

Bốn cấp độ phân tích dữ liệu trong kinh doanh.

Có 4 cấp độ phân tích tăng dần về giá trị và độ phức tạp: **Mô tả** → **Chẩn đoán** → **Dự đoán** → **Đề xuất**. Chúng trả lời lần lượt **Chuyện gì đã xảy ra?** → **Tại sao?** → **Có thể sẽ xảy ra gì?** → **Nên làm gì?**. Muốn làm tốt, cần kết hợp **bối cảnh kinh doanh**, **công nghệ** (SQL, Python, BI), và **khoa học dữ liệu** (thống kê, ML, tối ưu hóa).

1) Bức tranh tổng thể

Cấp độ	Câu hỏi trung tâm	Giá trị tạo ra	Đầu vào tối thiểu	Đầu ra điển hình	Công cụ/KT chính
Descriptive (Mô tả)	Chuyện gì đã xảy ra?	Nhìn rõ hiện trạng	Dữ liệu lịch sử đã làm sạch	Báo cáo, dashboard, KPI cơ bản	Thống kê mô tả, SQL GROUP BY , pivot, biểu đồ
Diagnostic (Chẩn đoán)	Tại sao nó xảy ra?	Tìm nguyên nhân gốc rễ	Dữ liệu + giả thuyết	Phát hiện yếu tố tác động	Tương quan, kiểm định (t-test/ANOVA), phát hiện bất thường, RCA (5 Whys)
Predictive (Dự đoán)	Điều gì có thể xảy ra?	Dự báo tương lai	Dữ liệu lịch sử + đặc trưng	Xác suất/khoảng dự báo	Hồi quy, phân loại, chuỗi thời gian (ARIMA/Prophet/LSTM)
Prescriptive (Đề xuất)	Nên làm gì?	Tối ưu quyết định	Dự báo + ràng buộc	Kế hoạch/khuyến nghị hành động	Quy hoạch tuyến tính, mô phỏng Monte Carlo, GA, RL

Quy trình học tập/triển khai nên đi **từ trái sang phải**: đảm bảo mô tả chính xác → chẩn đoán đáng tin → dự đoán ổn định → đề xuất hành động khả thi.

2) Cấp độ 1 — Phân tích Mô tả (Descriptive)

Mục tiêu: Tóm tắt dữ liệu thô thành thông tin dễ hiểu để nắm hiện trạng.

- **Kỹ thuật cốt lõi:**

- Thống kê mô tả: mean, median, mode, variance, standard deviation, range.
- SQL tổng hợp: **GROUP BY**, **COUNT**, **SUM**, **AVG**, **MIN/MAX**.
- Pivot/Dashboard: cột KPI, lọc theo thời gian, phân khúc.
- Phân cụm (k-means) & luật kết hợp (Apriori) khi cần mô tả cấu trúc/giỏ hàng.

Đầu ra kỳ vọng: Báo cáo định kỳ, KPI theo thời gian/kênh/SKU/khách hàng, biểu đồ xu hướng & phân bố.

3) Cấp độ 2 — Phân tích Chẩn đoán (Diagnostic)

Mục tiêu: Trả lời “Tại sao” bằng kiểm chứng nguyên nhân.

- **Kỹ thuật:**

- **Tương quan & ma trận tương quan** để sàng lọc mối liên hệ.
- **Kiểm định giả thuyết:** t-test, ANOVA, kiểm định tỷ lệ.
- **Phát hiện bất thường** (z-score, IQR, isolation forest) để tìm sự kiện gây lệch.
- **RCA** (5 Whys, Ishikawa) để hệ thống hóa nguyên nhân gốc.

Đầu ra kỳ vọng: Danh sách yếu tố tác động chính, mức độ ảnh hưởng, khuyến nghị thử nghiệm A/B.

4) Cấp độ 3 — Phân tích Dự đoán (Predictive)

Mục tiêu: Ước lượng tương lai với mức bất định định lượng được.

- **Kỹ thuật:**

- **Hồi quy** (Linear/Tree/Ensemble) cho biến liên tục (nhu cầu, doanh thu).

- **Phân loại** (Logistic, SVM, XGBoost, NN) cho kết quả nhị phân/đa lớp (churn, rủi ro).
- **Chuỗi thời gian**: ARIMA/SARIMA, Prophet, LSTM/Temporal Fusion Transformer.
- **Đánh giá mô hình**: hold-out, cross-val; chỉ số MAE/MAPE/RMSE/AUC.

Đầu ra kỳ vọng: Dự báo có **khoảng tin cậy**, danh sách khách hàng/SKU rủi ro cao, bản đồ nhu cầu theo vùng.

5) Cấp độ 4 — Phân tích Đề xuất (Prescriptive)

Mục tiêu: Ra quyết định tối ưu dưới ràng buộc thực tế.

- **Kỹ thuật:**
 - **Quy hoạch tuyến tính (LP/MILP)**: phân bổ ngân sách, lập kế hoạch sản xuất-vận tải, phân kho.

Đầu ra kỳ vọng: Kế hoạch vận hành/bố trí nguồn lực tối ưu, “what-if” theo kịch bản, **khuyến nghị có thể hành động** (actionable).

6) Liên kết 3 trụ cột: Kinh doanh – Công nghệ – Khoa học dữ liệu

- **Bối cảnh kinh doanh:** Mục tiêu (giảm chi phí, tăng OTD, tối ưu tồn kho), ràng buộc (ngân sách, SLA, năng lực).
- **Công nghệ:**
 - **SQL** cho trích xuất/tiền xử lý;
 - **Python/R** cho thống kê, ML, tối ưu;
 - **BI** (Power BI/Tableau) cho trực quan & theo dõi KPI.
- **Khoa học dữ liệu:** Thiết kế thí nghiệm, chọn mô hình & chỉ số, kiểm định, chuẩn hóa quy trình (CRISP-DM).