**Bài tập Pandas:**

*Ths. Phan Tuấn Anh*

**Bài 1:** Cho 1 dictionary sau:

data = {

    'Tên': ['John', 'Alice', 'Bob', 'Carol', 'David'],

    'Tuổi': [25, 30, 22, 35, 28],

    'Giới tính': ['Nam', 'Nữ', 'Nam', 'Nữ', 'Nam'],

    'Số điện thoại': ['123-456-7890', '987-654-3210', '555-123-4567', '999-888-7777', '333-444-5555'],

    'Điểm danh gia': [8.5, 9.0, 7.2, 8.9, 6.7]

}

1. Tạo DataFrame và in ra thông tin cơ bản của DataFrame
2. Xác định số lượng khách hàng nam và nữ
3. Tính tuổi trung bình của khách hàng
4. Lọc ra những khách hàng có điểm đánh giá cao hơn 8.0
5. Sắp xếp khách hàng theo tuổi tăng dần

**Bài 2:** Cho 1 dictionary sau:

data = {

'Mã đơn hàng': [101, 102, 103, 104, 105],

'Ngày đặt hàng': ['2023-09-01', '2023-09-02', '2023-09-02', '2023-09-03', '2023-09-04'],

'Sản phẩm': ['Áo thun', 'Quần jeans', 'Áo sơ mi', 'Áo khoác', 'Váy đầm'],

'Số lượng': [2, 1, 3, 2, 1],

'Đơn giá': [15.99, 29.99, 19.99, 49.99, 39.99]

}

1. In ra thông tin cơ bản của DataFrame
2. In ra DataFrame với cột mới 'Tổng tiền' (Tổng tiền=Số lượng\* Đơn giá)
3. Tính tổng doanh số bán hàng tổng cộng
4. Tạo biểu đồ cột hiển thị số lượng sản phẩm theo tên sản phẩm
5. Xác định ngày có số lượng đơn hàng cao nhất
6. Lọc ra các đơn hàng có tổng tiền trên 50 đô la

**Bài 3:** Cho data = {

'Học sinh': ['Alice', 'Bob', 'Carol', 'David', 'Eve'],

'Toán': [85, 90, 78, 92, 88],

'Văn': [80, 88, 76, 94, 86],

'Anh': [75, 92, 80, 88, 90] }

1. Thêm cột ĐTB vào DataFrame.
2. Xác định học sinh có ĐTB cao nhất.
3. Sắp xếp học sinh theo thứ tự cao nhất của ĐTB

**Bài 4:** Cho data = {

'Người': ['Alice', 'Bob', 'Alice', 'Carol', 'Bob', 'Alice'],

'Số tiền': [100, 200, 150, 300, 50, 250],

'Loại giao dịch': ['Mua', 'Bán', 'Mua', 'Bán', 'Mua', 'Bán'] }

1. Tính tổng số tiền giao dịch của mỗi người
2. Xác định người có tổng số tiền giao dịch lớn nhất
3. Tạo biểu đồ cột hiển thị phân phối số tiền giao dịch của mỗi người

**Bài 5:** Tạo một DataFrame giả lập với dữ liệu đơn hàng

data = {

'OrderID': range(1, 11),

'CustomerID': np.random.randint(1, 101, 10),

'ProductID': np.random.randint(1, 21, 10),

'Quantity': np.random.randint(1, 11, 10),

'Price': np.random.uniform(10, 100, 10),

'OrderDate': pd.date\_range(start='2022-01-01', periods=10),

'Country': np.random.choice(['USA', 'Canada', 'UK', 'Australia'], 10)

}

1. Tính tổng doanh số bán hàng theo quốc gia
2. Tạo cột mới tính tổng tiền từng đơn hàng
3. Tìm sản phẩm được mua nhiều nhất
4. Tạo một tập dữ liệu con chỉ chứa các đơn hàng có giá trị lớn hơn 500
5. Tính trung bình giá của các sản phẩm theo quốc gia