

# STT	Delivery Outcome	Topics	Questions	<input checked="" type="checkbox"/> Check
1	Jupyter notebook	EDA	Đã phân tích Target Analysis + visualization	<input checked="" type="checkbox"/>
2	Jupyter notebook	EDA	Đã đánh giá data quality (null/missing/outlier, data types...etc) - dạng table	<input checked="" type="checkbox"/>
3	Jupyter notebook	EDA	Đã phân tích Univariate Analysis + visualization	<input checked="" type="checkbox"/>
4	Jupyter notebook	EDA	Đã phân tích Multivariate Analysis + visualization	<input checked="" type="checkbox"/>
5	Jupyter notebook	EDA	Đã làm Multivariate Analysis cùng với target + visualization	<input checked="" type="checkbox"/>
6	Jupyter notebook	EDA	Đã comment nhận định về insight ở trên (Target/Uni/Multivariate Analysis)	<input checked="" type="checkbox"/>
7	Jupyter notebook	Data Split	Dataset có time-depedent không?	<input checked="" type="checkbox"/>
8	Jupyter notebook	Data Split	Đã chọn dataset split phù hợp (random split, cross-validation, time-series split,...etc)	<input checked="" type="checkbox"/>
9	Jupyter notebook	Data Split	Nếu feature/target bị thay đổi theo thời gian (time-dependent), đánh giá sự thay đổi đó (KS-statistic, PSI)	<input type="checkbox"/>
10	Jupyter notebook	Data Split	Trả lời các câu hỏi sau về cách chia train/test split mà bạn chọn cho dataset này: + Nếu là random split, thì expected performance của model trên tập train và trên tập test ra sao? (giống hay khác). Có rủi ro gì không trong quá trình đánh giá kết quả? + Nếu là time-series split và các chỉ số KS stats/PSI ở trên cho thấy có sự khác nhau rõ ràng giữa phân phối target (regression) hoặc tỷ lệ label (classification) của train và test. Thì expected performance của model trước khi train của bạn là gì? Có rủi ro gì trong quá trình đánh giá không? + Đối với random split, liệu bài toán của bạn có thể thay đổi thành bài toán theo time-series split không? Diễn giải thử cho người nghe cách thay đổi bài toán từ random split sang time-series split	<input type="checkbox"/>
11	Jupyter notebook	Data Pipeline	Đã chọn data pipeline phù hợp ứng với từng feature (category, number)	<input checked="" type="checkbox"/>
12	Jupyter notebook	Data Pipeline	Đã đánh giá về distribution change sau khi transform numerical features. Liệu có sự thay đổi quá lớn?	<input type="checkbox"/>
13	Jupyter notebook	Data Pipeline	Đã có bao nhiêu column mới sau khi transform categorical features. Data sparity mới là bao nhiêu?	<input type="checkbox"/>
14	Jupyter notebook	Pre-Modeling	Đã xác định được các metric phù hợp cho bài toán này chưa? Vì sao	<input checked="" type="checkbox"/>
15	Jupyter notebook	Pre-Modeling	Đã xác định được các giá trị hoặc ngưỡng mà metric đánh giá là tốt hoặc ko tốt chưa?	<input checked="" type="checkbox"/>
16	Jupyter notebook	Pre-Modeling	Đã chọn các mô hình phù hợp chưa? (ít nhất là 2 mô hình khác nhau)	<input checked="" type="checkbox"/>
17	Jupyter notebook	Pre-Modeling	Đã nhận định điểm mạnh/yếu các mô hình là gì chưa?	<input checked="" type="checkbox"/>
18	Jupyter notebook	Pre-Modeling	Đã dự đoán mô hình nào tốt nhất trước khi train chưa ?	<input checked="" type="checkbox"/>
19	Jupyter notebook	Post-Modeling	Đã train ít nhất 2 mô hình khác nhau chưa?	<input checked="" type="checkbox"/>
20	Jupyter notebook	Post-Modeling	Đã plot các kết quả evaluation cần thiết chưa? (AUC, Residual Plot,...etc)	<input type="checkbox"/>
21	Jupyter notebook	Post-Modeling	Đã tổng hợp kết quả thành bảng chưa? Bao nhiêu metric?	<input checked="" type="checkbox"/>
22	Jupyter notebook	Evaluation	Đã chọn 1 metric duy nhất phù hợp nhất với bài toán này chưa?	<input checked="" type="checkbox"/>
23	Jupyter notebook	Evaluation	Đã tìm ra mô hình nào tốt nhất chưa?	<input checked="" type="checkbox"/>
24	Jupyter notebook	Advanced Analytic	Đã đưa ra giải thích về mô hình chưa? (Feature Importance, Coefficient, WoE/IV, SHAP values...etc)	<input type="checkbox"/>
25	Jupyter notebook	Advanced Analytic	Đã đánh giá về overfitting/underfitting chưa? Nếu overfitting, liệu có khớp với nhận định về expected performance trước khi train không? (ý 10)	<input type="checkbox"/>
26	Jupyter notebook	Advanced Analytic	Đã đánh giá về feature statilibity chưa (KS stats, PSI)	<input type="checkbox"/>
27	Jupyter notebook	Advanced Analytic	Đã đánh giá về model stability chưa (KS-statistic, PSI)	<input type="checkbox"/>

# STT	Delivery Outcome	Topics	Questions	<input checked="" type="checkbox"/> Check
28	Deployment	Deploy	Đã hỏi Quản Độn g về triển khai chưa?	<input checked="" type="checkbox"/>
29	Deployment	Deploy	Đã vẽ lại workflow/diagram của hệ thống triển khai chưa? Lưu ý bản vẽ thể hiện rõ các component và workflow rõ ràng	<input checked="" type="checkbox"/>
30	Deployment	Python-Code	Đã viết script dựng data preprocessor + mô hình lên online serving chưa?	<input checked="" type="checkbox"/>
31	Deployment	Python-Code	Đã viết script gửi API request vào mô hình chưa?	<input checked="" type="checkbox"/>
32	Deployment	Python-Code	Đã có kết quả nhận được từ server chưa?	<input checked="" type="checkbox"/>
33	Deployment	Python-Code	Đã test thử thành công 1 ví dụ sample label chưa?	<input checked="" type="checkbox"/>
34	Presentation	Follow instruction đây để bạn có thể thiết kế slide tốt nhất	Bạn chỉ có 10 - 15p present và 15p còn lại để Q&A (format style Consultant Presentation) Bạn sẽ làm presentation theo mô hình McKinsey là Situation-Complication-Resolution (SCR) Framework Bạn sẽ để các ý trọng tâm ở trên, các ý detail/supporting evidence sẽ để phần Appendix sau cùng	<input type="checkbox"/>
35	Presentation	1. Giới thiệu & mục tiêu bài toán (1 slide)	Miêu tả thông tin liên qua về tính cấp thiết & cấp bách bài toán. Vì sao cần solution cho problem này	<input type="checkbox"/>
36	Presentation	2. Hiện trạng dữ liệu (1,2 slide)	Mô tả hiện trạng dữ liệu bao gồm các thống kê về data source, quality, error, các statistic cơ bản...etc	<input type="checkbox"/>
37	Presentation	3. Phương pháp luận (1 slide)	Dựa trên mục tiêu bài toán hãy đưa ra 1 tổng thể cách làm về quy trình và ước tính kết quả thu được. VD: Do nhu cầu tiêu dùng của KH, cho nên chúng ta cần có 1 mô hình dự đoán hàng tồn kho cần thiết để tối đa lợi nhuận. Và mô hình X này sẽ giúp chúng ta giải quyết dự đoán hàng tồn kho	<input type="checkbox"/>
38	Presentation	4. Thực hiện & Kết quả mô hình (2-5 slides)	+ Thực hiện mô hình nào? Vì sao? Và chất lượng mô hình ra sao? + Ứng dụng kết quả mô hình sẽ được gửi cho ai/phòng ban nào? + Vận hành kết quả từ mô hình như thế nào? (có thể giả sử tình huống cụ thể) + Business Impact thu được từ mô hình là gì? (có thể giả sử tình huống cụ thể)	<input type="checkbox"/>
39	Presentation	5. Các hướng phát triển khác (1 slide)	+ Các phương pháp làm khác hoặc bài toán có thể mở rộng hơn + Các nghiên cứu hoặc các bài viết liên quan cho bài toán này + Các chỉ số business impact có thể liên quan (có thể giả sử tình huống cụ thể)	<input type="checkbox"/>