

# COMPREHENSIVE CREDIT RISK ANALYSIS WITH EDA TECHNIQUES

Explore the vital role of Exploratory Data Analysis in credit risk assessment and how it aids in uncovering potential risks and enhancing decision-making processes.

KARAN LUBHANA



## CREDIT RISK ANALYSIS WITH EDA

Utilizing EDA to Mitigate Loan Default Risks

### OBJECTIVE: APPLY EDA IN BANKING

Utilize Exploratory Data Analysis to enhance decision-making in the banking sector.

### GOAL: UNDERSTAND CUSTOMER BEHAVIOR

Analyze patterns in customer data to gain insights into borrowing habits and preferences.

### AIM: MITIGATE LOAN DEFAULT RISKS

Identify factors contributing to loan defaults to develop effective risk management strategies.

### CONSUMER ATTRIBUTES ANALYSIS

Examine customer demographics, income levels, and credit histories to inform lending practices.

### LOAN ATTRIBUTES EXAMINATION

Evaluate loan characteristics such as amount, term, and interest rates to predict default probabilities.

### PATTERN RECOGNITION

Use EDA techniques to uncover significant trends and correlations in the dataset.

### DATA VISUALIZATION TECHNIQUES

Employ visual tools like charts and graphs to present findings effectively to stakeholders.

### IMPACT ON LENDING POLICIES

Adjust lending policies based on insights gained from EDA to minimize risk and enhance profitability.

### CONTINUOUS MONITORING

Implement ongoing analysis to adapt to changing market conditions and customer behaviors.

# STRATEGIES TO MITIGATE LOAN DEFAULT RISKS

Leveraging Exploratory Data Analysis (EDA) for Decisions

## CHALLENGE OF INSUFFICIENT CREDIT HISTORIES

Loan companies struggle with customers lacking credit histories, increasing the likelihood of defaults.

## REDUCING LOAN DEFAULT RISK

By employing Exploratory Data Analysis (EDA), we can identify patterns that help minimize the chances of loan defaults.

## ENSURING ACCEPTANCE OF REPAYING APPLICANTS

It's crucial to ensure applicants who are capable of repaying loans are not unfairly rejected during the approval process.

## INFORMED DECISION MAKING ON LOANS

Utilizing EDA allows for data-driven decisions regarding loan approvals, rejections, and necessary adjustments to loan terms.

# CLIENT DEFAULT RISK MANAGEMENT STRATEGIES

Actions Based on Identifying Default Patterns

## IDENTIFY HIGH-RISK CLIENTS

Recognize patterns that indicate potential loan defaults to mitigate risks.

## DENY LOANS TO HIGH-RISK APPLICANTS

Proactively deny loans to individuals identified as high-risk to prevent defaults.

## ADJUST INTEREST RATES

Implement higher interest rates for higher-risk borrowers to compensate for increased risk.

## REDUCE LOAN AMOUNTS

Limit loan amounts for those at risk of defaulting to minimize potential losses.

## KEY DATASETS OVERVIEW

Summary of Essential Data Files

### APPLICATION DATA

Contains vital information about clients at the time of their loan application.



### PREVIOUS APPLICATION

Includes records of clients' previous loan applications and their outcomes.



### COLUMNS DESCRIPTION

Serves as a data dictionary that explains the variables used in the datasets.



## ESSENTIAL LIBRARIES FOR DATA MANIPULATION AND VISUALIZATION

Key Libraries in Python

### PANDAS

Library for data manipulation and analysis, providing data structures like DataFrames.



### NUMPY

Library for numerical computations, enabling support for large, multi-dimensional arrays.



### SEABORN

Data visualization library based on matplotlib, offering a high-level interface for drawing attractive graphics.



### MATPLOTLIB

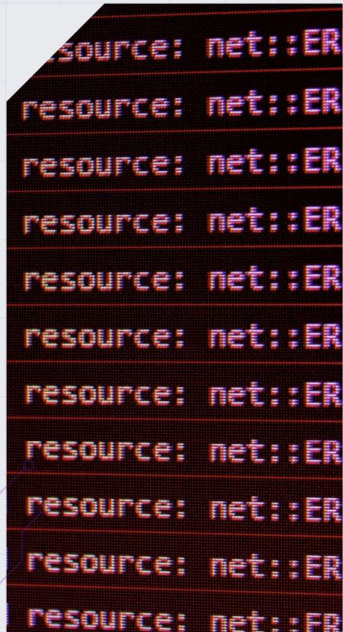
Comprehensive library for creating static, animated, and interactive visualizations in Python.



### RE

Library for regular expressions, useful for string manipulation and pattern matching.





resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER  
resource: net::ER

## LOADING DATASETS WITH ENCODING

Handling UnicodeDecodeError

### ■ LOAD APPLICATION DATA

We begin by loading the main application dataset using pandas.

### ■ LOAD PREVIOUS APPLICATION DATA

Next, we load the previous application dataset for comparative analysis.

### ■ HANDLE ENCODING WITH COLUMNS DESCRIPTION

Finally, we load the columns description with specified encoding to avoid errors.

## DATASET INSPECTION OVERVIEW

Key Actions Performed



### ■ INSPECTING APPLICATION DATA STRUCTURE

We begin by examining the structure of the application dataset to understand its components.

### ■ DISPLAYING APPLICATION DATA INFORMATION

The application data's information is displayed using the `info()` function for better comprehension.

### ■ INSPECTING PREVIOUS APPLICATION DATA STRUCTURE

Next, we inspect the previous application dataset to compare changes over time.

### ■ DISPLAYING PREVIOUS APPLICATION DATA INFORMATION

Similar to the current dataset, we utilize the `info()` function to present the previous data structure.

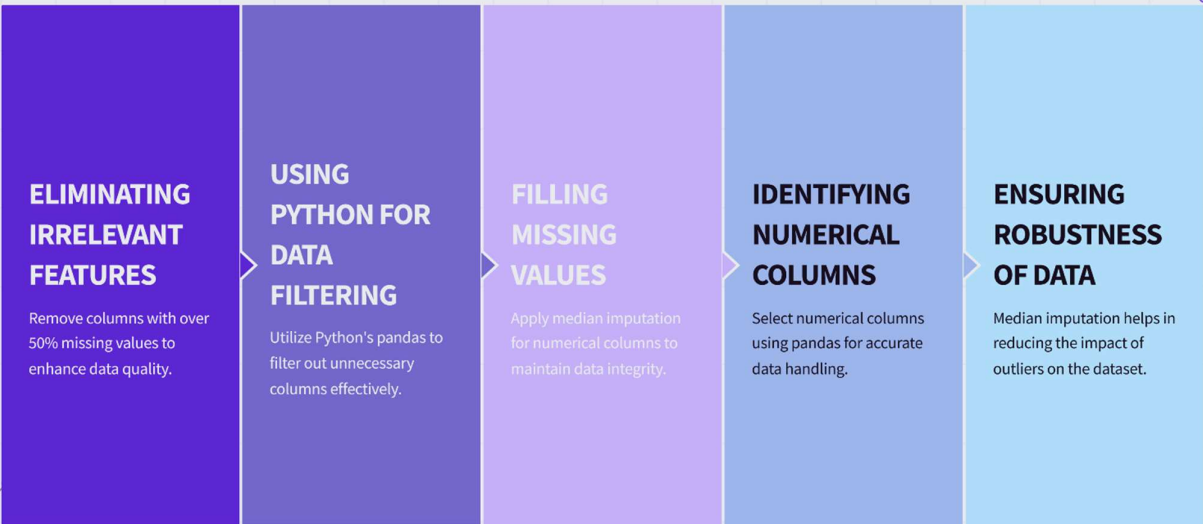
# COLUMN DESCRIPTIONS FROM CSV

An overview of the columns and their respective descriptions

COLUMN	DESCRIPTION
Column	The name of the data column.
Description	A brief explanation of what the column represents.
DataType	Specifies the type of data contained in the column.
UniqueValues	Indicates if the column has unique values.
MissingValues	Count of missing entries in the column.
Range	The range of values present in the column.
Default	Default value if applicable.

# DATA CLEANING TECHNIQUES OVERVIEW

Essential steps for effective data preparation





# DATA HANDLING STRATEGIES AND INSIGHTS

Essential Methods for Managing Data Challenges

## MISSING DATA TREATMENT

Identify and address missing data effectively to maintain dataset integrity.

01

## DATA IMBALANCE ANALYSIS

Analyze the imbalance between 'clients with payment difficulties' and 'all other cases' to adjust modeling.

03

## OUTLIERS IDENTIFICATION

Identify outliers and analyze their potential impact on overall dataset conclusions.

02



# ANALYSIS OF VARIABLE DISTRIBUTIONS

Key Insights on TARGET Variable and Client Income

## UNDERSTANDING THE TARGET VARIABLE

Analyzing the distribution of the TARGET variable helps in understanding the outcome of our predictions.

## VISUAL REPRESENTATION OF TARGET

Utilizing a countplot to visualize the TARGET variable's distribution reveals the frequency of each class.

## CLIENT INCOME DISTRIBUTION ANALYSIS

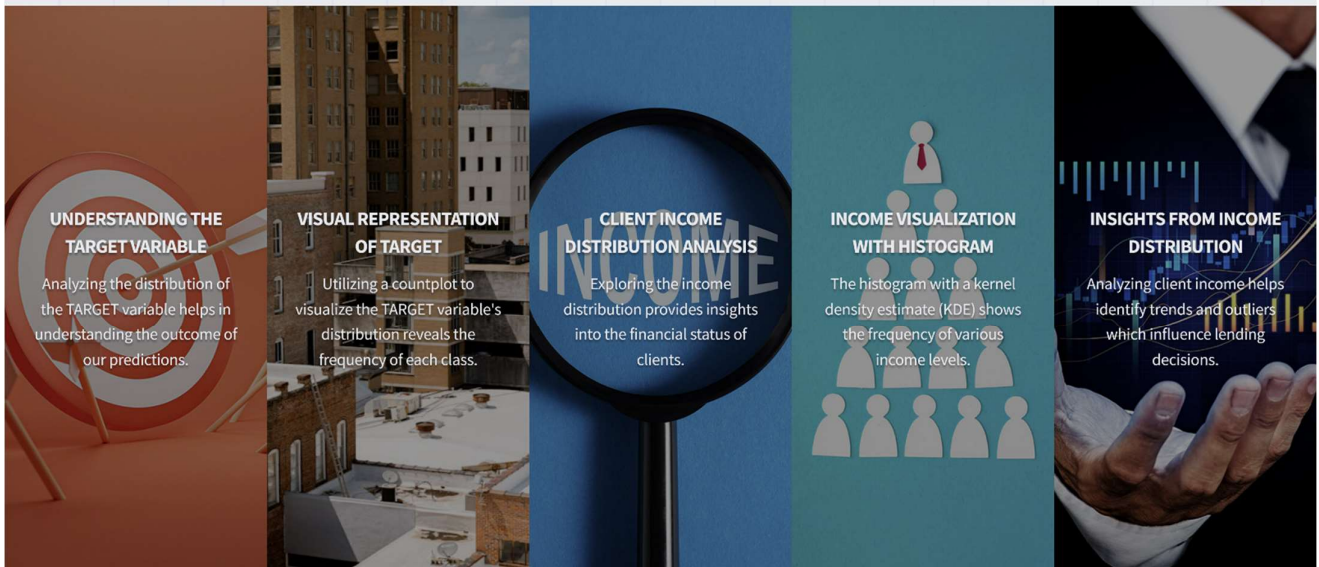
Exploring the income distribution provides insights into the financial status of clients.

## INCOME VISUALIZATION WITH HISTOGRAM

The histogram with a kernel density estimate (KDE) shows the frequency of various income levels.

## INSIGHTS FROM INCOME DISTRIBUTION

Analyzing client income helps identify trends and outliers which influence lending decisions.



## KEY INSIGHTS FROM ANALYSIS

Understanding loan default predictors and demographics



### LOAN AMOUNTS AND DEFAULT RISK

Higher loan amounts significantly increase the likelihood of default among clients.



### GENDER AND DEFAULT RATES

Female clients demonstrate lower default rates compared to their male counterparts.



### IMPACT OF EXTERNAL RISK SOURCES

EXT\_SOURCE\_1 and EXT\_SOURCE\_2 exhibit a strong negative correlation with default risk, indicating they are critical risk factors.

## TRANSFORMING CREDIT RISK ANALYSIS WITH EDA

Explore how Exploratory Data Analysis (EDA) can significantly enhance credit risk assessment and improve decision-making processes in financial services for better outcomes.

