**TRƯỜNG HỢP KINH DOANH 2: DỰ ÁN MARKETING BÁN HÀNG**

**Mục tiêu**

Mục tiêu chính của dự án dữ liệu bán hàng này là tận dụng dữ liệu bán hàng lịch sử để thu được những hiểu biết về xu hướng bán hàng, tối ưu hóa chiến lược marketing và nâng cao hiệu suất kinh doanh tổng thể. Bằng cách phân tích các khía cạnh khác nhau của hồ sơ bán hàng, chúng tôi hướng tới việc nâng cao việc ra quyết định, tăng doanh thu và cải thiện sự hài lòng của khách hàng.

**Mục tiêu**

* **Xác định Xu hướng Bán hàng:**
  + Phân tích dữ liệu bán hàng để xác định các mô hình theo mùa, tốc độ tăng trưởng và các xu hướng quan trọng khác theo thời gian.
  + Hiểu cách bán hàng, doanh thu và lợi nhuận biến động qua các giai đoạn khác nhau (ví dụ: hàng tháng, hàng quý).
* **Tối ưu hóa Hiệu suất Sản phẩm:**
  + Xác định sản phẩm nào (Loại Mặt hàng) có hiệu suất cao về số lượng bán được, tổng doanh thu và tổng lợi nhuận.
  + Xác định các sản phẩm có hiệu suất thấp và điều tra các lý do tiềm ẩn cho doanh thu thấp của chúng.
* **Nâng cao Nhận thức Khu vực và Quốc gia:**
  + Kiểm tra hiệu suất bán hàng trên các khu vực và quốc gia khác nhau để xác định các thị trường có hiệu suất cao nhất.
  + Hiểu các xu hướng và sở thích cụ thể của từng khu vực và quốc gia để điều chỉnh chiến lược marketing cho phù hợp.
* **Cải thiện Hiệu quả Kênh Bán hàng:**
  + So sánh hiệu suất của các kênh bán hàng khác nhau (Trực tuyến so với Ngoại tuyến) về doanh thu, lợi nhuận và số lượng bán được.
  + Xác định các kênh bán hàng hiệu quả nhất và phân bổ nguồn lực để tối đa hóa tác động của chúng.
* **Tinh chỉnh Nhận thức về Khách hàng và Phân khúc:**
  + Phân khúc khách hàng dựa trên hành vi mua hàng, ưu tiên đơn hàng và các yếu tố liên quan khác.
  + Phát triển các chiến dịch marketing có mục tiêu cho các phân khúc khách hàng khác nhau để tăng cường sự tham gia và lòng trung thành của khách hàng.
* **Đánh giá và Tối ưu hóa Chiến dịch Marketing:**
  + Phân tích tác động của các nỗ lực marketing khác nhau lên bán hàng, doanh thu và lợi nhuận.
  + Xác định các chiến dịch marketing và chiến lược hiệu quả nhất để tối ưu hóa chi phí marketing.
* **Dự báo Doanh thu trong Tương lai:**
  + Phát triển các mô hình dự đoán để dự báo doanh thu trong tương lai dựa trên dữ liệu lịch sử.
  + Sử dụng các dự báo này để lên kế hoạch cho hàng tồn kho, nhân sự và nỗ lực marketing hiệu quả hơn.
* **Tăng cường Hiệu quả Hoạt động:**
  + Xác định sự thiếu hiệu quả trong quy trình bán hàng và đề xuất cải tiến.
  + Tối ưu hóa quản lý hàng tồn kho bằng cách liên kết với xu hướng và dự báo bán hàng để ngăn ngừa tồn kho quá nhiều hoặc thiếu hàng.
* **Tạo ra những Nhận thức Hành động:**
  + Cung cấp những nhận thức hành động và đề xuất chiến lược dựa trên phân tích dữ liệu.
  + Hỗ trợ việc ra quyết định bằng các báo cáo rõ ràng và súc tích làm nổi bật những phát hiện chính và cơ hội phát triển.
* **Giao Báo cáo và Mã Code Chất lượng Cao:**
  + Đảm bảo phân tích được ghi lại đầy đủ, với mã code sạch và có thể tái tạo.
  + Tạo các báo cáo súc tích tóm tắt những hiểu biết chính, thông số mô hình và kết quả hiệu quả.

Bằng cách đạt được những mục tiêu này, dự án nhằm mục đích cung cấp một sự hiểu biết toàn diện về động lực bán hàng, cho phép doanh nghiệp đưa ra các quyết định dựa trên dữ liệu thúc đẩy tăng trưởng và hiệu quả.

**Mô tả Dữ liệu**

Bộ dữ liệu được cung cấp trong "100000 Sales Records.csv" chứa thông tin về giao dịch bán hàng. Dưới đây là mô tả chi tiết về mỗi cột trong bộ dữ liệu:

|  |  |
| --- | --- |
| **Tên Cột** | **Mô tả** |
| **Khu vực** | Khu vực địa lý nơi xảy ra giao dịch bán hàng (ví dụ: Châu Phi hạ Sahara, Châu Âu, Châu Á). |
| **Quốc gia** | Quốc gia nơi xảy ra giao dịch bán hàng. |
| **Loại Mặt hàng** | Loại sản phẩm được bán (ví dụ: Trái cây, Quần áo, Thiết bị điện tử). |
| **Kênh Bán hàng** | Kênh thông qua đó việc bán hàng được thực hiện (ví dụ: Trực tuyến, Ngoại tuyến). |
| **Ưu tiên Đơn hàng** | Mức độ ưu tiên của đơn hàng (ví dụ: Cao, Trung bình, Thấp). |
| **Ngày Đặt hàng** | Ngày đặt hàng. |
| **ID Đơn hàng** | Một mã định danh duy nhất cho đơn hàng. |
| **Ngày Giao hàng** | Ngày đơn hàng được giao. |
| **Số lượng Bán được** | Số lượng sản phẩm được bán. |
| **Giá Bán lẻ** | Giá bán lẻ cho mỗi sản phẩm. |
| **Chi phí Mỗi sản phẩm** | Chi phí cho mỗi sản phẩm. |
| **Tổng Doanh thu** | Tổng doanh thu từ giao dịch bán hàng (tính bằng Số lượng Bán được \* Giá Bán lẻ). |
| **Tổng Chi phí** | Tổng chi phí liên quan đến giao dịch bán hàng (tính bằng Số lượng Bán được \* Chi phí Mỗi sản phẩm). |
| **Tổng Lợi nhuận** | Tổng lợi nhuận từ giao dịch bán hàng (tính bằng Tổng Doanh thu - Tổng Chi phí). |

**Ý nghĩa của Cột**

* **Khu vực và Quốc gia:**
  + Phân tích dữ liệu bán hàng theo khu vực và quốc gia để hiểu hiệu suất của thị trường địa lý.
* **Loại Mặt hàng:**
  + Xác định các loại sản phẩm có hiệu suất tốt nhất và tệ nhất.
* **Kênh Bán hàng:**
  + Đánh giá hiệu quả của các kênh bán hàng khác nhau (Trực tuyến so với Ngoại tuyến).
* **Ưu tiên Đơn hàng:**
  + Hiểu tác động của ưu tiên đơn hàng đến bán hàng và hiệu quả giao hàng.
* **Ngày Đặt hàng và Ngày Giao hàng:**
  + Phân tích dữ liệu theo thời gian để xác định các xu hướng theo mùa và thời gian giao hàng.
* **Số lượng Bán được, Giá Bán lẻ và Chi phí Mỗi sản phẩm:**
  + Đánh giá chi tiết số lượng bán được, giá cả và chi phí để tính toán các chỉ số tài chính.
* **Tổng Doanh thu, Tổng Chi phí và Tổng Lợi nhuận:**
  + Các chỉ số tài chính chính để đánh giá hiệu suất kinh doanh tổng thể.

Bộ dữ liệu này cung cấp một cái nhìn toàn diện về hoạt động bán hàng của công ty, bao gồm các chỉ số địa lý, loại sản phẩm, kênh bán hàng và tài chính. Bằng cách phân tích những yếu tố này, công ty có thể đưa ra các quyết định chiến lược để tối ưu hóa nỗ lực bán hàng và marketing, cải thiện doanh thu và tăng cường lợi nhuận.

**YÊU CẦU:**

**Phân tích Khám phá Dữ liệu (EDA)**

* **Hiểu và Làm sạch Dữ liệu:**
  + Các biến chính trong bộ dữ liệu là gì? (ví dụ: Khu vực, Quốc gia, Loại Mặt hàng, Kênh Bán hàng, Ưu tiên Đơn hàng, Ngày Đặt hàng, ID Đơn hàng, Ngày Giao hàng, Số lượng Bán được, Giá Bán lẻ, Chi phí Mỗi sản phẩm, Tổng Doanh thu, Tổng Chi phí, Tổng Lợi nhuận)
  + Có bất kỳ giá trị bị thiếu hoặc ngoại lệ nào cần được giải quyết không?
* **Đường cong Giá và Khối lượng:**
  + Đường cong Giá Bán lẻ và Số lượng Bán được được vẽ và diễn giải như thế nào?
  + Những hiểu biết nào có thể rút ra từ các xu hướng về giá và khối lượng?
* **Phân tích theo Thời gian:**
  + Bán hàng, doanh thu và lợi nhuận thay đổi như thế nào theo thời gian (ví dụ: hàng tháng, hàng quý)?
  + Các xu hướng hoặc mô hình theo mùa nào được quan sát trong dữ liệu dựa trên Ngày Đặt hàng và Ngày Giao hàng?
* **Phân tích Khu vực và Quốc gia:**
  + Bán hàng, doanh thu và lợi nhuận khác nhau như thế nào trên các khu vực (ví dụ: Khu vực) và quốc gia (ví dụ: Quốc gia) khác nhau?
  + Các khu vực và quốc gia nào có hiệu suất cao nhất về Tổng Doanh thu và Tổng Lợi nhuận?

**Phát triển và Đánh giá Mô hình**

* **Chọn và Điều chỉnh Tham số Mô hình:**
  + Những mô hình nào đang được xem xét để dự đoán bán hàng, doanh thu hoặc lợi nhuận?
  + Các tham số mô hình tốt nhất được chọn như thế nào (ví dụ: thông qua điều chỉnh siêu tham số)?
* **Huấn luyện và Kiểm tra:**
  + Phương pháp chia dữ liệu nào được sử dụng để huấn luyện và kiểm tra các mô hình (ví dụ: chia tập huấn luyện-kiểm tra, xác thực chéo)?
  + Các mô hình được đánh giá như thế nào dựa trên kết quả huấn luyện và kiểm tra?
* **Chỉ số để Đánh giá:**
  + Các chỉ số chính được sử dụng để đánh giá các mô hình là gì (ví dụ: độ chính xác, độ chính xác, độ thu hồi, điểm số F1)?
  + Các chỉ số này được diễn giải và sử dụng như thế nào để so sánh hiệu suất của các mô hình khác nhau?

**Tiêu chí Đánh giá**

* **Kỹ năng Phân tích:**
  + Các nhóm phân tích dữ liệu và diễn giải kết quả hiệu quả như thế nào?
  + Những hiểu biết từ EDA có ý nghĩa và khả thi không?
* **Hiệu suất Mô hình:**
  + Các mô hình hoạt động tốt như thế nào dựa trên các chỉ số đã chọn?
  + Có bằng chứng nào cho thấy việc điều chỉnh siêu tham số một cách chu đáo để tối ưu hóa hiệu suất mô hình không?
* **Chất lượng Tổng thể:**
  + Báo cáo được cấu trúc và viết như thế nào?
  + Mã code có sạch sẽ, được ghi lại đầy đủ và dễ hiểu không?

**Tóm tắt các Yêu cầu Chính**

Mỗi nhóm cần:

* Thực hiện EDA kỹ lưỡng, bao gồm việc vẽ và diễn giải đường cong Giá Bán lẻ và Số lượng Bán được, và phân tích dữ liệu bán hàng theo thời gian (Ngày Đặt hàng, Ngày Giao hàng), khu vực (Khu vực) và quốc gia (Quốc gia) dựa trên Tổng Doanh thu và Tổng Lợi nhuận.
* Phát triển và đánh giá các mô hình, chọn các tham số tốt nhất thông qua điều chỉnh siêu tham số và chia dữ liệu để huấn luyện và kiểm tra.
* Chuẩn bị một báo cáo súc tích (dưới 3 trang A4) tập trung vào EDA, các tham số mô hình đã chọn và kết quả huấn luyện/kiểm tra.
* Nộp báo cáo và mã code cho ban tổ chức trước thời hạn quy định.
* Đảm bảo công việc của họ được đánh giá dựa trên kỹ năng phân tích, hiệu suất mô hình và chất lượng tổng thể của báo cáo và mã code.

Bằng cách giải quyết những yêu cầu và câu hỏi này, các nhóm có thể đảm bảo phân tích của họ toàn diện, các mô hình được điều chỉnh tốt và những phát hiện của họ được truyền đạt rõ ràng.

**Bảng Tiêu chí Đánh giá**

Dưới đây là bảng cho các tiêu chí đánh giá cùng với các điểm phần trăm tương ứng. Bảng này cung cấp một phân tích rõ ràng về cách mỗi khía cạnh của dự án sẽ được đánh giá.

|  |  |  |
| --- | --- | --- |
| **Tiêu chí Đánh giá** | **Mô tả** | **Điểm Phần trăm** |
| **Phân tích Khám phá Dữ liệu (EDA)** | Chất lượng và độ sâu của việc khám phá dữ liệu, bao gồm các biểu đồ và diễn giải. | 25% |
| **Phân tích Giá và Khối lượng** | Độ chính xác và những hiểu biết từ việc phân tích đường cong giá và khối lượng. | 10% |
| **Phân tích theo Thời gian** | Những hiểu biết và xu hướng được xác định từ dữ liệu theo thời gian (phân tích hàng tháng, hàng quý). | 10% |
| **Phân tích Khu vực và Quốc gia** | Phân tích dữ liệu bán hàng theo khu vực và quốc gia, bao gồm việc xác định các khu vực có hiệu suất cao nhất. | 10% |
| **Chọn và Điều chỉnh Tham số Mô hình** | Sự phù hợp của việc chọn mô hình, điều chỉnh tham số và giải thích các tham số đã chọn. | 15% |
| **Kết quả Huấn luyện và Kiểm tra** | Đánh giá hiệu suất mô hình dựa trên kết quả huấn luyện và kiểm tra, bao gồm phân tích chỉ số. | 15% |
| **Chất lượng Báo cáo** | Sự rõ ràng, súc tích và cấu trúc của báo cáo, tập trung vào EDA và kết quả mô hình. | 10% |
| **Chất lượng Mã Code và Tài liệu** | Sự sạch sẽ, tổ chức và tài liệu của mã code để có thể tái tạo. | 5% |

**Phân tích Chi tiết**

* **Phân tích Khám phá Dữ liệu (EDA) - 25%**
  + Sự kỹ lưỡng trong việc kiểm tra dữ liệu.
  + Chất lượng và sự liên quan của các biểu đồ (ví dụ: biểu đồ phân bố, biểu đồ đường).
  + Diễn giải và những hiểu biết hành động từ dữ liệu.
* **Phân tích Giá và Khối lượng - 10%**
  + Độ chính xác trong việc vẽ Giá Bán lẻ và Số lượng Bán được.
  + Những hiểu biết và xu hướng được xác định từ đường cong giá và khối lượng.
* **Phân tích theo Thời gian - 10%**
  + Xác định các xu hướng bán hàng theo thời gian.
  + Phân tích các mô hình theo mùa và tốc độ tăng trưởng.
* **Phân tích Khu vực và Quốc gia - 10%**
  + Phân tích chi tiết hiệu suất bán hàng theo khu vực và quốc gia.
  + Xác định các khu vực/quốc gia có hiệu suất cao nhất và thấp nhất.
* **Chọn và Điều chỉnh Tham số Mô hình - 15%**
  + Căn cứ để chọn các mô hình.
  + Hiệu quả của việc điều chỉnh siêu tham số.
  + Giải thích lý do tại sao các tham số cụ thể được chọn.
* **Kết quả Huấn luyện và Kiểm tra - 15%**
  + Phương pháp để chia dữ liệu (ví dụ: chia tập huấn luyện-kiểm tra, xác thực chéo).
  + Các chỉ số hiệu suất (ví dụ: độ chính xác, độ chính xác, độ thu hồi, điểm số F1) và cách diễn giải của chúng.
* **Chất lượng Báo cáo - 10%**
  + Sự rõ ràng và cấu trúc của báo cáo.
  + Sự súc tích (dưới 3 trang A4).
  + Tập trung vào những phát hiện chính từ EDA và kết quả mô hình.
* **Chất lượng Mã Code và Tài liệu - 5%**
  + Mã code sạch sẽ và được tổ chức.
  + Tài liệu và chú thích thích hợp để có thể tái tạo.
  + Sự dễ dàng trong việc hiểu và theo dõi quy trình phân tích.