

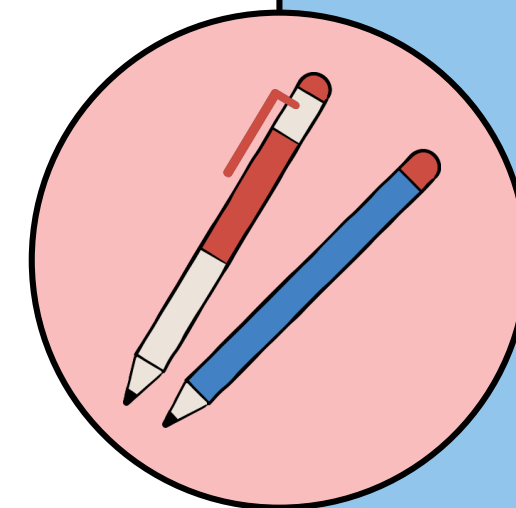
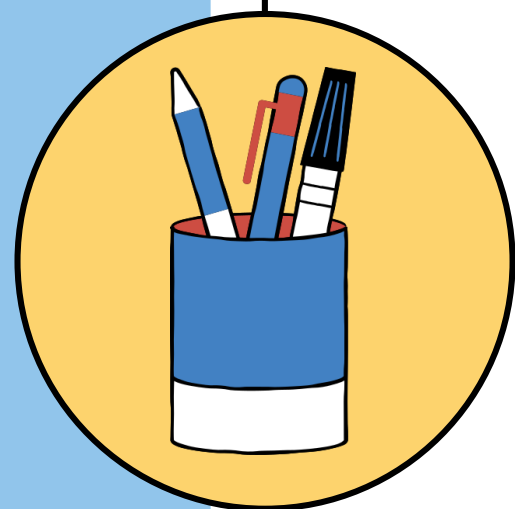
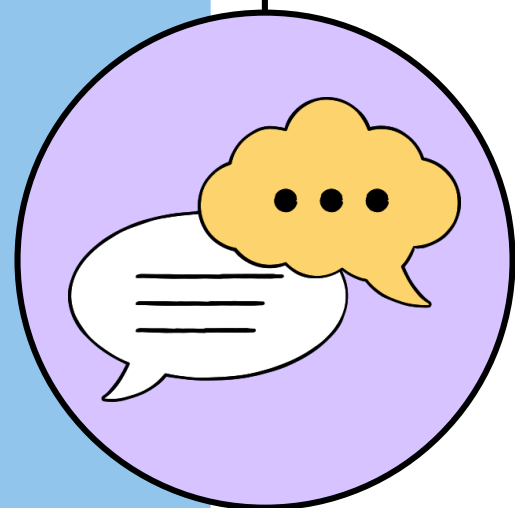


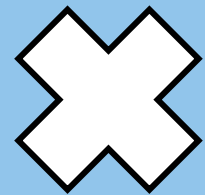
LẬP TRÌNH R CHO PHÂN TÍCH



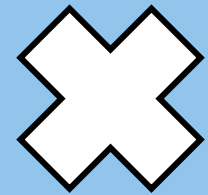
ĐỒ ÁN KẾT THÚC MÔN HỌC

Hướng dẫn nhanh



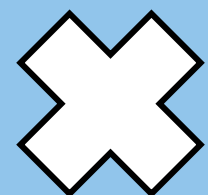
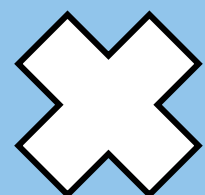
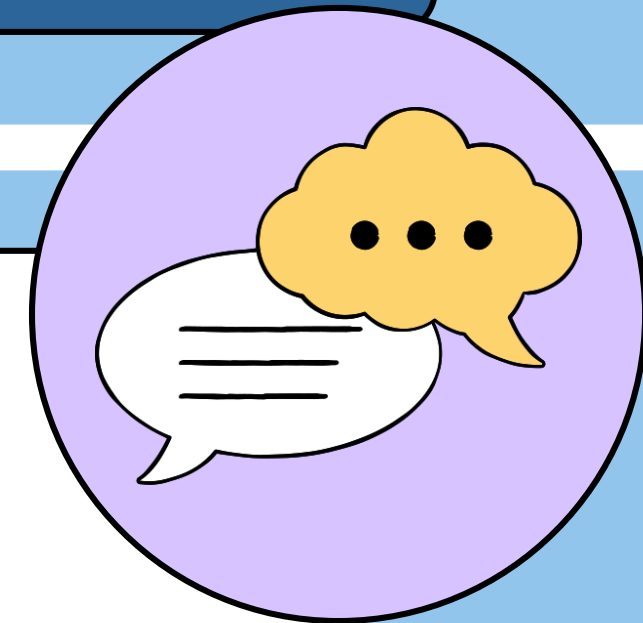


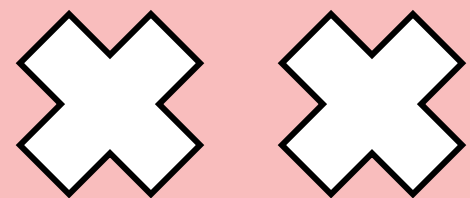
DANH SÁCH ĐỀ TÀI



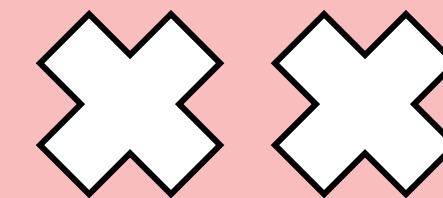
Một số đề tài gợi ý (bạn có thể chọn đề tài khác ngoài mục gợi ý này)

1. *Thực hiện bài toán phân tích dữ liệu với ngôn ngữ R tập dữ liệu liên quan: dữ liệu covid, dữ liệu bán hàng, dữ liệu cho thuê nhà.*
2. Machine learning validation.
3. Kiểm định giả thuyết thống kê
4. Hồi quy tuyến tính đơn biến, đa biến, logarit
5. Sentiment Analysis





Sử dụng Kaggle để chọn đề tài ngoài gợi ý



The screenshot shows a list of Kaggle items with the following details:

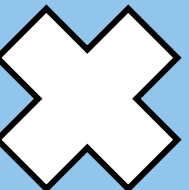
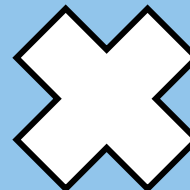
Item Title	Type	Author	Age	Comments
Analysis in R Notebook	Notebook	Zeeshan-ul-hassan Usmani	7y ago	23
HealthyR: R for health data analysis	Discussion Topic	Raju Kumar Mishra	5y ago	11
Customer Churn analysis & prediction, XGBoost R	Notebook	Kheirallah Samaha	3y ago	45
Reddit Vaccine Myths EDA and Text Analysis R	Notebook	Kheirallah Samaha	3y ago	97
Uber Data Analysis in R	Notebook	Prakhar Rathi	4y ago	48

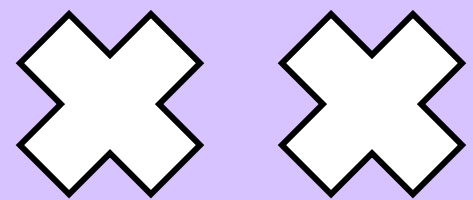




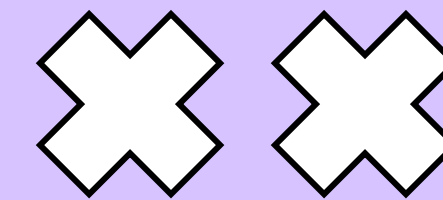
LƯU Ý!

Sinh viên chia thành các nhóm, chọn 1 đề tài và không quá 3 nhóm làm cùng 1 đề tài.





Cấu trúc bài đồ án



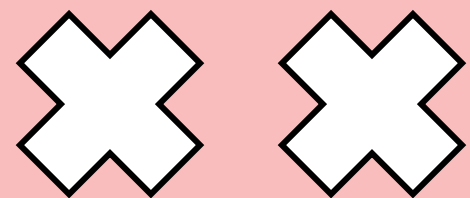
Nội dung: Xây dựng một phân tích dữ liệu sử dụng các kỹ thuật được học trong môn học như trực quan hóa dữ liệu, mô hình hóa dữ liệu. Sinh viên thực hiện đồ án trên Rmarkdown.

Lưu ý: Tập dữ liệu phải là tập kiểm chứng, có chất lượng (Các nguồn đã giới thiệu ở đầu học kỳ).

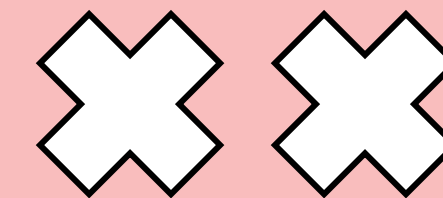
Hình thức: Sinh viên thực hiện theo nhóm (tối đa 4 SV).

Gợi ý cấu trúc của bài báo cáo

1. Tóm tắt (abstract)
2. Giới thiệu (introduction)
3. Dữ liệu (data)
4. Trực quan hóa dữ liệu (data visualization)
5. Mô hình hóa dữ liệu (data modeling)
6. Thực nghiệm, kết quả, và thảo luận
7. Kết luận (conclusions)
8. Phụ lục (appendices)
9. Đóng góp (contributions)
10. Tham khảo (references)
11. Peer assessment (Phần này nhóm sinh viên họp thảo luận đưa ra nhận xét về nội dung nhóm mình triển khai được; đánh giá mức hoàn thành, ưu, nhược điểm hạn chế của từng thành viên trong nhóm).



Tiêu chí đánh giá



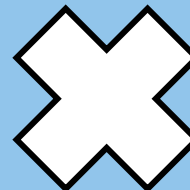
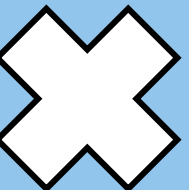
STT	Tiêu chí đánh giá	Mô tả mức chất lượng	
		A	B
		9-10	8-7
1	Hình thức báo cáo	<ul style="list-style-type: none">- Đầy đủ các phần theo được yêu cầu của cấu trúc đồ án.- Diễn đạt mạch lạc.	<ul style="list-style-type: none">- Đầy đủ các phần theo được yêu cầu của cấu trúc đồ án.- Đạt được yêu cầu chuẩn đầu ra môn học.
2	Kỹ năng làm việc nhóm	<ul style="list-style-type: none">- Mô tả đầy đủ nhiệm vụ đã phân chia và kết quả hoàn thành đúng theo tiến độ.- Tự đánh giá chất lượng bài báo cáo sát với năng lực.	<ul style="list-style-type: none">- Mô tả đầy đủ nhiệm vụ đã phân chia và kết quả hoàn thành đúng theo tiến độ.
3	Đánh giá chất lượng, mô tả về dữ liệu và xác định bài toán cần giải quyết	<ul style="list-style-type: none">- Hiểu và mô tả đầy đủ về tập dữ liệu, tập dữ thực tế và có chất lượng.- Đưa ra yêu cầu bài toán có chất lượng, thực tế, có giá trị khoa học.	<ul style="list-style-type: none">- Hiểu và mô tả đầy đủ về tập dữ liệu, Đưa ra yêu cầu bài toán có chất lượng.
4	Có khả năng sử dụng R cho các yêu cầu cơ bản, ứng dụng xác xuất thống kê	<ul style="list-style-type: none">- Sử dụng hợp lý, có hiệu quả các kỹ năng phân tích dữ liệu cơ bản trên tập dữ liệu.	<ul style="list-style-type: none">- Sử dụng các kỹ thuật phân tích dữ liệu đạt được yêu cầu bài toán trên dữ liệu, sử dụng hiệu quả các kiểu dữ liệu nâng cao: danh sách, ma trận, dataframe; sử dụng hàm,...
5	sử dụng các dạng đồ thị để phân tích dữ liệu	<ul style="list-style-type: none">- Sử dụng đa dạng các loại biểu đồ trên tập dữ liệu.- Mô tả được ý nghĩa phân tích sử dụng các biểu đồ giúp hiểu rõ tập dữ liệu và phát hiện tri thức mới.	<ul style="list-style-type: none">- Sử dụng đa dạng các loại biểu đồ trên tập dữ liệu.- Mô tả được ý nghĩa, phân tích cơ bản.
6	Sử dụng mô hình dữ liệu (time-seris, Hồi quy,...)	<ul style="list-style-type: none">- Sử dụng từ 03 mô hình phân tích dữ liệu trở lên.- Sử dụng hợp lý, có hiệu quả các kỹ năng phân tích dữ liệu cơ bản trên tập dữ liệu.	<ul style="list-style-type: none">- Sử dụng từ 02 mô hình phân tích dữ liệu trở lên, giải thích rõ ý nghĩa phân tích.- Giải thích rõ ý nghĩa phân tích trên tập dữ liệu.

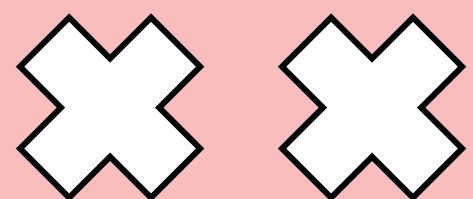




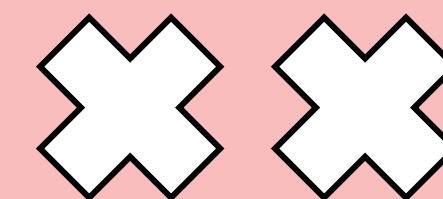
LƯU Ý!

CÁCH VÀ HẠN NỘP BÀI

- Cấu trúc: 01 file mã nguồn R và một file tóm tắt tổng kết báo cáo (.ppt).
 - Thời gian nộp: Trước ngày 15 tháng 05 năm 2025
- 
- 



Tóm lại:



Các nhóm chọn cho mình 1 đề tài (không quá 3 nhóm chọn cùng 1 đề tài).

Lưu ý về cấu trúc bài làm và các tiêu chí đánh giá để đạt điểm cao.

Lưu ý về cách thức nộp bài và nộp bài đúng hạn.



CHÚC CÁC BẠN LÀM BÀI THẬT TỐT!