**Yêu cầu các nội dung cần trình bày buổi bảo vệ**

1. Trình bày tổng quan về dữ liệu (hiểu dữ liệu, hiểu nghiệp vụ, hiểu ý nghĩa các thuộc tính)
2. Trình bày, show các báo cáo, biểu đồ, dashboad trên power BI theo mục A (thể hiện tình hình thực trạng doanh nghiệp của các chỉ số trên báo cáo, đủ 20 báo cáo)
3. Trình bày và giải thích vấn đề thể hiện qua các biểu đồ ( bằng kết quả phân tích dữ liệu trên Python đã lý giải được qua các báo cáo)
4. Trình bày cụ thể các bước để xây dựng được 1 trong các model mục B (kiểm tra, tiền xử lý dữ liệu, mã hóa thuộc tính, chuẩn hóa dữ liệu, lựa chọn model huấn luyện và kết quả các chỉ số đánh giá model):
   1. kết quả gồm các chỉ số đánh giá + đồ thị thể hiện hiện kết quả của mô hình:
      1. **model dự đoán:** đường giá trị thực tế và đường dự đoán
      2. **model phân loại:** đồ thị tỷ lệ % phân loại đúng sai
      3. **model phân cụm:** đồ thị thể hiện các điểm dữ liệu trong mỗi cụm
   2. đưa ra nhận xét và đánh giá kết quả output của mô hình
5. Tổng kết các công việc đã làm và đưa ra hướng phát triển trong tương lai

**Các yêu cầu công việc cụ thể**

### **Yêu cầu Báo cáo Dashboard và phân tích dữ liệu (SQL, Power BI, Python)**

1. **Tổng số tiền vay đăng ký ban đầu và số tiền giải ngân**:
   * Tạo biểu đồ tổng hợp để hiển thị số tiền vay đăng ký ban đầu (SoTienDKVayBanDau) và số tiền giải ngân (TienGiaiNgan) theo thời gian (tháng/năm).
   * Phân tích xu hướng tăng trưởng hoặc giảm sút của số tiền vay và giải ngân qua các kỳ.
2. **Số tiền còn lại phải trả theo trạng thái vay**:
   * Tạo biểu đồ phân loại trạng thái khoản vay (Trạng thái) và hiển thị số tiền còn lại phải trả (SoTienConLai).
   * Phân tích sự phân bố số tiền vay còn lại theo các trạng thái (đang vay, đã trả xong, có nợ xấu).
3. **Phân tích điểm tín dụng (TS\_CREDIT\_SCORE\_V2)**:
   * Tạo biểu đồ phân phối điểm tín dụng của khách hàng (TS\_CREDIT\_SCORE\_V2).
   * Phân tích mối quan hệ giữa điểm tín dụng và khả năng trả nợ hoặc tình trạng vay.
4. **Tỷ lệ khách hàng vay theo giới tính**:
   * Tạo biểu đồ cột hiển thị tỷ lệ vay theo giới tính (Gender).
   * Phân tích sự khác biệt trong việc vay của khách hàng nam và nữ.
5. **Độ tuổi khách hàng và số tiền vay**:
   * Tạo biểu đồ phân phối độ tuổi của khách hàng (Birthday) và phân tích mối quan hệ với số tiền vay (SoTienDKVayBanDau).
   * Xác định độ tuổi nhóm khách hàng vay nhiều nhất.
6. **Phân tích tình trạng vay (Trạng thái)**:
   * Tạo biểu đồ tròn hoặc cột để hiển thị tỷ lệ các trạng thái khoản vay (Trạng thái).
   * Phân tích các trạng thái như đang vay, đã trả xong, có nợ xấu, v.v.
7. **Phân tích khu vực khách hàng theo địa chỉ**:
   * Tạo bản đồ hiển thị phân bố khách hàng theo các địa chỉ như thành phố (CityName), quận (DistrictName), và phường (WardName).
   * Phân tích sự phân bố khách hàng theo khu vực địa lý.
8. **Mối quan hệ giữa thời gian sống và số tiền vay**:
   * Tạo biểu đồ phân tích mối quan hệ giữa thời gian đã sống (Thời gian đã sống) và số tiền vay đăng ký ban đầu (SoTienDKVayBanDau).
   * Xác định liệu thời gian sống lâu dài có ảnh hưởng đến khả năng vay hay không.
9. **Thống kê nghề nghiệp và thu nhập**:
   * Tạo biểu đồ phân phối nghề nghiệp (JobName) và thu nhập (Salary) của khách hàng.
   * Phân tích mối quan hệ giữa nghề nghiệp và mức thu nhập trong các khoản vay.
10. **Mối quan hệ giữa công ty và sản phẩm tín dụng**:
    * Tạo biểu đồ để phân tích mối quan hệ giữa công ty làm việc của khách hàng (NameCompany) và sản phẩm tín dụng (ProductCreditName).
    * Phân tích sự ưa chuộng của sản phẩm tín dụng tại các công ty khác nhau.
11. **Tỷ lệ khách hàng có nợ xấu (HasBadDebt)**:
    * Tạo biểu đồ cột để hiển thị tỷ lệ khách hàng có nợ xấu (HasBadDebt).
    * Phân tích mối quan hệ giữa nợ xấu và các yếu tố khác như thu nhập, điểm tín dụng, sản phẩm tín dụng.
12. **Phân tích thanh toán trễ (HasLatePayment)**:
    * Tạo biểu đồ phân tích tỷ lệ khách hàng có thanh toán trễ (HasLatePayment).
    * Phân tích mối quan hệ giữa thanh toán trễ và các yếu tố như thu nhập, trạng thái vay.
13. **Tình trạng khoản vay và số tiền vay ban đầu**:
    * Tạo biểu đồ phân tích số tiền vay đăng ký ban đầu (SoTienDKVayBanDau) theo trạng thái khoản vay (Trạng thái).
    * Phân tích xu hướng số tiền vay ban đầu theo các trạng thái khác nhau (đã trả xong, đang vay, có nợ xấu).
14. **Số lượng khoản vay và tình trạng tài chính**:
    * Tạo biểu đồ phân tích số lượng khoản vay (NumberOfLoans) của mỗi khách hàng.
    * Phân tích mối quan hệ giữa số lượng khoản vay và tình trạng tài chính của khách hàng (nợ xấu, thanh toán đúng hạn).
15. **Sản phẩm tín dụng và phương thức thanh toán**:
    * Tạo biểu đồ phân tích các phương thức thanh toán lãi suất (InterestPaymentType) theo từng sản phẩm tín dụng (ProductCreditName).
    * Phân tích sự liên kết giữa sản phẩm tín dụng và các phương thức thanh toán.
16. **Sự thay đổi trong mức nợ theo thời gian**:
    * Tạo biểu đồ để theo dõi sự thay đổi trong số tiền vay đăng ký ban đầu (SoTienDKVayBanDau) và số tiền còn lại (SoTienConLai) theo thời gian (theo FromDate, ToDate).
    * Phân tích xu hướng giảm hoặc tăng trong mức nợ qua các kỳ.
17. **Phân tích lịch sử nợ quá hạn (LongestOverdue)**:
    * Tạo biểu đồ phân tích lịch sử nợ quá hạn dài nhất (LongestOverdue) của khách hàng.
    * Phân tích mối quan hệ giữa nợ quá hạn và các yếu tố khác như điểm tín dụng, thu nhập, sản phẩm tín dụng.
18. **Tính toán thời gian trả nợ và số tiền vay**:
    * Tạo biểu đồ phân tích thời gian trả nợ (FromDate, ToDate) so với số tiền vay đăng ký ban đầu (SoTienDKVayBanDau).
    * Phân tích mối quan hệ giữa thời gian vay và số tiền vay, từ đó đưa ra các quyết định cho các khoản vay sau này.
19. **Thống kê theo khu vực công ty (CityCompany, DistrictNameCompany)**:
    * Tạo biểu đồ phân phối các công ty theo khu vực (thành phố CityCompany và quận DistrictNameCompany).
    * Phân tích sự phân bố công ty và khách hàng theo khu vực địa lý.
20. **Phân tích mối quan hệ giữa các yếu tố nhân khẩu học và khả năng vay**:
    * Tạo biểu đồ phân tích các yếu tố nhân khẩu học như giới tính (Gender), tuổi (Birthday), nghề nghiệp (JobName), và thu nhập (Salary) để dự đoán khả năng vay.
    * Phân tích mối quan hệ giữa các yếu tố nhân khẩu học và các khoản vay mà khách hàng có thể nhận được.

### **B. Yêu cầu về Dự đoán (Sử dụng toàn bộ các biến):**

1. **Dự đoán khả năng trả hết khoản vay**: Sử dụng toàn bộ thông tin khách hàng và khoản vay để dự đoán xác suất trả hết nợ đúng hạn.
2. **Dự đoán nợ xấu (HasBadDebt)**: Sử dụng toàn bộ các biến để dự đoán liệu khách hàng có khả năng rơi vào nợ xấu hay không.
3. **Dự đoán số ngày trễ hạn tối đa (LongestOverdue)**: Dự đoán số ngày trễ hạn tối đa cho khoản vay tiếp theo dựa trên tất cả các biến đầu vào.
4. **Dự đoán số tiền còn lại phải trả (SoTienConLai)**: Sử dụng toàn bộ dữ liệu để dự đoán số tiền còn lại mà khách hàng phải trả.
5. **Dự đoán số lần vay mới (NumberOfLoans)**: Dự đoán khả năng khách hàng sẽ vay thêm lần nữa dựa trên toàn bộ thông tin lịch sử.
6. **Dự đoán loại sản phẩm tín dụng (ProductCreditName)**: Sử dụng toàn bộ dữ liệu để dự đoán loại sản phẩm tín dụng mà khách hàng sẽ chọn.
7. **Dự đoán xác suất trả nợ đúng hạn**: Dựa trên toàn bộ thông tin về khách hàng, công việc, khoản vay, và lịch sử tín dụng.
8. **Dự đoán thu nhập của khách hàng (Salary)**: Sử dụng toàn bộ dữ liệu để dự đoán thu nhập của khách hàng.
9. **Dự đoán điểm tín dụng mới (TS\_CREDIT\_SCORE\_V2)**: Dựa trên toàn bộ dữ liệu để dự đoán điểm tín dụng sau một khoảng thời gian.
10. **Dự đoán tình trạng khoản vay (Trạng thái)**: Sử dụng toàn bộ thông tin để dự đoán tình trạng khoản vay (đã trả, đang trả, quá hạn).
11. **Dự đoán khả năng có khoản vay thành công**: Xác định xác suất khách hàng nhận được khoản vay mới dựa trên toàn bộ thông tin cá nhân và lịch sử tín dụng.
12. **Dự đoán hình thức thanh toán lãi suất (InterestPaymentType)**: Sử dụng toàn bộ dữ liệu để dự đoán khách hàng sẽ chọn hình thức thanh toán nào.
13. **Dự đoán khả năng nợ trễ hạn**: Sử dụng toàn bộ thông tin để dự đoán khả năng khách hàng trễ hạn thanh toán trong tương lai.
14. **Dự đoán thời gian sống tại địa chỉ hiện tại (Thời gian đã sống)**: Dựa trên toàn bộ thông tin cá nhân để ước tính thời gian khách hàng đã sống tại địa chỉ hiện tại.
15. **Dự đoán công việc mới của khách hàng (JobName)**: Dự đoán công việc mới của khách hàng dựa trên các thông tin hiện tại.
16. **Dự đoán địa chỉ công ty (AddressCompany)**: Sử dụng toàn bộ dữ liệu để dự đoán địa chỉ công ty khách hàng làm việc.
17. **Dự đoán nguy cơ rủi ro khi cho vay theo khu vực**: Dựa trên toàn bộ thông tin để dự đoán mức độ rủi ro theo khu vực cư trú.
18. **Dự đoán số lượng khoản vay cùng thời điểm (NumberOfLoans)**: Xác định số khoản vay đang tồn tại của khách hàng.
19. **Dự đoán khả năng khách hàng thay đổi công ty**: Dựa trên toàn bộ thông tin cá nhân và lịch sử tín dụng để dự đoán khả năng thay đổi công việc.
20. **Dự đoán khả năng gia tăng thu nhập**: Sử dụng toàn bộ dữ liệu để dự đoán khả năng thu nhập của khách hàng tăng trong tương lai.

**Yêu cầu về Phân loại:**

1. Phân loại khách hàng có nợ quá hạn thành các nhóm: không nợ, nợ ngắn hạn, và nợ dài hạn.
2. Phân loại khách hàng có khả năng nhận khoản vay tiếp theo dựa trên tình trạng hiện tại (Trạng thái).
3. Phân loại giới tính khách hàng (Gender) dựa trên thông tin cá nhân khác như địa chỉ và công việc.
4. Phân loại khách hàng có thể rơi vào nợ xấu dựa vào đặc điểm công ty và mức lương (Salary).
5. Phân loại sản phẩm tín dụng (ProductCreditName) thành các loại có rủi ro cao, trung bình và thấp.

### **Yêu cầu về Phân cụm:**

1. Phân cụm khách hàng dựa trên số lần vay (NumberOfLoans) và số tiền vay.
2. Phân cụm các khoản vay dựa trên thời gian trễ hạn (LongestOverdue) và số tiền còn lại (SoTienConLai).
3. Phân cụm khách hàng dựa trên điểm tín dụng (TS\_CREDIT\_SCORE\_V2) và thu nhập hàng tháng (Salary).
4. Phân cụm địa chỉ cư trú của khách hàng để xác định khu vực rủi ro cao trong việc cho vay.
5. Phân cụm các công ty có nhân viên vay vốn để xác định ngành nghề có rủi ro cao.

link download dữ liệu:

<https://drive.google.com/file/d/1IF1_2cCXOySoGCMX7ihf0YJjB4Gj8DdQ/view?usp=sharing>