**Phần 1: Đọc và khám phá dữ liệu**

1. **(1.0đ)** Đọc file dữ liệu diem\_thi\_thpt\_2022.csv vào một DataFrame.
2. **(0.5đ)** Hiển thị 10 dòng đầu tiên và 10 dòng cuối cùng của bảng dữ liệu.
3. **(0.5đ)** Xác định số lượng dòng và cột của dataset.
4. **(1.0đ)** Hiển thị thông tin chi tiết về dataset (kiểu dữ liệu, số lượng giá trị không null).
5. **(0.5đ)** Kiểm tra số lượng giá trị trùng lặp trong dataset dựa trên sbd.

**Phần 2: Xử lý dữ liệu**

1. **(1.5đ)** Kiểm tra và xử lý missing values:
   * Đếm số học sinh bị thiếu điểm môn **Ngữ Văn** và **Ngoại Ngữ**.
   * Điền giá trị thiếu của các môn học bằng **trung vị (median)** của từng môn.
   * Nếu cột **"Khoi"** bị thiếu, thay thế bằng "Chưa xác định".
2. **(1.0đ)** Tạo một cột mới "Diem\_TB" tính điểm trung bình của tất cả các môn thi (không tính NaN).
3. **(1.0đ)** Tạo một cột "XepLoai" dựa trên điểm trung bình:
   * >= 8.0: "Giỏi"
   * >= 6.5: "Khá"
   * >= 5.0: "Trung bình"
   * < 5.0: "Yếu"
4. **(1.0đ)** Đếm số học sinh trong mỗi nhóm "XepLoai".

**Phần 3: Phân tích thống kê**

1. **(1.5đ)** Với môn Toán (toan), tính:
   * Trung bình (mean)
   * Độ lệch chuẩn (std)
   * Khoảng (range) = max - min
   * IQR (Interquartile Range)
2. **(1.0đ)** Tính hệ số tương quan (correlation) giữa điểm **Toán** và **Lý, Hóa**.
3. **(1.0đ)** Liệt kê **10 học sinh có điểm Toán cao nhất** và **10 học sinh có điểm Toán thấp nhất**.
4. **(1.0đ)** Cho biết có bao nhiêu học sinh đạt **điểm 10** môn Toán.
5. **(1.5đ)** Xác định **outliers** của môn Toán bằng **IQR method** và liệt kê danh sách học sinh có điểm Toán nằm ngoài khoảng [Q1 - 1.5\*IQR, Q3 + 1.5\*IQR].

**Phần 4: Phân tích theo nhóm**

1. **(1.5đ)** Dựa vào cột Khoi, thực hiện:
   * Đếm số lượng học sinh thi khối "KHTN" và "KHXH".
   * Tạo bảng tổng hợp điểm trung bình từng môn theo từng khối.
2. **(1.5đ)** Tính điểm trung bình theo từng tỉnh thành (tentinh).
3. **(1.5đ)** Liệt kê 5 tỉnh có **điểm Toán trung bình cao nhất** và 5 tỉnh có **điểm Toán trung bình thấp nhất**.

**Phần 5: Trực quan hóa dữ liệu**

1. **(1.5đ)** Vẽ **biểu đồ Pie** thể hiện tỷ lệ học sinh thi các khối "KHTN", "KHXH", "Chưa xác định".
2. **(1.5đ)** Vẽ **biểu đồ Bar** thể hiện số lượng học sinh ở các tỉnh **Hà Nội, Hồ Chí Minh, Đà Nẵng**.
3. **(1.5đ)** Vẽ **biểu đồ Boxplot** để quan sát phân phối điểm Toán và phát hiện outliers.
4. **(1.5đ)** Vẽ **biểu đồ Histogram** thể hiện phân phối điểm Toán.