Data to Decisions Qlik Journey through LendingClub Issued Loans Analysis (Qlik)

1 Introduction

1.1 Overview Of The Project

The primary problem facing the company is that there is no data analysis of LendingClub loan data in the existing lending approach. Due to this, the institution finds it difficult to comprehend borrower patterns and market trends, which causes issues with risk assessment, makes it difficult to predict loan defaults, and makes it difficult to modify lending criteria in response to shifting market conditions.To meet the business needs, a robust data analytics system that can extract meaningful insights from the loan data provided by LendingClub must be developed. By employing this framework, the financial institution ought to be capable of recognizing borrower groups that pose a high risk, accurately projecting default rates, and establishing the foundation for promptly adjusting lending conditions.

1.2 Purpose Of The Project

The particular business issue is that the existing lending strategy is inadequate and not well-informed by thorough analysis of loan data from LendingClub. The organization has trouble accurately identifying risk, anticipating loan default rates, and dynamically adjusting lending criteria in response to changing market conditions because it finds it difficult to evaluate borrower behavior and market dynamics. The creation of a strong data analytics system capable of deriving significant insights from loan data given by LendingClub is necessary to meet the business requirements. With the use of this framework, the financial institution should be able to identify high-risk borrower categories, properly anticipate default rates, and get the basis in place for making real-time modifications to lending criteria. To guarantee a smooth installation, the solution should also be flexible, scalable, and able to integrate with current systems.

Lending Club may find that a data analysis methodology is an effective tool for achieving its objectives. Both investors and borrowers can gain from the platform by gaining important insights, organizing data better, and expediting access to it. Nonetheless, in order to preserve confidence within the Lending Club ecosystem, it's imperative to address any privacy and security problems.

2 Define Problem/ Problem Understanding

2.1 Specify The Business Problem

The specific business problem is that, despite careful examination of loan data from LendingClub, the current lending strategy is insufficient and ill-informed. The firm finds it challenging to assess borrower behavior and market dynamics, which makes it difficult to effectively identify risk, predict loan default rates, and dynamically alter lending criteria in response to shifting market conditions.

2.2 Business Requirements

Building a strong data analytics system capable of deriving valuable insights from loan data given by LendingClub is one of the business requirements. Through the use of this framework, the financial institution should be able to identify high-risk borrower groups, properly estimate default rates, and lay the groundwork for lending criterion modifications that may be made in real time. To enable smooth installation, the solution should also be flexible, scalable, and able to integrate with current systems.

2.3 Literature Survey

In the context of peer-to-peer lending platforms like LendingClub, a thorough literature review is necessary to comprehend current approaches, instruments, and best practices in risk management and lending strategy optimization. In order to improve financial institutions' decision-making processes, this survey seeks to find pertinent papers on peer-to-peer lending trends, data analytics in finance, and related analysis.

3 Data Collection

3.1 Collect the DataSet

The dataset was collected from the kaggle website.Data provides all of the meta information about the columns given in the CSV files. The data file named lc_2016_2017.csv

Column Description of the Dataset:

member_id: Contains unique member id of the members

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

3.2 Connect the data with Qlik Sense

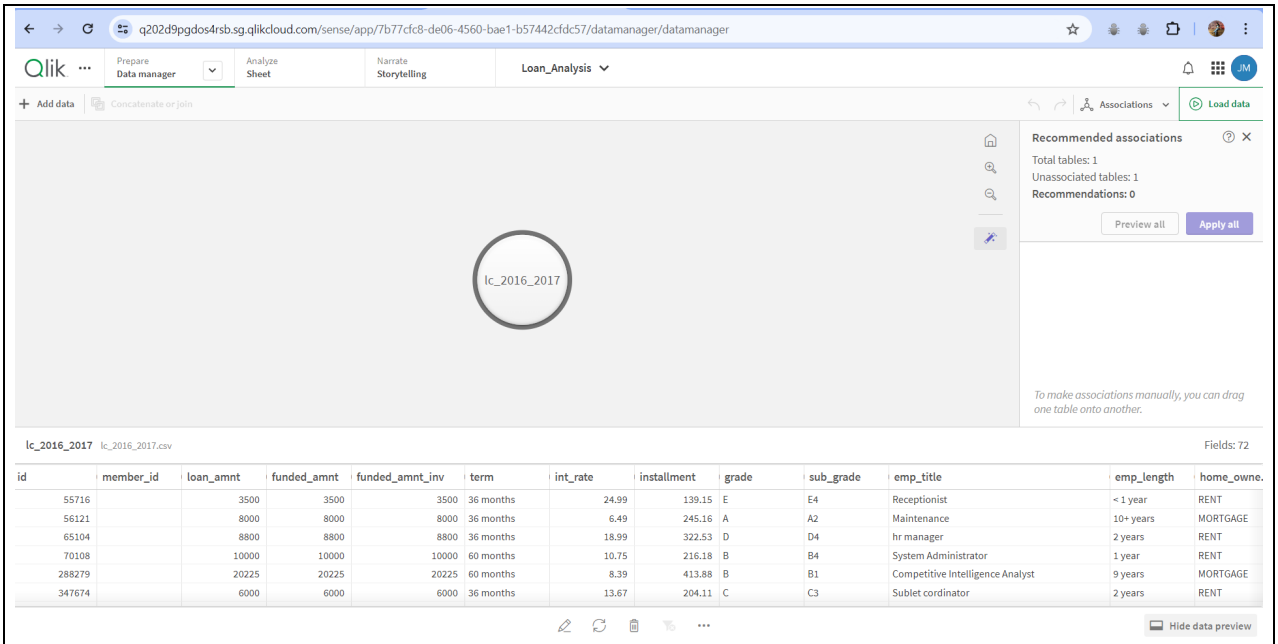
We must create a new analytics app, launch it in app preview, and access our Qlik account. Click on "file" and select "other sources." We must then upload the files from our local computer. Data files will then display.

4 Data Preparation

4.1 Prepare The data for the Visualisation

Data must be cleaned to eliminate unnecessary or incomplete information. It then needs to be formatted to facilitate visualization. Finally, it needs to be examined for patterns and trends, filtered to focus on specific data subsets, prepared for visualization software, and confirmed to be accurate and comprehensive. This procedure improves the data's readability and gets it ready for the development of visualizations that will reveal performance and efficiency trends. Having now cleaned the data, we can go on to the visualization phase.

Select the one you want to add from the list of data files we have, click "next," and our data files will be uploaded. To see the data we uploaded navigate to the data manager and we will see the result as below.



5 Data Visualisations

5.1 Visualisations

Bar Chart Visualisations

The bar chart is suitable for comparing multiple values. The dimension axis shows the category items that are compared, and the measure axis shows the value for each category item. The average loan amount taken out by members for a given tenure, such as 36 or 60 months, is displayed in this visualization. The performance and efficiency of banks can be examined using a variety of popular visualizations, such as scatter plots, pie charts, bar charts, line charts, heat maps, and more. Here we use bar charts for average loan amount for the term and Grade Wise-count of members.Figure 5(a), the dimension values are different terms: 36 months, 60 months. Each term represents a dimension value, and has a corresponding bar. The bar height corresponds to the measure value (Average(loan_amnt)) for the different grades.

Figure 5(a)-Bar chart that represents average loan amount for the term

Figure 5(b) represents, the dimension values are different grades: A, B, C, D, E, F, G. Each grade represents a dimension value, and has a corresponding bar. The bar height corresponds to the measure value (count of members) for the different grades.

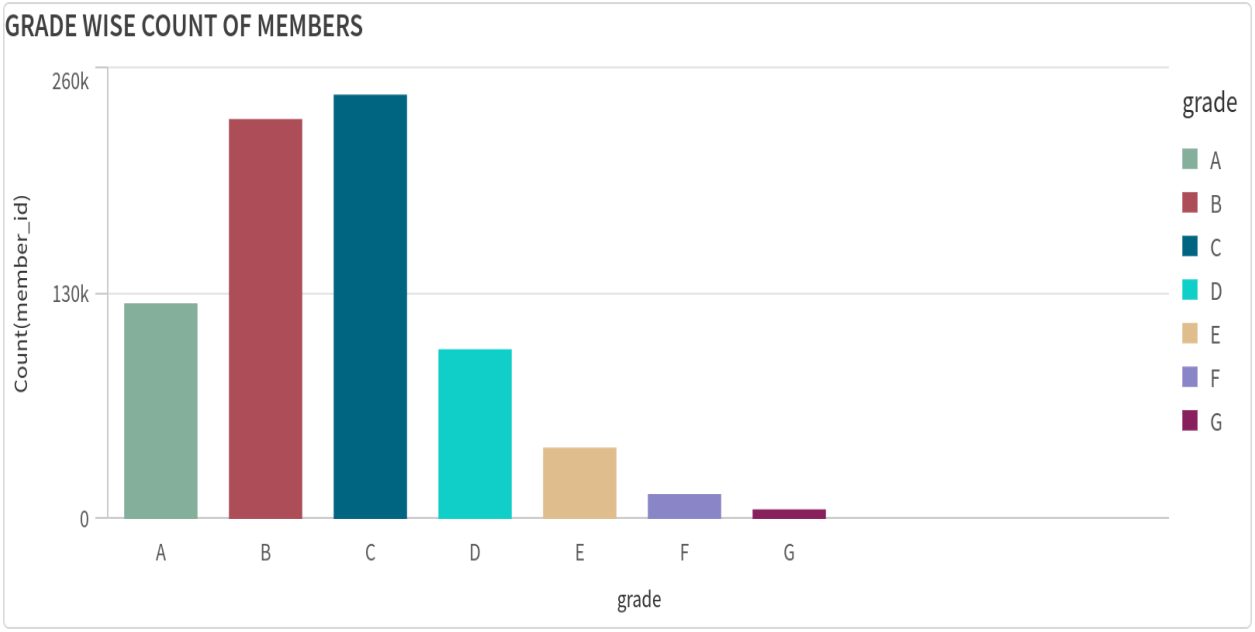


Figure 5(b)- Grade Wise-Count of Members

Pie Chart Visualisation

This type of visualisations organize and show data as a percentage of a whole. True to the name, this kind of visualization uses a circle to represent the whole, and slices of that circle, or “pie”, to represent the specific categories that compose the whole. The pie chart represents the average loan amount for the type applications i.e., individual or joint. Figure 5(c) represents i.e., individual account types are 41.7 percent and joint account types are 58.3 percent.

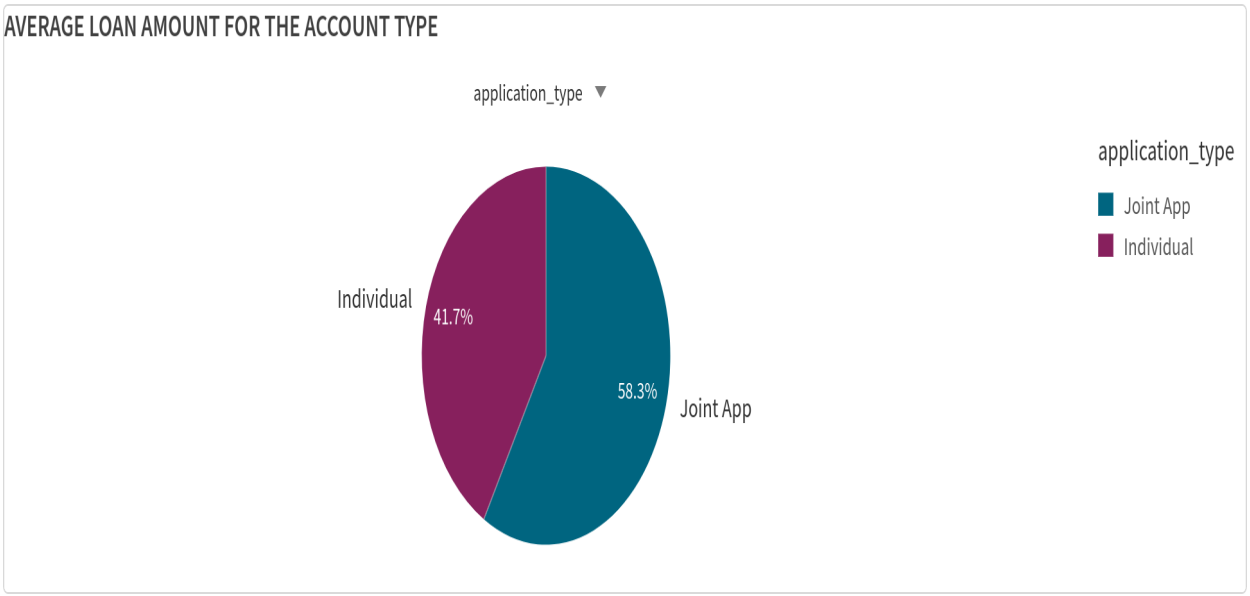


Figure 5(c)- Average loan amount for the account type

KPI visualisations

The KPI visualization can show one or two measure values, and is used to track performance. KPIs give a quick understanding of the performance within an area.Use KPIs to get an overview of performance values that are central to an organization. Use color coding and symbols to indicate how the figures relate to the expected results. Here Figure 5(d) & 5(e) represents two KPIs to measure total loan amount and total number of accounts.Here the total loan amount is 11.17G and total number of loan accounts are 759.3k. The KPI is somewhat limited when it comes to graphical components. You can use symbols to help illustrate the performance, but if you want a more conspicuous component, consider using a gauge.

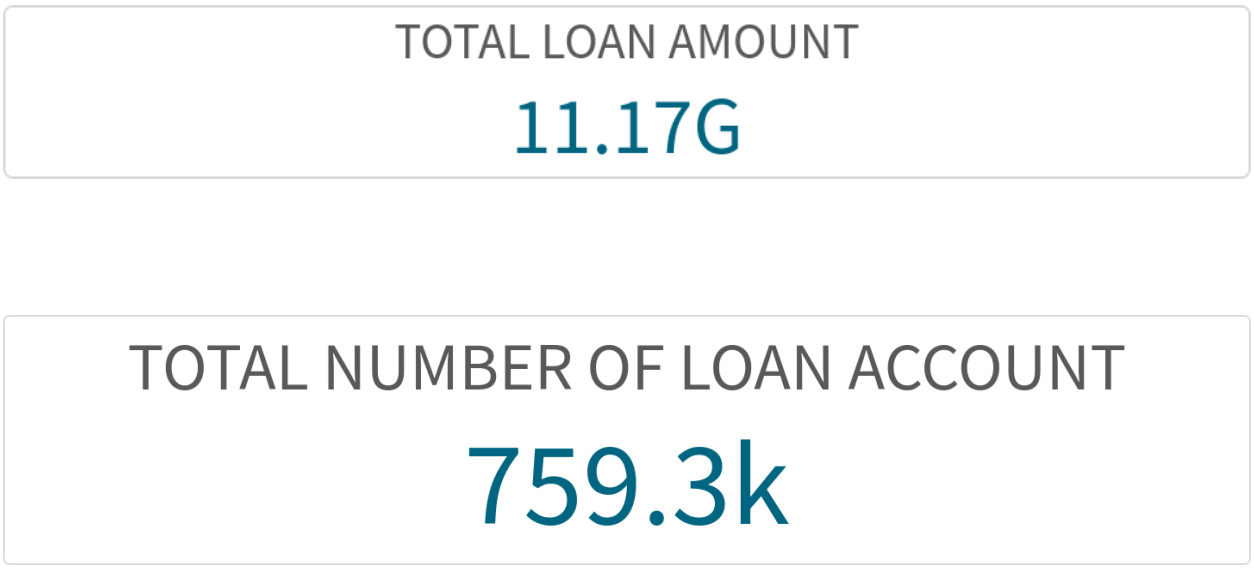


Figure 5(d) & 5(e)-Total loan amount and Total number of accounts

Tree Map Visualisations

The treemap visualization has got its name from the fact that data is referred to and handled as tree parts i.e. branches, nodes, leaves. It is used to display hierarchical data i.e. data that is present in many layers of logic. The hierarchical data is represented in nested rectangles or boxes i.e. small boxes within large boxes.A treemap is the best choice to make when you have loads of hierarchical data and not that much space on the sheet. Treemaps save a lot of space on the sheet as it can contain as many data values as it can in the form of small to smallest boxes (nested rectangles). The color scheming and size grading reveals the highs and lows of the situation and tells you at an instant which product is doing good and which needs more attention and strategy. The one major drawback of a treemap chart is that it does not show negative values.Here we use treemap for showing statewise average laon amount.

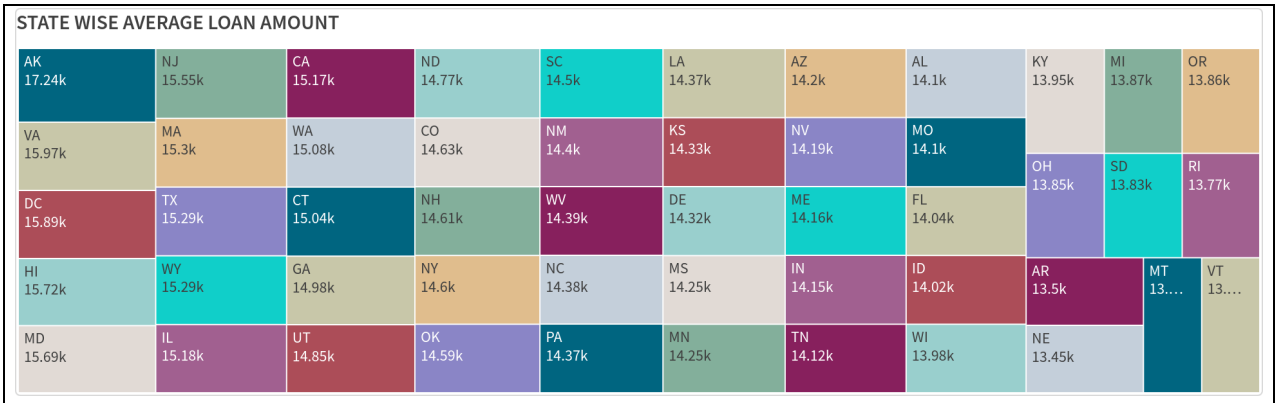


Figure 5(f)-State wise average loan amount

Bullet Chart Visualisation

The bullet chart displays a gauge with extended options. Bullet charts can be used to visualize and compare performance of a measure to a target value and to a qualitative scale, such as poor, average, and good.Bullet charts let you compare and measure performance with more enriched information than a common gauge. This is helpful when comparing performance according to a target and a simple performance rating. For example: you can show how sales relate to a target value, and in context of poor, good, and stretched performance.

Figure 5(g) represents bullet chart visualisation, the dimension values are different grades: A, B, C, D, E, F, G. Each grade represents a dimension value, and has a corresponding bar. The bar height corresponds to the measure value (count of members) for the different grades.

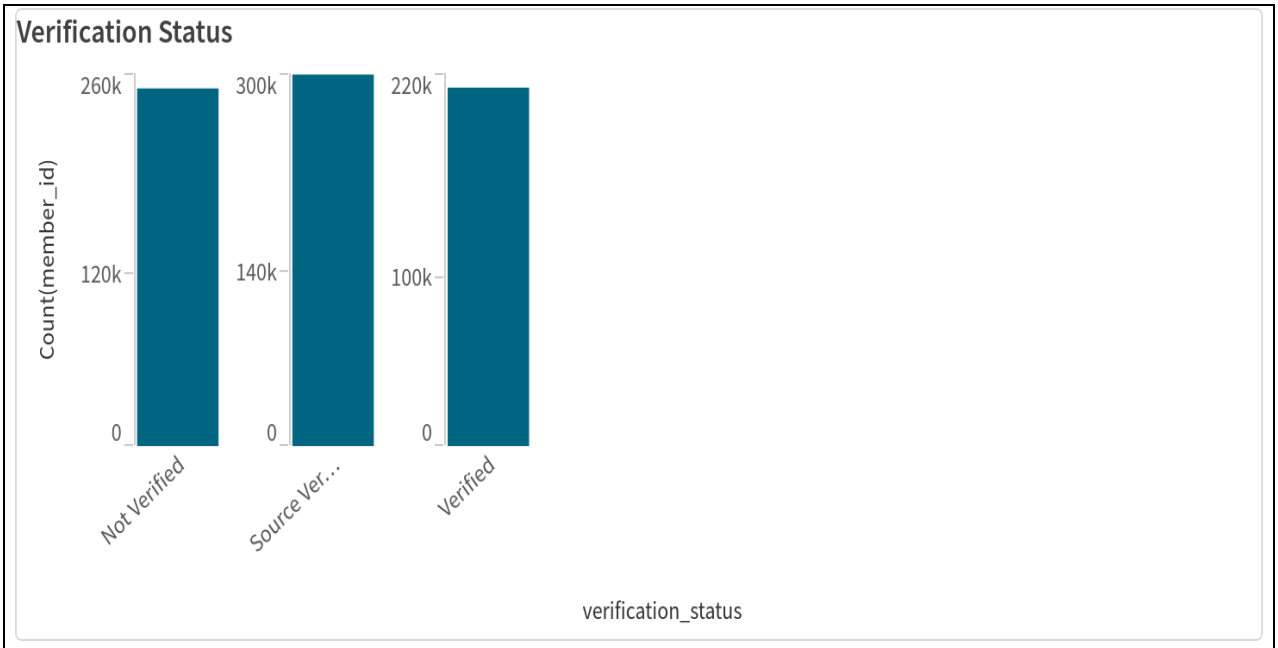
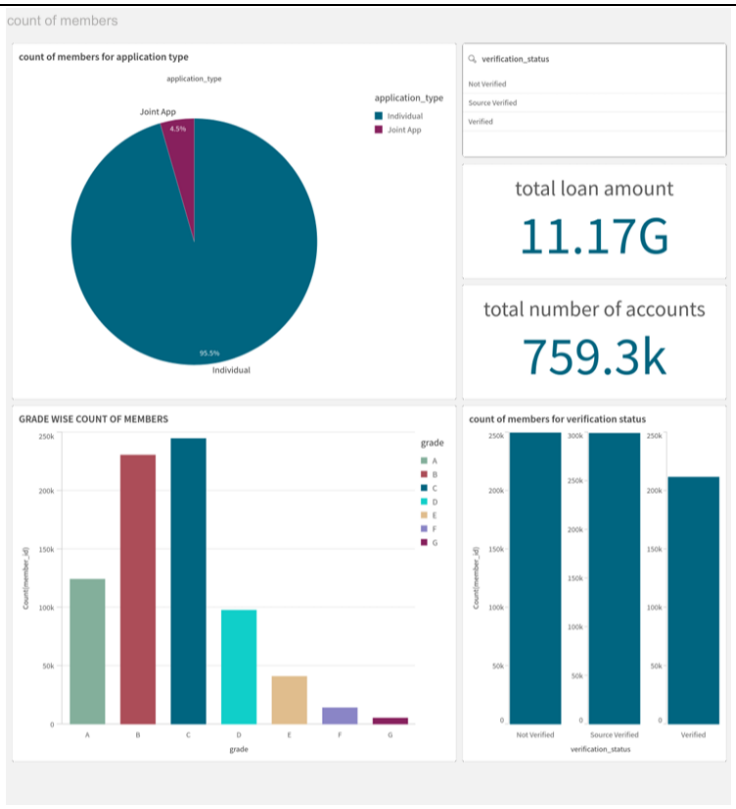


Figure 5(g)-Verification Status of members

6 Dashboard

6.1 Responsive and Design Of Dashboard

A responsive dashboard is a tool that gives consumers easy access to crucial data and is made to change its appearance and features according to the type of device it's being viewed on (desktop, tablet, mobile). This guarantees a positive user experience on all screen sizes. With Qlik Sense Cloud's integrated responsive design features, you can make dashboards that change automatically to fit various screen sizes. This gets rid of the need to create several versions for tablets, smartphones, and desktop computers.



LOAN ANALYSIS FOR LENDING CLUB

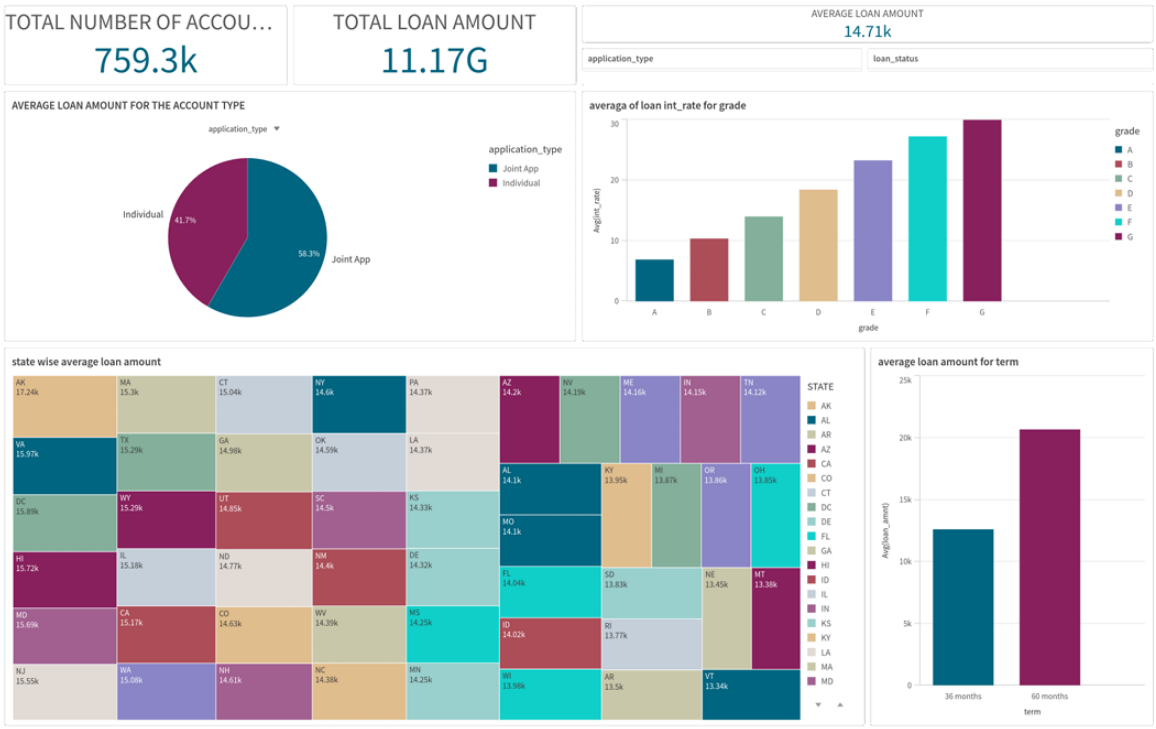


Figure 6(a) & 6(b)-count members & Loan Analysis for Lending club

Figure 6(a) & 6(b) represents two dashboards that tells about loan analysis for lending club and count of members for specific fields

7 Report

7.1 Report Creation

We employ storytelling in our report, and we create story telling slides with a snapshot of each visualization we created. We put them in order, use shapes to explain them, and create a PowerPoint.Figure 7(a) & 7(b) represents the reports for the data.You can take snapshots of selected visualizations and use them in your narrative together with text, shapes, and effects. You create slides and design the story with your particular audience in mind. In your narrative you focus on key elements and create a convincing story to make your message clear.

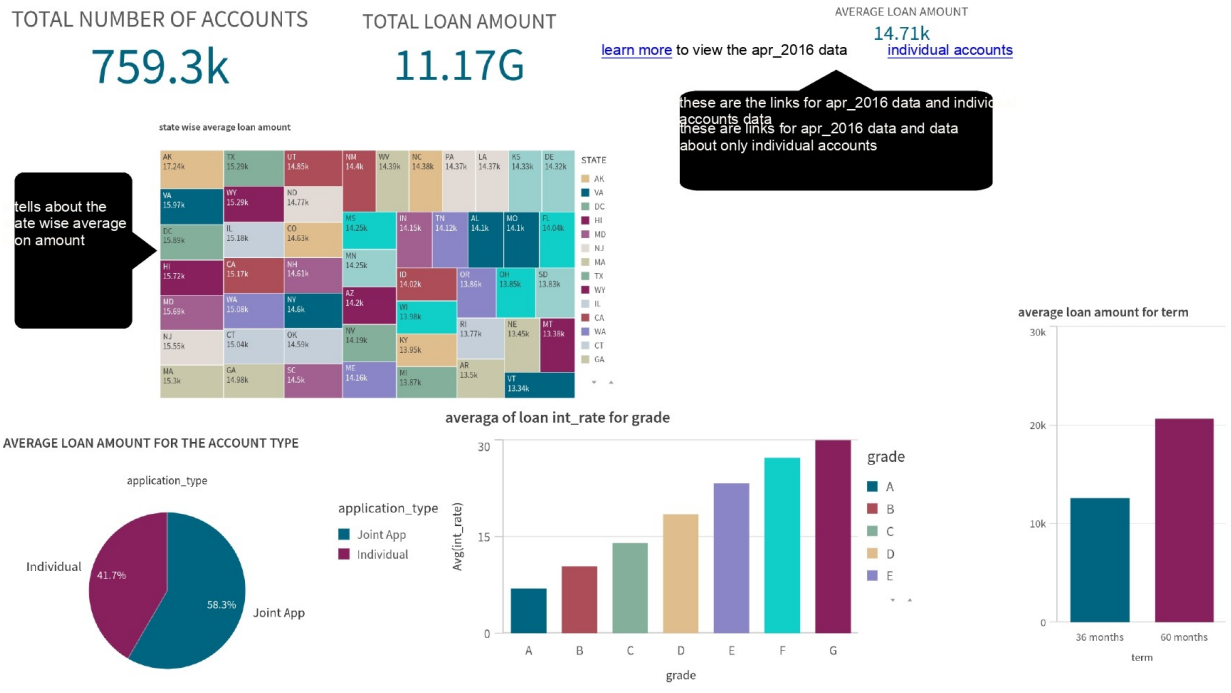


Figure 7(a)

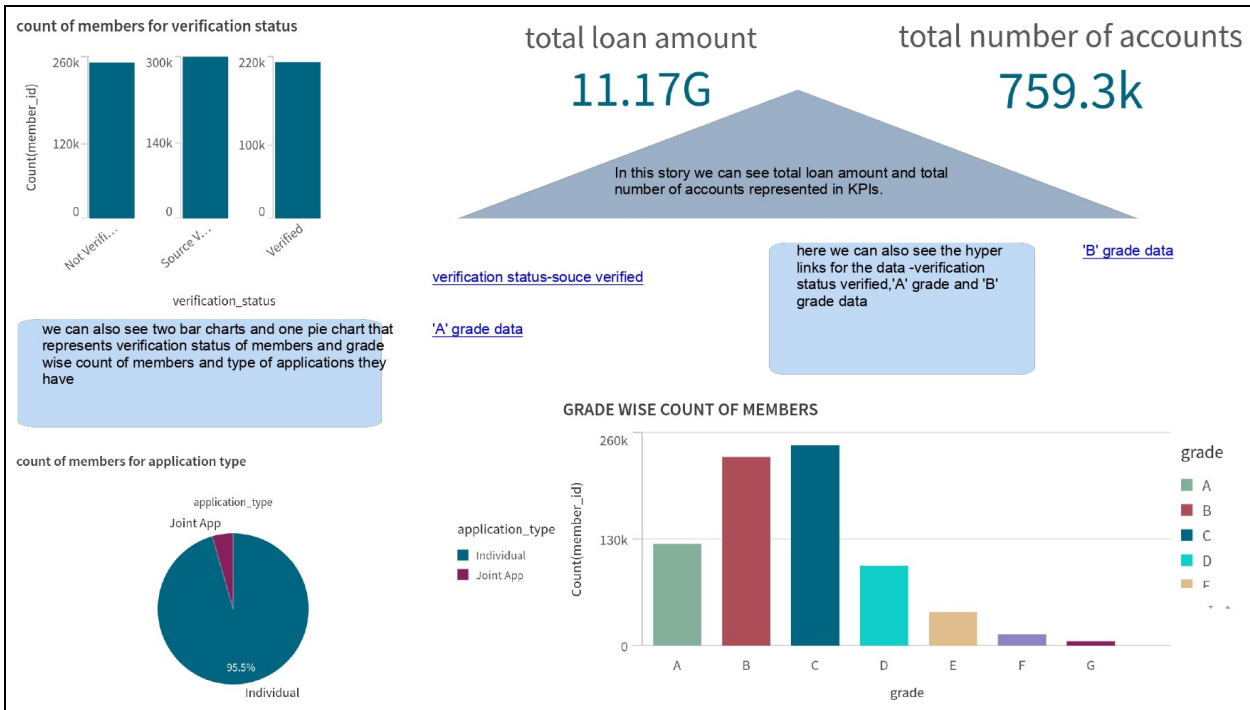


Figure 7(b)

8 Performance Testing

8.1 Amount Of Data Rendered

The volume or quantity of data that has been loaded into a system, software program, database, or any other data processing or storage environment is referred to as the "Amount of Data Loaded". This metric shows how much data has been effectively processed and is available for usage, alteration, or analysis within the system. It highlights the system's ability to handle and make data available for later operations and decision-making processes. It includes all forms of data that have been transmitted into the target environment from multiple sources.

8.2 Utilization of Filters

"Utilization of Filters" describes the purposeful placement, operation, or utilization of filtering mechanisms in a pipeline, software program, or system to selectively extract, modify, or analyze data according to predetermined standards or conditions. In this method, specified parameters are made to identify the important data

points that should be included in the analysis and the ones that shouldn't be, along with the criteria that must be met.

By limiting the scope of the data, filters are essential for managing enormous datasets and improving the efficacy and efficiency of data processing and analysis. Filters aid in the reduction of data noise, enhancement of insight clarity, and facilitation of more accurate decision-making by concentrating solely on pertinent information that satisfies predetermined criteria. Dynamic filter application ensures that the data is usable and relevant throughout the analysis process by letting users modify criteria in response to changing needs. In a variety of domains, such as business intelligence, data science, and analytics, where the capacity to promptly and precisely identify useful information can have a substantial impact on results and performance, this targeted approach to data management is indispensable.

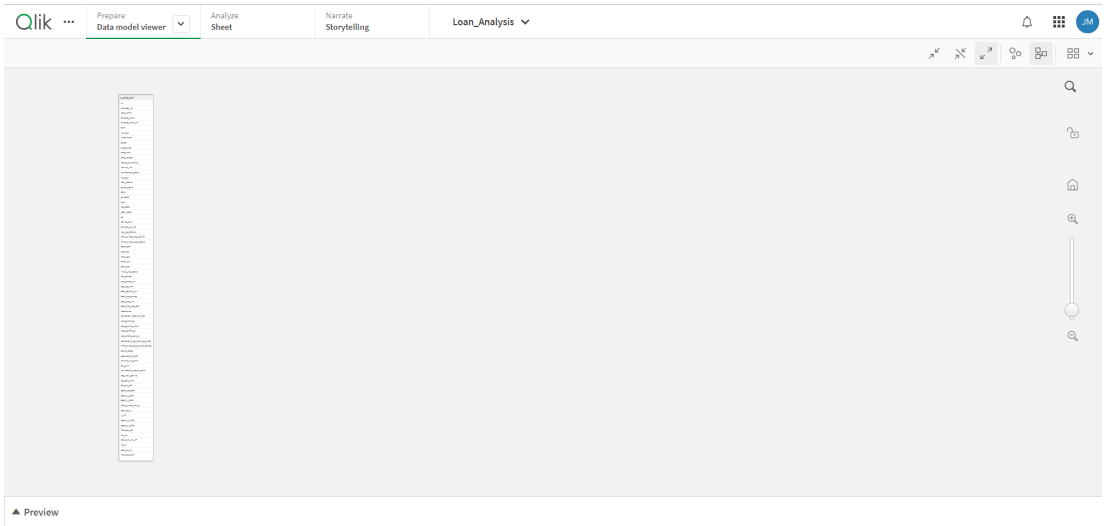


Figure 8(a)-Amount of data loaded

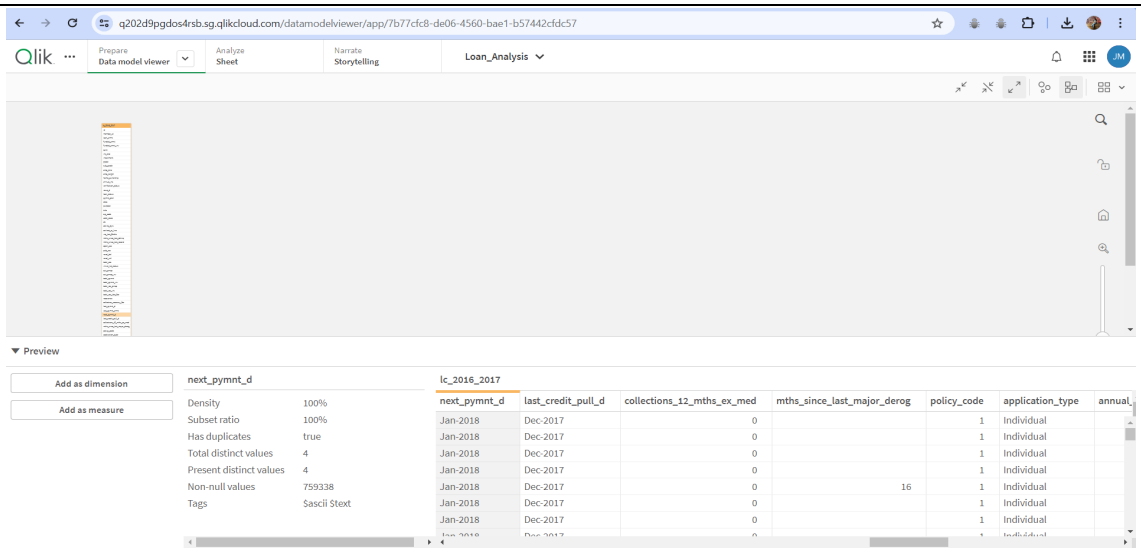


Figure 8(b)-Amount of data loaded

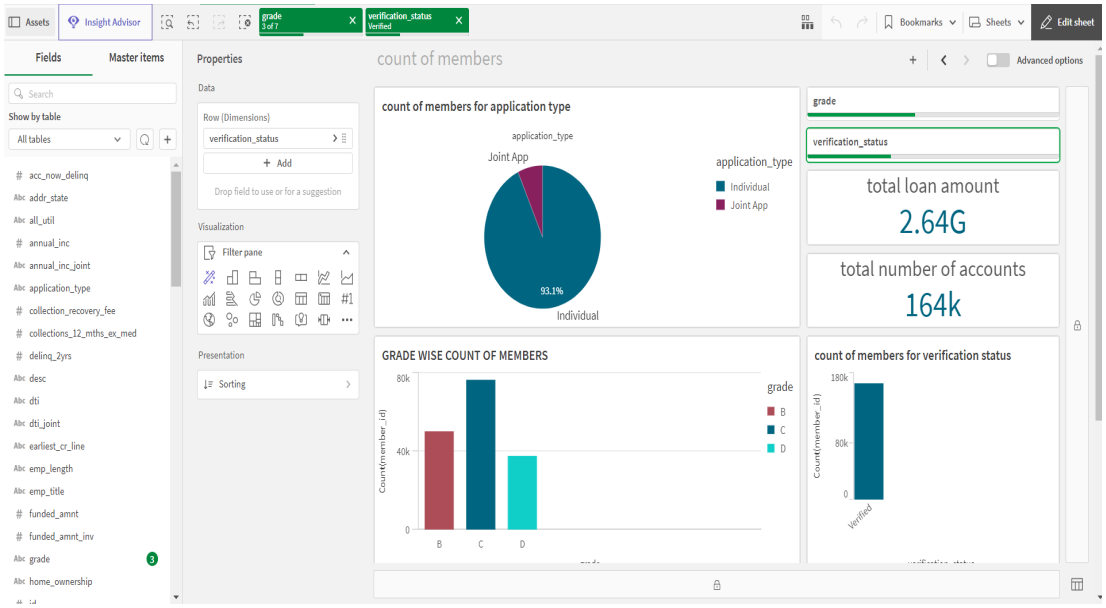


Figure 9(a)- Utilization of filters

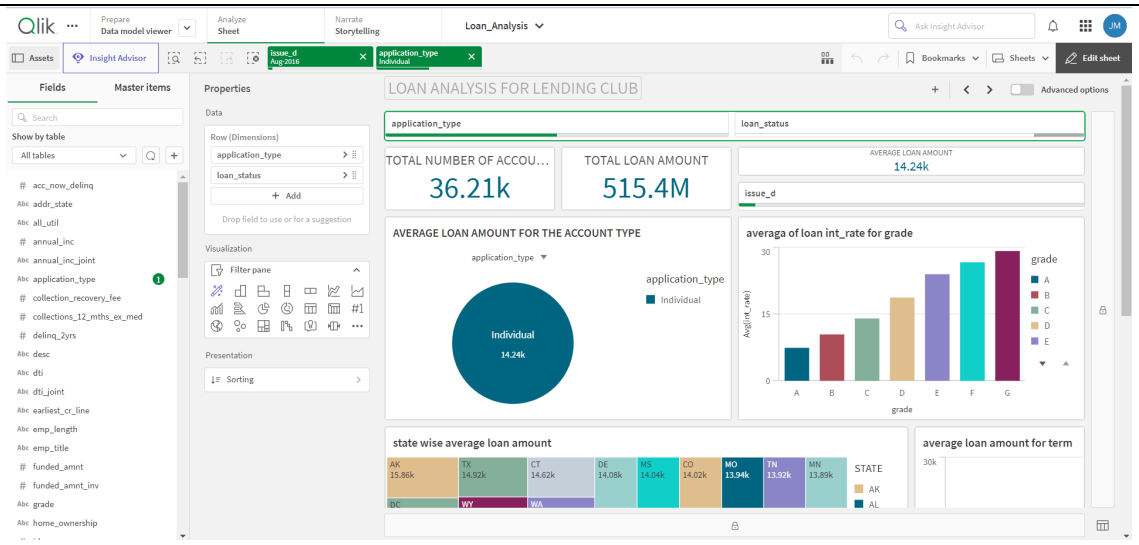


Figure 9(b)- Utilization of filters

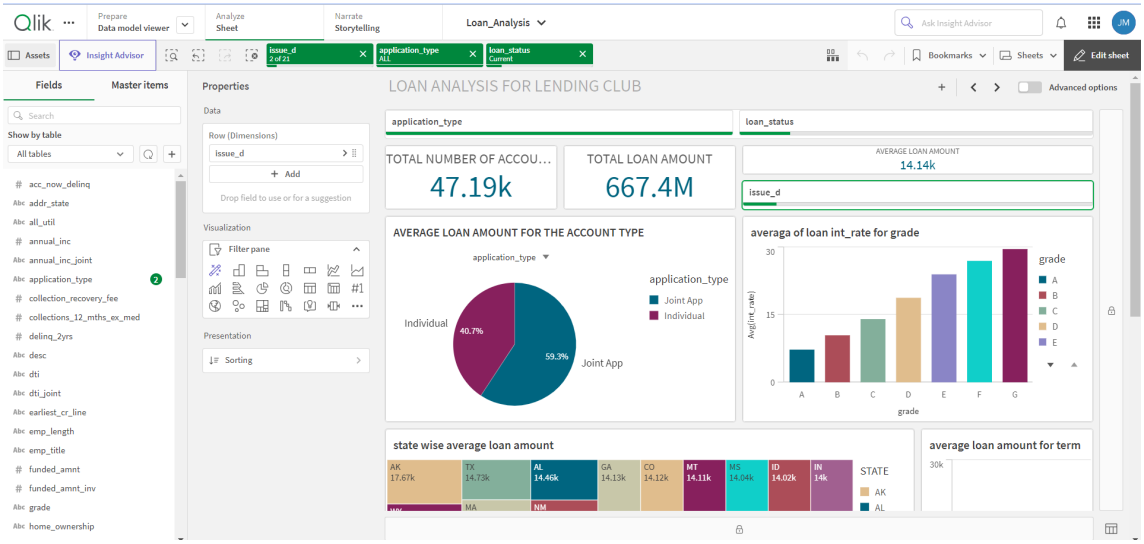


Figure 9(c)- Utilization of filters