Democratizing Access to Supply Chain Finance for Small & Medium Enterprises

by

Daniel Granados Nicholls
Master of Business Administration, Institut Européen d'Administration des Affaires (INSEAD), 2021
Bachelor of Science, Industrial Engineering, Technological of Monterrey, 2018

and

Emre Muzaffer Kulluk
Bachelor of Science, Business Administration, Babson College, 2015

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Signature of Author: ______________________________
Department of Supply Chain Management
May 6, 2022

Signature of Author: ______________________________
Department of Supply Chain Management
May 6, 2022

Certified by: ______________________________
Mehdi H. Farahani
Postdoctoral Associate
Capstone Advisor

Accepted by:
____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Democratizing Access to Supply Chain Finance for Small & Medium Enterprises

by

Daniel Granados Nicholls

and

Emre Muzaffer Kulluk

Submitted to the Program in Supply Chain Management
on May 6, 2022 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

Abstract

Small and Medium Enterprises (SMEs) are 2.6 times more likely to be rejected for a loan than a multinational, and the worldwide trade finance gap for SMEs is estimated at $1.7 Trillion USD. The main barrier to finance for SMEs is the high costs of due diligence during the financing process. Our research partner, a third-party logistics (3PL) provider was interested in exploring using their trade data to inform creditworthiness decisions for SMEs. Previous research has shown that alternative databases can be used to improve the risk assessment of SMEs' creditworthiness, however, we found no evidence that supply chain operational data from 3PLs can be used to improve the creditworthiness assessment of SMEs for trade financing. Through a partnership with a 3PL with a financial institution branch, we collect insights into the challenges and opportunities for 3PLs to leverage their databases to better inform credit scoring decisions for SMEs. We also use two publicly available databases to illustrate the methodology we propose in our research for 3PLs to build their own credit scoring methodologies. We document the proposed features to be explored by 3PLs which to build their own credit scoring models. Aligned with the research on alternative databases, we conclude that the use of operational supply chain data from 3PLs can be useful to strengthen credit scoring models for trade financing of SMEs. In addition, we propose solutions to common challenges drawn from the nature of a 3PL's data structure and initial model iterations (i.e., cold start problem, feature acquisition). Supply chain operational data from 3PLs can be leveraged to build a credit score model and could be a steppingstone for 3PLs to take a central role in the trade ecosystem.

Capstone Advisor: Mehdi H. Farahani
Title: Postdoctoral Associate

Acknowledgements

By Daniel Granados Nicholls

Thanks to my brother, father, mother, and grandparents for supporting me.

Thanks to my friends, especially Sada and Javier for bringing joy to my life.

Thanks to my most recent family members, Anapaula and Tom for their unconditional love, faith, and companionship.

"Knowledge is of no value unless you put into practice."
—Anton C.

By Emre Muzaffer Kulluk

I am grateful to all of those with whom I have had the pleasure to work during this research.

Special thanks to my family, friends, and colleagues for their unreserved support.

"Knowing yourself is the beginning of all wisdom."
—Aristotle

Contents

List of Figures

List of Tables

1. **Introduction**

"The modern era of international trade is one of increasingly complex interactions between people, firms, and organizations. Supply chains cross countries and regions. Trade has become a 24/7 business and good performance in trade requires connectivity along not only roads, rail, and sea, but in telecommunications, financial markets and information-processing. Having inefficient or inadequate systems of transportation, logistics, and trade-related infrastructure can severely impede a country's ability to compete on a global scale." (The World Bank, 2021)

Supply chain trade financing is a function that has evolved from being owned by financial institutions (FIs), into a function that crosses the trade finance ecosystem, requiring participation from FIs, technology providers, suppliers, buyers, and logistics providers (McKinsey & Company, 2021).

Trade finance is an umbrella term that represents the financial instruments and products used by companies to facilitate international trade and commerce. The main goal of trade finance products is to help reduce the risk associated with global trade by reconciling the divergent needs of exporters and importers (Barnier, 2020).

Unlike conventional financing mechanisms that are predominantly used to manage solvency or liquidity, trade financing is utilized to protect against international trade's inherent risks, such as currency fluctuations, political instability, issues of non-payment, or the creditworthiness of one of the parties involved. Examples of trade financial mechanisms include (Barnier, 2020; Beck et al., 2021; Lotte van Wersch, 2019):

- Lending lines of credit issued by banks to help both importers and exporters.

- Using letters of credit (LOCs) to reduce the risk associated with global trade where the buyer's bank guarantees payment to the seller for the goods shipped.
- Factoring through which companies are paid based on a percentage of their accounts receivables.
- Exporting credit or working capital supplied to exporters.
- Insuring shipments and the delivery of goods to protect the exporter from nonpayment by the buyer.

Small and mid-size enterprises: Small and mid-size enterprises (SMEs) are businesses that maintain revenues, assets, or employee levels below a certain threshold. Each country has its own definition of what constitutes an SME. Certain size criteria must be met and occasionally the industry in which the company operates is considered as well (Liberto, 2020; World Bank, n.d.). SMEs play an important role in the economy, employing vast numbers of people and helping to shape innovation. Governments regularly offer incentives, including favorable tax treatment and better access to loans, to help keep them in business (Liberto, 2020).

Our motivation for this study is to improve how trade financing decisions for SMEs are made. Trade finance for SMEs is relevant to promoting trade and consequently economic growth (Demirguc-Kunt et al., 2017; OECD, 2021b). However, SMEs face two key challenges with gaining access to trade financing: (1) counterparty risk is high[1], especially when they lack the access needed to conduct their due diligence; and (2) trade requires working capital investments while SMEs access to capital requirements is constrained. Traditional trade financing instruments

[1] "*Counterparty risk is the likelihood or probability that one of those involved in a transaction might default on its contractual obligation. Counterparty risk can exist in credit, investment, and trading transactions*" (Murphy, 2020).

allow large enterprises to overcome both challenges — the World Trade Organization (WTO) estimates that 80–90% of world trade relies on financing. However, the World Bank estimates that in emerging economies, SMEs account for up to 40% of the gross domestic product (GDP) and more than half of the jobs worldwide (OECD, 2021b; Trade Finance Global, 2020; World Bank, n.d.)[2].

The challenge for SMEs in accessing trade finance is that the relative transaction cost is high for both parties. For SMEs, it typically means that they must dedicate a high share of their human resources to meet the financial institution's requirements. On the other hand, financial institutions have two uneconomical options: provide a labor-intensive due-diligence process, or attempt to automate the process with limited information; both usually result in uneconomical outcomes for the SME and the financial institution (OECD, 2021b).

In this research, we seek to answer the following research question: can supply chain transactional data features[3] be leveraged by third-party logistics operators (3PL) to assess the risk of financing trade for SMEs? Credit scoring models incorporating the transactional data features could allow financial institutions to better assess trade finance riskiness for SMEs. As part of our research, we will identify the key factors that drive trade credit risk for SMEs. We will include supply chain operational variables and financial metrics in our feature selection process. With populated data on our proposed data features, 3PLs could develop a model to determine trade

---

[2] Today, SMEs have less access to trading than larger enterprises. For example, in the manufacturing sector in most OECD countries, only 10-25% of SMEs export whereas 90% of their larger counterparts export. (OECD, 2021, p.7). The contribution of SMEs to trade is higher if we account for indirect exports and other sectors beyond manufacturing, but there is still room for growth. (Ali et al., 2018; Auboin, 2021; International Finance Corporation, 2013; Kim et al., 2019; OECD, 2021b; World Bank, n.d.).

[3] In this capstone paper, we refer to data features as the input variables for credit models (i.e., financial rations, estimated revenue, country of origin, etc.)

financing options for SMEs, and, as a consequence, promote trade growth. Beyond business applications, this research is relevant from an academic point of view because it provides new insights on how to measure trade credit risk with new alternative data sources.

We believe that trade finance decisions could be supported with a credit scoring model that incorporates supply chain data from 3PLs, for example: gross merchandise value, country of origin, country of destination, etc. We also believe that these models would provide increased supply chain visibility and improved risk management practices in addition to improving financing decisions made by 3PLs with financial institutions branches to guarantee a scalable and economical method to make credit decisions for SMEs. 3PLs could also benefit from improved stickiness with SMEs in their trade ecosystem. By improving the measurement of trade risk of SMEs, 3PLs with financial institution branches can make more informed lending decisions and proactively manage the credit risk through the 3PLs. Furthermore, the operational data can be leveraged by financial institutions and 3PLs to improve risk management (Acero et al., 2021; Ng et al., 2021; Sáenz et al., 2018; Sáenz & Revilla, 2014).

We have conducted a collaborative research project with a 3PL with a financial institution branch to understand how to build a credit scoring model which leverages operational supply chain data to measure and manage the trade risk. Our methodology was to review the existing literature to identify the key features for credit score modeling, and to illustrate the credit score modeling process with two publicly available SMEs credit databases. In our methodology, we prioritized using explainable and interpretable models over other techniques such as deep learning to comply with the World Bank recommendation on policy for interpretability of credit decision processes (Brock, 2021; Euromoney Institutional Investor PLC, 2014; The World Bank Group, 2019; World Bank, n.d.).

This document is structured in five chapters: (1) Introduction, (2) Literature Review, (3) Data and Methodology, (4) Results and Analysis, (5) Discussion, and (6) Conclusion.

2. Literature Review

The research question driving our capstone is: can supply chain transactional data features be leveraged by third party logistics operators (3PL) to assess the risk of financing trade for SMEs? Today, SMEs[4] face barriers to access trade finance due to high transaction costs in the due-diligence process for financial institutions. The objective of this research is to understand how a credit scoring model can leverage supply chain operational data to evaluate trade financing riskiness for SMEs, making the due-diligence process scalable and less labor-intensive for the financial institutions in partnership with 3PLs (OECD, 2021a, 2021b).

In the body of the literature review, we synthesize the body of knowledge related to our research question from four angles: (2.1) the state of the trade finance industry; (2.2) the gap and barriers for SMEs' trade financing; (2.3) credit scoring methodologies; and (2.4) the value of operational supply chain data for trade financing.

2.1 The State of the Trade Finance Industry

"*Trade finance is essential to global trade. In many cases, goods simply cannot cross borders without it. This is particularly true in emerging market and developing economies (EMDEs), as risk perception, jurisdictional differences, unfamiliar counterparty relationships, and geographic distances, among other factors, create a need to document and share risk on*

[4] See motivation and problem statement for further details: Small and mid-size enterprises: Small and mid-size enterprises (SMEs) are businesses that maintain revenues, assets, or employee levels below a certain threshold. Each country has its definition of what constitutes an SME. Certain size criteria must be met and occasionally the industry in which the company operates is considered as well (Liberto, 2020; World Bank, n.d.). SMEs play an important role in the economy, employing vast numbers of people and helping to shape innovation. Governments

shipments."(International Finance Corporation, 2020, p. 2). Trade finance is tied with economic and job growth and is crucial to drive development (Beck et al., 2021; International Finance Corporation, 2020; Kim et al., 2019). In Figure 1 we can see that the global trade finance market is composed of core players and enablers (McKinsey & Company, 2021); logistics providers can play a key role in partnership with financial institutions.

Figure 1

The Global Trade Finance Ecosystem Segmented by Core Participants

Note. McKinsey & Company, (2021). *Reconceiving the global trade finance ecosystem. McKinsey & Company.* *https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/reconceiving%20the%20global%20trade%20finance%20ecosystem/reconceiving-the-global-trade-finance-ecosystem-final.pdf?shouldIndex=false*

Up to 80% of trade is financed by credit or credit insurance, but coverage is not uniform. A lack of trade finance is a significant non-tariff barrier to trade. Small and medium-sized enterprises (SMEs) face the greatest hurdles to access affordable trade financing because of barriers to transactions (i.e., Know Your Customer Costs "KYC"[5] and Anti-Money Laundering costs "AML"[6]) (Kim et al., 2019). In some large, developed countries, up to a third of SMEs face such challenges. SMEs account for 20% of US exports, and 40% of EU exports. Globally, over half of trade finance requests by SMEs are rejected, whereas just 7% for multinational companies. Global liquidity tends to be concentrated within the biggest institutions and their clients (WTO, n.d., 2016).

In response to the challenges resulting from the COVID-19 pandemic, governments are looking to their export credit agencies (ECAs) to fill any financing gaps left by the private market and to mitigate the impact of the crisis by engaging in both short-term (ST) and medium- and long-term (MLT) trade finance. In the absence of comprehensive data on trade finance, the brief from the OECD uses surveys to attempt to identify emerging trends. These indicators suggest that ST trade finance is facing access problems (increased costs of ST financing for SMEs and higher rates of rejected applications) while MLT trade finance appears to be relatively resilient (decrease of 34% in volume and 15% in the number of MLT export credit transaction). ECAs may therefore have a role to play in ST trade finance by acting on liquidity and increasing

---

[5] The know your customer costs (KYC) refers to the associated costs to gather data on a client during a due diligence process. The KYC process can be automated or conducted manually, but there is a trade off between accuracy and cost per client.

[6] The anti-money laundering costs "AML" refers to the costs and expenses related to audit and perform due diligences to reduce the risk of conducing transactions that support money laundering.

capacity. However, for MLT trade finance, ECAs might have fewer levers for action, especially if the pandemic is affecting the demand side and reducing the pipeline of projects (OECD, 2021a).

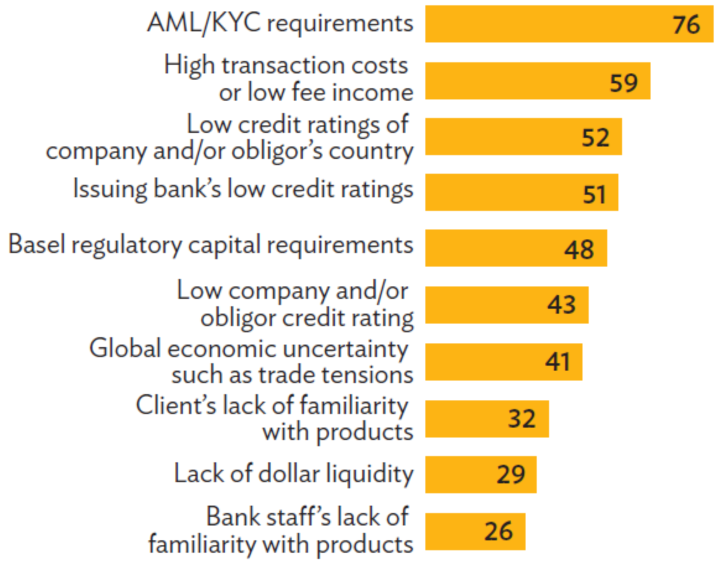
2.2 The Gap and Barriers for SMEs' Trade Financing

SMEs in developing countries face even greater challenges in accessing trade finance. The estimated value of unmet demand for trade finance in Africa is US$ 120 billion (one-third of the continent's trade finance market) and US$ 700 billion in developing Asia. Bridging these gaps in provision would unlock the trading potential of many thousands of individuals and small businesses around the world. Gaps in trade finance provision are highest in new "frontier" countries for trade, where trade opportunities are increasing as global production patterns evolve. The gaps in trade financing provision arise due to a mix of structural and development factors. These gaps are even greater after the 2008-09 financial crisis compounded them. And to stress the problem, local banking sectors are often not equipped to fill the market gap (WTO, 2016).

The main barriers of trade finance, as shown in Figure 2, are related to the lack or high costs of knowledge to accurately assess the creditworthiness of SMEs. According to market surveys, SMEs are 2.6x more likely to be rejected for trade financing than are multinationals (Kim et al., 2019, 2021).

Figure 2

Barriers to Trade Finance (% of Responses)



AML = anti-money laundering, KYC = know your customer.
Source: ADB. 2019 Trade Finance Gaps, Growth, and Jobs Survey—Banks.

Note. Kim et al., (2019). 2019 Trade Finance Gaps, Growth, and Jobs Survey (0 ed., ADB Briefs) [ADB Briefs]. Asian Development Bank. https://doi.org/10.22617/BRF190389-2

With so many businesses deprived of the support that they need to grow, action is needed to address these trade financing gaps. The United Nations Financing for Development agenda, who are already executing various steps to tackle this issue on three fronts, highlight that: first, to encourage global financial institutions to remain engaged and to ensure that regulations are not prohibitive; second, to increase the capacity of local financial institutions; and third, to provide support measures to increase the availability of trade finance via multilateral development banks (WTO, 2016).

2.3 Credit Scoring Methodologies

Credit scoring is a common statistical approach used by financial institutions and lenders of capital to determine the creditworthiness of an individual, business, institutions, and even countries. Credit scores can be used to accept or reject the offering of a loan, determine the riskiness of the counterparty, and determine the pricing of the credit (Brock, 2021; Lando, 2004). Credit scoring should not be confused with credit rating. The latter is mostly used for large enterprises and institutions to assess the counterparty risk of not completing its financial obligations. Credit scores tend to be in any numerical range, whereas credit ratings tend to be classified (e.g., AAA, AA, BB, etc.) (The World Bank Group, 2019).

Credit scoring has evolved. Modern credit scoring was spearheaded by the discriminant technique from Ronald A. Fisher in 1936, the technique is useful to classify observations into non-overlapping groups. The discriminant technique has been used to differentiate creditworthy loans from non-creditworthy loans. (Fisher, 1936; The World Bank Group, 2019). Another example of a mathematical technique is linear programming, currently utilized in modern algorithms to determine credit scores such as myFICO (The World Bank Group, 2019).

The Fair Credit Reporting Act (FCRA) was introduced in the 1970s in the United States to prevent financial institutions from using discriminatory data (i.e., race, gender, etc.) to estimate the credit score (Federal Register, 2012; The World Bank Group, 2019). Recent innovations in statistical methods combined with access to new data sets could enable new credit decisions, which suggests that incorporating new alternative data sources such as 3PL transactional data could lead to improved credit riskiness measurement (The World Bank Group, 2019).

Financial institutions that offer credits to SMEs for trade finance, are also subject to the Basel[7] regulations and could affect their capital requirements. Because of Bassel II,[8] multiple financial institutions realized that improving in-house credit scoring models was not only a requirement in some markets but also an opportunity to improve risk management and improve the profitability of the credit portfolio (Brock, 2021; Cucinelli et al., 2018; Cummings & Durrani, 2016; The World Bank Group, 2019).

In parallel to financial institutions seeking methods to improve their in-house credit scoring models, researchers have developed innovations in the credit scoring methodologies. Innovation in credit scoring is twofold: due to the availability from new data sources, researchers have been able to apply new statistical methodologies. Recent literature on credit scoring builds on such innovations, searching for methodologies that can improve model performance (Dastile et al., 2020; Dias Duarte et al., 2017; Roy & Shaw, 2021; The World Bank Group, 2019), exploring data sources that can provide improved predictive power and studying the effect of different financial instruments (Altman et al., 2020; Bedin et al., 2019; Dias Duarte et al., 2017; Schmidt-Eisenlohr, 2009), and feature engineering to reduce the feature acquisition costs for alternative credit scoring data sources (Angilella & Mazzù, 2015; Kou et al., 2021) The pace of innovation in trade finance has accelerated by the COVID-19 pandemic crisis (Auboin, 2021).

---

[7] "*The Basel Accords are a series of three sequential banking regulation agreements (Basel I, II, and III) set by the Basel Committee on Bank Supervision (BCBS).*

The Committee provides recommendations on banking and financial regulations, specifically, concerning capital risk, market risk, and operational risk. The accords ensure that financial institutions have enough capital on account to absorb unexpected losses." (Chen, 2011)

[8] Basel II incorporated additional requirements on capital adequacy for banks, specifically, it required financial institutions to assess credit risk of assets in the calculation of regulatory capital ratios (Chen, 2020).

Traditional credit scoring methodologies include linear regression, discriminant analysis, and logistic regressions; all are used to predict the outcome or probability of default or non-default on the credit (Beaver, 1966; Lando, 2004; The World Bank Group, 2019). New methodologies include supervised and unsupervised machine learning techniques for credit scoring (i.e., decision trees, random forests, gradient boosting, neural networks, clustering, automated feature engineering, natural language processing, etc.) (Dastile et al., 2020; The World Bank Group, 2019). The drawback of the new methodologies is that they commonly require a higher cardinality of features, and new regulations might restrict the use cases of the new methodologies as they will likely be too complicated to be able to explain the effect of each feature on the outcome (Dastile et al., 2020; The World Bank Group, 2019).

2.4 The Value of Operational Supply Chain Data for Trade Financing

Existing literature explores how alternative data sources (instead of data sources limited to financial metrics) can improve model performance (Gupta et al., 2018; Kou et al., 2021; Roy & Shaw, 2021; Shi et al., 2019). Roy & Shaw, (2021) explored in their research how non-financial data sources can be leveraged with new methodologies to improve baseline credit scoring models. Shi et al., (2019) analyzed the interaction of macroeconomic variables in the default riskiness of loans with farmers in China. Gupta et al., (2018) proved that the size and age of an SME can help to understand the survival probability and by consequence their implicit counterparty default risk. Lee et al., (2012) showed that international sales of Korean SMEs are associated with better survival rates, suggesting that country of origin and destination mix from 3PL could be useful for measuring credit riskiness. Grunert et al., (2005) show that financial institutions which use financial and alternative data sources have better predictive power than using only one of the sources. However, it is key to ensure that the features that are included are designed to reduce feature acquisition costs without sacrificing performance (Kou et al., 2021).

We found no evidence in existing research to prove or disprove that operational supply chain data from 3PLs can enhance trade finance decisions for financial institutions. However, research suggests that it could given the value of alternative sources of data (Ali et al., 2018, 2019; McGuinness et al., 2018; Ng et al., 2021; Song et al., 2016). Supply chain data as an alternative data source can improve trade financing decisions and can even go as far as making firms more creditworthy. Ng et al., (2021) explored how US conflict mineral firms with higher supply chain visibility have improved trade finance and are deemed to have higher creditworthiness. Moreover, Ali et al., (2018) showed that SMEs engaged in trade finance have improved performance, especially in the case of a high degree of trade digitalization. McGuinness et al., (2018) showed that trade finance helped decrease the probability of distress of European SMEs during the financial crisis in the period of 2003-12.

2.5 Conclusion of the Literature Review

Recent literature in credit scoring is focused on twofold innovation: the use of new statistical approaches and the leverage of alternative data sources. The existing literature shows that alternative data sources, combined with new statistical approaches, can improve the predictive power of creditworthiness scoring of SMEs (Altman & Sabato, 2007; Dias Duarte et al., 2017; Georgios, 2019; Gupta et al., 2015, 2018; Kou et al., 2021; Roy & Shaw, 2021; Shi et al., 2019).

We identified that the existing literature does not prove the potential value of incorporating operational supply chain data from 3PLs for SMEs trade finance. Our research will help to understand what the potential of this alternative data source is to improve SME creditworthiness scoring and to potentially reduce the gap for SME trade financing. Our findings, from the literature research, are summarized in Figure 3:

Figure 3

Framework of Findings from Literature Research: Market Structure and Players

	Core participants				Enablers	
Players	Suppliers (i.e., SMEs)	3PLs	Financial institutions	Buyers	Regulators	Trade organizations
Potential role in trade financing	Global trade is commonly part of their strategy for growth and will likely adopt solutions that give them **access to marketplaces** (McKinsey & Company, 2021).	Is naturally poised to **lead the ecosystem** with ownership of the physical & information flow.	Eagerly exploring alternative sources of revenue. **Credit cautious** since 2009-11 crisis and moreover with COVID-19 (OECD, 2021a).	Seeking to lower acquisition cost, specially trading costs. Will likely be interested in solutions **that drive trade costs down** (McKinsey & Company, 2021).	Interested in **promoting job growth** – 60-70% of employment is tied to MSMEs (McKinsey & Company, 2021). Drive credit regulation.	WTO promote **trade financing** for SMEs as a **key driver for economic growth** in terms of GDP and employment (WTO, n.d., 2016).
Relevance & implications for research	**High:** CRM & transactional database will provide direct features for the credit score modeling.	**Highest:** central transactional database used for deriving features for the supply chain credit scoring models.	**High:** central as they can complement transactional information with financial profiling from traditional credit scoring models.	**Medium:** direct features will capture traits of the buyers as they can serve as proxies of quality due-diligence from procurement functions.	**Low:** some of the engineered features will capture indirectly regulators effects (i.e., country of origin & destination).	**Low:** unlikely that some of the engineered features will capture directly or indirectly their effects.

Note. Designed based from McKinsey & Company's, OECD & WTO frameworks for market structuring (McKinsey & Company, 2021; OECD, 2021a; WTO, n.d., 2016).

3. Data and Methodology

Our research objective is to understand which data features of supply chain transactional data can be leveraged to improve trade financing decisions for SMEs. To do so, we have designed the following methodology to analyze and understand the effect of supply chain transactional data features on trade financing decisions for SMEs. We built the methodology to best fit the available data that a 3PL will likely have. First, we will elaborate the data availability constraints for our research and the likely data structure used by 3PLs. In the time frame of our capstone, we did not have access to the records of the database, but we will elaborate on the structure and potential features to be generated. Next, we will develop the methodology used in this capstone to identify potential features to be used by 3PLs to estimate creditworthiness. In this chapter we look at three major themes of the methodology: (1) performing exploratory data analysis; (2) engineering features from supply chain transactional data; and (3) building and testing credit scoring models.

3.1 Data

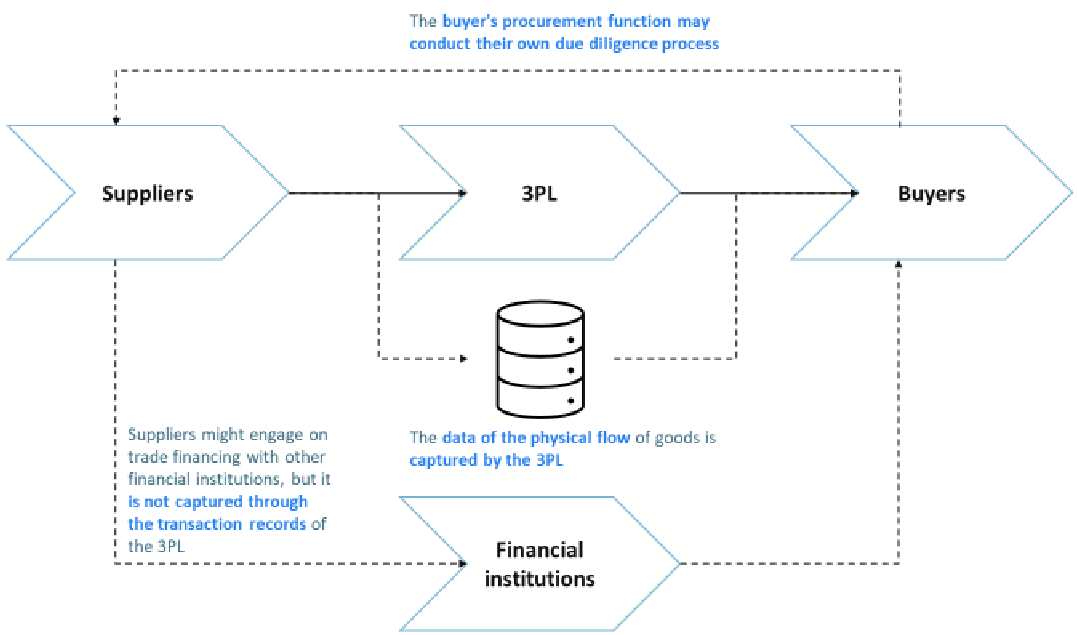
3.1.1 Understanding the Transactional Data Context of a 3PL

We have conducted this research in collaboration with an international 3PL to ensure that the insights are industry applicable. It is relevant to understand the context of the data availability for a 3PL to be able to design a credit score model from the 3PL data.

Based on our collaboration with the industry, we learnt that 3PLs will probably have transactional data records designed to capture the movement of goods as illustrated in Figure 4. Through the partnership, we conducted guided interviews with the data owners of the partner 3PL. The interviews allowed us to understand the structure of the databases and how they are linked with each other and updated as data is generated with each transaction.

Figure 4

Schematic of Typical Database Setting for Transaction Records of a 3PL



Note. Interviews with logistics provider, personal communication, (2021) These databases will tend to be exhaustive on the level of details regarding physical movements and the nature of the goods being transported. It is unlikely though, that a 3PL will have accurate and exhaustive logs of the characteristics of their clients (i.e., stored in Customer Relationship Management software, also known as CRM). Traditional credit scoring methodologies rely on financial ratios calculated from CRMs or databases built from due diligence and review processes (Brock, 2021; The World Bank Group, 2019). Hence, 3PLs will likely lack the traditional databases used for credit scoring (Roy & Shaw, 2021; The World Bank Group, 2017).

In addition, it may be possible to have access to additional databases from the 3PL to boost the predictive power of the supply chain operational data. For example, the payment records tied to the transactions could be leveraged to measure existing payment delays and defaults of SMEs even before being financed. In the case of our partner, we were able to leverage the payment records database and a CRM equivalent database. Lastly, 3PLs interested in trade financing can obtain more data features through purchasing credit scoring of SME databases, using publicly available databases, and building processes to obtain them.

3.1.2 Public Databases Used for Capstone

We used two publicly available databases to build classifier models to assess SME credit riskiness. The two databases are: (1) Polish companies bankruptcy data set from the UCI Machine Learning Repository (Tomczak et al., 2016), and (2) the United States of America national Small Business Administration (SBA) database (Li et al., 2018). We utilized these public databases to illustrate the potential methodologies 3PLs could use to assess the creditworthiness of SMEs and to identify features that could potentially be replicated or estimated from 3PLs' transactional databases.

The Polish companies database has 64 attributes, one dependent variable and 19,967 rows of data. The database records financial ratios and reports if the company defaulted or continued as a going concern. The defaulted companies are from 2000-2012 and the going concern companies are from 2007-2013 (Tomczak et al., 2016). This database was useful for our capstone to illustrate if financial ratios are useful for classification models. This database also helped us better understand if 3PLs could estimate some of the data features from the transactional database or record them in a new database from the feature acquisition process during the loan generation.

The SBA database, after cleaning, had 14 attributes, one dependent variable and 465,629 rows of data. The SBA database has attributes from multiple alternative data sources including geographic location, properties of the SME (e.g., the number of employees), and characteristics of the loan (e.g., size of the loan, portion guaranteed by collateral, etc.). The SBA database measures if an SME defaulted or repaid its loan obligation. The data used is from 1987 to 2005 (Li et al., 2018). This database was useful for our capstone to illustrate how 3PLs could leverage alternative data sources to build a credit score model.

The two databases were useful for this capstone project because we were able to illustrate a viable methodology to build the first iteration of a credit score model for SMEs. Additionally, we were able to identify that model selection is key to improve performance of the riskiness assessment. Finally, the databases were useful to understand to what degree could a 3PL estimate the key features used in traditional credit scoring models.

3.2 Methodology with Public Databases

3.2.1 Performing Exploratory Data Analysis

Our purpose of performing exploratory data analysis (EDA) was to understand the quality and exhaustiveness of the data. We performed three key EDAs on each database: (1) cleaning of null values, (2) distribution analysis, and (3) Pearson correlation analysis (Pearson, 2018; Tukey, 1977). We eliminated null values without imputation due to the large size of the dataset. We eliminated the null values during the data cleaning process. We performed distribution analysis to understand the risk of outliers and adjusted if needed. The Pearson correlation matrix allowed us to identify risks of multicollinearity in the models.

3.2.2 Engineering Features from Supply Chain Transactional Data

3PLs are unlikely to have access to financial data from the SMEs but they will commonly have access to transaction databases. Credit scoring models built from transactional databases require feature engineering to be able to leverage their supply chain transactional data for credit score modeling. Transactional databases are not designed to capture features directly relevant to credit scoring modeling. For example, some key features to determine SMEs' credit score are their age and the number of recent transactions, but these features are not directly found in a transaction database; instead, they must be must be engineered (Gupta et al., 2018; Kou et al., 2021). In Section 4.1.2 we present the synthesis of proposed features to be explored in future research to build a credit scoring model for SMEs with the transactional database of a 3PL.

3.2.3 Building and Testing Credit Scoring Models

For both databases, we built five models to predict the classification: (1) logistic regression for classification (Eq. 1) (Jurafsky & Martin, 2021), (2) single tree model (Eq. 2) (Romadhan, 2021; Therneau et al., 2022), (3) single tree model with weights for unbalanced dataset (Eq. 3) (Romadhan, 2021; Therneau et al., 2022), (4) Bayesian classifier (Eq. 4) (Malhotra, 2017), and (5) logistic regression with LASSO regularization (Eq. 5) (Jurafsky & Martin, 2021). For equations 1 to 5, we have defined the following variables:

- $\hat{\theta}$ is the loss function the model is trying to fit the data to.
- $y^{(i)}$ is the value the model predicts.
- $\hat{y}^{(i)}$ is the observed value.
- $x^{(i)}$ is the value the input variable vector.
- θ_j is a hyperparameter that the model learns during training and validation.
- α is a secondary hyperparameter that we can find using cross validation to specify the size of the regularization term.

Model (1) logistic regression

The logistic regression model estimates the probability of the discrete outcome observed given a set of input variables. The logistic regression models are commonly used with a cutoff to decide how to assign the predicted probability to an outcome (Jurafsky & Martin, 2021).

$$\hat{\theta} = argmax_{\theta} \sum_{i=1}^{m} logP\left(y^{(i)} \middle| x^{(i)}\right) \qquad \text{Eq. 1}$$

Model (2) single tree model (CART)

The single tree model is a predictive model that splits the outcomes based on cutoff rules for each variable and it branches to have multiple layers of cutoffs (Romadhan, 2021; Therneau et al., 2022).

$$\hat{\theta} = argmin_{\theta} \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})^2 \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})^2 \qquad \text{Eq. 2}$$

Model (3) single tree model (CART) with weights

The single tree model is similar to the single tree model but it has been adjusted with weights for each observation to influence the relevance of certain type of mistakes. It allows to build a model that penalizes one mistake more than other classification errors (Romadhan, 2021; Therneau et al., 2022).

$$\hat{\theta} = argmin_{\theta} \sum_{i=1}^{n} \theta_j (y^{(i)} - \hat{y}^{(i)})^2 \sum_{i=1}^{n} \theta_j (y^{(i)} - \hat{y}^{(i)})^2 \qquad \text{Eq. 3}$$

Model (4) Bayesian classifier

The Bayesian classifier model, as the name suggests, leverages the Bayesian principle to train a classifier model given an observed input vector $x^{(i)}$. The model maximizes the probability of correctly classifying an observation given an input vector (Malhotra, 2017).

$$\hat{\theta} = argmax_{\theta} P\left(y^{(i)} \middle| x^{(i)}\right)\left(y^{(i)} \middle| x^{(i)}\right) \qquad \text{Eq. 4}$$

Model (5) Logistic regression with LASSO regularization

The logistic regression model with LASSO regularization is similar to the logistic regression, except it includes the LASSO regularization term which reduces the weight of the parameters from the original model. This causes the new model to be less likely to overfitting and improves the possibility for a better out of sample performance (Jurafsky & Martin, 2021).

$$\hat{\theta} = argmax_{\theta}\left[\sum_{i=1}^{m} logP\left(y^{(i)}\middle|x^{(i)}\right)\right] - \alpha\sum_{j=1}^{n}\left|\theta_j\right| \quad \text{Eq. 5}$$

Overfitting can be a challenge when building predictive models in model training so we divided the data into training (70% of the data) and testing datasets (30% of the data) (Bishop, 2006; Cady, 2017; Kahloot & Ekler, 2021). We will test the performance of the five classification models using accuracy rate, sensitivity, specificity, and the area under the curve[9] in Eq. 6 through 9, respectively (Cady, 2017).

Performance metrics for classification models

$$Accuracy\ rate = \frac{True\ positive + True\ negative}{True\ \&\ false\ positive + True\ \&\ false\ negative} \quad \text{Eq. 6}$$

$$Sensitivity = \frac{True\ positive}{True\ positive + False\ negative} \quad \text{Eq. 7}$$

$$Specificity = \frac{True\ negative}{True\ negative + False\ positive} \quad \text{Eq. 8}$$

$$AUC = (1 - False\ positive\ rate) \cdot (True\ positive\ rate) \quad \text{Eq. 9}$$

---

[9] "*AUC is an effective way to summarize the overall diagnostic accuracy of the test. It takes values from 0 to 1, where a value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test.*" (Mandrekar, 2010)

4. Results and Analysis

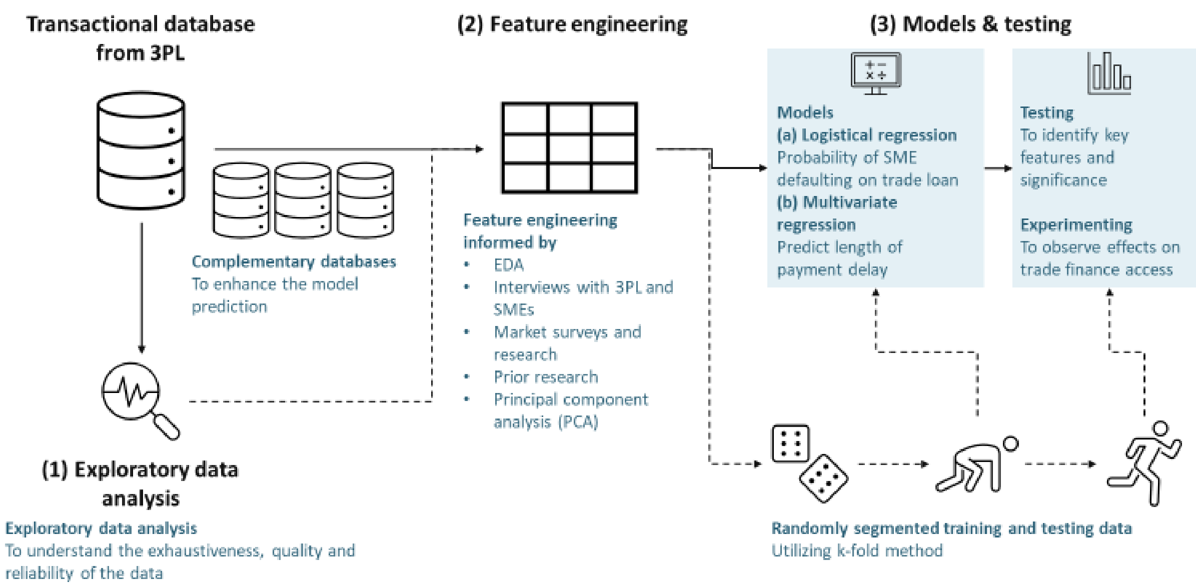
4.1 Results from Interviews

4.1.1 Proposed Methodology for a 3PL to Build a Credit Score Model

Based on the interview process and the literature review, in Figure 5 we propose how a 3PL could leverage a transactional database to build a credit scoring model to assess the riskiness of financing an SMEs trade.

Figure 5

Methodology Overview



A 3PL first performs their own exploratory data analysis, like the work we have done in Section 4.2.1. Then, the 3PL engineers features from the transactional database and alternative databases. The 3PL can build the features proposed in Section 4.1.2 and Section 4.2.3 Key features. Next, the 3PL decides on the dependent variables to be used for training the credit scoring models. Finally, the 3PL can train the credit scoring models using methodologies such as k-fold (we split the data into multiple subsets where we repeat the training and validation of the model). The resulting model would allow the 3PL to solve the cold start problem[10] for the first time they set up the credit scoring process and for new SMEs from which they lack financial information. In ongoing operations, the 3PL should also build into the credit scoring model the features captured from traditional credit databases; this information would need to be collected from the trade financing process (i.e., delay on payment on loans, amount securitized with goods).

4.1.2 Proposed Features for a 3PL to Use in a Credit Score Model

In this section we have aggregated the features found through the literature review that could be useful for 3PLs to explore when building a credit score model. Based on our literature review[11] and interviews with industry experts, we have built Table 1 which aggregates the key features identified by database source. These features are independent from our findings from the publicly available databases.

Table 1

Proposed Features for a Credit Scoring Model for Trade Financing of SMEs

Database source	Feature	Source
Transactions - shipment records	Age of SME - adopted from Kou et al., 2021 "Age of company"	Kou et al., 2021
	Trade financing to transaction value ratio - *decision variable*	Adapted from Bedin et al., 2019
	Size of SME estimated as GMV[12]	Adapted from Gupta et al., 2018

[10] Cold start refers to the issue of a model not being able to learn inferences without having gathered sufficient information
[11] See literature review for further details Figure 2, 2.3 Credit scoring methodologies, and 2.4 The value of operational supply chain data for trade financing
[12] GMV = Gross Merchandise Value

	Survived economic crisis (i)	Adapted from McGuinness et al., 2018
	Number of transactions in time (i)	**Proposed by our capstone**
	GMV in last (i) months(s)	
	Time since last transaction	
	Avg. GMV of transactions	
	Classification of goods shipped	
Transactions - payments records	**% of transactions paid on time**	**Proposed by our capstone**
	Avg. delay time of payment	
	Number of delayed payments	
Transactions - customer data fields	**Last update of customer fields**	**Proposed by our capstone**
	Reported values (Y/N)	
	% of values reported	
CRM	**Reported values (Y/N)**	**Proposed by our capstone**
	% of values reported	
	Days since last update on CRM	
	Business unit classification for trade operations	
	Internal customer classification for trade operations	
	Size reported by SMEs	
	Revenues reported by SMEs	
External databases	Outlook of industry	Adapted from Roy & Shaw, 2021
	Profitability of industry	
	GDP growth rate in SME home country	Adapted from Cucinelli et al., 2018
	Unemployment rate at SME home country	
	GDP growth rate at SME buyer country	
	Unemployment rate at SME buyer country	
	Enterprise tax records for *t* time	Chi & Zhang, 2017
	Breach of contract for t time	
	Years in industry	
	Industry sentiment index	
	Engel coefficient	
	Historical bankruptcy rate geography location	Kou et al., 2021
	Industry	
	Historical bankruptcy rate by geography and by business sector	
	Historical bankruptcy rate of entity's legal form	
	Type of firm	
Management - not available at time of research, to	Education and experience	Roy & Shaw, 2021
	Integrity commitment	
	Succession planning	
	Financial flexibility & group support	
	Credit history	

be acquired through due diligence process	Repayment period	
	Compliance	
	Government Approvals	
	Audit of account	
	Registered capital classification	Chi & Zhang, 2017
	Personal tax records	
	Number of managers	Kou et al., 2021
	Number of shareholders	
Financial - not available at time of research, to be acquired through due diligence process	EBITDA Margin	Georgios, 2019
	Profit per employee	
	Accounts receivable turnover	Chi & Zhang, 2017
	Cash conversion cycle	
	Revenue growth	
	Retained earnings growth	
	Current ratio	Roy & Shaw, 2021
	Quick ratio	
	Leverage	
	Debt-Equity ratio	
	Outside liabilities - net worth	
	Proprietary ratio	
	Debt service coverage	
	Interest coverage	
	Fixed assets coverage	
	Stock' turnover	
	Debtors' turnover	
	Creditor' turnover	
	Return on capital	
	Operating profit ratio	
	Net profit ratio	

Note. Features collected and adapted from multiple papers (Bedin et al., 2019; Chi & Zhang, 2017; Cucinelli et al., 2018; Georgios, 2019; Gupta et al., 2018; Kou et al., 2021; McGuinness et al., 2018; Roy & Shaw, 2021).

In addition, we identified a key challenge that most likely 3PLs will also face: how to build a credit scoring model with the lack of labeled data on completed loans and defaulted loans. The challenge is that 3PLs probably would not have labels on which SMEs went bankrupt or defaulted on their loans. From the literature review we concluded that 3PLs will require labels to measure the performance on their credit scoring model[13]. We recommend transforming the problem from an unlabeled dataset to a labeled dataset.

To do so, we recommend performing feature engineering to obtain labels that could measure the trade finance riskiness of SMEs by measuring how effectively the SME fulfills loan obligations and potentially if an SME that could have defaulted (Chopra & Packt Publishing, 2019; Dong & Liu, 2018; Duboue, 2020; Ozdemir, 2018). We propose the following features because 3PLs with access to transactional databases like the data structure we reviewed, should be able to generate the features even in cold start.

- (a) Probability of an SME defaulting: The 3PL could estimate if an SME defaulted based on their shipment behavior. For example, if an SME stopped activity with the 3PL abruptly they could label it as an SME which defaulted. A model trained on this dependent variable could learn patterns on the behavior of SMEs which will stop operations with the 3PL.
- (b) Expected delay in payment: From the interviews with our partner 3PL, we identified that in addition to the transactional database, the 3PL has a payment database on the shipments. The delay on payment could be used as a proxy for the delay on repayment of trade finance. This engineered feature could be used to solve cold start challenges of 3PLs. In the long term, 3PLs would have data on the delay of payment on the loans that they could use to label the default instead of using an estimate.

---

[13] We want to express our gratitude to the Senior lecturer in Operations Research and Statistics from the Massachusetts Institute of Technology, Dr. Mohammad Fazel Zarandi for his guidance between using supervised or unsupervised training (personal communication, November 15, 2021).

The designed target features have limitations. Feature (a) assumes that all SMEs which discontinued transactions with the 3PL were due to default or bankruptcy, however, it could be that the SME decided to opt out of international trade or chose to operate with another 3PL, also known as churning. Feature (b) assumes that the delay in payment was due to risk of default on the trade receipt.

Finally, we recommend that 3PLs perform principal component analysis to reduce the dimensionality of the features dataset and identify the key components driving the predictions of the model (Abdi & Williams, 2010; Jolliffe, 2002).

4.2 Results from Public Databases

4.2.1 Exploratory Data Analysis

As described in Section 3.2.1, we performed three key EDAs on each database: (1) cleaning of null values, (2) distribution analysis, and (3) Pearson correlation analysis (Pearson, 2018; Tukey, 1977).

(1) Cleaning of null values: We first eliminated all the rows with any null value and measured the dimensionality of the data base. In Table 2 we share the results on dimensionality after performing the data cleaning.

Table 2

Cleaning of Null Values Output

Database	Dimensionality	Null values	Resulting rows
Polish companies	$43{,}405 \times 65$	41,322	19,967
U.S. SBA SME	$535{,}774 \times 21$	982,030	465,629

(2) Distribution analysis: We plotted the histograms and continuous distributions of the numeric variables to look for any key outlier that could need to be excluded. From the histograms and distribution plots we observed that the parameter values are highly concentrated for the Poland companies database. For the SBA database, we observed that each value shows a different distribution with no outstanding cases of outliers. We will illustrate the learnings from the exploratory data analysis using one variable for each of the databases, the full plots are in the Appendix (i): histograms and distribution analysis of databases in Figures 14 to 17.

For example, in Figure 6 we observed the histogram analysis of the logarithm of total assets from the Polish database. In Figure 7 we observed the distribution analysis for the same variable. From both plots, we concluded that there were no outliers, and that the distribution resembles a normal distribution.

Figure 6

Histogram Analysis for Poland Database [example]

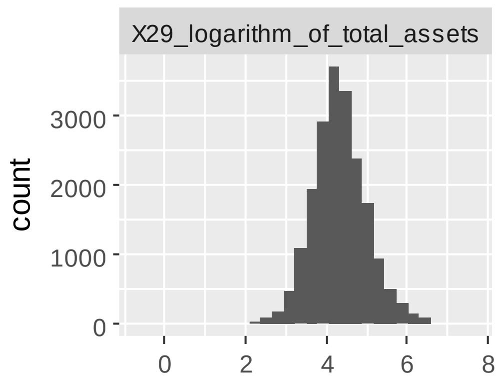
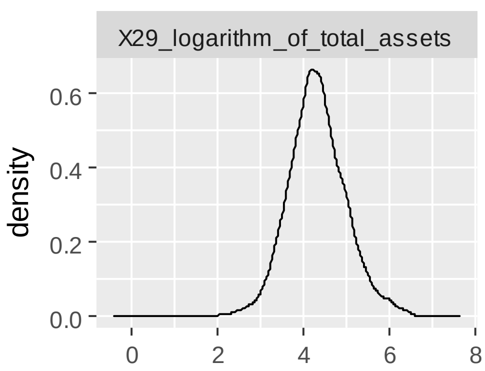


Figure 7

Distribution Analysis for Poland Database [example]



In Figure 8 we observed the histogram analysis of the industry classifier code from the U.S. SBA SME database. In Figure 7 we observed the distribution analysis for the same variable. From both plots, we concluded that there were no outliers, that the values are concentrated on specific codes that might be better represented as categorical data, and that there is no clear distribution pattern that follows a generalized distribution model.

Figure 8

Histogram Analysis for U.S. SBA SME

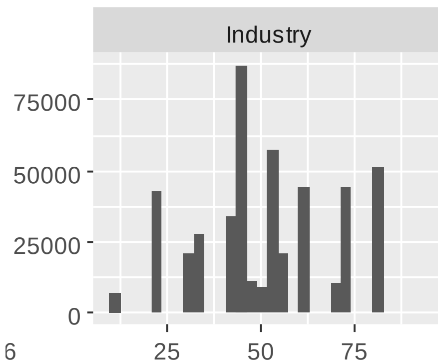
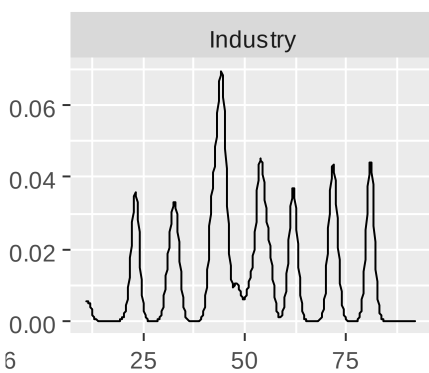


Figure 9

Distribution Analysis for U.S. SBA SME



(3) Pearson correlation analysis: In Figure 10 and Figure 11 we calculated the pearson correlation map between variables for each database. We did not observe cases of unexpected perfect correlation, suggesting that all features are linearly independent from each other. We observed few instances of high correlation, but these were not surprising (i.e., in the SBA SME database we identified that SBA_appv and DisbursementGross were highly correlated, but this was expected, as DisbursementGross is a non-linear transformation from SBA_appv since the guaranteed amount of the loan is a function of the loan amount).

Figure 10

Pearson Correlation Analysis for Poland Database

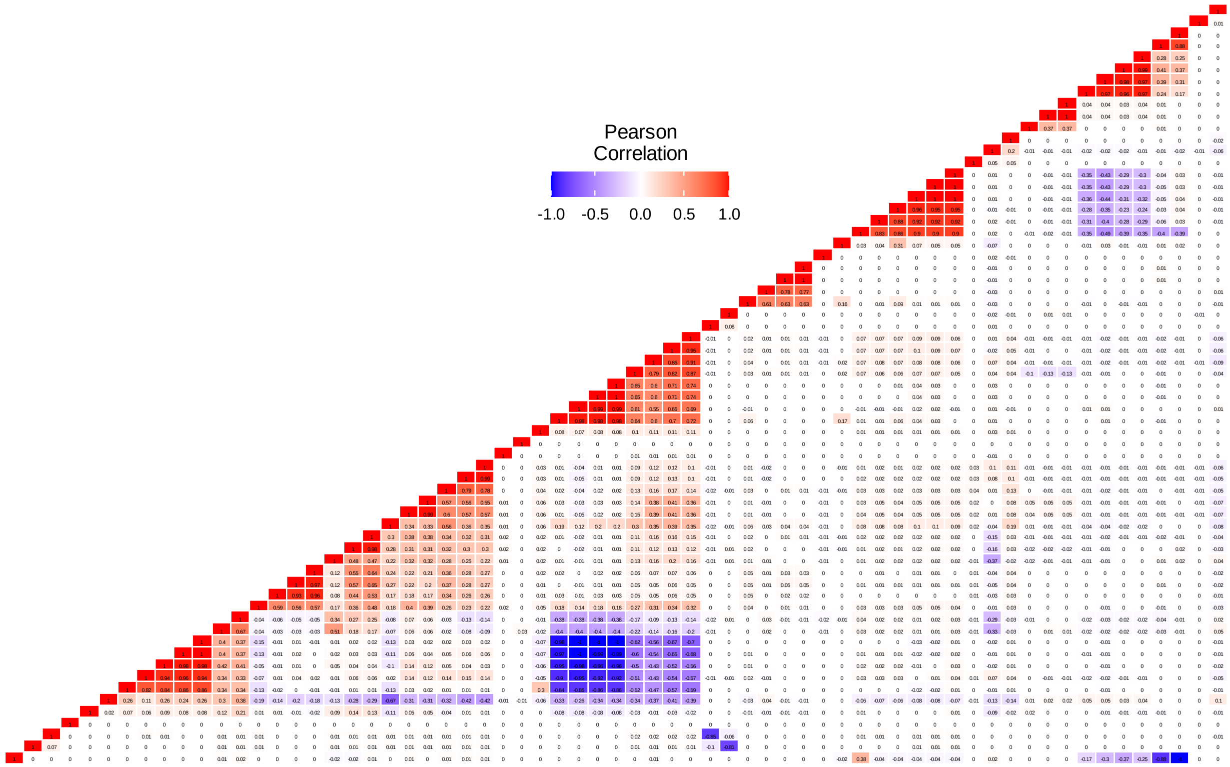
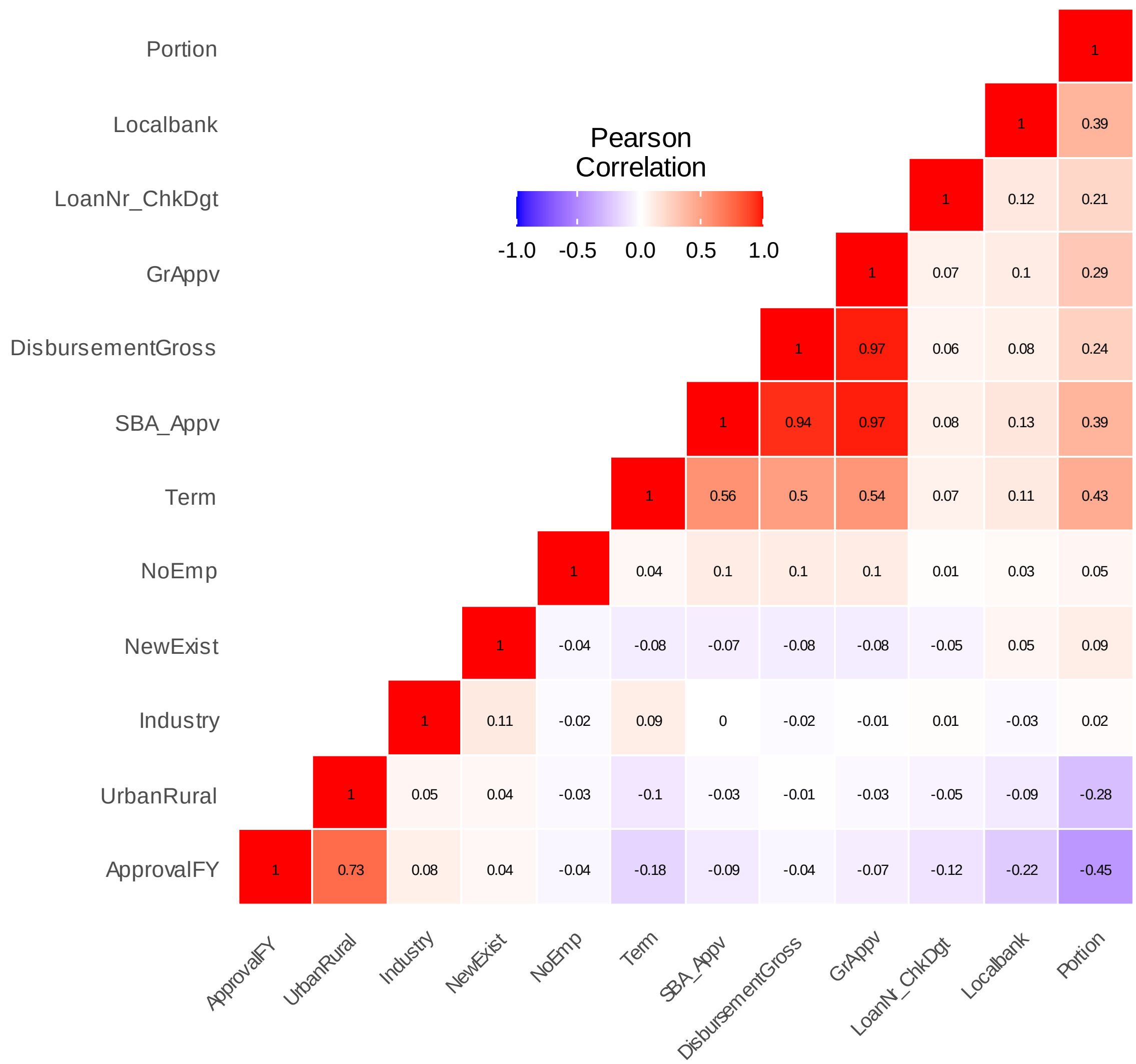


Figure 11

Pearson Correlation Analysis for U.S. SBA SME Database



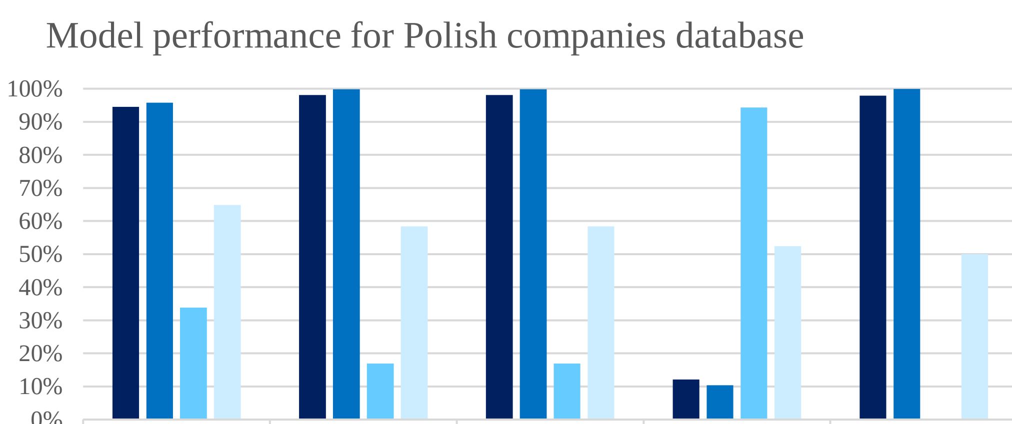
4.2.2 Model Performance

We built five models for each of the two databases. We measured the performance of each model with four metrics described in Section 3.2.3 Building and testing credit scoring models. The model with the best performance for the Polish companies database was the logistic regression with the highest area under the curve. The best model for the SBA database was the single tree (CART) with weights penalizing the errors.

In Figure 12 we show the performance metrics of the five models we built for the Polish companies database. We found that the logistic regression has the best area under the curve, whereas the other models like the CART and logistic regression with LASSO had higher accuracies. The logistic regression outperforms the other models because it has a higher area under the curve and the strongest specificity.

Figure 12

Performance Metrics of Models for Polish Companies

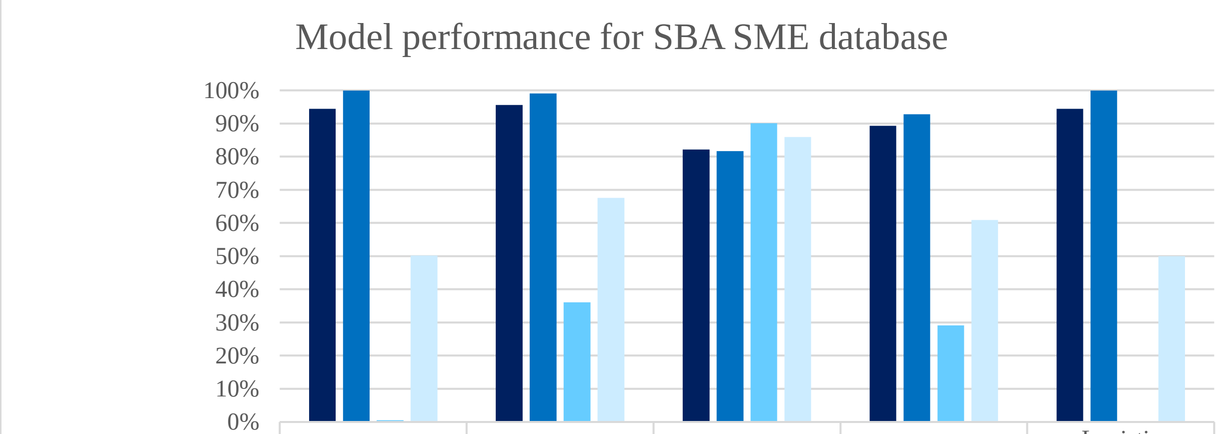


	Logistic regression	CART	CART with weights	Bayesian classifier	Logistic regression with LASSO
Accuracy	94.49%	98.13%	98.13%	12.14%	97.88%
Sensitivity	95.77%	99.85%	99.85%	10.40%	99.95%
Specificity	33.87%	16.94%	16.94%	94.36%	0.05%
Area under the curve	64.82%	58.39%	58.39%	52.38%	49.97%

In Figure 13 we show the performance metrics of the five models we built for the SBA SME database. We found that the CART model with weights has the best performance with the highest area under the curve. Even though the CART with weights has the worst accuracy, the model has the highest specificity and the best capacity to differentiate between both categories given the unbalanced nature of the dataset.

Figure 13

Performance Metrics of Models for SBA SME



	Logistic regression	CART	CART with weights	Bayesian classifier	Logistic regression with LASSO
Accuracy	94.41%	95.56%	82.17%	89.27%	94.44%
Sensitivity	99.93%	99.05%	81.70%	92.80%	99.97%
Specificity	0.52%	36.09%	90.08%	29.11%	0.06%
Area under the curve	50.22%	67.57%	85.89%	60.96%	50.02%

4.2.3 Key Features

In this section, we explore how some of the key features from the best models could be approximated by a 3PL leveraging their transactional databases and the potential risks of errors in the engineered feature estimates.

For the Polish companies database, the logistic regression had the best area under the curve. Some of the key features (as seen in Table 3) could be estimated by the 3PL by leveraging the transactional database.

Table 3

Key Features from Best Performing Model for Polish Companies

Feature	Coefficients
X15_.total_liabilities_._365...gross_profit_.depreciation.	2.73E+19
X8_book_value_of_equity_.total_liabilities	-2.72E+19
X19_gross_profit_.sales	-7.86E+16
X11_.gross_profit_.extraordinary_items_.financial_expenses..total_assets	4.21E+15
X3_working_capital_.total_assets	3.79E+15
X1_net_profit_.total_assets	-3.09E+15
X24_gross_profit_.in_3_years..total_assets	2.47E+15
X20_.inventory_._365..sales	-2.24E+15
X21_sales_.n..sales_.n.1.	-1.51E+15
X45_net_profit_.inventory	-1.51E+15
X44_.receivables_._365..sales	1.51E+15
X39_profit_on_sales_.sales	-1.37E+15
X12_gross_profit_.short.term_liabilities	1.32E+15
X49_EBITDA_.profit_on_operating_activities_.depreciation..sales	1.18E+15
X23_net_profit_.sales	-1.17E+15
X2_total_liabilities_.total_assets	-6.89E+14
X27_profit_on_operating_activities_.financial_expenses	-4.72E+14
X4_current_assets_.short.term_liabilities	4.60E+14
X17_total_assets_.total_liabilities	4.19E+14

For example, x44 and x21. x44 is the ratio of receivables divided by sales. This measure estimates how quickly can a company collect payment from its customers, in other words, to what degree is the company financing its customers. The 3PL could estimate this variable only for the trade being financed from the gross merchandise value from their existing transactional databases and the date of payment for the goods shipped in the invoice. The limitation of this estimate is that it could only be done for the portfolio of shipments seen by the 3PL. If the SME has other receivables and sales, the 3PL would not be able to estimate the ratio for the SME, but only for a subset of the SMEs' transactions. On the other hand, x21 is the growth of sales from period to period. This metric can be estimated by the 3PL by tracking the cumulative value of the gross merchandise shipped by the SME for each period. The estimate for x21 has the same limitation as that for x44, as it would only be possible for the 3PL to estimate it for the shipments that they see from the SME, not the full sales achieved by the SME.

However, the two most relevant features in the logistic regression model for the Polish company were x15 and x8, which cannot be estimated by the 3PL without acquiring features during the financing due diligence process. These two measures are both financial metrics that assess the liquidity of the companies and their capacity to pay debt obligations. x15 measures the ratio of the total liabilities divided by the companies' gross profit plus depreciation. The numerator is the companies' current level of debt, and the denominator is an estimate of the operational cash flow available in the period. This metric measures the capacity of the firm to cover its liabilities from its operating cash flow. x8 is the ratio of the book value of equity divided by the total liabilities. It is a ratio used to measure the leverage of the company, based on the distribution of the sources of financing. Higher financial leverage means a riskier company to lend to.

For the SBA SMEs database, the CART tree with weights had the best area under the curve. Some of the key features (as seen in Table 4) could be estimated by the 3PL leveraging the transactional database.

Table 4

Key Features from Best Performing Model for SBA SMEs

Feature	Variable Importance
Term	126,692.94
Portion	19,641.16
GrAppv	12,722.78
DisbursementGross	10,889.07
SBA_Appv	10,606.28
ApprovalFY	7,966.761
State	7,047.61
Localbank	6,489.52

Most of the key features from the SBA SMEs best performing model can be estimated by a 3PL. Three examples of features that could be captured by a 3PL during the financing process include: the term is the duration of the loan, meaning how much of the loan is covered by collateral, and the amount disbursed and the collateral size. There are other features that could also be captured by the 3PL from the current SME information at the transaction database, for example, the geographic location of the SME.

5. Discussion

5.1 Limitations

The main limitation of this capstone paper is that the databases we used for credit score modeling are not from a 3PL. Even though we have found evidence that suggests that a 3PL could build a trade finance risk assessment model for SMEs, our research does not prove that operational transactional data from a 3PL can be used. Our research suggests that a 3PL with trade financing processes could leverage their operational transactional data to build a credit score model with features that would approximate some of the most useful features we observed from financial and other alternative databases.

Furthermore, we have identified that a key constraint for 3PLs building credit scoring models for the first time will be to label the observation for the predicted variables on their training datasets. From our interviews, we have identified that 3PLs will likely not have labeled data for the SME's credit performance, but we have proposed two alternatives that 3PLs can likely use to reproduce the labels.

6. Conclusion

In this chapter we answer our research question: can supply chain transactional data features be leveraged by third-party logistics operators (3PLs) to assess the risk of financing trade for SMEs? The conclusion chapter of this capstone is split into three subsections where we discuss the key insights for management, the future research that should be explored in this topic, and the contribution of this academic capstone.

6.1 Insights and Management Recommendations

From our research, we synthesize our recommendations to management in three key insights: (1) there is strong evidence that transactional databases from 3PLs can be used to build features for a credit score model for SME trade financing; (2) data labeling is a key challenge for cold starting the credit score model; and (3) 3PLs will likely need to acquire features by collecting information during the SME trade financing process.

1. *Transactional databases from 3PLs can be used to build features for a credit score model for SME trade financing.* From the literature review and from the models we built from the publicly available databases, we observed that features from alternative datasets improve the credit score model performance. In the literature review, we found that features such as age and size of the SME, which can both be estimated from a 3PL database (i.e., CRM records or from the transactional database) can improve the assessment of SME riskiness. In the case of the publicly available datasets we evaluated, we found that multiple features could be estimated through the transactional database of a 3PL (i.e., growth of sales, the ratio of receivables to sales, geography).

2. *Data labeling is a key challenge for cold starting a credit score model.* From our interviews, we learned that 3PLs would likely have no records on SMEs that have defaulted or that have defaulted loans. Without data labeling, it would be challenging for the 3PL to build a credit scoring model for SMEs' trade financing riskiness. We have concluded that 3PLs could engineer features, such as delays in payments of receivables, to label their dataset to be able to train their baseline model.
3. *3PLs will likely need to acquire features by collecting information during the SME trade financing process*. From our literature review, we have learned that credit scoring models benefit from alternative data sources, but the models still rely on traditional financial datasets. We have not found evidence to claim that a credit score model built solely on alternative datasets is sufficient. Hence, we would recommend 3PLs to acquire both financial features and alternative features during the financing process.

6.2 Future Research

In this section, we discuss the three main hypotheses we believe future research on the topic of trade financing decisions for SMEs using 3PL supply chain data should focus on.

(1) Which are the key alternative features that a 3PL can generate from their own databases to support the assessment of the trade creditworthiness of an SME? Our research suggests that 3PLs can generate features from their databases to build a credit scoring model, but we have not identified which are the most relevant features. Future research could focus on generating multiple iterations of features and testing which are the most relevant for assessing trade financing risk for SMEs.

(2) Which are the key financial features that a 3PL needs to acquire to improve the assessment of trade creditworthiness of an SME? From our literature review, interviews, and research we have learned that traditional financial features are relevant to measuring the creditworthiness of SMEs (e.g., liability ratios, cash flow ratios, interest coverage ratios). However, different studies have different financial features and we have not found an exhaustive study that proves which financial features are the most relevant. Future research could also explore which financial features are the most relevant in the context of trade financing with the 3PL operational features.

(3) Does giving trade financing to SMEs improve their intrinsic creditworthiness of SMEs? In our literature review, we learned that SMEs' riskiness can benefit from financing, but there are multiple sources of financing (e.g., trade financing, revolving credit, loans). Future research could explore if there is a causality effect on creditworthiness between SMEs with and without trade financing.

6.3 Contribution

In this capstone project we aimed to answer the research question: *Can supply chain transactional data features be leveraged by third-party logistics operators (3PL) to assess the risk of financing trade for SMEs?* We believe they can. SMEs are marginalized from today's trade financing ecosystem due to the high costs of due diligence. As consequence they are 2.6x times more likely to be rejected for a loan than a multinational (Kim et al., 2019, 2021). 3PLs are a key player in the trade industry (McKinsey & Company, 2021) and we believe they could potentially build an ecosystem around financing to serve SMEs.

In our capstone project, we identified three key insights: (1) non-traditional features such as supply chain data-driven features can help strengthen credit scoring models; (2) 3PLs can engineer some of the traditional and non-traditional features for credit scoring models from their transactional and payments database; and (3) interpretable models can be trained to assess SMEs' creditworthiness.

These findings are relevant for 3PLs because it means that they could explore trade financing as a branch to their core trade business. 3PLs could build an ecosystem around the trade business by acquiring interaction points with their customers, such as trade financing. By owning more share of interactions and a higher share of the spend from their consumers on the trade ecosystem, they could potentially benefit from higher customer stickiness, higher organic growth, and higher revenue per customer of the ecosystem.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Acero, B., Saenz, M. J., & Luzzini, D. (2021). Introducing synchromodality: One missing link between transportation and supply chain management. *Journal of Supply Chain Management*, jscm.12269. https://doi.org/10.1111/jscm.12269

Ali, Z., Gongbing, B., & Mehreen, A. (2018). Does supply chain finance improve SMEs performance? The moderating role of trade digitization. *Business Process Management Journal*, *26*(1), 150–167. https://doi.org/10.1108/BPMJ-05-2018-0133

Ali, Z., Gongbing, B., & Mehreen, A. (2019). Supply chain network and information sharing effects of SMEs' credit quality on firm performance: Do strong tie and bridge tie matter? *Journal of Enterprise Information Management*, *32*(5), 714–734. http://dx.doi.org/10.1108/JEIM-07-2018-0169

Altman, E. I., Esentato, M., & Sabato, G. (2020). Assessing the credit worthiness of Italian SMEs and mini-bond issuers. *Global Finance Journal*, *43*, 100450. https://doi.org/10.1016/j.gfj.2018.09.003

Altman, E. I., & Sabato, G. (2007). Modelling Credit Risk for SMEs: Evidence from the U.S. Market. *Abacus*, *43*(3), 332–357. https://doi.org/10.1111/j.1467-6281.2007.00234.x

Angilella, S., & Mazzù, S. (2015). The financing of innovative SMEs: A multicriteria credit rating model. *European Journal of Operational Research*, *244*(2), 540–554. https://doi.org/10.1016/j.ejor.2015.01.033

Auboin, M. (2021). *Trade Finance, Gaps and the Covid-19 Pandemic: A Review of Events and Policy Responses to Date*. 30.

Barnier, B. (2020, July 23). *How Trade Finance Eases Transactions for Importers and Exporters*. Investopedia. https://www.investopedia.com/terms/t/tradefinance.asp

Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, *4*, 71–111. https://doi.org/10.2307/2490171

Beck, S., DiCaprio, A., & Pokharel, S. (2021). *Trade and Supply Chain Finance*. 17.

Bedin, A., Billio, M., Costola, M., & Pelizzon, L. (2019). Credit Scoring in SME Asset-Backed Securities: An Italian Case Study. *Journal of Risk and Financial Management*, *12*(2), 89. https://doi.org/10.3390/jrfm12020089

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Brock, T. (2021, April 8). *Credit Scoring*. Investopedia. https://www.investopedia.com/terms/c/credit_scoring.asp

Cady, F. (2017). *The data science handbook*. John Wiley & Sons, Inc.

Chen, J. (2011, March 10). *Basel Accord*. Investopedia. https://www.investopedia.com/terms/b/basel_accord.asp

Chen, J. (2020, October 31). *Basel II*. Investopedia. https://www.investopedia.com/terms/b/baselii.asp

Chi, G., & Zhang, Z. (2017). Multi Criteria Credit Rating Model for Small Enterprise Using a Nonparametric Method. *Sustainability*, *9*(10), 1834. https://doi.org/10.3390/su9101834

Chopra, S. & Packt Publishing. (2019). *Hands-on feature engineering with Python*. PACKT Publishing.

Cucinelli, D., Battista, M. L. D., Marchese, M., & Nieri, L. (2018). Credit risk in European banks: The bright side of the internal ratings based approach. *Journal of Banking & Finance*, *93*, 213–229. https://doi.org/10.1016/j.jbankfin.2018.06.014

Cummings, J. R., & Durrani, K. J. (2016). Effect of the Basel Accord capital requirements on the loan-loss provisioning practices of Australian banks. *Journal of Banking & Finance*, *67*, 23–36. https://doi.org/10.1016/j.jbankfin.2016.02.009

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 106263. https://doi.org/10.1016/j.asoc.2020.106263

Demirguc-Kunt, A., Klapper, L., & Singer, D. (2017). *Financial Inclusion and Inclusive Growth: A Review of Recent Empirical Evidence*. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-8040

Dias Duarte, F., Matias Gama, A. P., & Paulo Esperança, J. (2017). Collateral-based in SME lending: The role of business collateral and personal collateral in less-developed countries. *Research in International Business and Finance*, *39*, 406–422. https://doi.org/10.1016/j.ribaf.2016.07.005

Dong, G., & Liu, H. (2018). *Feature engineering for machine learning and data analytics* (First edition.). CRC Press/Taylor & Francis Group.

Duboue, P. (2020). *The art of feature engineering: Essentials for machine learning* (First edition.). Cambridge University Press.

Euromoney Institutional Investor PLC. (2014). *SMEs swept up in supply chain finance revolution*. Euromoney Institutional Investor PLC.

Federal Register. (2012). Bureau of Consumer Financial Protection: Fair Credit Reporting (Regulation V). *Bureau of Consumer Financial Protection: Fair Credit Reporting (Regulation V)*, *GAO-12-363R*, 1–5.

Fisher, R. A. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Georgios, K. (2019). Credit risk evaluation and rating for SMES using statistical approaches: The case of European SMES manufacturing sector. *Journal of Applied Finance and Banking*, *9*(5), 59–83.

Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, *29*(2), 509–531. https://doi.org/10.1016/j.jbankfin.2004.05.017

Gupta, J., Barzotto, M., & Khorasgani, A. (2018). Does size matter in predicting SMEs failure? *International Journal of Finance & Economics*, *23*(4), 571–605. https://doi.org/10.1002/ijfe.1638

Gupta, J., Gregoriou, A., & Healy, J. (2015). Forecasting bankruptcy for SMEs using hazard function: To what extent does size matter? *Review of Quantitative Finance and Accounting*, *45*(4), 845–869. https://doi.org/10.1007/s11156-014-0458-0

International Finance Corporation. (2013). *Closing the Credit Gap for Formal and Informal Micro, Small, and Medium Enterprises*.

International Finance Corporation. (2020). *Why Trade Finance Matters—Especially Now* (p. 18).

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

Jurafsky, D., & Martin, J. (2021). Logistic Regression. In *Speech and Language Processing* (3rd ed., p. Chapter 5). https://web.stanford.edu/~jurafsky/slp3/5.pdf

Kahloot, K. M., & Ekler, P. (2021). Algorithmic Splitting: A Method for Dataset Preparation. *IEEE Access*, *9*, 125229–125237. https://doi.org/10.1109/ACCESS.2021.3110745

Kim, K., Beck, S., & Latoja, M. C. (2021). *2021 Trade Finance Gaps, Growth, and Jobs Survey*. Asian Development Bank. https://www.adb.org/publications/2021-trade-finance-gaps-growth-jobs-survey

Kim, K., Beck, S., Tayag, M. C., & Latoja, Ma. C. (2019). *2019 Trade Finance Gaps, Growth, and Jobs Survey* (0 ed., ADB Briefs) [ADB Briefs]. Asian Development Bank. https://doi.org/10.22617/BRF190389-2

Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems*, *140*, 113429. https://doi.org/10.1016/j.dss.2020.113429

Lando, D. (2004). *Credit Risk Modeling*. Princeton. https://press.princeton.edu/books/hardcover/9780691089294/credit-risk-modeling

Lee, H., Kelley, D., Lee, J., & Lee, S. (2012). SME Survival: The Impact of Internationalization, Technology Resources, and Alliances. *Journal of Small Business Management*, *50*(1), 1–19. https://doi.org/10.1111/j.1540-627X.2011.00341.x

Li, M., Mickel, A., & Taylor, S. (2018). "Should This Loan be Approved or Denied?": A Large Dataset with Class Assignment Guidelines. *Journal of Statistics Education*, *26*(1), 55–66. https://doi.org/10.1080/10691898.2018.1434342

Liberto, D. (2020, November 14). *Small and Mid-size Enterprise (SME) Definition*. Investopedia. https://www.investopedia.com/terms/s/smallandmidsizeenterprises.asp

Lotte van Wersch, C. (2019). Statistical Coverage of Trade Finance—Fintechs and Supply Chain Financing. *IMF Working Papers*, *19*(165). https://doi.org/10.5089/9781498317030.001

Malhotra, A. (2017, July 3). *Naive Bayes Theorem*. https://medium.com/@akankshamalhotra24/naive-bayes-theorem-79832d506a63

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

McGuinness, G., Hogan, T., & Powell, R. (2018). European trade credit use and SME survival. *Journal of Corporate Finance*, *49*, 81–103. https://doi.org/10.1016/j.jcorpfin.2017.12.005

McKinsey & Company. (2021). *Reconceiving the global trade finance ecosystem*. McKinsey & Company. https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/reconceiving%20the%20global%20trade%20finance%20ecosystem/reconceiving-the-global-trade-finance-ecosystem-final.pdf?shouldIndex=false

Murphy, C. (2020, December 25). *Counterparty Risk*. Investopedia. https://www.investopedia.com/terms/c/counterpartyrisk.asp

Ng, J., Yeung, A., & Zhang, J. J. (2021). Does Supply Chain Visibility Enhance Trade Credit? Evidence through Operational Improvements. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3907784

OECD. (2021a). *Trade finance in the COVID era: Current and future challenges* (OECD Policy Responses to Coronavirus (COVID-19)) [OECD Policy Responses to Coronavirus (COVID-19)]. https://doi.org/10.1787/79daca94-en

OECD. (2021b). *Trade finance for SMEs in the digital era* (OECD SME and Entrepreneurship Papers No. 24; OECD SME and Entrepreneurship Papers, Vol. 24). https://doi.org/10.1787/e505fe39-en

Ozdemir, S. (2018). *Feature Engineering Made Easy* (1st edition). Packt Publishing.

Pearson, R. K. (2018). *Exploratory Data Analysis Using R*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315382111

Romadhan, A. (2021, August 12). Regression in Decision Tree—A Step by Step CART (Classification And Regression Tree). *Medium*. https://arifromadhan19.medium.com/regrssion-in-decision-tree-a-step-by-step-cart-classification-and-regression-tree-196c6ac9711e

Roy, P. K., & Shaw, K. (2021). A multicriteria credit scoring model for SMEs using hybrid BWM and TOPSIS. *Financial Innovation*, *7*(1), 77. https://doi.org/10.1186/s40854-021-00295-5

Sáenz, M. J., & Revilla, E. (2014). Creating More Resilient Supply Chains. *MIT SLOAN MANAGEMENT REVIEW*, 8.

Sáenz, M. J., Revilla, E., & Acero, B. (2018). Aligning supply chain design for boosting resilience. *Business Horizons*, *61*(3), 443–452. https://doi.org/10.1016/j.bushor.2018.01.009

Schmidt-Eisenlohr, T. (2009, December 7). *Towards a theory of trade finance | Elsevier Enhanced Reader*. *ECO 2009/43*. https://doi.org/10.1016/j.jinteco.2013.04.005

Shi, B., Zhao, X., Wu, B., & Dong, Y. (2019). Credit rating and microfinance lending decisions based on loss given default (LGD). *Finance Research Letters*, *30*, 124–129. https://doi.org/10.1016/j.frl.2019.03.033

Song, H., Yu, K., Ganguly, A., & Turson, R. (2016). Supply chain network, information sharing and SME credit quality. *Industrial Management & Data Systems*, *116*(4), 740–758. http://dx.doi.org/10.1108/IMDS-09-2015-0375

The World Bank. (2021). *Connectivity, Logistics & Trade Facilitation* [Text/HTML]. World Bank. https://www.worldbank.org/en/topic/trade-facilitation-and-logistics

The World Bank Group. (2017). *Alternative Data Transforming SME Finance*.

The World Bank Group. (2019). *Credit Scoring Approaches Guidelines*.

Therneau, T. M., Atkinson, E. J., & Foundation, M. (2022). *An Introduction to Recursive Partitioning Using the RPART Routines*. 60.

Tomczak, S., Zieba, M., & Tomczak, J. (2016, April 11). *UCI Machine Learning Repository: Japanese Credit Screening Data Set*. http://archive.ics.uci.edu/ml/datasets/Japanese+Credit+Screening

Trade Finance Global. (2020). *Trade finance explained an SME guide for importers and exporters*. Trade Financing Global.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley PubCo.

World Bank. (n.d.). *World Bank SME Finance: Development news, research, data* [Text/HTML]. World Bank. Retrieved October 10, 2021, from https://www.worldbank.org/en/topic/smefinance

WTO. (n.d.). *WTO | The WTO and World Customs Organization—The challenges of trade financing*. Retrieved October 10, 2021, from https://www.wto.org/english/thewto_e/coher_e/challenges_e.htm

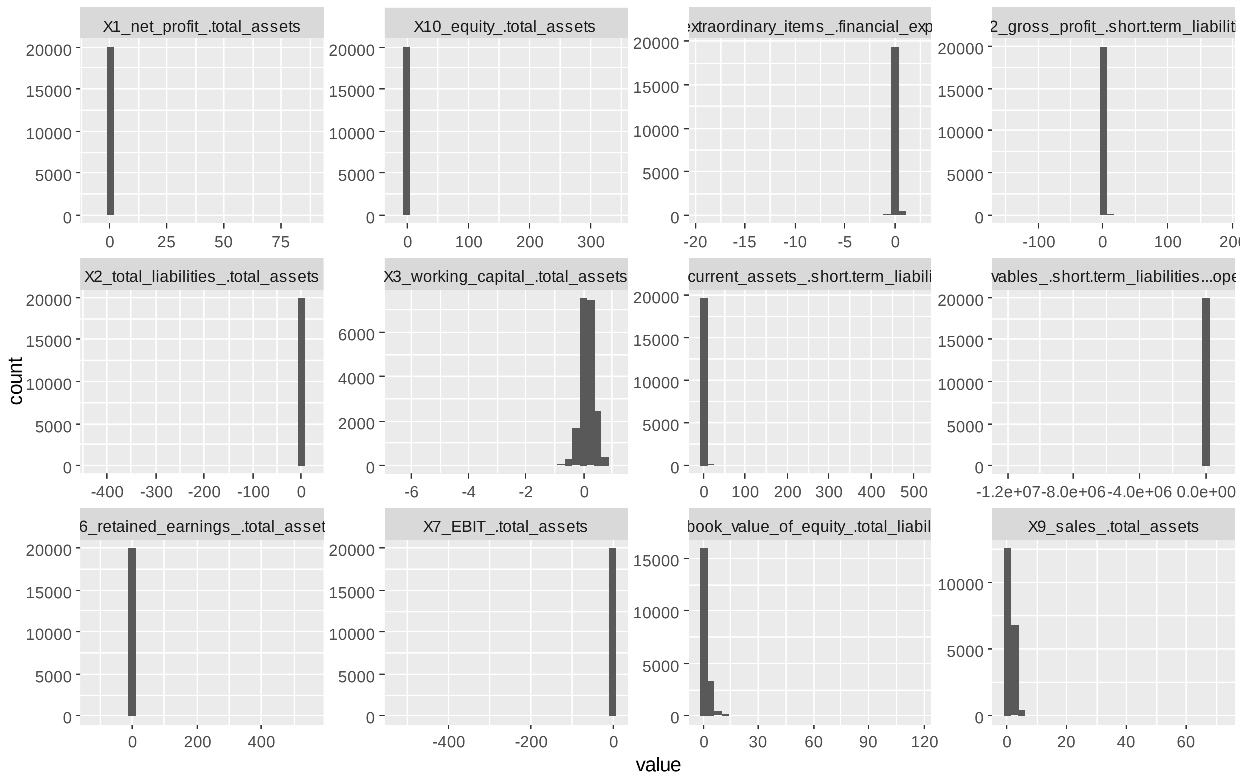
WTO. (2016). *Trade finance and SMEs, Bridging the gaps in provision* (p. 44). World Trade Organization.

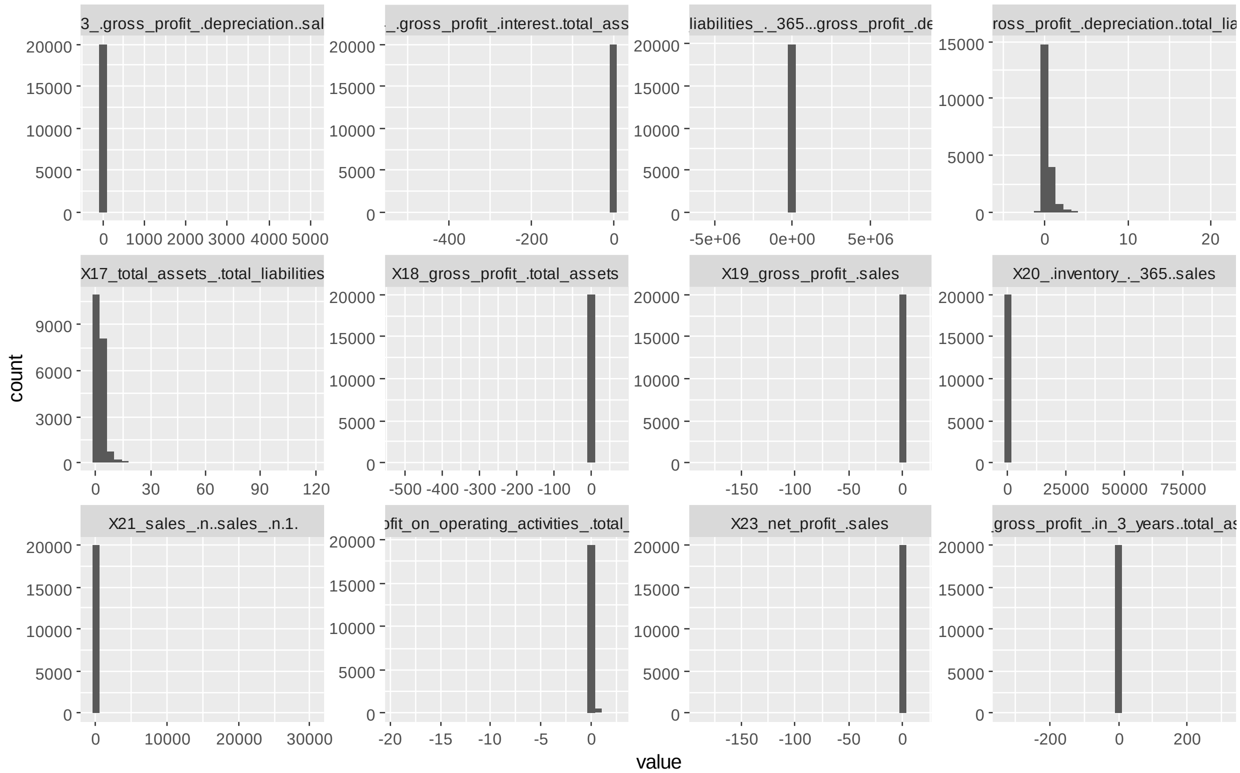
Appendix

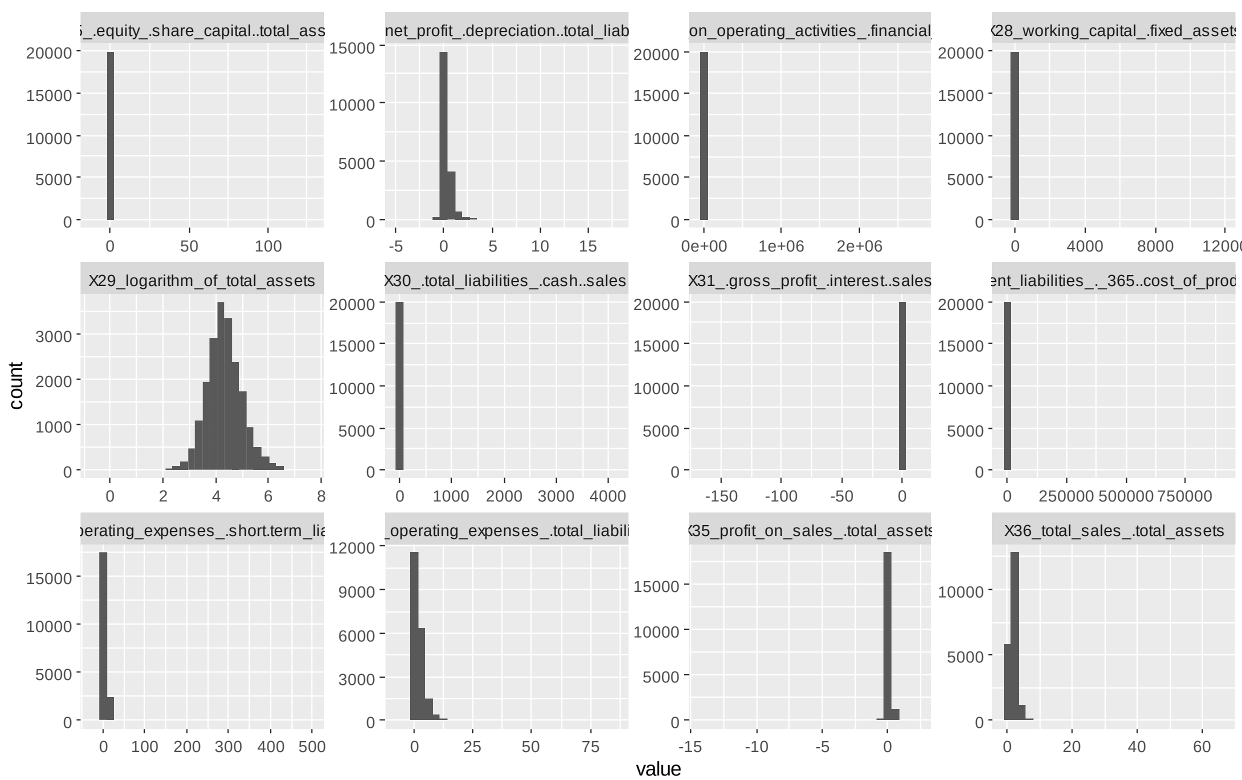
Appendix (A): Histograms and Distribution Analysis of Databases

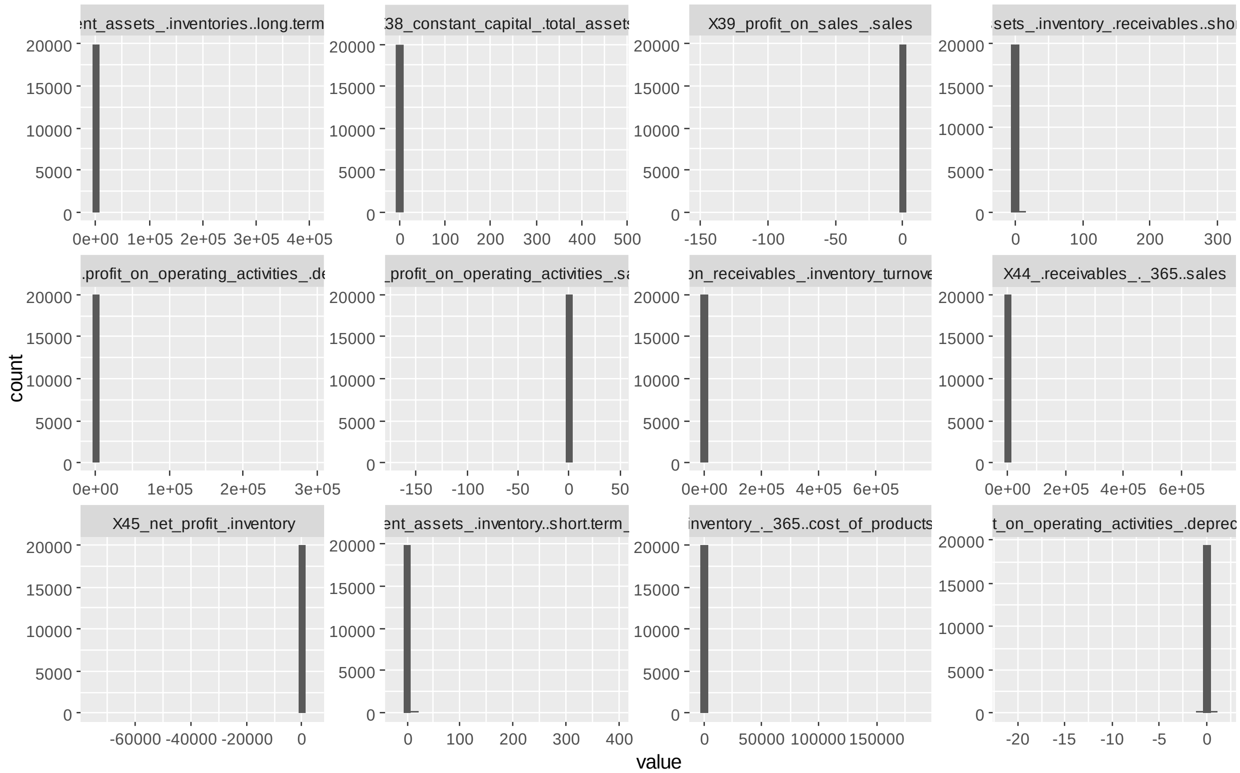
Figure 14

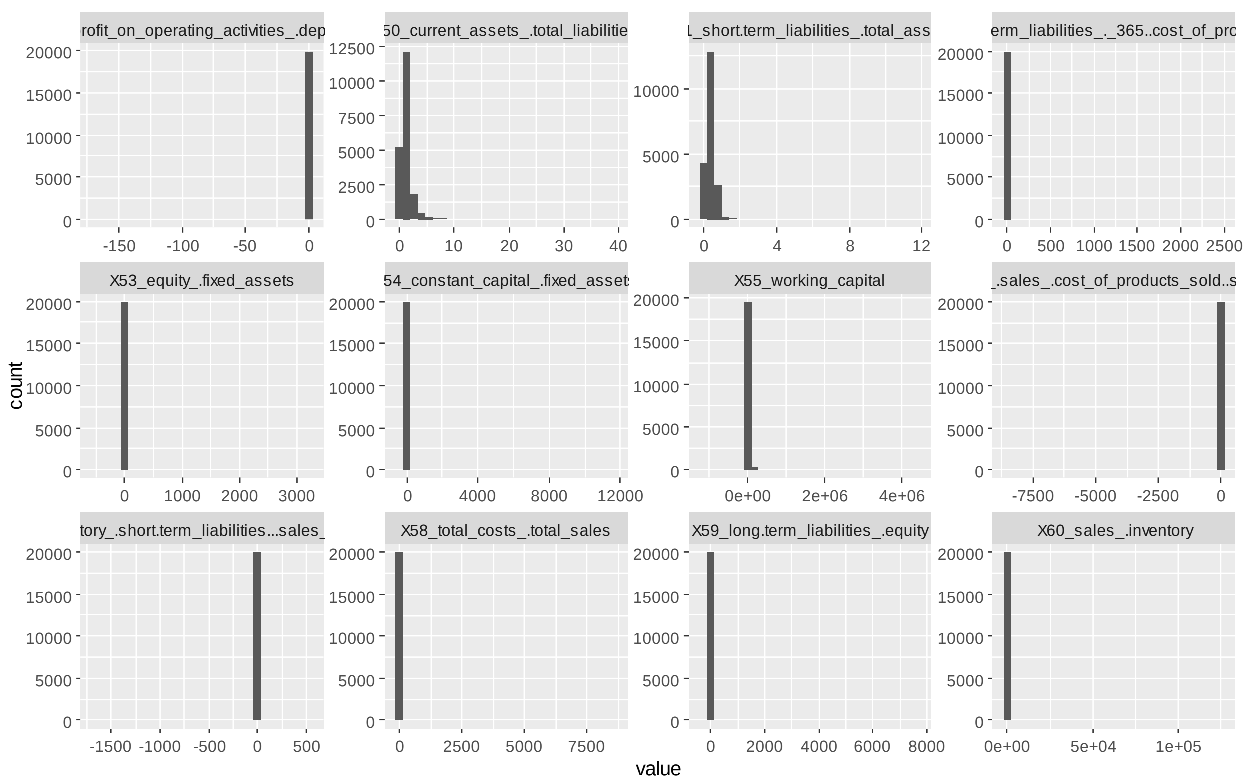
Histogram Analysis for Polish Companies




3_.gross_profit_.depreciation..sal
_.gross_profit_.interest..total_ass
liabilities_._365...gross_profit_.de
ross_profit_.depreciation..total_lia
X17_total_assets_.total_liabilities
X18_gross_profit_.total_assets
X19_gross_profit_.sales
X20_.inventory_._365..sales
X21_sales_.n..sales_.n.1.
ofit_on_operating_activities_.total_
X23_net_profit_.sales
gross_profit_.in_3_years..total_as
count
value


.equity.share_capital..total_ass
net_profit_.depreciation..total_liab
on_operating_activities_.financial
X28_working_capital_.fixed_assets
X29_logarithm_of_total_assets
X30_.total_liabilities_.cash..sales
X31_.gross_profit_.interest..sales
ent_liabilities_._365..cost_of_prod
perating_expenses_.short.term_lia
_operating_expenses_.total_liabili
X35_profit_on_sales_.total_assets
X36_total_sales_.total_assets
count
value


nt_assets_.inventories..long.term
38_constant_capital_.total_assets
X39_profit_on_sales_.sales
sets_.inventory_.receivables..shor
.profit_on_operating_activities_.de
_profit_on_operating_activities_.sa
on_receivables_.inventory_turnove
X44_.receivables_._365..sales
X45_net_profit_.inventory
ent_assets_.inventory..short.term_
inventory_._365..cost_of_products
t_on_operating_activities_.deprec
count
value


rofit_on_operating_activities_.dep
50_current_assets_.total_liabilitie
_short.term_liabilities_.total_ass
erm_liabilities_._365..cost_of_pro
X53_equity_.fixed_assets
54_constant_capital_.fixed_asset
X55_working_capital
sales.cost_of_products_sold..s
tory_.short.term_liabilities...sales_
X58_total_costs_.total_sales
X59_long.term_liabilities_.equity
X60_sales_.inventory
count
value

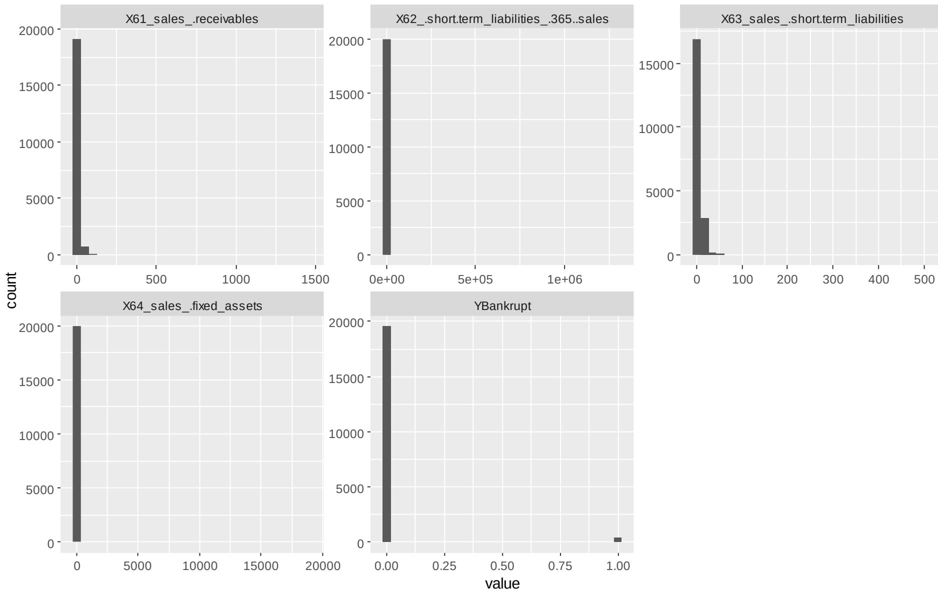
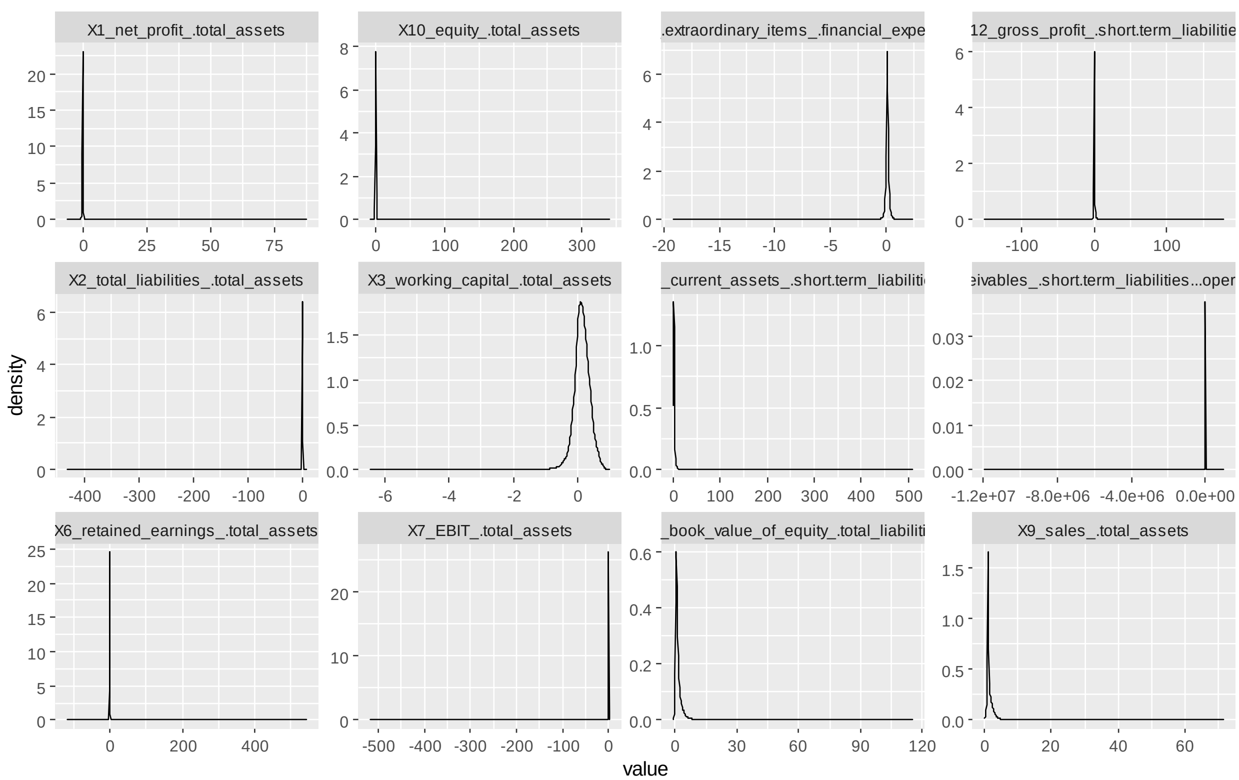
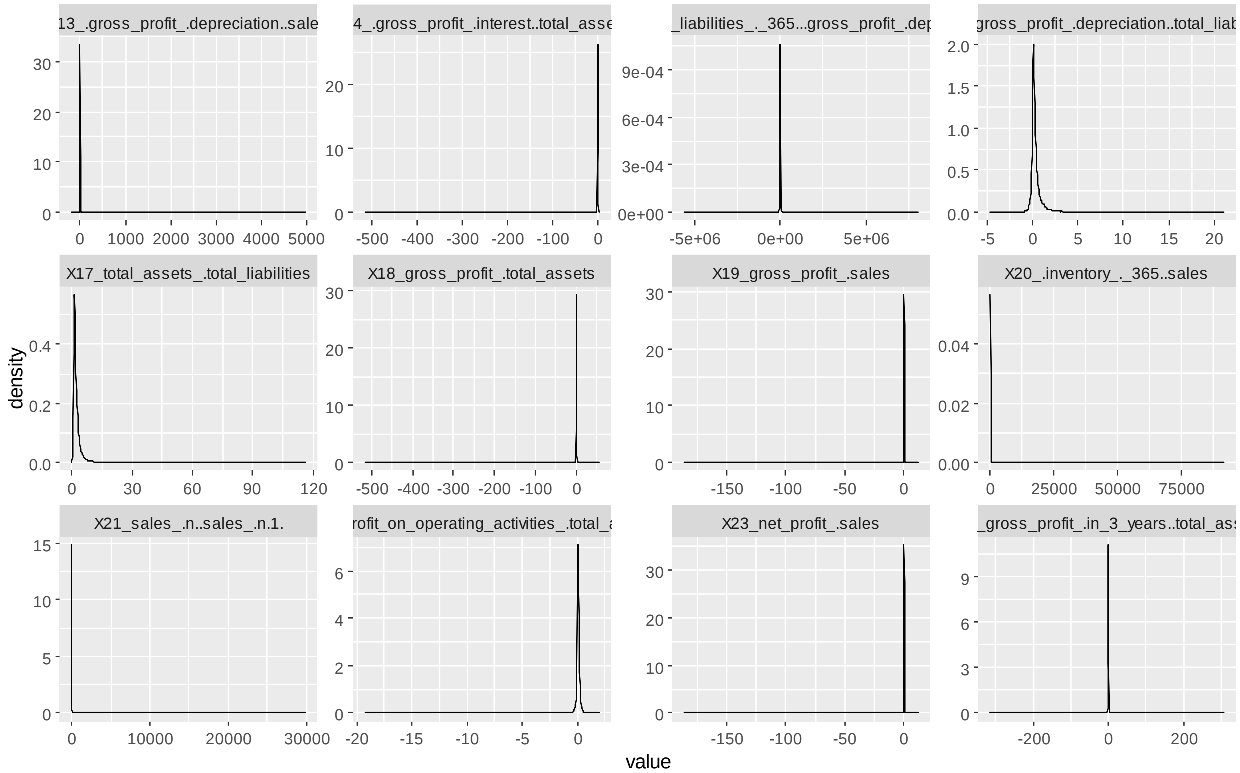
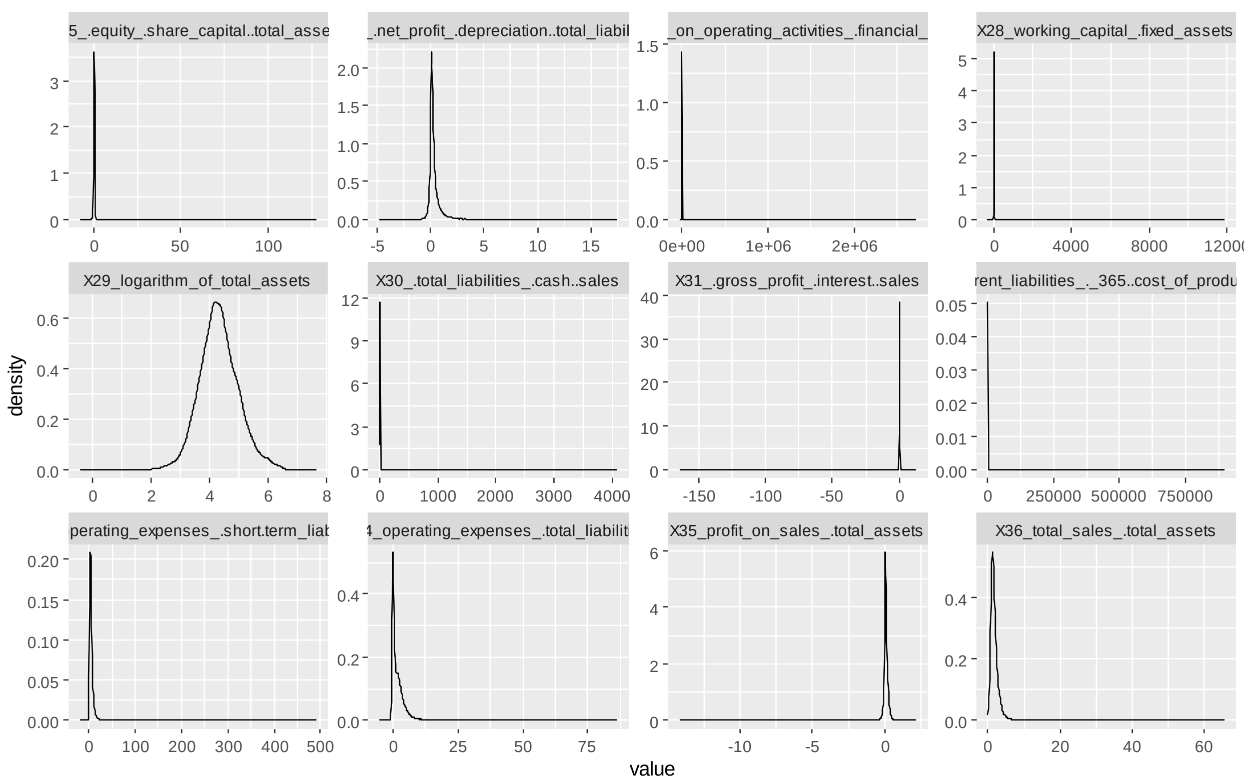


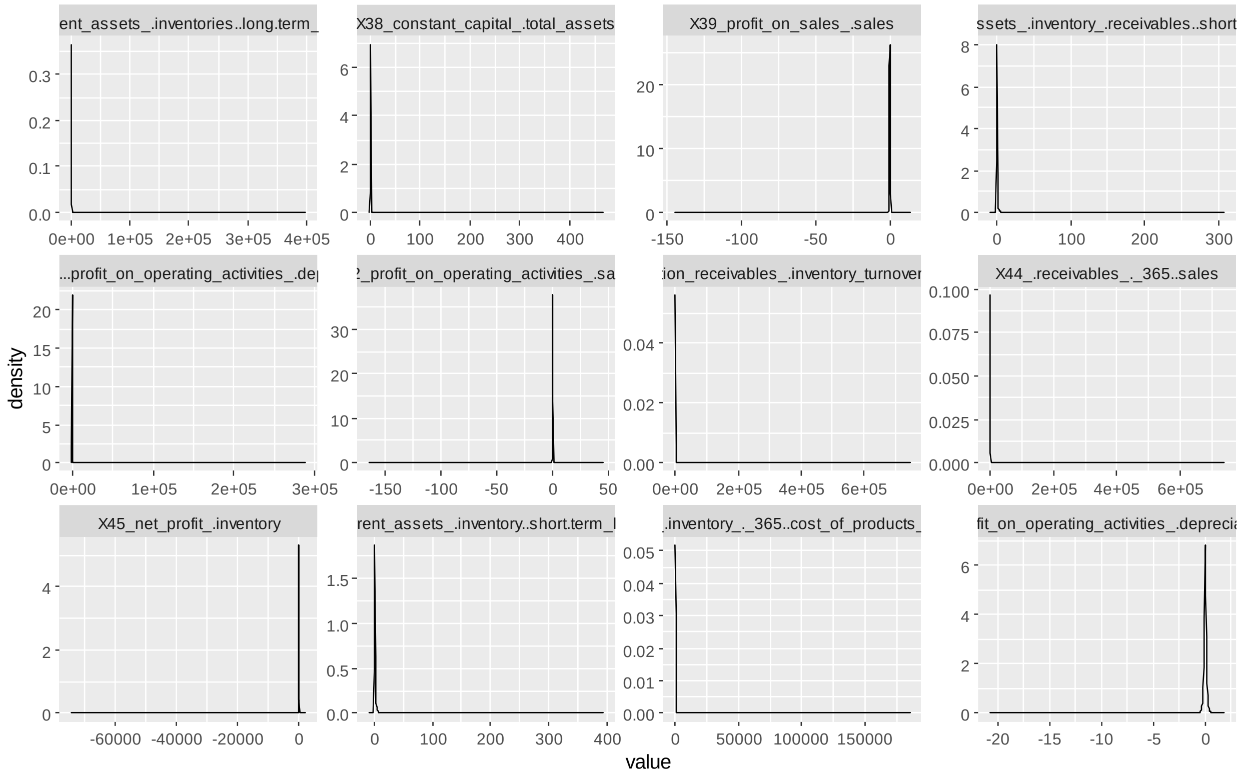
Figure 15

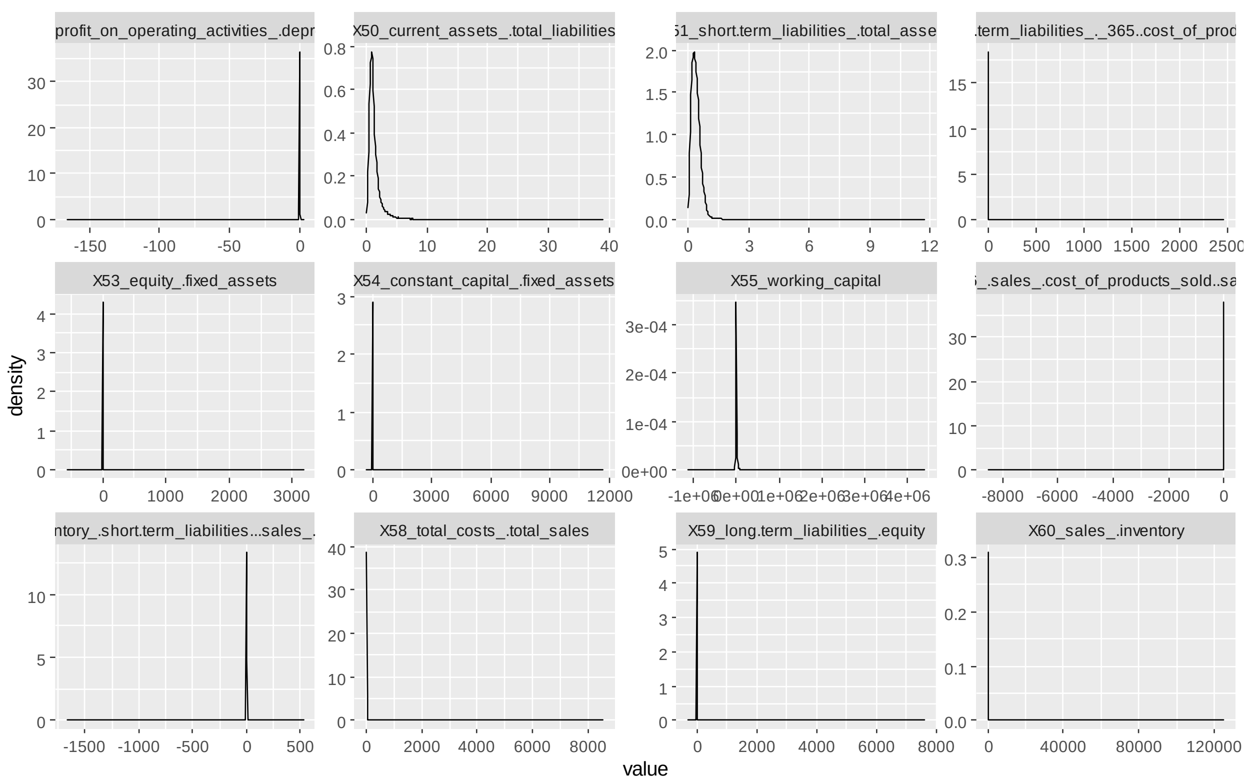
Distribution Analysis for Polish Companies




13_.gross_profit_.depreciation..sale
4_.gross_profit_.interest..total_asse
liabilities._365...gross_profit_.dep
gross_profit_.depreciation..total_liab
X17_total_assets_.total_liabilities
X18_gross_profit_.total_assets
X19_gross_profit_.sales
X20_.inventory_._365..sales
X21_sales_.n..sales_.n.1.
rofit_on_operating_activities_.total_a
X23_net_profit_.sales
_gross_profit_.in_3_years..total_ass
density
value


5_.equity_.share_capital..total_asse
_net_profit_.depreciation..total_liabil
_on_operating_activities_.financial_
X28_working_capital_.fixed_assets
X29_logarithm_of_total_assets
X30_.total_liabilities_.cash..sales
X31_.gross_profit_.interest..sales
rent_liabilities_._365..cost_of_produ
perating_expenses_.short.term_liab
4_operating_expenses_.total_liabiliti
X35_profit_on_sales_.total_assets
X36_total_sales_.total_assets
density
value


ent_assets_.inventories..long.term_
X38_constant_capital_.total_assets
X39_profit_on_sales_.sales
ssets_.inventory_.receivables..short
...profit_on_operating_activities_.dep
2_profit_on_operating_activities_.sa
ion_receivables_.inventory_turnover
X44_.receivables_._365..sales
X45_net_profit_.inventory
rent_assets_.inventory..short.term_l
.inventory_._365..cost_of_products_
fit_on_operating_activities_.deprecia
density
value


profit_on_operating_activities_.depr
X50_current_assets_.total_liabilities
51_short.term_liabilities_.total_asse
.term_liabilities_._365..cost_of_prod
X53_equity_.fixed_assets
X54_constant_capital_.fixed_assets
X55_working_capital
.sales.cost_of_products_sold..sa
ntory_.short.term_liabilities...sales_.
X58_total_costs_.total_sales
X59_long.term_liabilities_.equity
X60_sales_.inventory
density
value

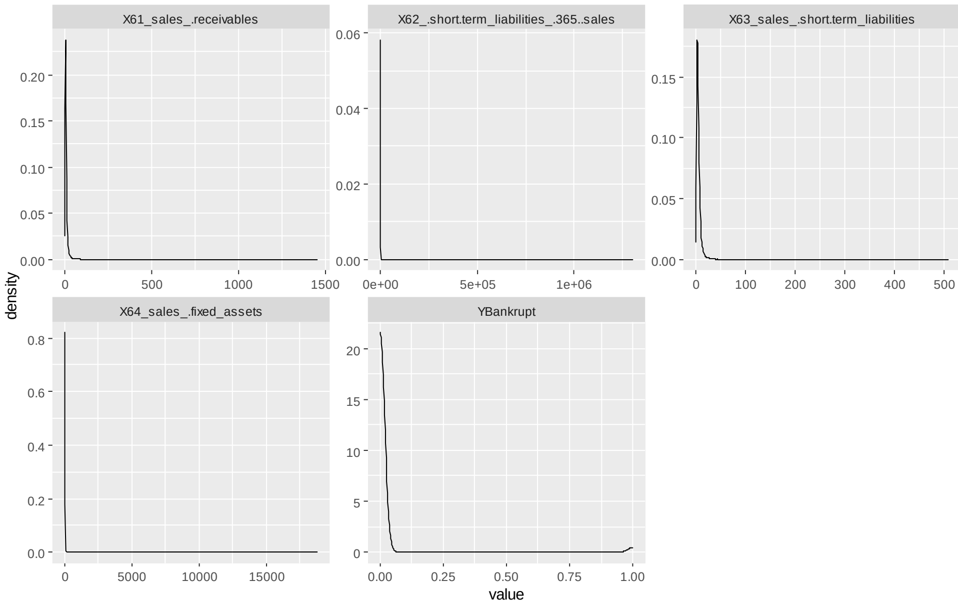

X61_sales_.receivables
X62_.short.term_liabilities_.365..sales
X63_sales_.short.term_liabilities
X64_sales_.fixed_assets
YBankrupt
density
value

Figure 16

Histogram Analysis for U.S. SBA SME

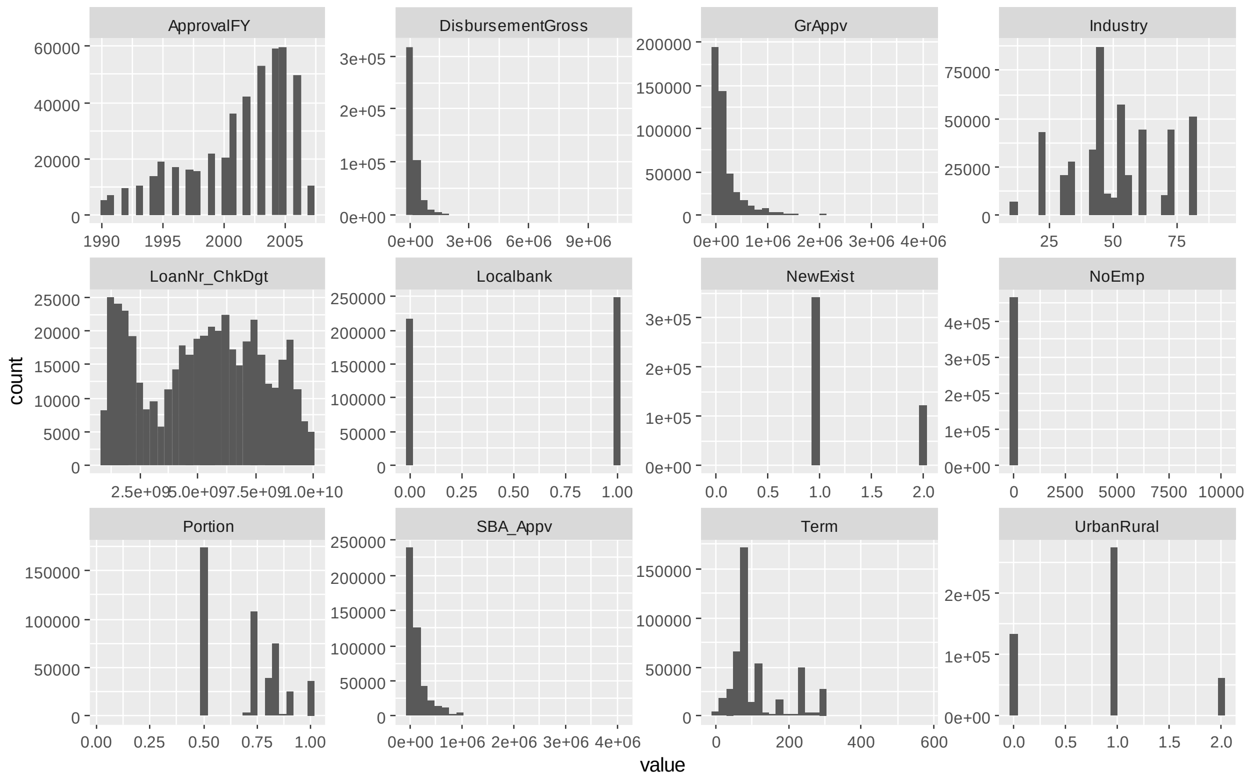
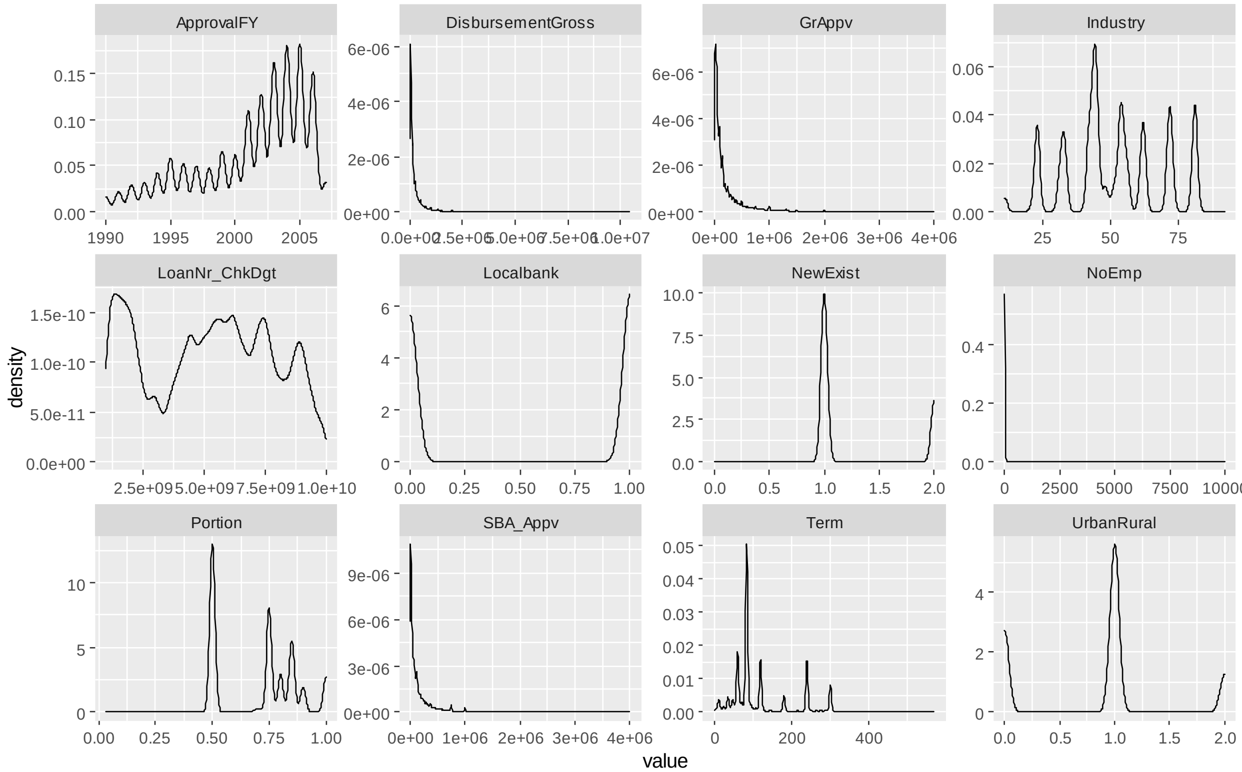


Figure 17

Distribution Analysis for U.S. SBA SME



Appendix (B): Ranking of Key Features Found from Literature Review

Table B1

Table of Key Features Found from Literature Review

Database source	Feature	Source
Transactions - shipment records	Age of SME - adopted from Kou et al., 2021 "Age of company"	Kou et al., 2021
	Trade financing to transaction value ratio - decision variable	Adapted from Bedin et al., 2019
	Size of SME estimated as GMV	Adapted from Gupta et al., 2018
	Survived economic crisis (i)	Adapted from McGuinness et al., 2018
External databases	Outlook of industry	Adapted from Roy & Shaw, 2021
	Profitability of industry	
	GDP growth rate in SME home country	Adapted from Cucinelli et al., 2018
	Unemployment rate at SME home country	
	Engel coefficient	Chi & Zhang, 2017
	Historical bankruptcy rate geography location	Kou et al., 2021
	Historical bankruptcy rate by geography and by business sector	
Management	Education and experience	Roy & Shaw, 2021
	Credit history	
	Registered capital classification	Chi & Zhang, 2017
	Personal tax records	
	Number of managers	Kou et al., 2021
	Number of shareholders	
Financial	EBITDA Margin	Georgios, 2019
	Accounts receivable turnover	Chi & Zhang, 2017
	Cash conversion cycle	
	Revenue growth	
	Retained earnings growth	
	Current ratio	Roy & Shaw, 2021
	Quick ratio	
	Debt-Equity ratio	
	Interest coverage	
	Fixed assets coverage	

Note. Features collected and adapted from multiple papers (Bedin et al., 2019; Chi & Zhang, 2017; Cucinelli et al., 2018; Georgios, 2019; Gupta et al., 2018; Kou et al., 2021; McGuinness et al., 2018; Roy & Shaw, 2021).