import pandas as pd

data = pd.read\_csv('bike.csv') *#使用pandas库读取bike.csv文件*

data = data.drop('id', axis=1) *#剔除id行*

data = data[data['city'] == 1].drop('city', axis=1) *#筛选出上海市的所有数据，并剔除city列*

data['hour'] = data['hour'].apply(lambda x: 1 if 6 <= x <= 18 else 0) *# 将hour列中原来6点-18点统一为1；19点-次日5点统一为0。*

y = data['y'].values

data = data.drop('y', axis=1) *# 提取y列并转换为numpy列向量，然后剔除原y列*

data\_array = data.values *# 将DataFrame对象转换为Numpy数组*

from sklearn.model\_selection import train\_test\_split *#按照8:2的比例划分训练集和测试集*

X\_train, X\_test, y\_train, y\_test = train\_test\_split(data\_array, y, test\_size=0.2, random\_state=42)from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

y\_train = scaler.fit\_transform(y\_train.reshape(-1, 1)).flatten()

y\_test = scaler.transform(y\_test.reshape(-1, 1)).flatten() *#对训练集数据、训练集标签、测试集数据和测试集标签进行归一化*

from sklearn.linear\_model import LinearRegression

model = LinearRegression()

model.fit(X\_train, y\_train) *# 构建一个线性回归模型，并利用训练集训练模型*

y\_pred = model.predict(X\_test) *#利用测试集对训练好的模型进行评估。*

from sklearn.metrics import mean\_squared\_error *#使用均方根误差（RMSE）作为评估指标，并输出RMSE值。*

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

rmse = mse \*\* 0.5print(f'RMSE: {rmse}') *#运行结果：RMSE: 0.1652799928539963*