

MẪU BÁO CÁO ĐÁNH GIÁ SƠ BỘ - MILESTONE 3

Capstone Project: Phân tích dữ liệu chuỗi thời gian

Mục đích: Sinh viên tự đối chiếu yêu cầu Milestone 3 và điền minh chứng trước buổi vấn đáp. Giảng viên sử dụng cùng biểu mẫu để đánh giá nhanh và đặt câu hỏi.

Hướng dẫn điền: (1) Điền ngắn gọn, ưu tiên chỉ rõ số trang/mục/Hình-Bảng trong báo cáo và section/cell trong notebook. (2) Nếu “Chưa thực hiện”, ghi lý do và kế hoạch bổ sung (nếu có).

1. Thông tin nhóm & phạm vi phân tích

Mã nhóm / Mã đề án	04
Thành viên (Họ tên - MSSV)	Trần Đình Sang - 24C11060 Phạm Phong Hào - 24C12006
Biến mục tiêu (target) & đơn vị đo	Doanh thu, đơn vị Bảng Anh
Tần suất chuỗi (daily/weekly/monthly/...) & lý do chọn	Chuỗi thời gian có tần suất theo ngày, do dữ liệu giao dịch được ghi nhận liên tục theo từng ngày và được chuẩn hóa về trục thời gian ngày trong quá trình tiền xử lý
Khoảng thời gian quan sát (từ - đến)	2010-12-01 đến 2011-12-09
Công cụ/chồng công nghệ (Python libs, notebook, ...)	Jupyter lab/notebook, Python, Numpy, Pandas, Seaborn, Matplotlib, statsmodels, prophet
Link nộp bài / tên file: báo cáo - slide - notebook	https://drive.google.com/drive/folders/1Geb1TOChDHtS45UEFd7NOngPhLwGeQpE?usp=sharing

2. Quy ước đánh giá nhanh

- Trạng thái: Done (hoàn tất) | Partial (chưa đầy đủ) | Not (chưa thực hiện)
- % hoàn thành: 0 - 100 (tự ước lượng theo tiêu chí tối thiểu của từng hạng mục)
- Tham chiếu báo cáo: ghi rõ Mục/Section + Trang/Page + Hình/Bảng (nếu có). Ví dụ: “3.2, tr. 8, Hình 5”.
- Tham chiếu notebook: ghi rõ file + section/heading hoặc cell number/tag. Ví dụ: “notebook.ipynb - Section 2.3 - Cell #12”.

3. Check-list bắt buộc Milestone 3 (Sinh viên tự điền trước vấn đáp)

3.1. Yêu cầu 1 - Phân tích khám phá dữ liệu (EDA)

ID	Công việc bắt buộc	Tiêu chí tối thiểu	Kết quả/Minh chứng (tóm tắt)	Trạng thái	%	Tham chiếu trong báo cáo	Tham chiếu trong notebook	Ghi chú
EDA-1	Trực quan hóa chuỗi thời gian	Có biểu đồ theo thời gian (line chart) cho biến mục tiêu (có ghi tần suất)	Biểu diễn bằng linechart, với trục Ox là thời gian, Trục Oy là biến mục tiêu. Chủ yếu biểu diễn các mẫu quan sát để thấy được chủ quan trend và seasonality.	Done	100	5.1, tr. 13, Hình 5.1 5.2, tr. 14, Hình 5.2 5.2, tr. 15, Hình 5.3 - 5.4 - 5.5 6.2.2, tr. 19, Hình 6.3 6.2.2, tr. 22, Hình 6.5 6.2.2, tr. 23, Hình 6.7 6.3.2, tr. 25, Hình 6.12 6.3.2, tr. 26, Hình 6.14 ...	dam-final.ipynb/Section 3/Cell 4, Cell 5 code/problem-02/CustomerAndRevenue.ipynb, mục Phân tích insights (5 mục con), Cell 16, Section Phân tích nhóm mua lẻ	
EDA-2	Mô tả đặc điểm chính	Nhận định về: xu hướng (trend), mùa vụ (seasonality), nhiễu/biến động (noise)	Thực hiện phân rã STL, phân tích từng phân khúc khách hàng, phân tích để thấy được đặc điểm của doanh thu ở quy mô chung toàn tập	Done	100	6.2.2, tr. 21, Hình 6.4 6.2.2, tr. 22, Hình 6.6 6.2.2, tr. 23, Hình 6.8 6.3.2, tr. 26, Hình 6.13 6.3.2, tr. 27, Hình 6.15 ...	dam-final.ipynb/Section 4/Cell 6 code/problem-02/CustomerAndRevenue.ipynb, mục Phân tích insights (5 mục con), Cell 18, Section Phân	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.

							tích nhóm mua lẻ Cell 27, Section Phân tích nhóm mua sỉ	
EDA-3	Nhận định sơ bộ khả năng dự báo	Nêu lý do chuỗi có/không phù hợp dự báo; lưu ý hạn chế	Dữ liệu biến động rất lớn ở giai đoạn cuối chuỗi, ví dụ mục phân tích bán lẻ, bán sỉ. Vì vậy phương sai trước dự báo biến động lớn nên khả năng dự báo sẽ có sai số	Done	100	6.3.3, tr 27-28	dam-final.ipynb/Section 3/Cell 4, Cell 5, Cell 6 code/problem-02/CustomerAndRevenue.ipynb, mục Phân tích insights (5 mục con)	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.

3.2. Yêu cầu 2 - Tiền xử lý dữ liệu chuỗi thời gian

ID	Công việc bắt buộc	Tiêu chí tối thiểu	Kết quả/Minh chứng (tóm tắt)	Trạng thái	%	Tham chiếu trong báo cáo	Tham chiếu trong notebook	Ghi chú
PRE-1	Chuẩn hóa trực thời gian	Xác định tần suất chuỗi và khoảng thời gian quan sát; giải thích tác động của tần suất đến dự báo	Cách khoảng cách được xác định theo daily, weekly, monthly.	Done	100	Chương 5-6, từ tr 13 đến tr 27. 5.2, tr 14, Hình 5.3 6.2.2, tr. 21, Hình 6.4 6.2.2, tr. 22, Hình 6.6	dam-final.ipynb/Section 2/Cell 3 code/problem-02/CustomerAndRevenue.ipynb, mục Phân tích	

						6.2.2, tr. 23, Hình 6.8 6.3.2, tr. 26, Hình 6.13 6.3.2, tr. 27, Hình 6.15	insights (5 mục con),	
PRE-2	Kiểm tra missing timestamps	Kiểm tra dữ liệu có đều tần suất; phát hiện mốc thời gian bị thiếu (nếu có)	Đã kiểm tra bằng cách in shape của df, trước và sau khi điền	Done	100	Mục 3.2, tr 8-9 (chỉ note giải thuật, không note thống kê)	dam-final.ipynb/Section 2/ Cell 3	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.
PRE-3	Xử lý giá trị bị thiếu (missing values)	Nhận diện loại missing và lập luận phương pháp xử lý (ffill/bfill/inte rpolation/seas onal mean/giữ missing...)	Đã điền thiếu các time_index bị thiếu, ví dụ ở mục phân tích phân khúc khách hàng đã viết hàm để xử lý việc điền thiếu bằng hàm trung bình trượt	Done	100	Mục 3.2, tr 8-9	dam-final.ipynb/Section 2/ Cell 3 code/problem -02/Customer AndRevenue. ipynb, mục Phân tích insights (5 mục con), Cell 13, Section Phân tích nhóm mua lẻ	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.
PRE-4	Phát hiện & xử lý outliers theo thời gian	Phát hiện outlier (trực quan/thống kê); giải thích là lỗi hay sự kiện; quyết định giữ/điều chỉnh/loại bỏ		Done	100	[Mục/Trang/ Hình]	dam-final.ipynb/Section 8/ Cell 10	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.

PRE-5	Biến đổi dữ liệu (transformations)	Cân nhắc log/Box-Cox/power...; nêu mục tiêu (ổn định phương sai, giảm biên độ mùa vụ...) và ảnh hưởng đến mô hình	Đã biến đổi theo hàm log. Mục đích phân tích tỷ trọng doanh thu/khách hàng để quan sát “đuôi dài” trên histogram	Done	100	6.2.1, tr. 20, Hình 6.1	dam-final.ipynb/Section 11/ Cell 16	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.
PRE-6	Kiểm tra tính dừng (stationarity) & chuẩn bị mô hình	Kiểm tra có xu hướng/mùa vụ; nêu cách chuẩn bị (differencing/seasonal differencing/..) nếu cần	Phương pháp đưa về bài toán phân rã STL. Kết quả đã xác định tất cả các chuỗi thời gian phân tích trong nghiên cứu đều có tính dừng	Done	100	6.2.2, tr. 21, Hình 6.4 6.2.2, tr. 22, Hình 6.6 6.2.2, tr. 23, Hình 6.8 6.3.2, tr. 26, Hình 6.13 6.3.2, tr. 27, Hình 6.15	dam-final.ipynb/Section 11/ Cell 15, 16, 17	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.

3.3. Yêu cầu 3 - Phân rã chuỗi thời gian

ID	Công việc bắt buộc	Tiêu chí tối thiểu	Kết quả/Minh chứng (tóm tắt)	Trạng thái	%	Tham chiếu trong báo cáo	Tham chiếu trong notebook	Ghi chú
DEC-1	Phân rã chuỗi thời gian	Tách Trend/Seasonality/Residual (ví dụ STL hoặc phương pháp phù hợp)	Chuỗi doanh thu theo ngày được phân rã bằng phương pháp STL. Chu kỳ mùa vụ được xác định là 7 ngày	Done	100	6.2.2, tr. 21, Hình 6.4 6.2.2, tr. 22, Hình 6.6 6.2.2, tr. 23, Hình 6.8 6.3.2, tr. 26, Hình 6.13 6.3.2, tr. 27, Hình 6.15	dam-final.ipynb/Section 4/ Cell 6 code/problem-02/ CustomerAndRevenue.ipynb, mục Phân tích insights (5 mục con)	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.

						5.2, tr. 14, Hình 5.2 5.2, tr. 14, Hình 5.3 5.2, tr. 15, Hình 5.4-5.5 ...		
DEC-2	Phân tích ý nghĩa các thành phần	Diễn giải ý nghĩa của Trend, Seasonality, Residual theo ngữ cảnh kinh doanh	Dựa đồ phân rã, đã trực quan thành 3 đồ thị ứng với Trend, Seasonality, Residual nhìn nhận và phát hiện insights, phát biểu insights dưới góc nhìn của business	Done	100	Mục 7.2, tr 34-36	dam-final.ipynb/Section 4/ Cell 6 code/problem-02/ Customer AndRevenue.ipynb, mục Phân tích insights (5 mục con)	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.
DEC-3	Liên hệ phân rã với lựa chọn mô hình	Giải thích vì sao kết quả phân rã hỗ trợ lựa chọn mô hình dự báo (SARIMA/ETS/Prophet/...)	Dựa vào biên độ dao động vào cuối chuỗi, cho thấy trend tăng, Mô hình SARIMA không đáp ứng được, nhưng Prophet thì đáp ứng được	Done	100	6.2.2, tr. 21, Hình 6.4 6.2.2, tr. 22, Hình 6.6 6.2.2, tr. 23, Hình 6.8 6.3.2, tr. 26, Hình 6.13 6.3.2, tr. 27, Hình 6.15	code/problem-02/ Customer AndRevenue.ipynb, mục Phân tích insights (5 mục con)	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.

3.4. Yêu cầu 4 - Xây dựng và so sánh mô hình dự báo

ID	Công việc bắt buộc	Tiêu chí tối thiểu	Kết quả/Minh chứng (tóm tắt)	Trạng thái	%	Tham chiếu trong báo cáo	Tham chiếu trong notebook	Ghi chú
MOD-1	Huấn luyện ít nhất 2 biến thể mô hình	Có tối thiểu 2 biến thể (vd: SARIMA cấu	Huấn luyện hai mô hình dự báo gồm SARIMA và	Done	100	7.1, tr 31-34	dam-log-model.ipynb/Section 3/ Cell 6/	

		hình khác nhau; ETS giả định khác nhau; Prophet có/không holiday...)	Prophet trên cùng tập dữ liệu. Các mô hình được cấu hình phù hợp với đặc điểm xu hướng và mùa vụ của chuỗi để phục vụ so sánh hiệu quả dự báo.			6.3.3, tr 27-30	Mục 3: Define 2 SARIMA variants code/problem -02/ Customer AndRevenue.ipynb, mục Phân tích insights (5 mục con) Cell 22-23, Section Phân tích nhóm mua lẻ Cell 29-30, Section Phân tích nhóm mua sỉ	
MOD-2	Thiết kế train-test / backtesting	Mô tả cách chia dữ liệu theo thời gian (time-based split) hoặc rolling validation (nếu có)	Trong bài toán dự báo 30 ngày tiếp theo, nghiên cứu đã chia tập data ban đầu theo tỷ lệ train: test = 11:1, sau đó dùng mô hình dự báo cho 30 ngày tiếp theo	Done	100	6.3.3, tr 27-30	dam-log-model.ipynb/Section 3/ Cell 6/ Mục 2: Train/Test split (time-based) code/problem -02/ Customer AndRevenue.ipynb, mục Phân tích insights (5 mục con) Cell 22-23, Section Phân	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.

							tích nhóm mua lẻ Cell 29-30, Section Phân tích nhóm mua sỉ	
MOD-3	So sánh mô hình theo độ chính xác	So sánh bằng metric phù hợp (vd MAE/RMSE/MAPE) + đồ thị dự báo vs thực tế	Các mô hình được so sánh bằng MAE, RMSE kết hợp với biểu đồ dự báo so với giá trị thực tế. Kết quả: Prophet có sai số thấp hơn hoặc tương đương SARIMA và theo sát xu hướng chuỗi tốt hơn ở giai đoạn cuối.	Done	100	[Mục/Trang/ Hình/Bảng]	dam-log-model.ipynb/Section 3/ Cell 6/ Mục 4: Fit + Forecast + Evaluate	Dam-final.ipynb, có code nhưng không hiển thị trong báo cáo.
MOD-4	So sánh theo khả năng diễn giải & mục tiêu kinh doanh	Bàn về interpretability và sự phù hợp với mục tiêu bài toán (trade-off)	Sau khi xem xét về khả năng mô phỏng lại trend của kết quả mô hình tạo ra, nhóm quyết định chọn Prophet vì kết quả prophet cho thấy rõ khả năng tạo trend của mô hình trên kết quả.	Done	100	7.2.3, tr 36 6.3.3, tr 27-30	code/problem-02/Customer AndRevenue.ipynb, mục Phân tích insights (5 mục con) Cell 22-23, Section Phân tích nhóm mua lẻ Cell 29-30, Section Phân tích nhóm mua sỉ	

3.5. Yêu cầu 4 (tiếp) - Phân tích kết quả, hạn chế & hướng phát triển

ID	Công việc bắt buộc	Tiêu chí tối thiểu	Kết quả/Minh chứng (tóm tắt)	Trạng thái	%	Tham chiếu trong báo cáo	Tham chiếu trong notebook	Ghi chú
RES-1	Chọn mô hình cuối cùng & thảo luận trade-off	Nêu lý do chọn mô hình cuối, cân bằng độ chính xác và khả năng giải thích	Chọn Prophet, lý do khả năng mô phỏng lại trend của mô hình này trên kết quả test.	[Done]	[100]	7.2.3, tr 36 6.3.3, tr 27-30	code/problem-02/Customer AndRevenue.ipynb, mục Phân tích insights (5 mục con) Cell 22-23, Section Phân tích nhóm mua lẻ Cell 29-30, Section Phân tích nhóm mua sỉ	
RES-2	Trình bày kết quả dự báo tương lai	Có dự báo cho giai đoạn tương lai, biểu đồ kèm khoảng tin cậy (nếu có)	Có dự báo tương lai, và so sánh với kết quả thực tế và trả về khoảng tin cậy	[Done]	[100]	6.3.3, tr 27-30	code/problem-02/Customer AndRevenue.ipynb, mục Phân tích insights (5 mục con) Cell 22-23, Section Phân tích nhóm mua lẻ Cell 29-30, Section Phân tích nhóm mua sỉ	
RES-3	Nêu rõ điều có thể/không thể suy ra từ mô hình	Trình bày giới hạn diễn giải, giả định mô hình,		[Not]	[0]	[Mục/Trang]	[File/Section/Cell]	

		nguy cơ overfitting/shift...						
RES-4	Hạn chế & hướng phát triển tiếp theo	Chỉ ra hạn chế dữ liệu/mô hình; đề xuất dữ liệu bổ sung/phương pháp khác/cải thiện	Hạn chế: Mô hình hiện tại chỉ khai thác thông tin nội tại của chuỗi doanh thu mà chưa xem xét các biến ngoại sinh như chương trình khuyến mãi, yếu tố kinh tế vĩ mô hoặc sự kiện đặc biệt.	[Done]	[100]	7.1.2, tr 34	[File/Section/Cell]	Chỉ ghi trong tài liệu, không có trong code

3.6. Sản phẩm nộp (đối chiếu yêu cầu nộp bài)

ID	Công việc bắt buộc	Tiêu chí tối thiểu	Kết quả/Mình chứng (tóm tắt)	Trạng thái	%	Tham chiếu trong báo cáo	Tham chiếu trong notebook	Ghi chú
DEL-1	Báo cáo phân tích chuỗi thời gian (PDF/DOCX)	Trình bày chuyên nghiệp, hướng tới nhà quản lý; đầy đủ các phần Milestone 1-3	Đã hoàn thành	[Done]	[100]	[N/A]	[N/A]	
DEL-2	Slide thuyết trình	Tóm tắt vấn đề - phương pháp - kết quả - khuyến nghị; số liệu/biểu đồ rõ ràng	Đã hoàn thành	[Done]	[100]	[N/A]	[N/A]	

DEL-3	Notebook Python	Code/Markdown rõ ràng; tái lập kết quả; có cấu trúc theo nội dung báo cáo	Đã hoàn thành	[Done]	[100]	[N/A]	[N/A]	
DEL-4	Bản in mang vào buổi vấn đáp	Báo cáo tổng kết + bảng checklist đã điền theo mẫu	50%	[Partial]	[50]	[N/A]	[N/A]	

4. Tự đánh giá tổng quan (tóm tắt 5-7 dòng)

- Điểm mạnh: Đã biết cách khai thác 1 vấn đề trọng tâm là doanh thu. Đã biết cách khai thác đi sâu vào vấn đề là khai thác phân khúc khách hàng hệ thống để thấu hiểu khách hàng nhiều hơn. Đã nhìn nhận được bài toán này doanh thu dao động mạnh nên việc phân rõ chọn phương pháp cộng, thay vì phương pháp nhân. Việc làm nhóm khá thuận lợi, các thành viên đều biết sử dụng công cụ như notebook, git, latex nhuần nhuyễn, ...
- Điểm yếu: Vẫn chưa phân tích được outliers. Chưa kịp khai thác các vấn đề liên quan đến doanh thu trên đối tượng khách vắng lai. Việc segmentation customers vẫn còn hạn chế một phương pháp, chưa đa dạng hóa.

5. Xác nhận

Sinh viên/Đại diện nhóm ký xác nhận: _____ Ngày: _31_/_01_/2026_____

Giảng viên ghi nhận (sau vấn đáp): _____ Điểm milestone 3 (nếu có):
