**Project Proposal**

### Tiêu đề dự án

Phân tích tình hình kinh doanh bán hàng tại siêu thị tại Mỹ và dự báo doanh thu.

### Mục đích và Kết quả mong đợi

**Mục đích:** Nghiên cứu dữ liệu bán hàng của Siêu thị nhằm đánh giá hiệu suất kinh doanh, xác định các xu hướng quan trọng, và đề xuất các chiến lược tối ưu hóa doanh thu.

**Kết quả mong đợi:**

* Cung cấp insight quan trọng về hiệu suất bán hàng, hành vi khách hàng.
* Đưa ra các đề xuất cải tiến kinh doanh dựa trên dữ liệu.

### Nguồn dữ liệu

**Nguồn:** Kaggle.com  
**Link dataset:** <https://www.kaggle.com/datasets/ishanshrivastava28/superstore-sales>

**Mô tả:** Dữ liệu bán hàng của Superstore, bao gồm thông tin về đơn hàng, khách hàng, sản phẩm, khu vực và lợi nhuận.

Dữ liệu bao gồm 9.994 bản ghi với 21 cột, trong đó có các thông tin quan trọng như:

* **Order ID**: Thông tin đơn hàng
* **Order Date**: Ngày đặt
* **Ship Date**: Ngày ship
* **Ship Mode**: Phương thức giao hàng
* **Customer ID**: Mã khách hàng
* **Customer Name**: Tên khách hàng
* **Segment**: Loại khách hàng
* **Country**: Quốc gia
* **City**: Thành phố
* **State**: Bang
* **Region**: Khu vực
* **Product ID**: Mã sản phẩm
* **Product Name**: Tên sản phẩm
* **Category**: Danh mục
* **Sub-Category**: Danh mục con
* **Sales**: Doanh thu sau khi áp dụng giảm giá
* **Quantity**: Số lượng sản phẩm
* **Discount**: Giảm giá
* **Profit**: Lợi nhuận

### Kế hoạch phân tích dữ liệu

1. **Phân tích doanh số theo thời gian**
   * Thống kê doanh thu theo tháng, quý, năm.
   * Xác định xu hướng bán hàng và thời điểm cao điểm.
   * Đánh giá thời gian giao hàng trung bình và tác động đến khách hàng.
2. **Phân tích khách hàng**
   * Phân loại khách hàng theo phân khúc (Consumer, Corporate, Home Office).
   * Xác định nhóm khách hàng đóng góp nhiều nhất vào doanh thu.
   * Phân tích hành vi mua sắm theo khu vực địa lý.
3. **Phân tích danh mục sản phẩm**
   * Xác định sản phẩm bán chạy nhất và ít được ưa chuộng.
   * So sánh lợi nhuận giữa các danh mục sản phẩm.
   * Đánh giá tác động của giảm giá đến lợi nhuận.
4. **Phân tích theo khu vực địa lý**
   * So sánh doanh thu giữa các khu vực và thành phố.
   * Xác định khu vực có lợi nhuận cao nhất.

### Phương pháp tiếp cận

### **Làm sạch dữ liệu bằng Python**: Kiểm tra và xử lý giá trị bị thiếu, định dạng lại dữ liệu. Bổ sung các cột tính toán

* **Doanh thu sau giảm giá(Net Sales)** = Sales - Discount
* **Đơn giá sản phẩm** = Doanh sau trước giảm giá / Quantity
* **Giá vốn hàng bán (COGS - Cost of Goods Sold)** = Doanh thu sau giảm giá - Profit
* **Lợi nhuận gộp trên mỗi đơn vị (Gross Profit per Unit)** = (Doanh thu sau giảm giá - COGS) / Quantity
* **Tỷ suất lợi nhuận gộp (Gross Margin)** = (Doanh thu sau giảm giá - COGS) / Doanh thu sau giảm giá
* **Lợi nhuận trên mỗi đơn vị (Profit per Unit)** = Profit / Quantity
* **Số ngày chờ giao hàng** = Ship Date - Order Date
* **Tạo mô hình dữ liệu (Data Modeling) bằng Power BI hoặc SQL**: Thiết kế bảng dữ liệu quan hệ, xây dựng các truy vấn để tổng hợp thông tin.
* **Phân tích dữ liệu khám phá (EDA)**: Sử dụng Power BI để trực quan hóa dữ liệu.
* **Phân tích thống kê**: Xác định mối quan hệ giữa các biến và doanh số.
* **Dự báo doanh số** : Sử dụng Power BI
* **Xây dựng bảng điều khiển Power BI**: Trình bày kết quả phân tích trực quan.

### Trực quan hóa dữ liệu

* **Biểu đồ cột**: So sánh doanh thu theo danh mục sản phẩm, khu vực.
* **Card** : Thể hiện các con số tổng, trung bình…
* **Biểu đồ line**: Theo dõi xu hướng doanh số theo thời gian.
* **Biểu đồ tròn**: Phân tích tỷ lệ khách hàng theo phân khúc.
* **Heatmap**: Xác định thời gian cao điểm bán hàng.
* **Scatter plot** : Để xem mối quan hệ của các dữ liệu liên quan
* **Biểu đồ line** : Dự báo doanh thu trong tương lai
* **Dashboard Power BI**: Tổng hợp các chỉ số quan trọng.

### Lộ trình thực hiện

1. Thu thập và làm sạch dữ liệu bằng Python.
2. Phân tích dữ liệu khám phá (EDA) và trực quan hóa bằng Power BI.
3. Tạo mô hình dữ liệu (Data Modeling) bằng Power BI hoặc SQL.
4. Phân tích thống kê và đưa ra kết luận.
5. Dự báo doanh số trên Power BI.
6. Hoàn thành báo cáo cuối cùng và trình bày kết quả.

### Thuyết trình

Trình bày kết quả phân tích và cung cấp các khuyến nghị thực tế nhằm tối ưu hóa doanh thu, cải thiện dịch vụ khách hàng và nâng cao hiệu quả kinh doanh.