**Credit Scoring Analytics for Credit Card Product using**

**Logistic Regression Algorithm**

**Link Github : https://github.com/axeltanjung/credit\_scoring**

1. **Introduction & Background**
2. **Introduction**

In the contemporary financial landscape, the effective management of credit risk is imperative for the sustained growth and stability of banking institutions. Amara Bank, a prominent player in the banking sector, acknowledges the pivotal role of its Credit Card Product division in catering to the diverse financial needs of its clientele. To ensure the continued viability and profitability of its credit card services, Amara Bank's Credit Card Product division endeavors to develop a sophisticated credit scoring model using Logistic Regression.

The pervasive nature of credit card usage in the modern economy has underscored the necessity for Amara Bank to accurately assess the creditworthiness of its prospective and existing cardholders. As the banking industry undergoes rapid digital transformation, the demand for efficient and secure credit card services has become more pronounced. Consequently, Amara Bank's Credit Card Product division recognizes the urgency to leverage advanced analytics and statistical modeling techniques to refine its credit evaluation process, thereby minimizing the risks associated with extending credit facilities.

This journal seeks to outline the comprehensive business understanding behind the development of a credit scoring model tailored explicitly for Amara Bank's Credit Card Product division. By elucidating the underlying business context, objectives, constraints, stakeholders, and success criteria, this study aims to establish a holistic framework that aligns the technical intricacies of Logistic Regression modeling with the overarching business imperatives of Amara Bank. Through this comprehensive understanding, the Credit Card Product division endeavors to fortify its risk management practices, enhance customer satisfaction, and sustain a competitive edge in the dynamic credit card market landscape.

**B. Business Context**

The Credit Card Product division of Amara Bank operates within an increasingly dynamic and competitive financial services landscape. As digitalization continues to reshape the banking industry, the demand for convenient, secure, and seamless credit card services has witnessed an unprecedented surge. The division operates in an environment characterized by evolving consumer behaviors, changing regulatory landscapes, and heightened expectations for personalized and user-friendly financial experiences.

Amara Bank's Credit Card Product division acknowledges the significance of maintaining a balanced credit card portfolio to safeguard the institution from potential financial risks and losses. The proliferation of digital payment solutions and the growing preference for cashless transactions have amplified the importance of ensuring the responsible and judicious extension of credit to cardholders. Against the backdrop of increasing incidents of fraudulent activities and identity theft, the division is compelled to enhance its risk management practices to safeguard both the interests of the bank and the financial well-being of its customers.

Furthermore, the competitive landscape within the credit card market necessitates Amara Bank's Credit Card Product division to differentiate its offerings through a combination of attractive rewards programs, competitive interest rates, and, most importantly, a seamless and efficient credit approval process. Thus, the development of an advanced credit scoring model using Logistic Regression serves as a strategic imperative for the division, enabling it to leverage data-driven insights to make well-informed credit decisions while ensuring customer satisfaction and loyalty.

The division recognizes that in the current era of heightened digital connectivity, a robust credit scoring model is not only a tool for risk assessment but also a key enabler for fostering enduring customer relationships. By leveraging advanced analytics and machine learning techniques, the division aims to strike a delicate balance between prudent risk management and the facilitation of seamless and personalized credit experiences for its diverse customer base. This business context underscores the critical role of the credit scoring model in fortifying Amara Bank's position as a customer-centric and technologically adept financial institution within the competitive credit card market.

1. **Business Objective**

The development of an advanced credit scoring model using Logistic Regression within Amara Bank's Credit Card Product division is driven by a multifaceted set of business objectives, aimed at enhancing risk management practices, optimizing credit card approval processes, and fostering a culture of data-driven decision-making.

**Main Objectives:**

* **Minimize Non-Performing Loans (NPL) and Bad Debt Ratio:**

Amara Bank's Credit Card Product division is committed to minimizing the incidence of Non-Performing Loans and reducing the overall Bad Debt Ratio within its credit card portfolio. By implementing a robust credit scoring model, the division aims to accurately assess the creditworthiness of applicants, thereby mitigating the risk of default and delinquency. This objective aligns with the division's strategic focus on maintaining a healthy and sustainable credit card portfolio, ensuring long-term profitability, and preserving the trust and confidence of its stakeholders.

**Metric 1**: Reduction in Non-Performing Loans (NPL) by at least 15% within the first year of implementing the credit scoring model.

**Metric 2**: Decrease in the Bad Debt Ratio by a minimum of 10% within the initial two years post-implementation.

* **Increase Approval Rates and Market Share:**

The division aims to leverage the insights derived from the credit scoring model to streamline and expedite the credit approval process. By accurately identifying creditworthy applicants, the division seeks to increase its approval rates, thereby expanding its customer base and bolstering its market share within the competitive credit card market. This objective reflects the division's commitment to fostering inclusive and accessible financial services while capitalizing on growth opportunities and strengthening its position as a market leader in the credit card segment.

**Metric 3**: Increase in the Credit Card Approval Rate by 20% in the first six months after the deployment of the credit scoring model.

**Metric 4**: Expansion of the Market Share within the credit card segment by 8% within the first year of adopting the credit scoring model.

* **Transition from Manual Lending to Scorecards:**

To enhance operational efficiency and ensure consistency in credit decision-making, the division aims to transition from traditional manual lending practices to automated credit scoring methodologies. By adopting a data-driven approach to credit assessment through the utilization of scorecards, the division seeks to standardize and optimize its decision-making processes, thereby reducing the reliance on subjective evaluations and manual underwriting. This objective underscores the division's commitment to embracing technological advancements and fostering a culture of data-driven decision-making, ensuring greater accuracy, efficiency, and transparency in the credit approval process.

**Metric 5**: Reduction in the average time taken for credit assessment and approval by 30% within the initial three months of transitioning to the credit scoring model.

**Metric 6**: Implementation of the credit scoring model in at least 90% of credit card application evaluations within the first year, facilitating a comprehensive shift from manual lending practices to automated scorecard-based assessments.

By setting these specific target numbers for the key performance metrics, the Credit Card Product division of Amara Bank can effectively gauge the success and impact of the credit scoring model's implementation. These measurable KPIs not only provide a clear benchmark for evaluating the model's effectiveness but also serve as guiding parameters for the division to track its progress in achieving the outlined business objectives and strategic goals.

1. **Deliverables:**

* **Customized Credit Scoring Model:**

Development of a tailored Logistic Regression-based credit scoring model designed specifically for the Credit Card Product division of Amara Bank.

Comprehensive documentation outlining the model's variables, parameters, and underlying assumptions for transparent and auditable credit evaluations.

* **Performance Evaluation Framework:**

Establishment of a robust performance evaluation framework to monitor the effectiveness and accuracy of the credit scoring model in predicting credit risk and minimizing defaults.

Regular performance reports highlighting the model's predictive power, stability, and calibration, enabling continuous refinement and optimization of the scoring model.

* **Implementation Strategy and Training Program:**

Detailed implementation strategy outlining the phased deployment of the credit scoring model within the existing credit card approval process of Amara Bank's Credit Card Product division.

Comprehensive training program for relevant stakeholders, including credit analysts, underwriters, and decision-makers, to ensure seamless integration and effective utilization of the new credit scoring model.

* **Change Management Plan:**

Structured change management plan to facilitate a smooth transition from manual lending practices to automated scorecard-based assessments within the Credit Card Product division.

Communication strategy outlining the key benefits and implications of the new credit scoring model to foster a culture of data-driven decision-making and enhance organizational readiness for technological advancements.

* **Compliance Documentation and Regulatory Alignment:**

Compliance documentation ensuring adherence to relevant regulatory guidelines and data privacy standards in the development and deployment of the credit scoring model.

Alignment with industry best practices and regulatory requirements to foster transparency, fairness, and accountability in the credit evaluation process, thereby instilling trust and confidence among stakeholders and customers.

These deliverables serve as essential components for the successful development, implementation, and integration of the credit scoring model within the Credit Card Product division of Amara Bank, fostering a culture of data-driven decision-making and reinforcing the division's commitment to prudent risk management and customer-centric financial services.

1. **Dataset & Features**

For create the analysis, we use the dataset with features as follows

* + - **Age** : Represents the age of the person
    - **Occupation** : Represents the occupation of the person
    - **Annual\_Income** : Represents the annual income of the person
    - **Monthly\_Inhand\_Salary** : Represents the monthly base salary of a person
    - **Num\_Bank\_Accounts** : Represents the number of bank accounts a person holds
    - **Num\_Credit\_Card** : Represents the number of other credit cards held by a person
    - **Interest\_Rate** : Represents the interest rate on credit card
    - **Num\_of\_Loan** : Represents the number of loans taken from the bank
    - **Type\_of\_Loan** : Represents the types of loan taken by a person
    - **Delay\_from\_due\_date** : Represents the average number of days delayed from the payment date
    - **Changed\_Credit\_Limit** : Represents the average number of payments delayed by a person
    - **Num\_Credit\_Inquiries** : Represents the number of credit card inquiries
    - **Credit\_Mix** : Represents the classification of the mix of credits
    - **Outstanding\_Debt** : Represents the remaining debt to be paid (in USD)
    - **Credit\_Utilization\_Ratio** : Credit\_History\_Age
    - **Credit\_History\_Age** : Represents the age of credit history of the person
    - **Payment\_of\_Min\_Amount**: Represents whether only the minimum amount was paid by the person
    - **Total\_EMI\_per\_month** : Represents the monthly EMI payments (in USD)
    - **Amount\_invested\_monthly**: Represents the monthly amount invested by the customer (in USD)
    - **Payment\_Behaviour** : Represents the payment behavior of the customer (in USD)
    - **Monthly\_Balance** : Represents the monthly balance amount of the customer (in USD)

Source of original dataset can be access through this link:

https://www.kaggle.com/datasets/parisrohan/credit-score-classification/data

1. **Scorecard Development**

The data acquisition process plays a critical role in the development of the credit scoring model for the Credit Card Product division of Amara Bank. The following are the key components involved in the data acquisition stage:

1. **Data Acquisition**

**Application Form:**

The application form serves as a primary source of customer-provided data, capturing essential information such as personal details, employment history, income, and other relevant financial information. This data provides insights into the applicant's financial standing, employment stability, and credit requirements, forming the initial dataset for the credit evaluation process.

**Credit Bureau:**

The Credit Bureau data encompasses information obtained from various external sources, including SLIK (Sistem Layanan Informasi Keuangan), BI Checking (Bank Indonesia Checking), and APPI Checking (Asosiasi Perusahaan Pembiayaan Indonesia Checking). These sources provide comprehensive credit information, including the applicant's credit history, outstanding debts, existing credit accounts, and repayment patterns, enabling a thorough assessment of the applicant's creditworthiness and risk profile.

**Internal System - Behavioral Data:**

Behavioral data obtained from internal systems includes a comprehensive record of the applicant's interactions with Amara Bank, particularly within the Credit Card Product division. This data encompasses the applicant's historical credit card usage patterns, transactional behavior, repayment history, and any instances of repeat or additional orders. The analysis of this data provides valuable insights into the applicant's credit utilization habits and repayment discipline, contributing to a holistic assessment of the applicant's credit risk profile.

**Performance Data:**

Performance data refers to the outcome data associated with each credit application, encompassing critical information such as the applicant's payment history, instances of default, frequency of late payments, and any history of delinquency. This data enables the division to assess the applicant's credit performance and behavior, facilitating the identification of potential high-risk applicants and the establishment of risk mitigation strategies to minimize the incidence of non-performing loans and defaults within the credit card portfolio.

By effectively acquiring and analyzing these diverse data sources, the Credit Card Product division of Amara Bank can construct a comprehensive and multidimensional dataset, laying the groundwork for the subsequent development and implementation of the credit scoring model using Logistic Regression.

1. **Good/Bad Definition**

**Observation/Selection Statuses:**

**Observation Exclude:**

This status involves the exclusion of certain observations from the dataset, primarily to address sample bias issues. These exclusions may arise due to data inconsistencies, data quality issues, or other sample selection biases that could potentially skew the analysis and compromise the accuracy of the credit scoring model.

**Reject:**

The 'Reject' status refers to cases where applicants are not selected for credit approval, as determined by the lender. These cases represent instances where the lender deems the applicant to be ineligible for credit, typically due to factors such as a high-risk profile, poor credit history, or inadequate financial standing.

**Not Taken Up (NTU):**

'Not Taken Up' signifies cases where applicants are selected for credit but do not proceed with the credit facility, despite being approved by the lender. This status often arises when the borrower chooses not to utilize the credit product, even after meeting the lender's criteria for approval.

**Mutual Accept:**

'Mutual Accept' represents cases where the selected applicants have accepted and utilized the credit facility provided by the lender, either responsibly or in a manner that indicates potential credit misuse or abuse.

**Outcome/Performance Statuses:**

**Good:**

The 'Good' status signifies desirable outcomes, such as timely payments, responsible credit utilization, and a positive credit performance, reflecting a borrower's ability to manage credit responsibly and meet their financial obligations effectively.

**Bad:**

The 'Bad' status refers to undesirable outcomes, including instances of default, delinquency, or any other adverse credit events that indicate a high level of credit risk and potential financial instability.

**Indeterminate:**

The 'Indeterminate' status represents outcomes that fall between the 'Good' and 'Bad' categories, indicating a moderate level of credit risk or uncertainty in the borrower's credit behavior. This optional category allows for the nuanced classification of cases that do not distinctly fit into either the 'Good' or 'Bad' classification.

**Excludes:**

The 'Excludes' category encompasses any outcomes that lie outside the intended scope of the credit scoring model, such as operational-risk events or fraudulent activities that may impact the overall risk assessment but are not directly related to the borrower's credit behavior.

In the final scorecard development, the use of the 'Good' and 'Bad' classifications enables the Credit Card Product division to establish clear benchmarks for evaluating the creditworthiness of applicants, facilitating more accurate risk assessments and informed credit decisions.

1. **Performance and Sample Windows**

**Performance Window:**

The performance window refers to the designated time frame during which the credit performance of accounts is continuously monitored and assessed to assign specific performance targets. This window allows for the evaluation of the borrower's credit behavior over a defined period, enabling the identification of patterns, trends, and potential risks associated with the borrower's repayment patterns and credit utilization habits.

**Sample Window:**

The sample window denotes the specific time frame from which known 'good' and 'bad' cases are selected for the sample dataset. This window serves as the basis for selecting historical data to train and validate the credit scoring model, facilitating the identification of key variables and patterns that contribute to the differentiation between creditworthy and high-risk applicants.

Furthermore, there are several approaches to determining the sample and performance windows, including:

**Basel II:**

This approach, as defined by the Basel II framework, emphasizes the use of historical data and specific time frames to assess credit risk and establish adequate capital requirements, ensuring that financial institutions maintain sufficient capital reserves to cover potential credit losses within a defined time horizon.

**"Decision Horizon" Approach from IFRS 9:**

The "Decision Horizon" approach, as outlined by the International Financial Reporting Standards (IFRS) 9, emphasizes the evaluation of the credit performance of accounts within a specific time frame to assess expected credit losses accurately and facilitate effective risk management practices within the banking sector.

**Portfolio Maturity Using Cohort/Vintage Analysis:**

This approach involves the analysis of portfolio maturity through cohort or vintage analysis, enabling the identification of specific cohorts or groups of accounts with similar origination dates to assess their credit performance and behavior over time. This analysis aids in the understanding of the credit risk dynamics associated with different borrower segments within the portfolio.

By leveraging these diverse approaches, the Credit Card Product division of Amara Bank can effectively determine the performance and sample windows, enabling the accurate assessment of credit risk, the identification of key trends, and the development of a robust credit scoring model tailored to the unique dynamics of the credit card market.

1. **Data Sampling**

**Sample Types:**

**Development Set:**

The development set comprises a significant portion, typically 70 to 80 percent, of the total sample, and is used to construct and develop each scorecard within the credit scoring model. This set allows for the iterative refinement and optimization of the model, enabling the identification of key variables and patterns that contribute to the accurate differentiation between 'good' and 'bad' credit applicants.

**Validation Set:**

The validation set constitutes the remaining 20 to 30 percent of the sample and is utilized to independently test and validate the performance and predictive accuracy of the scorecard. This set enables the division to assess the model's generalizability and robustness, ensuring that the developed model exhibits consistent performance and accuracy across diverse datasets and real-world scenarios.

**Sample Sizes:**

According to Anderson, the literature on credit scoring recommends a minimum sample size of:

1,500 'bads'

1,500 'goods'

1,000 'rejects'

These sample size requirements ensure an adequate representation of both high-risk and low-risk credit applicants, as well as rejected applicants, facilitating a comprehensive analysis of the credit risk dynamics within the Credit Card Product division's portfolio.

**Sampling Methods:**

**Simple Random Sample:**

The simple random sampling method involves the random selection of observations from the population, ensuring that each element within the population has an equal chance of being selected for the sample. This method enables the division to obtain a representative sample and minimize potential sampling biases or distortions within the dataset.

**Stratified Random Sample:**

The stratified random sampling method involves the division of the population into distinct strata or subgroups based on specific characteristics or attributes, followed by the random selection of observations from each stratum. This method allows for the representation of diverse borrower segments within the sample, ensuring that the model adequately captures the unique credit risk profiles of different customer segments.

By leveraging these sample types, sample sizes, and sampling methods, the Credit Card Product division of Amara Bank can construct robust and reliable datasets for the development and validation of the credit scoring model, facilitating accurate risk assessments and informed credit decision-making within the dynamic credit card market landscape.

In the process of developing a robust credit scoring model for the Credit Card Product division of Amara Bank, preventing leakage of test set information and conducting exploratory data analysis (EDA) on the training set are critical steps to ensure the model's predictive performance on unseen data. The following elaborates on the key actions taken during the EDA phase:

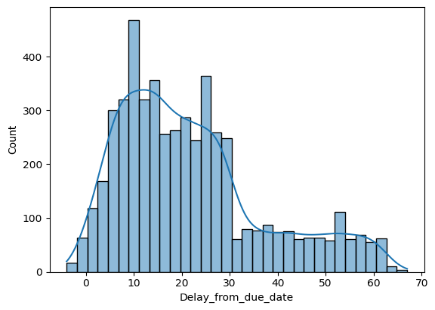
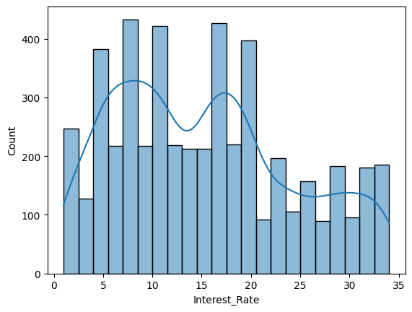
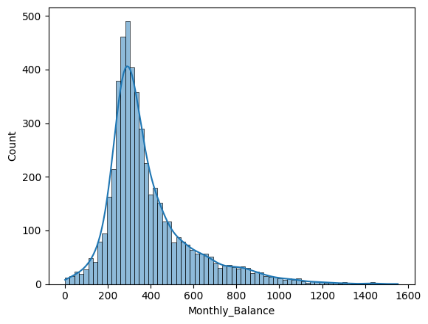
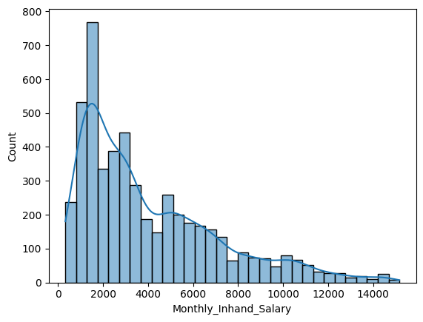
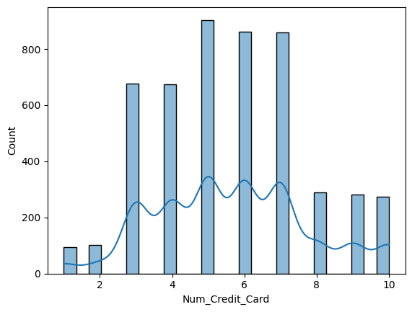
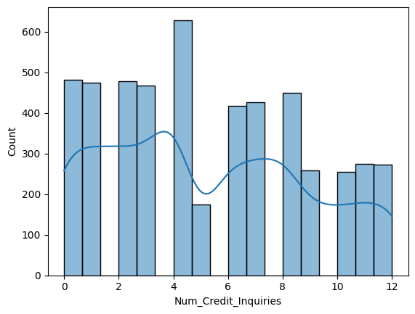
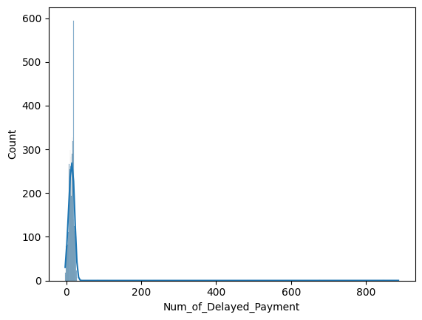
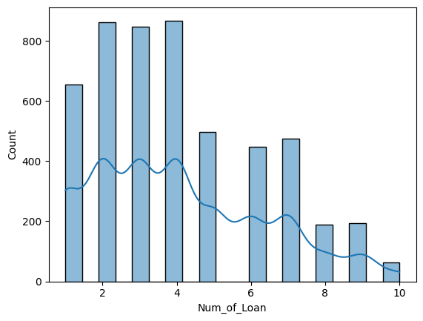
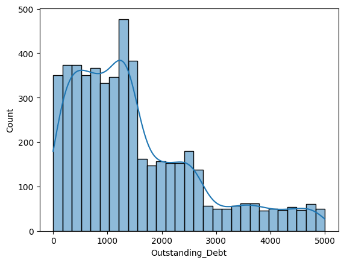
1. **Exploratory Data Analysis (EDA)**

**1. Check Data Integrity:**

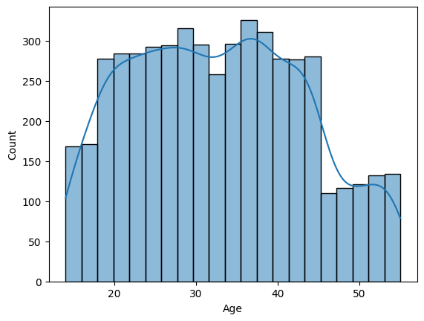
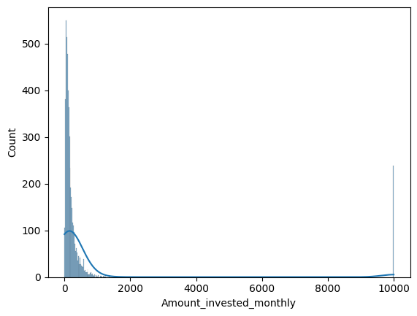
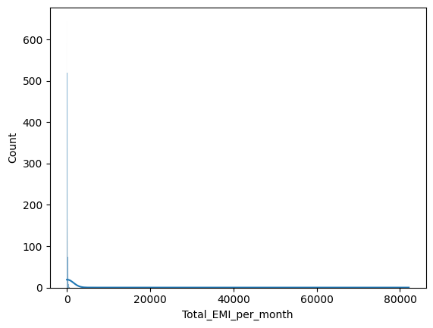
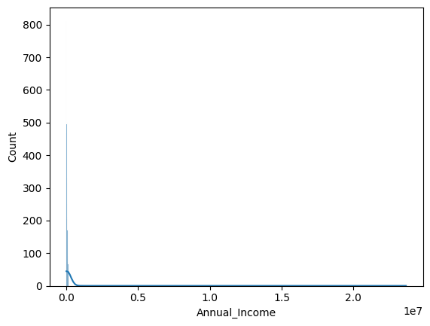
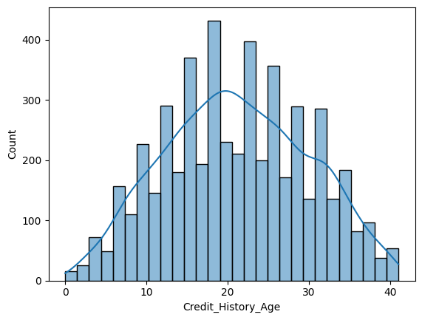
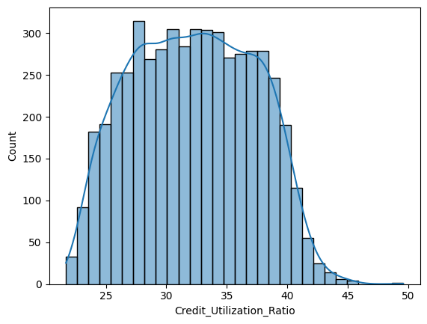
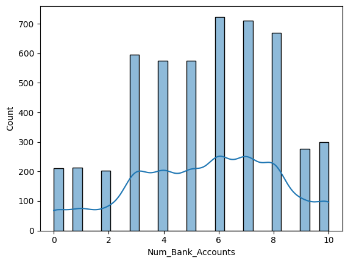
Conducting a thorough check of the data integrity within the training set is crucial to identify and address any potential data quality issues or inconsistencies. This step involves verifying the completeness, accuracy, and consistency of the data, ensuring that the dataset is free from any erroneous or misleading information that could impact the model's performance and reliability.

**2. Check for Any Insights in the Data:**

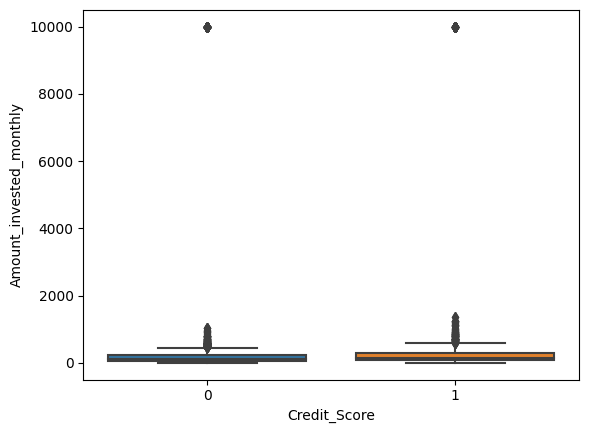
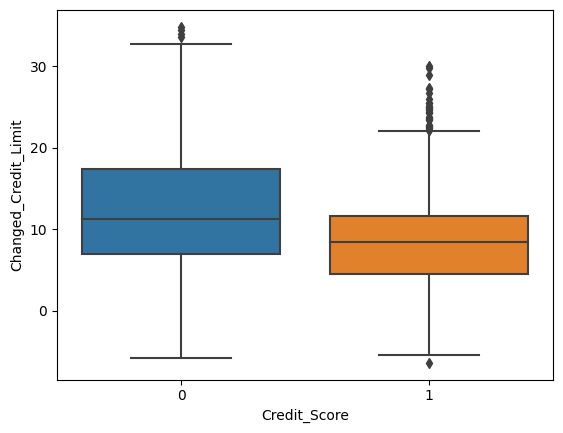
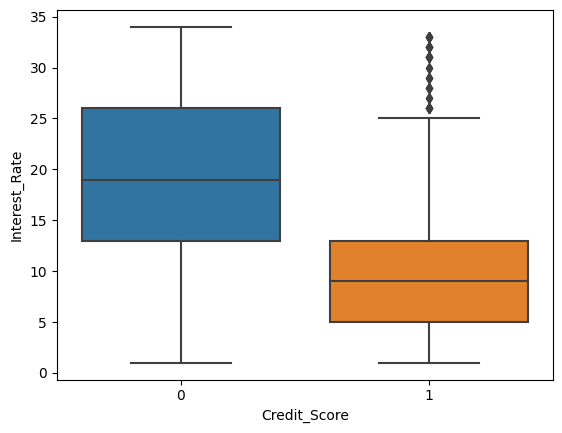
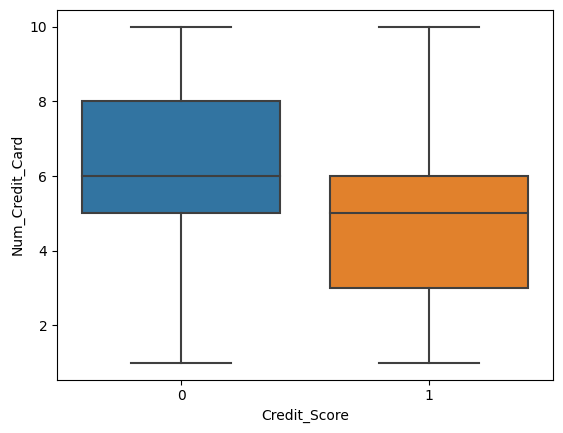
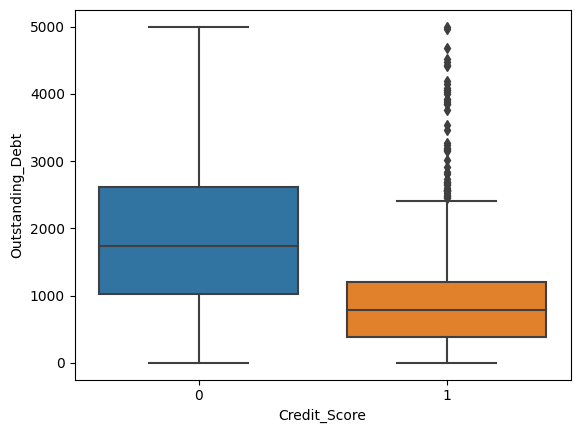
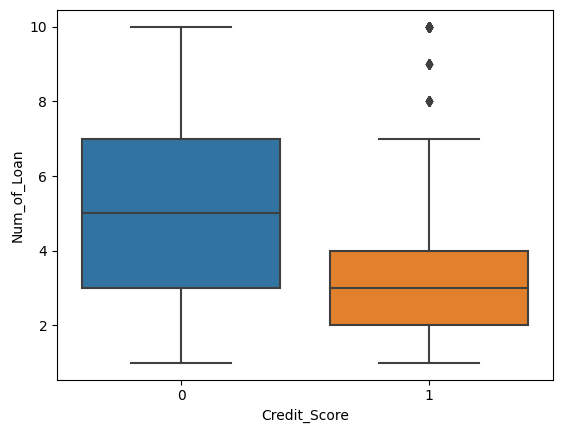
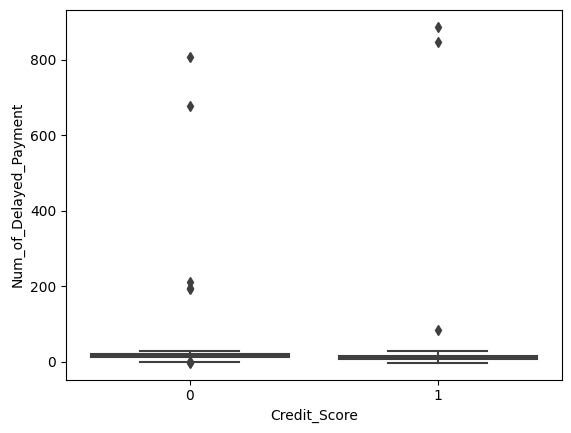
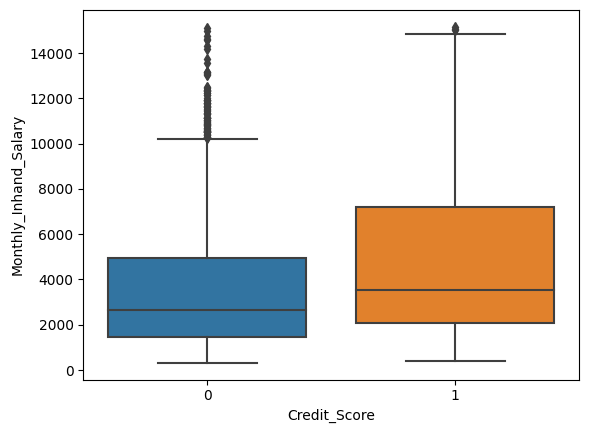
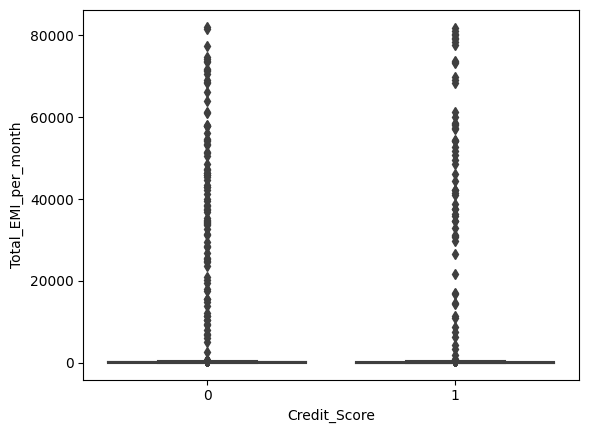
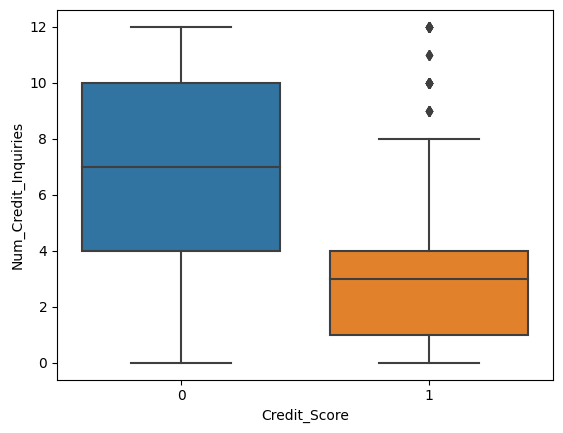
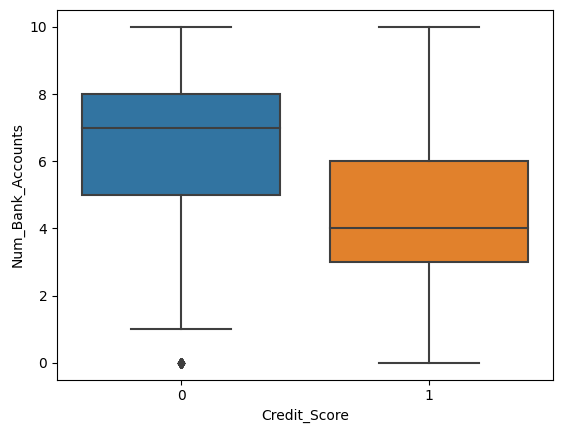
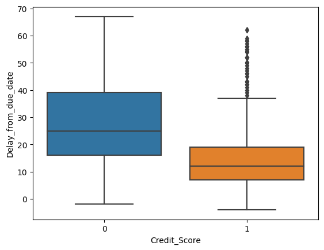
Exploring the distribution, proportion, presence of outliers, and patterns within the training set allows for a comprehensive understanding of the underlying data characteristics. Analyzing the distribution of key variables, identifying any imbalances in the data, detecting outliers, and assessing the extent of missing values provides valuable insights that can guide the subsequent data preprocessing and feature engineering processes.



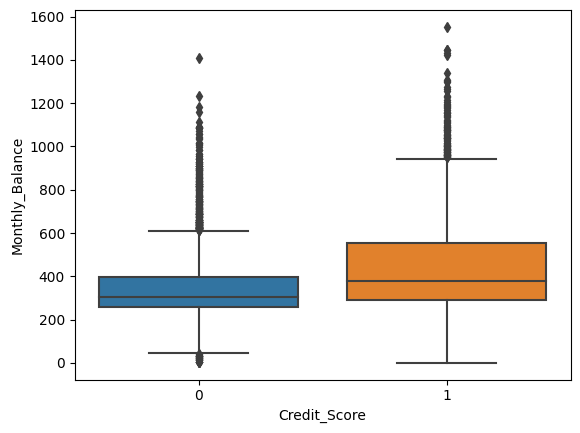
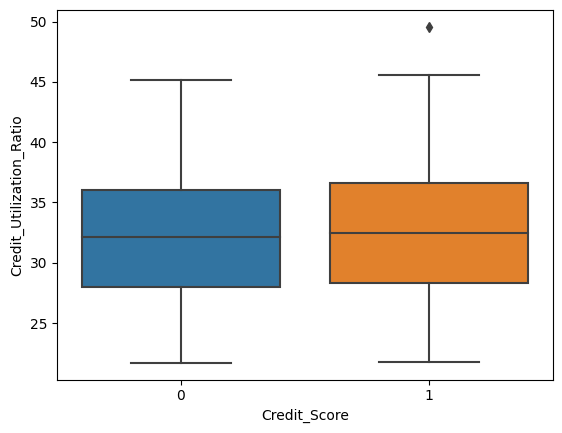
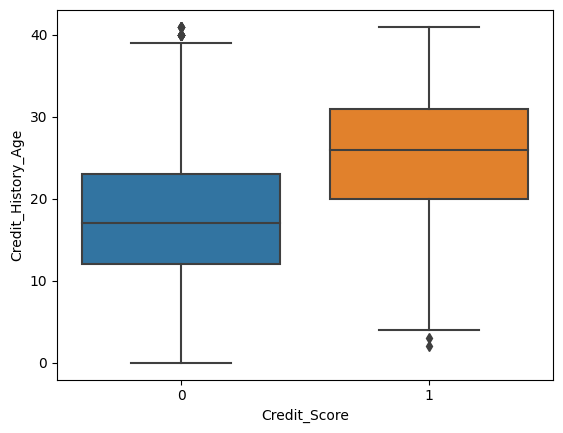
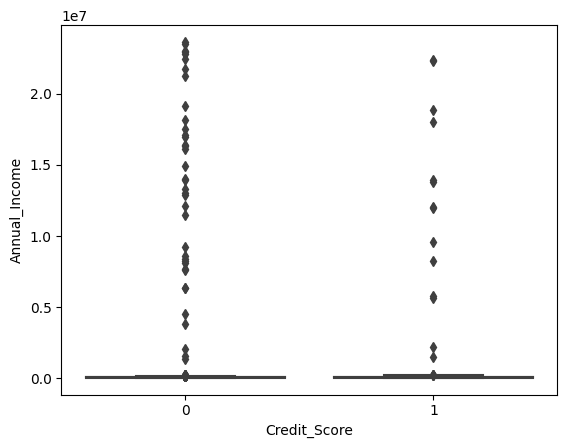
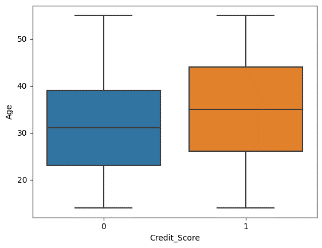
**Distribution Plot of Numerical Predictor**



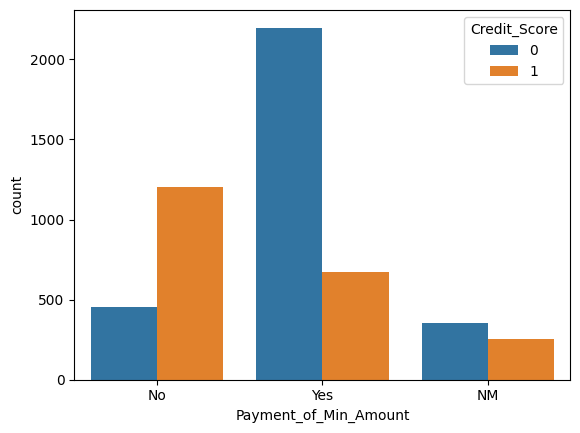
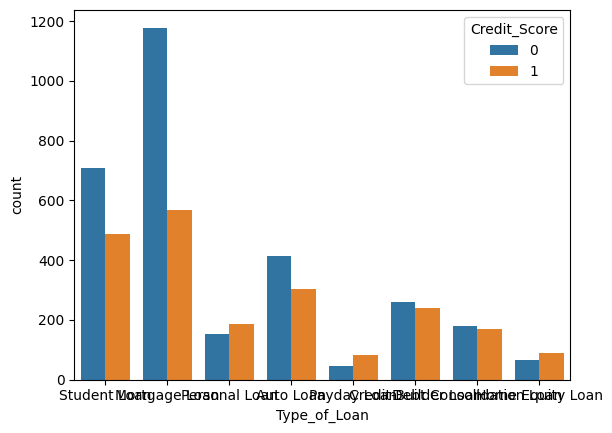
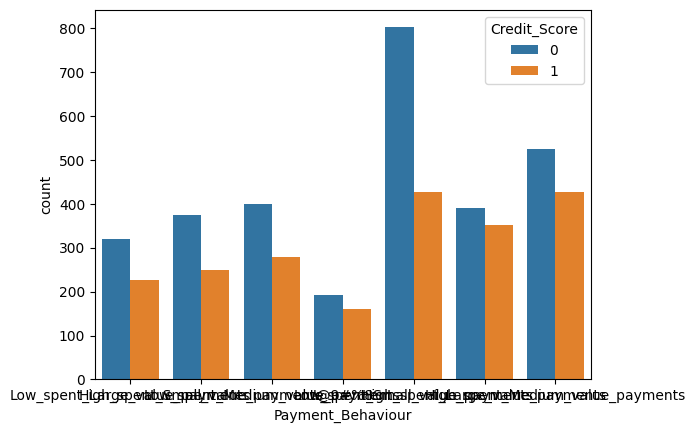
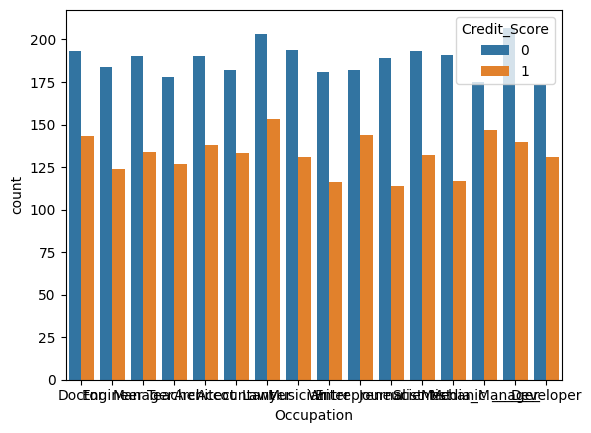
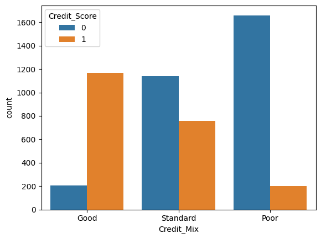
**Distribution Plot of Numerical Predictor**



**Box-Plot of Numerical Predictor**



**Box-Plot of Numerical Predictor**



**Statistical Description of Categorical Columns**

A chart of numbers and a graph

Description automatically generated with medium confidence

**Correlation of Numerical Feature**

**3. Make a Plan for Data Pre-processing:**

Formulating a robust plan for data pre-processing is essential to address data inconsistencies, handle missing values, treat outliers, and standardize the data for effective model training. This plan involves defining the steps for data cleaning, feature scaling, handling missing data through imputation strategies, and implementing data transformation techniques to ensure the data is appropriately structured and optimized for model development.

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

Description automatically generated

**Check missing values Check for data type**

There are missing values in `Age`, `Num\_Bank\_Accounts`, `Num\_Credit\_Card`, `Interest\_Rate`, `Num\_of\_Loan`, `Num\_of\_Delayed\_Payment`, `Num\_Credit\_Inquiries` a categorical variable and numerical.

A screen shot of a computer program

Description automatically generatedA computer screen shot of a program code

Description automatically generated

**Splitting dataset input and output**

1. **Initial Characteristic Analysis**
2. **Binning Numerical Characteristics**

Naeem Siddiqi's recommendation advocates for the binning of numerical characteristics into approximately 20 equal groups, each comprising approximately 5 percent of the total accounts. This approach enables the division to categorize and group numerical variables, such as income, credit utilization, or credit scores, into discrete intervals, facilitating a more intuitive and simplified representation of the data distribution and variability.

**Significance of Optimal Binning:**

**Mitigating Complexity and Overfitting:**

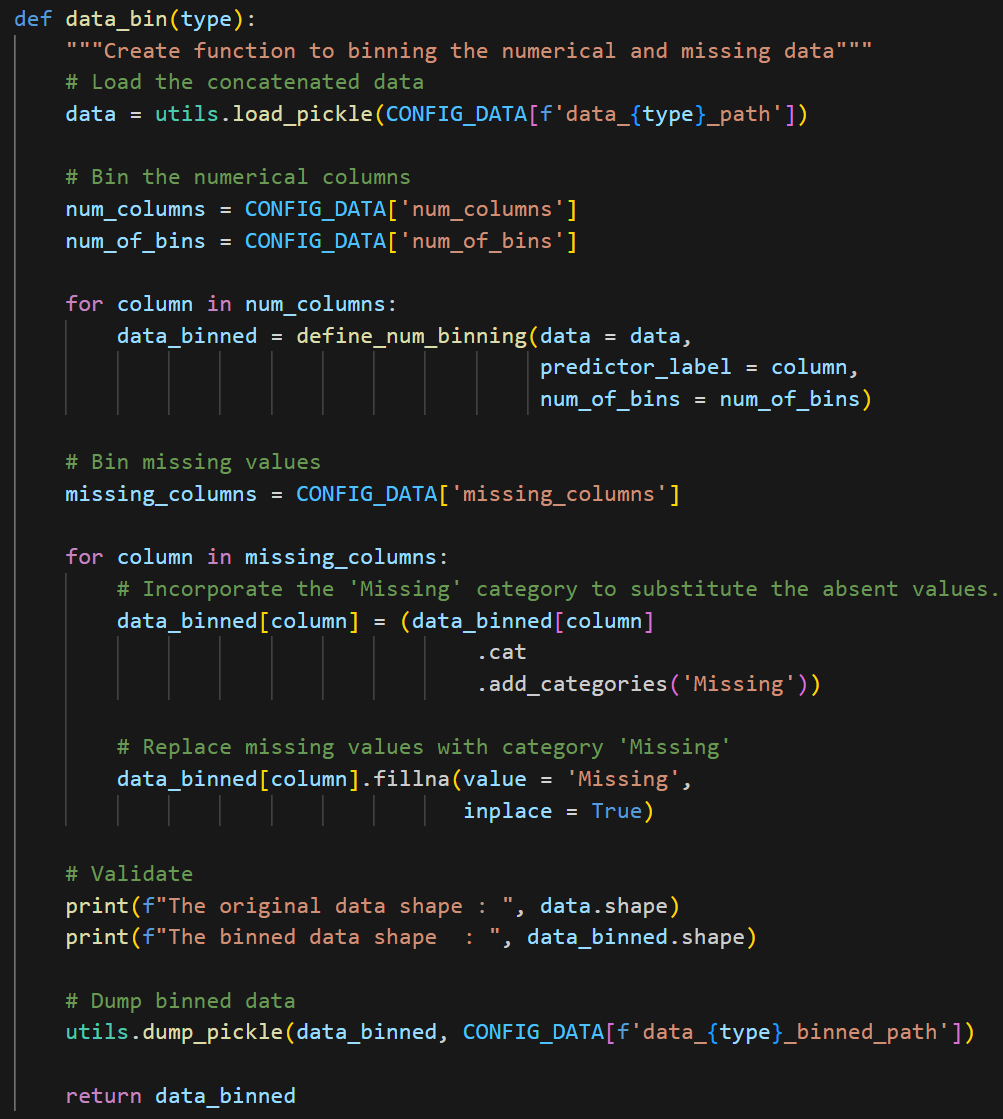
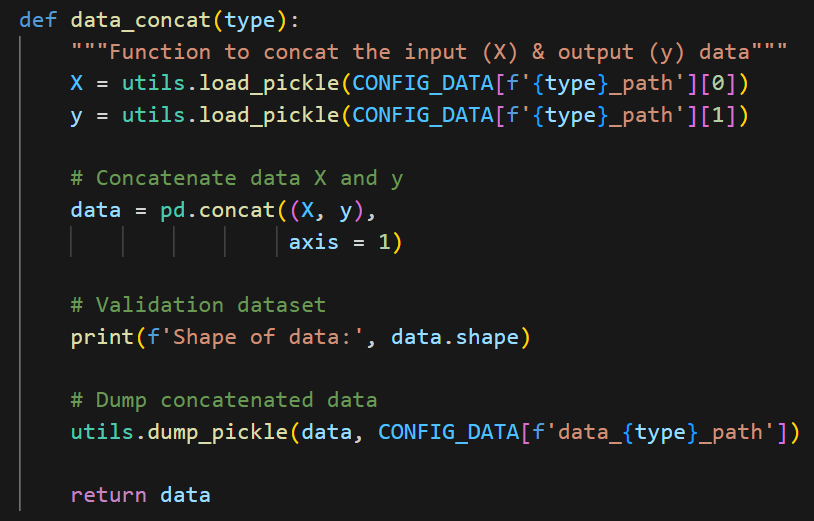
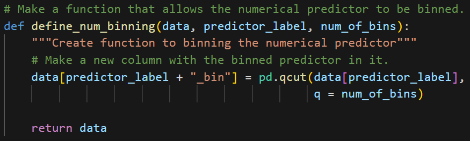
The suggested approach emphasizes the importance of balancing the number of bins to mitigate the risk of model overfitting. By limiting the number of bins to a reasonable count, the division can prevent the model from capturing noise or irrelevant patterns within the data, ensuring that the model's predictive capabilities remain focused on the essential information necessary for accurate credit risk assessment.

**Preserving Information:**

The approach highlights the need to avoid an excessively low number of bins, as this can lead to a loss of valuable information and intricacies present within the dataset. By maintaining an optimal number of bins, the division can effectively preserve the essential characteristics and patterns embedded within the data, enabling the model to capture the underlying nuances of the credit risk landscape and make informed and accurate credit decisions.

By leveraging this binning approach, the Credit Card Product division of Amara Bank can effectively simplify the representation of numerical characteristics, reduce data complexity, and enhance the interpretability of the credit scoring model, while preserving the critical information necessary for robust risk assessment and informed credit evaluation. This approach contributes to the development of a more streamlined and reliable credit scoring model tailored to the unique dynamics of the Credit Card Product division's customer base.

Using data\_concat function, input (X) and output (y) has been concatenaned. After that, create function to binning the numerical predictors and fill the null value with missing data with define\_num\_binning and data\_bin function.



**B. Understanding the Weight of Evidence (WoE):**

**1. Measure of Predictive Power:**

WoE quantifies the predictive power of each attribute by assessing its ability to differentiate between 'good' and 'bad' credit applicants. It calculates the relative likelihood of a specific attribute value leading to either a positive or negative outcome, enabling the division to identify the attributes that significantly influence the credit risk assessment process.

**2. Assessment of Relative Risk:**

WoE facilitates the assessment of the relative risk associated with different attributes within a characteristic. By assigning a numerical value to the relative risk of each attribute, the technique enables the division to prioritize the attributes that have a more substantial impact on the overall credit risk profile, allowing for a more accurate and targeted evaluation of credit applicants.

**3. Contribution to Model Development:**

WoE serves as a fundamental component in the development of the credit scoring model, aiding in the selection of relevant attributes and the construction of the model's predictive framework. By analyzing the WoE values of various attributes, the division can identify the most influential factors driving credit risk, enabling the model to incorporate these attributes effectively and enhance its discriminatory power and accuracy in assessing creditworthiness.

A white background with black text

Description automatically generated

1. **Understanding Information Value (IV)**

* **Comprehensive Strength Measurement:**

IV enables the comprehensive measurement of the total predictive strength of a characteristic, considering all possible values and their respective impact on the predicted outcome. By aggregating the predictive power of individual values within the characteristic, IV provides a holistic assessment of the characteristic's ability to differentiate between 'good' and 'bad' credit applicants, contributing to the overall robustness of the credit scoring model.

* **Quantification of Predictive Relationship:**

IV quantifies the strength of the relationship between the characteristic and the predicted outcome, offering a numerical representation of the characteristic's ability to contribute to the accurate assessment of credit risk. It considers both the magnitude and direction of the relationship, enabling the division to identify the characteristics that significantly influence the model's predictive accuracy and discriminatory power.

* **Contribution to Model Assessment:**

A math equations on a white background

Description automatically generatedIV plays a pivotal role in the evaluation and assessment of the credit scoring model's efficacy, facilitating the identification of the most impactful characteristics and their relative importance in the credit risk evaluation process. By analyzing the Information Value of various characteristics, the division can prioritize the key drivers of credit risk and refine the model's predictive framework to enhance its accuracy and reliability in predicting creditworthiness.

To evaluate each trait's strength separately as a predictor of credit performance. Create function define\_crosstab\_list that (contingency table) needed for the computation of WOE and IV in training data only.

A screen shot of a computer program

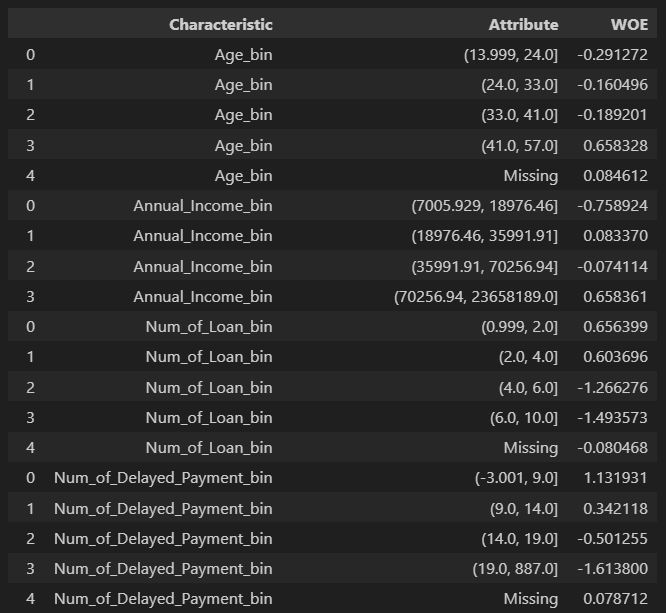
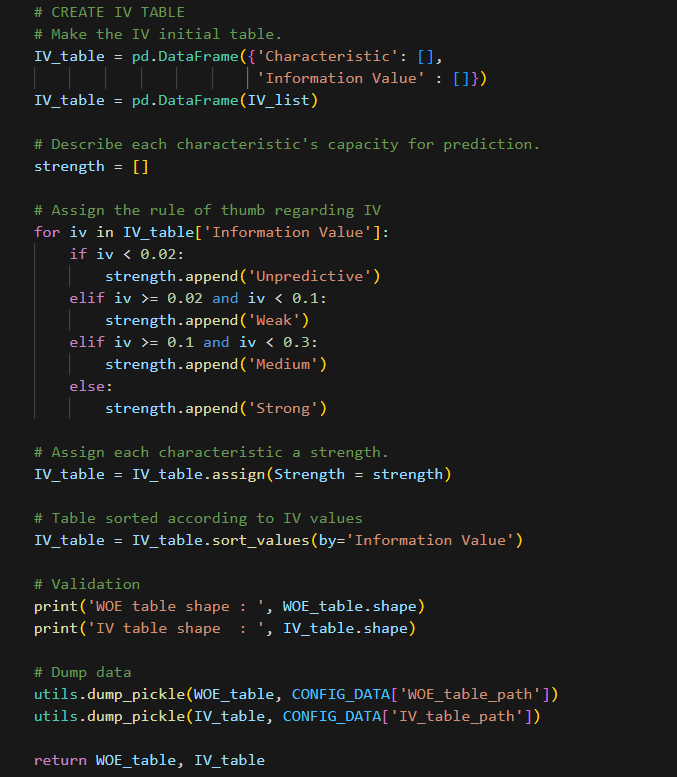
Description automatically generated

A black screen with white dots

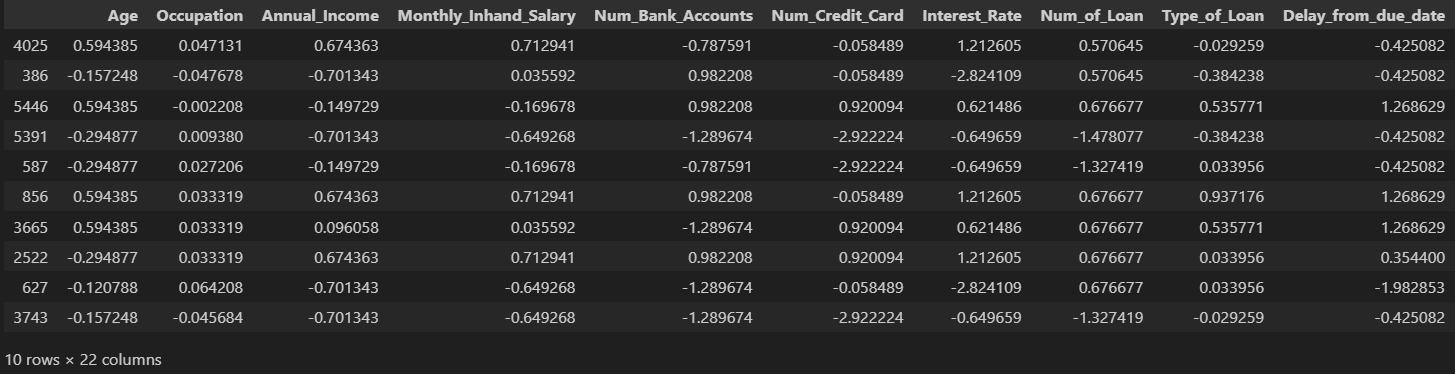
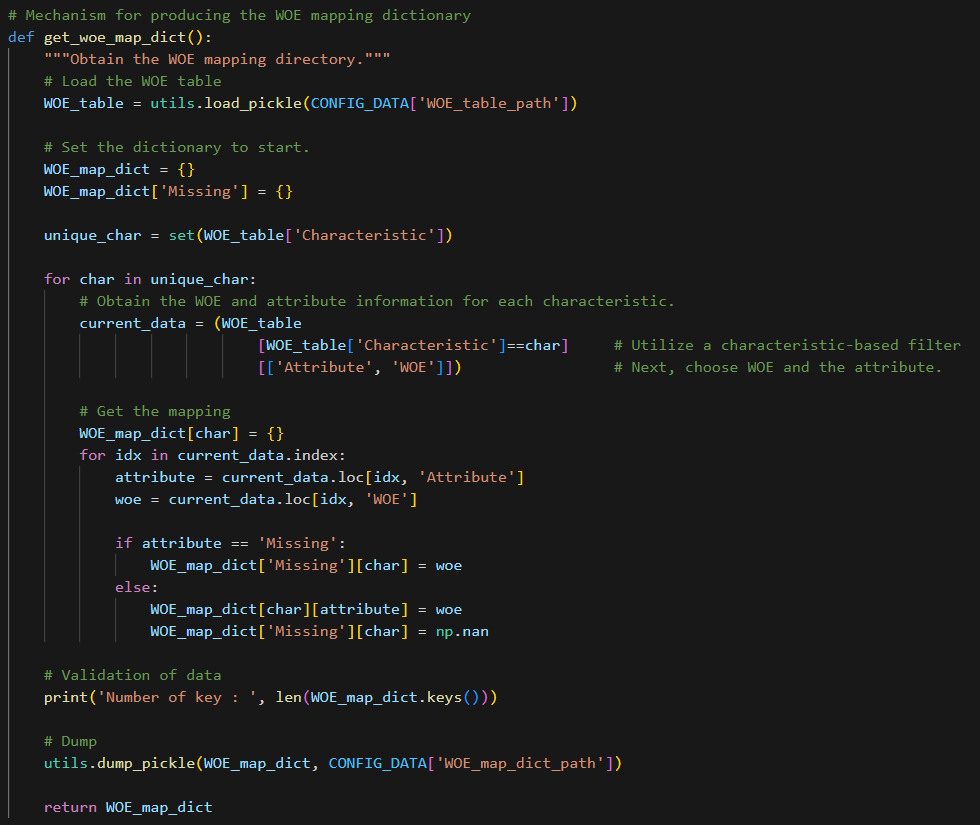
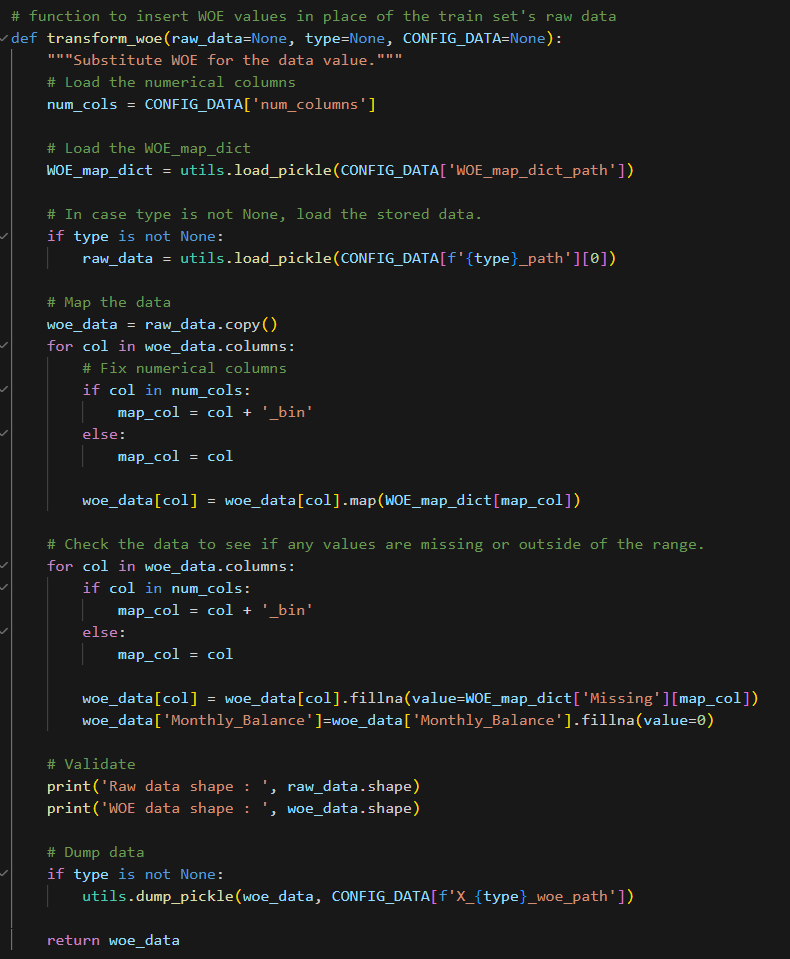
Description automatically generated

**Example of crosstab\_list**

After that, create function to calculate Information Value and Weight of Evidence as follows:



**Value of WoE and IV for each Characteristic**



**Create WoE mapping dictionary & Insert WOE values to the train set's raw data**

1. **Importance of Business-Based Approach (Logical Trend):**
   1. **Ensuring Sensible Weightings and Scores:**

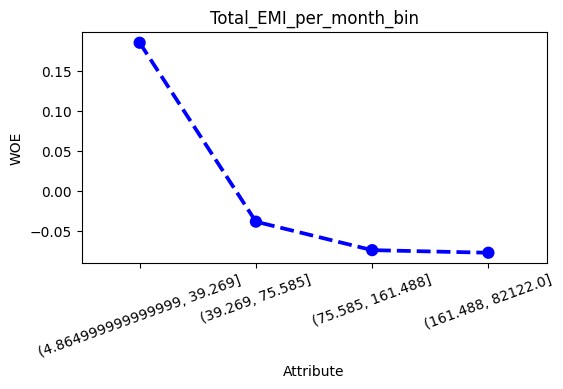
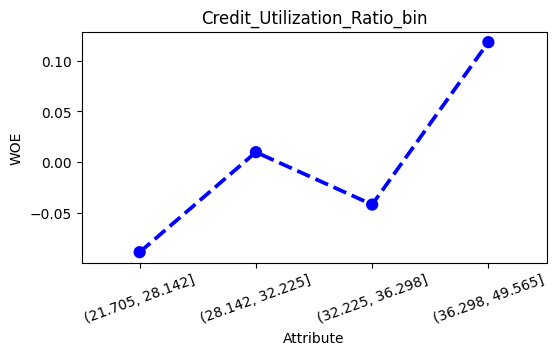
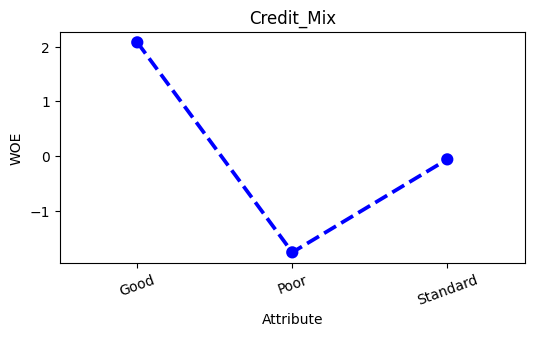
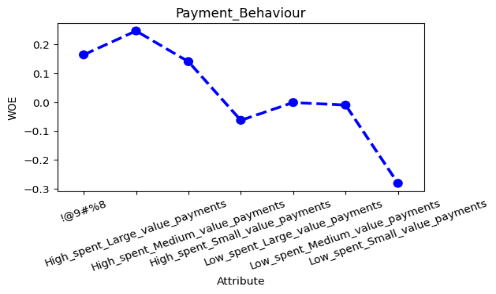
By incorporating a business-based approach, the division can verify that the final weightings and scores derived from the regression analysis align with logical trends and operational sensibilities within the banking industry. This alignment ensures that the model's outputs resonate with the practical realities of credit assessment, facilitating the development of a more intuitive and reliable credit scoring framework.

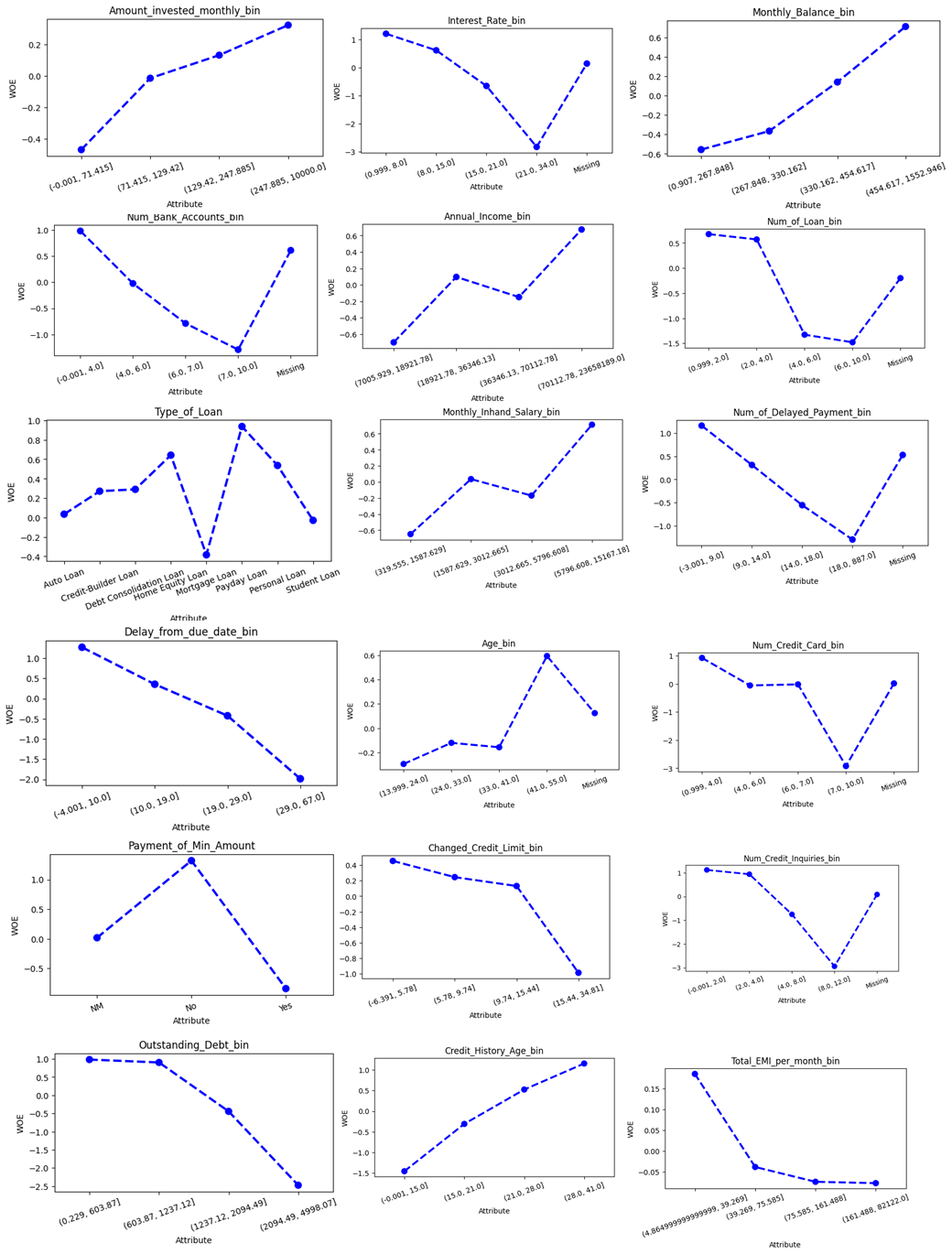
* 1. **Securing Buy-In from Internal End Users:**

The business-based approach serves to secure buy-in from internal end users, including risk managers, adjudicators, and other key stakeholders involved in the credit evaluation process. By integrating business expertise and insights into the model development, the division can foster a shared understanding and acceptance of the model's methodology and outputs, fostering a collaborative and informed approach to credit risk management within the organization.

* 1. **Validating Business Experience and Expertise:**

The business-based approach goes beyond a purely statistical evaluation, emphasizing the validation of business experience and expertise in guiding the development of the credit scoring model. By leveraging the collective knowledge and insights of industry experts and stakeholders, the division can ensure that the model's design and outputs reflect the nuanced dynamics of the credit card market, facilitating a more comprehensive and contextually relevant assessment of credit risk.





T

* **Test of Independence**

Assume that the operational/business sense and logical trend are supported by the WOE for each attribute in all characteristics. The Information Value (IV) of each trait can then be used to determine how well it predicts credit performance. Features range in their ability to predict outcomes from weak to nonexistent. Before modeling, we next determine how each characteristic and the response variable Credit\_Score are dependent on one another to determine whether the weak and unpredictable trait is unrelated to the default probability.

Create a funtion to perform independence test.

A screenshot of a computer

Description automatically generated

A screenshot of a black and white table

Description automatically generatedA computer screen shot of a program

Description automatically generated

1. **Design Scorecards**

All characteristics except Credit\_Utilization\_Ratio\_bin and Occupation are not independent of the response variable (probability of default).

* + - **Preliminary Scorecard**

**Initial Model Prototype:**

The Preliminary Scorecard serves as an initial model prototype, outlining the key variables, weightings, and scoring parameters that contribute to the assessment of credit risk. This version of the scorecard provides a foundational structure for evaluating the predictive capacity of the model and identifying potential areas for refinement and enhancement during subsequent stages of model development. For this project, we use Logistic Regression model to create credit scoring analysis.

**2. Iterative Refinement Process:**

The Preliminary Scorecard undergoes an iterative refinement process, where the division continuously evaluates the performance and effectiveness of the model based on real-world data and feedback from stakeholders. This iterative approach enables the division to identify and address any discrepancies, inconsistencies, or areas of improvement within the scorecard, ensuring that the final version of the model reflects the most accurate and reliable representation of credit risk within the credit card portfolio.

**3. Alignment with Business Objectives:**

The Preliminary Scorecard is designed to align with the strategic business objectives of Amara Bank's Credit Card Product division, ensuring that the model's variables and scoring parameters reflect the division's risk management priorities, customer-centric approach, and operational requirements. This alignment enables the division to develop a scorecard that not only accurately assesses credit risk but also supports the division's broader business goals and objectives within the competitive credit card market.

* + - **Modelling**

For create credit scoring model using Logistic Regression Algorithm, we performing forward selection procedure and best model adjustment. To fit a model on the train set and determine its CV score from the validation set, define the function `forward()`. And run forward selection to do selection based one every attribute that give higher Information value.

A screen shot of a computer program

Description automatically generatedA screenshot of a computer program

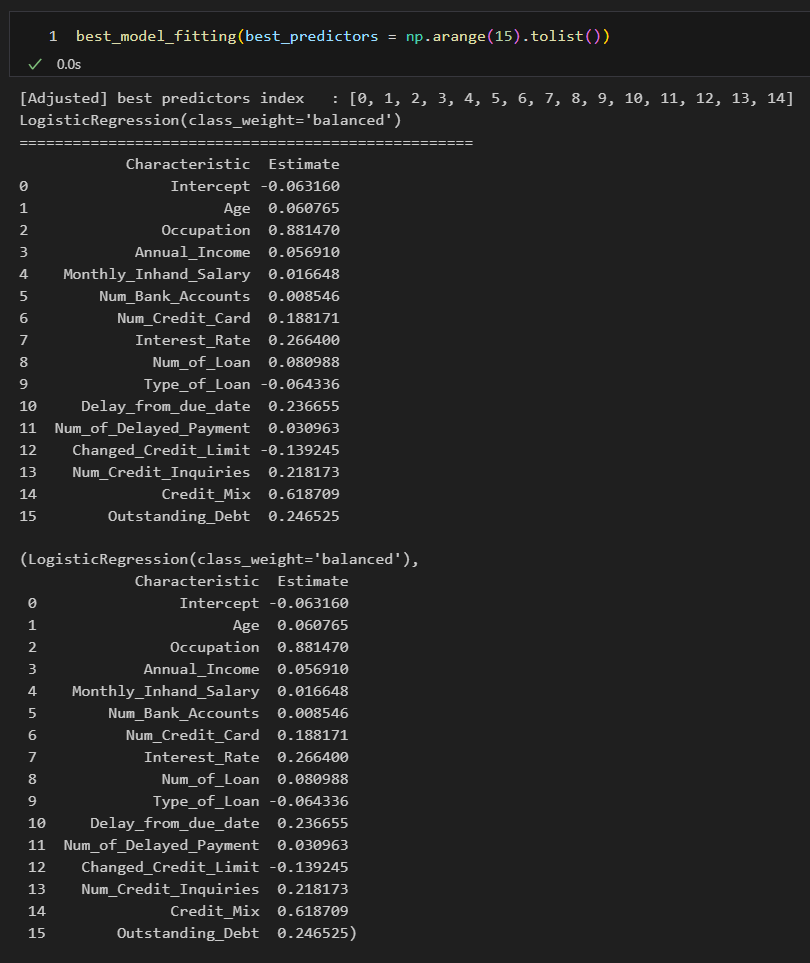
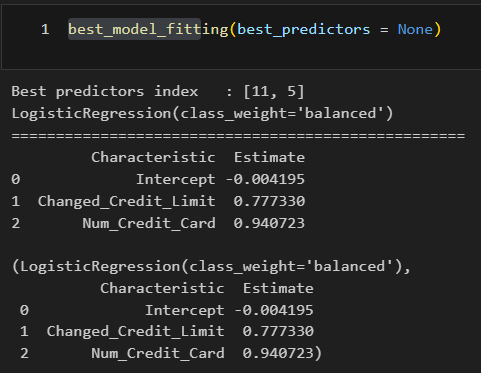
Description automatically generated

A screen shot of a computer program

Description automatically generatedA screenshot of a computer screen

Description automatically generated

Then, create function to fit optimal model across the entire X\_train using function best\_model\_fitting which give return best\_model, and best\_model\_summary with the output and functions as follows:



A graph of a graph

Description automatically generatedThen we can using the ROC AUC curve to see the performance of the model by AUC value and TPR and FPR rate. Apply that calculation for train data and test data to get the performance and make a comparation. Also we can use non-parametric testing Kolmogorov Smirnov to determine wether two distributions differ, or whether an underlying probability distribution differes from a hypothesized distribution

A graph of a curve

Description automatically generated

**ROC AUC Curve and KS Statistic Plot for Data Training**

* + - **Best Model Adjustment**

Too-simple scorecards typically cannot stand the test of time because:

* They are easily affected by slight alterations in the applicant profile.
* An adjudicator of quality would never base their decision on merely two features from an application form.

Every feature will be present in the finished model.

* The independence test indicates that no attribute is independent of the response variable (default probability).
* Typically, a final scorecard has eight to fifteen characteristics.
  + - **Scaling and Final Scorecard Production**

In this chapter, we’ll construct the scorecard, create the dictionary for the points map and estimate the credit score based on an input.

* 1. Create Scorecard

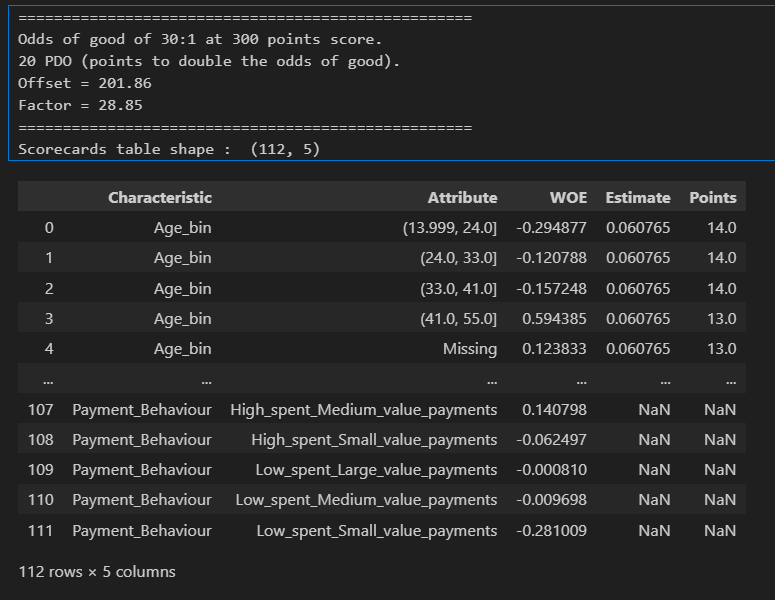
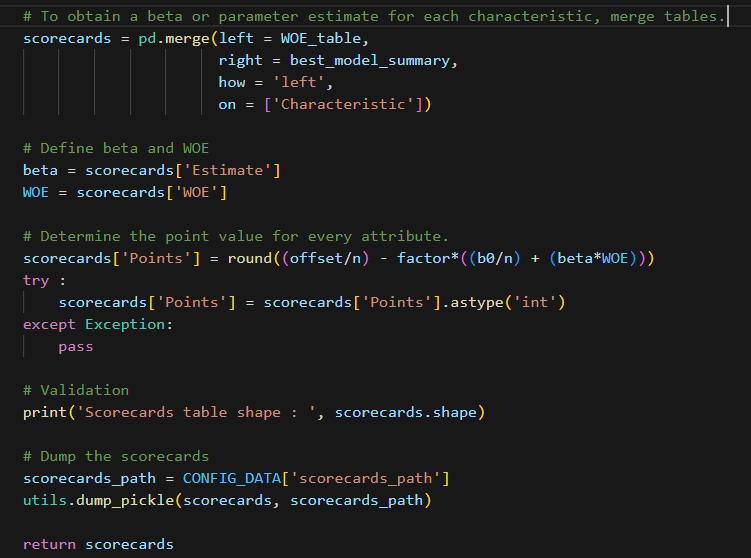
Give a score to each attribute by indicating: - 20 PDO (points to double the odds of good) and - Odds of good of 30:1 at 300 points score.

Thus, we can calculate the offset and factor:

A black background with white text

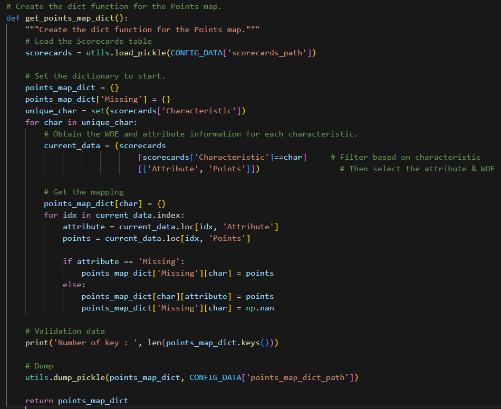
Description automatically generated

Then we create function that transforms the model's output into points as follows:

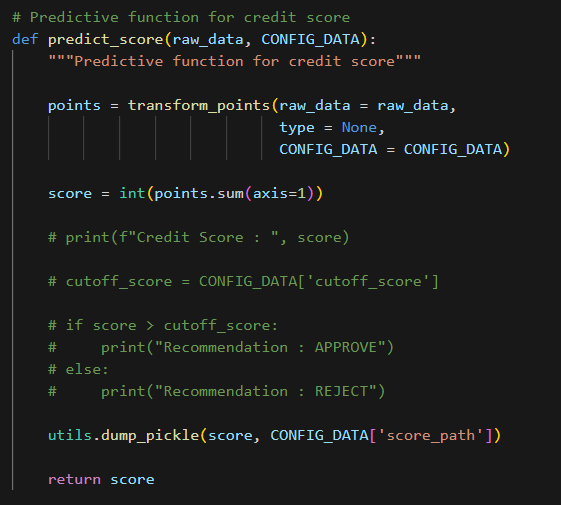


* + - **Predict the Credit Score**

In order to predict the credit score from an input, we must create the points map dictionary in this case.



Next, convert the unprocessed input data into points for scoring.



Next, include a function that determines the credit score using predict\_score function and use test\_input to test the machine learning model and we can get the output of the credit scoring modelling as follows:

A computer screen with text

Description automatically generatedA screenshot of a computer program

Description automatically generated

1. **Implementation Plan**
   1. **Level of Automation**

**Initial Integration Phase**: Integrate the credit scoring model into the existing credit evaluation infrastructure, enabling seamless data flow and automated credit decision-making processes.

**Automation Testing and Validation**: Conduct comprehensive testing and validation of the automated credit evaluation system to ensure accurate data processing, model execution, and consistent decision outputs.

**Enhanced Automation Features**: Implement advanced automation features to streamline credit assessment, improve operational efficiency, and facilitate real-time credit approvals and disbursements for qualified applicants.

A screenshot of a credit score

Description automatically generated

A screenshot of a computer

Description automatically generated

For more convinient, we create the user interface for Risk Officer who want to create a calculation in Scoreboard of credit scoring using FastAPI and Streamlit. For example, the applicant name Bima propose to grant the credit card in Bank Amara. And Risk Officer input the data to the apps to see the output of the Credit Score.

* 1. **Setting Credit Score Cutoff**

**Risk Threshold Analysis**: Analyze historical data and risk tolerance levels to determine the optimal credit score cutoff that aligns with the division's risk management objectives and regulatory requirements.

**Stakeholder Consultation**: Engage with key stakeholders, including risk managers, compliance officers, and senior leadership, to gather insights and inputs for setting the appropriate credit score cutoff that balances risk exposure and business growth objectives.

**Performance Monitoring and Adjustment**: Establish a monitoring mechanism to assess the effectiveness of the credit score cutoff in managing credit risk, and make necessary adjustments based on real-time performance data and market dynamics.

In order to determine the cutoff score, we require data on:

Anticipated approval percentage from:

- The most recent month or quarter's applicant distribution.

Anticipated failure rate derived from either the development sample or the most current sample

– Utilize the performance window for the scorecard.

Since we don't have access to the most recent information on credit applications,

we will:

- Calculate the test set's expected approval rate.

- Calculate the train set's expected bad rate.

A graph with blue lines and white text

Description automatically generatedA screenshot of a black screen

Description automatically generatedA screenshot of a black screen

Description automatically generated

A blue line graph with numbers

Description automatically generated

* 1. **Credit Process Workflow**

**Process Mapping and Redesign**: Map the end-to-end credit evaluation process, identify potential bottlenecks or inefficiencies, and redesign the workflow to incorporate the automated credit scoring model seamlessly.

**Training and Change Managemen**t: Conduct comprehensive training programs for relevant stakeholders to ensure a smooth transition to the new credit process workflow, emphasizing the benefits of the automated credit scoring model and its impact on operational efficiency and risk management.

**Continuous Process Improvement**: Implement a continuous process improvement framework to solicit feedback, monitor workflow performance, and identify opportunities for further optimization and enhancement of the credit evaluation process based on customer feedback and industry best practices.

A diagram of a data flow

Description automatically generated

1. **Conclusion and Recommendation**
   1. **Conclusion from the Scorecard and Analysis:**

Based on the comprehensive analysis and deployment of the credit scoring model within the Credit Card Product division of Amara Bank, the following key conclusions can be drawn:

* + - The implemented credit scoring model effectively assesses the creditworthiness of applicants, enabling the division to make informed and accurate credit decisions while minimizing the incidence of non-performing loans and defaults within the credit card portfolio.
    - The model demonstrates robust predictive capabilities, aligning with the division's risk management priorities, operational requirements, and strategic objectives, thereby reinforcing the division's commitment to prudent risk management and customer-centric financial services.
    - The integration of the credit scoring model has streamlined the credit evaluation process, facilitating efficient and automated decision-making, and enhancing the overall operational efficiency and effectiveness within the Credit Card Product division.
  1. **Recommendations for the User:**

**Risk Manager:**

Continuously monitor the model's performance and effectiveness in managing credit risk, making necessary adjustments to the credit score cutoff and risk thresholds based on real-time data and market dynamics.

**Operational Team:**

Leverage the automated credit scoring system to streamline the credit evaluation process, enhance operational efficiency, and ensure consistent and timely credit approvals and disbursements for qualified applicants.

**Marketing Team:**

Utilize the insights generated from the credit scoring model to develop targeted marketing strategies and tailored credit card offerings that cater to specific customer segments based on their credit risk profiles and financial needs.

* 1. **Recommendation for the Next Project:**

**Enhanced Model Refinement:**

Explore the possibility of incorporating alternative data sources and advanced analytical techniques, such as machine learning algorithms and artificial intelligence, to enhance the predictive power and accuracy of the credit scoring model and further improve its ability to assess credit risk in a dynamic and evolving market landscape.

**Expanded Data Integration:**

Consider integrating additional data sources and variables, such as socio-economic indicators, spending patterns, and behavioral insights, to develop a more comprehensive and nuanced understanding of customer credit behavior and risk management dynamics, enabling the division to refine the model's assessment criteria and decision-making parameters for more accurate and personalized credit evaluations.

1. **References Works**

* Intelligent Credit Scoring – *Naeem Siddiqi*
* The Credit Scoring Toolkit – *Raymond Anderson*
* Credit Scoring, Response Modelling and Insurance Rating – *Steven Finlay*
* An Introduction to Categorical Data Analysis – *Alan Agresti*
* Learning from Imbalanced Data Set – *Alberto Fernandez, S. Garcia, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera*
* Applied Logistic Regression – *David W. Hosmer and Stanley Lemeshow*
* Multivariate Data Analysis – *Joseph F. Hair Jr., William C. Black, Barry J. Babin, Rolph E. Enderson*
* An Introduction to Statistical Learning – *Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani*
* CS 109 – Stanford University