Lending Club Issued Loan Analysis

INTRODUCTION

Overview

This project focuses on improving the lending strategy of a financial institution by leveraging advanced data analytics on LendingClub loan data. The institution's current strategy faces significant challenges, including:

- **Inaccurate Risk Identification:** Struggles to accurately identify high-risk borrowers.
- **Difficulty Predicting Defaults:** Challenges in forecasting loan default rates effectively.
- **Static Lending Criteria:** Inability to adapt lending criteria dynamically based on evolving data insights.

By implementing a comprehensive data analytics framework, the institution can gain valuable insights into borrower behaviour, predict default rates more precisely, and adjust lending criteria in real-time, thereby enhancing risk management and decision-making processes.

Purpose

The primary objectives of this project include:

1. **Enhance Risk Identification:** Improve the precision in identifying high-risk borrowers to minimize loan defaults and associated financial losses.
2. **Predict Loan Default Rates:** Develop sophisticated predictive models to accurately forecast default rates, enhancing risk management.
3. **Dynamic Lending Criteria:** Enable real-time modifications to lending criteria based on evolving market conditions and borrower behaviour insights.
4. **Optimize Lending Strategy:** Make the overall lending strategy more responsive and data-driven, increasing the institution's competitive advantage.
5. **Seamless Integration:** Ensure that the analytics framework integrates smoothly with existing systems to facilitate easy implementation and scalability.

Define Problem

Specify The Business Problem

The financial institution's current lending strategy is inadequate due to a lack of comprehensive insights derived from LendingClub loan data. This results in:

- **Ineffective Borrower Assessment:** Difficulty in accurately assessing borrower behaviour and market dynamics.
- **Inaccurate Risk Identification:** Challenges in identifying high-risk borrowers.
- **Inflexible Lending Criteria:** Inability to adjust lending criteria dynamically in response to changing market conditions.

Business Requirements

The institution needs a robust data analytics framework that can:

- Extract meaningful insights from LendingClub loan data.
- Provide a deep understanding of borrower behaviour.
- Identify high-risk segments.
- Accurately predict default rates.
- Support real-time adjustments to lending criteria.
- Be scalable, adaptable, and integrate seamlessly with existing systems.

Literature Survey

1. **Dashboards and Visualization in Finance**

    - **Overview:** Effective dashboards and visualization techniques are critical in finance for translating complex data into actionable insights.

    - **Key Insights:**
        - *Study by Few (2006):* Demonstrated that well-designed dashboards enhance data interpretation and decision-making speed.

        - *Research by Heer et al. (2010):* Highlighted the importance of interactive visuals for data exploration.

2. **Qlik for Dashboard Creation**

- **Overview:** Qlik, including QlikView and Qlik Sense, offers powerful tools for creating interactive dashboards.
- **Key Features:**
  - *Associative Data Model:* Enables free data exploration without predefined queries.
  - *Self-Service BI:* Allows non-technical users to generate reports and dashboards independently.
- **Case Studies:**
  - *Eckerson Group (2017):* Showcased Qlik's impact on financial analytics, providing real-time insights and facilitating swift decision-making.
  - *Qlik Case Study (2019):* Demonstrated a financial institution streamlining reporting processes and gaining comprehensive financial performance insights using Qlik Sense.

3. **Trends in Financial Visualization Tools**

- **Overview:** Financial visualization tools are evolving to include real-time updates, AI integration, and enhanced user experience.

- **Key Trends:**
  - *Real-Time Dashboards:* Offer up-to-the-minute insights into financial data.
  - *AI Integration:* Utilize AI and machine learning for predictive analytics and advanced insights.
  - *Enhanced User Experience:* Prioritize user-friendly interfaces and interactive features for intuitive data exploration.

Data Collection

Collect the Dataset

Data collection involves gathering and measuring information on variables of interest systematically to answer research questions, test hypotheses, evaluate outcomes, and generate insights.

- **LendingClub Issued Loans | Kaggle:**
  - LendingClub's complete loan data issued from 2007-2017.
  - Dataset on Kaggle

Understand The Data

Data contains all the meta information regarding the columns described in the CSV files

Column Description of the Dataset:

member_id: Contains unique member id of the member
loan_amnt: Contains the loan amount taken by members
term: Contains the tenure for the loan_amount
int_rate: Rate of Interest for the loan_amount

grade: Grades of the members

Data Preparation

Prepare The Data For Visualization

Preparing data for visualization involves several key steps to ensure it is suitable for creating insightful visual representations:

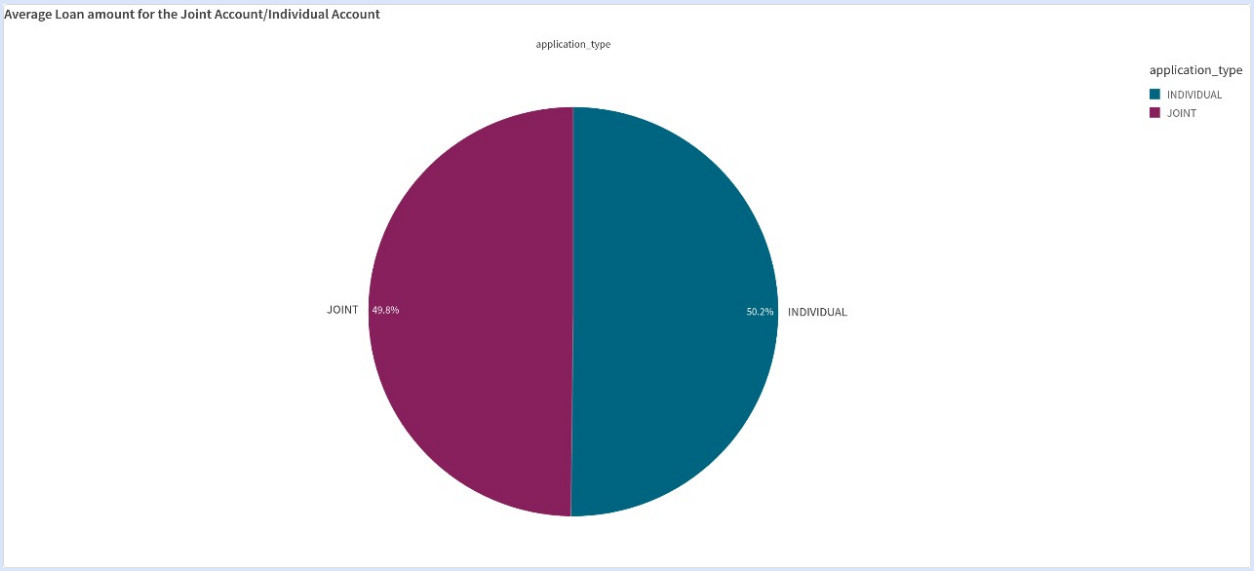
1. **Data Cleaning:** Remove irrelevant or missing data.
2. **Data Transformation:** Format the data appropriately for visualization.
3. **Data Exploration:** Identify patterns and trends within the data.
4. **Data Filtering:** Focus on specific subsets of data.
5. **Software Preparation:** Make sure the data is ready for use in visualization software.
6. **Data Accuracy Check:** Verify the data's accuracy and completeness.

Data Visualization

Data visualization involves creating graphical representations of data to help users understand and explore complex information. The goal is to make data more accessible, intuitive, and easier to interpret. By using visual elements like charts, graphs, and maps, data visualizations enable users to quickly identify patterns, trends, and outliers, leading to more informed decision-making.

1. **Loan Amount Analysis**

2. Average Loan Amount For The Account_Type



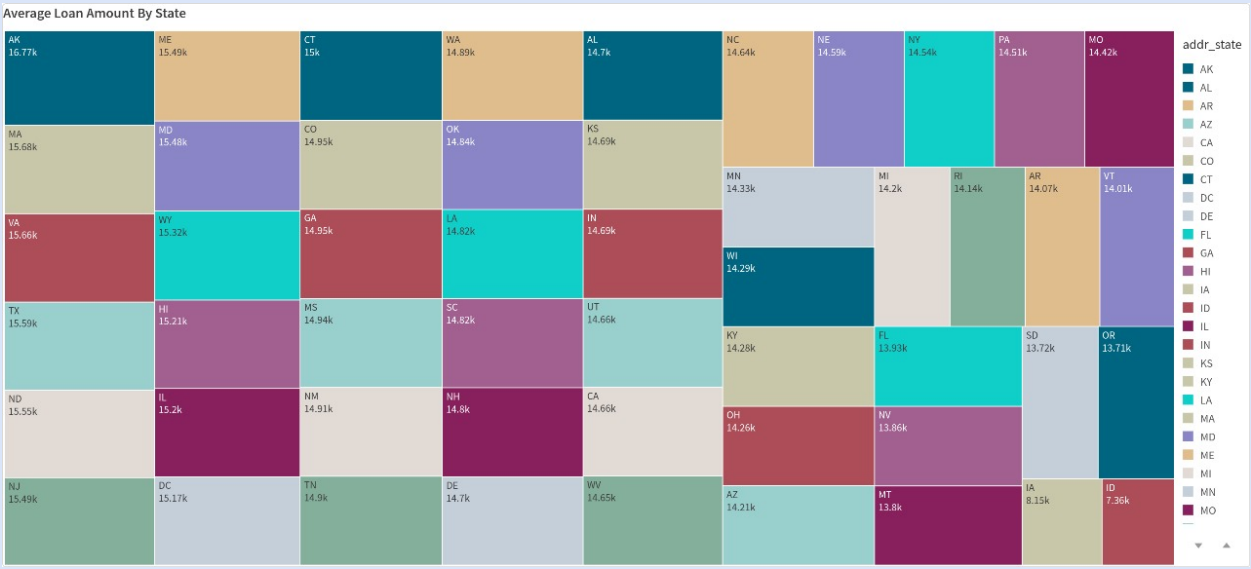
3. Total Loan Amount

4. Total Number Of Loan Account

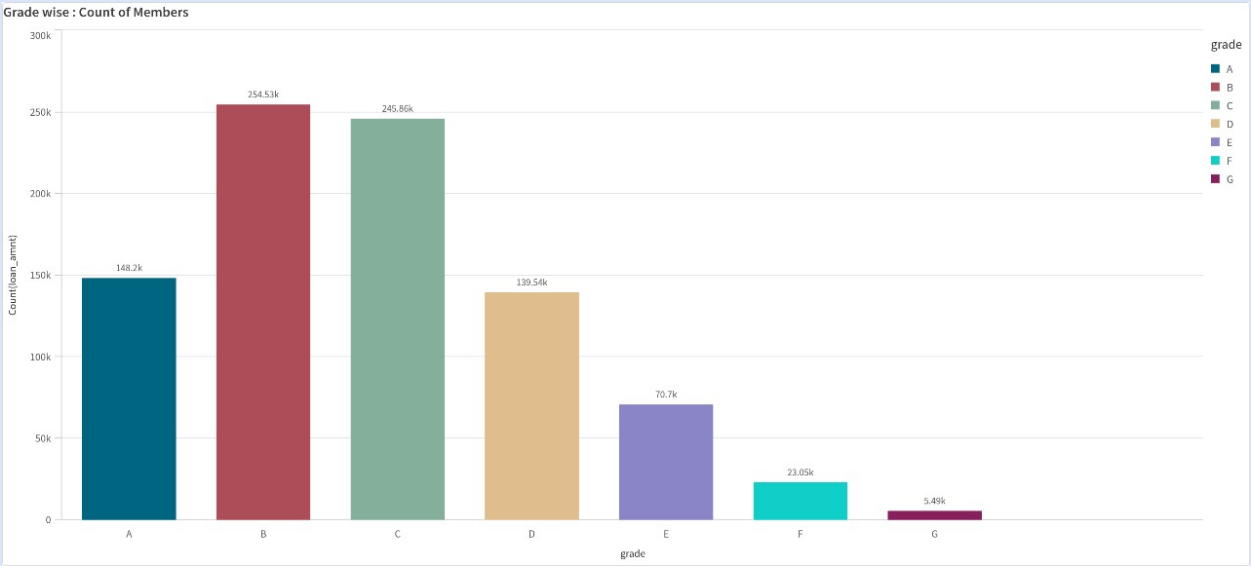
Total Number of Account

887.4k

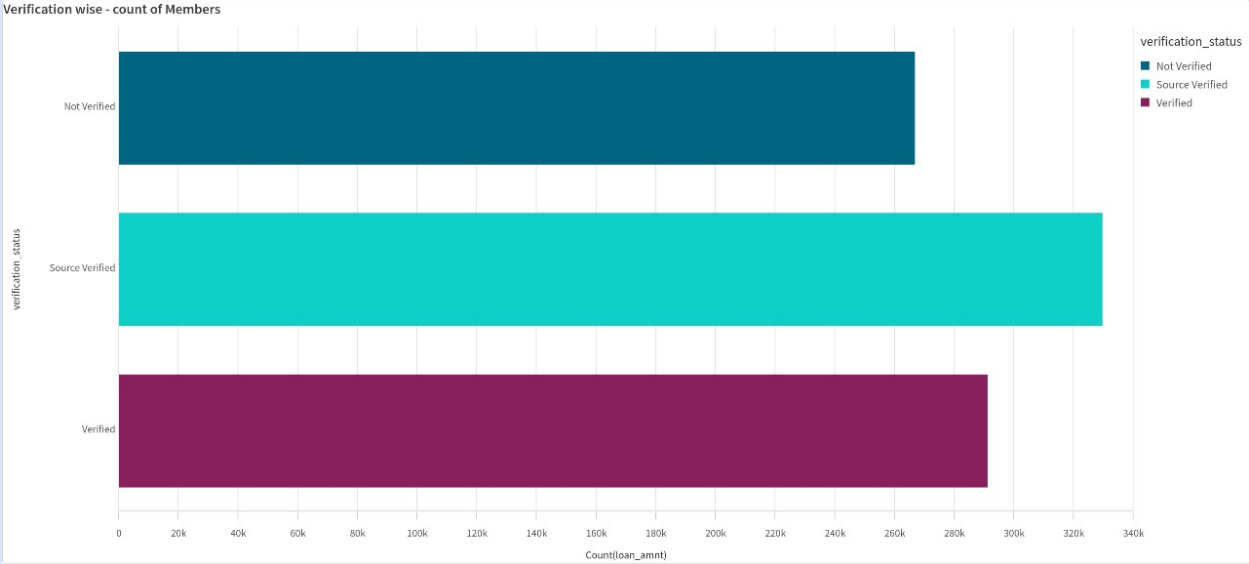
5. Average Loan Amount - State Wise



6. Grade Wise - Count Of Members



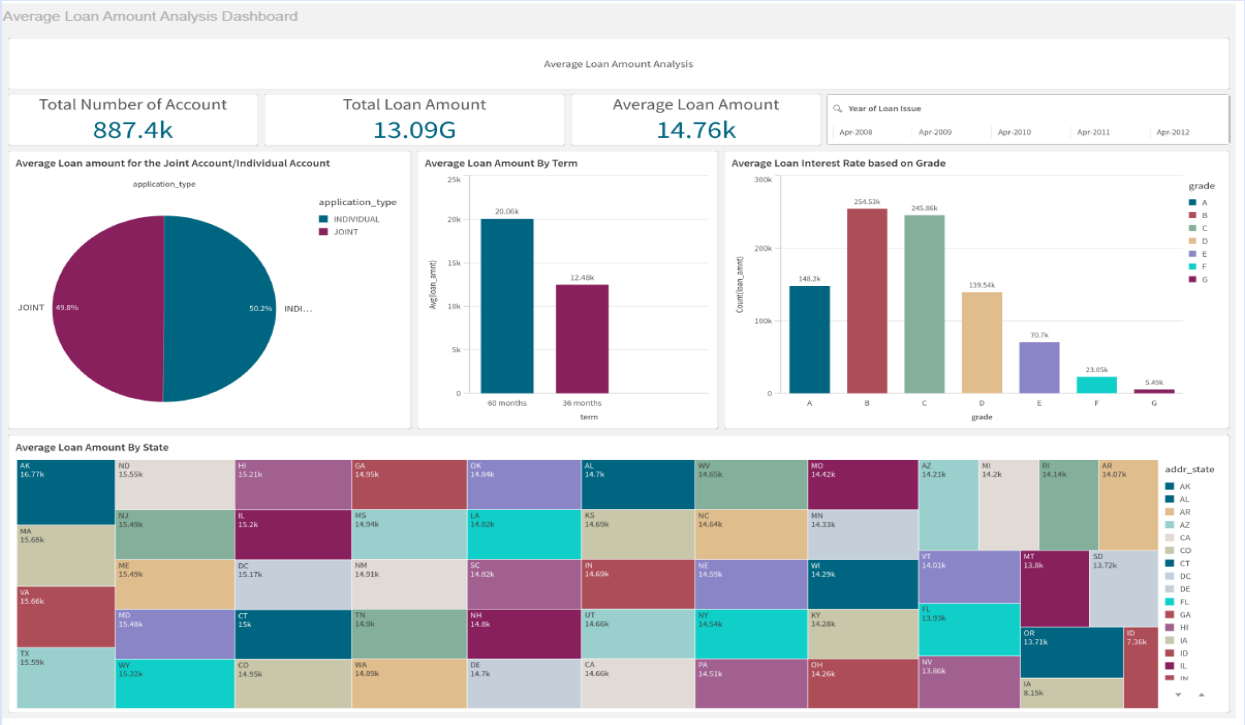
7. **Verification Status**



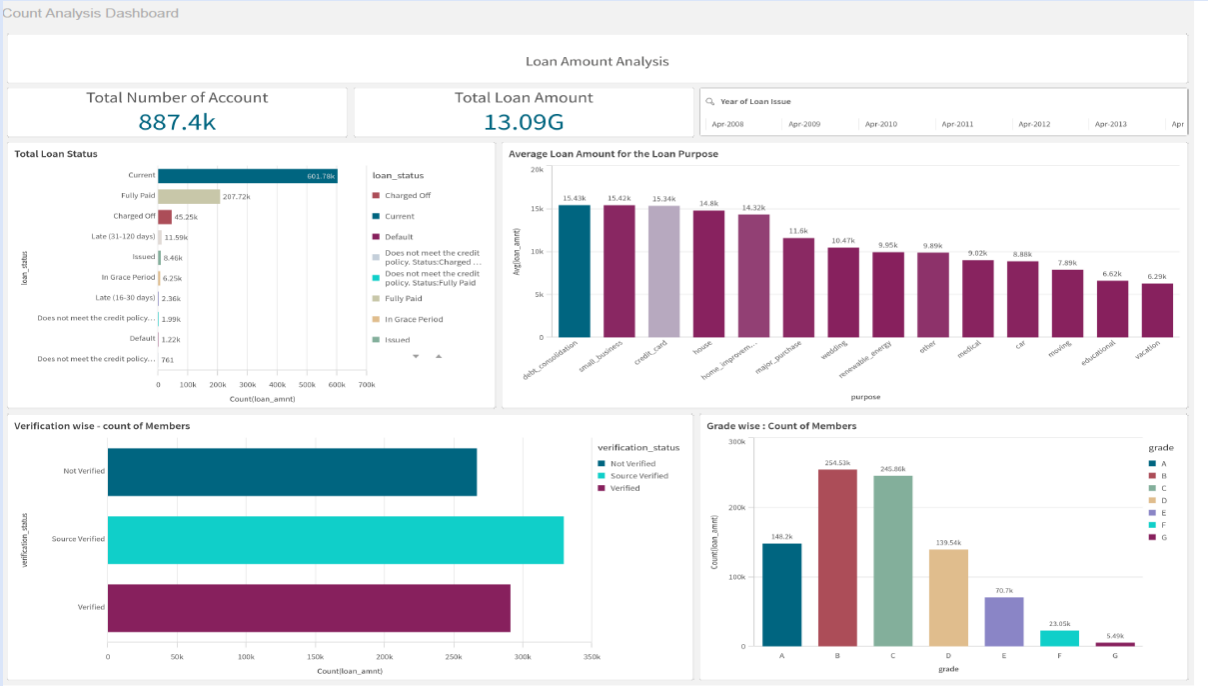
Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Dashboard 1: *Average Loan Amount Analysis*

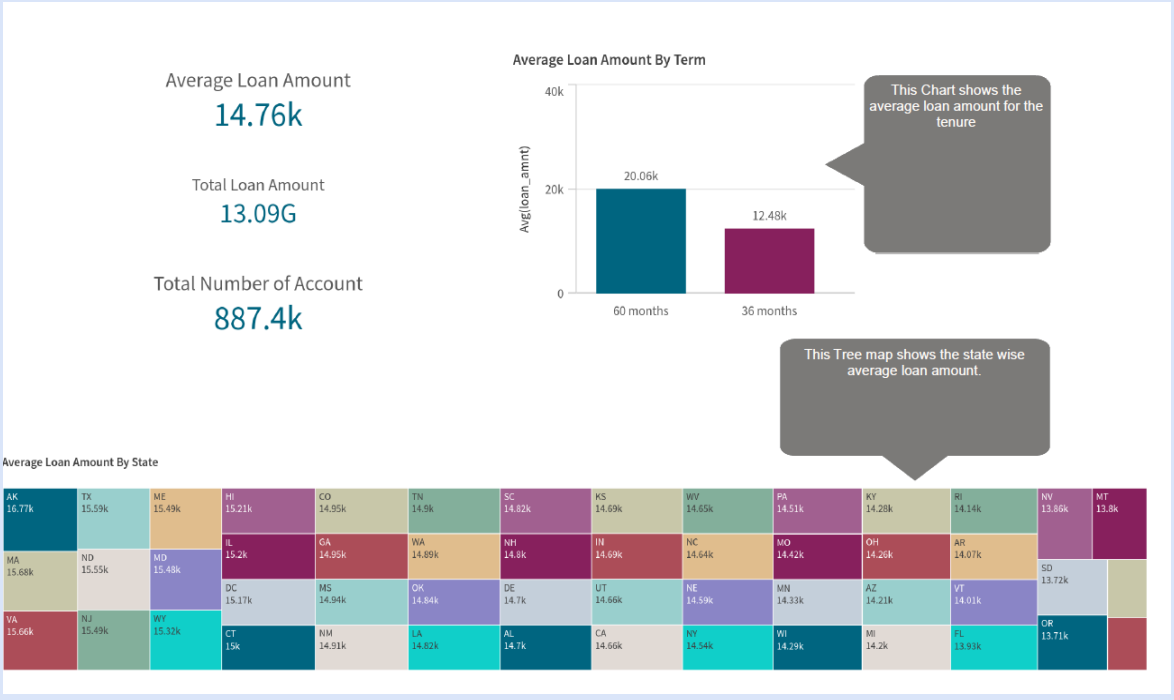


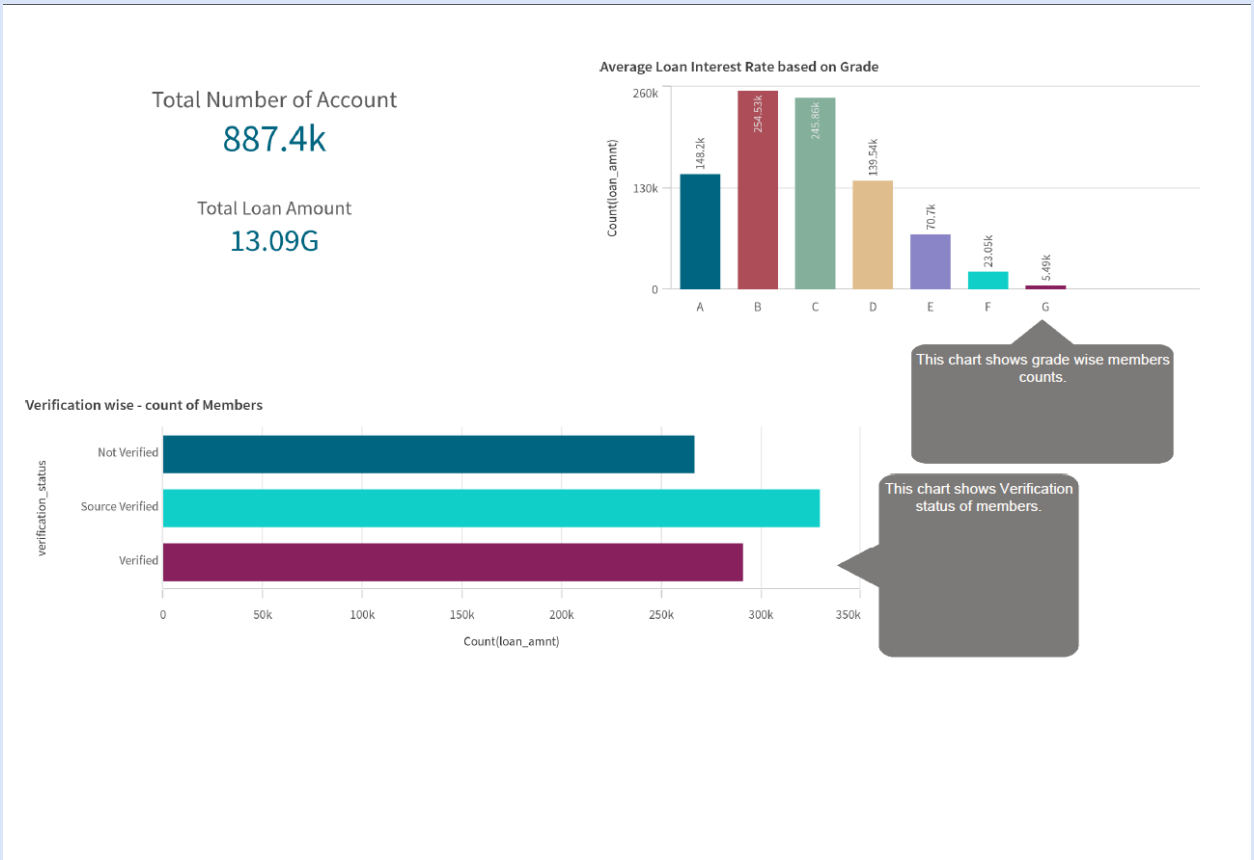
Dashboard 2: *Count Analysis*



Story

A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

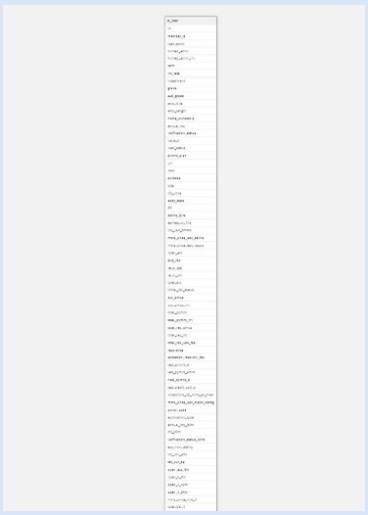


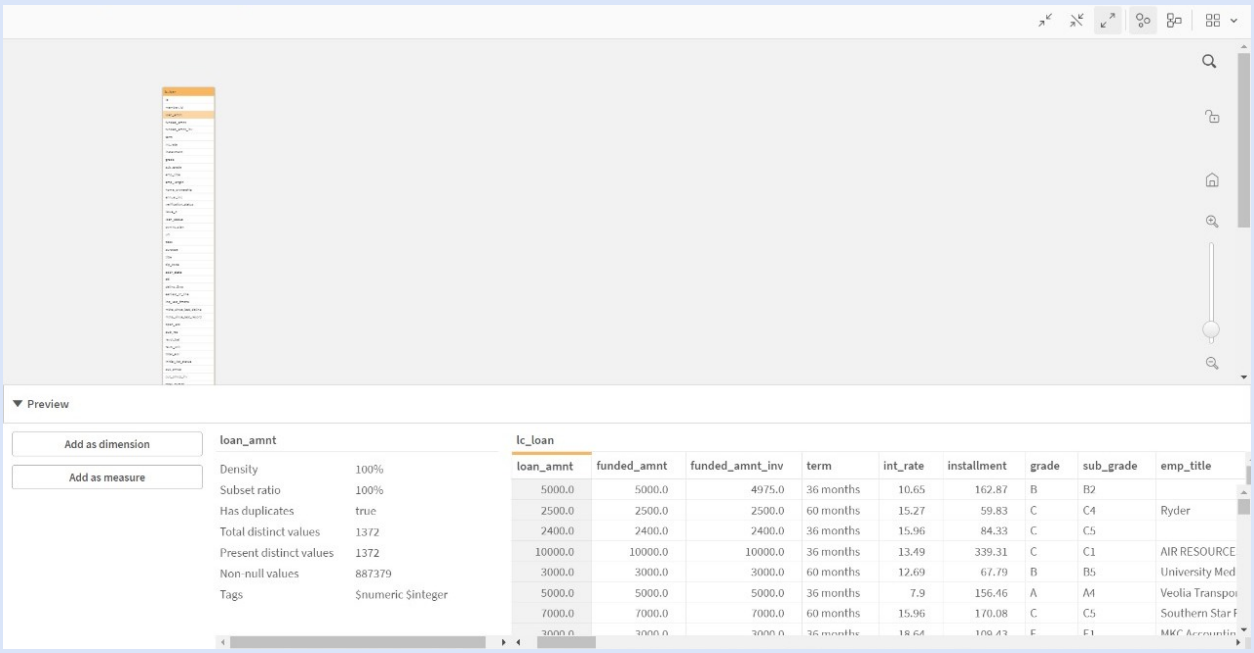


Performance Testing

Amount of data loaded

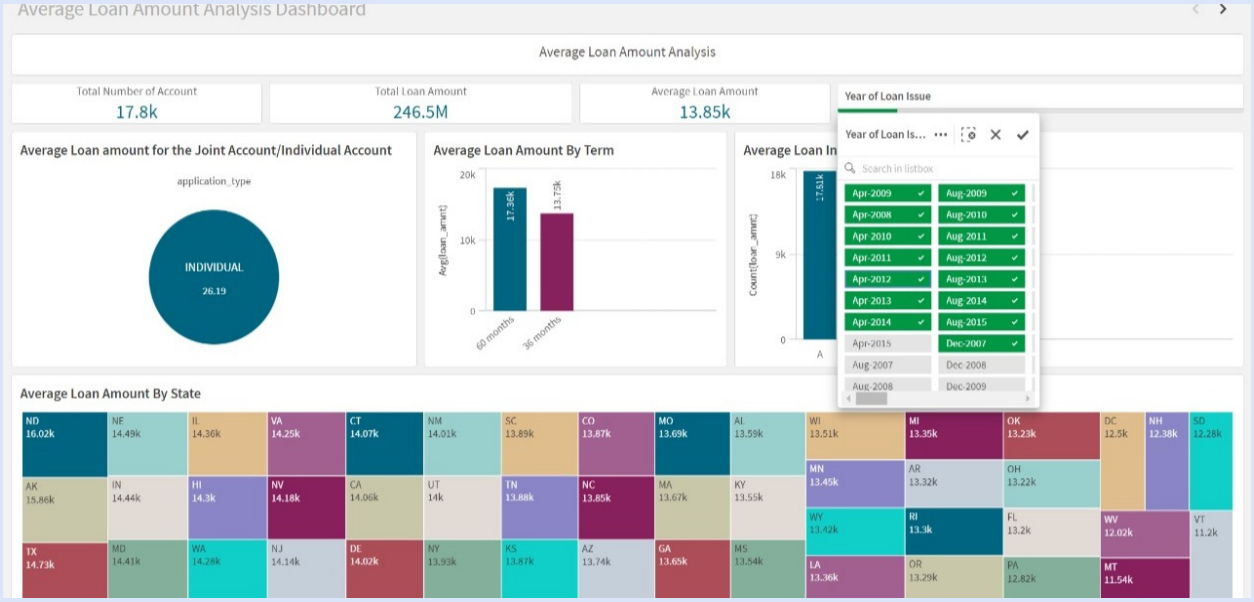
Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system





Utilization of Filters

"Utilization of Filters" refers to the application or use of filters within a system, software application, or data processing pipeline to selectively extract, manipulate, or analyze data based on specified criteria or conditions. Filters are used to narrow down the scope of data, focusing only on the relevant information that meets certain predefined criteria.


Average Loan Amount Analysis Dashboard
Average Loan Amount Analysis
Total Number of Account
17.8k
Total Loan Amount
246.5M
Average Loan Amount
13.85k
Year of Loan Issue
Average Loan amount for the Joint Account/Individual Account
application_type
INDIVIDUAL
26.19
Average Loan Amount By Term
20k
10k
0
Avg([loan_amnt])
17.58k
13.75k
60 months
36 months
Average Loan In
18k
9k
0
Count([loan_amnt])
17.51k
A
Year of Loan Is...
Search in listbox
Apr-2009
Aug-2009
Apr-2008
Aug-2010
Apr-2010
Aug-2011
Apr-2011
Aug-2012
Apr-2012
Aug-2013
Apr-2013
Aug-2014
Apr-2014
Aug-2015
Apr-2015
Dec-2007
Aug-2007
Dec-2008
Aug-2008
Dec-2009
Average Loan Amount By State
ND 16.02k
NE 14.49k
IL 14.36k
VA 14.25k
CT 14.07k
NM 14.01k
SC 13.89k
CO 13.87k
MO 13.69k
AL 13.59k
WI 13.51k
MI 13.35k
OK 13.23k
DC 12.5k
NH 12.38k
SD 12.28k
AK 15.86k
IN 14.44k
HI 14.3k
NV 14.18k
CA 14.06k
UT 14k
TN 13.88k
NC 13.85k
MA 13.67k
KY 13.55k
MN 13.45k
AR 13.32k
OH 13.22k
WY 13.42k
RI 13.3k
FL 13.2k
WV 12.02k
VT 11.2k
TX 14.73k
MD 14.41k
WA 14.28k
NJ 14.14k
DE 14.02k
NY 13.93k
KS 13.87k
AZ 13.74k
GA 13.65k
MS 13.54k
LA 13.36k
OR 13.29k
PA 12.82k
MT 11.54k

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. Tenure wise Average Loan Amount
8. The number of Accounts (Individual/Joint)
9. The number of members – Grade wise
10. The number of members – Verification Status