**Assignment 01: Phương pháp luận Khoa học Dữ liệu**

1. **Đề tài**

Khách hàng của bạn là một siêu thị lớn và hội đồng quản trị đặt ra cho bạn một thách thức để giúp tăng trưởng lợi nhuận cho việc kinh doanh của họ.

Vì có một lí do khiến lợi nhuận siêu thị chưa tăng trưởng như kì vọng đó là có những mặt hàng bị tồn kho nhiều và một số mặt hàng có nhu cầu cao nhưng giá cao hơn thị trường chung mà chính sách hậu mãi cũng không quá khác biệt so với đối thủ. nhóm đang nghiên cứu mô hình dự đoán nhu cầu mua những sản phẩm trong siêu thị để giúp ban lãnh đạo có những quyết định kinh doanh tốt hơn.

1. **Business Understanding**

**Nhóm sẽ nghiên cứu mô hình dự đoán những sản phẩm có doanh số bán hàng cao trong khoảng 6 tháng cuối năm.**

**Vậy việc dự đoán doanh số các sản phẩm bán chạy thì có ý nghĩa gì đối với siêu thị ?**

**Việc dự đoán doanh số các sản phẩm giúp siêu thị như sau:**

* **Đưa ra được chiến lược về giá hợp lí cho các sản phẩm bán chạy và các sản phẩm không bán chạy**
* **Lên kế hoạch nhập hàng, tránh tình trạng hết hàng khi doanh số mặt hàng đó tăng cao**
* **Chuẩn bị ngân sách để có những chiến lược bán hàng, nhập hàng, matketing, vận chuyển hợp lí nhất**
* **Đưa ra chiến lược chăm sóc khách hàng tốt hơn theo từng loại sản phẩm**

1. **Analytic Approach**

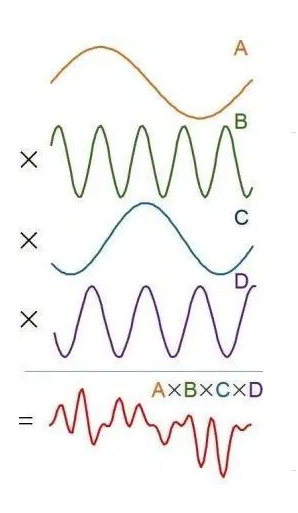
**Trong việc nghiên cứu giải pháp, nhóm sử dụng mô hình dự đoán để đưa ra xác suất mua hàng của từng sản phẩm trong 1 tháng tới và sẽ ưu tiên bán và nhập những sản phẩm có xác suất cao nhất.**

**Có nhiều mô hình dự đoán để áp dụng vào việc dự đoán về triển vọng sản phẩm và nhóm nghiên cứu quyết định sử dụng mô hình time series, mô hình được ứng dụng nhiều trong việc dự báo trong đời sống**

**Mô hình time series**

* **Giao thoa của sóng :**

**Sự giao thoa của 4 sóng hình sin với chu kì và pha khác nhau, kết quả thu được một quỹ đạo lên xuống rất phức tạp. Có thể thấy, các dữ liệu trong thực tế thường gập ghềnh và khó phát hiện xu hướng**

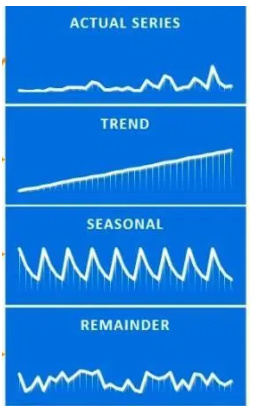


**Hình 1. Sự giao thoa của 4 sóng hình sin**

* **Phân tích time series:**

**Chúng ta có thể phân rã một dữ liệu chuỗi thời gian thành 4 thành phần con độc lập như sau:**

* **Trend: thành phần này chỉ ra xu hướng tổng quan của dữ liệu theo thời gian: lên hoặc xuống, tăng hoặc giảm.**
* **Seasonality: thành phần chỉ ra các xu hướng theo quý , tháng.**
* **Cycle: thành phần chu kì, nó khác Seasonality ở việc thành phần này có sự vận động trong khoảng thời gian dài hơn (nhiều năm)**
* **Irregular remainder: thành phần nhiễu còn lại sau khi đã trích xuất hết các thành phần ở trên, nó chỉ ra sự bất thường của các điểm dữ liệu.**



**Hình 2. Phân tích time series**

**Nguồn:** <https://insights.magestore.com/posts/giai-thuat-time-series-forecasting>

**Thay vì đưa ra dự báo trên Actual Series – là đường cong gập ghềnh và khó dự đoán, ta phân tích các thành phần con có tính quy luật với xu hướng vận động rõ ràng hơn rồi kết hợp các thành phần đó lại với nhau như việc kết hợp của 4 sóng hình sin như trên.**

**Ứng dụng của time series:**

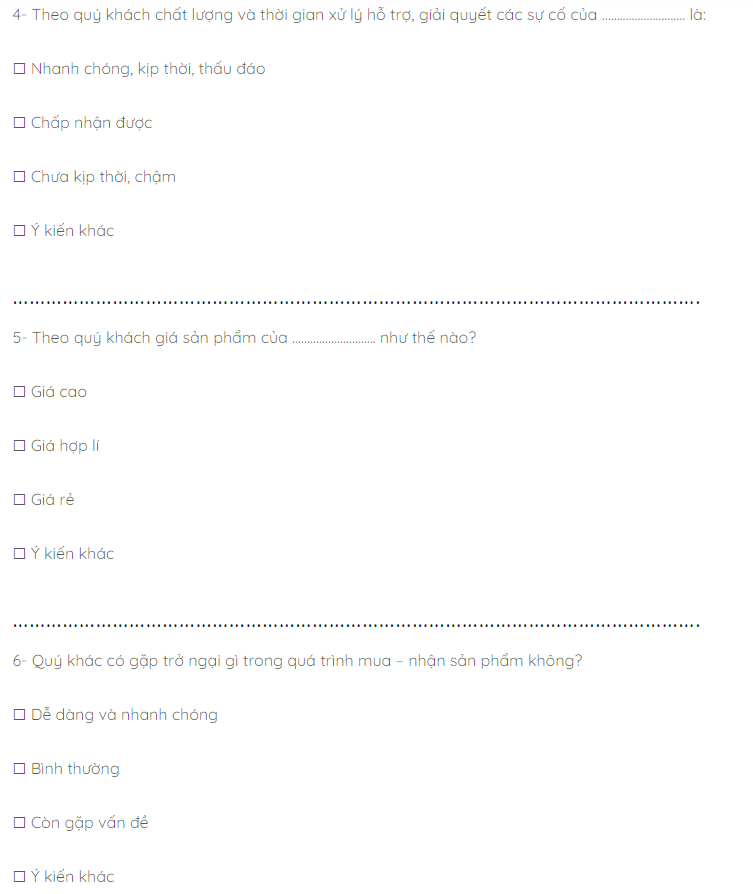
* **Dự đoán số lượng khách ăn pizza của một nhà hàng vào 2 ngày cuối tuần**
* **Dự báo nhu cầu của sản phẩm mới ra mắt của công ty để có phương án mua nguyên liệu sản xuất**
* **Dự đoán số lượng người mua iphone 14 pro max trong 1 tháng tớis**

1. **Data Requirements**

* **Loại dữ liệu:**
* **Phiếu khảo sát**

****

**Hình 3: Mẫu phiếu khảo sát**

****

**Hình 4: Mẫu phiếu khảo sát**

**Nguồn:** <https://stringeex.com/vi/blog/post/phieu-khao-sat-y-kien-khach-hang>

* **Các file nhập, xuất kho của siêu thị**
* **Thông tin hành vi (Lịch sử mua hàng, tỉ lệ quay lại mua hàng, giá trị đơn hàng, …) trên web siêu thị**
* **Lịch sử mua hàng cả online và offline**
* **Dự báo của các chuyên gia kinh tế về một số mặt hàng**
* **Nơi thu thập dữ liệu**
* **Các mẫu phiếu khảo sát qua link Google form, khách hàng có thể quét mã để truy cập**
* **Lịch sử mua hàng, hóa đơn nhập/ xuất kho từ máy chủ siêu thị**
* **Thông tin hành vi trên web của siêu thị được xử lí thông Google Analytics**
* **Dự báo của các chuyên gia kinh tế về một số mặt hàng ở các mặt báo kinh tế có nhiều lượt truy cập như CafeF, CaféBiz,…**
* **Số lượng dữ liệu**
* **Những thông tin đã lưu trữ trên máy chủ hoặc trong dịch vụ đám mây của siêu thị trong khoảng 3 - 5 năm trở lại đây**