

## **PYTHON CHO PHÂN TÍCH DỮ LIỆU**

### **Bài 1. Các khái niệm cơ bản**

1. Biến và đối tượng
2. Các loại dấu
  - 2.1. Dấu toán học
  - 2.2. Dấu gán
  - 2.3. Dấu so sánh
3. Thư viện
  - 3.1. Truy cập thư viện
  - 3.2. Cài đặt thư viện
4. Một số khái niệm khác
  - 4.1. Ký tự đặc biệt
  - 4.2. Magic commands

### **Bài 2. Vòng lặp và điều kiện**

1. Điều kiện rẽ nhánh
2. Vòng lặp while
3. Vòng lặp for
4. Xử lý lỗi

### **Bài 3. Kiểu dữ liệu (1)**

1. Kiểu dữ liệu số
  - 1.1. Khởi tạo
  - 1.2. Dấu và các hàm toán học
  - 1.3. Một số phương thức và thuộc tính
2. Kiểu dữ liệu Boolean
  - 2.1. Ứng dụng của kiểu dữ liệu Boolean
  - 2.2. Toán tử Boolean
  - 2.3. Chuyển đổi sang Boolean
3. Kiểu dữ liệu None

### **Bài 4. Kiểu dữ liệu (2)**

1. Kiểu dữ liệu chuỗi (String)
  - 1.1. Khởi tạo string
  - 1.2. Các kỹ thuật đặc biệt
  - 1.3. Index và slicing
  - 1.4. Các hàm với string
  - 1.5. Phương thức format
2. Kiểu dữ liệu ngày tháng (Datetime)
  - 2.1. Các kiểu dữ liệu ngày tháng
  - 2.2. Kiểu dữ liệu timedelta
  - 2.3. Định dạng ngày tháng

### **Bài 5. Cấu trúc dữ liệu**

1. List
  - 1.1. Thêm và xóa phần tử

- 1.2. Kỹ thuật slicing
  - 1.3. Các thao tác với list
  - 1.4. Ứng dụng của vòng lặp for với list
  - 1.5. List comprehension
- 2. Tuple
  - 2.1. Khởi tạo tuple
  - 2.2. Kỹ thuật unpacking
- 3. Dictionary
  - 3.1. Khởi tạo dictionary
  - 3.2. Thao tác với key và value
  - 3.3. Phương thức update
  - 3.4. Hàm zip
  - 3.5. Ứng dụng của vòng lặp for với dictionary
  - 3.6. Dictionary comprehension

## **Bài 6. Hàm**

- 1. Built-in function
  - 2. Hàm người dùng định nghĩa
  - 3. Sử dụng hàm làm đối số
- Tổng kết về hàm, phương thức, hằng số và thuộc tính

## **Bài 7. Numpy array**

- 1. Giới thiệu Numpy
  - 1.1. Hằng số
  - 1.2. Array (Mảng nhiều chiều)
  - 1.3. Khái niệm tính toán từng phần tử
  - 1.4. Các phép toán trong Numpy
  - 1.5. Mảng ngẫu nhiên
- 2. Thao tác với Array
  - 2.1. Mô tả array
  - 2.2. Slicing
  - 2.3. Tái cấu trúc array

## **Bài 8. Thống kê mô tả**

- 1. Các đại lượng cơ bản
- 2. Phương sai (Variance)
  - Ý nghĩa của phương sai
- 3. Tứ phân vị
  - Trung vị
- 4. Hệ số tương quan

## **Bài 9. Khám phá dữ liệu**

- 1. Giới thiệu chung
  - 1.1. Series và các kiểu dữ liệu
  - 1.2. Dataframe
  - 1.3. Quy trình phân tích dữ liệu
  - 1.4. Một số cài đặt thông dụng cho Pandas

## 2. Data Exploratory

### 2.1. Overview

Một số thuộc tính của dataframe

Các đại lượng thống kê

Tỉ lệ dữ liệu thiếu

### 2.2. Tìm hiểu chi tiết dữ liệu

Truy cập cột và hàng

Sắp xếp dữ liệu

Lọc dữ liệu

Các giá trị phân biệt (unique) trong cột

Phân nhóm dữ liệu

## **Bài 10. Làm sạch dữ liệu (1)**

### 1. Làm sạch sơ bộ dữ liệu

1.1. Xóa cột và hàng thừa

1.2. Đổi tên cột

1.3. Chuẩn hóa kiểu dữ liệu

1.4. Xử lý giá trị trùng lặp (duplicates)

1.5. Lọc dữ liệu

1.6. Ánh xạ (Mapping)

### 2. Xử lý dữ liệu văn bản

2.1. Xử lý dấu cách và xuống dòng

2.2. Sửa lỗi chính tả

2.3. Xử lý số điện thoại và mã số

2.4. Ghép cột

### 3. Xử lý dữ liệu ngày tháng

Các ký hiệu định dạng ngày tháng

3.1. Thuộc tính ngày tháng gốc

3.2. Thay đổi định dạng ngày tháng (Optional)

3.3. Trích xuất thuộc tính năm, quý, tháng, tuần

3.4. Trích xuất các thuộc tính chu kỳ

3.5. Khoảng thời gian (Timedelta)

## **Bài 11. Làm sạch dữ liệu (2)**

### 1. Xử lý missing data

1.1. Nguyên nhân gây thiếu dữ liệu

1.2. Nhận dạng dữ liệu trống

1.3. Xóa cột

1.4. Xóa dòng

1.5. Điền một giá trị cụ thể

1.6. Nội suy

### 2. Loại bỏ giá trị ngoại biên

2.1. Sử dụng trung bình và độ lệch chuẩn

2.2. Sử dụng khoảng tứ phân vị

## **Bài 12. Biến đổi dữ liệu**

### 1. Pivot và Unpivot

- 1.1. Bảng dài và bảng rộng
- 1.2. Unpivot
- 1.3. Pivot
  - Pivot cơ bản
  - Pivot với một phần dữ liệu
  - Pivot với 2 value
- 2. Liên kết dữ liệu
  - 2.1. Append
  - 2.2. Merge
- 3. Tính toán dữ liệu
  - 3.1. Lũy tiến
  - 3.2. Tính toán trên các cột
  - 3.3. Mức độ tăng trưởng

### **Bài 13. Trực quan hoá dữ liệu**

- 1. Giới thiệu trực quan hoá dữ liệu
  - 1.1. Tại sao cần trực quan hoá dữ liệu
  - 1.2. Các thành phần của biểu đồ
  - 1.3. Bảng màu Matplotlib và Seaborn
  - 1.4. Ba kiểu dữ liệu phổ biến trong Visualization
- 2. Biểu đồ cơ bản
  - 2.1. Pie chart
  - 2.2. Line chart
  - 2.3. Bar chart
  - 2.4. Heatmap
- 3. Biểu đồ chuyên sâu
  - 3.1. Histogram
    - Histogram cơ bản
    - Histogram với 2 chiều thông tin
  - 3.2. Box plot
    - Boxplot đơn với 1 thuộc tính
    - Boxplot với 3 hoặc 4 chiều thông tin
  - 3.3. Scatter plot
    - Scatter plot cơ bản
    - Scatter plot với 2 thuộc tính định lượng
    - Scatter plot với 3 và 4 thuộc tính
    - Swarm plot - biến thể của Scatter plot