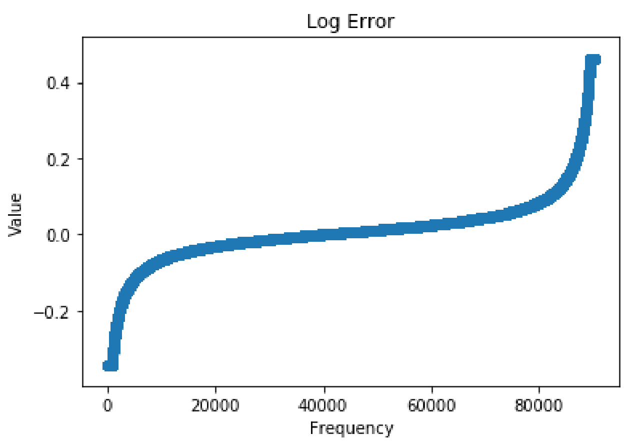
```
# eliminate outliers
LE1=np.percentile(df_final.iloc[:,1],1,interpolation='nearest')
LE99=np.percentile(df_final.iloc[:,1],99,interpolation='nearest')
logerr=np.array(df_final.iloc[:,1])
logerr[np.where(logerr>=LE99)]=LE99
logerr[np.where(logerr<=LE1)]=LE1

df_final.iloc[:,1]=logerr
```

In [134]:

```
# scatter of logerr
a = df_final.iloc[:,1]
plt.scatter(np.arange(len(df_final)), np.sort(a))
plt.xlabel('Frequency')
plt.ylabel('Value')
plt.title('Log Error')
```

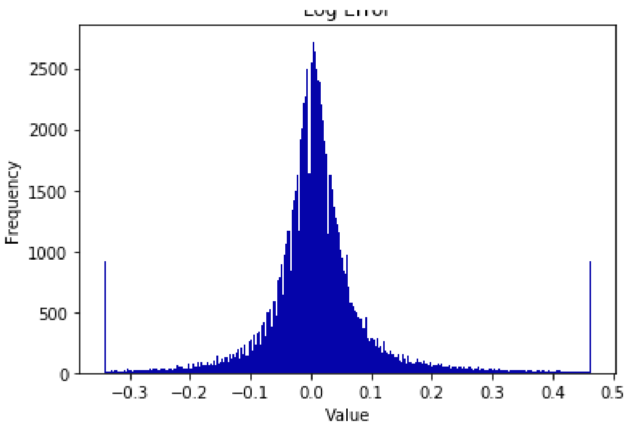
Out[134]:

```
Text(0.5, 1.0, 'Log Error')
```



In [135]:

```
# histogram of logerr
n, bins, patches = plt.hist(df_final.iloc[:,1], bins='auto', color='#0504aa')
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.title('Log Error')
maxfreq = n.max()
```



b) data cleaning

In [136]:

```
# build new data frame
df_1=np.array(df_final.columns)
df_2=np.array(df_final.isnull().sum(axis = 0))
df_11=np.column_stack((df_1,df_2))
df_12=pd.DataFrame(df_11)
df_12.columns = ["column_name", "missing_count"]
df_12['missing_ratio'] = df_12.iloc[:,1]/len(df_final)
print(df_12)
```

```
                     column_name missing_count missing_ratio
0                             id             0             0
1                       logerror             0             0
2                transactiondate             0             0
3          airconditioningtypeid         80113      0.887433
4       architecturalstyletypeid         90178      0.998926
5                   basementsqft         90261      0.999845
6                    bathroomcnt         58550      0.648574
7                     bedroomcnt         58550      0.648574
8            buildingclasstypeid         90267      0.999911
9          buildingqualitytypeid         70038      0.775829
10             calculatedbathnbr         58964       0.65316
11                    decktypeid         90052       0.99753
12      finishedfloor1squarefeet         87931      0.974035
13  calculatedfinishedsquarefeet         58778      0.651099
14          finishedsquarefeet12         60197      0.666818
15          finishedsquarefeet13         90261      0.999845
16          finishedsquarefeet15         89004      0.985921
17          finishedsquarefeet50         87931      0.974035
18           finishedsquarefeet6         90141      0.998516
19                          fips         58550      0.648574
20                  fireplacecnt         86924       0.96288
21                   fullbathcnt         58964       0.65316
22                  garagecarcnt         79830      0.884298
23               garagetotalsqft         79830      0.884298
24                hashottuborspa         89479      0.991182
25         heatingorsystemtypeid         70512       0.78108
26                      latitude         58550      0.648574
27                     longitude         58550      0.648574
28             lotsizesquarefeet         62072      0.687588
29                       poolcnt         84004      0.930534
30                   poolsizesum         89944      0.996333
31                 pooltypeid10         89887      0.995702
32                  pooltypeid2         89867       0.99548
33                  pooltypeid7         84412      0.935054
34     propertycountylandusecode         58550      0.648574
35         propertylandusetypeid         58550      0.648574
36         propertyzoningdesc         69685      0.771919
37         rawcensustractandblock         58550      0.648574
38                  regionidcity         59216      0.655951
39                regionidcounty         58550      0.648574
40          regionidneighborhood         77632       0.85995
41                   regionidzip         58562      0.648707
42                       roomcnt         58550      0.648574
```

43	storytypeid	90261	0.999845
44	threequarterbathnbr	86021	0.952877
45	typeconstructiontypeid	90163	0.998759
46	unitcnt	69677	0.771831
47	yardbuildingsqft17	89364	0.989909
48	yardbuildingsqft26	90241	0.999623
49	yearbuilt	58810	0.651454
50	numberofstories	83076	0.920255
51	fireplaceflag	90181	0.998959
52	structuretaxvaluedollarcnt	58678	0.649992
53	taxvaluedollarcnt	58551	0.648585
54	assessmentyear	58550	0.648574
55	landtaxvaluedollarcnt	58551	0.648585
56	taxamount	58551	0.648585
57	taxdelinquencyflag	89662	0.99321
58	taxdelinquencyyear	89662	0.99321
59	censustractandblock	58758	0.650878

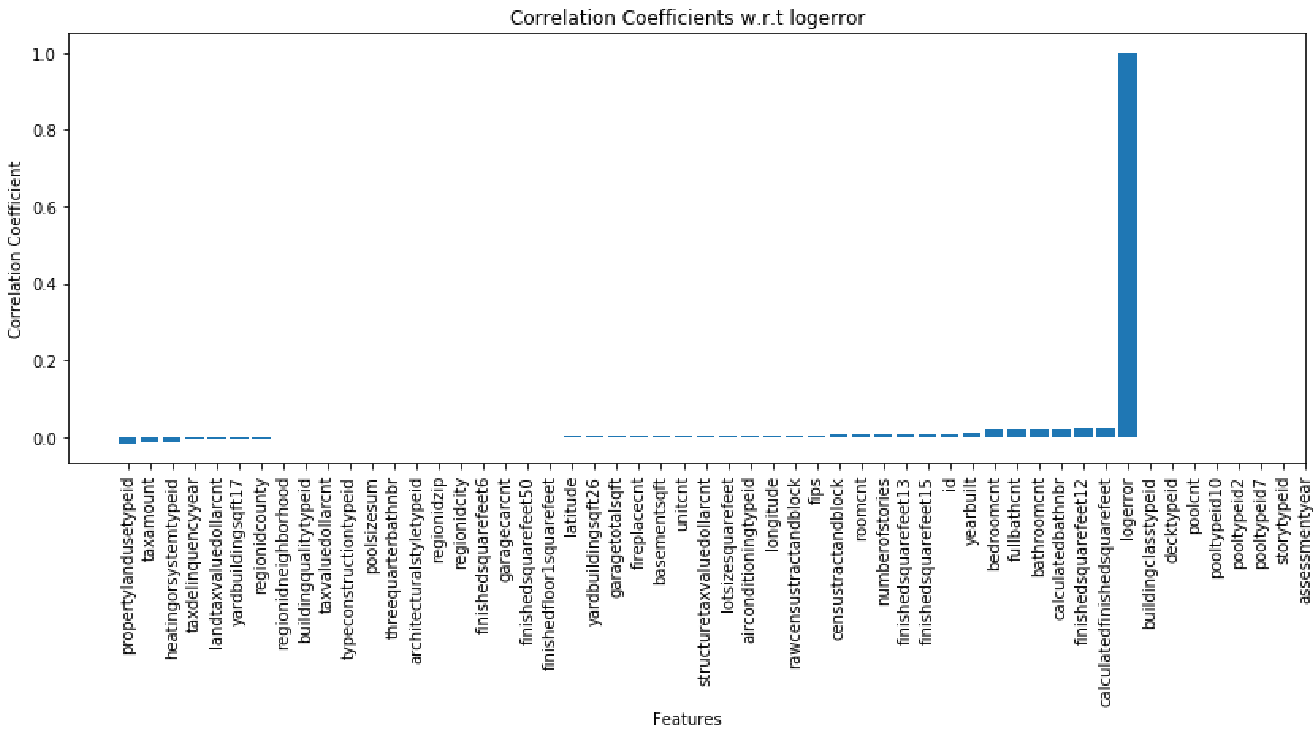
In [137]:

```
df_final=df_final.fillna(df_final.mean())
```

c) univariate analysis

In [138]:

```
# make bar chart
correlation = df_final.corr().iloc[1,:]
correlation.sort_values(ascending=True, inplace=True)
plt.figure(figsize = (14,5))
plt.bar(np.arange(54),np.sort(correlation.values))
plt.title('Correlation Coefficients w.r.t logerror')
plt.xlabel('Features')
plt.ylabel('Correlation Coefficient')
plt.xticks(np.arange(54),correlation.keys(),rotation = 'vertical')
plt.show()
```



explain reason

From the graph we can see that there are columns without any value of correlation. If we look at the data of these features, we find all the values of these fatures are the same. There is no change in values (variance) hence the correlation coefficient value with logerror is is NaN resulting in missing correaltion coefficient.

d) non-linear regression model

In [139]:

```
df_final=df_final.drop(["hashottuborspa", "propertycountylandusecode", "propertyzoningdesc",
"fireplaceflag", "taxdelinquencyflag","id","transactiondate"], axis=1)
y=df_final.iloc[:,0]
x=df_final.iloc[:,1:]
```

In [140]:

```
# split and train
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x,y, test_size=0.30)
```

In [141]:

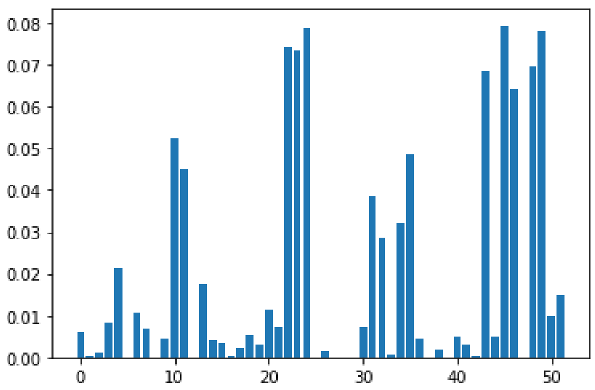
```
from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor()
rf.fit(X_train, y_train)
fi=rf.feature_importances_
y_pred=rf.predict(X_test)
```

```
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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 [142]:

```
# report importances and mse
plt.bar(np.arange(len(fi)),fi)
from sklearn.metrics import mean_squared_error as mse
MSE1=mse(y_test,y_pred)
print(MSE1)
```

```
0.010475270588728618
```



e) KFold

In [143]:

```
# KFold, k = 5
datax=np.array(x.iloc[:500,:])
datay=np.array(y.iloc[:500])
from sklearn.model_selection import KFold
kf=KFold(n_splits=5,shuffle=True)
MSE=[]
for a,b in kf.split(datax):
    trainx=datax[a,:]
    trainy=datay[a]
    testx=datax[b,:]
```

```
    testy=datay[b]
    rf.fit(trainx,trainy)
    fi=rf.feature_importances_
    ypred=rf.predict(testx)
    MSE2=mse(testy,ypred)
    MSE.append(MSE2)

MSE=np.array(MSE)

print('MSE for KFold:',MSE,'Avg MSE:',np.mean(MSE))
```

```
MSE for KFold: [0.01251576 0.01148013 0.0064085  0.00962729 0.01155723] Avg MSE:
0.0103177817836436
```

In [144]:

```
# Run d2 for 100 times
MSE_list=[]
for i in range(100):
    Xtrain,Xtest,ytrain,ytest=train_test_split(datax,datay,test_size=0.3,random_state=i)
    rf = RandomForestRegressor(random_state = i)
    rf.fit(Xtrain,ytrain)
    fi=rf.feature_importances_
    ypred1=rf.predict(Xtest)
    MSE3=mse(ytest,ypred1)
    MSE_list.append(MSE3)

MSE_list
```

```
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
```

```
  "10 in version 0.20 to 100 in 0.22.", FutureWarning)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.
```

```
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
```

```
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
```

```
  "10 in version 0.20 to 100 in 0.22.", FutureWarning)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
C:\Users\mohan\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:245: FutureWarning: The defa
ult 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)
```

Out[144]:

```
[0.012061375227502174,
 0.01238404839507565,
 0.009757195579127425,
 0.009982592264602654,
 0.00971437781237707,
 0.013108918708190255,
 0.014059471380348455,
 0.014387240840575277,
 0.010612758897975855,
 0.009974540830847768,
 0.010939235820424357,
 0.010001761960604707,
 0.010056262720859542,
 0.006474531468530929,
 0.01251384120369906,
 0.007644469659575878,
 0.009421651911738157,
 0.009488837650338604,
 0.00942500872646759,
 0.008601529906074129,
 0.01364430968431742,
 0.008876130717341536,
 0.012079045482951984,
 0.011608889062973674,
 0.008581063169112045,
 0.0098001584848041,
 0.00852647815424317,
 0.008063283538660662,
 0.01230811828921441,
 0.010171141670555484,
 0.013186805373276805,
 0.009124226564229655,
 0.013089466763846657,
 0.007594882713935564,
 0.007493272174276747,
 0.014200948194069973,
 0.01173791198165334,
 0.010199170168845 4,
 0.007482180298520437,
 0.012568526973940413,
 0.010071199285177936,
 0.010030386369471385,
 0.0117320916054 5139,
 0.008680805311553276,
 0.009242410377122886,
 0.012772586617029732,
 0.01349017594799377,
 0.015920203689387867,
 0.011356913324458272,
```

```
 0.01066205190936414,
 0.013761994379965196,
 0.011481746891735234,
 0.007590576965291412,
 0.01057620419103451,
 0.010160726466442605,
 0.009434500685336128,
 0.012743664862470097,
 0.010108746780599785,
 0.009480276645307081,
 0.011953487365196304,
 0.007705424919251795,
 0.012476046939482488,
 0.013244709672437495,
 0.009804180750642825,
 0.008683195485729203,
 0.008076365347297231,
 0.01176499407616651 7,
 0.008843204335689733,
 0.011676443925342589,
 0.00845126794442038,
 0.012869709284217394,
 0.01038527085854723,
 0.013809789874356717,
 0.010916277233244074,
 0.012120531482074 9,
 0.011567570528670667,
 0.009483813063117902,
 0.00790075924057199,
 0.011053234846193461,
 0.01092066588911933,
 0.011764254731947189,
 0.009870679277340753,
 0.01186040072747374,
 0.009476224367568347,
 0.009723240422728615,
 0.009266430904649937,
 0.010304594942030457,
 0.009144412915857714,
 0.01068377703270343,
 0.006882690887960816,
 0.011000535874831898,
 0.014355711716805341,
 0.014716527205212687,
 0.011581487273745705,
 0.009807680569712455,
 0.009745388416854388,
 0.010508018994221156,
 0.013541312634639103,
 0.013120929662676776,
 0.008980044725439814]
```

In [145]:

```
minMSE=min(MSE_list)
maxMSE=max(MSE_list)
print("Min MSE from list"+str(minMSE))
print("Max MSE from list"+str(maxMSE))
```

```
Min MSE from list0.006474531468530929
Max MSE from list0.015920203689387867
```

Findings

MSE is different for each seed. This variation can affect the understanding of the results. To overcome this, Cross Validation is used. Cross-validation splits dataset into train and test in KFold and every data points get to be tested exactly once and is used in training k-1 times. This helps in generalization and reduces the bias.