Data To Decisions Qlik Journey Through LendingClub Issued Loans Analysis (Qlik)

Project Report

1. Introduction

1.1 Overview

In the current data-driven era, making informed decisions based on vast amounts of data is crucial for business success. This project, "Data to Decisions: Qlik Journey Through LendingClub Issued Loans Analysis," focuses on leveraging Qlik's powerful data visualization and analysis capabilities to extract meaningful insights from the dataset of loans issued by LendingClub. LendingClub is a well-known peer-to-peer lending platform that connects borrowers with investors, providing a wide array of loan products. By analyzing the loan data, we aim to uncover trends, identify risk factors, and optimize lending strategies.

1.2 Purpose

The primary purpose of this project is to demonstrate the utility of Qlik's data analytics platform in transforming raw data into actionable insights. Specifically, this project aims to:

Identify Trends: Analyze trends in loan issuance, such as popular loan categories, borrower demographics, and seasonal patterns.

Risk Assessment: Evaluate risk factors associated with different types of loans and borrower profiles to enhance risk management strategies.

Performance Optimization: Provide insights to improve loan performance by identifying factors that contribute to successful repayment and those that lead to defaults.

Decision Support: Offer a robust decision support system for stakeholders, including investors, financial analysts, and LendingClub itself, to make informed decisions based on data-driven insights.

1.3 Technical Architecture

The technical architecture of this project encompasses the **data extraction, transformation, loading (ETL) process,** and the utilization of **Qlik for data visualization and analysis.** Here's a detailed breakdown of the architecture:

Data Sources:

LendingClub Loan Dataset: This dataset includes detailed information about loans issued by LendingClub, including loan amounts, interest rates, borrower information, repayment status, and more.

Supplementary Data: Additional datasets may include economic indicators, credit score distributions, and demographic information to enrich the analysis.

ETL Process: Extraction: Data is extracted from the primary sources, ensuring that the most recent and relevant data is used for analysis.

Transformation: The extracted data is cleaned, normalized, and transformed to ensure consistency. This includes handling missing values, encoding categorical variables, and aggregating data where necessary.

Loading: The transformed data is loaded into Qlik's data repository for further analysis.

Qlik Environment

Qlik Sense/QlikView: The primary tools used for data visualization and analysis. These tools provide an interactive interface for exploring data, creating dashboards, and generating reports.

Data Model: A well-structured data model is created within Qlik to facilitate efficient querying and visualization. This model integrates various data sources and ensures optimal performance.

Dashboards and Reports: Customized dashboards and reports are developed to present the insights in an intuitive and accessible manner. These visualizations help stakeholders to quickly grasp the key findings and make informed decisions.

Security and Compliance

Data Security: Ensuring the security of sensitive data through encryption, access controls, and compliance with relevant data protection regulations.

Compliance: Adhering to industry standards and regulations, including GDPR and other financial regulations, to ensure the ethical use of data.

Deployment and Maintenance

Deployment: The final solution is deployed within the organization's IT infrastructure or on a cloud-based platform, depending on the requirements.

Maintenance: Regular updates and maintenance are performed to ensure the system remains up-to-date with the latest data and continues to perform optimally.

2. Define Problem/Problem Understanding

2.1 Specify the Business Problem

LendingClub, as a peer-to-peer lending platform, facilitates a significant volume of loans between borrowers and investors. Despite its success, the company faces several critical challenges that need to be addressed to optimize its operations and enhance its value proposition. These challenges include:

Risk Management: Identifying and mitigating the risks associated with loan defaults. Unforeseen defaults can significantly impact investors' returns and the platform's reputation.

Loan Performance Optimization: Understanding the factors that contribute to successful loan repayments and using this knowledge to improve the overall performance of issued loans.

Investor Confidence: Providing investors with reliable data-driven insights to boost their confidence in the platform and encourage more investments.

Competitive Edge: Staying ahead in a competitive market by leveraging advanced data analytics to make more informed lending decisions and offer better services to borrowers and investors.

2.2 Business Requirements

To address the business problems outlined above, the following business requirements need to be fulfilled:

Comprehensive Data Analysis:

i. Perform in-depth analysis of the loan data to identify patterns, trends, and anomalies.
ii. Evaluate borrower profiles to determine risk factors associated with loan defaults.
iii. Analyze the performance of different loan types and identify key success factors.

Risk Assessment and Management:
Develop risk assessment models to predict the likelihood of loan defaults.
Create risk profiles for different borrower categories and loan types.
Implement strategies to mitigate identified risks and reduce default rates.

Performance Metrics and Dashboards:
Design interactive dashboards to visualize key performance metrics such as default rates, repayment schedules, and loan distributions.
Enable stakeholders to easily access and interpret these metrics for better decision-making.

Decision Support System:
Provide tools and reports to help investors make informed decisions regarding their investment portfolios.
Offer insights and recommendations to LendingClub's management to refine lending policies and practices.

Regulatory Compliance:
Ensure all data analysis and reporting processes comply with relevant financial regulations and data protection laws.
Maintain audit trails and documentation to support regulatory audits and reviews.

2.3 Literature Survey
A thorough literature survey was conducted to understand the current state

of peer-to-peer lending analysis and the application of data analytics in this domain. Key findings from the literature include:

Risk Prediction Models:

Studies have explored various machine learning models, such as logistic regression, decision trees, and neural networks, for predicting loan defaults. The use of ensemble methods and hybrid models has shown improved accuracy in risk prediction .

Loan Performance Analysis:

Research indicates that borrower characteristics, such as credit scores, income levels, and employment status, are significant predictors of loan performance .

Macro-economic factors, such as interest rates and economic cycles, also influence loan repayment behaviors .

Data Visualization Techniques:

Effective visualization techniques are crucial for interpreting complex loan data. Interactive dashboards and real-time data visualizations enhance stakeholders' ability to make data-driven decisions .

Tools like Qlik have been recognized for their robust data integration and visualization capabilities, making them suitable for financial data analysis .

Regulatory Considerations:

Compliance with regulations such as the General Data Protection Regulation (GDPR) and financial industry standards is essential for maintaining trust and legal standing .

Studies emphasize the need for transparent data practices and robust security measures to protect sensitive financial information .

3. Data Collection

3.1 Collect the Dataset

The primary dataset used in this project consists of detailed information on loans issued by LendingClub. This dataset is publicly available and includes various attributes such as loan amount, interest rate, borrower information, and repayment status. It provides a comprehensive view of the lending activities and borrower characteristics, enabling a thorough analysis of loan performance and risk factors.

The dataset can be accessed via the following link:

https://drive.google.com/file/d/13-q5tTVNZRkEaXu5kggTbwJjbbTla3K7/view

Dataset Features

Loan Attributes: Includes information such as loan amount, term, interest rate, installment amount, loan status, and issue date.

Borrower Information: Contains details like the borrower's employment status, annual income, credit score, and debt-to-income ratio.

Repayment Data: Provides data on loan payments, including payment dates, payment amounts, and remaining balance.

Additional Attributes: Includes other relevant variables such as grade, subgrade, purpose of the loan, and more.

This dataset will serve as the foundation for our analysis, providing the necessary data to explore trends, assess risks, and optimize loan performance.

3.2 Connect Data with Qlik Sense

Once the dataset has been collected, the next step is to connect it with Qlik Sense to enable data visualization and analysis. Qlik Sense offers a user-friendly interface for importing and integrating datasets, allowing for seamless analysis and insightful visualizations. The process to connect the data with Qlik Sense involves the following steps:

Set Up Qlik Sense Environment:

Ensure that Qlik Sense Desktop or Qlik Sense Enterprise is installed and properly configured on your system.

Open Qlik Sense and create a new app or use an existing one where the data will be loaded.

Load the Dataset:

Open the Qlik Sense app and navigate to the "Data load editor" from the main menu.

Click on "Create new connection" and select "Folder" to connect to the local storage where the dataset file is saved.

Upload the dataset file (CSV, Excel, or other supported formats) to Qlik Sense.

Once uploaded, you can preview the dataset to ensure all fields are correctly recognized.

Data Modeling:

After loading the data, create the necessary associations between different data tables within Qlik Sense to ensure a coherent data model.

This step is crucial for enabling accurate and efficient data querying and visualization.

Create Visualizations and Dashboards:

Use Qlik Sense's drag-and-drop interface to create various visualizations such as bar charts, line charts, pie charts, and tables.

Build interactive dashboards that allow users to filter data, drill down into details, and explore different aspects of the loan dataset.

4. Data Preparation
4.1 Prepare the Data for Visualization

Effective data preparation is a crucial step in ensuring that the data is clean, consistent, and ready for visualization. This process involves data cleaning, transformation, and the creation of necessary fields to support

insightful visualizations. Here's a detailed guide on preparing the LendingClub loan dataset for visualization in Qlik Sense:

Data Cleaning:
Handling Missing Values: Identify and address missing values in the dataset. Common strategies include filling in missing values with the mean, median, or mode, or dropping rows/columns with significant amounts of missing data.

Removing Duplicates: Check for and remove any duplicate records to ensure data integrity.

Standardizing Formats: Ensure consistency in data formats, such as dates and categorical variables. For instance, standardize date formats to a common structure (e.g., YYYY-MM-DD).

Data Transformation:
Encoding Categorical Variables: Convert categorical variables into numerical values using techniques such as one-hot encoding or label encoding. This is particularly useful for machine learning models and some visualizations.

Normalizing Numerical Values: Normalize or scale numerical values to ensure that they are on a comparable scale. This can be done using min-max scaling or z-score normalization.

Creating New Features: Generate new features that can provide additional insights. For example:

Loan-to-Income Ratio: Calculate the ratio of loan amount to the borrower's annual income.

Debt-to-Income Ratio (DTI): If not already present, calculate this key indicator of a borrower's financial health.

Loan Duration in Months: Convert loan term into a numerical value (e.g., 36 or 60 months).

Data Aggregation:

Summarize Data: Aggregate data at different levels, such as by loan grade, purpose, or borrower state. This helps in understanding overall trends and patterns.

Grouping Data: Group data by relevant categories for easier analysis and visualization. For instance, group loans by issue date to analyze trends over time.

Data Validation:

Verify Data Consistency: Ensure that the transformed data maintains its integrity and consistency. Cross-verify key metrics and statistics with the original dataset.

Check for Anomalies: Identify and address any anomalies or outliers that might skew the analysis. This can involve setting thresholds or using statistical methods to detect unusual data points.

5. Data Visualizations

5.1 Visualization

Data visualization is a key step in transforming prepared data into actionable insights. Using Qlik Sense, various visualizations can be created to explore different aspects of the LendingClub loan dataset. Below are some suggested visualizations along with a brief description of their purpose and the insights they aim to provide:

Total Number of Accounts:
Visualization Type: Single Number or Bar Chart
Description: Display the total count of accounts.
Example: "Total Number of Accounts: 10,000"

Total Loan Amount:
Visualization Type: Single Number or Bar Chart
Description: Show the total loan amount issued.
Example: "Total Loan Amount: $100,000,000"

Average Loan Amount:
Visualization Type: Single Number or Bar Chart
Description: Present the average loan amount across all loans.
Example: "Average Loan Amount: $10,000"

Average Loan Amount for Account Type:
Visualization Type: Bar Chart
Description: Show the average loan amount for each account type (e.g., individual, joint).
Example: "Average Loan Amount for Individual Accounts: $9,000"
"Average Loan Amount for Joint Accounts: $12,000"

Average Loan Interest Rate Based on Grade:
Visualization Type: Bar Chart
Description: Display the average loan interest rate for each loan grade (e.g., A, B, C).
Example: "Average Loan Interest Rate for Grade A: 8%"
"Average Loan Interest Rate for Grade B: 10%"

State-wise Average Loan Amount:
Visualization Type: Map Chart or Bar Chart
Description: Represent the average loan amount for each state.
Example: "California: $11,000"
"Texas: $9,500"

Tenure-wise Average Loan Amount:
Visualization Type: Bar Chart
Description: Show the average loan amount for different loan tenures (e.g., 36 months, 60 months).
Example: "Average Loan Amount for 36 Months: $9,800"
"Average Loan Amount for 60 Months: $11,200"

The Number of Accounts (Individual/Joint):

Visualization Type: Pie Chart or Bar Chart

Description: Show the count of individual and joint accounts.

Example: "Individual Accounts: 8,000"

"Joint Accounts: 2,000"

The Number of Members – Grade wise:

Visualization Type: Bar Chart

Description: Represent the count of members for each loan grade.

Example: "Grade A: 3,000 members"

"Grade B: 4,000 members"

The Number of Members – Verification Status:

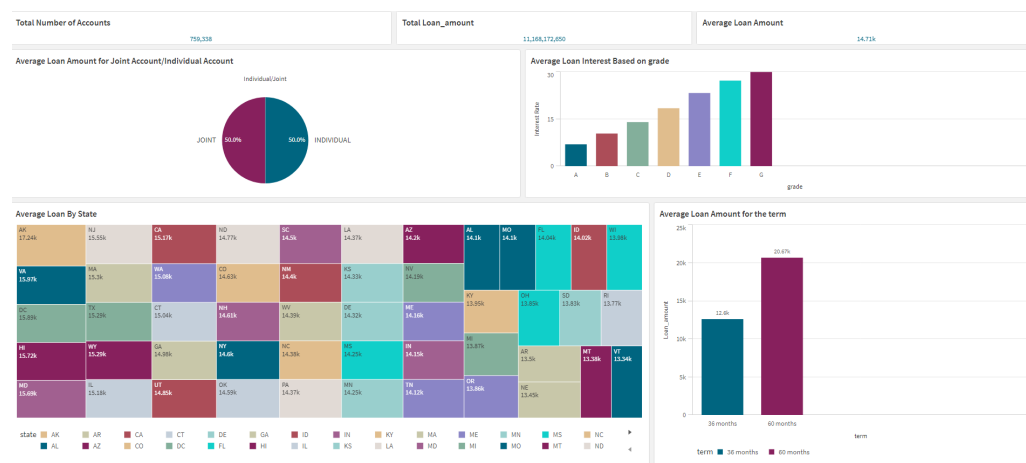Visualization Type: Bar Chart

Description: Display the count of members for each verification status (e.g., verified, not verified).
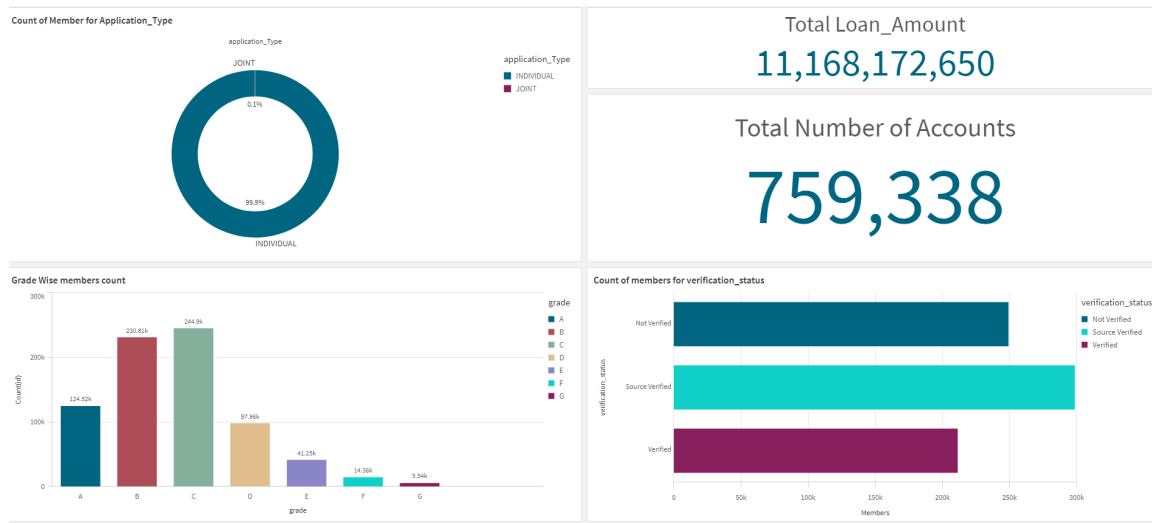
Example: "Verified Members: 6,000"

"Not Verified Members: 4,000"

6. Dashboard Design
Dashboard Design 1: Average Loan Amount Analysis

Dashboard Design 2: Count Analysis

Responsive Layout: Ensure the dashboard adapts to different devices.

Clear Navigation: Make it easy to move around the dashboard.

Consistent Branding: Use the same look and feel throughout.

Whitespace and Hierarchy: Use space effectively and prioritize important info.

Interactive Elements: Include filters and interactive features.

Data Integrity: Ensure accuracy and provide context for visuals.

Performance Optimization: Keep the dashboard fast and efficient.

Accessibility: Design with all users in mind, including those with disabilities.

7 . Performance Testing

Performance testing is crucial to ensure that the dashboard is responsive and efficient, even when handling large datasets. This involves assessing how well the dashboard performs under different conditions and optimizing as necessary.

7.1 Amount of Data Rendered

Objective: To test the performance of the dashboard when rendering varying amounts of data.

Steps:

Determine Data Volumes:
Start with a small subset of data.
Gradually increase the volume of data (e.g., 10%, 25%, 50%, 100% of the dataset).

Measure Load Times:
Record the time taken to load the dashboard for each data volume.
Note any significant delays or performance degradation.

Evaluate Responsiveness:
Assess the responsiveness of interactive elements (filters, drill-downs, etc.) with different data volumes.
Check for any lag or slow response times.

Optimize Data Loading:
Use techniques such as data aggregation, indexing, and caching to improve performance.
Consider loading only necessary data initially and fetching additional data

on demand.

8.2 Utilization of Data Filters
Objective: To test how effectively data filters are utilized and their impact on dashboard performance.

Steps:

Apply Filters:
Test various filter combinations (e.g., date ranges, loan grades, account types).
Ensure that filters work correctly and return accurate results.

Measure Filter Performance:
Record the time taken for the dashboard to update after applying filters.
Check the consistency and accuracy of the filtered data.

Optimize Filter Queries:
Simplify filter queries to reduce processing time.
Use pre-calculated fields or indexed columns where possible.

Test Concurrent Usage:
Simulate multiple users applying filters simultaneously.
Monitor the system's performance under concurrent usage.

Performance Testing Summary

1. Amount of Data Rendered:
Gradually increase data volume and measure performance.
Optimize data loading strategies to handle large datasets efficiently.

2. Utilization of Data Filters:

Test various filter combinations and measure performance.

Optimize filter queries and test concurrent usage.