

5G00ET68-3001 Data-analyysi ja tekoälyn perusteet

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
from sklearn.linear_model import LogisticRegression

from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler

df = pd.read_csv('housing.csv')

X = df.iloc[:, [0,1,2,3,4,5,6,7,8]]
y = df.iloc[:, [-1]]

# dummyt
ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(drop='first'), ['ocean_proximity'])], remainder='passthrough')
X = ct.fit_transform(X)

# opetus ja testidata
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

model = LogisticRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)

print(f'r2: {r2}')
print(f'mae: {mae}')
print(f'rmse: {rmse}')
```

Tämän errorin kanssa juminen pahasti. Googlen kautta en löytänyt vastausta.

```
ValueError: A given column is not a column of the dataframe
```