In [7]:
```
from sklearn import tree
import pandas as pd
import os
```

In [8]:
```
df = pd.read_csv("main_data_tree.csv")
df.head(10)
```

Out[8]:

	EOG_quintile	percent_EDS	child_abuse_rate	children_conc_pov	elevated_lead	juv_delinquent	No_HSdegree	parent_unemployed	county_poc	county_tier	poverty_county	median_inc_co
0	4	96.0	4.2	21.2	1.7	13.7	13.5	11.3	50.3	1	37.2	4
1	4	96.0	6.3	18.9	1.3	11.5	6.3	8.3	21.6	3	20.0	5
2	3	96.0	2.0	20.5	0.9	9.9	11.5	8.6	56.6	3	18.9	5
3	4	96.0	9.9	39.3	1.4	19.1	19.8	19.3	74.2	1	34.7	3
4	4	4.0	1.1	27.9	1.5	23.3	9.8	9.5	42.7	2	23.0	5
5	1	4.0	3.0	0.0	2.6	22.8	10.2	2.8	11.5	3	12.4	6
6	1	4.0	2.2	22.5	1.9	18.4	16.7	12.7	52.5	1	29.6	4
7	1	4.0	1.8	3.1	1.2	6.7	5.9	4.0	39.1	3	9.7	8
8	3	4.0	0.8	11.8	1.0	4.8	12.6	9.2	35.7	2	18.7	5
9	1	4.0	1.8	3.1	1.2	6.7	5.9	4.0	39.1	3	9.7	8

In [9]:
```
target = df["EOG_quintile"]
target_names = ["1", "2", "3", "4", "5"]
```

In [10]:
```
data = df.drop("EOG_quintile", axis=1)
feature_names = ["percent_EDS","child_abuse_rate","children_conc_pov","elevated_lead","juv_delinquent","No_HSdegree","parent_unemployed","county_poc","co
unty_tier","poverty_county","median_inc_county"]
data.head()
```

Out[10]:

	percent_EDS	child_abuse_rate	children_conc_pov	elevated_lead	juv_delinquent	No_HSdegree	parent_unemployed	county_poc	county_tier	poverty_county	median_inc_county
0	96.0	4.2	21.2	1.7	13.7	13.5	11.3	50.3	1	37.2	40433
1	96.0	6.3	18.9	1.3	11.5	6.3	8.3	21.6	3	20.0	53419
2	96.0	2.0	20.5	0.9	9.9	11.5	8.6	56.6	3	18.9	59329
3	96.0	9.9	39.3	1.4	19.1	19.8	19.3	74.2	1	34.7	35407
4	4.0	1.1	27.9	1.5	23.3	9.8	9.5	42.7	2	23.0	50112

In [11]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data, target, random_state=42)
```

In [12]:
```
clf = tree.DecisionTreeClassifier()
clf = clf.fit(X_train, y_train)
clf.score(X_test, y_test)
```

Out[12]:
```
0.4478330658105939
```

In [13]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators=200)
rf = rf.fit(X_train, y_train)
rf.score(X_test, y_test)
```

Out[13]:
```
0.48154093097913325
```

In [14]:
```
sorted(zip(rf.feature_importances_, feature_names), reverse=True)
```

Out[14]:
```
[(0.7440865643535888, 'percent_EDS'),
 (0.048522987433570135, 'county_poc'),
 (0.027533033640828344, 'median_inc_county'),
 (0.027302117085389502, 'parent_unemployed'),
 (0.025978038750146027, 'poverty_county'),
 (0.025539733650452, 'child_abuse_rate'),
 (0.02530911585474701, 'children_conc_pov'),
 (0.02520278537106873, 'juv_delinquent'),
 (0.025013662221660224, 'No_HSdegree'),
 (0.019505733394715777, 'elevated_lead'),
 (0.006006227691391518, 'county_tier')]
```

In [ ]: