**Phần 1: Khám phá dữ liệu**

1. **(1.0đ)** Đọc file CSV vào một DataFrame và hiển thị thông tin tổng quan (info(), describe()).
2. **(1.0đ)** Kiểm tra xem có **dòng trùng lặp** trên sbd không. Nếu có, hãy loại bỏ.
3. **(1.0đ)** Kiểm tra và đếm số lượng giá trị NaN trong mỗi cột.
4. **(1.0đ)** Điền giá trị NaN của các môn học bằng **trung vị (median)** của từng môn.
5. **(1.0đ)** Tạo một cột "Diem\_TB" là điểm trung bình của tất cả các môn học.

**Phần 2: Phân tích thống kê**

1. **(1.0đ)** Tìm học sinh có **tổng điểm cao nhất và thấp nhất**.
2. **(1.0đ)** Đếm số lượng học sinh đạt **điểm 10 môn Toán**.
3. **(1.0đ)** Xác định **5 tỉnh có số lượng học sinh tham gia thi nhiều nhất**.
4. **(1.0đ)** Tính **trung bình điểm Toán** theo từng tỉnh (tentinh).
5. **(1.0đ)** Tính hệ số tương quan giữa **Toán và các môn KHTN (Lý, Hóa, Sinh)**.

**Phần 3: Phát hiện outliers**

1. **(1.5đ)** Dùng phương pháp **IQR** để phát hiện outliers của môn Toán:

* Tính **Q1, Q3, IQR**.
* Xác định **Lower Bound (Q1 - 1.5\*IQR)** và **Upper Bound (Q3 + 1.5\*IQR)**.
* Lọc danh sách học sinh có điểm Toán nằm ngoài khoảng [Q1 - 1.5\*IQR, Q3 + 1.5\*IQR].

1. **(1.5đ)** Vẽ **Boxplot** cho điểm Toán để trực quan hóa outliers.
2. **(1.5đ)** Tạo một cột "Outlier\_Toan" với giá trị:

* True nếu điểm Toán là outlier.
* False nếu không.

**Phần 4: Phân tích nhóm dữ liệu**

1. **(1.5đ)** So sánh phân phối điểm Toán giữa các khối "KHTN" và "KHXH" bằng Boxplot.
2. **(1.5đ)** Đếm số lượng học sinh theo từng khối (Khoi).
3. **(1.5đ)** Xác định khối nào có điểm Toán trung bình cao hơn.
4. **(1.5đ)** Xác định **5 tỉnh có điểm trung bình môn Toán cao nhất và thấp nhất**.

**Phần 5: Trực quan hóa dữ liệu**

1. **(1.5đ)** Vẽ **Histogram** thể hiện phân phối điểm trung bình (Diem\_TB).
2. **(1.5đ)** Vẽ **Bar chart** thể hiện số lượng học sinh theo từng tỉnh (tentinh).
3. **(1.5đ)** Vẽ **Pie chart** thể hiện tỷ lệ học sinh theo từng khối (Khoi).