

# The Search For the GentriValue

Using machine learning to classify and  
predict gentrified neighborhoods in New  
York City in the next 10 Years

By Riley Krisch, Micah Achmoody,  
Nicholas Tan and David Ospina

**BOSTON**  
UNIVERSITY

# Gentrification Has Multiple Definitions

The Freeman model:

median household income + share of housing built in the prior 20 years are both less than the metro-area.

The Ellen & O'Regan Model:

ratio of the area's household income, compared to the metro average household income, is less than 0.7

## Gentrification

The McKinnish, et al model:

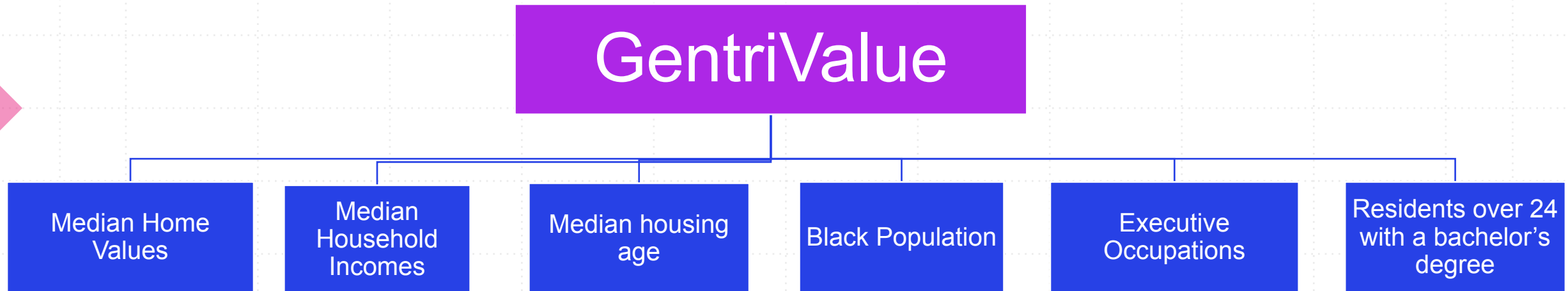
neighborhood average family income is in the bottom 20 percent of all neighborhoods and leads to an increase of at least \$10,000 in the neighborhood's average family income within the last decade.

National Geographic:

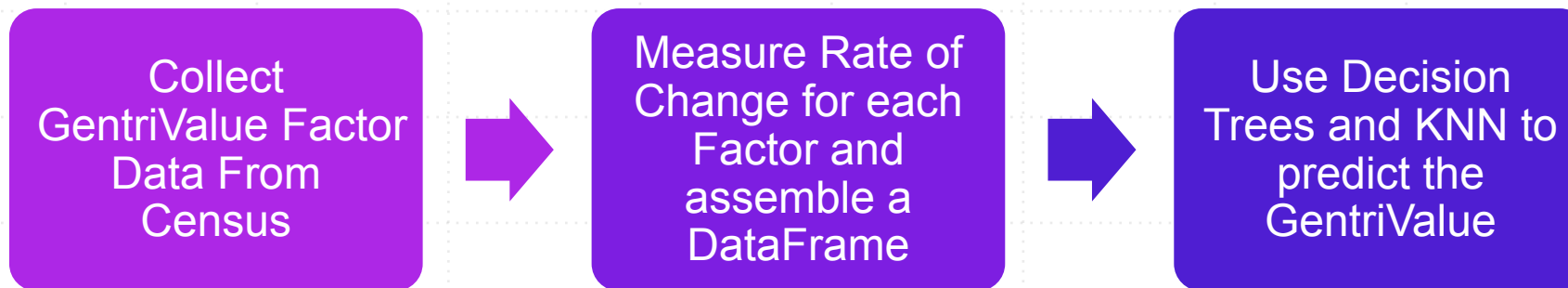
Gentrification is a process of change currently underway in many American cities. Gentrification brings conflict between longtime residents of old neighborhoods and new arrivals.



# Defining the Gentrivalue



## Process For Calculating the Gentrivalue



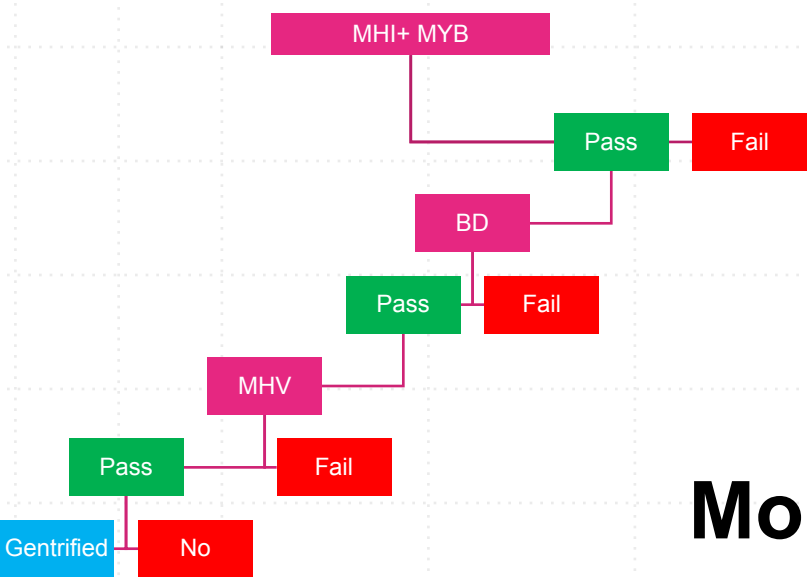
# New Methodology

The current problem is that using fixed metrics to determine if gentrification has occurred isn't actually machine learning cus it's just gg thru a flowchart. Hence the new steps are:

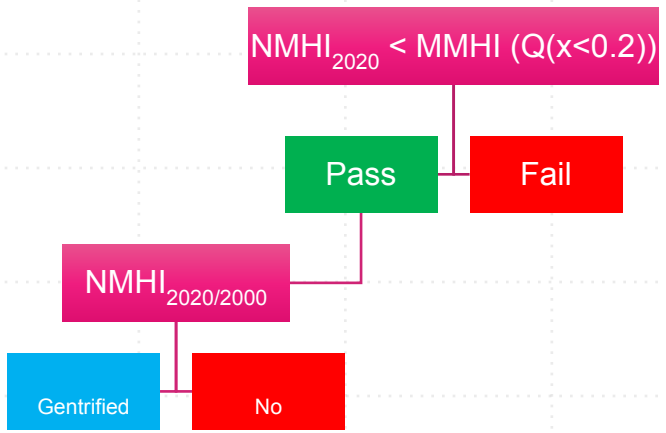
1. Determine if each neighborhood has been gentrified based on the three models by hard coding the criteria for each
2. Utilize machine learning to test how accurate our model is by running it on 80% of the data for training and testing if the last 20% is accurate (evaluation)
3. There is thus no need to actually predict anything to test if the machine learning algorithm is accurate, although we can once we have proved its accuracy

# Overview: Output will yield the GentrifValue

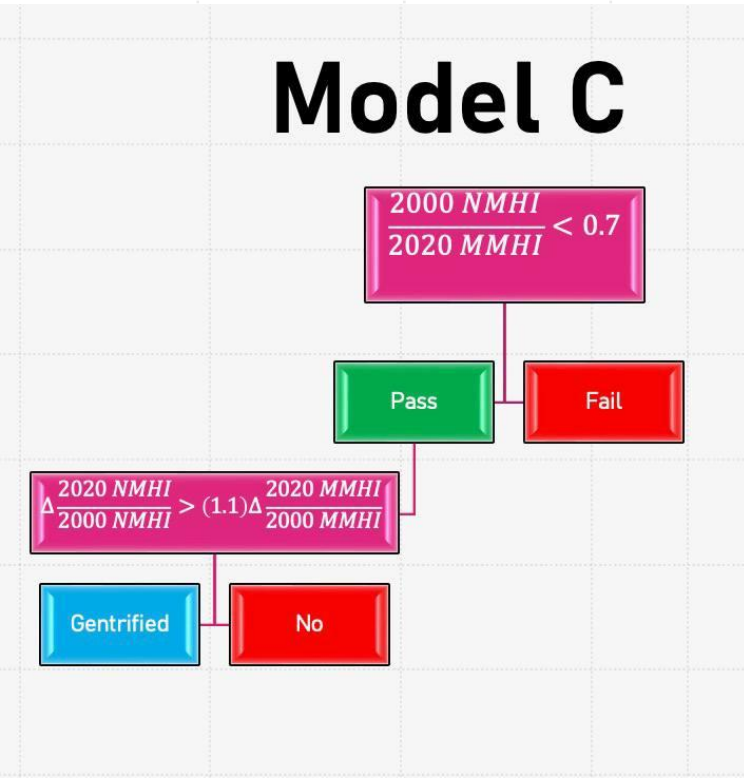
## Freeman



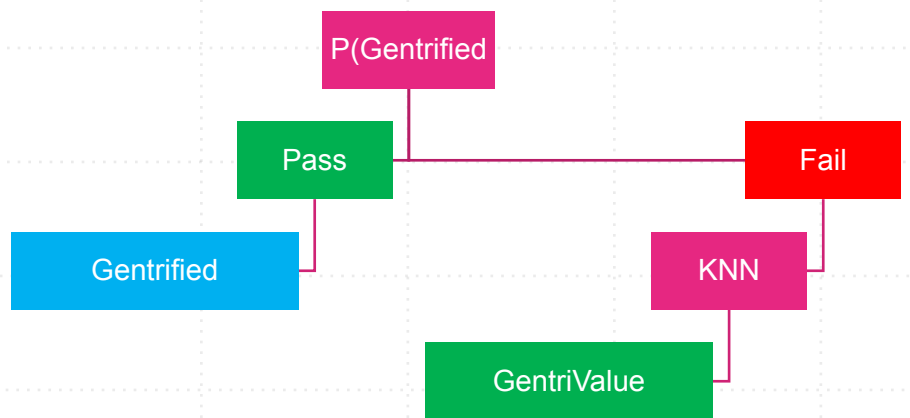
## Mckinnish

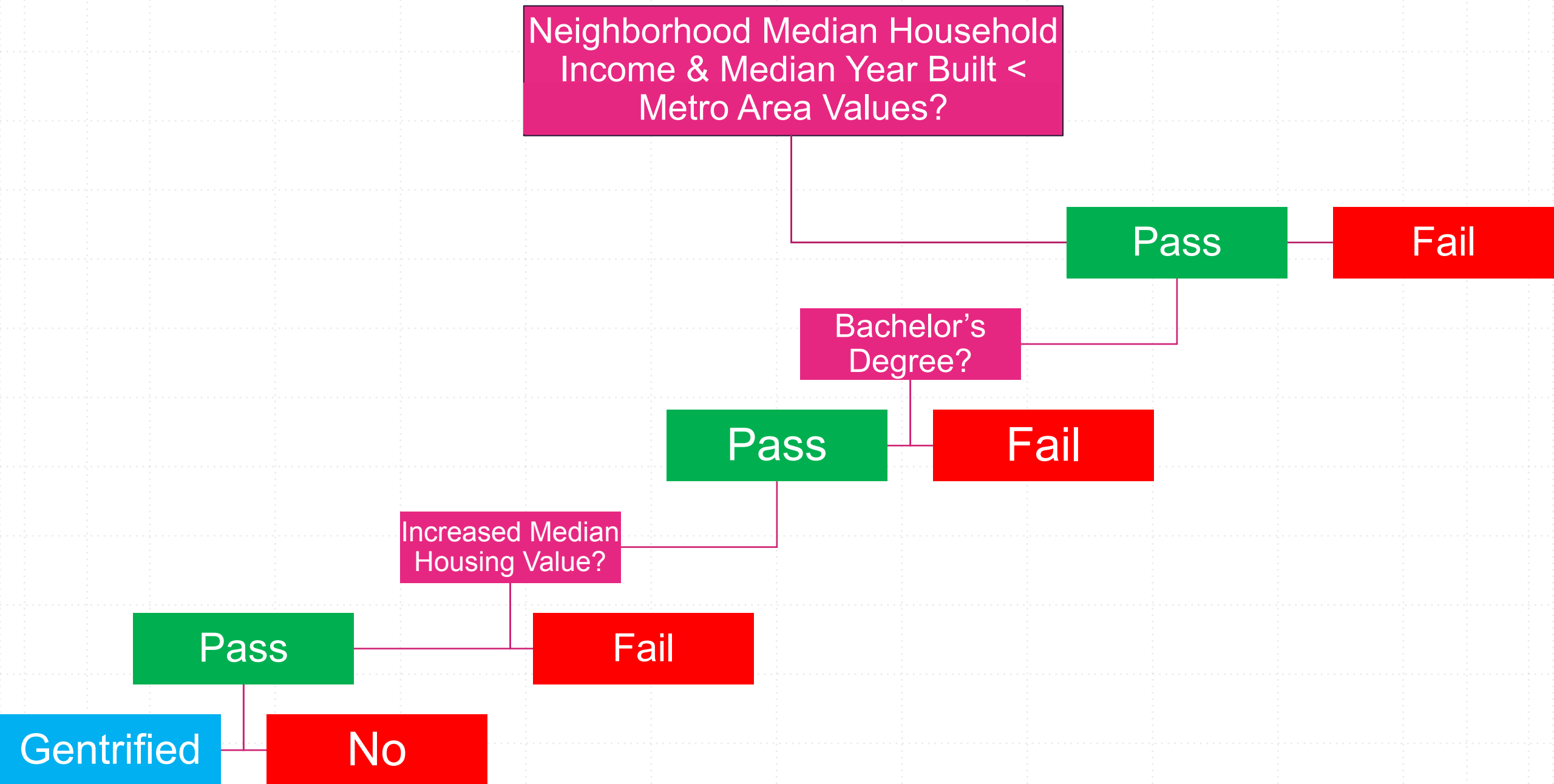


## Model C



Model D = {Model A, B, C}







# Freeman Pseudocode

for MHI2000 and MYB2000:

if NMHI2000<MMHI2000 and NMYB2000<MMYB2000:

# Ellen and O' Regan

2000 NMHI/2020 NMHI

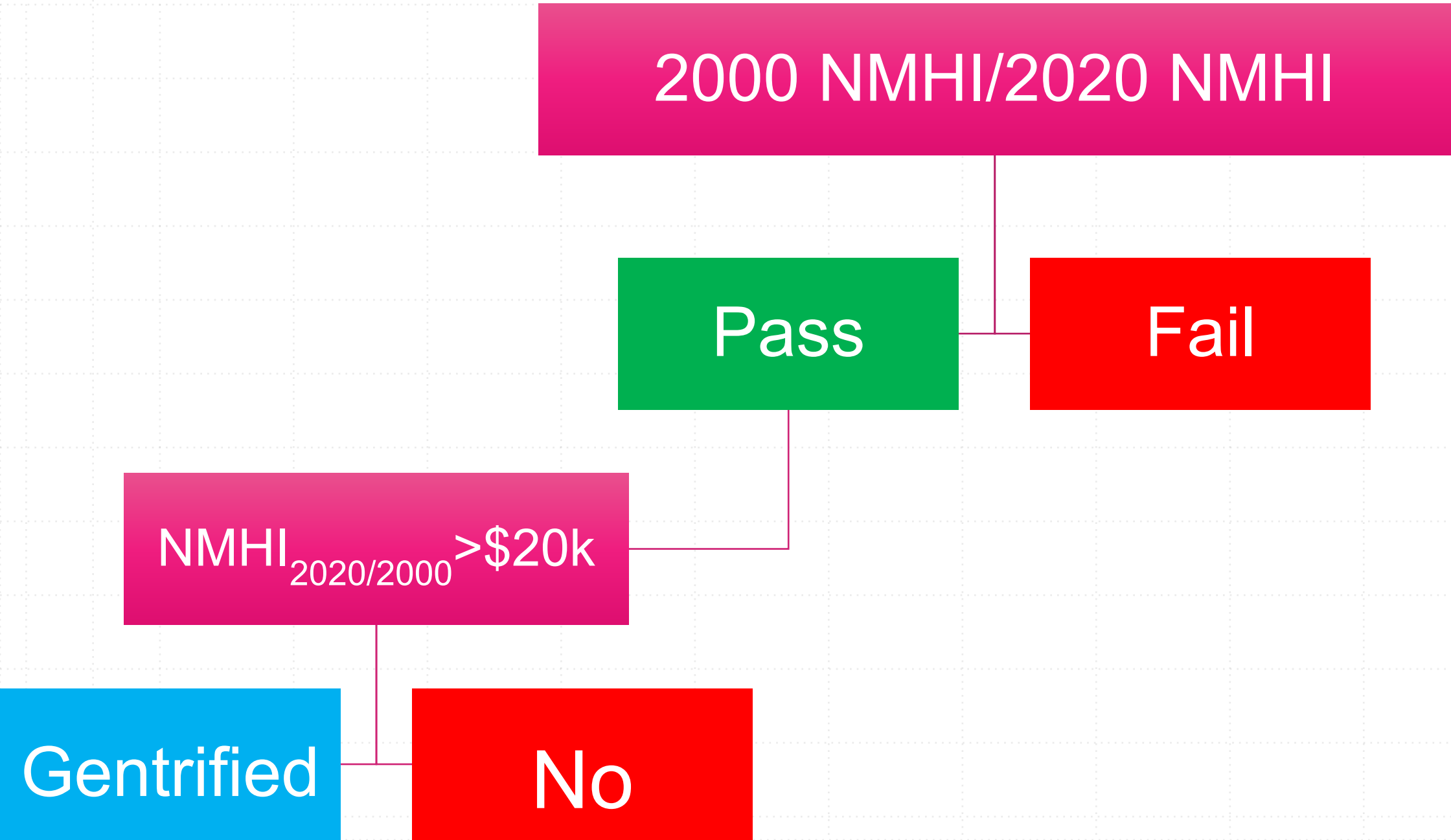
Pass

Fail

$\text{NMHI}_{2020/2000} > \$20k$

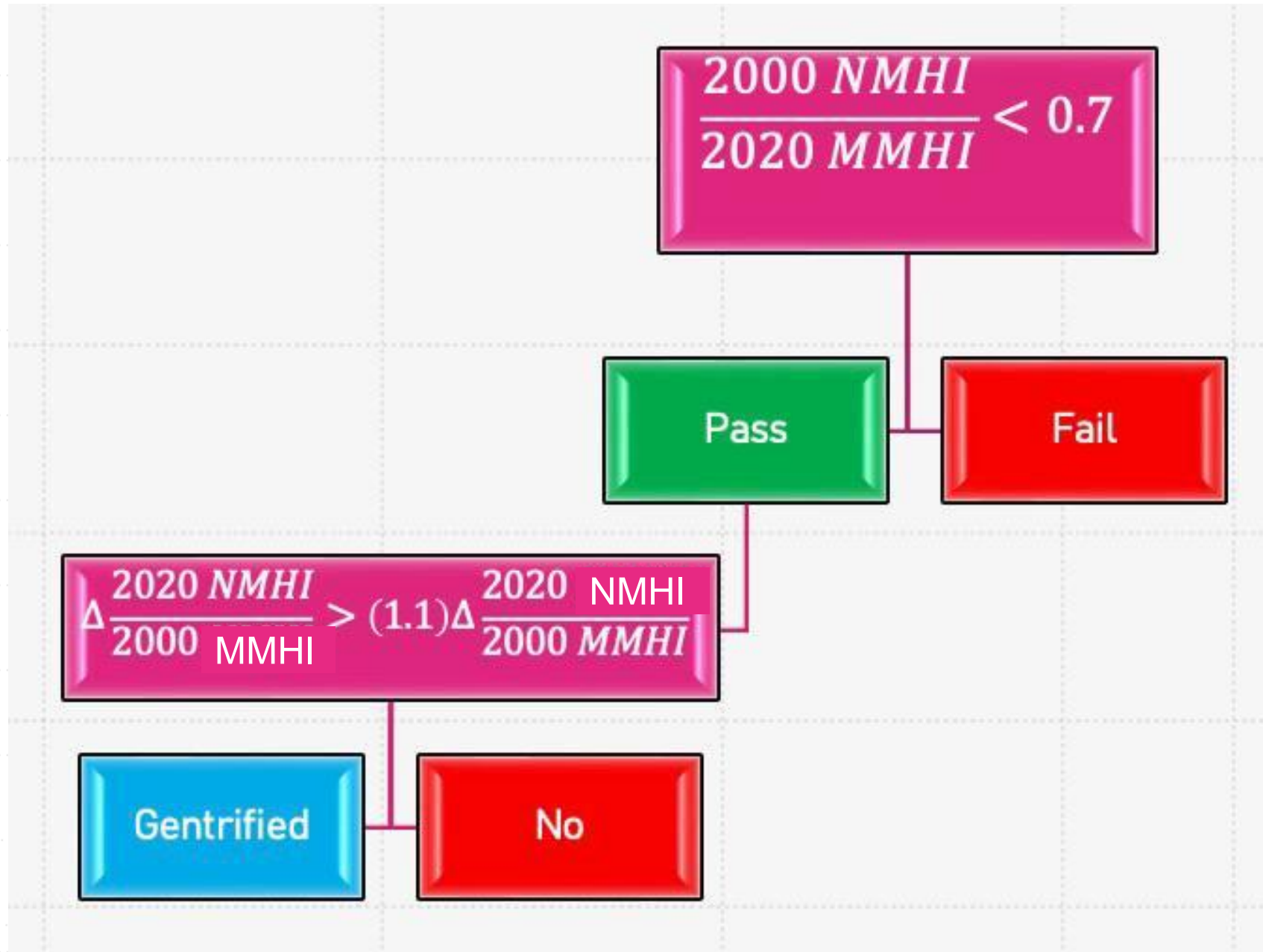
Gentrified

No





# Ellen and O' Reagan





# Ellen Pseudocode

```
def ellen():
```

```
    if (NMHI2000/NMH12020)<0.7 and \
```

```
    (NMHI2020/MMHI2020)>1.1(2000NMHI/2000MMHI):
```

```
        return Gentrified
```

```
    else return None
```

# Mckinnish

$NMHI_{2020} < MMHI (Q(x < 0.2))$

Pass

Fail

$NMHI_{2020/2000} > \$20k$

Gentrified

No



# Mckinnish Pseudocode

```
def mckinnish():
```

```
    for 20th percentile in MMHI2020:
```

```
        if NMHI2020 < MMH12020 and (NMHI2020 - NMHI2000) > 20k:
```

```
            return Gentrified
```

```
        else return None
```



# knn code

```
#knn
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
data_train, data_test, label_train, label_test =
train_test_split(df5.data, df5.target, test_size=0.2)
nbrs = KNeighborsClassifier(n_neighbors=N).fit(data_train,
label_train)

nbrs.score(data_test, label_test)
```



# decision trees code

```
#decision trees
from sklearn.model_selection import train_test_split
import numpy as np

features_train, features_test, labels_train, labels_test = \
train_test_split(df5.data, df5.target, test_size=0.2)

from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import cross_val_score

dtree = DecisionTreeClassifier(criterion="entropy", random_state=110)
dtree.fit(features_train, labels_train)
dtree.score(features_test, labels_test)

import matplotlib.pyplot as plt
from sklearn import tree

plt.figure(figsize=(14,10))
tree.plot_tree(dtree, feature_names = df5.feature_names, class_names =
df5.target_names)
```

# A Vague Definition Leads To Inconsistencies

- Due to the relatively unclear definition of Gentrification based on the three Gentrification models, potential discontinuities in the GentrValue Model consist of:
  1. Other Race Demographics Affected by GentrFactors
  2. Year gaps in the U.S. Census Bureau's Timeline
  3. Government Policies and future Infrastructure Reformation
  4. Inflation of Housing Prices and Income

# Thank You!

- Special thanks to Basilio from the Pardee Library for introducing Policy Map
- Policy Map for providing Census Data
- Keith Galli on YouTube for supplementing our Data Science learning

