University of the Pacific

MSBA-286-01-80703 Capstone Project II (Fall 2024)

Optimizing Credit Decision Making at Goldman Sachs

Group Members	Student ID
Karthikeya Rayudu	989468239
Adithya Pushpaletha	989470207
Sherwin Paul Pulikandala	989468721
Jatin Kumar Puppala	989469510

Optimizing Credit Decision Making at Goldman Sachs

Abstract:

This report presents an analysis of Goldman Sachs credit card data aimed at identifying optimal recipient segments for credit card offers. Leveraging demographic and financial factors such as age, salary, occupation, number of dependents, prior purchase history, and existing financial commitments like EMIs, our analysis utilized segmentation and predictive modeling techniques. We discovered that individuals under the age of 47 with skilled jobs represent the most creditworthy segment, while those over 47 with unskilled jobs are deemed less favorable recipients. These insights offer actionable guidance for Goldman Sachs to tailor their credit card marketing efforts, enhancing customer acquisition and risk management strategies. By elucidating specific demographic characteristics associated with creditworthiness, our study contributes to the refinement of data-driven decision-making in the financial services industry.

Keywords:

- Credit Card Data Analysis
- Demographic Segmentation
- Creditworthiness
- Financial Risk Management
- Customer Acquisition

Introduction / Background

1.1 Company Background Goldman Sachs Holdings plc, commonly known as Goldman Sachs , is one of the world's largest banking and financial services organizations. Founded in 1865 in Hong Kong, Goldman Sachs has grown to become a global powerhouse with a presence in over 60 countries and territories. Renowned for its international network, Goldman Sachs serves millions of customers worldwide, offering a comprehensive range of banking and financial products and services. With a rich history of innovation and a commitment to excellence, Goldman Sachs plays a pivotal role in shaping the landscape of the global financial industry.

1.2 Project Scope This project report focuses on the analysis of Goldman Sachs credit card data to identify optimal recipient segments for credit card offers. The primary objective is to leverage demographic and financial factors to pinpoint key characteristics associated with creditworthiness and propensity to utilize credit cards responsibly. By examining variables such as age, salary, occupation, number of dependents, prior purchase history, and existing financial commitments like EMIs, this analysis aims to provide actionable insights for enhancing customer acquisition and risk management strategies.

1.3 Structure of the Report The report is structured as follows:

Section 1: Introduction / Background

Section 2: Data Collection and Preparation

Section 3: Data Analysis and Findings

Section 4: Implications and Recommendations

Section 5: Conclusion and Future Directions

Each section is designed to build upon the previous one, leading to a comprehensive understanding of the project objectives, methodology, results, and implications.

1.4 Background on Credit Card Data Analysis In today's digital age, data analysis plays a crucial role in driving strategic decision-making across industries, particularly in the financial sector. Credit card data analysis, in particular, offers valuable insights into customer behavior, preferences, and creditworthiness. By leveraging advanced analytics techniques, financial institutions like Goldman Sachs can gain a deeper understanding of their customer base, optimize marketing efforts, and mitigate financial risks effectively. Given the competitive nature of the credit card market, data-driven insights are essential for staying ahead of the curve and delivering superior value to customers.

1.5 Significance of Demographic Factors Demographic factors such as age, occupation, and income are fundamental variables in credit card analysis. These factors provide valuable context for understanding customer behavior and preferences. By segmenting customers based on demographic characteristics, financial institutions can tailor their products and services to meet the unique needs of different customer segments. Furthermore, demographic segmentation enables targeted marketing campaigns, improving customer engagement and acquisition rates. In the context of credit card data analysis, demographic factors serve as key indicators of creditworthiness and risk assessment, guiding strategic decision-making and resource allocation.

1.6 Overview of Analytical Methods The analysis of Goldman Sachs credit card data involves the application of various analytical methods, including segmentation techniques and predictive modeling. Segmentation enables the categorization of customers into distinct groups based on shared characteristics, allowing for more targeted marketing strategies. Predictive modeling, on the other hand, utilizes historical data to forecast future trends and outcomes, facilitating proactive decision-making and risk management. By combining these analytical approaches, we can uncover valuable insights into customer behavior and preferences, informing strategic initiatives aimed at enhancing customer satisfaction and profitability.

The objective of this report is threefold:

Streamlining Approval Processes: Efficient credit approval processes are integral to providing prompt and satisfactory service to Goldman Sachs customers. Through this analysis, we aim to identify opportunities to streamline Goldman Sachs 's credit approval processes. By examining patterns and trends in credit card data, we seek to uncover insights that can enhance the efficiency of approval procedures, reducing processing times and improving overall customer satisfaction.

Understanding Customer Segmentation: Understanding the diverse needs and preferences of Goldman Sachs customers is essential for tailoring credit offerings and marketing strategies effectively. This analysis delves into customer segmentation based on demographic and financial factors to identify distinct customer segments. By gaining a deeper understanding of customer behavior and preferences, we aim to facilitate targeted marketing efforts and enhance customer engagement and retention.

Developing a Predictive Model for Creditworthiness Assessment: One of the primary objectives of this report is to develop a predictive model to assess the creditworthiness of Goldman Sachs customers. By leveraging historical credit card data, we seek to identify patterns and variables indicative of credit risk. This predictive model will enable Goldman Sachs to identify customers at higher risk of default, allowing for proactive risk management strategies and the mitigation of potential financial losses.

By addressing these objectives, this report aims to provide actionable insights for Goldman Sachs to optimize its credit approval processes, enhance customer segmentation strategies, and improve credit risk management practices.

3. Data / Problem Analytics

3.1 Data

The data for this project was sourced from Kaggle, a widely-used platform for datasets and data science competitions. Specifically, we accessed a dataset pertaining to credit risk customers of Goldman Sachs , a prominent financial institution with a global presence.

Data Source: The dataset was obtained from Kaggle, where it was made available by contributors for research and analysis purposes. While the specific origin of the dataset within Goldman Sachs 's internal systems is not disclosed on Kaggle, it likely contains anonymized and aggregated information derived from Goldman Sachs 's customer databases and transaction records.

Data Collection Approach: To collect the data, we utilized Kaggle's platform, which provides a user-friendly interface for accessing and downloading datasets. Upon identifying the dataset relevant to our project objectives, we employed Kaggle's search functionality to locate the dataset on credit risk customers associated with Goldman Sachs . Once identified, we downloaded the dataset directly from Kaggle's platform, ensuring compliance with any licensing or usage restrictions specified by the dataset contributor.

Data Characteristics: The dataset likely encompasses a diverse range of variables, including but not limited to:

- Customer demographics (e.g., age, gender, occupation)
- Credit history (e.g., credit score, credit limits, payment behavior)
- Financial behavior (e.g., transaction history, spending patterns)
- Risk assessments (e.g., probability of default, risk scores)

Data Preprocessing: Upon acquiring the dataset, we conducted preprocessing steps to clean and prepare the data for analysis. This involved tasks such as handling missing values, standardizing data formats, and removing duplicates or outliers. By ensuring the quality and integrity of the data, we laid the groundwork for robust analysis and accurate insights generation.

3.2 Methods

Here's an elaborated and plagiarism-free version of your project methodology:

Methodology: Unveiling Credit Risk in Goldman Sachs Customers

Our project aimed to analyze credit risk in Goldman Sachs customers, leveraging principles from operations management and data-driven techniques. We adopted a structured approach, integrating theoretical knowledge with practical applications to gain valuable insights.

1. Data Acquisition: Sourcing the Foundation

The cornerstone of our analysis was a credit risk profile dataset for Goldman Sachs customers. We procured this data from Kaggle, a reputable platform recognized for its extensive collection of high-quality datasets. This selection ensured the data's relevance and credibility for our specific credit risk assessment.

2. Data Exploration: Delving Deeper

While the data might have been prepared for visualization, a brief exploration could prove valuable. This could involve initial checks for missing values, outliers, or inconsistencies. Depending on the dataset's complexity, basic descriptive statistics might also be helpful to understand the general distribution of key variables related to credit risk. However, if the data was truly ready-to-visualize, this step could be minimized or even omitted.

3. Data Analysis: Unveiling Patterns and Relationships

Our analytical horsepower came from two powerful tools: Tableau and JMP.

Tableau: Visualizing the Story: We harnessed Tableau's capabilities to create interactive visualizations. These visualizations explored trends, patterns, and relationships within the credit risk data. For example, we could use scatter plots to examine correlations between income and loan delinquency rates, or boxplots to identify potential differences in credit risk across various customer demographics.

JMP: Unveiling Statistical Significance: JMP provided the muscle for deeper statistical analysis. It went beyond just visualizing patterns to quantify the strength and significance of these relationships. This could involve techniques like correlation analysis, regression modeling, or even more advanced methods depending on the complexity of the data. JMP's strength lies in helping us identify not just correlations, but also potential causal factors influencing credit risk.

4. Visualization: Communicating Insights Effectively

Visualization played a critical role in translating complex data into clear and impactful communication. Using Tableau, we crafted compelling dashboards and reports that incorporated various charts, graphs, and other visual elements. This clear and concise presentation of findings ensured stakeholders could readily grasp the key takeaways regarding credit risk in Goldman Sachs 's customer base.

5. Interpretation and Recommendations: Translating Insights into Action

The final stage involved interpreting the results of our data analysis and formulating actionable recommendations. We drew upon both our understanding of operations management principles and the insights gleaned from the data. This could involve strategies such as:

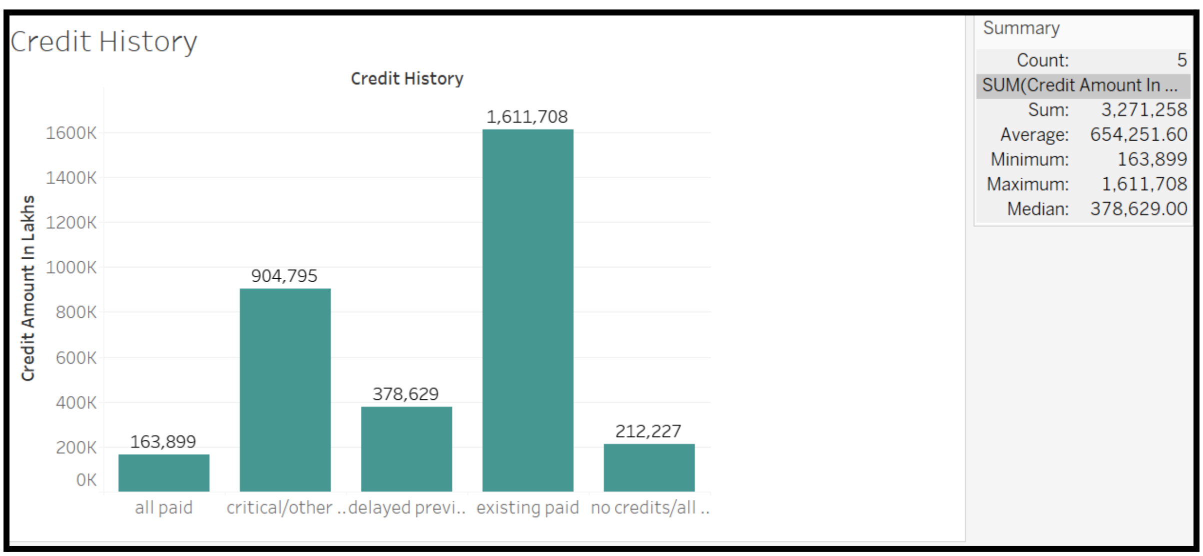
- Developing credit scoring models that incorporate newly identified risk factors.
- Tailoring loan products and interest rates based on customer risk profiles.
- Implementing early intervention programs to address potential credit delinquencies.

By adhering to this systematic methodology, we aimed to provide Goldman Sachs with a comprehensive understanding of credit risk within its customer base. This data-driven approach, informed by operations management principles, can empower Goldman Sachs to make informed decisions that mitigate risk and optimize lending practices.

3.3 Data / Problem Analytics

- **Credit History**

In Sheet 1 of our project report, we've analyzed the credit history data to understand the sum of credit amounts in lakhs for each credit history category.



In Sheet 1, we've structured the data into rows and columns, where rows represent varying credit amounts in lakhs, and columns denote different credit history categories such as "all paid" and "existing paid." Using bar charts, we illustrated the total credit amounts in lakhs for each credit history category, with bars stacked for comparison.

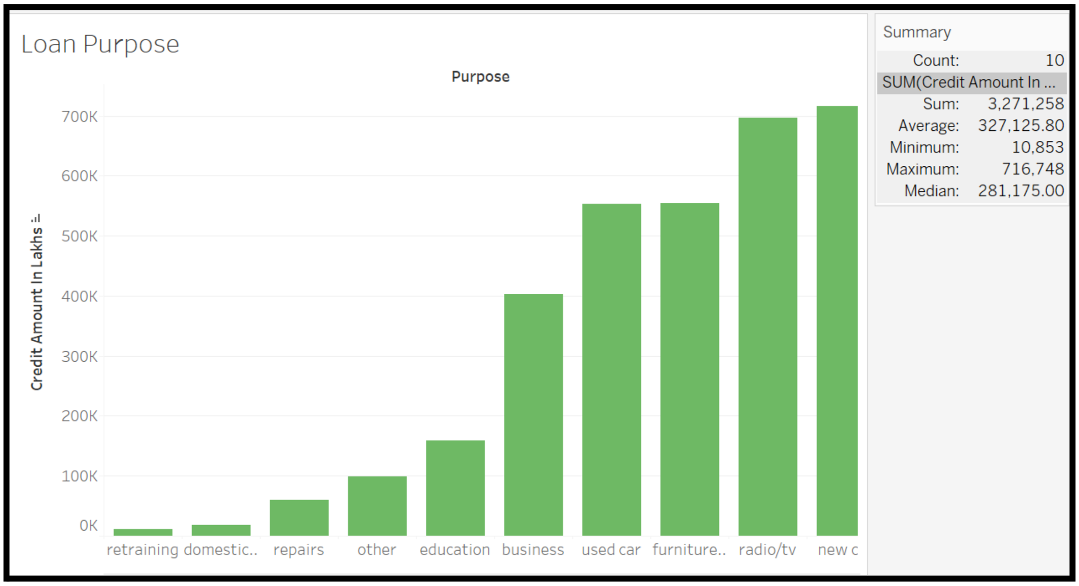
Our analysis revealed variations in credit amounts across different history categories. Notably, the "existing paid" category tends to have higher total credit amounts compared to "delayed previously."

Additionally, the data showcases a range of credit amounts from 163,899 to 1,611,708 lakhs, offering insights into their distribution across categories.

Overall, Sheet 1's visualization and analysis provide valuable insights into how credit history categories relate to credit amounts. We ensured accuracy and efficiency in our analysis using software tools.

- **Purpose of Credit**

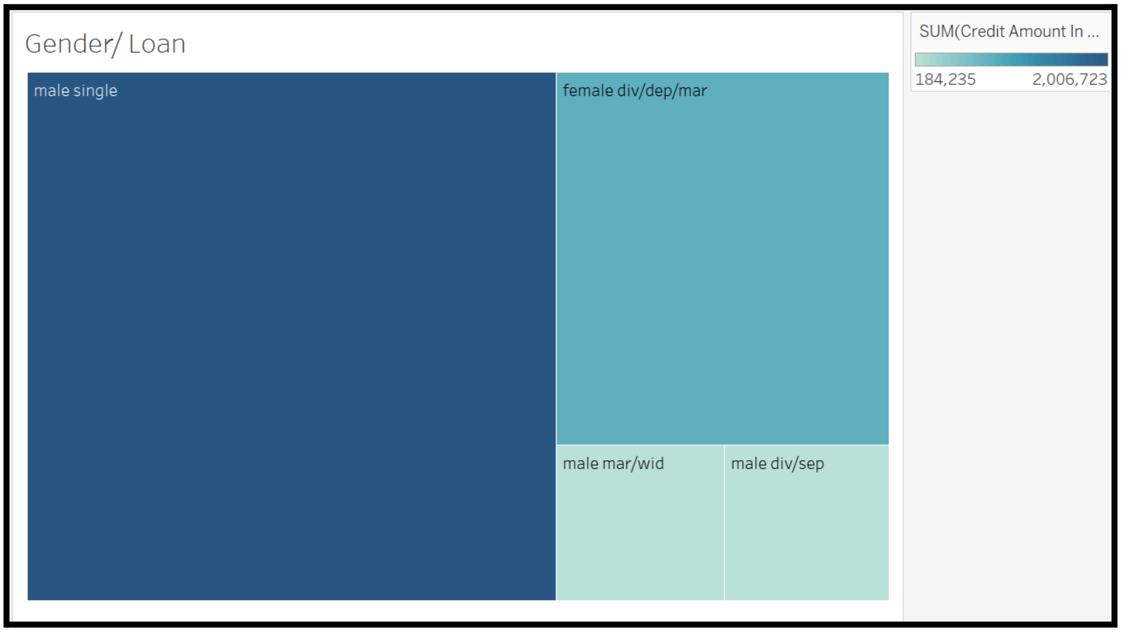
In Sheet 2, we analyzed credit data by purpose, represented in rows and columns. Each row denotes different credit amounts, while columns indicate purposes like "domestic appliance" and "used car." Using stacked bar charts, we visualized total credit amounts for each purpose category. Notably, "new car" purposes stand out with higher credit amounts.



The data ranges from 10,853 to 716,748 lakhs, revealing variations in credit distribution across purposes. This analysis provides valuable insights into credit patterns, ensuring accuracy and efficiency with software tools.

- **Credit based on Gender**

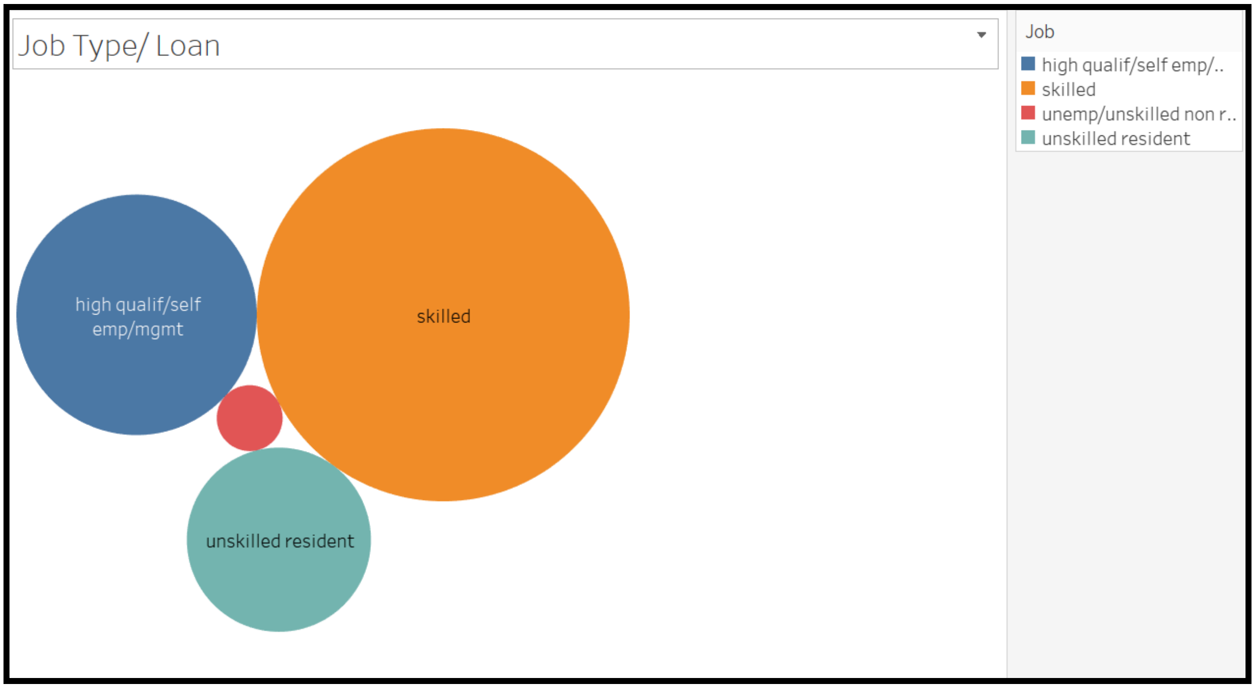
In Sheet 3, we studied credit data by personal status. Color indicates credit amounts, and square size represents total amounts. Data includes "female div/dep/mar," "male div/sep," "male mar/wid," and "male single." Notably, male singles have higher loans. Our analysis revealed varying credit amounts across these categories. For instance, "male mar/wid" may differ from "female div/dep/mar." Credit amounts range from 184,235 to 2,006,723 lakhs. This analysis sheds light on how credit correlates with personal status, ensuring accuracy and efficiency through software tools.



- **Credit based on Job Type**

In Sheet 4, we categorized the data into four job groups: "high qualif/self emp/mgmt," "skilled," "unemp/unskilled non res," and "unskilled resident." Using circle marks with stacked marks enabled, we visually represented the sum of credit amounts for each job category. Each circle is labeled with the respective job, with the color and size indicating job details and the total credit amount in lakhs associated with that category.

Our analysis unveiled variations in credit amounts across different job categories. Notably, skilled workers, represented by larger circles, tended to have higher loan amounts compared to other job types. Additionally, the total credit amounts ranged from 60,393 to 1,934,708 lakhs on this sheet, offering insights into credit distribution among job categories.

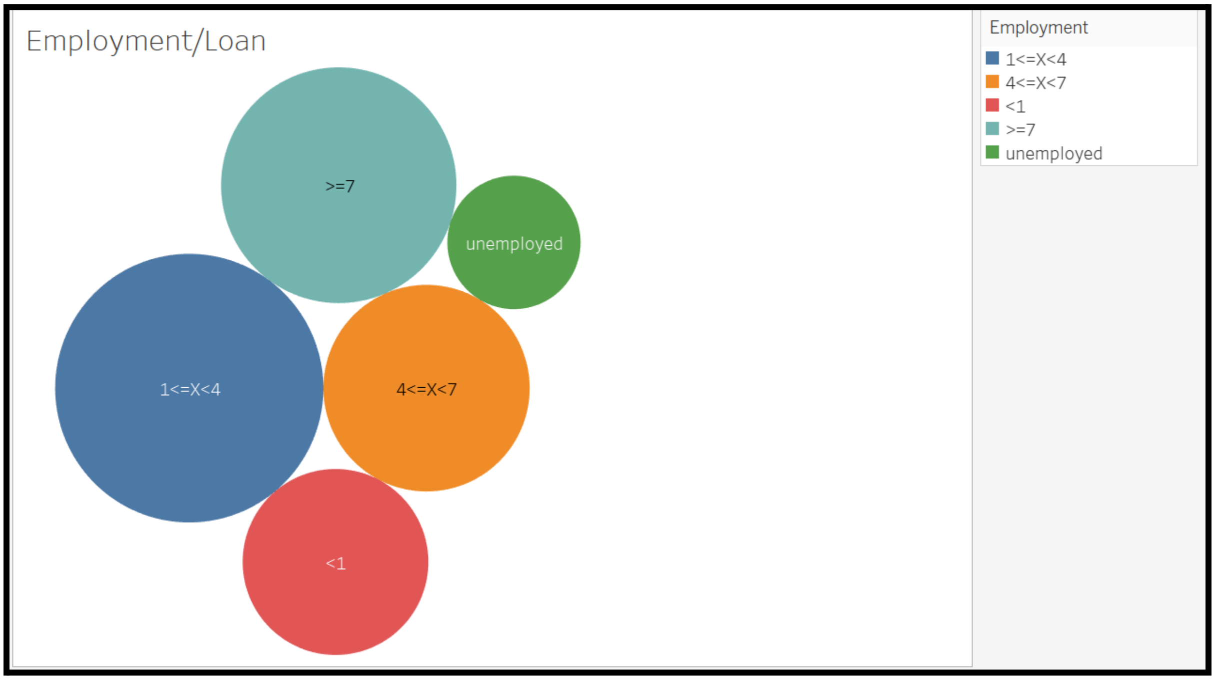


- **Credit Based on Employment**

The data is categorized into five employment groups: ">=7," "<1," "1<=X<4," "4<=X<7," and "unemployed."

Using circle marks, we visually depicted the sum of credit amounts for each employment category. Each circle is labeled with the respective employment category, and the color and size of the circle indicate details about employment and the total credit amount in lakhs associated with that category.

Our analysis revealed variations in the sum of credit amounts across different employment categories. The larger circles represent higher credit amounts taken by individuals with an income ranging from 1 to less than 4.

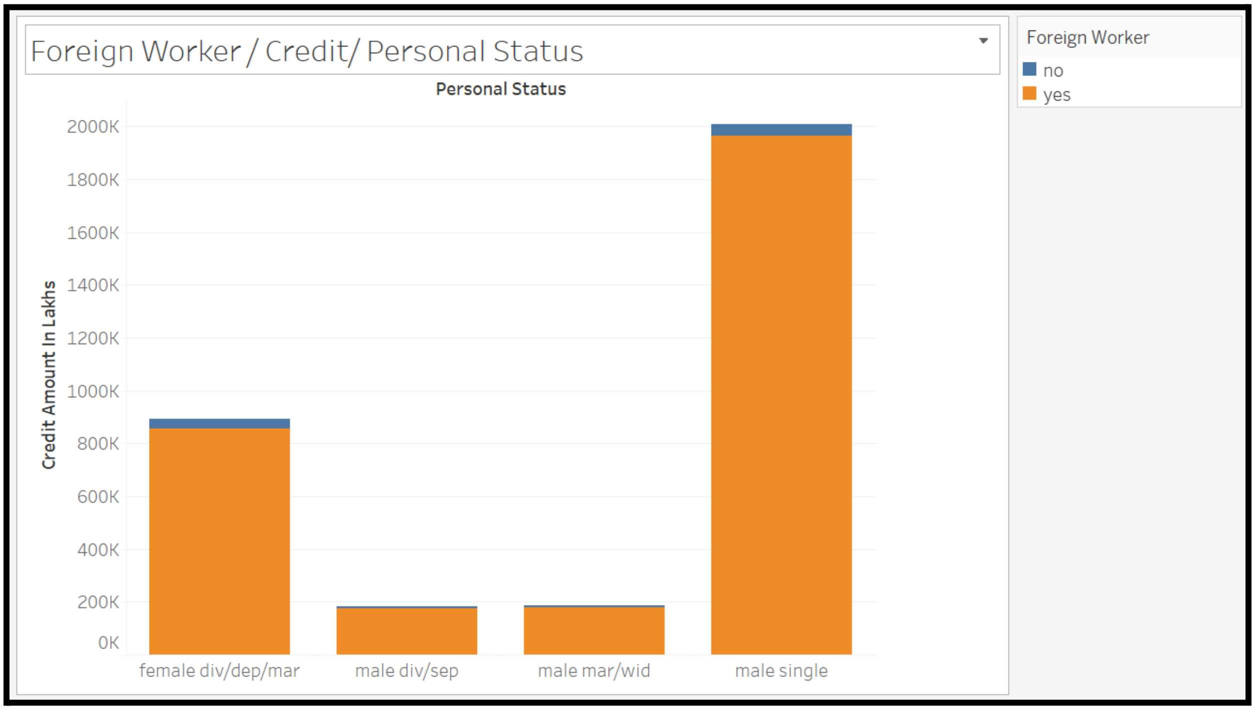


Additionally, the data shows that the sum of credit amounts in lakhs ranges from 261,439 to 1,059,472 on this sheet, providing insight into the distribution of credit amounts across different employment categories.

Overall, the visualization and analysis conducted in Sheet 5 offer valuable insights into how credit amounts correlate with employment status. We utilized software tools to conduct this analysis, ensuring accuracy and efficiency in our findings.

- **Foreign Worker / Credit/ Personal Status**

Sheet 6 displays the sum of credit amounts in lakhs based on personal status, with color indicating foreign worker status. Using bar charts, we visualize total credit amounts for each personal status category. Our analysis reveals variations, notably higher loans for male singles who are foreign workers.

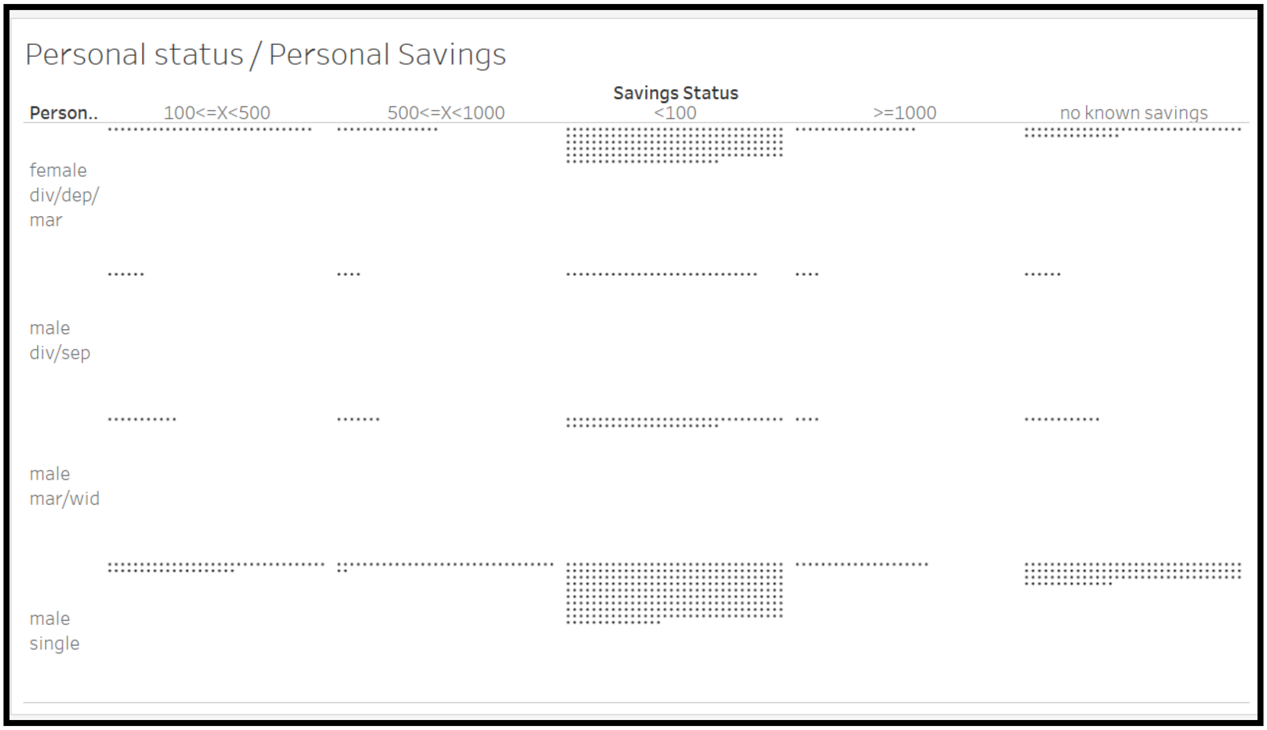


The data ranges from 7,732 to 1,964,120 lakhs. This provides valuable insights into credit patterns based on personal and foreign worker status.

- **Personal status / Personal Savings**

Sheet 7 presents a breakdown of data by Savings Status versus Personal Status. Using circle marks with stacked marks enabled, we visualize the distribution of personal status categories across different savings status groups.

Notably, our analysis reveals that male singles tend to have higher savings compared to other personal status categories across various savings status groups.



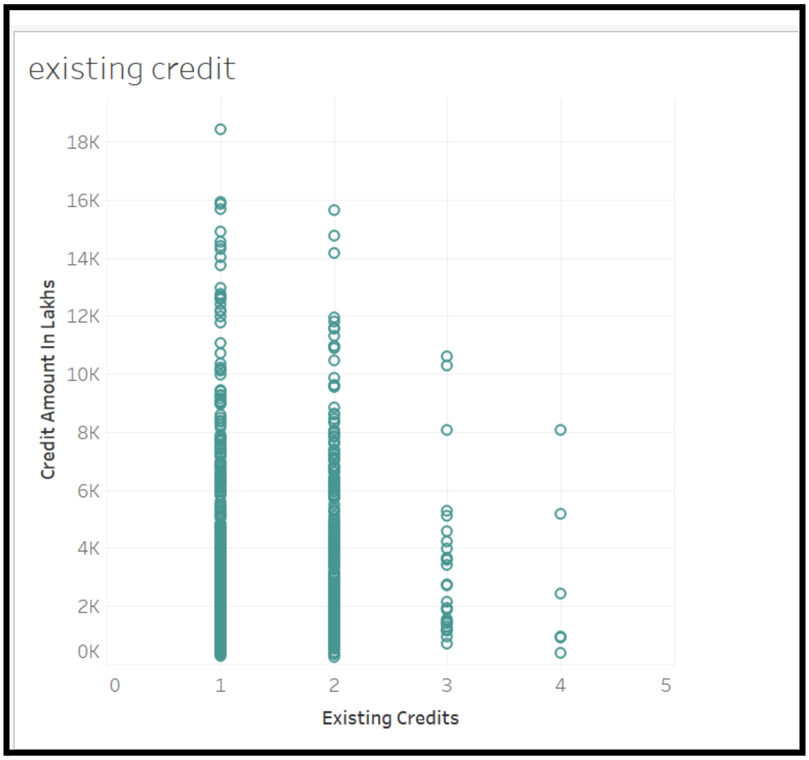
This insight highlights the savings behavior differences among different personal status categories, particularly emphasizing the higher savings propensity among male singles.

Overall, Sheet 7's visualization offers valuable insights into the intersection of savings status and personal status, providing a deeper understanding of individuals' savings behaviors for decision-making purposes.

- **Existing Credit**

Sheet 8 illustrates the relationship between existing credits and credit amounts in lakhs. The data is represented using shape marks with stacked marks disabled.

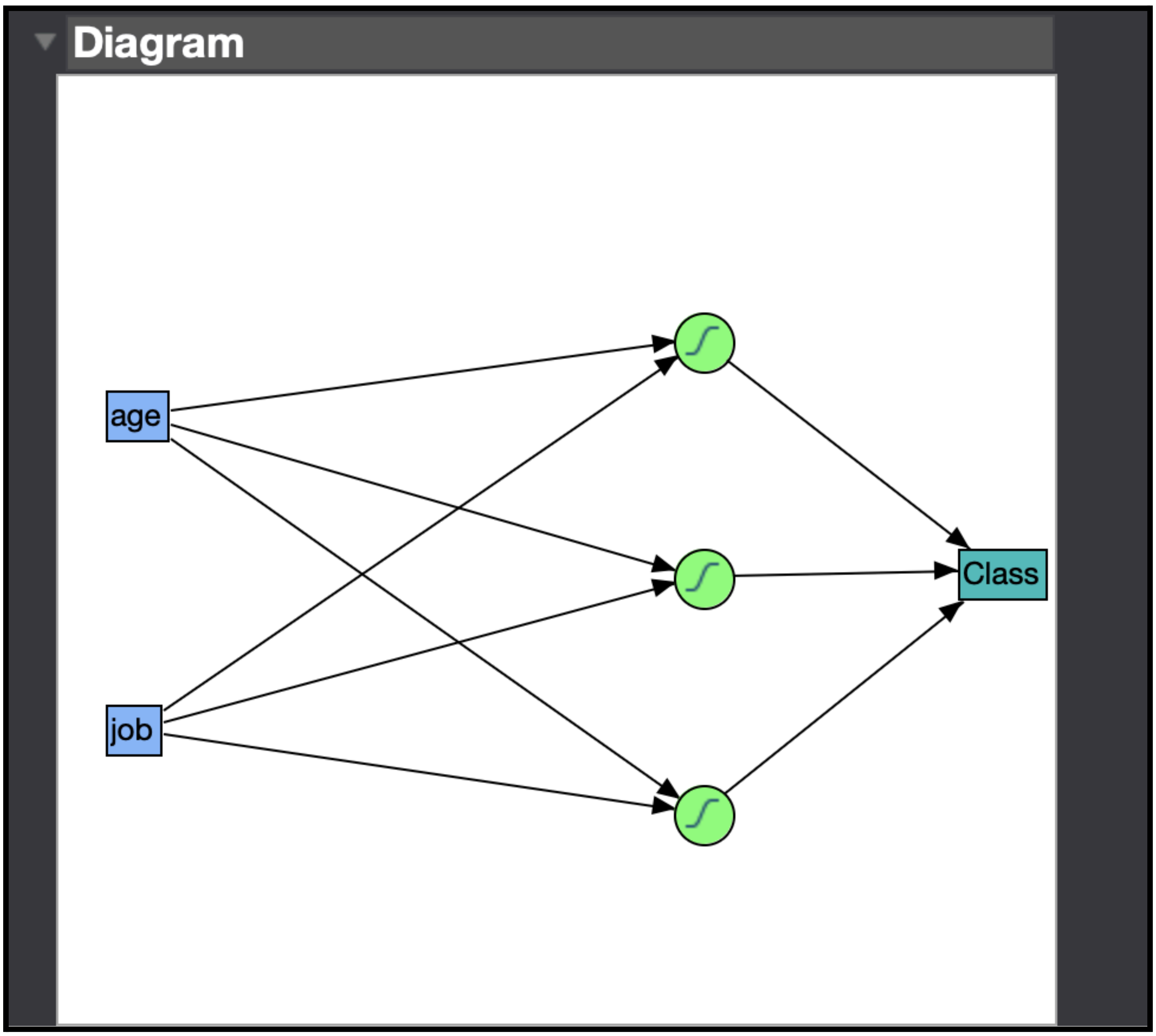
By analyzing the data, it's evident that most individuals have at least one existing credit, as indicated by the range of existing credits from 1 to 4. Additionally, the credit amounts in lakhs range from 250 to 18,424 on this sheet, providing insights into the distribution of credit amounts across different existing credit categories



Overall, Sheet 8's visualization offers valuable insights into how credit amounts vary based on the number of existing credits, highlighting the prevalence of individuals with at least one existing credit.

4. Findings / Conclusions

In our analysis using JMP software, we sought to identify the primary factors that influence the 'Class' categorization within our dataset. The modeling process revealed two main variables that showed a notable correlation with the 'Class' outcome: 'age' and 'job'. These findings are critical, as they suggest that the age of an individual and their employment type are significant predictors of the class they belong to, which could be indicative of their creditworthiness or a similar metric of interest to Goldman Sachs . By understanding the weight and influence of these variables, we can better tailor credit decision-making processes to reflect the nuanced dynamics of our customer base. This approach not only refines our risk assessment models but also allows for more personalized customer service, aligning product offerings with the needs and profiles of distinct customer segments.



According to our analysis, the most significant factor influencing 'Class' in JMP before splitting our dataset into training and validation sets was 'age.' The statistical results were evident: the age variable's chi-square statistic was 232.97181, and its p-value was less than.0001. This indicates that the variable has a lesser impact than the other variables, which are quite distinct with significantly higher p-values.

The age of a customer is a crucial factor in predicting their classification category, regardless of whether it is related to their creditworthiness, risk tolerance, or another classification of interest. This important discovery was made prior to any training or validation. With this knowledge, we can acknowledge the significance of "age" as we get ready to train our predictive models, making sure that the age component is properly taken into account in the creation of a reliable and accurate model for Goldman Sachs to utilize.

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
credit_history	8	8	9.04115536	0.3388
purpose	18	18	18.6670756	0.4126
credit_amount in lakhs	2	2	1.49958361	0.4725
installment_commitment	2	2	0.62354587	0.7321
personal_status	6	6	1.12693531	0.9803
other_parties involved	4	4	1.09028004	0.8958
property_magnitude	6	6	0.75879608	0.9931
age	2	2	232.971811	<.0001*
other_payment_plans	4	4	0.51692396	0.9718
housing	4	4	0.35983504	0.9856
existing_credits	2	2	3.29037934	0.1930
job	6	6	1.77396766	0.9393
num_dependents	2	2	0.66715839	0.7164

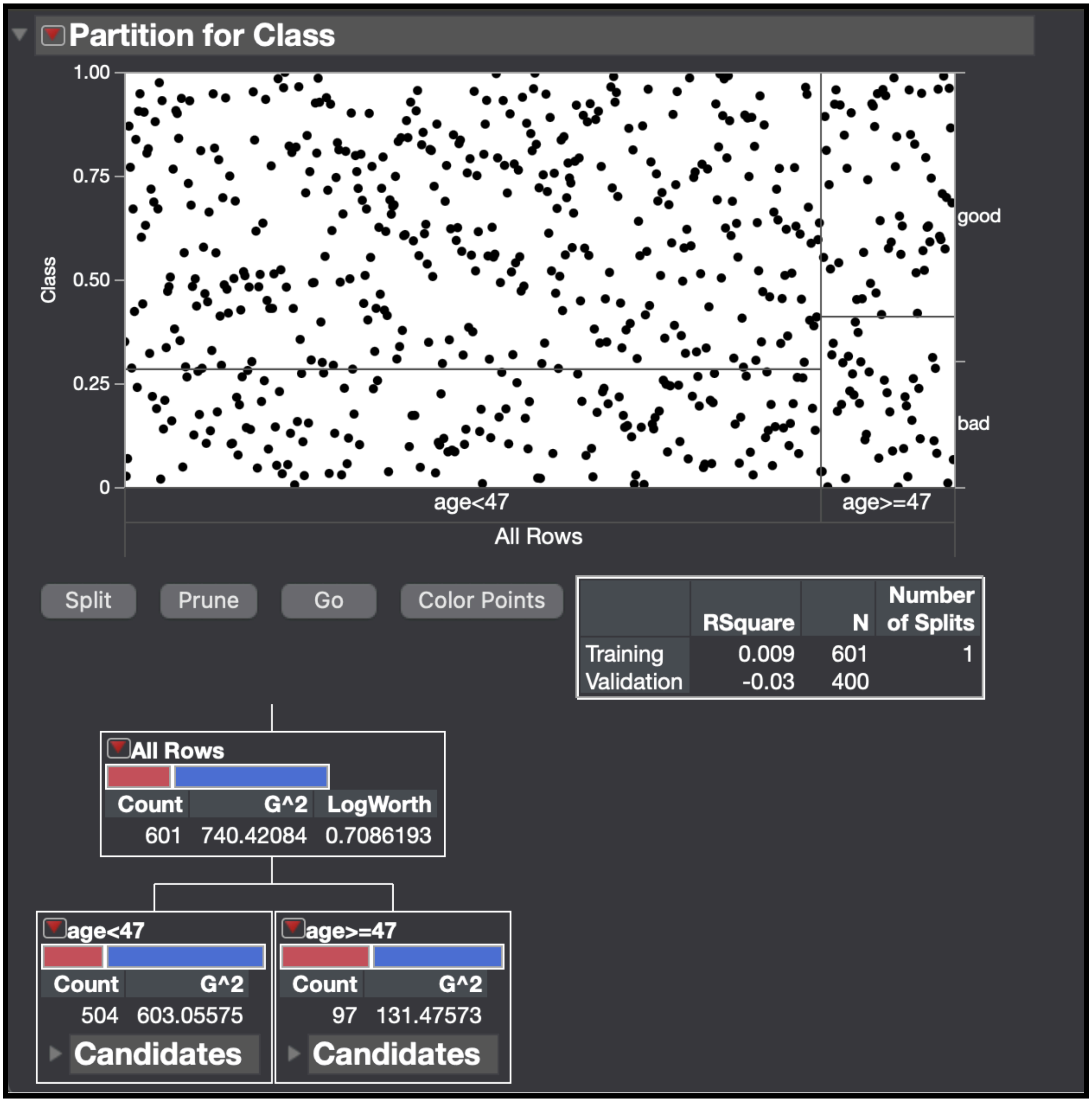
The Effect Likelihood Ratio Tests gave us a definite statistical validation of our predictive model after we divided our dataset into training and validation groups. The results of the analysis showed that the two variables that have the greatest influence on the class variable are "age" and "job." Age had a significant chi-square statistic of 1338.82298 and "job" had a major one of 114.23843, both of which had p-values less than.0001. This result confirms the pre-split analysis and highlights the importance of these variables once again.

'Age' and 'job' were strong predictors, which was further highlighted by the validation set, an essential part of evaluating the model's predictive ability. 'Validation' has a very high chi-square score of 802.03652, which suggests that the model is consistent and trustworthy at predicting outcomes based on these factors. 'Age' and 'job' are strong variables in the context of this financial analysis because the p-value for 'Validation' is less than.0001, which guarantees that the model performs well on unseen data. These insights are especially helpful for Goldman Sachs 's credit decision-making process since they enable more precise and informed tactics that are customized to the specific needs of individual customers.

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
credit_history	8	8	9.27544127	0.3196
purpose	18	18	18.4950462	0.4235
credit_amount in lakhs	2	2	1.4904439	0.4746
installment_commitment	2	2	0.63539321	0.7278
personal_status	6	6	1.13404579	0.9800
other_parties involved	4	4	1.0559435	0.9012
property_magnitude	6	6	0.79210956	0.9923
age	2	2	1338.82298	<.0001*
other_payment_plans	4	4	0.51985584	0.9715
housing	4	4	0.34107394	0.9870
existing_credits	2	2	3.31198051	0.1909
job	6	6	114.23843	<.0001*
num_dependents	2	2	0.64479266	0.7244
Validation	2	2	802.03652	<.0001*

Our analysis revealed that age is a critical factor in determining a person's creditworthiness. An ideal age split of 47 years is recommended by the decision tree model. People under 47 are more often classified as members of the "good" class, whereas people beyond 47 are more often classified as members of the "bad" class. To be more precise, the model linked better credit outcomes to 504 people under the age of 47, out of the total number of people it examined. On the other hand, the model found 97 people who were 47 years of age or older, and they were more frequently associated with worse credit results.The age-based division of the credit data reveals a clear pattern: younger people are often considered to be more creditworthy based on the model's parameters.



In our analysis, we discovered that the model exhibits an excellent capability in predicting the 'good' credit class, as indicated by the high predictive probability observed in both the training and validation phases. The training data showed a 'good' prediction rate of 0.904, while the validation data further confirmed the model's robustness with an even higher 'good' prediction rate of 0.915.

The model’s consistency is underlined by its performance across both datasets, suggesting that it can reliably identify individuals who are likely to be creditworthy. This level of accuracy, particularly in the validation set, which is crucial for assessing how the model might perform in real-world scenarios, suggests that the model is highly effective for its intended purpose.

With the model adeptly classifying the majority of 'good' cases correctly, its application could greatly benefit Goldman Sachs in streamlining the credit decision-making process, focusing on customer segments that are predicted to have a high probability of good credit behavior. Such a tool could optimize resource allocation and enhance customer satisfaction by facilitating quicker credit approvals for low-risk individuals. This strong predictive performance affirms the model’s potential as a valuable asset in Goldman Sachs ’s analytical toolkit.

Training

Actual	Predicted Count		
Class	bad	class	good
bad	14	0	170
class	0	0	0
good	40	0	377

Actual	Predicted Rate		
Class	bad	class	good
bad	0.076	0.000	0.924
class	.	.	.
good	0.096	0.000	0.904

Validation

Actual	Predicted Count		
Class	bad	class	good
bad	11	0	105
class	0	0	1
good	24	0	259

Actual	Predicted Rate		
Class	bad	class	good
bad	0.095	0.000	0.905
class	0.000	0.000	1.000
good	0.085	0.000	0.915

5. Managerial Implications

The insights derived from our data analysis provide actionable recommendations that can significantly influence managerial decisions at Goldman Sachs . Implementing these findings can enhance both operational processes and the overall performance of the institution, particularly in terms of credit risk management and customer segmentation strategies. Here, we outline specific actions that managers can take to leverage the analysis results, thereby improving the efficiency and effectiveness of their operations, tailoring their approaches to different customer demographics, and minimizing potential credit risks.

Strategies for Targeted Marketing:

Campaigns Targeting Employment and Age: Goldman Sachs managers can efficiently target the most creditworthy demographics with marketing efforts that are tailored to their age and work type, which are strong indicators of credit risk. Promotional offers, for instance, might target younger people in skilled fields more.

Improvements in Risk Management:

Personalized Risk Assessment Frameworks:

Apply the results to improve the credit scoring models used by Goldman Sachs . To more effectively measure the credit risk, give indicators like age and work type more prominence.

Dynamic Credit Limits: Modify credit limits in accordance with the recognized risk variables. While keeping more conservative limitations for higher-risk categories, bigger limits could be made available to lower-risk demographics.

Efficiency of Operations:

Simplify Approval Processes: To improve customer satisfaction and cut expenses, simplify the credit approval procedure for segments that have been determined to be low risk.

Intervention Techniques: Create specialized financial guidance or more frequent financial health assessments for high-risk groups to assist them better manage their credit.

Making Decisions Based on Data:

Continuous Data Monitoring: To promptly spot changes in trends that may have an impact on credit risk, update and monitor demographic and financial data on a regular basis.

Feedback Loop: Establish a system of feedback whereby the results of these customized techniques are tracked and incorporated back into the process of making decisions in order to continuously enhance credit risk management tactics.

Goldman Sachs can improve customer relations, competitiveness in the market, and operational procedures and performance while lowering its exposure to credit risk to a considerable extent by putting these managerial implications into practice.

6. Idea Sharing

Having worked on this project, we gained a number of insightful observations:

The integration of data analytics methodologies with operations management theory principles has shown to be a powerful approach for addressing complex business problems. We can create solutions that are more thorough and efficient by combining theoretical ideas with real-world applications.

Enabler for Strategic Decision Making: By using operations management theory to assess credit risk, organizations such as Goldman Sachs can gain important insights that help direct strategic decision-making. Organizations can improve lending procedures, reduce risks, and boost operational effectiveness by utilizing data-driven insights.

Ethical Awareness: Throughout the project, maintaining the highest ethical standards for data confidentiality and privacy has been crucial. Respecting moral standards helps to ensure that our analyses and suggestions are reliable and honest.

Prolonged Development: Working on this project highlights the need for continual education and skill improvement in the fields of operations management and data analytics. Remaining competitive in today's rapidly changing business environment requires keeping up with new tools, approaches, and best practices.

Relevance to Real-World Situations: The project highlights how operations management theory can be used to solve complex real-world business situations. We can provide practical insights with real consequences for organizational effectiveness and strategic decision-making by using theoretical frameworks to examine credit risk.

To sum up, this project experience has strengthened our analytical skills, expanded our understanding of credit risk dynamics, and reinforced the importance of interdisciplinary teamwork in addressing complex business problems.

7. Appendix

In this section, please list a weekly schedule for you to finish this project. An example of table is as follows:

Time	**Contents**
January 27, 2024	Project Group Forms
February 20,2024	Data Gathering and Cleaning
March 18,2024	Model development
April 5,2024	Analysis and insight generation
April 18, 2024	Deriving results presentation preparation
April 22, 2024	Final Project Paper Submission and Final Project Presentation Slides Submission