Hàm \_init\_(self): là hàm khởi tạo trong một lớp (class). Hàm này được dùng khi một đối tượng mới được tạo ra trong lớp (class).

Hàm set(): là một kiểu dữ liệu để lưu trữ tập hợp các phần tử. Một số đặc điểm của set,

Không có thứ tự: các phần tử trong set() không có thứ tự cụ thể

Không thể trùng lặp: Set không chứa các phần tử giống nhau

Không thể thay đổi: khi tạo set, bạn không thể thay đổi các phần tử trong nó

Hàm isdigit(): trả về true nếu chuỗi chỉ chứa các chữ số nếu không là false

Hàm isalpha(): trả về true nếu chuỗi có ít nhất 1 ký tự và tất cả là kí tự chữ cái. Nếu không đúng nó sẽ trả về false

Hàm isupper() trả về true nếu chuỗi có chữ viết hoa

Hàm islower() trả về true nếu chuỗi có chữ viết thường

Hàm random.random() để tạo ra một số ngẫu nhiên trong phạm vi từ 0 đến 1

Hàm random.randint(x,y) để tạo ra một số nguyên ngẫu nhiên trong đoạn từ x đến y

Hàm random.uniform(x,y) để tạo ra một số nguyên ngẫu nhiên trong đoạn từ x đến y

**Các hàm trong thư viện Pandas**

1. Hàm series

pd.Series([data, index,type, name, copy,…])

Series là mảng một chiều giống như mảng Numpy, hay như là một cột, một hàng của bảng. nhưng nó bao gồm thêm một bản đánh label. Series có thể được khởi tạo thông qua Numpy, Dic hoặc các kiểu dữ liệu vô hướng bình thường.

1. DataFrame

Pd.DataFrame([data, index, columns, dtype, copy])

Chèn thêm cột vào DataFrame

df[‘ tên cột’]= giá trị

df.insert(index, tên cột, [data]

Nhân bản thêm cột vào DataFrame

df.inset(index, tên cột, [data], allow\_duplicates=True)

Xóa cột bằng lệnh

del df[‘ tên cột’]

df.drop([‘tên cột’],axis=1, inplace= True)

df.pop(‘tên cột)

Chọn cột: df[col]

Chọn dòng theo label: df.loc[label]

Chọn dòng theo vị trí nguyên: df.iloc[loc]

Cắt dòng: df[số:số]

Chọn các dòng theo vector Boolean:df[bool\_vector]

df[‘Tên cột’].str.lstrip(‘data’) để xóa data ở bên trái của dữ liệu trong cột

df[‘Tên cột’].str.rstrip(‘data’) để xóa data ở bên phải của dữ liệu trong cột

df[‘Tên cột’].str.strip(‘data’) để xóa data ở bất kì vị trí của dữ liệu trong cột

df.merge(): để gộp các trường(cột) hay các hàng có chung dữ liệu(kiểu thêm bớt dữ liệu chứ không phải nghĩa tách dữ liệu) xem code Alex The Analyst để hiểu rõ

hàm any() trong python kiểm tra bất kì mục nào trong danh sách