**Chapter 2 – Literature Review**

The theoretical foundations of this research draw upon established works on credit-risk measurement and performance assessment.  
Resti and Sironi (2007), Saunders and Allen (2022), and Bessis (2015) provide a conceptual framework linking credit-risk metrics — such as probability of default, loss given default, and risk-adjusted returns — to overall portfolio performance.  
These frameworks justify the selection of KPIs in this thesis (Default Rate, Average Interest Rate, Average Loan Amount, Portfolio Growth, and Loan Distribution by Grade) as the core indicators of loan-portfolio health.  
Complementary perspectives from performance-measurement literature (Franco-Santos et al., 2012; Brynjolfsson et al., 2011; Chen et al., 2012) further explain how data-driven KPI systems enhance organizational decision-making and financial performance, forming the analytical basis for modern FinTech credit monitoring.

**2.1 Introduction to Peer-to-Peer Lending Literature**

Peer-to-peer (P2P) lending has emerged as one of the most significant financial innovations within the broader FinTech ecosystem, transforming the way credit is originated, priced, and distributed.  
Unlike traditional banking, where financial intermediation occurs through regulated institutions, P2P platforms directly connect individual lenders and borrowers through online marketplaces (Patwardhan, 2018).  
This disintermediation process enhances financial inclusion by offering credit to individuals and small businesses that are often underserved by banks, while simultaneously allowing investors to diversify their portfolios and earn risk-adjusted returns (Atz & Bholat, 2016).

The academic literature on P2P lending spans multiple dimensions — from credit-risk determinants and borrower behavior to portfolio performance and systemic implications.  
Early empirical studies primarily focused on U.S. markets, leveraging open datasets from platforms such as Lending Club and Prosper, to examine the financial and behavioral characteristics of loans (Emekter et al., 2015; Serrano-Cinca & Gutiérrez-Nieto, 2015).  
Subsequent research expanded geographically to include China, the world’s largest P2P market prior to its regulatory reforms, and the United Kingdom, where a more structured legal and supervisory framework enabled sustainable growth (Patwardhan, 2018).

Across all markets, scholars have attempted to define key indicators for measuring portfolio health and performance — such as default rate, average interest rate, average loan amount, return on investment (ROI), portfolio growth, and loan distribution by credit grade.  
These KPIs form the analytical foundation of this thesis, serving as standardized measures for comparing loan portfolios and monitoring credit performance over time.

**2.2 Empirical Evidence from the United States, China, and the United Kingdom**

**United States – Lending Club and Prosper**

The majority of empirical P2P lending literature originates from the United States, where Lending Club and Prosper became pioneering platforms after the 2008 financial crisis.  
Using early loan-level data from Lending Club (2007–2012), Emekter et al. (2015) demonstrated that *interest rate, debt-to-income ratio, and credit grade* are significant predictors of loan default.  
Their findings confirmed a positive correlation between interest rate and default probability, implying that higher-risk borrowers pay higher rates but also default more often.  
Similarly, Serrano-Cinca and Gutiérrez-Nieto (2015) analyzed 25,000 Lending Club loans and found that ROI varied considerably across grades, reinforcing the risk–return trade-off inherent in P2P markets.

Later studies employed advanced machine-learning and econometric models to improve predictive accuracy and KPI monitoring.  
For instance, Kim and Cho (2019) applied deep learning models to 850,000 Lending Club loans (2007–2016), identifying loan grade, income, and loan purpose as dominant features for default prediction.  
More recently, Davaadorj et al. (2025) and Sam’an, Deris, and Farikhin (2025) utilized expanded datasets (2008–2020) to explore *feature selection* and *employment disclosure effects*, emphasizing how non-financial borrower characteristics (e.g., job title or employment history) can significantly affect the Default Rate KPI.

Overall, U.S.-based research established a robust analytical framework for monitoring loan portfolios using standardized KPIs.  
Default rates in Lending Club data typically range between 12–17%, with higher-risk grades (E–G) driving both increased ROI and volatility (Kim & Cho, 2019; Davaadorj et al., 2025).  
These studies collectively form the methodological backbone of this thesis’s KPI system, as they define the metrics and modeling approaches most commonly adopted in P2P risk analysis.

**China – High Growth and Credit Risk in P2P Lending**

China’s P2P lending industry grew exponentially during the 2010s, accounting for over 75% of global P2P loan volume before the regulatory crackdown of 2018 (Patwardhan, 2018).  
Academic studies in this context often focus on macro-level credit-risk patterns, platform governance, and systemic vulnerabilities.  
Lin et al. (2016) and Yin et al. (2023) both analyzed Chinese P2P datasets, revealing that interest rate and borrower credibility are primary drivers of default rates.  
In Yin et al. (2023), using RenRenDai data (126,000 loans), the average interest rate reached 14%, with default rates between 20–30%, highlighting the trade-off between high profitability and fragile borrower quality.

Rao et al. (2020) proposed a comprehensive KPI system for evaluating Chinese P2P platforms, introducing 21 quantitative indicators grouped into categories such as *loan quality, credit concentration, and operational sustainability*.  
Their framework formalized the measurement of credit risk through multidimensional performance indices rather than single metrics.  
Tian et al. (2023) later extended this by integrating *machine-learning–based risk scoring* with macroeconomic indicators to capture systemic exposure across multiple platforms.

Compared to the United States, Chinese studies emphasize *scale and system-level dynamics* rather than individual loan-level behavioral variables.  
The high default rates and regulatory failures in China’s market demonstrate the importance of standardized KPI monitoring and transparent data pipelines — a gap this thesis addresses through its structured, reproducible framework.

**United Kingdom – Stability and KPI Comparability**

The United Kingdom represents the third major ecosystem for P2P lending, characterized by regulated platforms such as Zopa and Funding Circle.  
Atz and Bholat (2016) conducted one of the earliest comprehensive analyses, using anonymized data from Zopa, RateSetter, and Funding Circle covering nearly 14 million loan agreements (2010–2013).  
Their study revealed a pronounced *North–South divide* in credit allocation: London and the South East acted as net lenders, whereas Northern regions were net borrowers.  
Average interest rates ranged from 6.1% to 6.7%, and the median loan amount was approximately £4,000, illustrating moderate yields compared with U.S. platforms.  
While default data were not yet available, their results documented rapid portfolio growth — from £0 to £500 million in just three years — and established early benchmarks for the Interest Rate, Loan Amount, and Portfolio Growth KPIs.

Building upon this foundation, Xu, Su, and Celler (2021) conducted a detailed risk and ROI analysis using Funding Circle data (16,476 loans, 2010–2018).  
Their findings revealed that *interest rate, loan term, and purpose (capital expansion)* are positively correlated with default probability, whereas *loan amount* and *firm age* reduce risk.  
Average loan size was £73,344, with mean interest rates around 9.37% and default rates reaching 15–20%, depending on sector.  
Importantly, they also calculated Return on Investment (ROI), confirming that riskier loans yield higher returns — aligning with findings from U.S. studies (Emekter et al., 2015).

Together, U.K. research complements U.S. and Chinese evidence by illustrating that platform maturity, regulation, and borrower type significantly influence KPI patterns.  
The Funding Circle data — though focusing on SMEs rather than consumer loans — still operates under a P2P model, providing a valuable comparative reference for KPI-based portfolio monitoring across regions.

**2.3 Behavioral and Theoretical Perspectives on Credit Risk**

Beyond financial indicators, P2P lending research increasingly examines *behavioral, informational, and moral-hazard dimensions* of borrower and investor behavior.  
The decentralized nature of P2P markets introduces asymmetries of information and incentive misalignments that differ from those in conventional banking.

Alsabah and Alibrahim (2024) explored *moral hazard* in online lending environments, arguing that the lack of institutional oversight amplifies the risk of borrower opportunism, particularly when screening mechanisms are weak.  
They found that borrowers may strategically misreport income or employment information when risk-based pricing mechanisms rely heavily on self-declared data — directly linking to studies such as Davaadorj et al. (2025) on employment disclosure effects.

Similarly, Mesly and Ivanaj (2024) analyzed *deceptive behaviors and psychological biases* among borrowers and investors, emphasizing that platform design (e.g., reputation systems, transparency levels) can mitigate but not eliminate moral hazard.  
Their findings reinforce the importance of incorporating *non-financial KPIs* — such as borrower disclosure completeness and verification ratio — into credit-risk frameworks.

These behavioral insights complement the quantitative literature by clarifying why certain KPIs (e.g., default rate, ROI) may deviate from expected risk-return patterns.  
The integration of financial and behavioral perspectives thus supports a holistic approach to loan-portfolio monitoring, in which both numerical metrics and informational asymmetries shape credit outcomes.

**2.4 Technical and Methodological Framework for KPI-Based Loan Monitoring**

While the preceding sections examined the financial and behavioral dimensions of peer-to-peer (P2P) lending, implementing a data-driven loan-portfolio monitoring system also requires a strong technical foundation.  
End-to-end analytical pipelines depend on robust data-engineering frameworks that ensure the reliability, traceability, and scalability of financial KPIs.

These regulatory and methodological foundations are consistent with the risk-management principles articulated by Resti and Sironi (2007) and Bessis (2015), which emphasize data accuracy, aggregation consistency, and alignment between risk measurement and capital performance.  
This section reviews the principal technical and methodological sources that support the thesis’s architecture — encompassing database modeling, data quality, ETL processes, regulatory alignment, and API exposure.

**Data Modeling and Database Design**

The relational database implemented for this thesis follows the dimensional modeling principles established by Kimball and Ross (2013) in *The Data Warehouse Toolkit*.  
The *star schema* design separates fact tables (containing KPI values such as Default Rate, Loan Amount, and Interest Rate) from dimension tables (containing borrower, grade, and time-related attributes).  
This approach allows for efficient aggregation, filtering, and temporal comparisons — essential for portfolio tracking over time.  
In addition, the data-cleaning and transformation logic adheres to the data-quality framework defined by Batini and Scannapieco (2016), who classify information quality through dimensions such as *accuracy, completeness, consistency,* and *timeliness*.  
These principles ensure that KPI computations rely on validated and harmonized loan records, minimizing analytical bias.

**ETL and Data Transformation Processes**

Modern ETL pipelines are grounded in the design philosophy introduced by Zaharia et al. (2016) in the *Apache Spark* framework, which unified batch and streaming data processing under a single computational model.  
Although the implementation in this thesis uses Python and PostgreSQL rather than distributed clusters, the same concepts — reproducible dataflows, in-memory transformations, and modular stages for extraction, transformation, and loading — are applied.  
Complementing this, Vandermarliere et al. (2022) illustrate how financial institutions can design scalable pipelines with automated validation, metadata management, and audit trails, ensuring transparency and reproducibility in KPI reporting.  
The thesis adapts these principles to create a maintainable and replicable workflow for computing credit-risk metrics such as Default Rate, Average Interest Rate, and Portfolio Growth.

**Risk Data Aggregation and Regulatory Compliance**

Beyond technical efficiency, financial data systems must comply with supervisory standards.  
The Basel Committee on Banking Supervision (2019), through *BCBS 239: Principles for Effective Risk Data Aggregation and Reporting*, defines requirements for accuracy, completeness, and consistency in risk data.  
These standards provide the conceptual rationale for the structured storage and aggregation of loan-level data.  
Moreover, the International Association of Credit Portfolio Managers (IACPM, 2023) formalizes KPI definitions — such as *Default Rate, Loan Distribution by Grade, and Portfolio Growth* — and advocates for standardized data governance frameworks across financial institutions.  
Together, BCBS 239 and IACPM serve as the methodological link between credit-risk metrics and data-engineering implementation.

**API Layer and Data Exposure**

The final layer of the thesis pipeline concerns data exposure through an Application Programming Interface (API).  
Following Tiangolo’s (2023) *FastAPI* framework, KPI outputs are published via RESTful endpoints, allowing external systems or dashboards to query indicators such as /api/kpis/default\_rate or /api/kpis/portfolio\_growth.  
FastAPI’s asynchronous architecture and automatic schema validation through *Pydantic* enable both scalability and integrity of the responses, aligning with open-data and transparency principles promoted in financial analytics.  
This design ensures that computed KPIs are not static outputs but dynamic, reusable services — directly supporting real-time portfolio monitoring.

**Integration of Technical and Financial Perspectives**

Integrating these technical frameworks with the financial literature bridges the gap between theoretical research and practical implementation.  
Dimensional modeling ensures analytical efficiency, data-quality controls guarantee validity, ETL architecture delivers scalability, regulatory standards impose consistency, and API exposure promotes transparency.  
Together, these layers produce a unified system for calculating, validating, and sharing KPIs — transforming credit-risk metrics from static analytical results into dynamic decision-support tools.

**2.5 Summary and Research Gap**

The literature reviewed above highlights a dynamic and multidisciplinary body of research on P2P lending and loan-portfolio analytics.  
From the U.S. empirical studies (Emekter et al., 2015; Kim & Cho, 2019) to Chinese macro-analyses (Rao et al., 2020; Yin et al., 2023) and U.K. evidence (Atz & Bholat, 2016; Xu et al., 2021), scholars have established strong associations between KPIs such as default rate, interest rate, loan term, and ROI.  
The findings consistently confirm the *risk–return trade-off* across markets and loan segments, while also underscoring the influence of regulatory and behavioral factors.

However, several gaps remain in the existing literature:

1. Most studies focus on *modeling or prediction* of credit risk, not on *monitoring frameworks* that continuously compute and expose KPIs.
2. Few works integrate *data-engineering methods* — such as ETL pipelines, schema design, and API services — into academic analyses of financial risk.
3. The majority of empirical studies are *platform-specific* (e.g., Lending Club, Funding Circle) and lack reproducible architectures that can generalize to different datasets or financial institutions.

This thesis addresses these limitations by developing an end-to-end data-engineering pipeline for KPI calculation and portfolio monitoring using the Lending Club dataset.  
It operationalizes the KPIs defined in the literature (Default Rate, Average Loan Amount, Average Interest Rate, Portfolio Growth, and Loan Distribution by Grade) within a transparent, reproducible, and API-accessible architecture. By combining insights from financial research and data-engineering frameworks, the study contributes to both domains — offering a scalable methodology for real-time loan-portfolio performance analysis and bridging the existing divide between academic theory and applied financial analytics.