Bài 1

Nhập các thư viện

import numpy as np

import pandas as pd

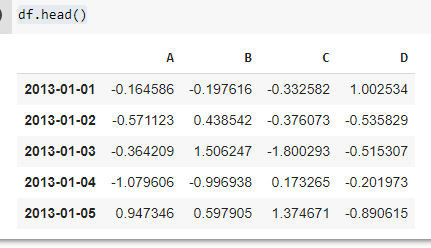
Series và DataFrame

s = pd.Series([1, 3, 5, np.nan, 6, 8])

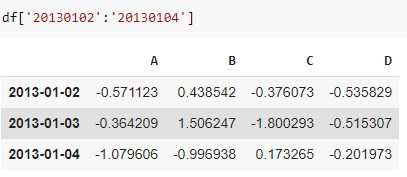
dates = pd.date\_range('20130101', periods=6)

df = pd.DataFrame(np.random.randn(6, 4), index=dates, columns=list('ABCD'))

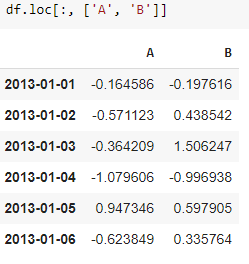
Bảng dữ liệu



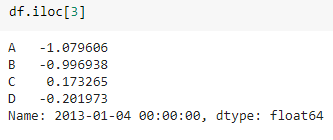
Chọn 1 vùng dữ liệu



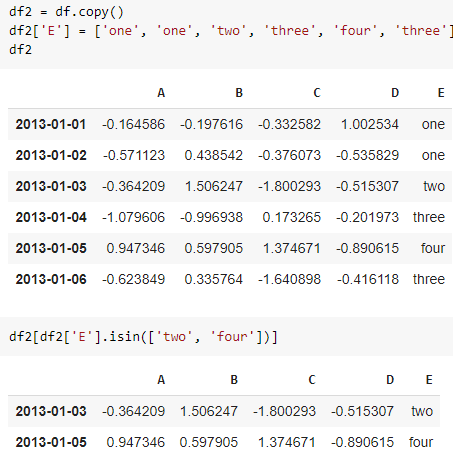
Chọn theo nhãn



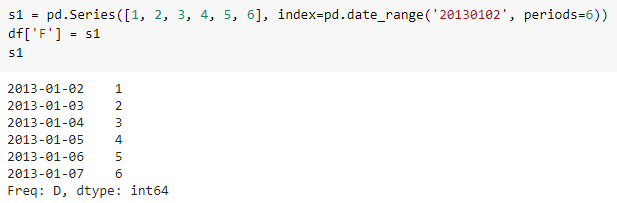
Chọn theo vị trí



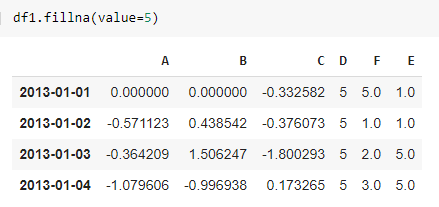
Lập chỉ mục



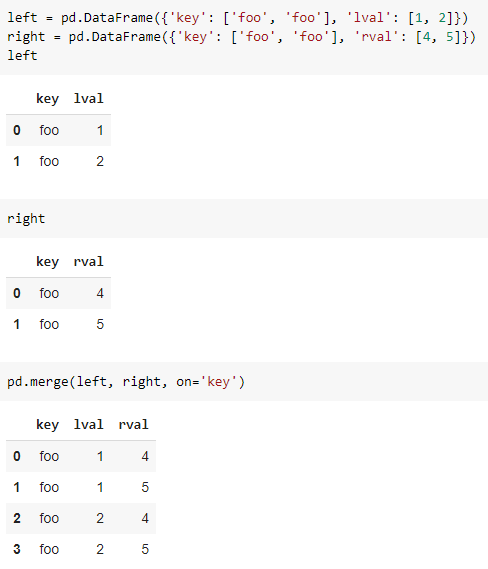
Tạo cột mới



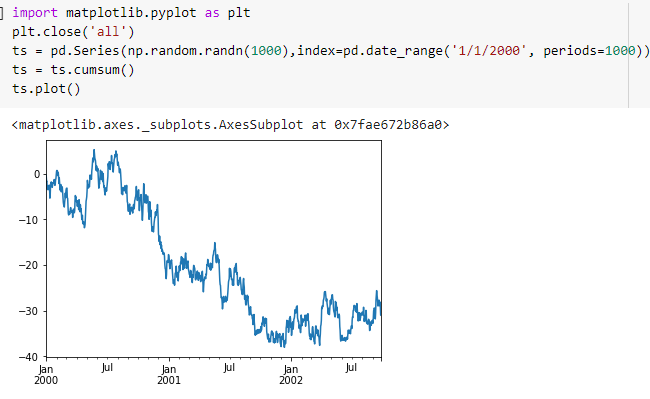
Điền những điểm dữ liệu bị thiếu

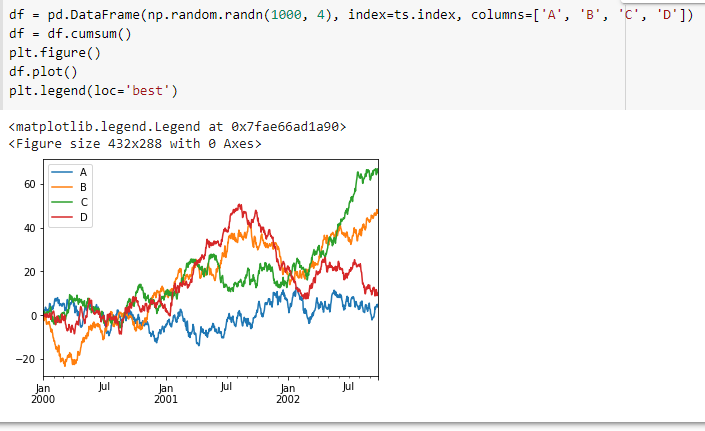


Kết nối 2 bảng

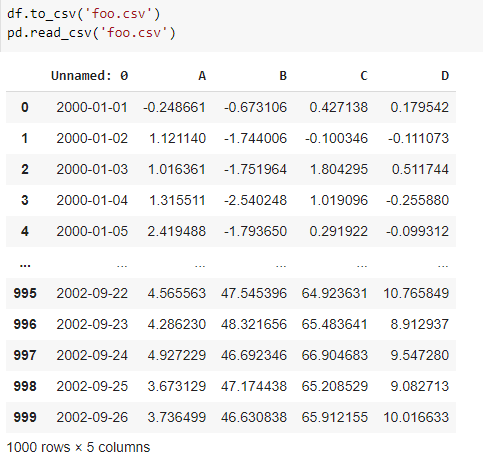


Vẽ biểu đồ



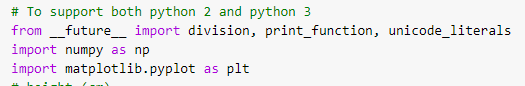


Xuất file csv

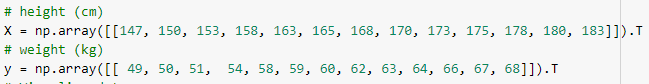


Bài 2

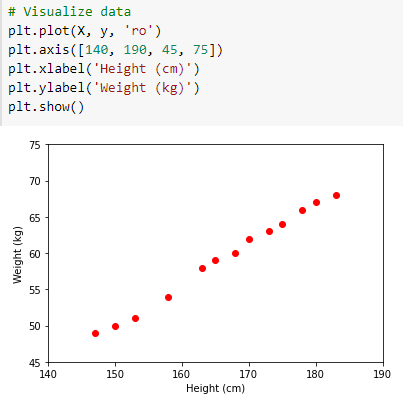
Nhập các thư viện



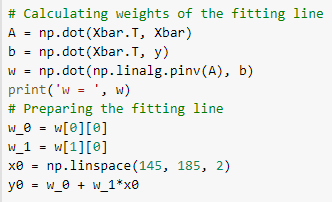
Nhập dữ liệu



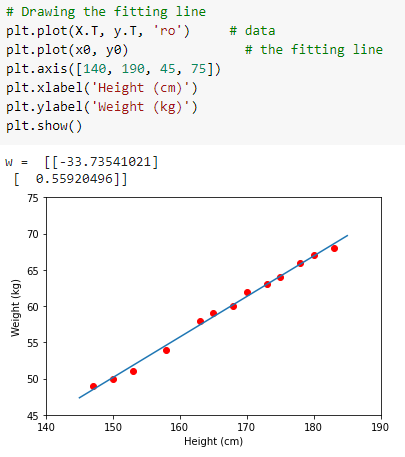
Vẽ biểu đồ



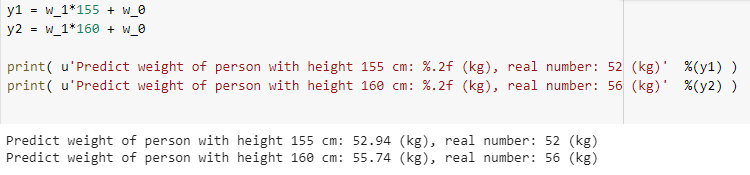
Tạo đường dự đoán



Vẽ đường dự đoán

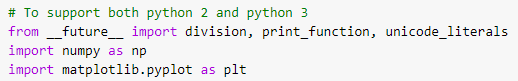


Kiểm tra dự đoán

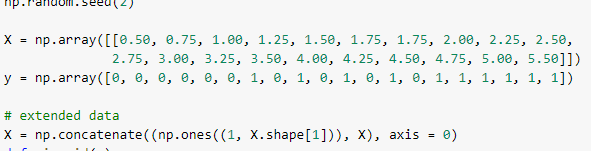


Bài 3

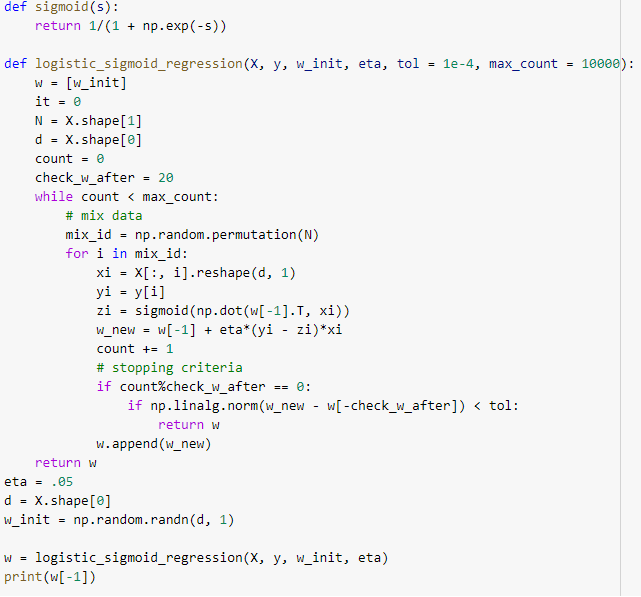
Nhập các thư viện



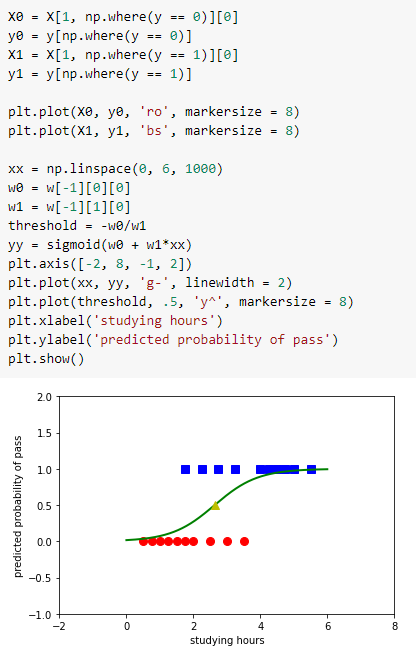
Tạo dữ liệu



Tạo hàm hồi quy

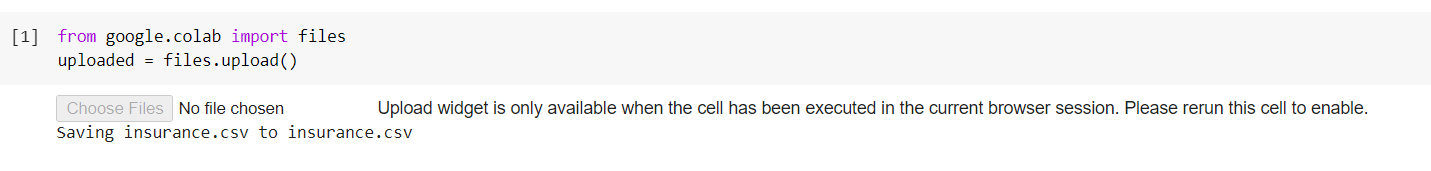


Vẽ biểu đồ



Bài 4:

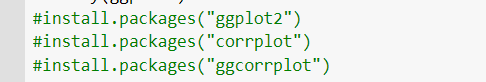
Upload file và install thư viện R.



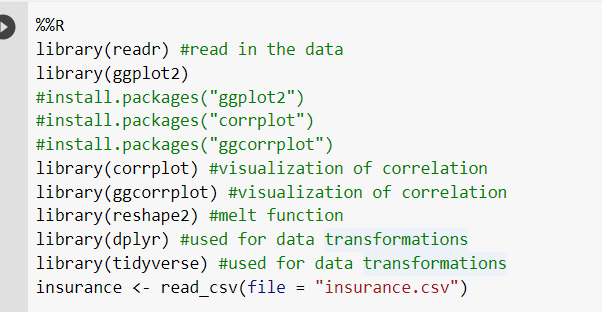
Cài đặt môi trường R



Cài đặt thư viện:

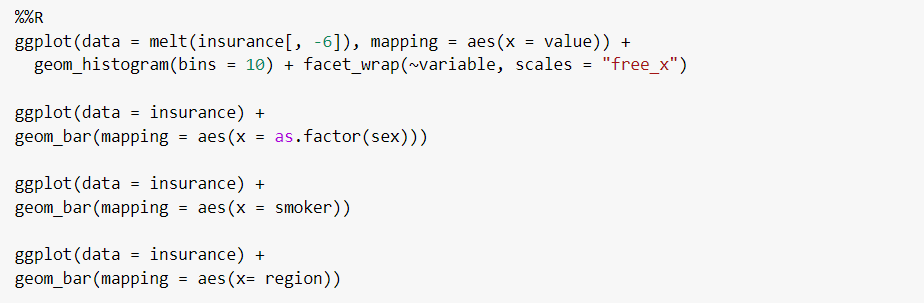


Import thư viện:

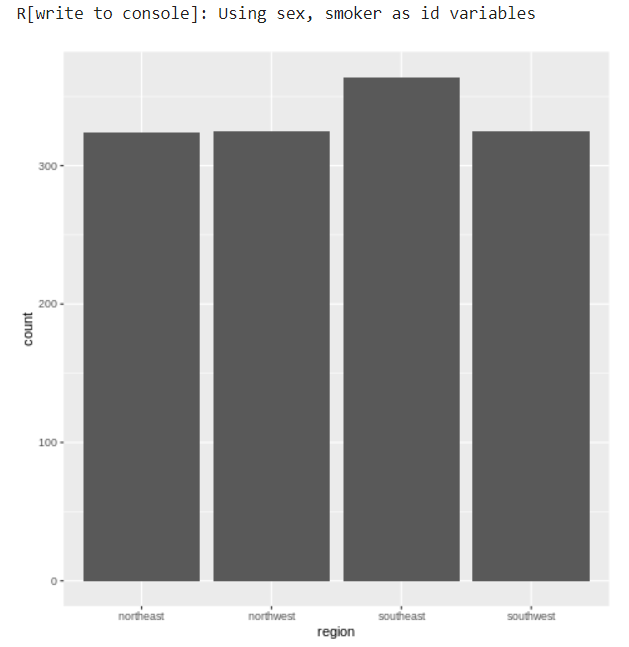


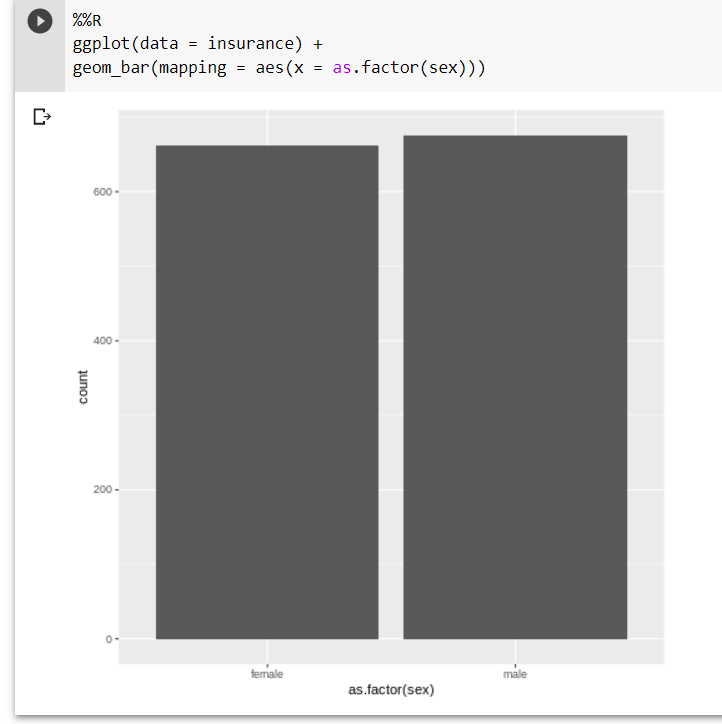


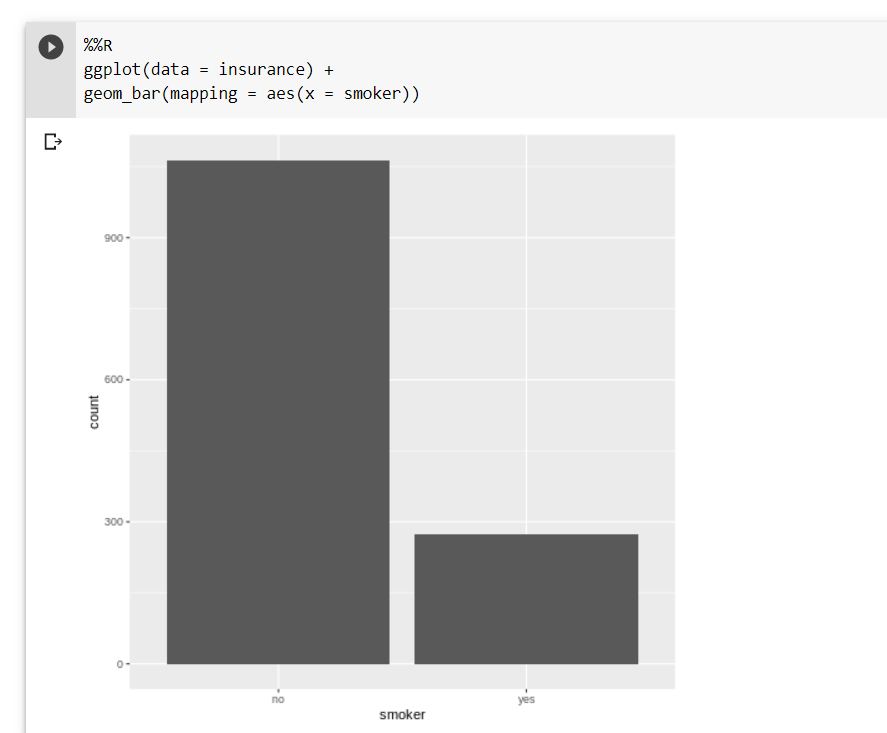
Using sex, smoker as id variables

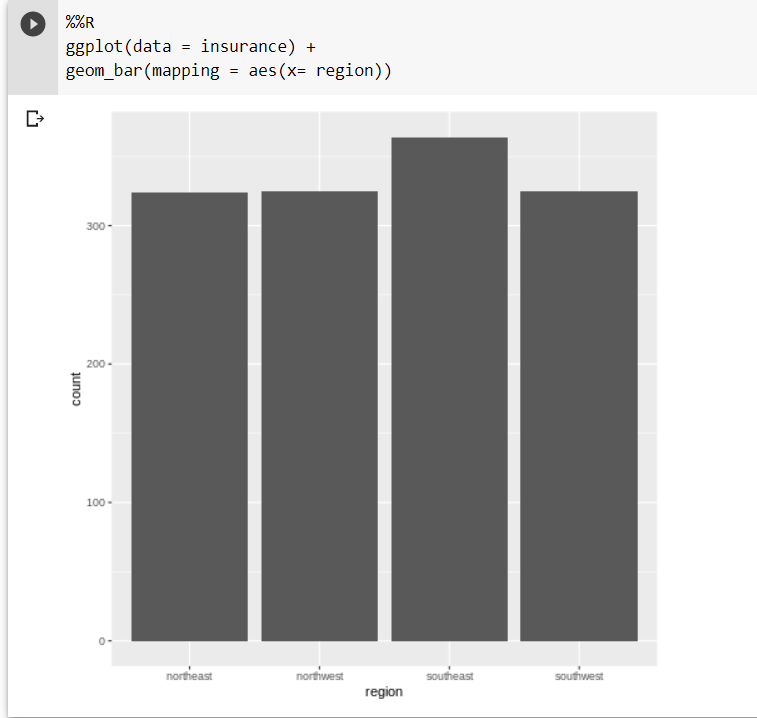


Kết quả:











Bài 6

Import thư viện và read file:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

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

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.svm import SVC

from sklearn.linear\_model import LogisticRegression

from sklearn.preprocessing import StandardScaler

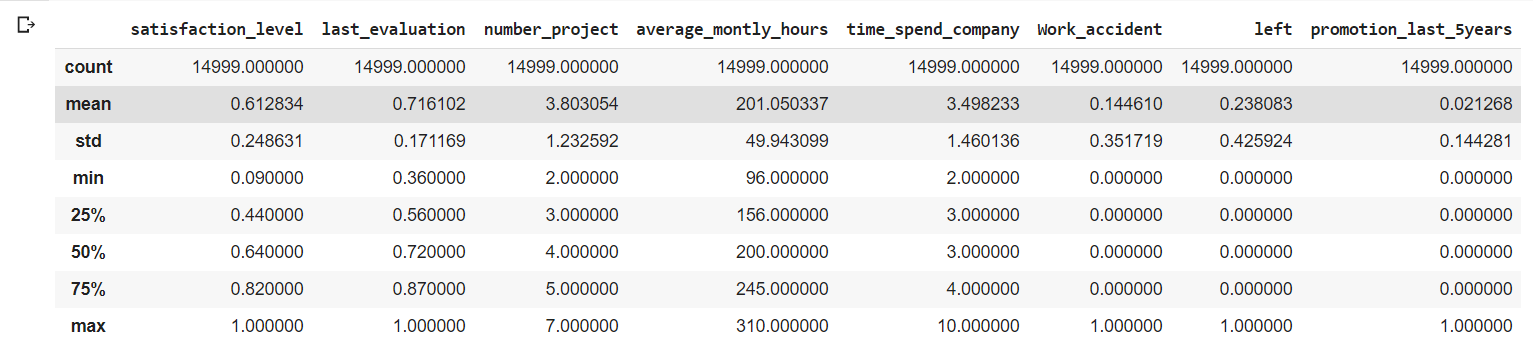
from sklearn.metrics import classification\_report, accuracy\_score, roc\_auc\_score

plt.style.use('ggplot')

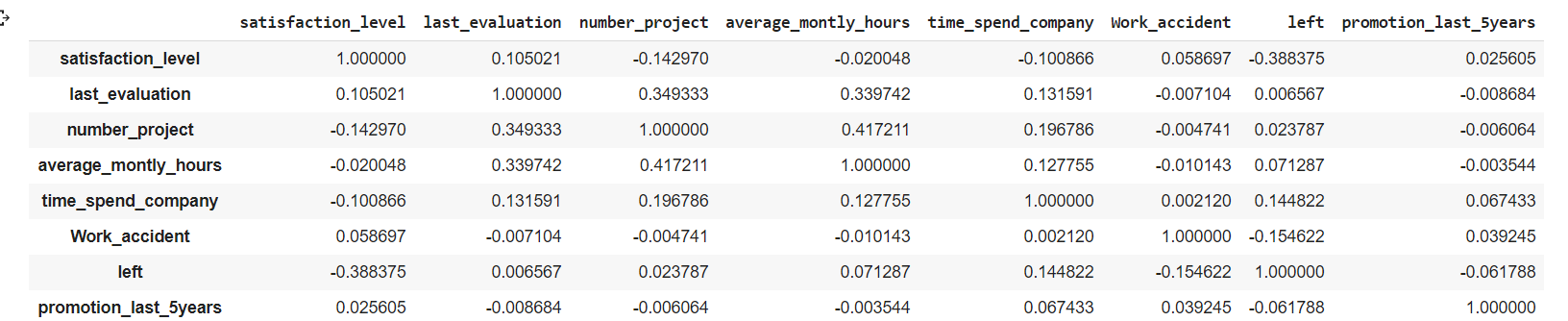
df = pd.read\_csv("/content/sample\_data/HR\_comma\_sep.csv", delimiter=',')

df.info()

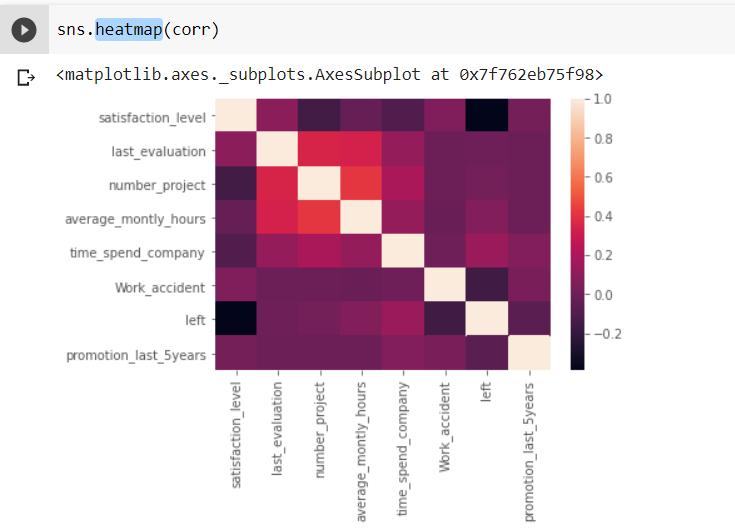
Sử dụng hàm describe() cho các thống kê cơ bản về dữ liệu:



Sử dụng hàm corr() để thể hiện sự tương quan giữa các giá trị



sns.heatmap(corr): sử dụng heatmap để vẽ biểu đồ sử dụng màu sắc để thể hiện độ lơn của giá trị



Dùng sns.countplot() để vẽ biều đồ trục



