### CONCLUSION

The data preparation process is a fundamental step in the Lending Club loan analysis, ensuring that the dataset is clean, consistent, and ready for predictive modeling. By systematically handling missing values through imputation, we addressed potential data gaps that could bias the analysis. Cleaning the data to remove duplicates and outliers further enhanced the dataset's reliability. Data transformation through normalization and encoding ensured that numerical and categorical variables were appropriately scaled and formatted for analysis.

To address key risk factors like economic downturns and borrower instability, we should continuously update the models and incorporating new data sources such as social media activity and payment behavior. These thorough data preparation steps provide a solid foundation for further analysis and model development, ultimately improving Lending Club's risk assessment and decision-making processes.

### **8.2 No Of Visualizations/ Graphs**

1. Total Number of Accounts
2. Total Loan Amount
3. Average Loan Amount
4. Average Loan Amount for Account type
5. Average Loan Interest rate based on Grade
6. State wise Average Loan Amount
7. The number of Accounts (Individual/Joint)
8. The number of members – Grade wise

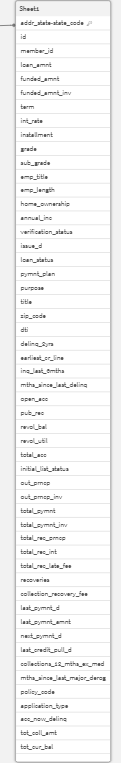
9.The number of members – Verification Status

10.Funded Amount vs sum of loan amount

11. Grade vs annual Income

8. Performance testing

8.1 Amount of Data used for the visualisation Insights.



7.2 Story 2

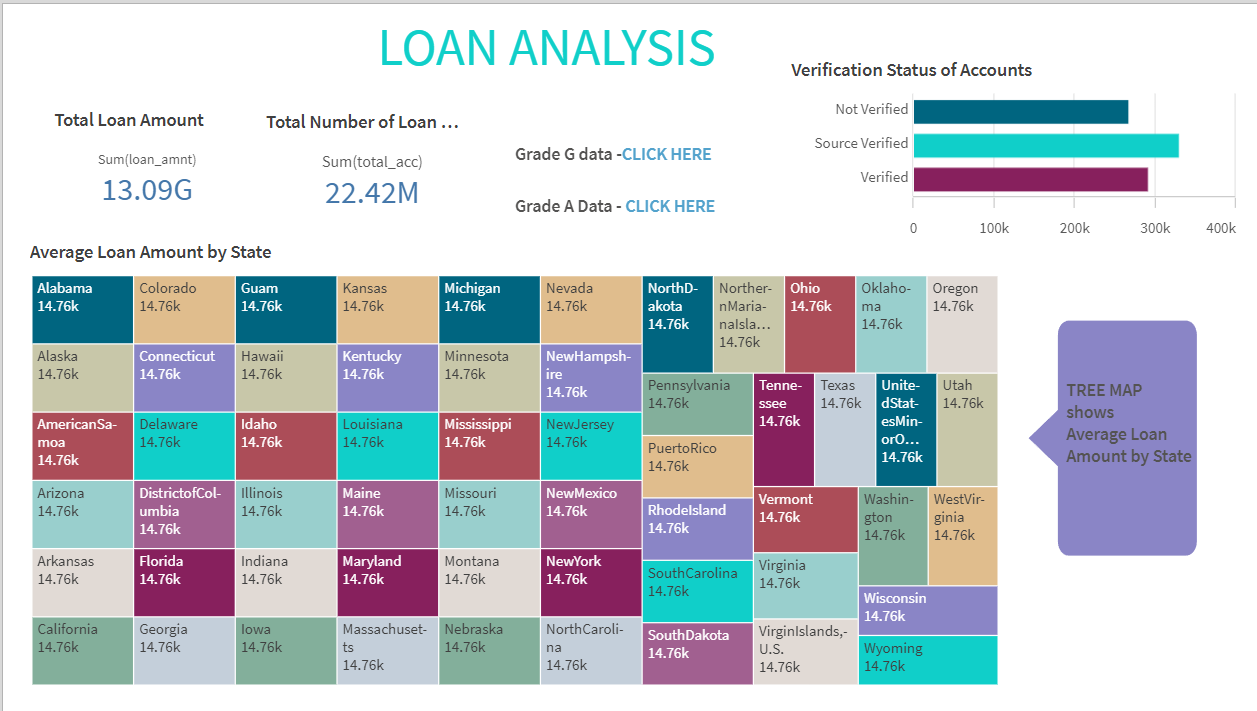


Fig : Loan Analysis Story 2 on Lending Club

1. DESIGN OF STORY

7.1 . Story 1

We can do the story for easy understanding of data.Here are some insights of story creation on Lending Club Loan Analysis.

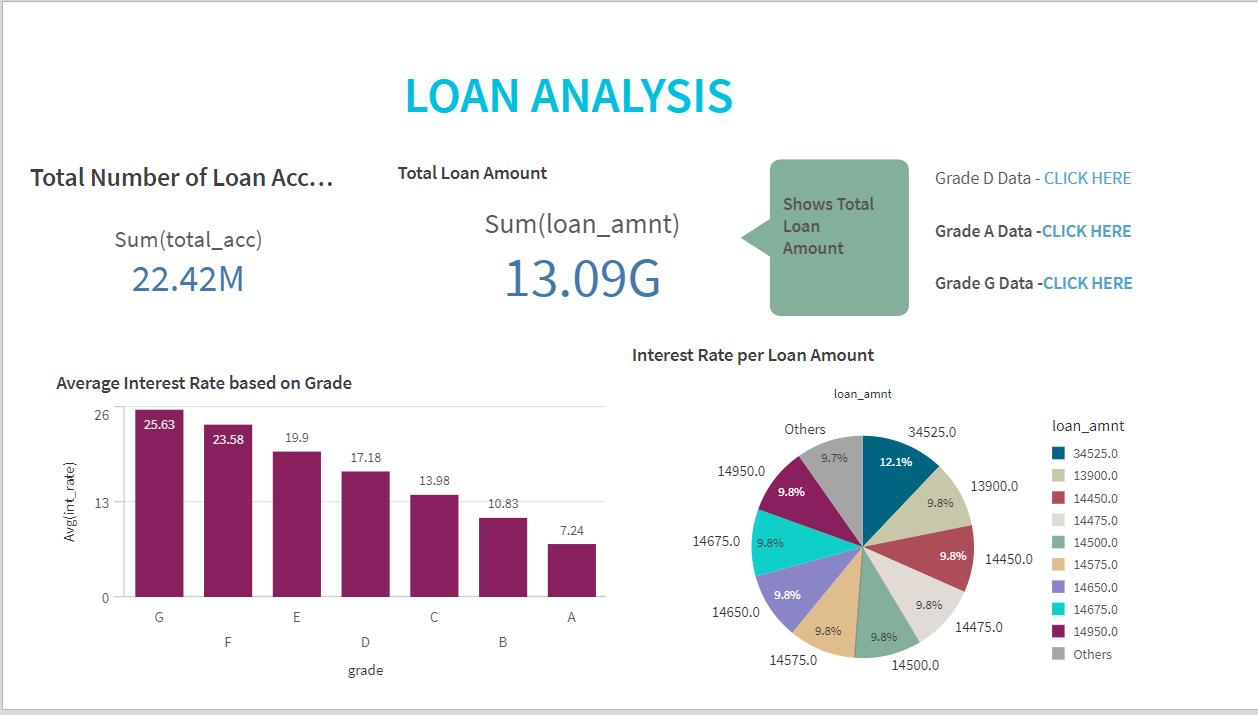


Fig : Loan Analysis Story 1 on Lending Club

1. DASHBOARD

Here on the qliksense we can create different type of visualisation based on different parameters.

We can also use the responsive dashboard, where we can find the specfic data we need using responsive dashboards. I have created the responsive dashboard based on the grade, where we can get the data on different grade analysis.

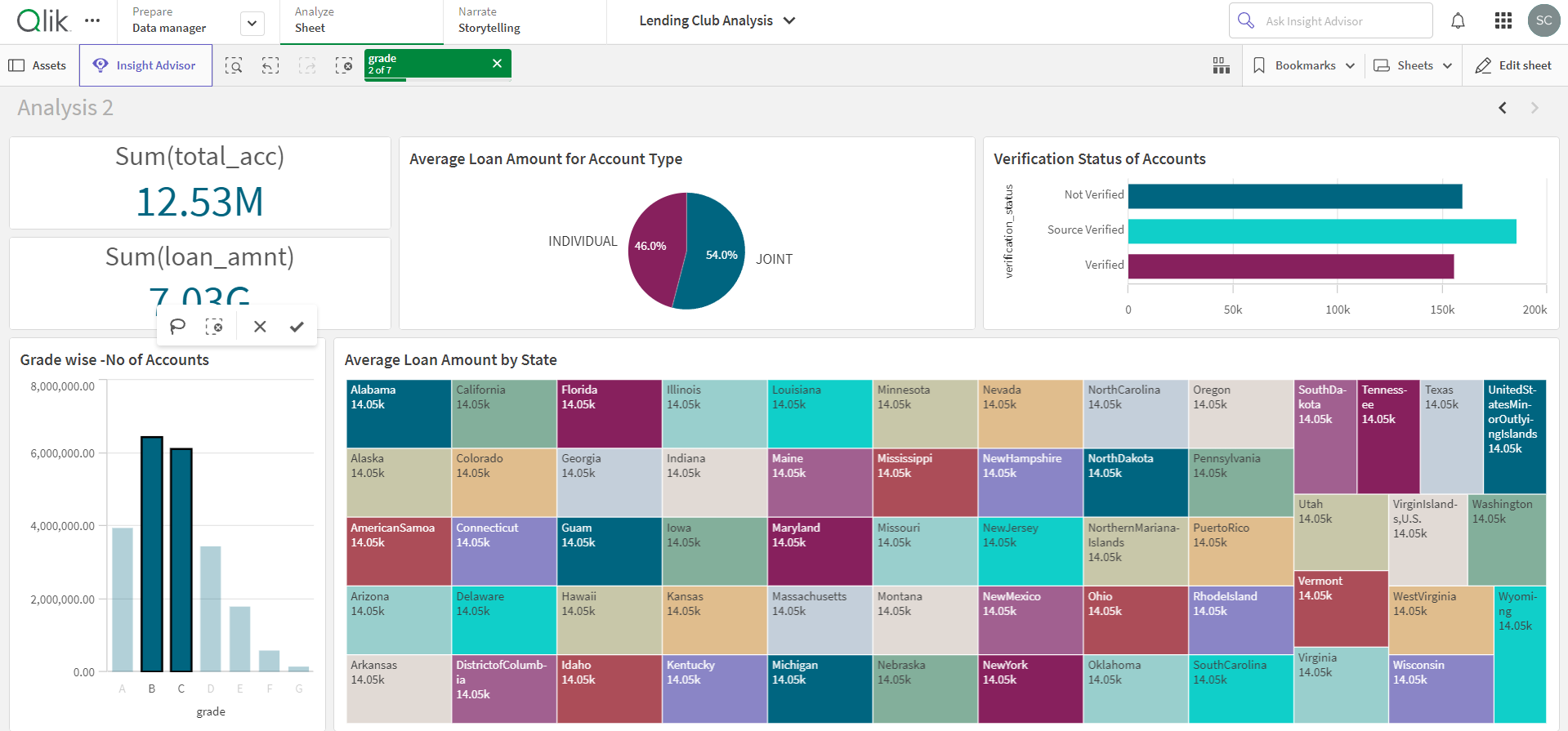


Fig: Dash Board of Analysis of Grade B & C data.

The above Dashboard shows only the data of Grade B & Grade C data.

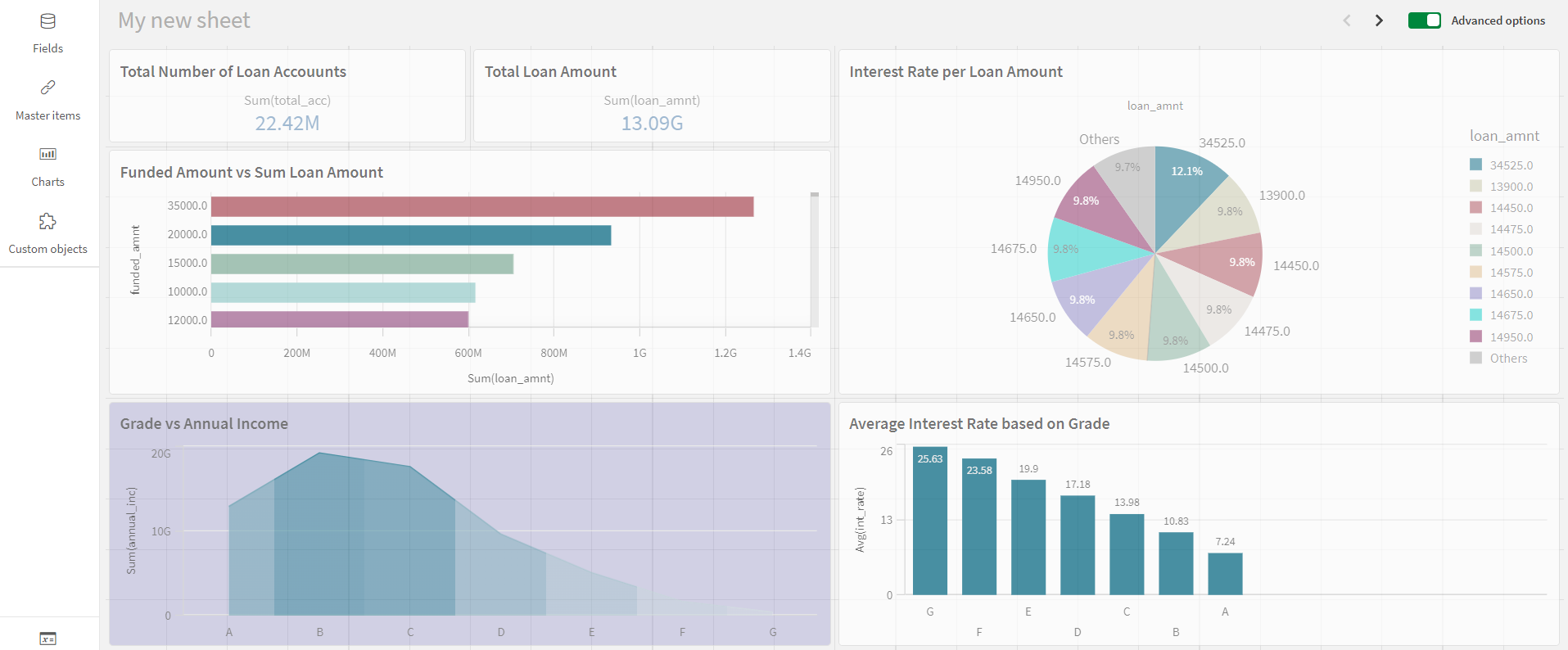


Fig: Data visualisation of 2nd dashboard

This visualisation shows the insights of:

1. Total Loan Accounts

2. Total Loan Amount

3. Grade Wise Annual Income

4.Interest rate per Loan Amount

1. Funded amount vs Sum of Loan Amount
2. Average Interest Rate based on Grade

5. DATA VISUALISATION

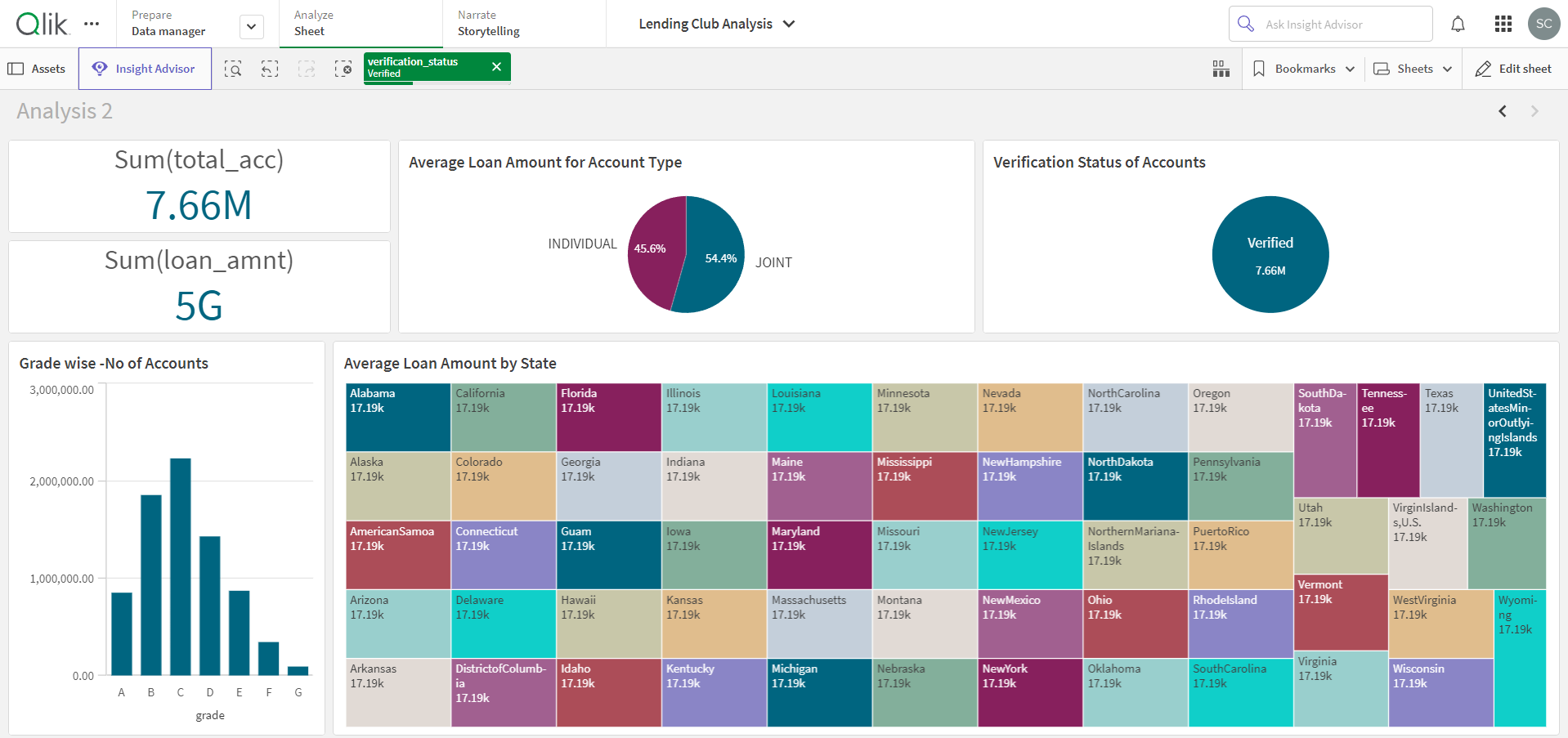


Fig : Data Visualisation of 1st Dashboard

This visualisation shows the insights of:

1. Total Accounts
2. Sum Loan Accounts
3. Grade wise - No of Accounts
4. Average Loan amount for Account Type
5. Verification Status
6. Average Loan Amount by State

#### 3. Data Cleaning

Data cleaning involved several steps to ensure the data quality and consistency:

Removing Duplicates: Duplicate records were identified and removed to prevent bias in the analysis.

Handling Outliers**:** Outliers can distort the analysis and model performance. The IQR (Interquartile Range) method was used to identify and handle outliers.

Standardizing Categorical Values**:** Categorical values were standardized to ensure consistency, for instance, converting '10+ years' and '10 years' into a single category.

#### 4. Data Transformation

Data transformation involved normalizing numerical features and encoding categorical variables:

Normalization: Numerical features were normalized to have a mean of 0 and a standard deviation of 1. This step is essential for algorithms that are sensitive to the scale of data.

5. Visualisation

Doing the visualisation based on the values we take for the for the desired outcome.

4. DATA PREPERATION

4.1 Prepare the data for visualisation

#### Data preparation is a crucial step in the data analysis process. It involves cleaning and transforming raw data into a format suitable for analysis. This section outlines the steps taken to prepare the Lending Club loan data for modeling, including handling missing values, data cleaning, data transformation, and feature encoding.

#### 1. Data Collection

The dataset used in this analysis contains detailed information about loans issued by Lending Club from 2007 to 2018. The dataset includes the following attributes:

* Loan details (e.g., loan amount, interest rate, term)
* Borrower information (e.g., annual income, employment length, home ownership)
* Loan status (e.g., fully paid, charged off)

#### 2. Handling Missing Values

Missing data is a common issue in real-world datasets. Several strategies were employed to handle missing values:

Imputation for Numeric Variables**:** Missing values in numeric variables were imputed using the median, as it is less sensitive to outliers than the mean.

Missing values in categorical variables were imputed using the most frequent category

3.2 CONNECT DATA WITH LINK

Using the dataset I have uploaded the dataset into the qlik sense software. Then steps:

1. Open Qlik Sense.
2. Create a new app.
3. Name the app Scripting Tutorial, and then click Create.
4. Open the app.
5. Open Data load editor from the drop-down menu in the top toolbar.
6. Click.

3. DATA COLLECTION

3.1 Collect the dataset

I have collected the dataset from the Kaggle. Where we can find the data for the data visulaisations. It had different aspects for the analysis.

Column Description of the Dataset:

* loan\_amnt: The total amount of the loan.
* funded\_amnt: The amount funded by investors.
* term: The term of the loan (36 or 60 months).
* int\_rate: Interest rate on the loan.
* grade: Lending Club assigned loan grade

LINKS:

Dataset Links- 1. <https://drive.google.com/file/d/18VOlqd1DjouviSW6W_qUNnHS6qjwOBJl/view?usp=sharing>

Dataset Linkk 2. <https://drive.google.com/file/d/1I8nkFhTkAlOpbFU6QDEWOStodu7ayM1U/view?usp=sharing>

Dataset Link 3. <https://drive.google.com/file/d/1SLI0JUqtuS2X6TvyMAyyKNfVffjVzrKg/view?usp=sharing>

#### Default Prediction Models

Machine learning models have been extensively applied to predict loan defaults in P2P lending. Random forests, gradient boosting, and neural networks have shown promising results. Research by Malekipirbazari and Aksakalli (2015) demonstrated that ensemble methods like random forests significantly outperform traditional logistic regression in predicting defaults. Their study emphasized the role of features like loan amount, interest rate, and debt-to-income ratio in determining loan outcomes.

CHALLENGES AND FUTURE DIRECTIONS

Despite these advancements, challenges remain in the form of data quality, model interpretability, and regulatory compliance. The dynamic nature of economic conditions necessitates continuous model updates and validations. Future research is expected to focus on integrating alternative data sources, such as social media activity and mobile phone usage, to enhance prediction accuracy.

In conclusion, the literature underscores the importance of robust risk assessment models in P2P lending. The integration of advanced machine learning techniques has significantly improved the accuracy of default predictions, helping platforms like Lending Club manage credit risk more effectively. This study aims to build on these foundations, leveraging historical loan data to develop predictive models that enhance Lending Club’s risk assessment capabilities.

2.3 LITERATURE REVIEW:

Peer-to-peer (P2P) lending, exemplified by platforms like Lending Club, has revolutionized the financial sector by connecting borrowers directly with investors, bypassing traditional banking intermediaries. This model offers benefits such as higher returns for investors and more accessible credit for borrowers. However, it also introduces challenges,

particularly in assessing and managing credit risk. A substantial body of research has explored various aspects of P2P lending, including risk assessment, default prediction, and the impact of borrower characteristics on loan performance.

#### Risk Assessment in P2P Lending

Credit risk assessment in P2P lending is critical, given the absence of traditional banking safeguards. Numerous studies have employed statistical and machine learning techniques to enhance risk prediction. For instance, Serrano-Cinca, Gutierrez-Nieto, and Lopez-Palacios (2015) analyzed default risk using logistic regression, decision trees, and support vector machines. They highlighted the importance of borrower characteristics such as credit history and income levels in predicting defaults.

#### Borrower Characteristics and Loan Performance

The influence of borrower characteristics on loan performance has been a focal point of several studies. Research indicates that factors such as employment length, annual income, and home ownership status significantly affect the likelihood of loan repayment. Emekter, Tu, Jirasakuldech, and Lu (2015) found that loans to borrowers with higher credit grades and verified incomes have a lower default rate. Additionally, Zhang, Jia, and Diao (2016) explored the impact of social and economic factors on loan performance, revealing that economic downturns increase default rates among lower-income borrowers.

2. PROBLEM STATEMENT

2.1 Business Problem :

Lending Club faces the challenge of minimizing loan defaults while maximizing returns for investors, a critical balance to sustain its peer-to-peer lending model. The primary business problem is accurately identifying high-risk loans before they default, which impacts financial stability and investor confidence. High default rates can deter investors, leading to reduced capital availability and increased costs for Lending Club.

This analysis aims to understand the key factors influencing loan defaults and develop predictive models that accurately classify loans as likely to default or be fully paid. By leveraging historical loan data, including borrower details and loan characteristics, the goal is to enhance Lending Club's risk assessment capabilities. Improved prediction models will enable better decision-making, reduce the incidence of default, and optimize the loan approval process, ultimately leading to increased investor trust and a more robust lending platform. This analysis is crucial for maintaining a competitive edge in the rapidly evolving financial technology sector.

2.2 Business Requirements:

Create an efficient and user-friendly platform for users to interact with data or perform specific tasks.

* User authentication and authorization.
* Intuitive user interface (UI) design.
* Data visualization capabilities (charts, graphs, dashboards).
* Task-specific modules (e.g., loan approval, risk assessment).
* Integration with other systems

1.3. Technical Architecture

The Technical Architecture for a Lending Club Loan Analysis is as follows :

1.Data Collection and Preprocessing :

* Gather historical loan data from Lending Club’s platform.
* Cleanse and pre-process the data by handling missing values, encoding categorical variables, and scaling features

2.Exploratory Data Analysis (EDA):

* Visualize distributions, correlations, and key statistics.
* Understand the data’s characteristics and identify patterns.

3.Feature Engineering:

* Select relevant features.
* Reduce conditionality if needed.

4.Modeling:

* Implement modals using QlikSense
* Train the model on historical data to predict loan defaults.

5.Evaluation and Interpretation:

* Assess model performance using metrics.
* Interpret feature importance and identify key predictors of repayment behavior.

6.Deployment:

* Deploy the trained model in a production environment

Monitor model performance and retrain periodically.

1.INTRODUCTION

1.1 Overview

Lending Club is the world’s largest peer-to-peer lending company acting as a loan originator and a web platform between borrowers and investors.They claim to transform the banking system by operating at a lower cost than a traditional bank and thereby making credit more affordable and investing more rewarding. Over the last 8 years, the number of loans in the marketplace has increased exponentially, yet little is known about the algorithms that determine if a loan is approved, and if it is, the interest rate a loan is offered at. In this paper made an attempt to demystify the inner workings of this marketplace by applying Qlik Analytics techniques to Lending Club’s publicly available dataset.

1.2 Purpose

This project aims to conduct a comprehensive analysis of peer-to-peer lending data from Lending Club to gain insights into loan performances, borrower behavior, and risk factors.

A Report on

LENDING CLUB LOAN ANALYSIS

**Submitted by:**

CH SHYAM

**Under the Esteemed Guidance of**

Mr.Revanth Reddy (Mentor)