**ML Assignment 1**

import pandas as pd

df = pd.read\_csv('/content/Metro\_zhvi\_uc\_sfrcondo\_tier\_0.33\_0.67\_sm\_sa\_month.csv')

df.head()

df= df.fillna(df.mean())

# categorical\_cols = ['SizeRank', 'RegionType', 'StateName']

# X = pd.get\_dummies(X, columns=categorical\_cols)

X = df.drop(['RegionID'], axis=1)

y = df['RegionID']

# from sklearn.preprocessing import LabelEncoder

# label\_encoder = LabelEncoder()

# for col in categorical\_cols:

# X[col] = label\_encoder.fit\_transform(X[col])

categorical\_cols = ['RegionName', 'RegionType', 'StateName']

X = pd.get\_dummies(X, columns=categorical\_cols)

from sklearn.model\_selection import train\_test\_split

# X = df[['SizeRank', 'RegionType', 'StateName']]

from sklearn.linear\_model import LinearRegression

model = LinearRegression()

model.fit(X\_train, y\_train)

from sklearn.metrics import mean\_squared\_error, r2\_score

y\_pred = model.predict(X\_test)

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

print(f"Mean Squared Error: {mse}")

print(f"R-squared: {r2}")

# y = df['2000-01-31']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)