

| **TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP**  **THÀNH PHỐ HỒ CHÍ MINH**    **KHOA CÔNG NGHỆ THÔNG TIN** |  |  |
| --- | --- | --- |

Họ tên: ……………………………Mã SV:…………………Số máy:………

Sinh viên nộp bài một file duy nhất ***HoTen\_MaSV.\****

Cho *dataset* điểm **PISA** năm 2015 của học sinh Việt Nam tham gia kỳ thi đánh giá năng lực **Toán học, Đọc hiểu và Khoa học**. Anh/chị hãy thực hiện các yêu cầu sau:

1. ***(3đ)*** Đọc file dữ liệu ***PISA\_VN\_2015.csv*** và thực hiện:
   1. **(0.25đ)** Cho biết trong bảng dữ liệu gồm những cột nào
   2. **(0.75đ)** Hiển thị 5 dòng đầu tiên và 5 dòng cuối cùng của bảng dữ liệu.
   3. **(0.75đ)** Cho biết bảng dữ liệu có bao nhiêu dòng và bao nhiêu cột?
   4. **(0.75đ)** Cho biết thông tin về các cột dữ liệu và kiểu của từng cột trong bảng
   5. **(0.5đ)** Cho biết số lượng thông tin bị thiếu của cột **JOYSCIE** trong bảng
2. ***(2.25đ)*** Với cột điểm thi **PV1READ,** hãy tính các trị thống kê:
3. **(0.75đ)** Trung binh (*mean*),
4. **(0.75đ)** Độ lệch chuẩn (*standard deviation*)
5. **(0.75đ)** Và khoảng dữ liệu (*range*)
6. ***(1.5đ)*** Dựa vào dữ liệu cột **Gender**, anh/chị hãy thực hiện:
   1. **(0.75đ)** Đếm số lượng học sinh Nam (*Boys*) và học sinh Nữ (*Girls*)
   2. **(0.75đ)** Vẽ đồ thị **Pie** để so sánh mối tương quan giữa Boy và Girl.
7. ***(1.75đ)***
8. **(0.75đ)** Liệt kê 10 học sinh có điểm Toán (**pv1MATH**) cao nhất
9. **(0.75đ)** Liệt kê 10 học sinh có điểm Khoa học (**pv1SCIE**) thấp nhất.
10. **(0.25đ)** Cho biết số tuổi trung bình của 10 học sinh có điểm khoa học (**pv1SCIE)** thấp nhất
11. ***(1.5đ)***
12. **(0.75đ)** Thống kê để hoàn thành bảng số liệu dưới đây
13. **(0.75đ)** Từ đó trực quan hóa dữ liệu **Bảng 1** bằng đồ thị **Bar**.

**Bảng 1**

| **Vùng (REGION)** | **Số học sinh** |
| --- | --- |
| SOUTH |  |
| CENTRAL |  |
| NORTH |  |

------------------------- Hết ------------------------------

*Sinh viên* ***tuyệt đối*** *không trao đổi bài, các bài giống nhau bị điểm 0 Thực hành*