```python
import pandas as pd
import numpy as np

def log(data):
    # 去掉 0
    data = data.drop(data[data['pm2.5'] == 0].index).reset_index(drop=True)
    # 对数转换数据
    data['pm2.5_log'] = np.log(data['pm2.5'])

    return data

def train_test_split(data):
    # 测试数据日期
    test_date = pd.date_range(start='2010-01-07', freq='W-Thu', end='2014-12-25').strftime('%Y-%m-%d').tolist()
    # 测试数据索引
    test_index = []
    for i in range(len(data['date'])):
        if data.loc[i, 'date'] in test_date:
            test_index.append(i)
    # 得到测试集和训练集
    test_data = data.iloc[test_index, :].reset_index(drop=True)
    train_data = data.drop(index=data.index[test_index]).reset_index(drop=True)

    return test_data, train_data
```