Credit Analysis Using EDA

AIDS SKILL PROJECT

Submitted in the partial fulfillment of the requirements for the award of the degree of Bachelor of Technology

in

Department of Computer Science And Engineering

by

2000030091 – B.Sai Harshith
2000031187 – M.S.K Prasanth
2000030003 - Daasanjaneya Kumar
2000030213 – Vamsi

under the supervision of

Dr.Raju Anitha
(Associate Professor)

Koneru Lakshmaiah Education Foundation
(Deemed to be University estd., u/s 3 of UGC Act 1956)
Greenfields, Vaddeswaram, Guntur (Dist.), Andhra Pradesh - 522502

November, 2019

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
(DST-FIST Sponsored Department)

DECLARATION

The Project Report entitled "Credit Analysis Using EDA" is a record of bondfide work of B.Sai Harshith (2000030091) , M.S.K Prasanth (2000031187) , Daasanjaneya Kumar (2000030003), Vamsi (2000030213) in partial fulfillment of the requirement for the award of degree "BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING"during the academic year 2021-2022.The results embodied in this report have not been copied from any other departments/University/Institute .

2000030091-B.Sai Harshith

2000031187-M.S.K Prasanth

2000030213-Vamsi

2000030003-Daasanjaneya Kumar

(DST-FIST Sponsored Department)

CERTIFICATE

This is to certify that the Project Report entitled “Credit Analysis Using EDA” is a record of bonafide work of B.Sai Harshith (2000030091) , M.S.K Prasanth (2000031187) , Daasanjaneya Kumar (2000030003), Vamsi (2000030213) in partial fulfillment of the requirement for the award of degree “BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING” is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other departments/University/Institute

Signature of the Supervisor	Signature of the HOD
Dr. Raju Anitha	Dr.T.Pavan Kumar
(Associate Professor)	(Professor)

Signature of the EXTERNAL EXAMINER

ACKNOWLEDGMENT

My sincere thanks to **Dr.Raju Anitha** for her outstanding support throughout the project for the successful completion of the work

We express our gratitude to **Dr.T.Pavan Kumar,** Head of the Department for Computer Science and Engineering for providing us with adequate facilities, ways andmeans by which we are able to complete this project based report.

We express our gratitude to**,** Professor In-Charge for Computer Science and Engineering for providing us with adequate facilities, ways and means by which we are able to complete this project based report.

We would like to place on record the deep sense of gratitude to the honorable Vice Chancellor, K L University for providing the necessary facilities to carry the concluded project based report.

Last but not the least, we thank all Teaching and Non-Teaching Staff of our department and especially my classmates and my friends for their support in the completion of our project based report.

CONTENT

Credit Analysis Using EDA

ABSTRACT

Capital markets facilitate the appropriate exchange of money. For lenders and borrowers, the interest rate is the primary pricing mechanism that markets use in this exchange. For borrowers deemed to have higher credit risk, lenders add a risk premium to the interest rate to compensate for higher risk. Credit risk is the risk that the borrower will default on (not pay) the payments agreed upon in the loan. Such risk-based interest rates ensure that money flows appropriately between lenders and borrowers.

Credit analysis is the process of determining a potential borrower's credit risk. This note explores credit analysis and its relation to the credit-risk premium in interest rates. Credit analysis is also used to estimate whether the credit rating of a bond issuer is about to change. By identifying companies that are about to experience a change in debt rating, an investor or manager can speculate on that change and possibly make a profit.

Traditionally most banks have relied on subjective judgment to assess the credit risk of a corporate borrower. Essentially, bankers used information on various borrower characteristics – such as character (reputation), capital (leverage), capacity to pay (volatility of earnings), conditions of the customer's business (purpose of the loan), and collateral – in deciding whether or not to make a given loan. These characteristics are commonly referred to as the 5 Cs. Developing this type of expert system is time-consuming and expensive. Incorporating certain soft (qualitative) data in a risk model is particularly demanding, however successful implementation eliminates human error and reduces potential for misuse. That is why, from time to time, banks have tried to clone their decision-making process. Even so, in the granting of credit to corporate customers, many banks continue to rely primarily on their traditional expert system for evaluating potential borrowers.

Key Words: capital market, interest rates, credit risk, risk premium, credit analysis, Iceland, fishing, financial statement, income statement, volatility of earnings, collateral, leverage, reputation, Qualitative, traditional expert.

Dataset:(https://www.kaggle.com/kapoorshivam/credit-analysis)

ID	Name
2000030091	B.Sai Harshith
2000031187	M.S.K Prasanth
2000030003	Daasanjaneya Kumar
2000030213	Vamsi

I. INTRODUCTION

Credit analysis is a type of financial analysis that an investor or bond portfolio manager performs on companies, governments, municipalities, or any other debt-issuing entities to measure the issuer's ability to meet its debt obligations. Credit analysis seeks to identify the appropriate level of default risk associated with investing in that particular entity's debt instruments.

Credit analysis evaluates the riskiness of debt instruments issued by companies or entities to measure the entity's ability to meet its obligations.The credit analysis seeks to identify the appropriate level of default risk associated with investing in that particular entity.The outcome of the credit analysis will determine what risk rating to assign the debt issuer or borrower

To judge a company's ability to pay its debt, banks, bond investors, and analysts conduct credit analysis on the company. Using financial ratios, cash flow analysis, trend analysis, and financial projections, an analyst can evaluate a firm's ability to pay its obligations. A review of credit scores and any collateral is also used to calculate the creditworthiness of a business.

Not only is the credit analysis used to predict the probability of a borrower defaulting on its debt, but it's also used to assess how severe the losses will be in the event of default.

The outcome of the credit analysis will determine what risk rating to assign the debt issuer or borrower. The risk rating, in turn, determines whether to extend credit or loan money to the borrowing entity and, if so, the amount to lend. An example of a financial ratio used in credit analysis is the debt service coverage ratio (DSCR). The DSCR is a measure of the level of cash flow available to pay current debt obligations, such as interest, principal, and lease payments. A debt service coverage ratio below 1 indicates a negative cash flow.

For example, a debt service coverage ratio of 0.89 indicates that the company's net operating income is enough to cover only 89% of its annual debt payments. In addition to fundamental factors used in credit analysis, environmental factors such as regulatory climate, competition, taxation, and globalization can also be used in combination with the fundamentals to reflect a borrower's ability to repay its debts relative to other borrowers in its industry.

Credit analysis is important for banks, investors, and investment funds. As a corporation tries to expand, they look for ways to raise capital. This is achieved by issuing bonds, stocks, or taking out loans. When investing or lending money, deciding whether the investment will pay off often depends on the credit of the company. For example, in the case of bankruptcy, lenders need to assess whether they will be paid back.

Similarly, bondholders who lend a company money are also assessing the chances they will get their loan back. Lastly, stockholders who have the lowest claim priority access the capital structure of a company to determine their chance of being paid. Of course, credit analysis is also used on individuals looking to take out a loan or mortgage

1.1 Theoretical Background

In recent decades, a number of objective, quantitative systems for scoring credits have been developed. In univariate (one variable) accounting-based credit-scoring systems, the credit analyst compares various key accounting ratios of potential borrowers with industry or group norms and trends in these variables.

Today, Standard & Poor's, Moody's, and Risk Management Association can all provide banks with industry ratios. The univariate approach enables an analyst starting an inquiry to determine whether a particular ratio for a potential borrower differs markedly from the norm for its industry. In reality, however, the unsatisfactory level of one ratio is frequently mitigated by the strength of some other measure. A firm, for example, may have a poor profitability ratio but an above-average liquidity ratio. One limitation of the univariate approach is the difficulty of making trade-offs between such weak and strong ratios. Of course, a good credit analyst can make these adjustments. However, some univariate measures – such as the specific industry group, public versus private company, and region – are categorical rather than ratio-level values. It is more difficult to make judgments about variables of this type.

Although univariate models are still in use today in many banks, most academics and an increasing number of practitioners seem to disapprove of ratio analysis as a means of assessing the performance of a business enterprise. Many respected theorists downgrade the arbitrary rules of thumb (such as company ratio comparisons) that are widely used by practitioners and favor instead the application of more rigorous statistical techniques.

Fuzzy logic and neural networks are examples of novel methods of developing credit scoring expert systems that deliver greater accuracy in estimates of future performance of a business enterprise. Beside hard data present in traditional ratio analysis, fuzzy logic can easily incorporate linguistic terms and expert opinions which makes it more adapted to cases with insufficient and imprecise hard data, as well as for modeling risks that are not fully understood.[2]

Typical education credentials often require a business related bachelor's degree majoring in finance, business, statistics, or accounting (to include an emphasis in finance or economics). An MBA is not required, however is increasingly being held or pursued by analysts, often to become more competitive for advancement opportunities. Commercial bankers also undergo intense credit training provided by their bank or a third-party company.

II. SOFTWARE REQUIREMENTS

Operating System : Windows 10 and Above

Processor : Intel i10 above and Ryzen Series 4000 Above

RAM : 4 GB

Software : Anaconda – Jupyter.

Language : Python3

III. FLOWCHART

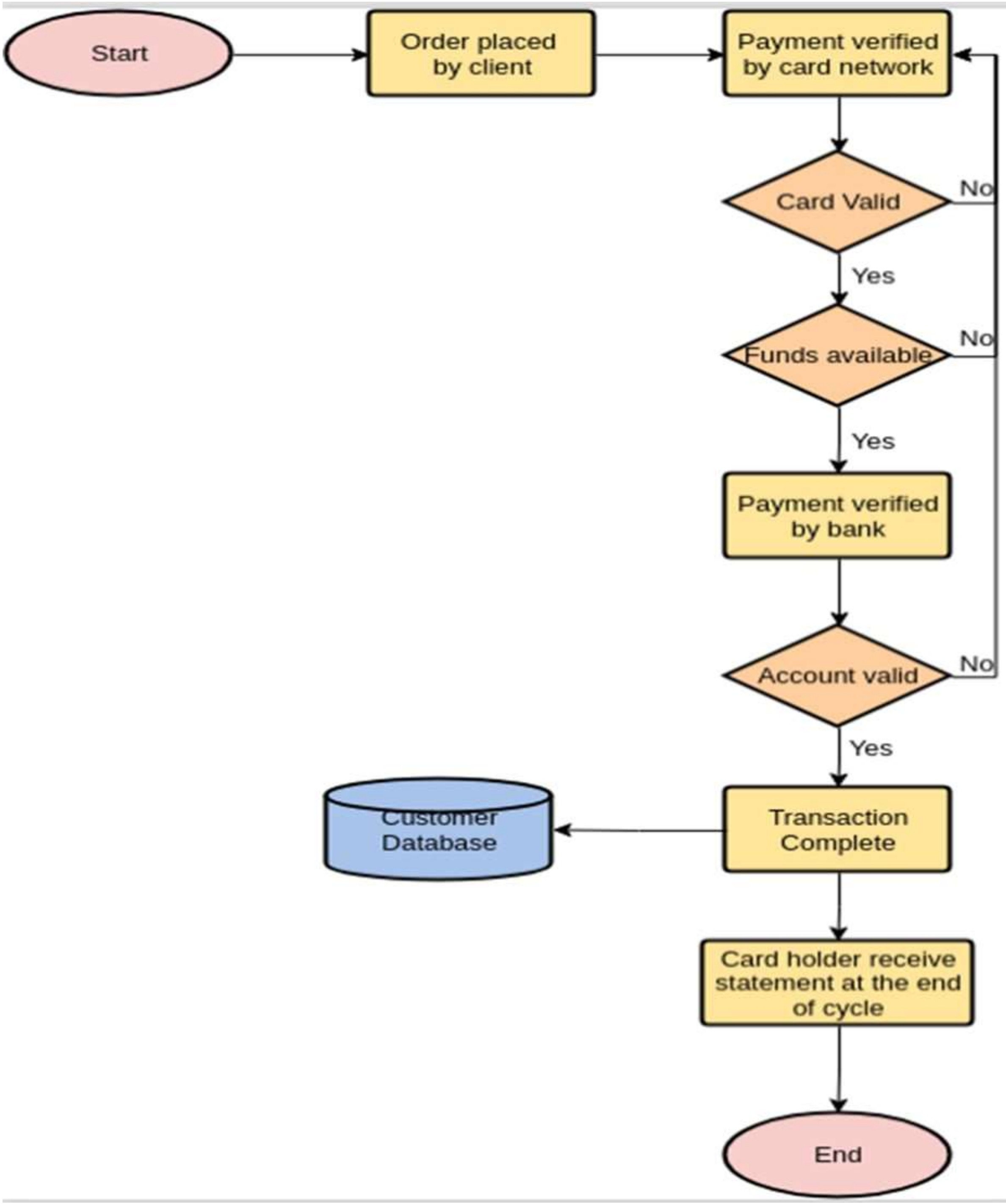


Fig3.1: Flow Chart Of credit analysis using EDA

3.1. Explanation of Flow Chart

DataSet:

he term data set refers to a file that contains one or more records. The record is the basic unit of information used by a program running on z/OS. Any named group of records is called a data set.

Importing DataSet

We have been using data sets already stored as R objects. A data scientist will rarely have such luck and will have to import data into R from either a file, a database, or other sources. Currently, one of the most common ways of storing and sharing data for analysis is through electronic spreadsheets. A spreadsheet stores data in rows and columns. It is basically a file version of a data frame. When saving such a table to a computer file, one needs a way to define when a new row or column ends and the other begins. This in turn defines the cells in which single values are stored.

Data Pre Processing

Data preprocessing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying machine learning or data mining algorithms.

Why Data Pre Processing

Preprocessing of data is mainly to check the data quality. The quality can be checked by the following

- Accuracy: To check whether the data entered is correct or not.
- Completeness: To check whether the data is available or not recorded.
- Consistency: To check whether the same data is kept in all the places that do or do not match.
- Timeliness: The data should be updated correctly.
- Believability: The data should be trustable.
- Interpretability: The understandability of the data.

Passing Parameters To Classify

We have to pass some parameters, so that depending on the parameters the data is classified into two parts. It maybe linear or non linear.

The parameters that we passed to classify must be inside the dataset I.e. csv file

credit analysis using EDA:

Credit analysis is a process of drawing conclusions from available data (both quantitative and qualitative) regarding the creditworthiness of an entity, and making recommendations regarding the perceived needs, and risks. Credit Analysis is also concerned with the identification, evaluation, and mitigation of risks associated with an entity failing to meet financial commitments.

IV. EXPLORATORY DATA ANALYSIS

In statistics, exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.
A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

Exploratory data analysis has been promoted by John Tukey since 1970 to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments.

EDA is different from initial data analysis (IDA) which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA..

EDA explained using sample Data set

To share my understanding of the concept and techniques I know,I'll take an example of white variant of Wine Quality data set which is available on UCI Machine Learning Repository and try to catch hold of as many insights from the data set using EDA.

To starts with, I imported necessary libraries (for this example pandas,numpy,matplotlib and seaborn) and loaded the data set.

Note : Whatever inferences I could extract,I've mentioned with bullet points.

Get a better understanding of data

EDA helps to bring out points from datasets that may not be analyzed by standard data science algorithms. EDA helps in better data understanding.

Understanding data patterns

EDA is known for capturing and analyzing uncommon data patterns that will be skipped by typical machine learning algorithms.

Drawing charts and graphs for better understanding

EDA is all about data visualization. EDA analyzes data sets from different angles and projects the results as charts and graphs.

To get a better understanding of the problem statement

With graphs and charts and other forms of data visualization using ML, EDA gives a sound picture of the problem.

Some Good EDA and Data Pre-processing Practices

Have a well-defined problem statement

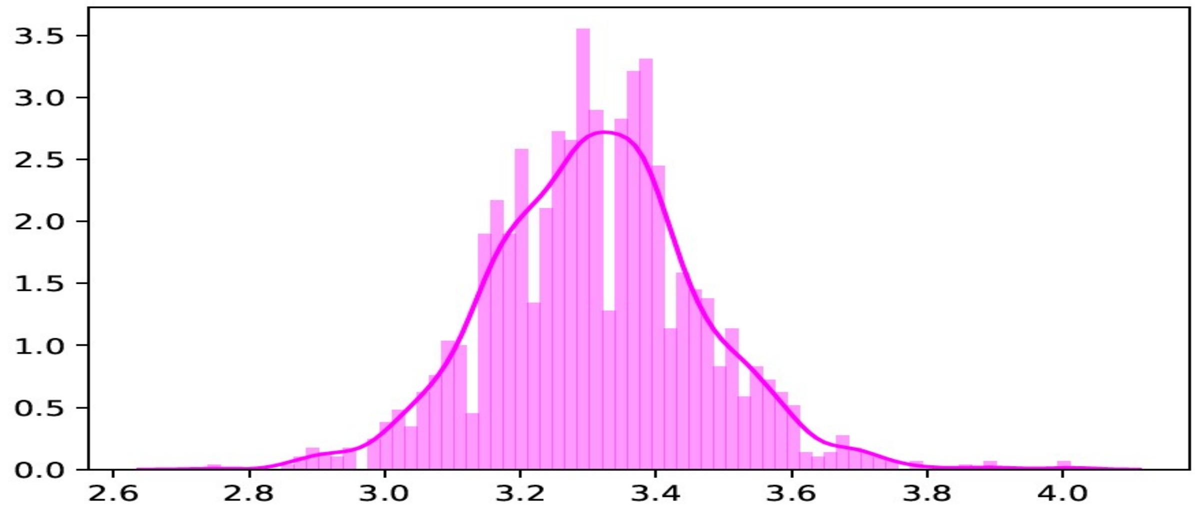
It is always advised to have a clear goal and not add items to it. In data science terms, "*controlled exploration yields best results*."

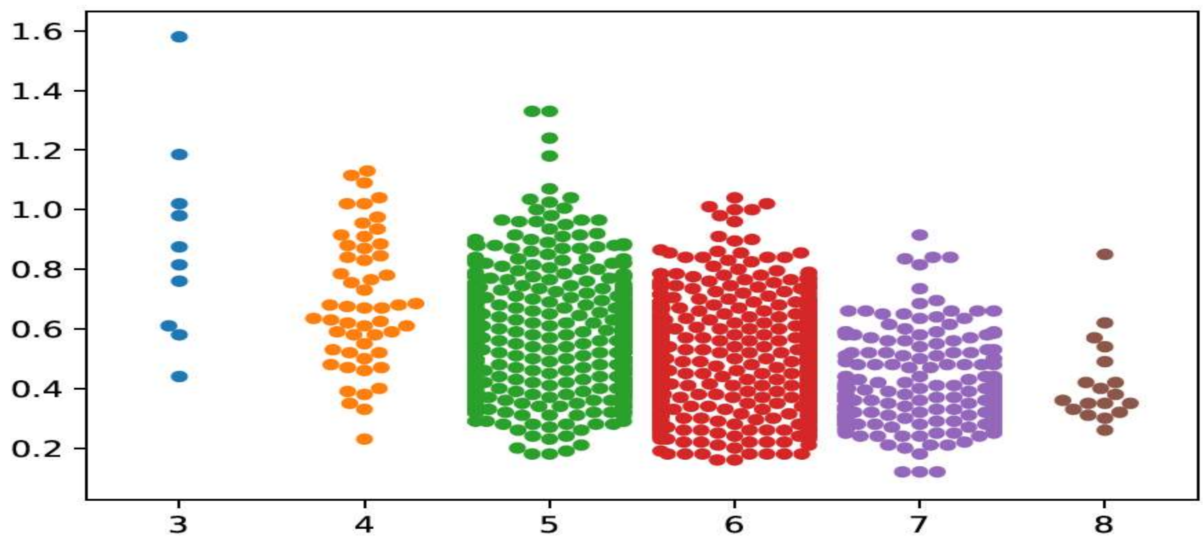
Level the data set and re-iterate

When EDA reveals interesting findings and patterns that are not related to the original problem statement, do not get diverted. You can note it down for later use but immediately return to the initial exploration.

Conclusion

The pre-processing step of EDA not only makes the entire machine learning data analytics process easier but also leads to better results. Hence, it is highly recommended that data scientists use it if they aren't already.


3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0


1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
3
4
5
6
7
8

BASIC TERMINOLOGY

Support Vectors — the points which are closest to the hyperplane

Hyperplane — a subspace with dimension 1 lower than its ambient space . It serves to divide the space into multiple sections.
Given a 3-dimensional space, the subsequent hyperplane would be 2-dimensional plane. Similarly, in a 2-dimensional plane, the hyperplane would be a 1-dimensional line.

Margin — the distance between the hyperplane and the nearest data point from either side

Kernel — a mathematical function used to transform input data into a different form. Common kernel functions include *linear, nonlinear, polynomial, etc.*

Scatter Plot: A scatter plot is a chart type that is normally used to observe and visually display the relationship between variables. The values of the variables are represented by dots. The positioning of the dots on the vertical and horizontal axis will inform the value of the respective data point; hence, scatter plots make use of Cartesian coordinates to display the values of the variables in a data set. Scatter plots are also known as scattergrams, scatter graphs, or scatter charts.

Matplotlib: Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI
toolkits like Tkinter, wxPython, Qt, or GTK. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged.[3] SciPy makes use of Matplotlib.

Matplotlib was originally written by John D. Hunter. Since then it has an active development community[4] and is distributed under a BSD-style license. Michael Droettboom was nominated as matplotlib's lead developer shortly before John Hunter's death in August 2012[5] and was further joined by Thomas Caswell.[6][7] Matplotlib is a NumFOCUS fiscally sponsored project.[8]

Matplotlib 2.0.x supports Python versions 2.7 through 3.10. Python 3 support started with Matplotlib 1.2. Matplotlib 1.4 is the last version to support Python 2.6.[9] Matplotlib has pledged not to support Python 2 past 2020 by signing the Python 3 Statement.[10]

V. ADVANTAGES AND LIMITATIONS

5.1 ADVANTAGES OF CREDIT ANALYSIS

Firms, governments and all types of creditors engage in credit analysis to determine to what extent they face credit risk associated with their investments. In weighing the advantages and disadvantages of making a certain type of investment, firms utilize in-house computer programs to advise on reducing and avoiding risk (or transferring it elsewhere) or use third party help, like examining rating agencies' estimations of creditworthiness from companies like Standard & Poor's, Moody's, Fitch Ratings and others. After lenders use their own models and the advice of others to rank customers according to risk, they apply this knowledge to reduce credit risk.

On the other hand, if a company has no debt at all, then investors will wonder if the company has the ability to expand and grow. If not, then stock prices will not appreciate. Credit analysis helps determine both the growth potential and stability of a company.

The second concern for equity holders about credit quality is the claim on assets. Equity holders have the least claim on assets of a company in the case of bankruptcy. If the company goes bankrupt, shareholders will get their claim only if secured and unsecured creditors did not already take all the remaining assets. This is why the level of existing debt is important for equity holders as well.

5.2 LIMITATIONS OF QUANTITATIVE CREDIT ANALYSIS

Credit analysis, by definition, looks forward. All looks into the future will be subjective and involve many unique industry and issuer factors. This process cannot be reduced to a formulaic methodology or quantitative model.

Drawing from the past, credit analysts attempt to focus on a longer-term, future perspective. They must look through the cycle and through any potential accounting distortions. Finally, they must examine a variety of reasonably adverse scenarios to incorporate sensitivity to risk. Sovereign, economic, and industry factors influence risk. Soft facts, or subjective judgments, add valuable information to the credit analysis process.3 Quantitative and qualitative pitfalls exist in the quantitative portion of any analysis.
Finally, it is also important to note the difference between *relative* and *absolute* quantitative credit analysis. Relative analysis of comparable credits allows us to understand whether, after controlling for all differences in (for example) size, financial leverage, profitability, cash flow and liquidity, the credit is rated consistently with its group of comparables. But this does *not* answer whether the entire group of comparables is rated appropriately. This question may only be answered through an absolute default and recovery analysis to gauge investor risk and reward.

Understanding Credit Analysis Basics

Credit analysis draws conclusions by evaluating the available quantitative and qualitative data regarding the creditworthiness of a client and making recommendations on whether or not to approve the loan application. The objective of credit analysis is to determine the risk of default that a client presents and assign a risk rating to each client. The risk rating will determine if the company will approve (or reject) the loan application, and if approved, the amount of credit to be granted.

Credit Analysis: Traditional Approach vs. Modern Approach

Traditionally, banks relied on subjective methods to evaluate the creditworthiness of their clients. The methods focus on the character of the borrower, and the credit analyst is required to assess the level of credit risk associated with lending to the borrower. The traditional approach evaluates the main characteristics of the borrower, commonly referred to as the 5 C's of Credit. An analysis that is based purely on the characteristics of the borrower is subject to human error and misuse. Banks continue to rely on traditional credit analysis approaches when evaluating potential borrowers.

Unlike traditional approaches, modern credit analysis approaches are based on qualitative credit scoring systems. In such an approach, credit analysts use the univariate accounting-based credit scoring systems to compare key accounting ratios of specific clients versus industry ratios to show how a client's ratio differs from the industry standards or trends.

Credit scoring systems assign scores to several aspects associated with the creditworthiness of a borrower. The scores can range from 300 to 850, with the latter being the highest credit rating that a borrower can get. The key aspects of a borrower that determine their credit score include payment history, current debt, length of debt, type of debt, and the payment interest. A bank can establish its own credit scoring system or use third party services such as FICO.

1. Financial Assessment

Financial assessment involves assessing the current cash flows that the business generates to determine if the borrower is in a position to service the debt. The lender can evaluate the financial performance of a company either in isolation or in comparison to other companies operating in the same industry as the lender.

Ideally, a company will be considered to possess strong financials for credit purposes if its cost structure allows it to produce consistently higher than average profits throughout the financial period. It means that the company maintains above-average revenues during the peak season, as well as during a downturn. Such a company will find it easier to get loan approval because it demonstrates a low risk of default.

Conversely, a company with irregular revenues across all phases of the business cycle will be considered weak for credit purposes. It may refer to a company that reports above-average results during the peak season but struggles to generate profits when there is a business downturn because of low market demand. The lender may either reject the loan application on the basis of weak cash flows or extend a lower credit than what the borrower needed and at higher costs.

2. Industry Competitiveness

How competitive is the company in the industry? A bank is interested in lending to a company with a strong competitive position in its specific industry. When assessing the competitiveness of a company, the lender looks at the company’s business strategy and how consistent and adaptable it is to the current trends and changes in the market.

Dominant companies with an upper hand in the market have strong barriers that limit new entrants into the market. The barriers may be licensing agreements, strong patents, copyright protection, and franchises, which may be too expensive for new entrants to afford. Weak companies often have poorly structured business strategies that are not in line with market trends and changes. Such companies also tend to see high rates of customer dissatisfaction and low-rates of reinvestment.

3. Business Environment

The lender looks at the business environment in which a business operates to identify any challenges that the business may be facing and that may affect its ability to honor its financial obligations. For example, the lender can consider the industry risk to determine how the industry dynamics and regulatory climate can affect the performance of the business.

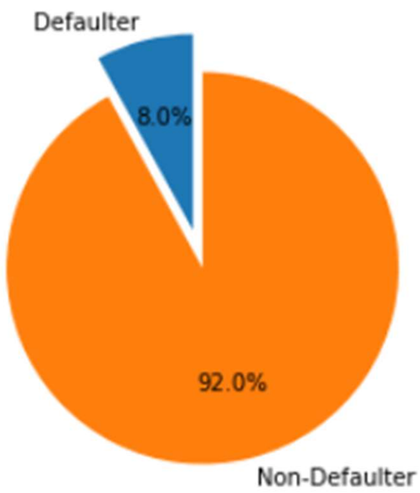
Another consideration is country risk, specifically how the political, legal, and tax climates in the country where the company operates affect its business activities. A company may be considered a viable borrower if it can isolate itself from such risks and create strategic alternatives to hedge against them. On the other hand, a business may be defined as weak if its earnings are greatly affected by changes in business cycles and factors within the business environment.

VI. RESULT

Code:

```
#Checking for data imbalance for the Target Column
Defaulter = round((appdf['TARGET'].value_counts()[1]/len(appdf)),2)
Non_Defaulter = round((appdf['TARGET'].value_counts()[0]/len(appdf)),2)
explode= (0.1,0.1)
client = [Defaulter, Non_Defaulter]
labels = 'Defaulter', 'Non-Defaulter'
plt.pie(client, labels=labels, explode=explode, autopct='%1.1f%%', startangle=90)
plt.show()
```

output:

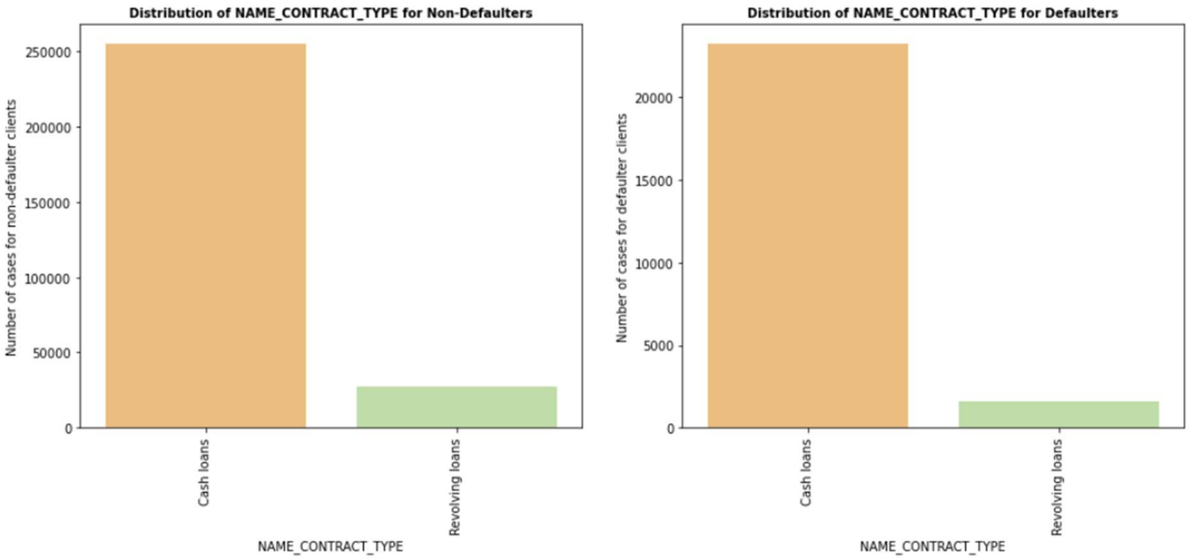


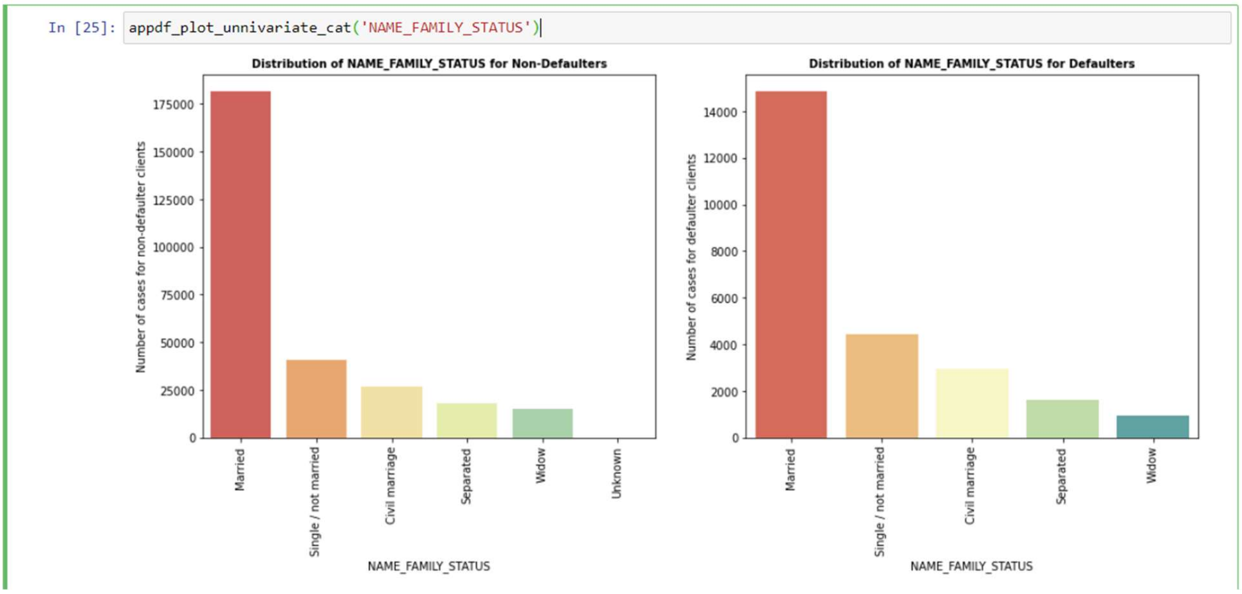
Code:

```
def appdf_plot_unnivariate_cat(var):
    plt.figure(figsize=(16,6))
    plt.subplot(1, 2, 1)
    sns.countplot(var, data=appdf0, palette= 'Spectral', order= appdf0[var].value_counts().index)
    #Order keyword is used above to have the order of the values remain the same in both the subplots
    plt.title('Distribution of '+ '%s' %var +' for Non-Defaulters', weight='bold', fontsize=10)
    plt.xlabel(var)
    plt.xticks(rotation=90)
    plt.ylabel('Number of cases for non-defaulter clients')
    plt.subplot(1, 2, 2)
    sns.countplot(var, data=appdf1, palette='Spectral', order= appdf1[var].value_counts().index)
    plt.title('Distribution of '+ '%s' %var +' for Defaulters', weight='bold',fontsize=10)
    plt.xlabel(var)
    plt.xticks(rotation=90)
    plt.ylabel('Number of cases for defaulter clients')
```

```
    plt.show()
    appdf_plot_univariate_cat('NAME_CONTRACT_TYPE')
```

output:



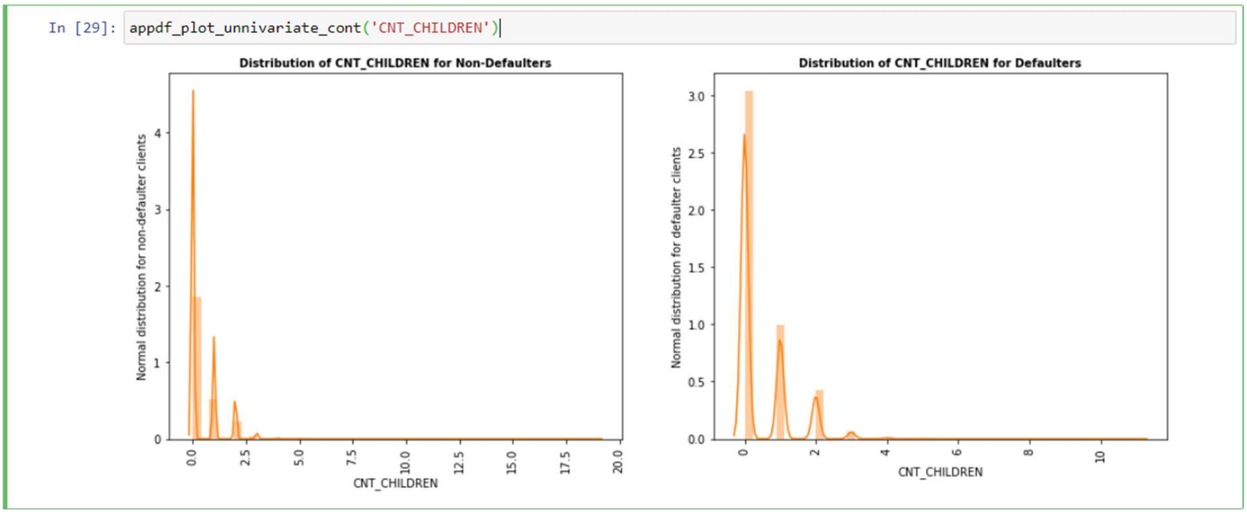


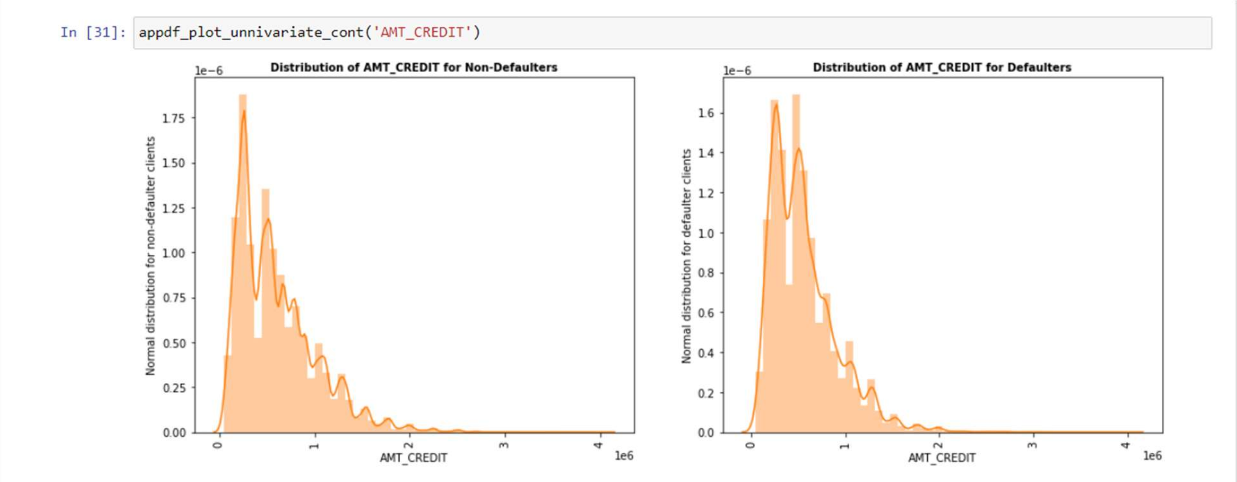
Code:

```
def appdf_plot_univariate_cont(var):

    plt.figure(figsize=(16,6))

    plt.subplot(1, 2, 1)
    sns.distplot(appdf0[var],color='tab:orange')
    plt.title('Distribution of '+ '%s' %var +' for Non-Defaulters', weight='bold', fontsize=10)
    plt.xlabel(var)
    plt.xticks(rotation=90)
```

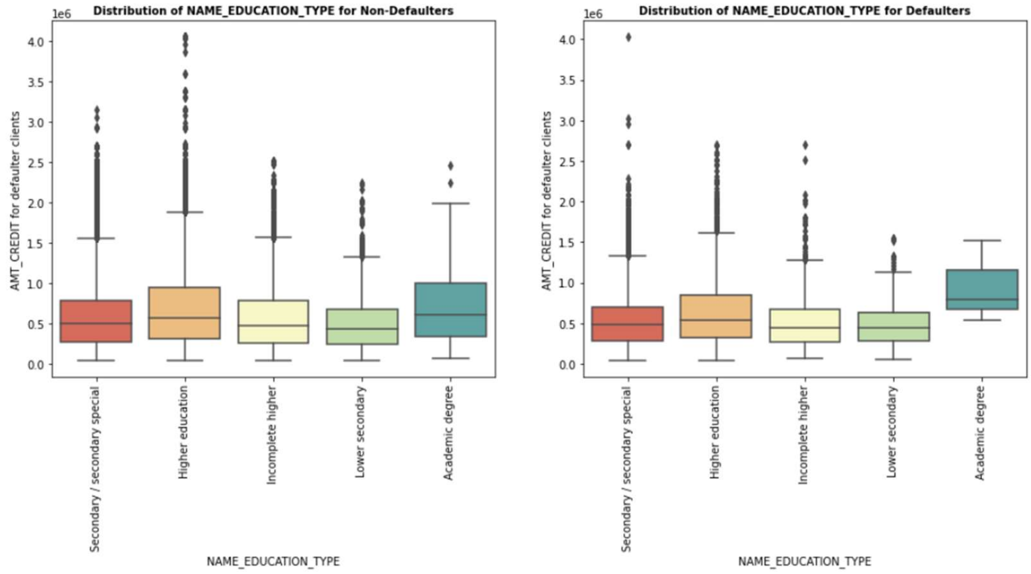
```
plt.ylabel('Normal distribution for non-defaulter clients')
plt.subplot(1, 2, 2)
sns.distplot(appdf1[var],color='tab:orange')
plt.title('Distribution of '+ '%s' %var +' for Defaulters', weight='bold',fontsize=10)
plt.xlabel(var)
plt.xticks(rotation=90)
plt.ylabel('Normal distribution for defaulter clients')
plt.show()
```

output:

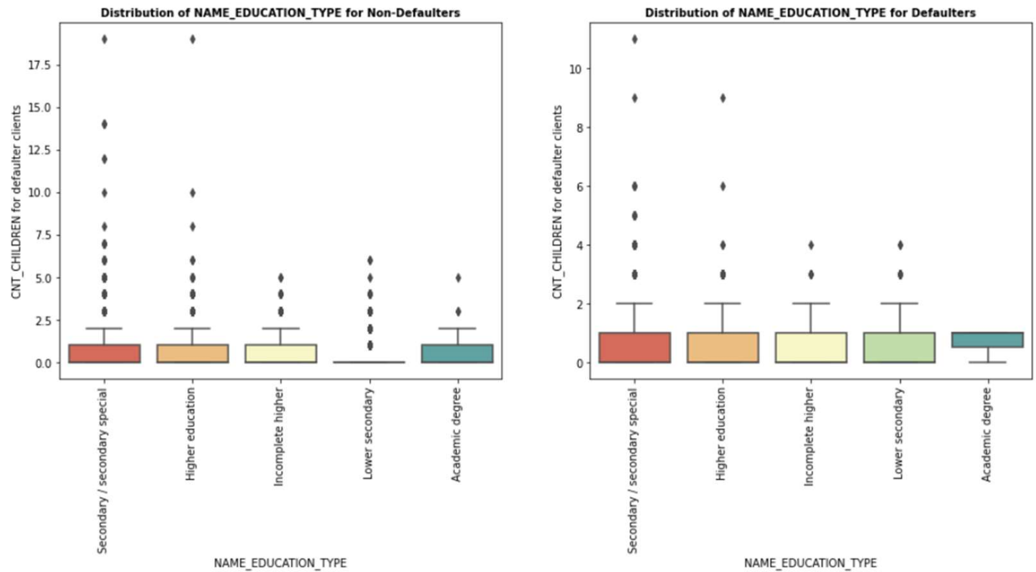




```
In [42]: appdf_plot_bivariate_cat_cont('NAME_EDUCATION_TYPE','AMT_CREDIT')
```




```
In [45]: appdf_plot_bivariate_cat_cont('NAME_EDUCATION_TYPE','CNT_CHILDREN')
```

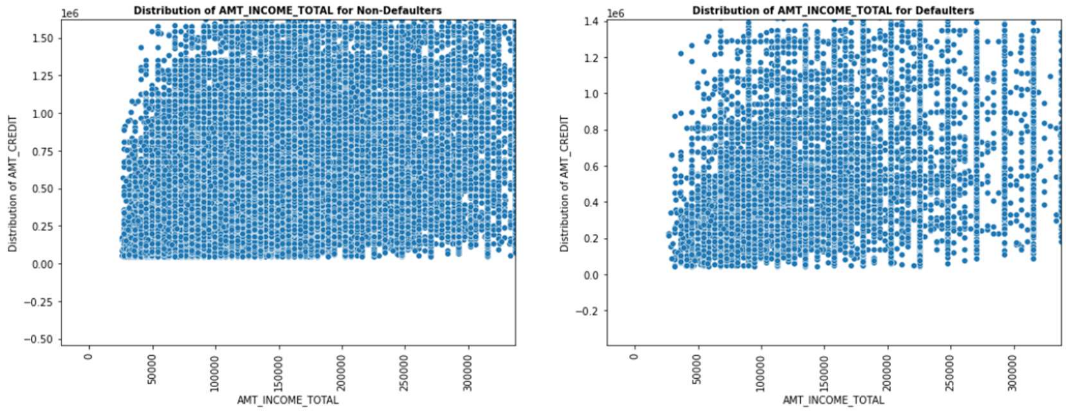



```
def appdf_plot_bivariate_cont_cont(var_cont1,var_cont2):
    plt.figure(figsize=(18,6))
    plt.subplot(1, 2, 1)
    sns.scatterplot(x=var_cont1,y=var_cont2, data=appdf0, palette='Spectral')
    plt.title('Distribution of '+ '%s' %var_cont1 +' for Non-Defaulters', weight='bold', fontsize=10)
   # plt.xlabel(var_cont1)
    plt.xticks(rotation=90)
    plt.ylabel('Distribution of '+ '%s' %var_cont2)
    #Below steps are to fing the IQR range to ignore the Outliers
```

```
#For X-Axis
xIQR=1.5*(appdf0[var_cont1].quantile(.75)-appdf0[var_cont1].quantile(.25))
xlowerlim=appdf0[var_cont1].quantile(.25)-xIQR
xupperlim=appdf0[var_cont1].quantile(.75)+xIQR
#For Y-Axis
yIQR=1.5*(appdf0[var_cont2].quantile(.75)-appdf0[var_cont2].quantile(.25))
ylowerlim=appdf0[var_cont2].quantile(.25)-yIQR
yupperlim=appdf0[var_cont2].quantile(.75)+yIQR
#Applying the limits on the Axis range
plt.ylim(ylowerlim,yupperlim)
plt.xlim(xlowerlim,xupperlim)
#print(xlowerlim,xupperlim)
plt.subplot(1, 2, 2)
sns.scatterplot(x=var_cont1,y=var_cont2, data=appdf1, palette='Spectral')
plt.title('Distribution of '+ '%s' %var_cont1 +' for Defaulters', weight='bold',fontsize=10)
#plt.xlabel(var_cont1)
plt.xticks(rotation=90)
plt.ylabel('Distribution of '+ '%s' %var_cont2)
#Below steps are to fing the IQR range to ignore the Outliers
#For X-Axis
xIQR=1.5*(appdf1[var_cont1].quantile(.75)-appdf1[var_cont1].quantile(.25))
xlowerlim=appdf1[var_cont1].quantile(.25)-xIQR
xupperlim=appdf1[var_cont1].quantile(.75)+xIQR
#For Y-Axis
yIQR=1.5*(appdf1[var_cont2].quantile(.75)-appdf1[var_cont2].quantile(.25))
ylowerlim=appdf1[var_cont2].quantile(.25)-yIQR
yupperlim=appdf1[var_cont2].quantile(.75)+yIQR
#print(xlowerlim,xupperlim)
#Applying the limits on the Axis range
plt.ylim(ylowerlim,yupperlim)
plt.xlim(xlowerlim,xupperlim)
plt.show()
```

output:

```
In [48]: appdf_plot_bivariate_cont_cont('AMT_INCOME_TOTAL','AMT_CREDIT')
```



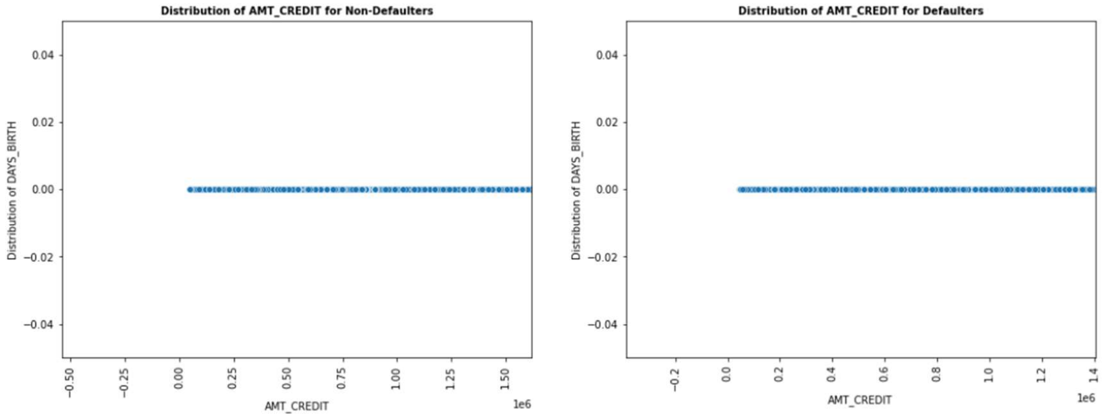
Code:

```
In [46]: def appdf_plot_bivariate_cont_cont(var_cont1,var_cont2):
             plt.figure(figsize=(18,6))
             plt.subplot(1, 2, 1)
             sns.scatterplot(x=var_cont1,y=var_cont2, data=appdf0, palette='Spectral')
             plt.title('Distribution of '+ '%s' %var_cont1 +' for Non-Defaulters', weight='bold', fontsize=10)
           # plt.xlabel(var_cont1)
             plt.xticks(rotation=90)
             plt.ylabel('Distribution of '+ '%s' %var_cont2)
             #Below steps are to fing the IQR range to ignore the Outliers
             #For X-Axis
             xIQR=1.5*(appdf0[var_cont1].quantile(.75)-appdf0[var_cont1].quantile(.25))
             xlowerlim=appdf0[var_cont1].quantile(.25)-xIQR
             xupperlim=appdf0[var_cont1].quantile(.75)+xIQR
             #For Y-Axis
             yIQR=1.5*(appdf0[var_cont2].quantile(.75)-appdf0[var_cont2].quantile(.25))
             ylowerlim=appdf0[var_cont2].quantile(.25)-yIQR
             yupperlim=appdf0[var_cont2].quantile(.75)+yIQR
             #Applying the limits on the Axis range
             plt.ylim(ylowerlim,yupperlim)
             plt.xlim(xlowerlim,xupperlim)
             #print(xlowerlim,xupperlim)
             plt.subplot(1, 2, 2)
             sns.scatterplot(x=var_cont1,y=var_cont2, data=appdf1, palette='Spectral')
             plt.title('Distribution of '+ '%s' %var_cont1 +' for Defaulters', weight='bold',fontsize=10)
             #plt.xlabel(var_cont1)
             plt.xticks(rotation=90)
             plt.ylabel('Distribution of '+ '%s' %var_cont2)
             #Below steps are to fing the IQR range to ignore the Outliers
             #For X-Axis
             xIQR=1.5*(appdf1[var_cont1].quantile(.75)-appdf1[var_cont1].quantile(.25))
             xlowerlim=appdf1[var_cont1].quantile(.25)-xIQR
             xupperlim=appdf1[var_cont1].quantile(.75)+xIQR
```

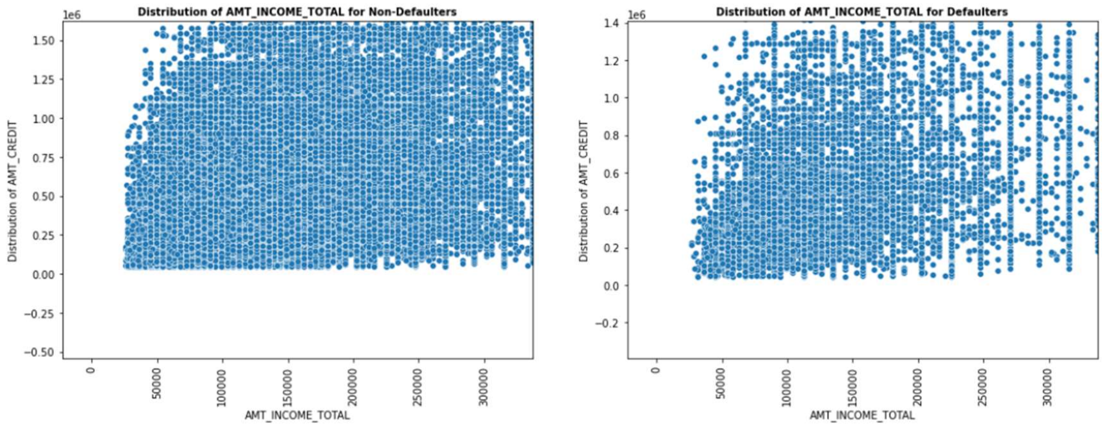
```
             #For X-Axis
             xIQR=1.5*(appdf1[var_cont1].quantile(.75)-appdf1[var_cont1].quantile(.25))
             xlowerlim=appdf1[var_cont1].quantile(.25)-xIQR
             xupperlim=appdf1[var_cont1].quantile(.75)+xIQR
             #For Y-Axis
             yIQR=1.5*(appdf1[var_cont2].quantile(.75)-appdf1[var_cont2].quantile(.25))
             ylowerlim=appdf1[var_cont2].quantile(.25)-yIQR
             yupperlim=appdf1[var_cont2].quantile(.75)+yIQR
             #print(xlowerlim,xupperlim)
             #Applying the limits on the Axis range
             plt.ylim(ylowerlim,yupperlim)
             plt.xlim(xlowerlim,xupperlim)
             plt.show()

In [47]: appdf_plot_bivariate_cont_cont('AMT_CREDIT','DAYS_BIRTH')
```

Output:

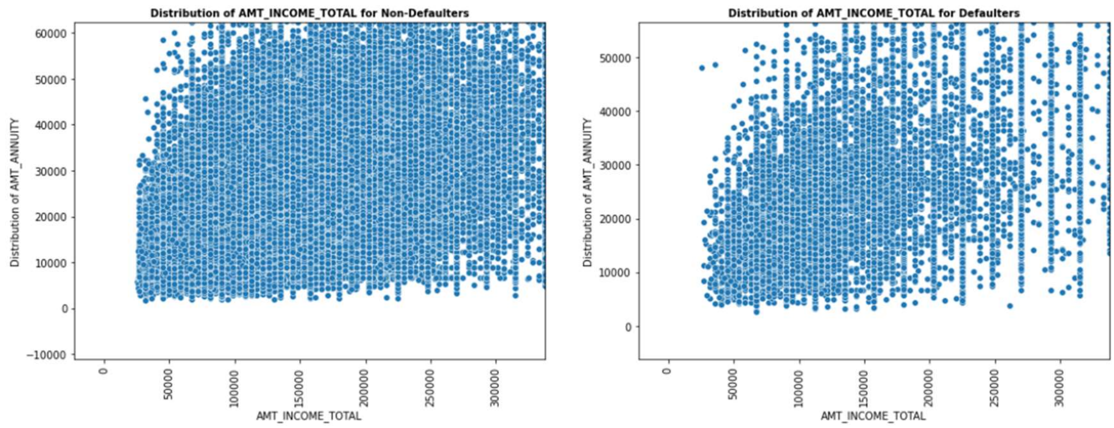
```
In [47]: appdf_plot_bivariate_cont_cont('AMT_CREDIT','DAYS_BIRTH')
```




```
In [48]: appdf_plot_bivariate_cont_cont('AMT_INCOME_TOTAL','AMT_CREDIT')
```



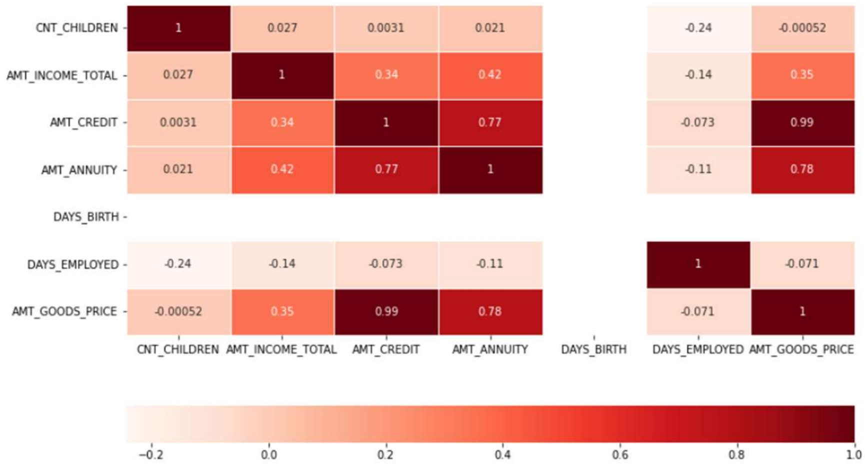

```
In [49]: appdf_plot_bivariate_cont_cont('AMT_INCOME_TOTAL','AMT_ANNUITY')
```



Code:

```
In [52]: def correlation_heatmap(var):
             plt.figure(figsize=(12,8))
             cor = var.corr()
             sns.heatmap(cor,annot=True,linewidths=.5,cbar_kws={"orientation": "horizontal"},cmap="Reds")
             plt.show()
             #Steps to obtain the top correlation.
             indices = np.where(cor > -1)
             indices = [(cor.index[x], cor.columns[y],abs(cor.iloc[x,y])) for x, y in zip(*indices) if x != y and x < y]
             a=sorted(indices, key=lambda x: x[2],reverse=True)
             print("Top Ten Correlations are:")
             for i in range(0,10):
                 print('%d. '%(i+1)+a[i][0]+' and '+a[i][1])
```

```
In [53]: correlation_heatmap(appdf0[['CNT_CHILDREN','AMT_INCOME_TOTAL','AMT_CREDIT','AMT_ANNUITY','DAYS_BIRTH','DAYS_EMPLOYED','AMT_GOODS_
```



output:

```
Top Ten Correlations are:
1. AMT_CREDIT and AMT_GOODS_PRICE
2. AMT_ANNUITY and AMT_GOODS_PRICE
3. AMT_CREDIT and AMT_ANNUITY
4. AMT_INCOME_TOTAL and AMT_ANNUITY
5. AMT_INCOME_TOTAL and AMT_GOODS_PRICE
6. AMT_INCOME_TOTAL and AMT_CREDIT
7. CNT_CHILDREN and DAYS_EMPLOYED
8. AMT_INCOME_TOTAL and DAYS_EMPLOYED
9. AMT_ANNUITY and DAYS_EMPLOYED
10. AMT_CREDIT and DAYS_EMPLOYED
```

VII. CONCLUSION

Credit Analysis is about delivering conclusions based on past, present, and the future situations. The role of credit analysis offers a plenty of opportunities to discover and experience different types of businesses as one retains with a number of clients from different sectors. The career is not only monetarily worthwhile but also helps an individual to improve along with rendering good opportunities to build one's career.

In reality, however, the unsatisfactory level of one ratio is frequently mitigated by the strength of some other measure. A firm, for example, may have a poor profitability ratio but an above-average liquidity ratio. One limitation of the univariate approach is the difficulty of making trade-offs between such weak and strong ratios. Of course, a good credit analyst can make these adjustments. However, some univariate measures – such as the specific industry group, public versus private company, and region – are categorical rather than ratio-level values. It is more difficult to make judgments about variables of this type.

Although univariate models are still in use today in many banks, most academics and an increasing number of practitioners seem to disapprove of ratio analysis as a means of assessing the performance of a business enterprise. Many respected theorists downgrade the arbitrary rules of thumb (such as company ratio comparisons) that are widely used by practitioners and favor instead the application of more rigorous statistical techniques.

Traditionally most banks have relied on subjective judgment to assess the credit risk of a corporate borrower. Essentially, bankers used information on various borrower characteristics – such as character (reputation), capital (leverage), capacity to pay (volatility of earnings), conditions of the customer's business (purpose of the loan), and collateral – in deciding whether or not to make a given loan. These characteristics are commonly referred to as the 5 Cs.[1] Developing this type of expert system is time-consuming and expensive. Incorporating certain soft (qualitative) data in a risk model is particularly demanding, however successful implementation eliminates human error and reduces potential for misuse. That is why, from time to time, banks have tried to clone their decision-making process. Even so, in the granting of credit to corporate customers, many banks continue to rely primarily on their traditional expert system for evaluating potential borrowers.

Credit Analysis entails researching and analyzing the debt profile and debt servicing abilities of individuals, companies or even sovereigns (i.e. countries). A Credit Analyst, therefore, is someone who finds out the creditworthiness of an entity (either an individual or company or country) depending on the demands of the situation. In the case of issuing loans, companies / individual borrowers are appraised to see if they have the ability to service the debt and also if it is safe to give out the loan. In the case of a credit card application, income streams and previous defaults etc. will be analyzed. In the case of countries, although more complex to analyze, the end result is the same – an assessment of risk – also called a 'Credit Rating'. Below is a chart with rating categories used by 3 leading global Credit Rating Agencies.

VIII. Future Scope

A credit analyst mainly focuses on analyzing the financial data of the customer, companies that are applying for credit or loans to determine the risk that the bank, or other lending or credit-granting institution will not recoup funds loaned.

A credit analyst is responsible for assessing a loan applicant's ability to repay the loan and recommending that it be approved or denied. Credit analysts are employed by commercial and investment banks, credit card companies, credit rating agencies, and investment companies. They may also work in the credit departments of a wide range of companies.

References:

Abor, J. & Biekpe, N. 2005. What determines the capital structure of listed firms in Ghana? African Finance Journal, 7(1): 37-48.

Adam, T. & Goyal, V.K. 2008. The investment opportunity set and its proxy variables. The Journal of Financial Research, 31(1): 41-63.

Akinboade, O.A. & Makina, D. 2006. Financial sector development in South Africa, 1970-2002. Studies in Economics and Econometrics, 30(1): 101-127.