**Mô tả tập dữ liệu**

* **Business operations survey 2022: information and communications technology**: file CSV ghi lại dữ liệu kinh doanh hoạt động khảo sát doanh nghiệp trong Công nghệ thông tin và truyền thông trong năm 2022
  + Tập dữ liệu gồm có các trường:
    - Description (Mô tả)
    - Industry (Doanh nghiệp)
    - Level (mức độ)
    - Size (kích thước nhân sự)
    - line\_code
    - value
* **Environmental tax account 1999-2021**: file CSV ghi lại dữ liệu về thuế môi trường trong khoảng năm 1999-2021
  + Tập dữ liệu gồm có các trường:
    - Year (Năm)
    - Industry\_Code (Mã doanh nghiệp)
    - Industry\_Desc (Mô tả doanh nghiệp)
    - Nation\_Account\_Var (Tài khoản quốc gia)
    - Units (Đơn vị tiền tệ)
    - Magnitude (Độ lớn tiền tệ)
    - Source (Nguồn tài khoản)
    - Data\_value (Trị giá dữ liệu)
    - Flag (Đánh dấu)

**Mô tả và diễn giải quá trình đánh giá 4 giả định trong HQTT**

* Đối với file **Business operations survey 2022: information and communications technology:**

*# ===========================================================*

*#Hồi quy tuyến tính*

*#khởi tạo mô hình hồi quy tuyến tính*

linearModel = lnm.LinearRegression()

*#Biến độc lập*

bdl = dataFrame[['value']]

*#Biến phụ thuộc*

bpt = dataFrame['level']

*#Huấn luyện mô hình trên dữ liệu*

linearModel.fit(bdl, bpt)

*#Hệ số hồi quy*

hshq = linearModel.coef\_

print('Hệ số hồi quy:', hshq)

*#Sai số*

ss = linearModel.intercept\_

print('Sai số:', ss)

****

*#Dự đoán value của level trên mô hình hồi quy tuyến tính*

print(linearModel.predict([[24515]]))

****

* Đối với file **Environmental tax account 1999-2021:**

*# ===========================================================*

*#Hồi quy tuyến tính*

*#khởi tạo mô hình hồi quy tuyến tính*

linearModel = lnm.LinearRegression()

*#Biến độc lập*

bdl = dataFrame[['data\_value']]

*#Biến phụ thuộc*

bpt = dataFrame['source']

*#Huấn luyện mô hình trên dữ liệu*

linearModel.fit(bdl, bpt)

*#Hệ số hồi quy*

hshq = linearModel.coef\_

print('Hệ số hồi quy:', hshq)

*#Sai số*

ss = linearModel.intercept\_

print('Sai số:', ss)

****

*#Dự đoán data\_value của source trên mô hình hồi quy tuyến tính*

print(linearModel.predict([[24515]]))

****

**Hình ảnh trực quan hóa**

* Đối với file **Business operations survey 2022: information and communications technology:**

A screenshot of a computer

Description automatically generated

* Đối với file **Environmental tax account 1999-2021:**

A graph with a red line

Description automatically generated