In [5]:
```
#California house price prediction

#Python · Regression with a Tabular California Housing Dataset
```

In [2]:
```
#Data
```

In [4]:
```
DATAPATH = Path('C:/Users/zizhe/Desktop/Leo Zhao/California house price prediction')
```

In [ ]:
```
train_set = pd.read_csv(DATAPATH/'train.csv')
test_set = pd.read_csv(DATAPATH/'test.csv')
```

In [6]:
```
# load original dataset
original = fetch_california_housing()
features = train_set.columns[1:-1]
df_original = pd.DataFrame(original['data'], columns=features)
df_original['MedHouseVal'] = original['target']
```

In [7]:
```
train_set.head()
```

Out[7]:

	id	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	MedH
0	0	2.3859	15.0	3.827160	1.112100	1280.0	2.486989	34.60	-120.12	
1	1	3.7188	17.0	6.013373	1.054217	1504.0	3.813084	38.69	-121.22	
2	2	4.7750	27.0	6.535604	1.103175	1061.0	2.464602	34.71	-120.45	
3	3	2.4138	16.0	3.350203	0.965432	1255.0	2.089286	32.66	-117.09	
4	4	3.7500	52.0	4.284404	1.069246	1793.0	1.604790	37.80	-122.41	

In [24]:
```
#EDA

#Data Size
#The dataset is small. We can try out more advanced training and ensembling techniques
```

In [25]:
```
print('train_set shape: ', train_set.shape)
print('test_set shape: ', test_set.shape)
print('original shape: ', df_original.shape)
```

```
train_set shape:  (37137, 10)
test_set shape:  (24759, 9)
original shape:  (20640, 9)
```

In [26]:
```
train_set.describe()
```

Out[26]:

	id	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOcc
count	37137.00000	37137.000000	37137.000000	37137.000000	37137.000000	37137.000000	37137.0000
mean	18568.00000	3.851029	26.057005	5.163124	1.062204	1660.778919	2.8312
std	10720.67281	1.803167	12.158221	1.206242	0.096490	1302.469608	2.7024
min	0.00000	0.499900	2.000000	0.851064	0.500000	3.000000	0.9500
25%	9284.00000	2.602300	17.000000	4.357522	1.020305	952.000000	2.3944
50%	18568.00000	3.515600	25.000000	5.068611	1.054545	1383.000000	2.7448
75%	27852.00000	4.699700	35.000000	5.858597	1.088825	1856.000000	3.1253
max	37136.00000	15.000100	52.000000	28.837607	5.873181	35682.000000	502.9906

In [27]:

```
#Missing Data
#here are no missing values!
```

In [28]:

```
total = train_set.isnull().sum().sort_values(ascending=False)
percent = (train_set.isnull().sum()/train_set.isnull().count()).sort_values(ascending=
missing_data = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_data.head(10)
```

Out[28]:

	Total	Percent
id	0	0.0
MedInc	0	0.0
HouseAge	0	0.0
AveRooms	0	0.0
AveBedrms	0	0.0
Population	0	0.0
AveOccup	0	0.0
Latitude	0	0.0
Longitude	0	0.0
MedHouseVal	0	0.0

In [29]:

```
# original data
total = df_original.isnull().sum().sort_values(ascending=False)
percent = (df_original.isnull().sum()/df_original.isnull().count()).sort_values(ascenc
missing_data = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_data.head(10)
```

Out[29]:

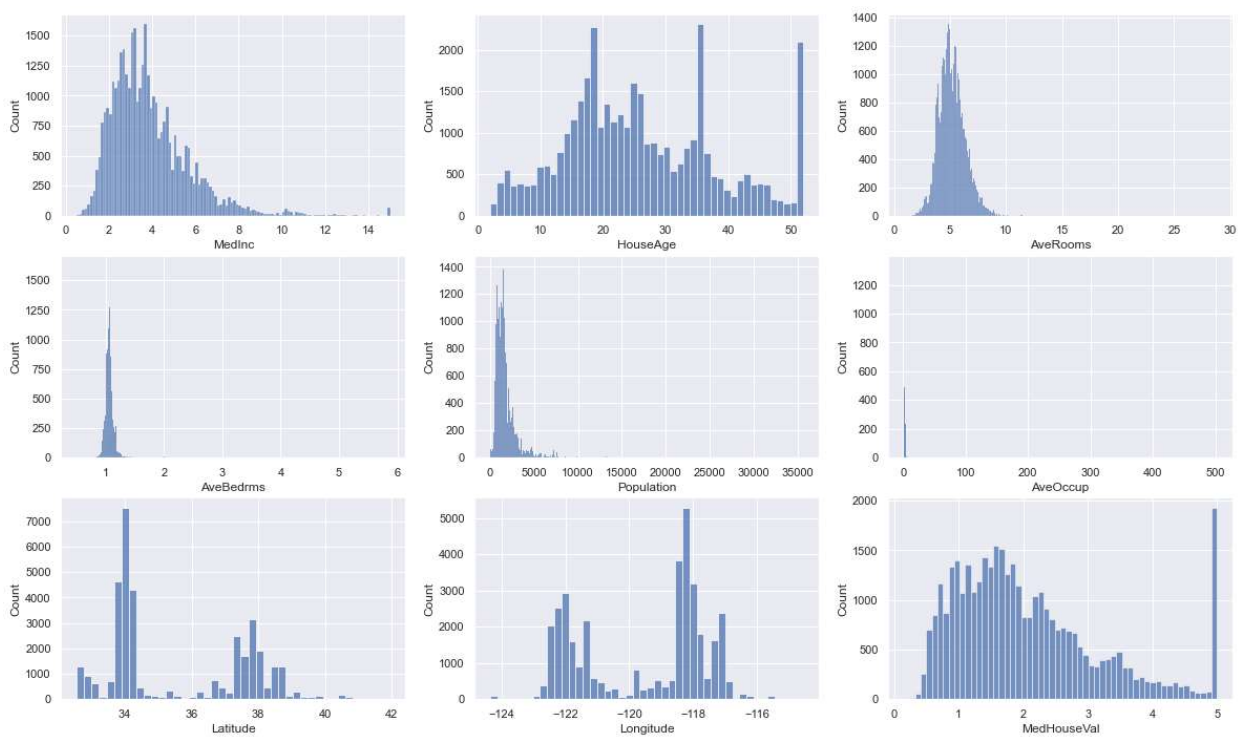
	Total	Percent
MedInc	0	0.0
HouseAge	0	0.0
AveRooms	0	0.0
AveBedrms	0	0.0
Population	0	0.0
AveOccup	0	0.0
Latitude	0	0.0
Longitude	0	0.0
MedHouseVal	0	0.0

In [30]:
```
#Data visualisation
```

In [31]:
```
features = ['MedInc', 'HouseAge', 'AveRooms', 'AveBedrms', 'Population', 'AveOccup', '
target = ['MedHouseVal']
```

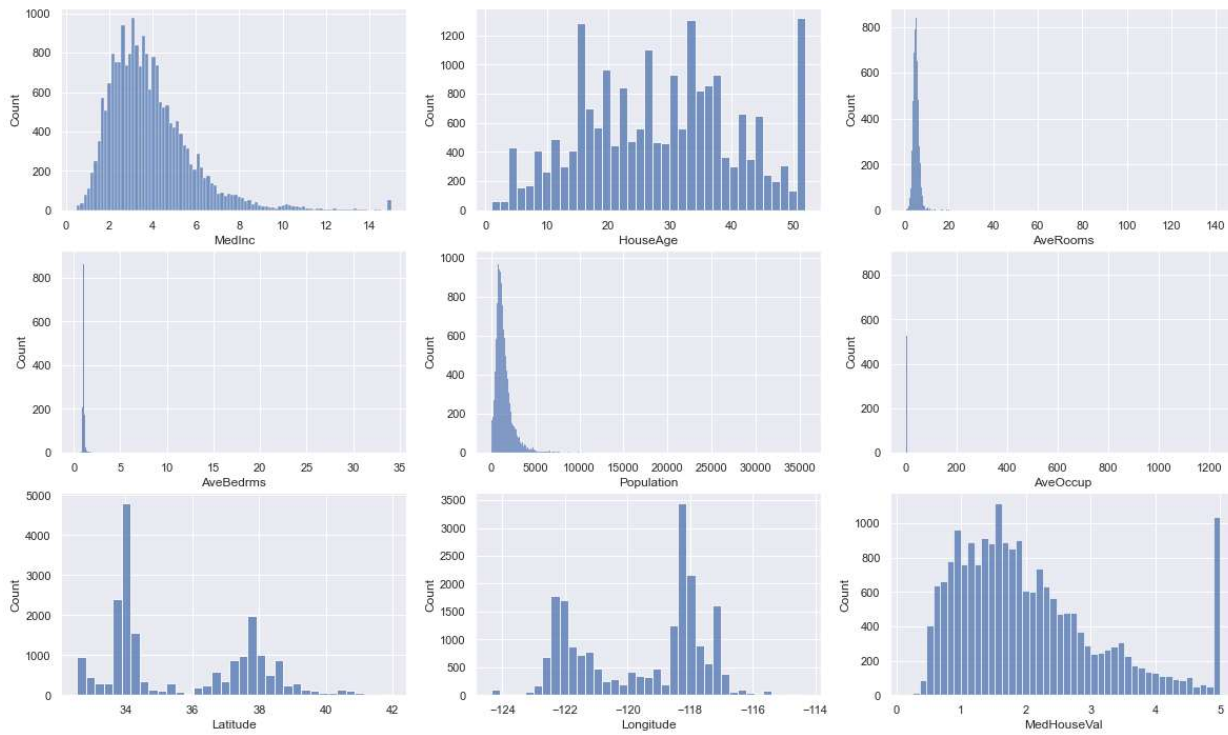
In [32]:
```
#Data Distributions
#There isn't notable differences between train and test
```

In [33]:
```
# Train set
fig, axes = plt.subplots(3,3, figsize=(20, 12))
for i, j in zip(features+target, axes.flatten()):
    sns.histplot(train_set[i], ax=j)
plt.show()
```

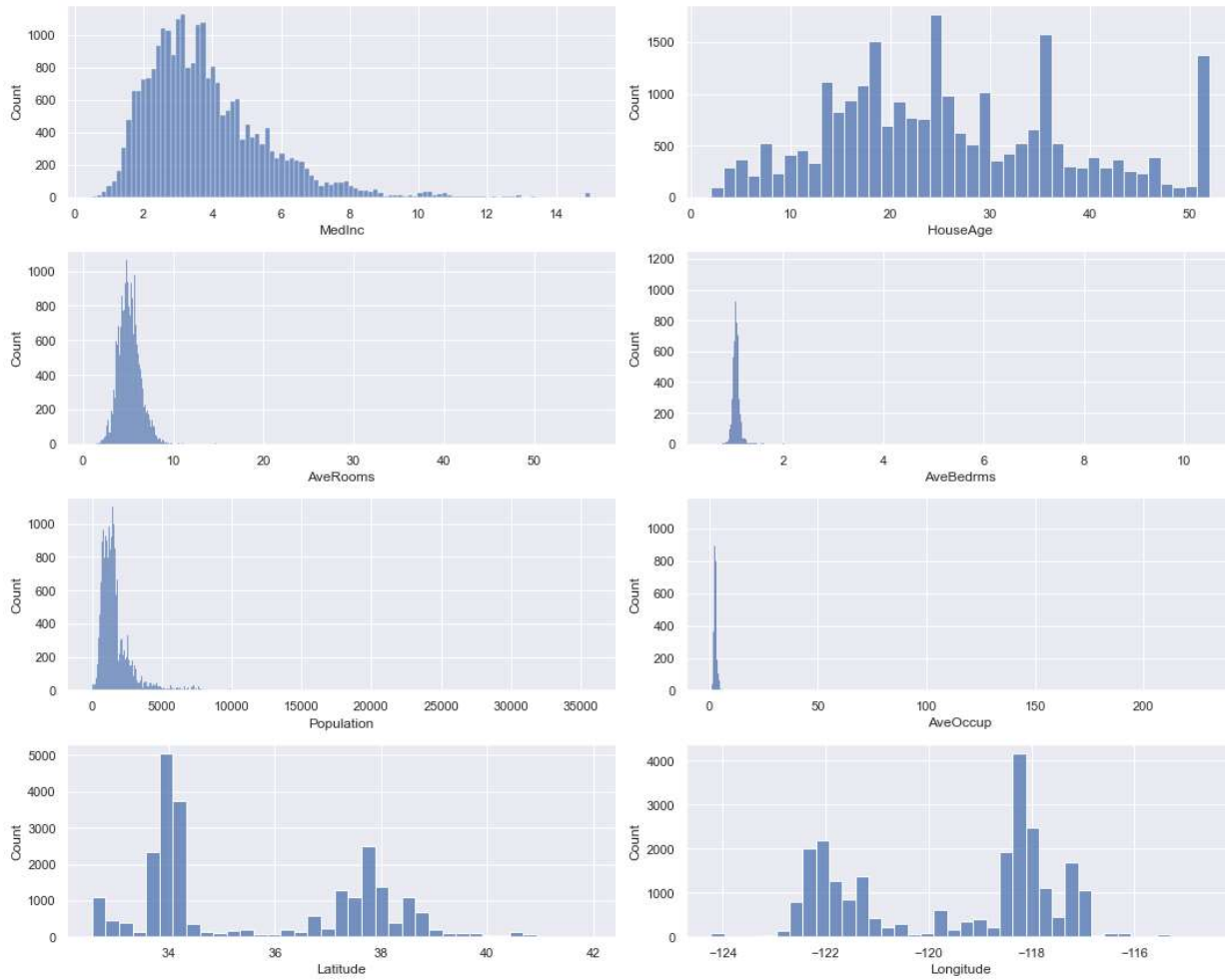


In [34]:
```
# original data
fig, axes = plt.subplots(3,3, figsize=(20, 12))
```

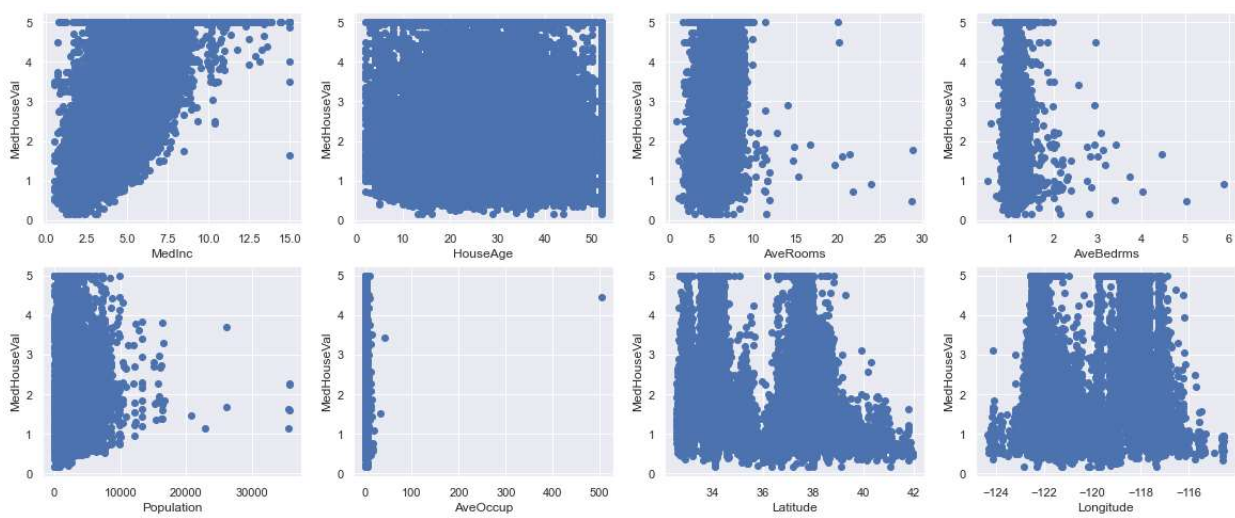
```
for i, j in zip(features+target, axes.flatten()):
    sns.histplot(df_original[i], ax=j)
plt.show()
```



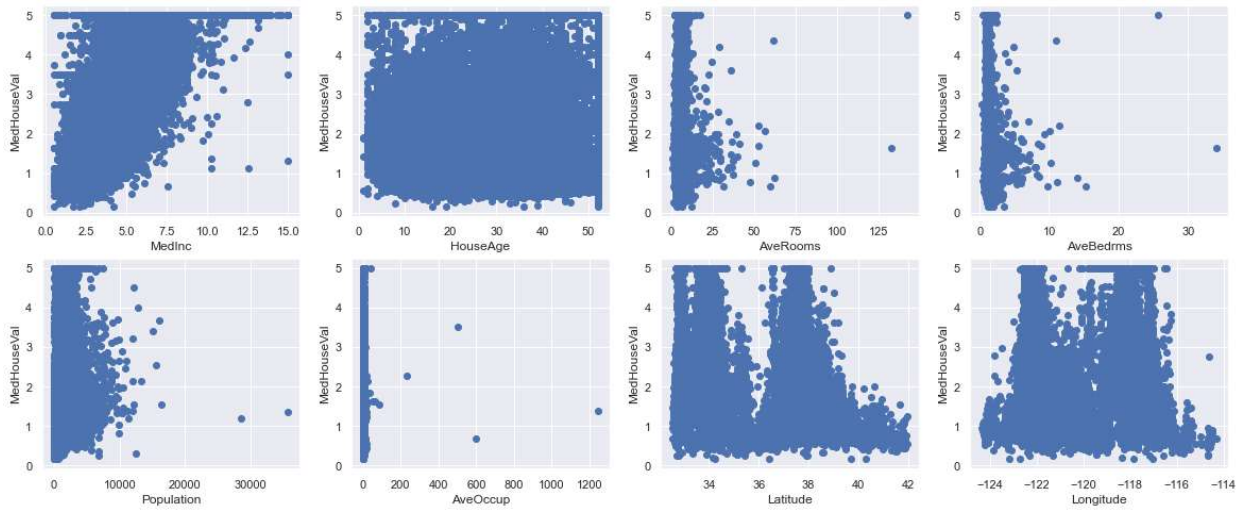
In [35]:
```
# Test set
fig, axes = plt.subplots(4,2, figsize=(15, 12))
for i, j in zip(features, axes.flatten()):
    sns.histplot(test_set[i], ax=j)
fig.tight_layout()
plt.show()
```



In [36]:
```
fig, axs = plt.subplots(2,4, figsize=(20, 8))
for i, j in enumerate(features):
    var = j
    plt.subplot(2,4,i+1)
    data = pd.concat([train_set['MedHouseVal'], train_set[var]], axis=1)
    plt.scatter(x=data[var], y=data['MedHouseVal'])
    plt.xlabel(j)
    plt.ylabel('MedHouseVal')
plt.show()
```



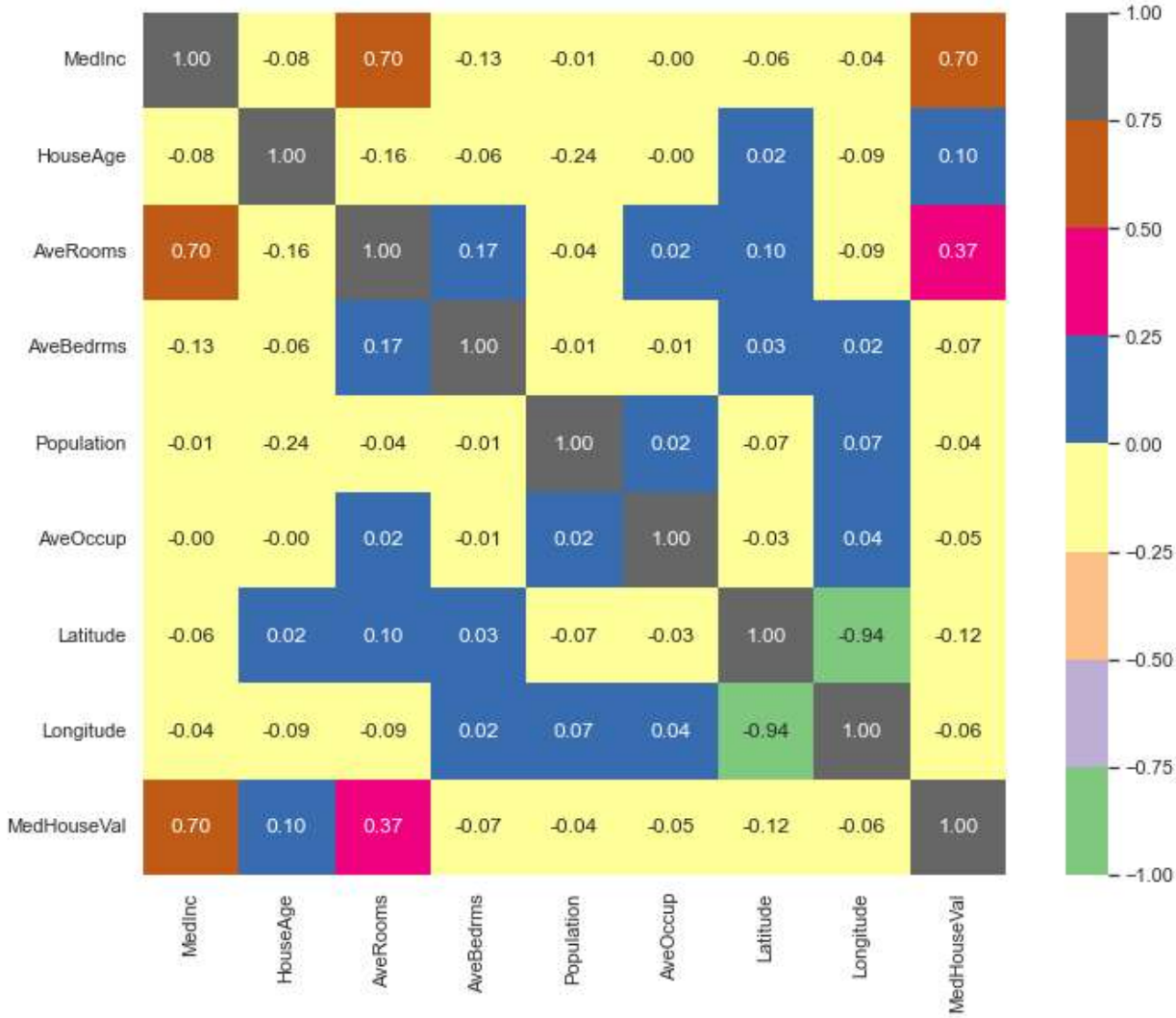
In [37]:
```
fig, axs = plt.subplots(2,4, figsize=(20, 8))
for i, j in enumerate(features):
    var = j
    plt.subplot(2,4,i+1)
    data = pd.concat([df_original['MedHouseVal'], df_original[var]], axis=1)
    plt.scatter(x=data[var], y=data['MedHouseVal'])
    plt.xlabel(j)
    plt.ylabel('MedHouseVal')
plt.show()
```



In [38]:
```
#Correlation matrix (Heatmap Style)¶
#Greater MedInc the greater the MedHouseVal, validated from correlations and from bivc
#Correlations from train and original datasets are different.
```

In [39]:
```
# train set
corrmat = train_set[features+target].corr()
f, ax = plt.subplots(figsize=(12, 9))
sns.heatmap(corrmat, square=True, annot=True, fmt='.2f', cmap='Accent', vmin=-1, vmax=
```

Out[39]:
```
<AxesSubplot:>
```

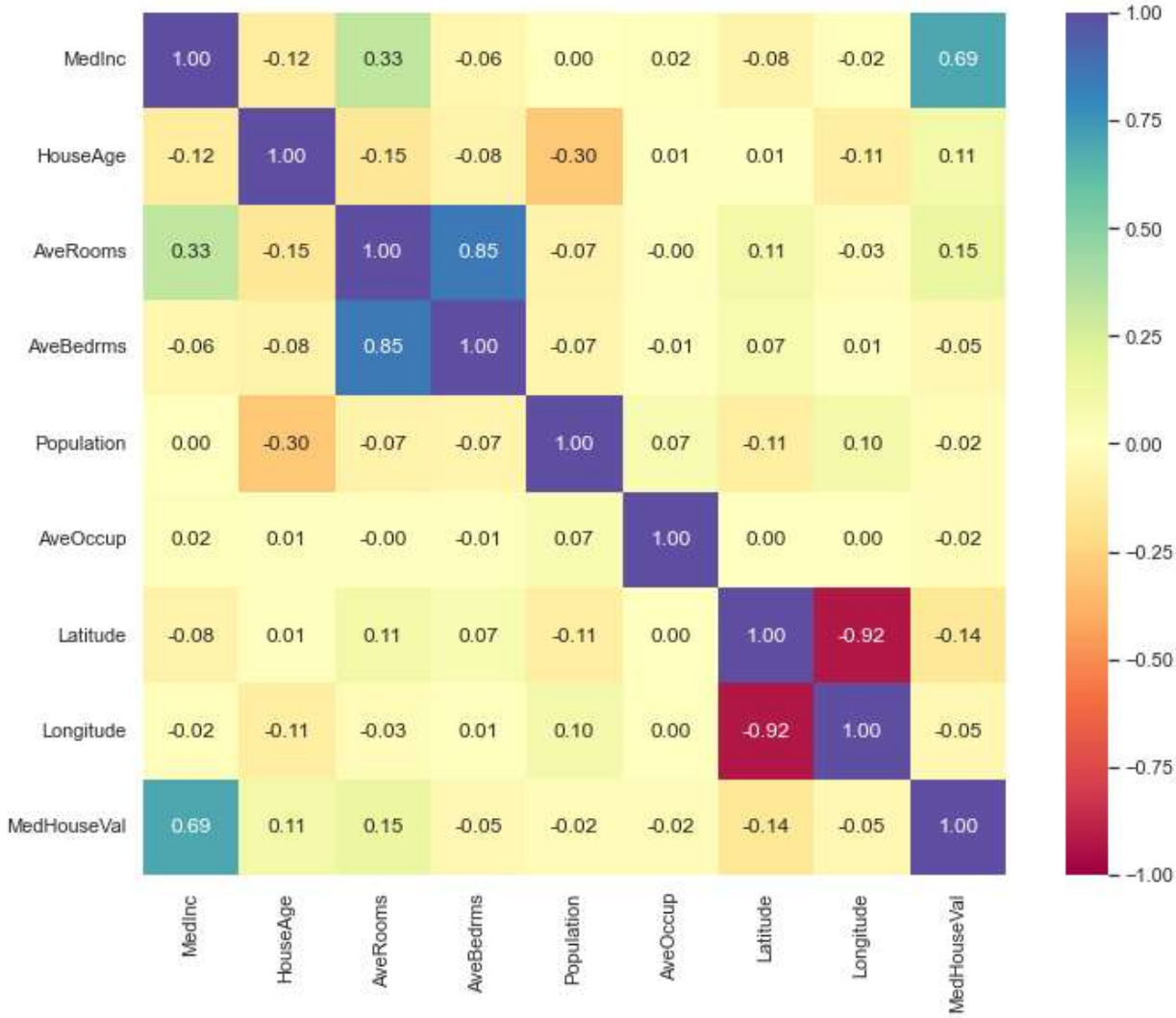


In [40]:

```
# original data
corrmat = df_original[features+target].corr()
f, ax = plt.subplots(figsize=(12, 9))
sns.heatmap(corrmat, square=True, annot=True, fmt='.2f', cmap='Spectral', vmin=-1, vma
```

Out[40]:

```
<AxesSubplot:>
```

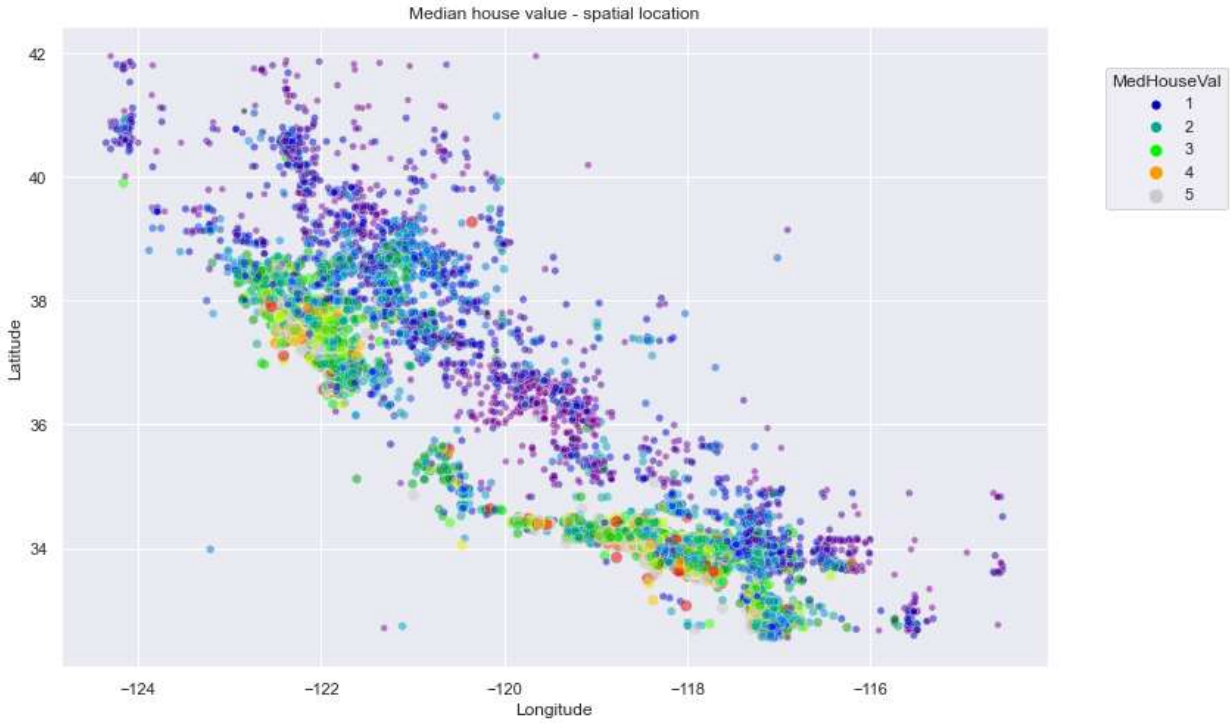


In [41]:
```
#Longitude and Latitude.
#By rotating the coordinates, they may provide more spatial information!
```

In [42]:
```
plt.figure(figsize = (12, 8))

sns.scatterplot(data = train_set, x = "Longitude", y = "Latitude",
                size = "MedHouseVal", hue = "MedHouseVal",
                palette = "nipy_spectral", alpha = 0.5)

plt.legend(title = "MedHouseVal", bbox_to_anchor = (1.05, 0.95), loc = 2)
plt.title("Median house value - spatial location")
plt.show()
```



In [44]:
```
def rt_crds(df):

    df['rot_15_x'] = (np.cos(np.radians(15)) * df['Longitude']) + \
                      (np.sin(np.radians(15)) * df['Latitude'])

    df['rot_15_y'] = (np.cos(np.radians(15)) * df['Latitude']) - \
                      (np.sin(np.radians(15)) * df['Longitude'])

    df['rot_30_x'] = (np.cos(np.radians(30)) * df['Longitude']) + \
                      (np.sin(np.radians(30)) * df['Latitude'])

    df['rot_30_y'] = (np.cos(np.radians(30)) * df['Latitude']) - \
                      (np.sin(np.radians(30)) * df['Longitude'])

    df['rot_45_x'] = (np.cos(np.radians(45)) * df['Longitude']) + \
                      (np.sin(np.radians(45)) * df['Latitude'])

    df['rot_45_y'] = (np.cos(np.radians(45)) * df['Latitude']) - \
                      (np.sin(np.radians(45)) * df['Longitude'])

    return df

df_original = rt_crds(df_original)
train_set = rt_crds(train_set)
test_set = rt_crds(test_set)
```

In [45]:
```
#Rotating the coordinates
```

In [46]:
```
plt.figure(figsize = (12, 8))

sns.set_theme(style='darkgrid')
sns.scatterplot(data = train_set, x = "rot_45_x", y = "rot_45_y",
                size = "MedHouseVal", hue = "MedHouseVal",
                palette = "gist_rainbow", alpha = 0.5)
```

```
plt.legend(title = "MedHouseVal", bbox_to_anchor = (1.05, 0.95), loc = 2)
plt.title("Median house value - spatial location")
plt.show()
```

In [47]:
```
#Train Model
```

In [ ]:
```
#Adding a feature to denote if the data is generated or not
```

In [48]:
```
df_original['is_generated'] = 0
train_set['is_generated'] = 1
test_set['is_generated'] = 1
```

In [49]:
```
train_concat = pd.concat([train_set.drop('id', axis=1), df_original]).reset_index(drop
```

In [50]:
```
features = ['MedInc', 'HouseAge', 'AveRooms', 'AveBedrms', 'Population', 'AveOccup', '
            'is_generated', 'rot_15_x', 'rot_15_y', 'rot_30_x', 'rot_30_y', 'rot_45_x'
target = ['MedHouseVal']
```

In [51]:
```
#XGBRegressor
```

In [ ]:
```
import warnings
warnings.filterwarnings('ignore')

kf = KFold(n_splits=10, random_state=42, shuffle=True)
clfs = []
err = []

for i, (train_index, val_index) in enumerate(kf.split(train_concat)):
    X_train, X_val = train_concat[features].loc[train_index], train_concat[features].l
    y_train, y_val = train_concat['MedHouseVal'][train_index], train_concat['MedHouse\

    clf = XGBRegressor(n_estimators=10000,
                       max_depth=9,
                       learning_rate=0.001,
```

```
                        colsample_bytree=0.66,
                        subsample=0.9,
                        min_child_weight=20,
                        reg_lambda=16,
                        tree_method='hist',
                        seed=42)

    clf.fit(X_train.values, y_train,
            early_stopping_rounds=50,
            eval_set=[(X_val.values, y_val)],
            verbose=1000)

    preds = clf.predict(X_val.values)

    rmse = mean_squared_error(y_val, preds, squared=False)
    err.append(rmse)
    clfs.append(clf)
    print(f'RMSE on fold {i}: {rmse}')
    print('-'*50)

    print(f'Average RMSE (five fold): {sum(err)/10}')
```

```
[0]     validation_0-rmse:1.14230
[1000]  validation_0-rmse:0.68309
[2000]  validation_0-rmse:0.55228
[3000]  validation_0-rmse:0.51881
[4000]  validation_0-rmse:0.50865
[5000]  validation_0-rmse:0.50454
[6000]  validation_0-rmse:0.50256
[7000]  validation_0-rmse:0.50143
[8000]  validation_0-rmse:0.50068
[9000]  validation_0-rmse:0.50014
[9999]  validation_0-rmse:0.49968
RMSE on fold 0: 0.4996774562956203
--------------------------------------------------
Average RMSE (five fold): 0.049967745629562035
[0]     validation_0-rmse:1.16186
[1000]  validation_0-rmse:0.69043
[2000]  validation_0-rmse:0.55528
[3000]  validation_0-rmse:0.51983
[4000]  validation_0-rmse:0.50830
[5000]  validation_0-rmse:0.50322
[6000]  validation_0-rmse:0.50040
[7000]  validation_0-rmse:0.49878
[8000]  validation_0-rmse:0.49764
[9000]  validation_0-rmse:0.49677
[9999]  validation_0-rmse:0.49598
RMSE on fold 1: 0.49597305451216167
--------------------------------------------------
Average RMSE (five fold): 0.0995650510807782
[0]     validation_0-rmse:1.16169
[1000]  validation_0-rmse:0.70253
[2000]  validation_0-rmse:0.57241
[3000]  validation_0-rmse:0.53869
[4000]  validation_0-rmse:0.52811
[5000]  validation_0-rmse:0.52358
[6000]  validation_0-rmse:0.52111
[7000]  validation_0-rmse:0.51959
[8000]  validation_0-rmse:0.51850
[9000]  validation_0-rmse:0.51763
[9999]  validation_0-rmse:0.51692
RMSE on fold 2: 0.516924680866617
--------------------------------------------------
Average RMSE (five fold): 0.1512575191674399
[0]     validation_0-rmse:1.13612
[1000]  validation_0-rmse:0.68556
[2000]  validation_0-rmse:0.55976
[3000]  validation_0-rmse:0.52821
[4000]  validation_0-rmse:0.51864
[5000]  validation_0-rmse:0.51455
[6000]  validation_0-rmse:0.51234
[7000]  validation_0-rmse:0.51106
[8000]  validation_0-rmse:0.51016
[9000]  validation_0-rmse:0.50947
[9999]  validation_0-rmse:0.50893
RMSE on fold 3: 0.5089256623568088
--------------------------------------------------
Average RMSE (five fold): 0.2021500854031208
[0]     validation_0-rmse:1.16650
[1000]  validation_0-rmse:0.71461
[2000]  validation_0-rmse:0.58459
[3000]  validation_0-rmse:0.54897
```

```
[4000]	validation_0-rmse:0.53664
[5000]	validation_0-rmse:0.53117
[6000]	validation_0-rmse:0.52809
[7000]	validation_0-rmse:0.52626
[8000]	validation_0-rmse:0.52500
[9000]	validation_0-rmse:0.52407
[9999]	validation_0-rmse:0.52340
RMSE on fold 4: 0.5233998461498987
--------------------------------------------------
Average RMSE (five fold): 0.25449007001811064
[0]	validation_0-rmse:1.16069
[1000]	validation_0-rmse:0.68574
[2000]	validation_0-rmse:0.54723
[3000]	validation_0-rmse:0.51041
[4000]	validation_0-rmse:0.49872
[5000]	validation_0-rmse:0.49355
[6000]	validation_0-rmse:0.49080
[7000]	validation_0-rmse:0.48937
[8000]	validation_0-rmse:0.48845
[9000]	validation_0-rmse:0.48767
[9999]	validation_0-rmse:0.48706
RMSE on fold 5: 0.48705757203858846
--------------------------------------------------
Average RMSE (five fold): 0.3031958272219695
[0]	validation_0-rmse:1.17007
[1000]	validation_0-rmse:0.71327
[2000]	validation_0-rmse:0.58098
[3000]	validation_0-rmse:0.54522
[4000]	validation_0-rmse:0.53343
[5000]	validation_0-rmse:0.52824
[6000]	validation_0-rmse:0.52539
[7000]	validation_0-rmse:0.52379
[8000]	validation_0-rmse:0.52265
[9000]	validation_0-rmse:0.52169
[9999]	validation_0-rmse:0.52095
RMSE on fold 6: 0.5209485166700063
--------------------------------------------------
Average RMSE (five fold): 0.35529067888897015
[0]	validation_0-rmse:1.16027
[1000]	validation_0-rmse:0.70778
[2000]	validation_0-rmse:0.58117
[3000]	validation_0-rmse:0.54896
[4000]	validation_0-rmse:0.53908
[5000]	validation_0-rmse:0.53498
[6000]	validation_0-rmse:0.53277
[7000]	validation_0-rmse:0.53137
[8000]	validation_0-rmse:0.53040
[9000]	validation_0-rmse:0.52968
[9999]	validation_0-rmse:0.52907
RMSE on fold 7: 0.5290734573071948
--------------------------------------------------
Average RMSE (five fold): 0.4081980246196896
[0]	validation_0-rmse:1.16559
```

In [ ]:

```
# LGBMRegressor
```

In [75]:

```
err = []

for i, (train_index, val_index) in enumerate(kf.split(train_concat)):
```

```
    X_train, X_val = train_concat[features].loc[train_index], train_concat[features].l
    y_train, y_val = train_concat['MedHouseVal'][train_index], train_concat['MedHouseV

    clf = lgbm.LGBMRegressor(learning_rate=0.01,
                             max_depth=9,
                             num_leaves=100,
                             colsample_bytree=0.8,
                             subsample=0.9,
                             subsample_freq=5,
                             min_child_samples=36,
                             reg_lambda=28,
                             n_estimators=10000,
                             metric='rmse',
                             random_state=42)

    clf.fit(X_train.values, y_train, eval_set=[(X_val.values, y_val)], callbacks=[lgbm
    preds = clf.predict(X_val.values)

    rmse = mean_squared_error(y_val, preds, squared=False)
    err.append(rmse)
    clfs.append(clf)
   # print(f'RMSE on fold {i}: {rmse}')
   # print('-'*50)

print(f'Average RMSE (five fold): {sum(err)/10}')
```

```
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was
0.001999 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 51999, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.075399
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2491]  valid_0's rmse: 0.499428
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.008176 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 51999, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.076688
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2393]  valid_0's rmse: 0.494617
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.008197 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 51999, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.075723
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[3499]  valid_0's rmse: 0.514738
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.007571 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 51999, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.077851
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2067]  valid_0's rmse: 0.511197
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.006447 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 51999, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.073901
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2720]  valid_0's rmse: 0.521998
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.008004 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 51999, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.076867
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2908]  valid_0's rmse: 0.484474
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.006795 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 51999, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.075375
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2518]  valid_0's rmse: 0.52052
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.006646 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 52000, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.073274
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2168]  valid_0's rmse: 0.527056
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.007566 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3369
[LightGBM] [Info] Number of data points in the train set: 52000, number of used featu
res: 15
```

```
[LightGBM] [Info] Start training from score 2.074941
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
```

```
[2095]  valid_0's rmse: 0.515944
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was
0.007199 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3370
[LightGBM] [Info] Number of data points in the train set: 52000, number of used featu
res: 15
[LightGBM] [Info] Start training from score 2.077508
Training until validation scores don't improve for 100 rounds
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
Early stopping, best iteration is:
[2650]  valid_0's rmse: 0.503438
Average RMSE (five fold): 0.5093410286803228
```

In [ ]:

```
#CatBoostRegressor
```

In [76]:

```
err = []

for i, (train_index, val_index) in enumerate(kf.split(train_concat)):
    X_train, X_val = train_concat[features].loc[train_index], train_concat[features].l
    y_train, y_val = train_concat['MedHouseVal'][train_index], train_concat['MedHouseV

    clf = CatBoostRegressor(iterations=20000,
                            depth=9,
                            learning_rate=0.01,
                            rsm=0.88,
                            subsample=0.795,
                            min_data_in_leaf=35,
                            l2_leaf_reg=8,
```

```
                            random_strength=0.63,
                            bootstrap_type='Bernoulli',
                            grow_policy='SymmetricTree',
                            loss_function='RMSE',
                            eval_metric='RMSE',
                            task_type="CPU",
                            random_state=42)

    clf.fit(X_train.values, y_train, eval_set=(X_val.values, y_val), early_stopping_r
    preds = clf.predict(X_val.values)

    rmse = mean_squared_error(y_val, preds, squared=False)
    err.append(rmse)
    clfs.append(clf)
    #print(f'RMSE on fold {i}: {rmse}')
    #print('-'*50)

print(f'Average RMSE (five fold): {sum(err)/10}')
```

```
0:      learn: 1.1510268        test: 1.1356808 best: 1.1356808 (0)     total: 190ms
remaining: 1h 3m 10s
1000:   learn: 0.5186973        test: 0.5208454 best: 0.5208454 (1000)  total: 42.3s
remaining: 13m 22s
2000:   learn: 0.4882728        test: 0.5103300 best: 0.5103288 (1994)  total: 1m 23s
remaining: 12m 34s
3000:   learn: 0.4671714        test: 0.5067127 best: 0.5067074 (2996)  total: 2m 6s
remaining: 11m 55s
4000:   learn: 0.4496150        test: 0.5051487 best: 0.5051487 (4000)  total: 2m 47s
remaining: 11m 11s
5000:   learn: 0.4341397        test: 0.5045348 best: 0.5045348 (5000)  total: 3m 29s
remaining: 10m 29s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.504432304
bestIteration = 5093

Shrink model to first 5094 iterations.
0:      learn: 1.1488185        test: 1.1550003 best: 1.1550003 (0)     total: 33.9ms
remaining: 11m 17s
1000:   learn: 0.5181564        test: 0.5218663 best: 0.5218663 (1000)  total: 41.1s
remaining: 13m
2000:   learn: 0.4881404        test: 0.5093097 best: 0.5093097 (2000)  total: 1m 22s
remaining: 12m 23s
3000:   learn: 0.4668441        test: 0.5042525 best: 0.5042525 (3000)  total: 2m 4s
remaining: 11m 45s
4000:   learn: 0.4495871        test: 0.5013661 best: 0.5013633 (3997)  total: 2m 46s
remaining: 11m 4s
5000:   learn: 0.4342321        test: 0.4997434 best: 0.4997430 (4998)  total: 3m 28s
remaining: 10m 23s
6000:   learn: 0.4202195        test: 0.4985413 best: 0.4985413 (6000)  total: 4m 10s
remaining: 9m 43s
7000:   learn: 0.4072754        test: 0.4979785 best: 0.4979785 (6993)  total: 4m 51s
remaining: 9m 1s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.4979540494
bestIteration = 7033

Shrink model to first 7034 iterations.
0:      learn: 1.1488885        test: 1.1549739 best: 1.1549739 (0)     total: 42.8ms
remaining: 14m 15s
1000:   learn: 0.5163697        test: 0.5407506 best: 0.5407506 (1000)  total: 40.1s
remaining: 12m 41s
2000:   learn: 0.4862467        test: 0.5289167 best: 0.5289167 (2000)  total: 1m 23s
remaining: 12m 31s
3000:   learn: 0.4652340        test: 0.5235524 best: 0.5235524 (3000)  total: 2m 17s
remaining: 13m 1s
4000:   learn: 0.4480588        test: 0.5202607 best: 0.5202607 (4000)  total: 3m 4s
remaining: 12m 18s
5000:   learn: 0.4327992        test: 0.5185224 best: 0.5185161 (4971)  total: 3m 58s
remaining: 11m 55s
6000:   learn: 0.4188847        test: 0.5172477 best: 0.5172477 (6000)  total: 4m 49s
remaining: 11m 16s
7000:   learn: 0.4062073        test: 0.5163654 best: 0.5163510 (6987)  total: 5m 34s
remaining: 10m 20s
8000:   learn: 0.3945274        test: 0.5156279 best: 0.5156266 (7998)  total: 6m 17s
remaining: 9m 26s
Stopped by overfitting detector  (100 iterations wait)
```

```
bestTest = 0.5156002147
bestIteration = 8012

Shrink model to first 8013 iterations.
0:      learn: 1.1516022        test: 1.1295898 best: 1.1295898 (0)     total: 56.3ms
remaining: 18m 45s
1000:   learn: 0.5169159        test: 0.5335716 best: 0.5335716 (1000)  total: 44.8s
remaining: 14m 9s
2000:   learn: 0.4871937        test: 0.5221757 best: 0.5221732 (1997)  total: 1m 29s
remaining: 13m 28s
3000:   learn: 0.4662199        test: 0.5173530 best: 0.5173520 (2999)  total: 2m 19s
remaining: 13m 8s
4000:   learn: 0.4488644        test: 0.5146415 best: 0.5146415 (4000)  total: 3m 15s
remaining: 13m 2s
5000:   learn: 0.4336453        test: 0.5129856 best: 0.5129848 (4999)  total: 4m 6s
remaining: 12m 19s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.5127962971
bestIteration = 5289

Shrink model to first 5290 iterations.
0:      learn: 1.1482878        test: 1.1598421 best: 1.1598421 (0)     total: 47.2ms
remaining: 15m 43s
1000:   learn: 0.5154419        test: 0.5501140 best: 0.5501140 (1000)  total: 44.5s
remaining: 14m 4s
2000:   learn: 0.4861621        test: 0.5370743 best: 0.5370743 (2000)  total: 1m 27s
remaining: 13m 3s
3000:   learn: 0.4652378        test: 0.5312214 best: 0.5312214 (3000)  total: 2m 10s
remaining: 12m 18s
4000:   learn: 0.4476437        test: 0.5281118 best: 0.5281118 (4000)  total: 2m 55s
remaining: 11m 41s
5000:   learn: 0.4321193        test: 0.5262094 best: 0.5262094 (5000)  total: 3m 39s
remaining: 10m 57s
6000:   learn: 0.4180330        test: 0.5253666 best: 0.5253611 (5996)  total: 4m 23s
remaining: 10m 15s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.5248532067
bestIteration = 6614

Shrink model to first 6615 iterations.
0:      learn: 1.1489474        test: 1.1537773 best: 1.1537773 (0)     total: 34.3ms
remaining: 11m 25s
1000:   learn: 0.5196643        test: 0.5089444 best: 0.5089444 (1000)  total: 46.9s
remaining: 14m 50s
2000:   learn: 0.4889925        test: 0.4969537 best: 0.4969537 (2000)  total: 1m 29s
remaining: 13m 20s
3000:   learn: 0.4680934        test: 0.4924403 best: 0.4924403 (3000)  total: 2m 16s
remaining: 12m 55s
4000:   learn: 0.4509560        test: 0.4899942 best: 0.4899888 (3997)  total: 3m
remaining: 12m 1s
5000:   learn: 0.4356222        test: 0.4884451 best: 0.4884256 (4996)  total: 3m 45s
remaining: 11m 16s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.4881320492
bestIteration = 5234

Shrink model to first 5235 iterations.
```

```
0:      learn: 1.1478804        test: 1.1633784 best: 1.1633784 (0)     total: 33.4ms
remaining: 11m 8s
1000:   learn: 0.5159978        test: 0.5474049 best: 0.5474049 (1000)  total: 45.7s
remaining: 14m 27s
2000:   learn: 0.4859440        test: 0.5350625 best: 0.5350625 (2000)  total: 1m 30s
remaining: 13m 33s
3000:   learn: 0.4649369        test: 0.5294985 best: 0.5294985 (3000)  total: 2m 10s
remaining: 12m 20s
4000:   learn: 0.4473505        test: 0.5263113 best: 0.5263113 (4000)  total: 2m 51s
remaining: 11m 24s
5000:   learn: 0.4320024        test: 0.5244669 best: 0.5244669 (5000)  total: 3m 35s
remaining: 10m 45s
6000:   learn: 0.4181115        test: 0.5231560 best: 0.5231518 (5956)  total: 4m 23s
remaining: 10m 15s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.5224440599
bestIteration = 6873

Shrink model to first 6874 iterations.
0:      learn: 1.1488975        test: 1.1534051 best: 1.1534051 (0)     total: 60.1ms
remaining: 20m 2s
1000:   learn: 0.5152049        test: 0.5501805 best: 0.5501805 (1000)  total: 42.7s
remaining: 13m 30s
2000:   learn: 0.4851286        test: 0.5393591 best: 0.5393591 (2000)  total: 1m 41s
remaining: 15m 17s
3000:   learn: 0.4638463        test: 0.5349751 best: 0.5349751 (3000)  total: 2m 29s
remaining: 14m 5s
4000:   learn: 0.4468288        test: 0.5326781 best: 0.5326781 (4000)  total: 3m 15s
remaining: 13m 2s
5000:   learn: 0.4316068        test: 0.5311558 best: 0.5311558 (5000)  total: 4m
remaining: 12m 2s
6000:   learn: 0.4179426        test: 0.5303755 best: 0.5303488 (5943)  total: 4m 50s
remaining: 11m 17s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.5300834639
bestIteration = 6420

Shrink model to first 6421 iterations.
0:      learn: 1.1483298        test: 1.1587789 best: 1.1587789 (0)     total: 32.9ms
remaining: 10m 57s
1000:   learn: 0.5164777        test: 0.5380456 best: 0.5380456 (1000)  total: 51.3s
remaining: 16m 14s
2000:   learn: 0.4860092        test: 0.5265079 best: 0.5265079 (2000)  total: 1m 39s
remaining: 14m 51s
3000:   learn: 0.4651502        test: 0.5222773 best: 0.5222525 (2995)  total: 2m 25s
remaining: 13m 43s
4000:   learn: 0.4476267        test: 0.5198428 best: 0.5198414 (3997)  total: 3m 7s
remaining: 12m 30s
5000:   learn: 0.4321043        test: 0.5183399 best: 0.5183366 (4993)  total: 3m 59s
remaining: 11m 59s
6000:   learn: 0.4181603        test: 0.5174564 best: 0.5174556 (5984)  total: 4m 41s
remaining: 10m 55s
7000:   learn: 0.4053941        test: 0.5169071 best: 0.5169020 (6994)  total: 5m 28s
remaining: 10m 10s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.516495099
bestIteration = 7881
```

```
Shrink model to first 7882 iterations.
0:	learn: 1.1516169	test: 1.1296979 best: 1.1296979 (0)	total: 55.9ms
remaining: 18m 37s
1000:	learn: 0.5181125	test: 0.5274349 best: 0.5274349 (1000)	total: 46.4s
remaining: 14m 41s
2000:	learn: 0.4882176	test: 0.5149727 best: 0.5149727 (2000)	total: 1m 42s
remaining: 15m 21s
3000:	learn: 0.4674137	test: 0.5098138 best: 0.5098138 (3000)	total: 2m 29s
remaining: 14m 8s
4000:	learn: 0.4500764	test: 0.5070716 best: 0.5070612 (3998)	total: 3m 13s
remaining: 12m 52s
5000:	learn: 0.4346835	test: 0.5053247 best: 0.5053247 (5000)	total: 4m 5s
remaining: 12m 16s
6000:	learn: 0.4207104	test: 0.5044618 best: 0.5044592 (5998)	total: 5m 7s
remaining: 11m 56s
7000:	learn: 0.4081277	test: 0.5037707 best: 0.5037489 (6974)	total: 5m 57s
remaining: 11m 4s
8000:	learn: 0.3961660	test: 0.5032402 best: 0.5032221 (7943)	total: 6m 49s
remaining: 10m 14s
Stopped by overfitting detector  (100 iterations wait)

bestTest = 0.5030916
bestIteration = 8329

Shrink model to first 8330 iterations.
Average RMSE (five fold): 0.5115882331164175
```

In [ ]:

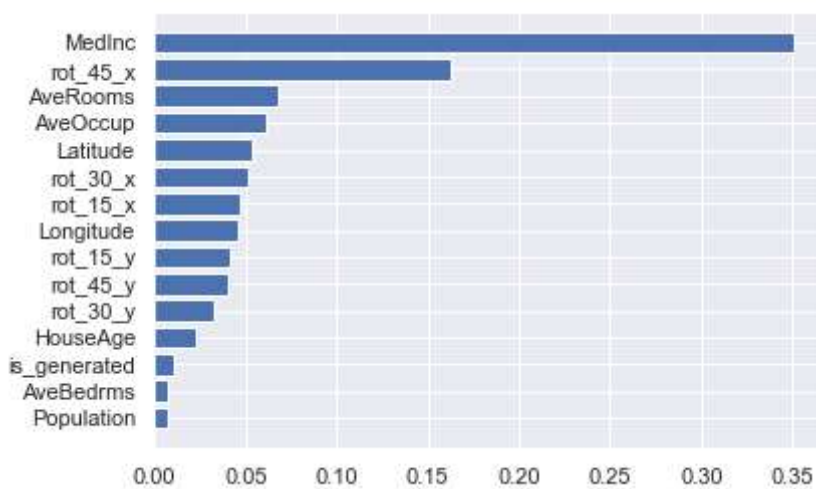
```
#Feature Importance
```

In [77]:

```
imp = np.zeros(15)
for clf in clfs[:10]:
    imp+= clf.feature_importances_

print('XGBoost')
plt.barh([features[i] for i in np.argsort(imp/10)], sorted(imp/10))
plt.show()
```

```
XGBoost
```

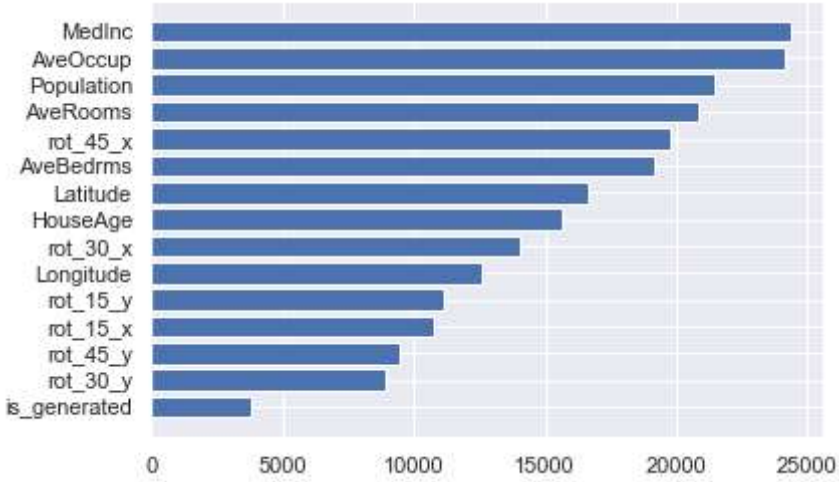


In [78]:

```
imp = np.zeros(15)
for clf in clfs[10:20]:
    imp+= clf.feature_importances_
```

```
print('LGBM')
plt.barh([features[i] for i in np.argsort(imp/10)], sorted(imp/10))
plt.show()
```

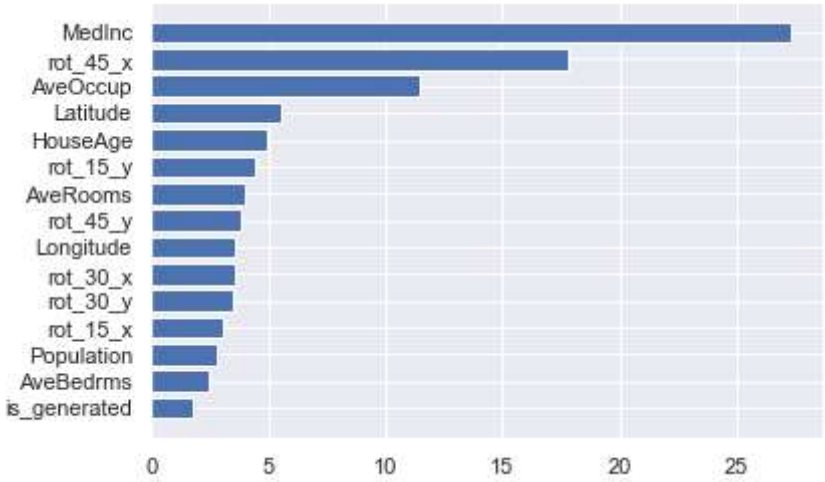
LGBM



In [79]:
```
imp = np.zeros(15)
for clf in clfs[20:]:
    imp+= clf.feature_importances_

print('CatBoost')
plt.barh([features[i] for i in np.argsort(imp/10)], sorted(imp/10))
plt.show()
```

CatBoost



In [81]:
```
#Blending
```

In [82]:
```
test_preds1 = []
test_preds2 = []
test_preds3 = []

for clf in clfs[:10]:
    preds = clf.predict(test_set[features].values)
    test_preds1.append(preds)

for clf in clfs[10:20]:
    preds = clf.predict(test_set[features].values)
    test_preds2.append(preds)
```

```
for clf in clfs[20:]:
    preds = clf.predict(test_set[features].values)
    test_preds3.append(preds)
```

In [ ]:
```
#Stacking
```

In [83]:
```
test_preds1 = np.stack(test_preds1).mean(0)
test_preds2 = np.stack(test_preds2).mean(0)
test_preds3 = np.stack(test_preds3).mean(0)
```

In [84]:
```
test_preds = test_preds1*0.5 + test_preds2*0.4 + test_preds3*0.1
```

In [85]:
```
Final result = pd.DataFrame(data={'id': test_set.id, 'MedHouseVal': test_preds})
Final result.head()
```

Out[85]:

	id	MedHouseVal
0	37137	0.652187
1	37138	1.038345
2	37139	4.010434
3	37140	3.230649
4	37141	2.461173

In [86]:
```
vals = train_set['MedHouseVal'].unique().tolist()
Final result['MedHouseVal'] = Final result['MedHouseVal'].apply(lambda x: min(vals, ke
Final result.MedHouseVal.clip(0, 5, inplace=True)
Final result.head()
```

Out[86]:

	id	MedHouseVal
0	37137	0.652
1	37138	1.038
2	37139	4.010
3	37140	3.231
4	37141	2.461