TRƯỜNG ĐẠI HỌC VĂN LANG

**KHOA: KỸ THUẬT CƠ – ĐIỆN VÀ MÁY TÍNH**

**ĐỀ THI KẾT THÚC HỌC PHẦN**

**Học kỳ 3, năm học 2021 - 2022**

**Mã học phần:** DDS0170

**Tên học phần:** Phân tích dữ liệu trực quan

**Mã nhóm lớp học phần:** 213\_DDS0170\_01

**Thời gian làm bài:** 10 ngày

**Hình thức thi:** Bài tập lớn

**Cách thức nộp bài:** Sinh viên Upload bài làm (Word/pdf)

**Đề bài**

**Bài tập 1:** Thực hiện các câu hỏi bên dưới về nội dung Thay đổi khí hậu.

Sinh viên sử dụng các thư viện sau:

1. Lượng khí thải cacbon năm vừa rồi lớn hơn năm đầu tiên bao nhiêu lần?
2. Nhiệt độ đã tăng lên bao nhiêu qua các năm? So sánh giữa những năm gần đây và những năm cũ (nhất).
3. Vẽ time series plot của lượng phát thải carbon từ tập dữ liệu *temp\_carbon*. Trục y là tấn khí thải carbon của mỗi năm.
4. Vẽ line plot về nồng độ CO2 qua các nắm. Tô màu bằng cách sử dụng biến .

**Bài tập 2:** Thực hiện các câu hỏi bên dưới về nội dung Properties of Stars.

Sinh viên sử dụng các thư viện sau:

1. mean magnitude bằng bao nhiêu?
2. Độ lệch chuẩn của magnitude bằng bao nhiêu?
3. Vẽ density plot của magnitude.
4. Có bao nhiêu peaks trong dữ liệu?
5. Vẽ phân bố (distribution) của star temperature.
6. Statement nào biểu diễn tốt nhất đặc tính của phân bố temperature (temperature distribution)?
7. Vẽ scatter plot của dữ liệu với temperature trên trục và magnitude trên trục .

**Bài tập 3:** Thực hiện các câu hỏi bên dưới về nội dung Titanis Survival.

Sinh viên định nghĩa thư viện sử dụng code bên dưới:

1. Vẽ density plot của age và được tô màu bởi tình trạng survival (survival status).
2. Nhóm nào có tỉ lệ sống nhiều hơn chết?
3. Lọc (filter) dữ liệu để loại bỏ người không trả tiền vé.
4. Vẽ boxplot của giá vé (fare) dựa trên thông tin về sống sót (survival status).
5. Biến đổi (transform) bằng log2 của giá vé (fares).
6. Biến Pclass dùng phân loại tầng lớp hành khách. Hãy vẽ 3 biểu đồ barplot.

- Vẽ barplot cơ bản về lớp hành khách được tô màu bởi tình trạng sống sót (filled by survival).

- Vẽ barplot như trên nhưng sử dụng đối số position = position\_fill() để mô tả mối liên hệ giữa mỗi nhóm thay vì đếm (count).

- Vẽ barplot của người sống sót được tô màu bởi lớp hành khách sử dụng position = position\_fill().

**Bài tập 4:** (Đề mở)

Dùng thuật toán phù hợp để crawl dữ liệu về kết quả kỳ thi tốt nghiệp THPT năm 2022. Sử dụng package để trực quan hoá dữ liệu và phân tích dữ liệu của kỳ thi năm nay (Lưu ý: nên phân tích theo hướng dữ liệu thu thập được đã truyền tải những nội dung gì quan trọng/hấp dẫn đến người đọc).

**Hướng dẫn sinh viên trình bày:**

Bản báo cáo (Word/PDF), sinh viên trình bày trang bìa, bài tập và bài làm. Lưu ý: Mỗi bài tập mới phải trình bày qua trang mới.

**Tiêu chí chấm điểm và thang điểm (Rubric):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tiêu chí** | **Trọng số (%)** | **Tốt**  **100%** | **Khá**  **75%** | **Trung bình**  **50%** | **Kém**  **0%** |
| Bài tập lớn | 100 | * Nộp đúng hạn; * Các bước thực hiện đầy đủ, kết quả chính xác; * Trình bày rõ ràng, có thứ tự; * Chèn công thức toán học vào các biểu thức, phép tính. | * Nộp đúng hạn; * Các bước thực hiện khá đầy đủ, kết quả chính xác; * Trình bày rõ ràng, chưa có thứ tự; * Một số biểu thức, phép tính chưa chèn công thức toán học. | * Trễ hạn không quá 3 ngày; * Các bước thực hiện tương đối đầy đủ, một số kết quả chưa đúng; * Trình bày khá lung tung nhưng còn theo dõi được; * Nhiều biểu thức, phép tính chưa chèn công thức toán học.. | * Không nộp/ trễ hạn hơn 4 ngày; * Các bước thực hiện không đầy đủ, kết quả sai; * Trình bày lung tung, rối mắt; * Không chèn công thức toán học vào các biểu thức, phép tính. |

*Ngày biên soạn: 15/07/2022*

**Giảng viên biên soạn đề thi:** Trương Quốc Trí

*Ngày kiểm duyệt:*

**Trưởng (Phó) Khoa/Bộ môn kiểm duyệt đề thi:** Trương Quốc Trí

Sau khi kiểm duyệt đề thi, **Trưởng (Phó) Khoa/Bộ môn** gửi về Trung tâm Khảo thí qua email:[khaothivanlang@gmail.com](mailto:khaothivanlang@gmail.com)bao gồmfile word và file pdf.