Lending Club Issued Loan Analysis

Report prepared by Anmol Singh

1. INTRODUCTION

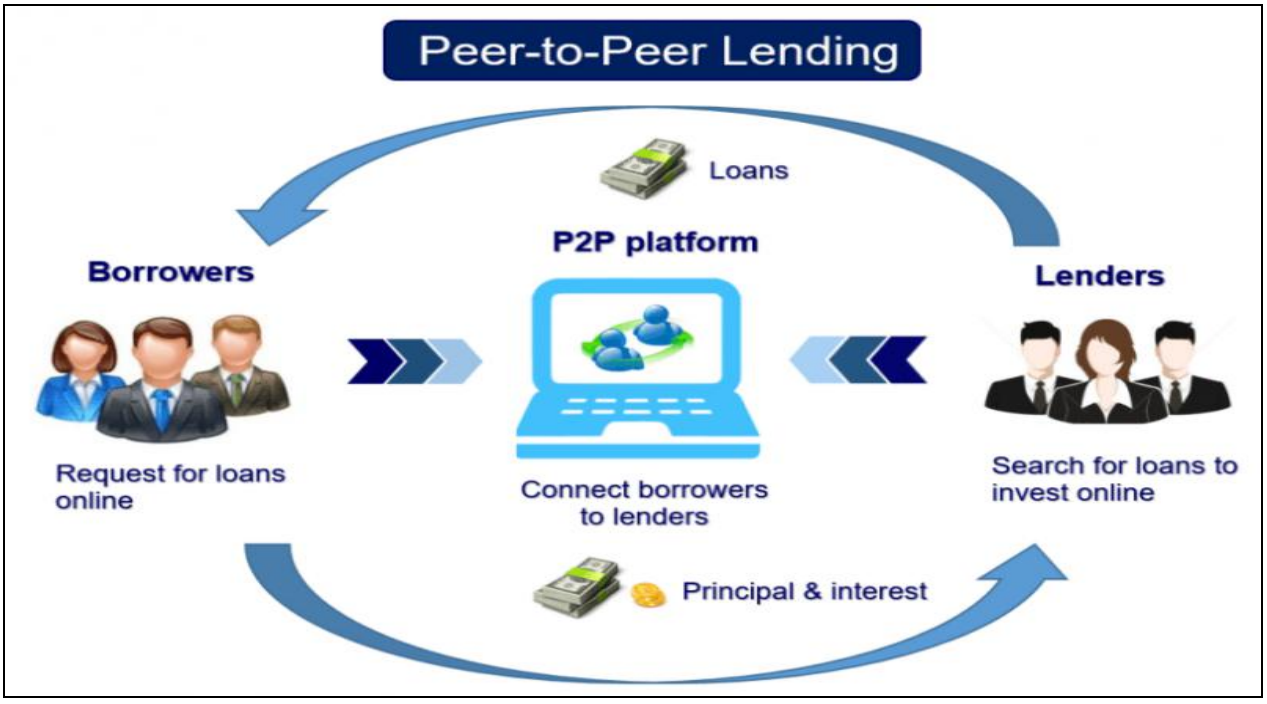
Project Title: LendingClub Issued Loans Analysis

1.1 Overview

The "LendingClub Issued Loans Analysis" project focuses on examining loan data from LendingClub, a leading peer-to-peer lending platform. The objective is to gain insights into borrower profiles, loan characteristics, and repayment performance, as well as to identify risk factors associated with loan defaults.

Using LendingClub's publicly available dataset, this analysis involves data cleaning, exploratory data analysis (EDA), and predictive modeling. The aim is to develop models that can predict loan outcomes, thus aiding in better risk assessment and decision-making for both LendingClub and its investors.

This project seeks to provide actionable insights that can enhance lending practices, improve investor strategies, and guide borrowers in understanding loan approval criteria.



1.2 Purpose

The purpose of the "LendingClub Issued Loans Analysis" project is to leverage data analytics to gain a comprehensive understanding of the dynamics involved in peer-to-peer lending through LendingClub. By analyzing detailed loan data, the project aims to:

1. **Identify Trends and Patterns**: Understand the characteristics and behaviors of borrowers, as well as the terms and conditions of the loans they receive.
2. **Assess Loan Performance**: Evaluate the repayment performance of loans to identify factors that contribute to successful repayments or defaults.
3. **Enhance Risk Management**: Develop predictive models to assess the risk of loan defaults, helping LendingClub and investors make informed lending decisions.
4. **Inform Strategic Decisions**: Provide insights that can help LendingClub optimize their lending practices, reduce default rates, and improve overall financial outcomes.
5. **Guide Borrowers**: Offer potential borrowers an understanding of the factors that influence loan approvals and terms, aiding them in making better financial decisions

1.3 Technical Architecture

The technical architecture for the "LendingClub Issued Loans Analysis" using Qlik involves several stages, from data ingestion to decision-making dashboards. Below is a detailed outline of the architecture:

1. Data Ingestion

- **Source Data**: LendingClub's publicly available loan dataset, which includes multiple variables such as loan amount, interest rate, borrower's credit score, employment length, and loan status.
- **Data Storage**: Raw data is stored in a centralized data repository, such as a data warehouse (e.g., Amazon Redshift, Google BigQuery, or Azure SQL Database).

2. Data Preprocessing

- **Data Cleaning**: Handle missing values, outliers, and inconsistencies using ETL (Extract, Transform, Load) tools such as Talend, Apache Nifi, or Qlik Data Catalyst.
- **Data Transformation**: Normalize and standardize data, perform feature engineering, and create derived metrics that are relevant for analysis.

3. Data Integration

- **Data Integration Platform**: Use Qlik Sense or QlikView for integrating various data sources and preparing data for analysis. Qlik's associative engine helps in connecting disparate data sources and enabling seamless exploration.
- **ETL Workflow**: Define ETL workflows to automate data cleaning, transformation, and loading into Qlik's in-memory data model.

4. Data Analysis and Modeling

- **Exploratory Data Analysis (EDA)**: Perform EDA using Qlik's visualization capabilities to uncover trends, patterns, and anomalies in the data.
- **Predictive Modeling**: Utilize machine learning tools like Python or R, integrated with Qlik, to build and train models that predict loan defaults and other outcomes. Libraries such as scikit-learn, TensorFlow, or H2O.ai can be used for this purpose.

5. Data Visualization

- **Dashboards and Reports**: Develop interactive dashboards and reports in Qlik Sense or QlikView to visualize key metrics, trends, and predictive model results.
- **Visualization Components**: Use charts, graphs, KPIs, and other visual elements to represent data insights effectively. Qlik's drag-and-drop interface simplifies the creation of these components.

2. DEFINE PROBLEM/PROBLEM UNDERSTANDING

2.1 Business problem

Lending Club faces the task of optimizing its underwriting criteria to attract creditworthy borrowers while minimizinng default risk. The business problem in analyzing Lending Club loans could be centered around optimizing investment decisions, Risk management, investor confidence, pricing strategy and borrower satisfaction.

Difficulties in predicting loan default rates exacerbate this problem hindering the institution's capacity to proactively manage risk and optimize portfolio performance

Impacts:

- **Financial Losses**: Inaccurate risk assessments and failure to predict defaults lead to financial losses due to bad debts.
- **Operational Inefficiencies**: Inefficient lending processes consume more time and

resources, reducing overall operational efficiency.

- **Competitive Disadvantage**: Without data-driven insights, the institution risks falling behind competitors who leverage advanced analytics to optimize their lending strategies.
- **Customer Dissatisfaction**: Inconsistent and inaccurate lending decisions can lead to customer dissatisfaction and loss of trust in the institution.

2.2 Business Requirements

The business requirements revolve around setting up a solid data analytics framework to dig out valuable insights from Lending Club loan data. Loan Requirement for Business in order to effectively analyze Lending Club issued loans data, the business requires a sturdy data analytics framework. The business requirements might also encompass the need for advanced machine learning algorithms to enhance predictive modeling, streamlinel decision making processes and optimize loan approval procedures.

Lenders nearly always evaluate the owner's personal credit when a small business owner wants money. Most lenders only lend to enterprises having a track record of at least two years making it difficult for business to obtain finance. lenders consider debt backed by valuable assets to be less risky. Lenders evaluate the risk of your type of business throughout the loan approval procedure.

2.3 Literature survey

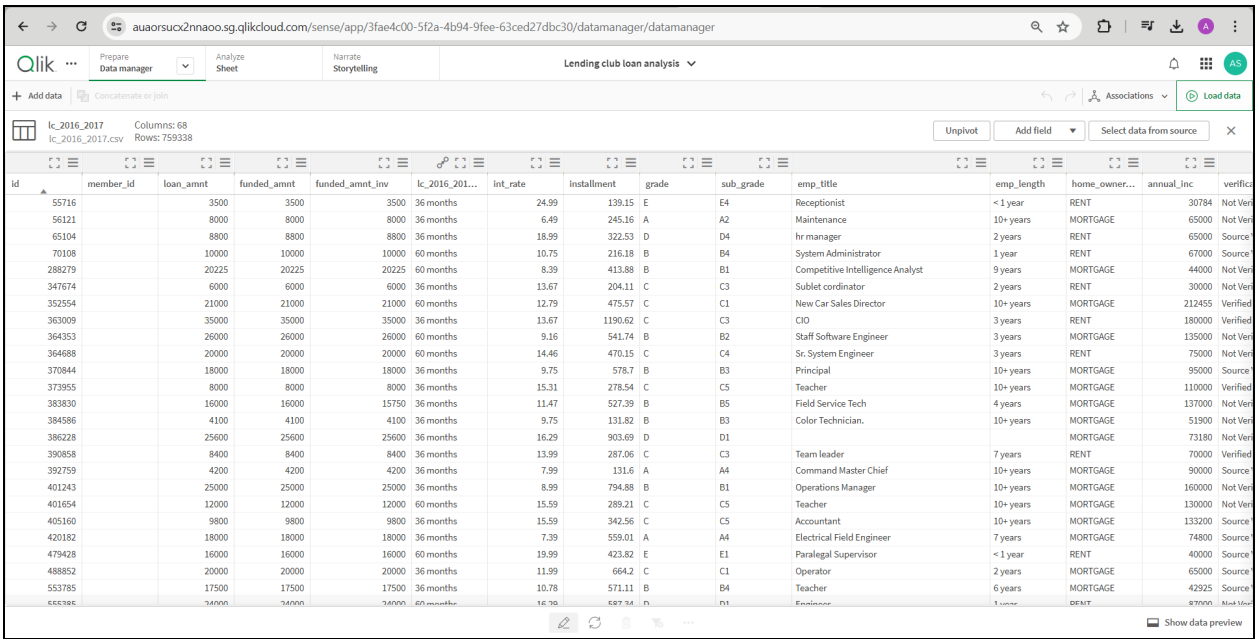
A comprehensive literature survey is essential to understand existing methodologies, tools, and best practices in optimizing lending strategies and risk management in the context of peer-to-peer lending platforms such as LendingClub. This survey aims to identify relevant studies on data analytics in finance, peer-to-peer lending trends, and similar analyses conducted by financial institutions to enhance their decision-making processes. The challenges faced in analyzing lending club loan data and purpose future research direction to enhance predective modeling and decision making

3. DATA COLLECTION

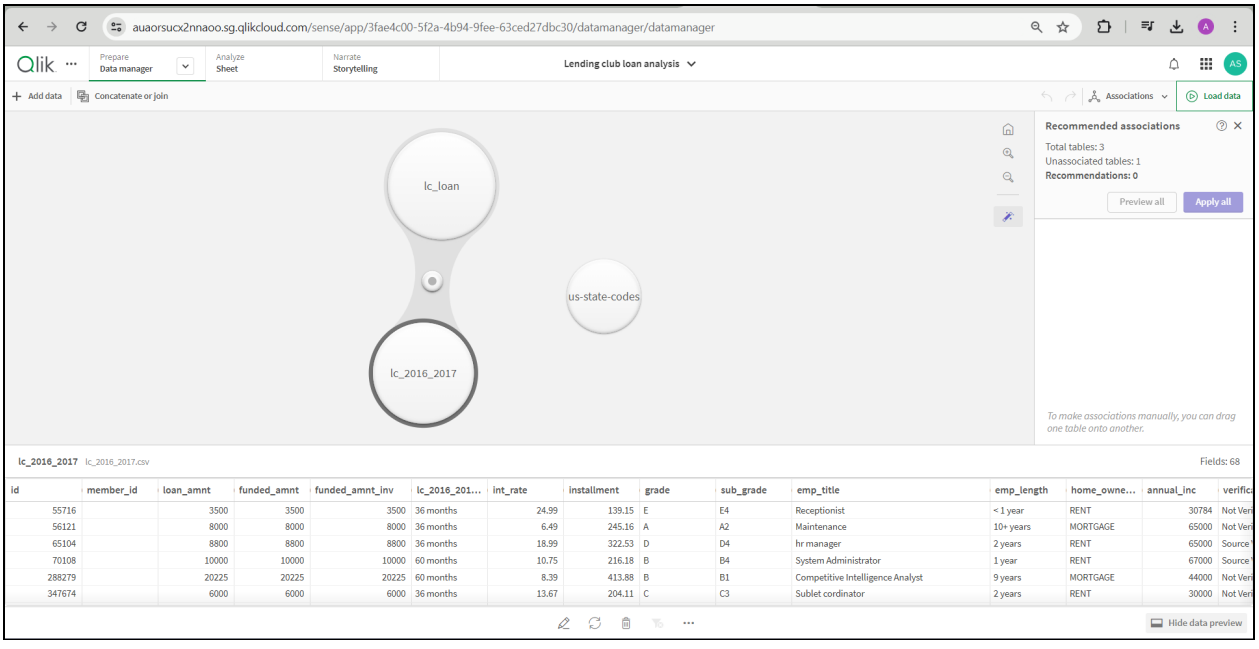
Data collection and extraction from a database involve gathering information stored in a database for analysis or other purposes. We need understand the data porperly in lending club loan analysis, data collection from a database would entail retrieving loan related data from the Lending Club database.

Data collection is process of gathering information from various sources such as the lending club website, data repositories like kaggle and other sources. This data typically inculdes detailes about the loan issued by Lendingclub. Such as loan amount, interest rate, term length, borrower's credit score, loan status, verification status, employment length, loan purpose and more. The dataset showing in the given below.

3.1 Dataset



3.2 Connect data with qlik sense



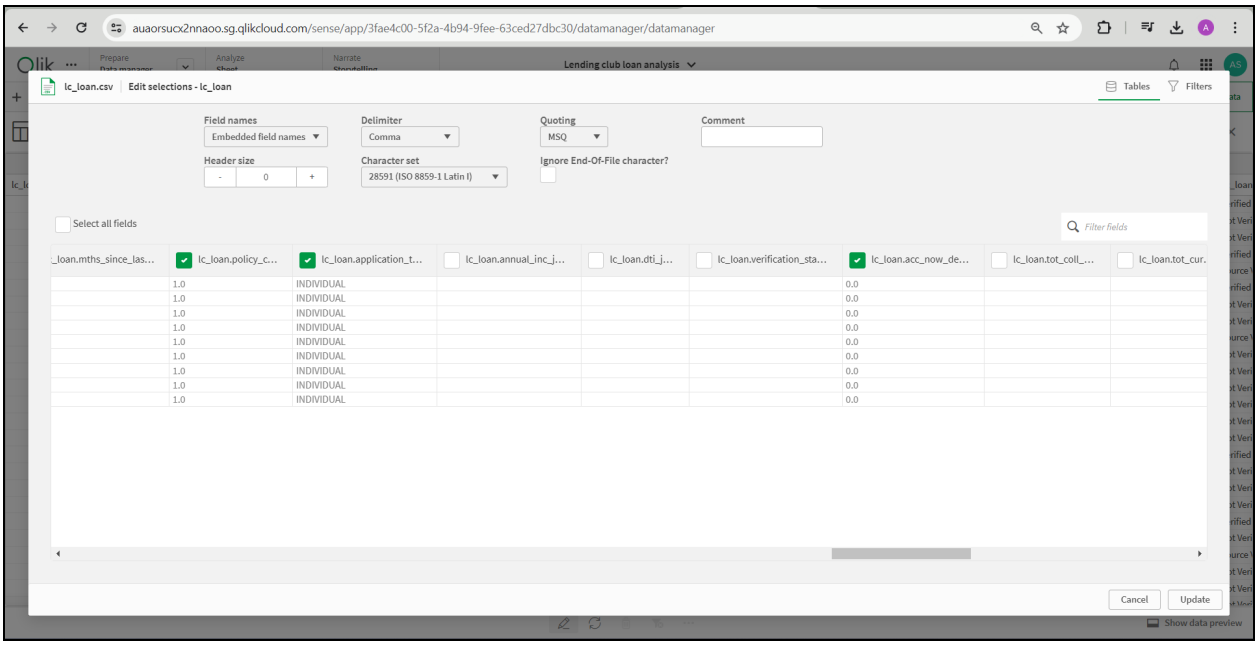
4. DATA PREPARATION

4.1 Preparing data for visualization

To prepare the data for visualization in Lending club loan analysis, first start cleaning the data to handle missing values, outliers, inconsistencies and remove the unnecessary data or unwanted data. This involves tasks like imputing missing values, removing duplicates and standardzizing data formats.

Next step is transform the data by creating new variables, aggregating information for better visualization. This step helps in shaping the data to highlight the key insights we want to visualize.

Organize the data in a format suitable for the visualization tool you plan to use. such as a spreadsheet. This will make it easier to create meaningful and insightful visualizations to communicate your findings effectively.



5. DATA VISUALIZATIONS

Data visualization is all about representing data in a visual format like chart, graphs and maps to help people understand the information more eaily. When it comes to Lending club issued loan analysis we can use data visualization to show patterns, trnds and relationships in the loan data. The graph is easily understood by everyone without any prior knowledge and it is also save the time, it allowa us to relate and compare the data for different periods. It always depends on

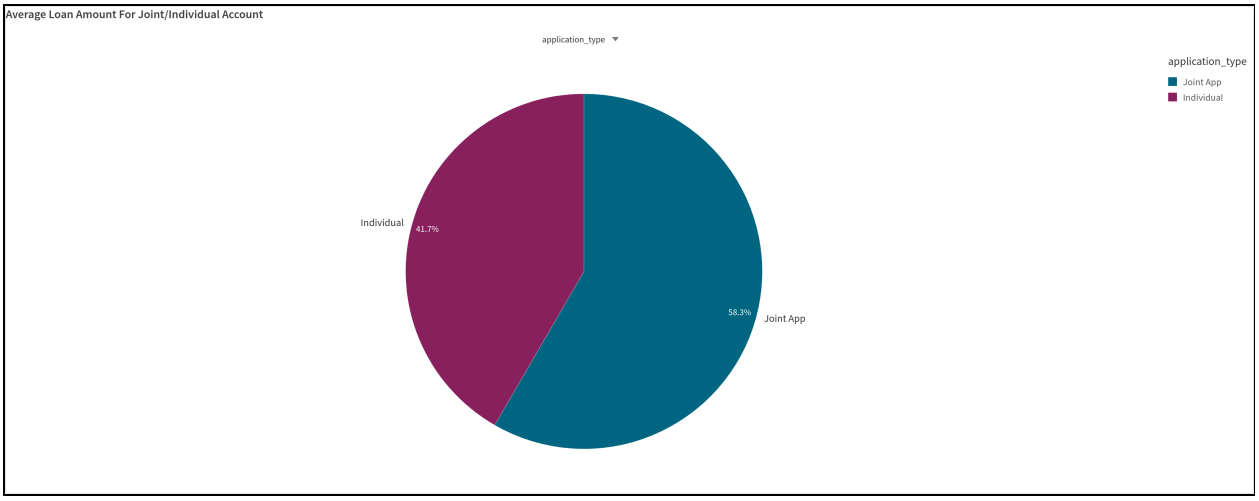
the type of information in a particular domain. There are different types of graphical representation. Some of them are as follows.

- Bar Chart
- Bullet Chart
- Combo Chart
- Gauge
- Histogram
- KPI
- Line Chart
- Map
- Pie Chart
- NL insights
- Pivot table
- Scatter plot
- Table
- Treemap

5.1 Visualization

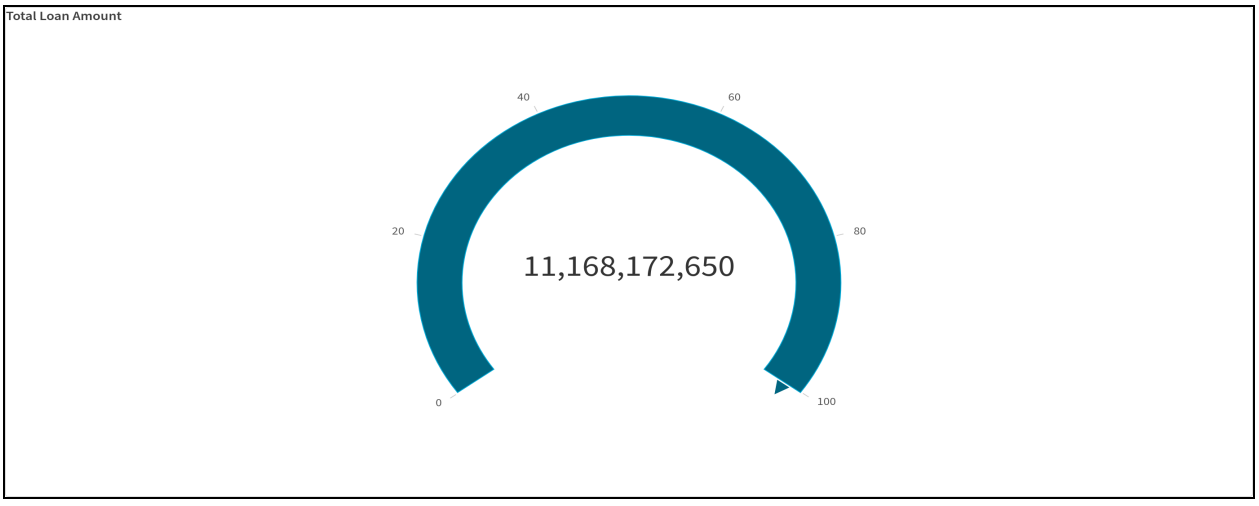
Average Loan Amount For The Term

Average Loan Amount For Joint/Individual Account



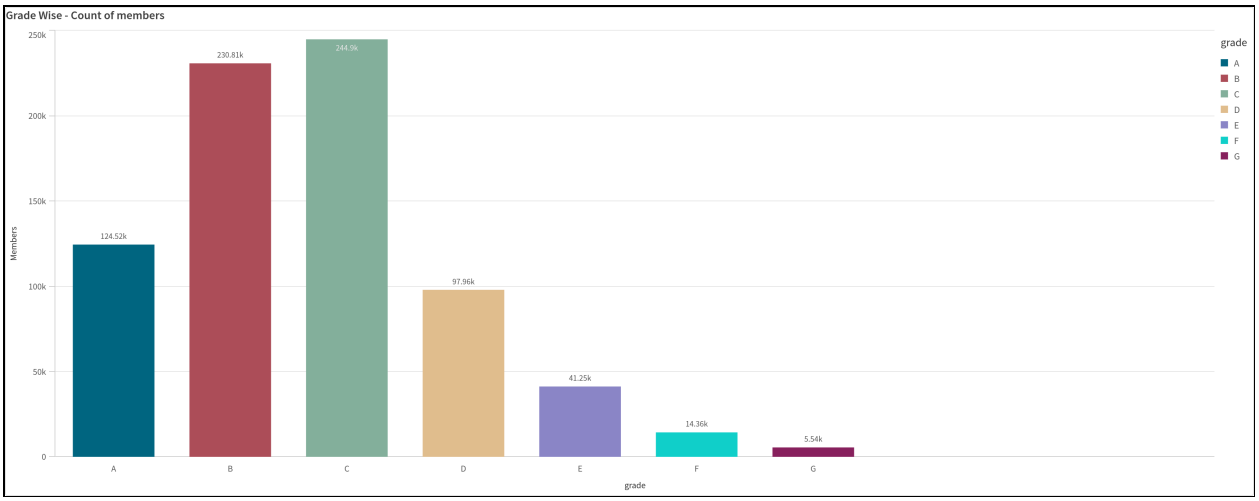
Total Number of Loan Account

Total Loan Amount

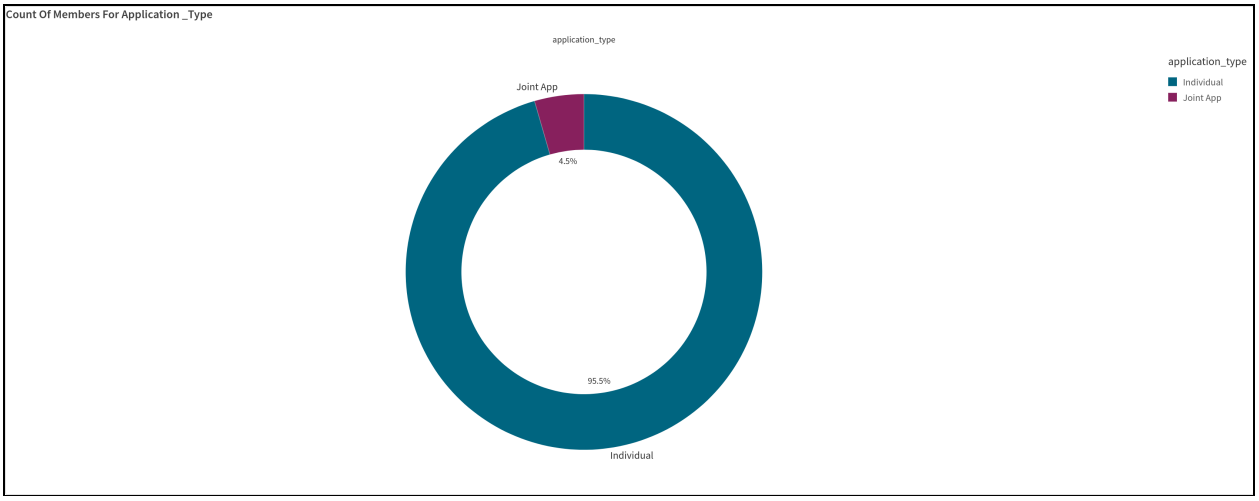


Average Loan Amount by State

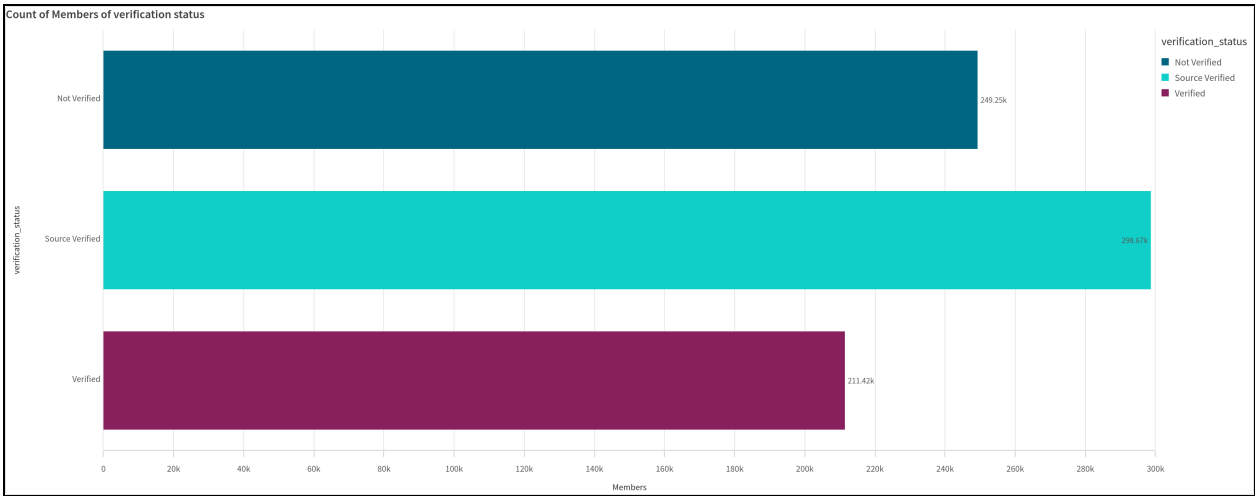
Grade wise - count of members



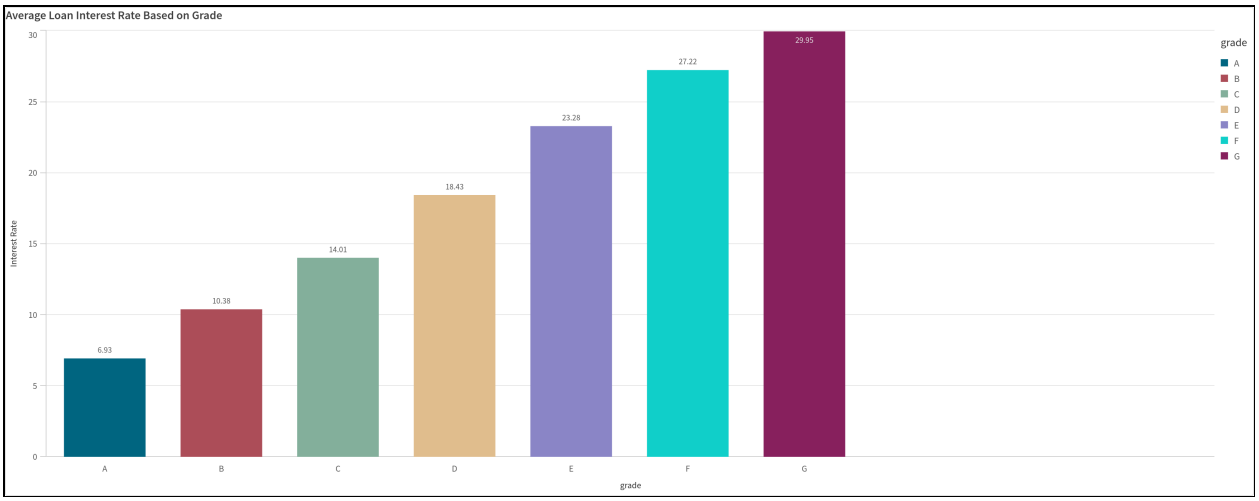
Count of Members for application_type



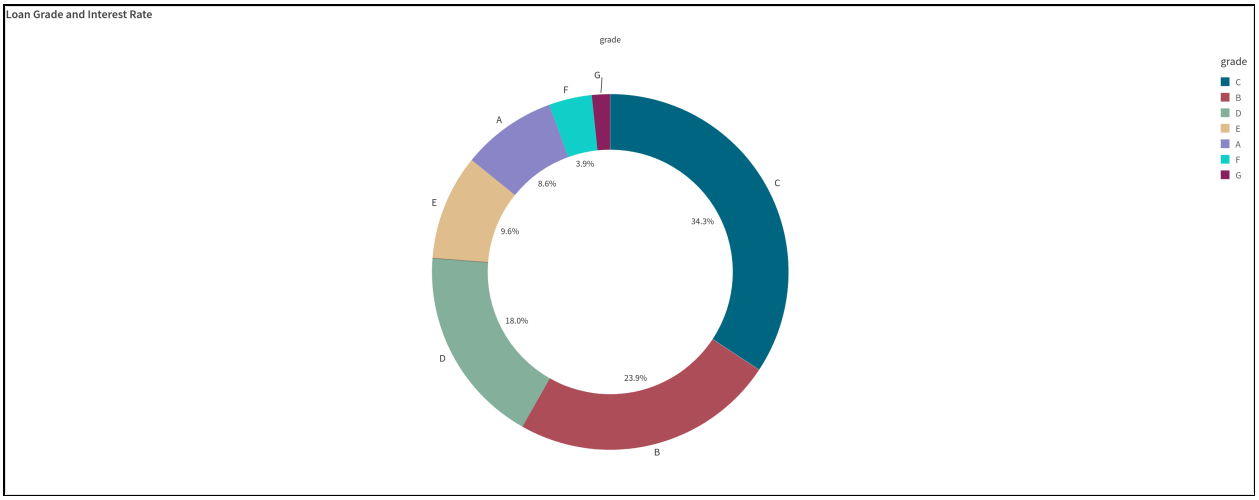
Count of members of verification status



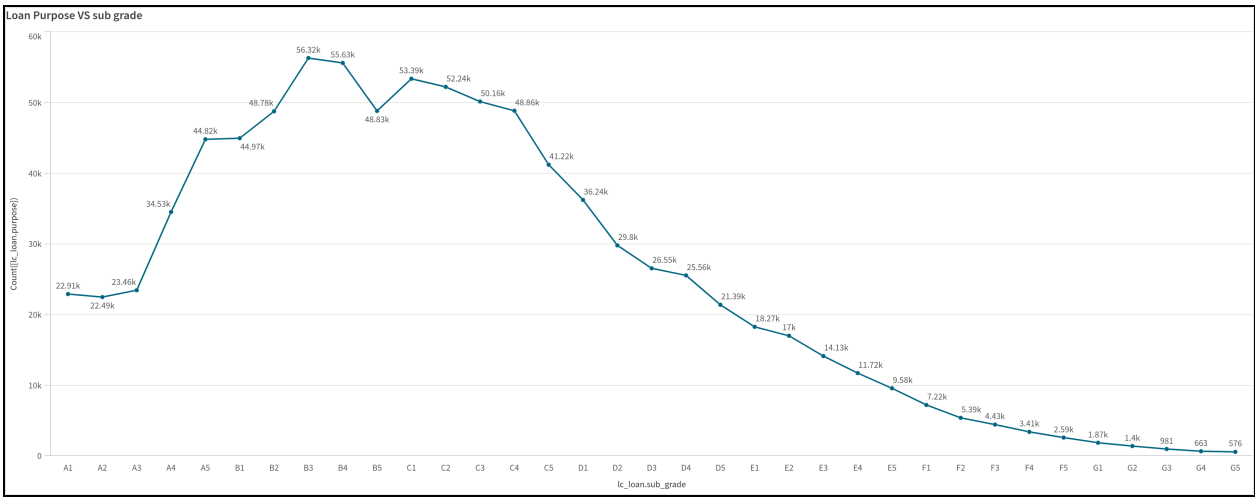
Average Loan Interest rate based on grade



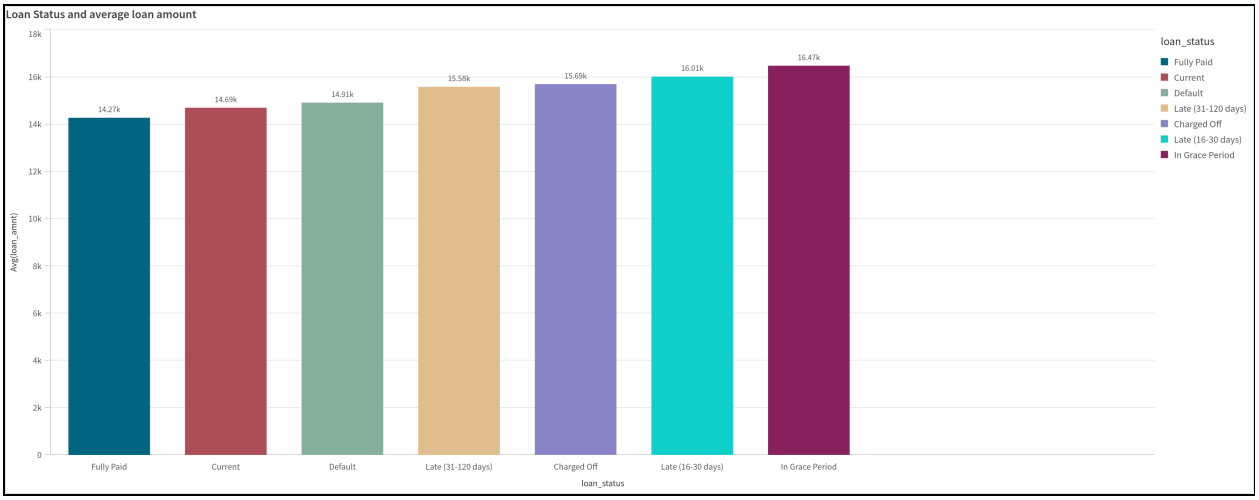
Loan grade and interest rate



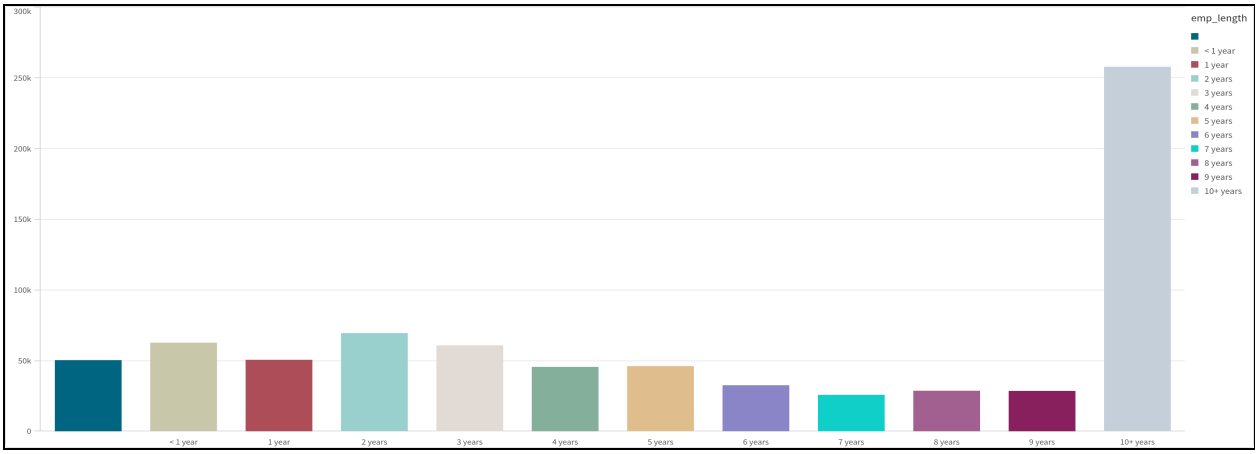
Loan purpose vs sub grade



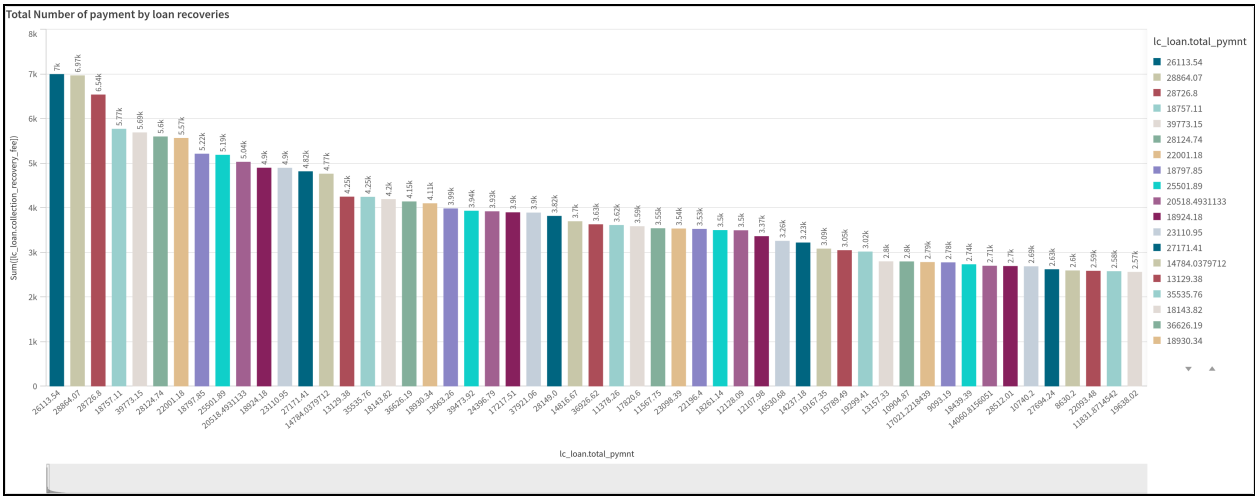
Loan status and average loan amount



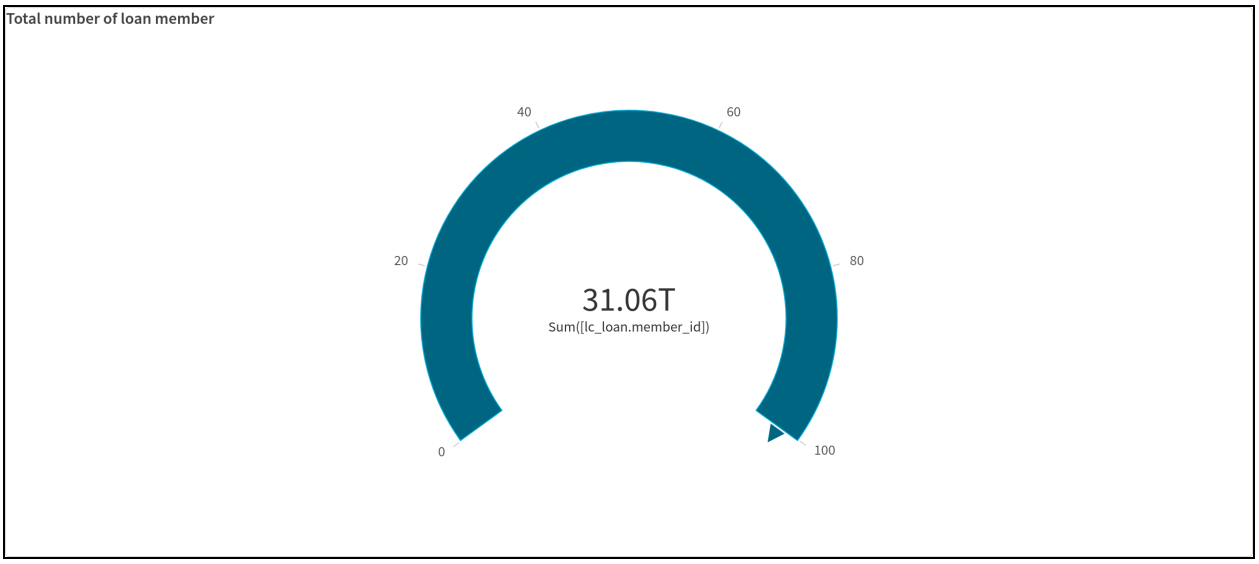
Employee length vs all util



Total number of payment by loan recoveries



Total number of loan member



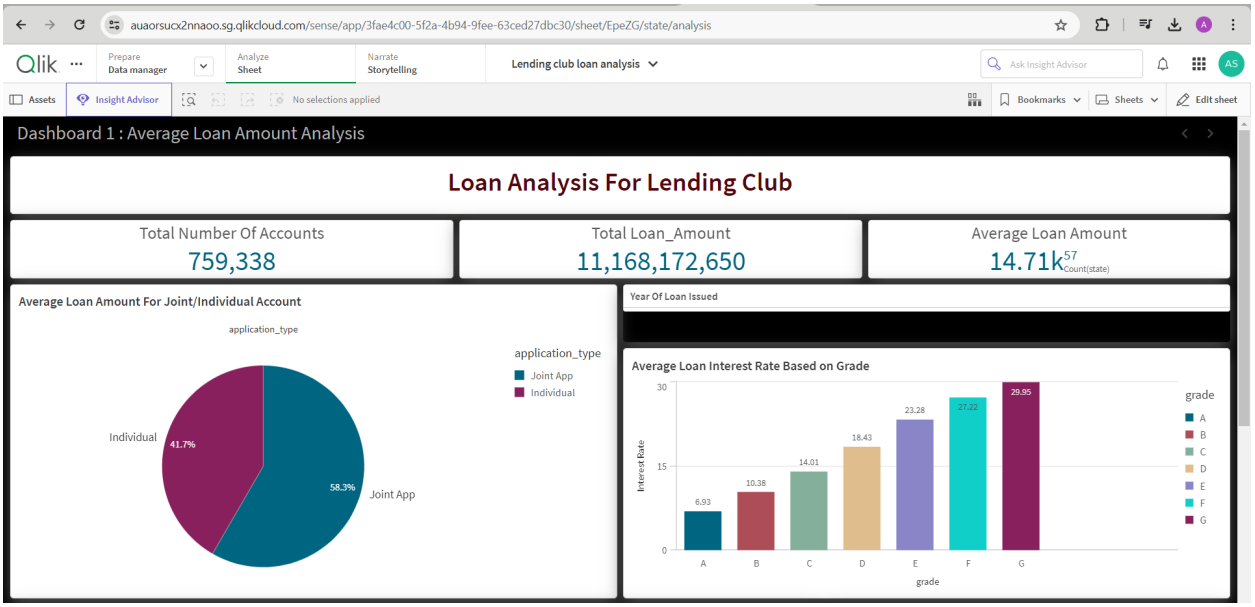
6. DASHBOARD

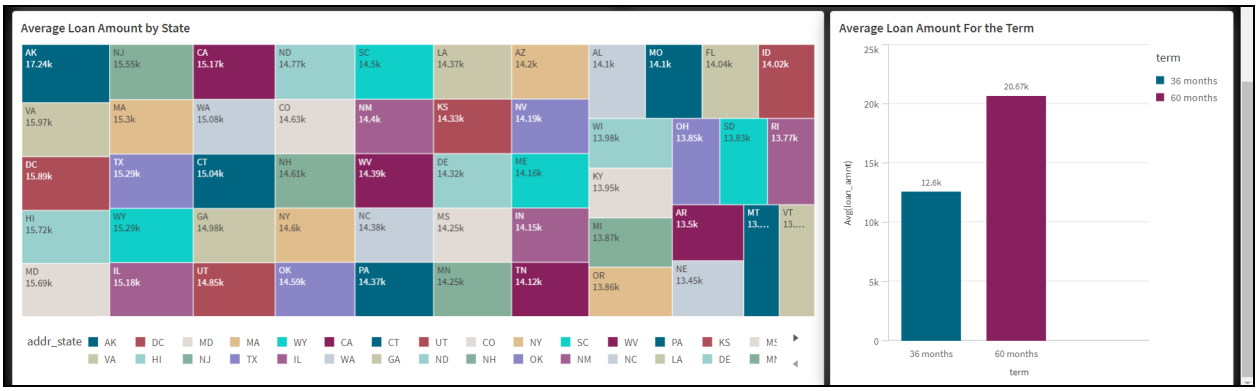
A Lending Club loan analysis dashboard is a visual representation of data related to loans issued by the Lending Club platform. It typically includes various mertics such as loan amount, interest rates, borrower information, loan status and other relevent data points. To ctreate a Lending club loan analysis dashboard we have follow these steps.

Data preparation, define metrics, create visualizations,add filters, Dashboard layout, interactivity, trend analysis and comparative analysis. By presenting the information through visualizations like charts and graphs.

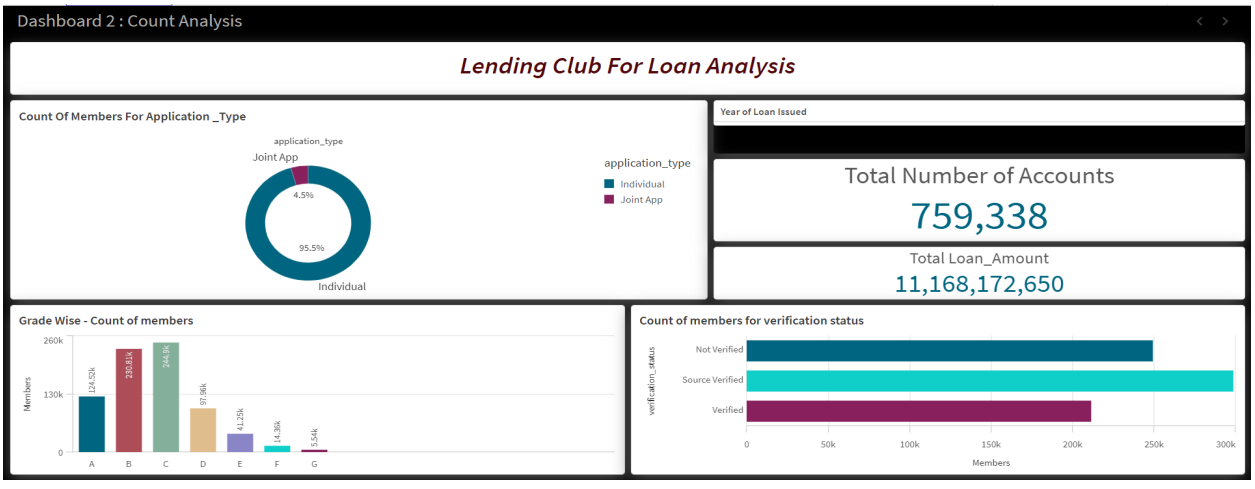
6.1 Responsive and Design of Dashboard

Dashboard 1 : Average loan Amount Analysis

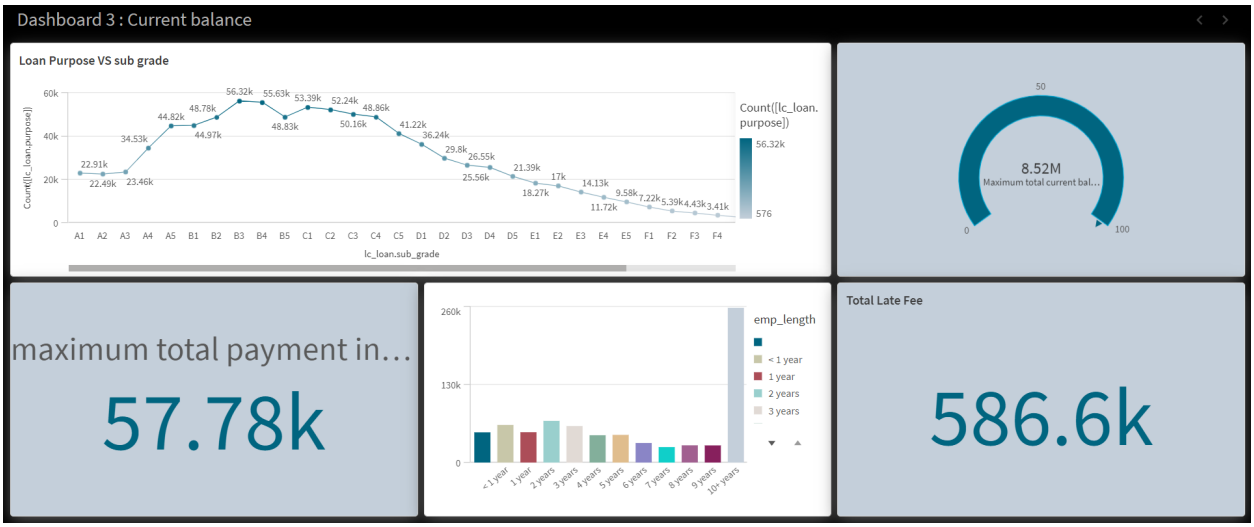




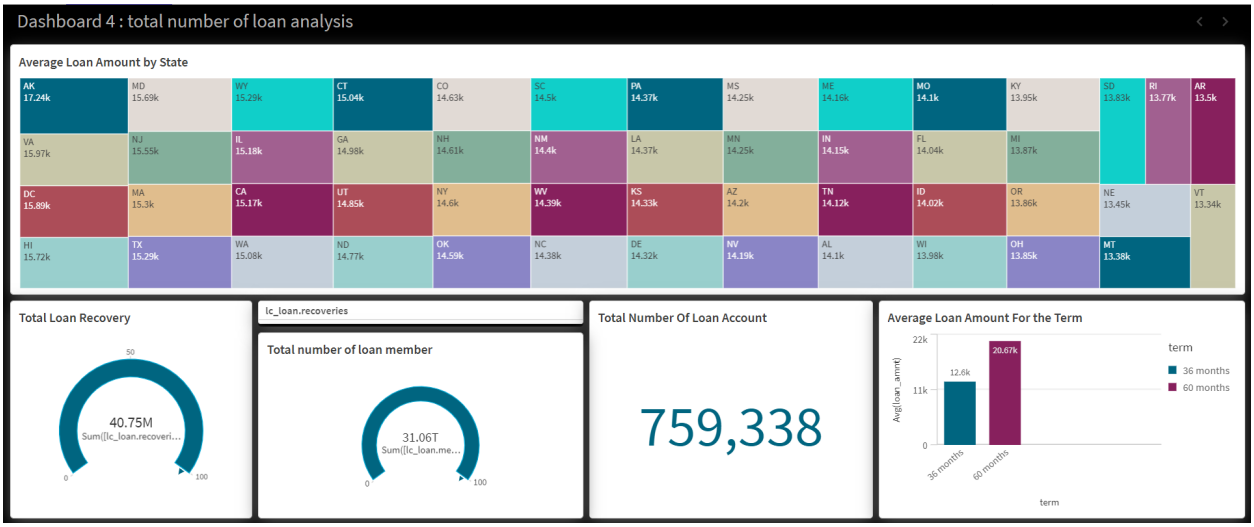
Dashboard 2 : Count Analysis



Dashboard 3 : Current balance



Dashboard 4 : Total number of loan analysis

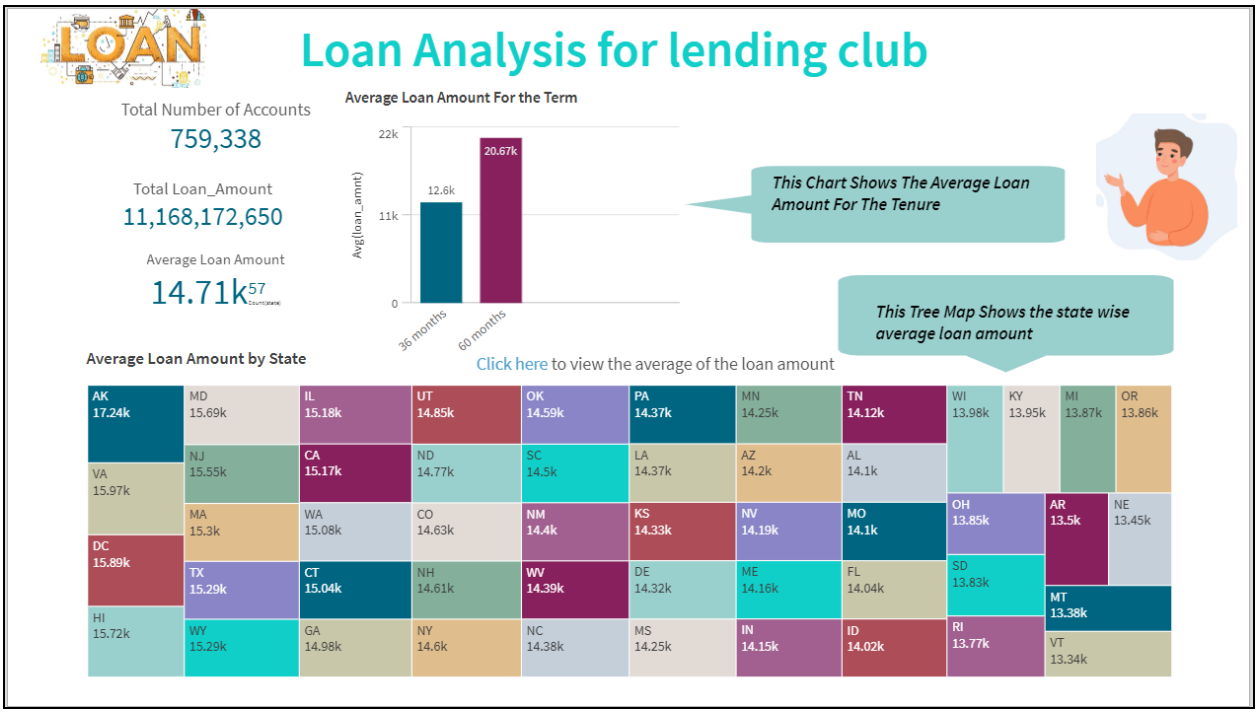


7. STORYTELLING

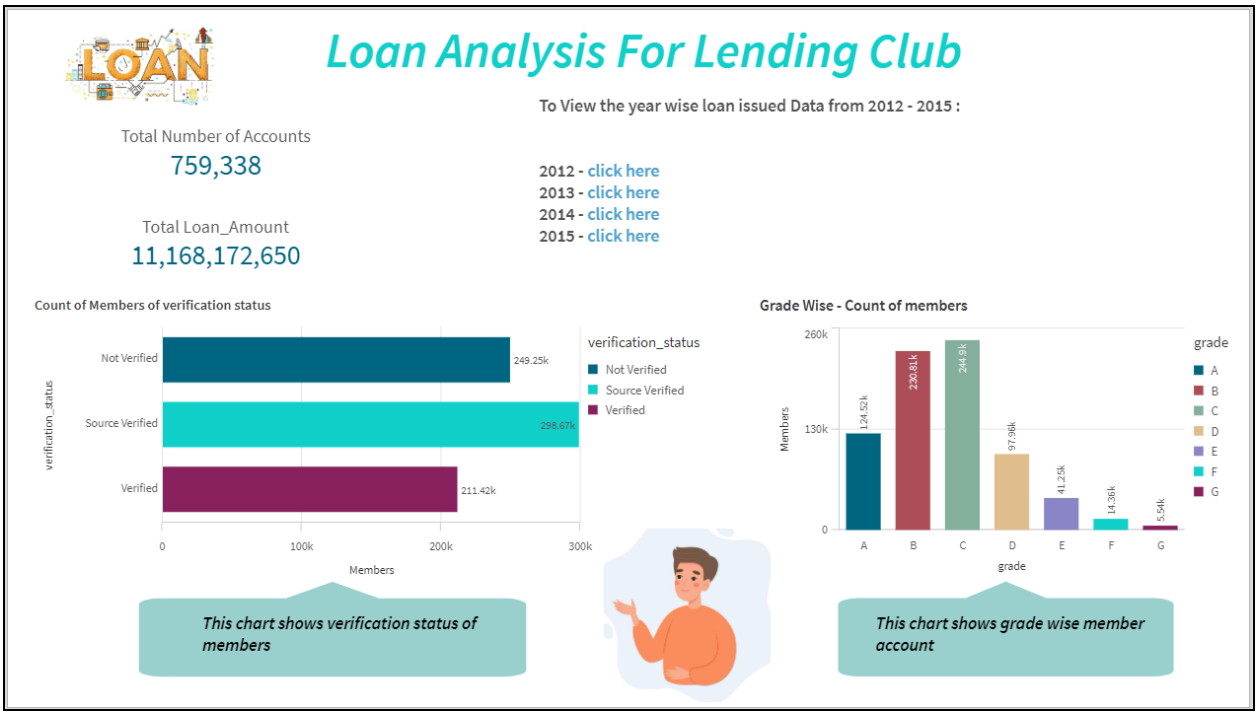
A data story is like telling a story with data and analysis. it helps make the information more interesting. We can start the present the data and analysis logically with conclusion summarizing the key findings. We can use reports, presentations, visuals to tell your data story. It should be organized in a logical and systematic way and analysis step by step. This often involves using visualizations, chart, graphs and other tools. It's like storytelling with numbers and facts to make the informationand understandable.

Data stories can be presented using various mediums including reports, presentations,interactive visualizations and videos.

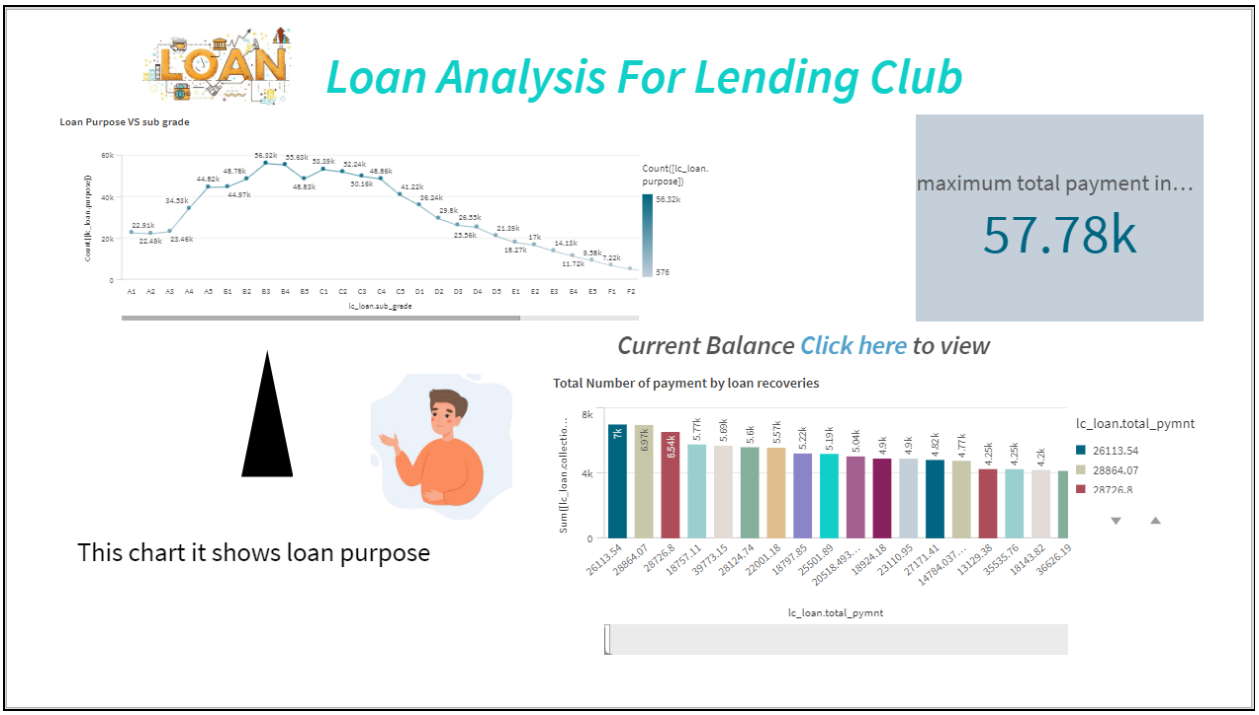
StoryTelling 1:



StoryTelling 2:



StoryTelling 3:

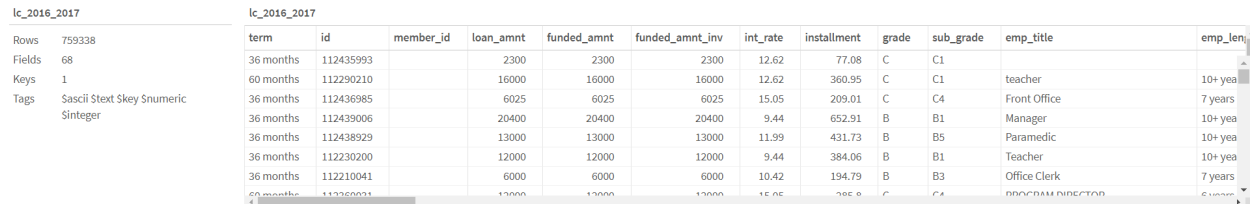


8. PERFORMANCE TESTING

8.1 Amount of Data Rendered

In the lending club loan analysis using the dataset downloaded from kaggle the amount of data loaded can vary depending on the specific datasest I downloaded. The dataset from kaggle contains a significant amount of information, including details about loans, borrowers, loan status, loan amount, interest rates, and more.

Amount of data loaded in the context of the Lending club loan analysis using the dataset from kaggle refers to the volume or quantity of data that has been imported from the datasset into analysis environment.



lc_2016_2017

Rows	759338
Fields	68
Keys	1
Tags	Sascii $text $key $numeric $integer

lc_2016_2017

term	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	installment	grade	sub_grade	emp_title	emp_len
36 months	112435993		2300	2300	2300	12.62	77.08	C	C1		
60 months	112290210		16000	16000	16000	12.62	360.95	C	C1	teacher	10+ yea
36 months	112436985		6025	6025	6025	15.05	209.01	C	C4	Front Office	7 years
36 months	112439006		20400	20400	20400	9.44	652.91	B	B1	Manager	10+ yea
36 months	112438929		13000	13000	13000	11.99	431.73	B	B5	Paramedic	10+ yea
36 months	112230200		12000	12000	12000	9.44	384.06	B	B1	Teacher	10+ yea
36 months	112210041		6000	6000	6000	10.42	194.79	B	B3	Office Clerk	7 years

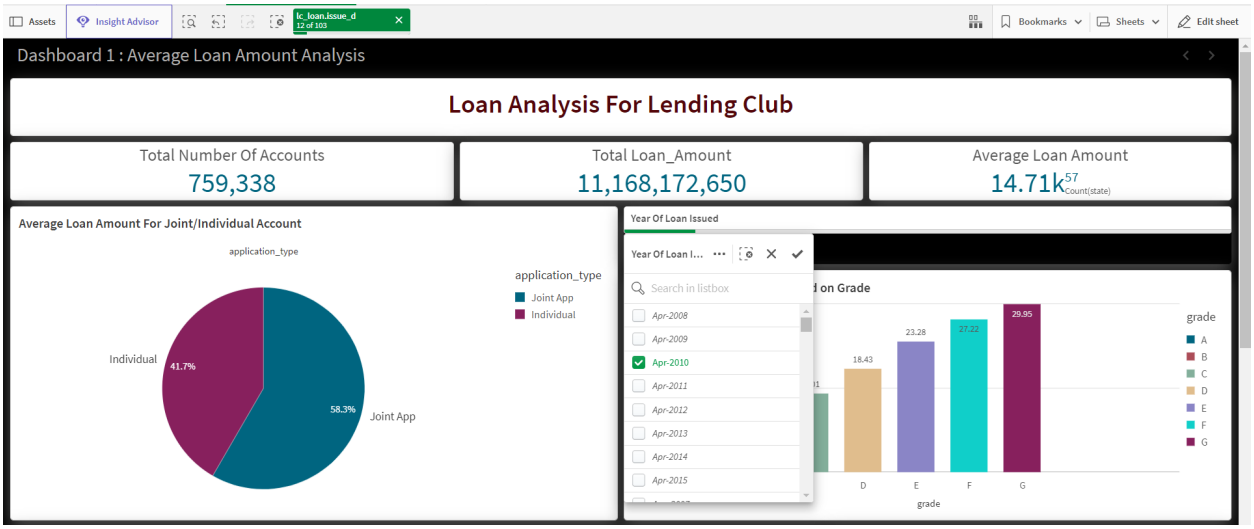
hom...	annu...	verifi...	issue_d	loan_...	pym...	purp...	title	zip_c...	addr...	dti	delin...	earli...	inq_l...
OWN	10000	Not Verified	Jun-2017	Current	n	credit_card	Credit card refina	148xx	NY	21.61	0	Sep-1985	
MORTGAGE	94000	Not Verified	Jun-2017	Current	n	debt_consolidati	Debt consolidatic	021xx	MA	25.61	0	Jun-1992	
MORTGAGE	46350	Not Verified	Jun-2017	Current	n	home_improvem	Home improvem	018xx	MA	8.88	0	Jun-2002	

Performance testing is crucial for ensuring that the "LendingClub Issued Loans Analysis" project meets its objectives of efficient, scalable, and reliable data processing and visualization. By systematically testing and optimizing each component of the system, you can ensure that the analysis and insights derived from LendingClub's loan data are delivered in a timely and efficient manner, supporting informed decision-making and strategic planning.

8.2 Utilization of Data Filters

Data filters are crucial for refining and analyzing large datasets by focusing on specific subsets of data. In the context of the LendingClub Issued Loans Analysis, data filters help isolate relevant information, streamline the analysis process, and enhance the accuracy of insights. Below is an overview of how data filters can be effectively utilized in this project

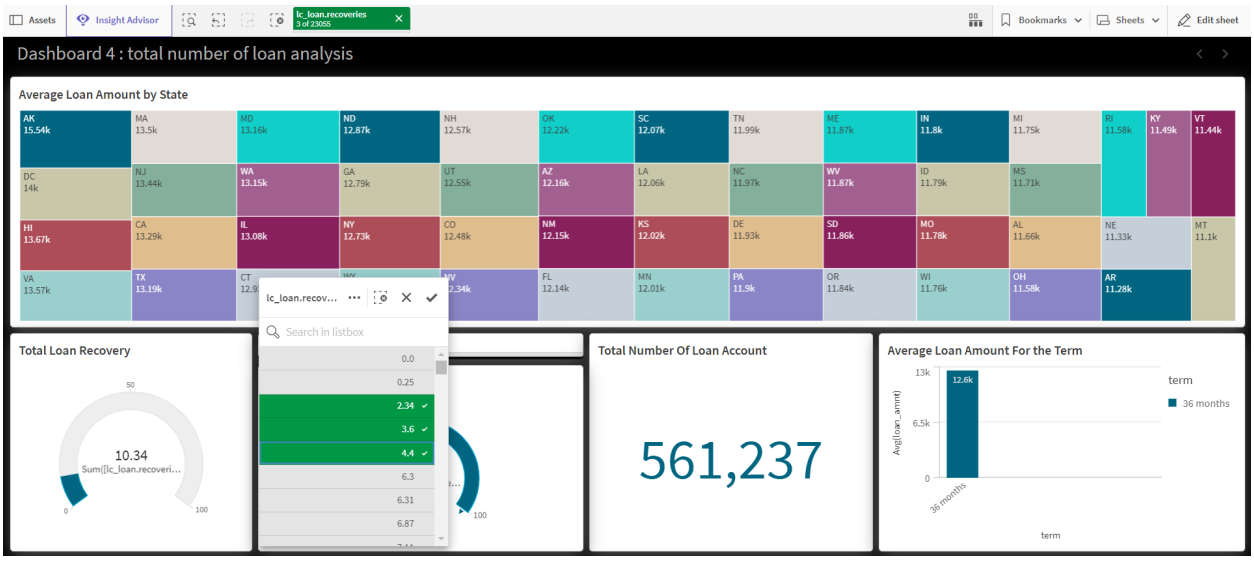
Filter 1:



In this Filter I have used "year of loan issued" which filter out the year I am looking for from the other data and thus it uses data more effectively

Filter 2:

In this Filter I have used "loan recoveries" this filter out the data and give me the data of loan recovery for a specific input I am looking for



Number of Visualization and graph

- Average Loan Amount For The Term
- Average Loan Amount For Joint/Individual Account
- Total Loan Amount
- Total Number of Loan Account
- Average Loan Amount by State
- Grade Wise - Count of members
- Count Of Members For Application _Type
- Count of Members of verification status
- Average Loan Amount
- Average Loan Interest Rate Based on Grade
- Loan Issued in 2012 - 2015
- Loan Grade and Interest rate
- Loan Purpose VS sub grade
- Maximum total current balance
- Loan Status and average loan amount
- Employee length vs all util
- Total Number of payment by loan recoveries
- Average Loan Amount
- Tenure wise average loan amount

Conclusion

The "Data to Decisions: Qlik Journey Through LendingClub Issued Loans Analysis" project exemplifies the transformative potential of data analytics in optimizing lending strategies and risk management for peer-to-peer lending platforms. By leveraging Qlik's advanced capabilities for data integration, analysis, and visualization, the project successfully addressed core business challenges and enhanced decision-making processes. Through comprehensive ETL processes and effective data filters, we ensured clean, reliable datasets and focused analysis, enabling targeted insights and improving efficiency. Advanced predictive models enhanced risk assessment, accurately identifying high-risk borrowers and predicting loan default rates, thereby optimizing lending criteria. Qlik's dynamic, interactive dashboards facilitated real-time data visualization, supporting quick decision-making and adaptability to market changes. Performance testing confirmed the system's scalability and reliability, ensuring it can handle increasing data volumes and user loads efficiently. This project not only improved operational efficiency and strategic decision support but also laid a robust foundation for continuous innovation in data analytics, positioning LendingClub for sustained growth and success in the competitive peer-to-peer lending market.