A Project Report On Lending Club Issued Loans Analysis

Introduction:

Lending Club, a prominent peer-to-peer lending platform, has revolutionized the way individuals and small businesses secure loans. By connecting borrowers directly with investors, Lending Club eliminates traditional banking intermediaries, offering a more streamlined and often more affordable borrowing process. Analyzing the loans issued by Lending Club provides valuable insights into the lending trends, borrower demographics, loan performance, and risk factors associated with peer-to-peer lending. This analysis involves examining a variety of loan attributes such as loan amount, interest rates, loan grades, borrower credit scores, and repayment status. By leveraging data analytics and machine learning techniques, we can uncover patterns and correlations that inform better lending decisions, enhance risk management strategies, and ultimately contribute to the efficiency and reliability of the lending marketplace. Understanding these dynamics not only benefits investors seeking to maximize their returns but also supports borrowers in obtaining fair and transparent loan terms.

1. Define Problem / Problem Understanding:-

1.1:- Specify the business problem:-

The specific business problem revolves around the inadequacy of the current lending strategy, which is not sufficiently informed by comprehensive insights derived from LendingClub loan data. The institution struggles to assess borrower behavior and market dynamics effectively, resulting in challenges such as inaccurate risk identification, difficulties in predicting loan default rates, and the inability to dynamically adjust lending

criteria to respond to evolving market conditions.

1.2:- Business requirements:-

The business requirements involve the establishment of a robust data analytics framework that can extract meaningful insights from LendingClub issued loans data. This framework should enable the financial institution to gain a deep understanding of borrower behavior, identify high-risk segments, predict default rates accurately, and provide the necessary foundation for real-time adjustments to lending criteria. Additionally, the solution should be scalable, adaptable, and capable of integrating with existing systems to ensure seamless implementation.

1.3:- Literature Survey :-

1.Default Prediction

Several studies have focused on predicting loan defaults using machine learning and statistical methods:

- Malekipirbazari and Aksakalli (2015) utilized ensemble learning techniques, such as Random Forests and Gradient Boosting Machines, to predict loan defaults. They found that these methods outperformed traditional logistic regression models in terms of predictive accuracy.
- Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015) analyzed the use of logistic regression, decision trees, and neural networks for predicting defaults. Their study highlighted the importance of feature selection and the use of borrower's credit score as a significant predictor.
- Ravindra and Bhatnagar (2019) applied deep learning techniques, specifically Long Short-Term Memory (LSTM) networks, to model time-series data of loan repayments. Their approach improved prediction accuracy over conventional methods.

2. Risk Assessment

Assessing the risk associated with Lending Club loans has been another crucial area of research:

- Emekter, Tu, Jirasakuldech, and Lu (2015) explored the risk-return profile of P2P

lending by analyzing the determinants of loan performance. They concluded that interest rates, loan grades, and debt-to-income ratios significantly affect the risk of default.
- Iyer, Khwaja, Luttmer, and Shue (2016) examined the impact of soft information, such as borrower narratives, on loan performance. They found that qualitative information provided by borrowers can be a useful supplement to quantitative credit data.

3. Loan Performance Analysis

Analyzing the performance of loans issued by Lending Club involves examining various factors that influence repayment behavior and investor returns:

- Lin, Prabhala, and Viswanathan (2013) studied the influence of social networks on loan outcomes. They found that borrowers with a higher number of friends on the platform are more likely to have their loans funded and exhibit better repayment behavior.
- Herzenstein, Dholakia, and Andrews (2011) investigated the effects of borrower characteristics on loan performance. Their findings indicated that demographic factors, such as age and employment status, significantly influence the likelihood of loan approval and repayment.
- Morse (2015) focused on the role of geographic factors in loan performance. The study highlighted that borrowers from regions with higher economic stability and lower unemployment rates tend to have better loan outcomes.

4. Comparative Studies

Comparative studies between Lending Club and other lending platforms provide insights into the broader P2P lending market:

- Hulme and Wright (2006) compared the performance of loans on Lending Club with those on Prosper, another P2P lending platform. They found differences in default rates and investor returns, attributing these to platform-specific policies and borrower demographics.
- Bachmann, Becker, Buerckner, Hilker, and Kock (2011) analyzed the differences in regulatory frameworks between the US Europe and their impact on P2P lending platforms like Lending Club. Their study suggested that regulatory environment plays a significant role in shaping the risk and return profiles of loans.

2. Data Collection & Extraction From Database:-

2.1:- Data Collection:-

Kaggle Platform :-

Kaggle is a popular platform that hosts a variety of datasets, including those related to finance and peer-to-peer lending. The Lending Club dataset on Kaggle contains comprehensive information on loans issued by the platform.

Accessing the Dataset:-

1. Kaggle Account: Ensure you have a registered account on Kaggle.

2. Dataset URL: Navigate to the Lending Club dataset page on Kaggle. The dataset can be found here.

3. Google Drive Link: https://drive.google.com/file/d/12UMiBk_LRMzXfgCGNp-Jlkq5_NKCxmch/view?usp=drive_link

Downloading the Dataset:-

1. Download Data: Click on the "Download" button to download the dataset. The dataset typically comes in CSV format, which is suitable for data analysis and visualization.

2. API Access: Alternatively, you can use Kaggle's API to download the dataset programmatically.

3. Data Preparation for Visualization:-

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for

creating visualizations to gain insights into the performance and efficiency. Since the data is already cleaned, we can move to visualization.

4. Data Visualization:-

Data visualization is the process of creating graphical representations of data to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

4.1. Loan Amount Analysis :-

This visualization contains the average loan amount taken by members for the tenure such as 36 months or 60 months. Some common types of visualizations that can be used to analyze the performance and efficiency of banks include bar charts, line charts, heat maps, scatter plots, pie charts, Maps etc.

Total Number of Accounts:-

Total Number of Accounts

Sum(total_acc)

78.21k

Total Loan Amount

Total Loan_Amount

Sum(loan_amnt)

3.81M

Average Loan Amount

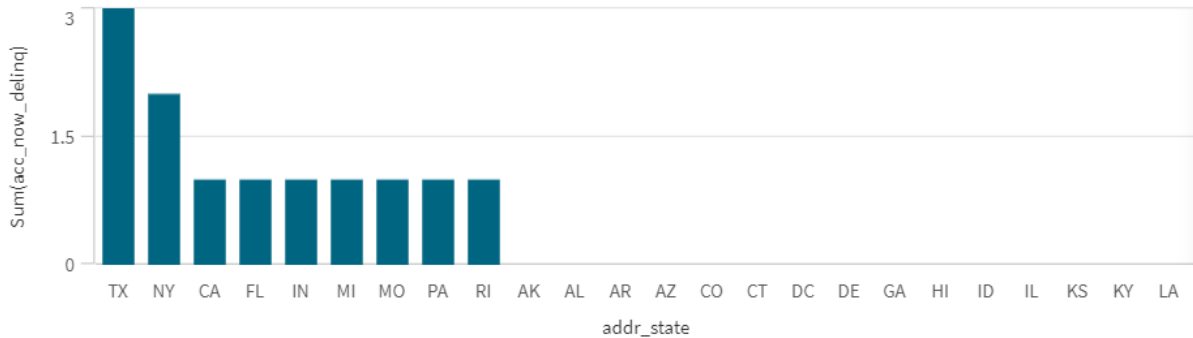## Average Loan Amount

Avg(loan_amnt)

1k

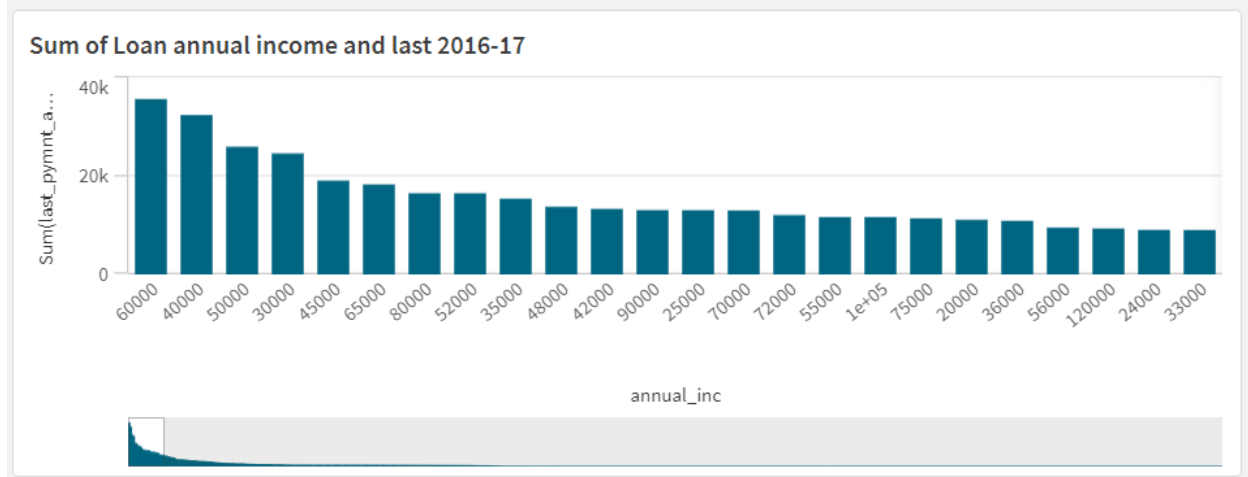Sum of All Utilization

Sum of All utilisations

Sum(all_util)

230.8k

Sum of Loan 2016-17 Accounts – State Wise
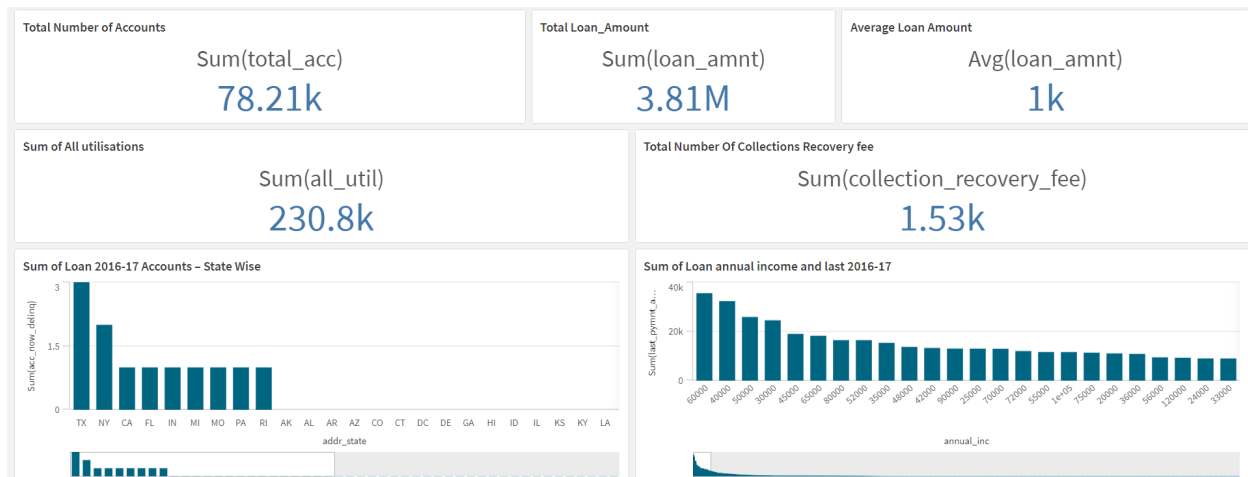

Sum of Loan annual income and last 2016-17

Sum of Loan annual income and last 2016-17

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.

Total Number of Accounts	Total Loan_Amount	Average Loan Amount
Sum(total_acc)	Sum(loan_amnt)	Avg(loan_amnt)
78.21k	3.81M	1k

Sum of All utilisations	Total Number Of Collections Recovery fee
Sum(all_util)	Sum(collection_recovery_fee)
230.8k	1.53k

Sum of Loan 2016-17 Accounts – State Wise

Sum of Loan annual income and last 2016-17

avg funded_amnt_inv by id

id	avg funded_amnt_inv
111835534	40k
112436524	40k
112052761	40k
111816681	40k
111743616	40k
111670729	40k

Distribution of avg loan_amnt for zip_code

zip_code
Area layer

avg loan_amnt
35k

© OpenStreetMap contributors

avg loan_amnt by term and loan_status

60 months

Default 21.6k	Late (31-120 days) 20.7k	Fully Paid 20.69k	Current 20.66k
In Grace Period 21.26k			
Late (16-30 days) 20.92k	Charged Off 20.66k		

36 months

In Grace Period 14.32k	Late (31-120 days) 13.3k	Current 12.54k
Late (16-30 days) 13.84k		
Charged Off 13.32k	Fully Paid 12.51k	
	Default 11.96k	

Comparison of avg loan_amnt and avg total_acc for zip_code

Story:-

LENDING CLUB ANALYSIS

Total Number of Accounts

Sum(total_acc)

78.21k

Total Loan_Amount

Sum(loan_amnt)

3.81M

Average Loan Amount

Avg(loan_amnt)

1k

Sum of All utilisations

Sum(all_util)

230.8k

Sum of Loan 2016-17 Accounts – State Wise
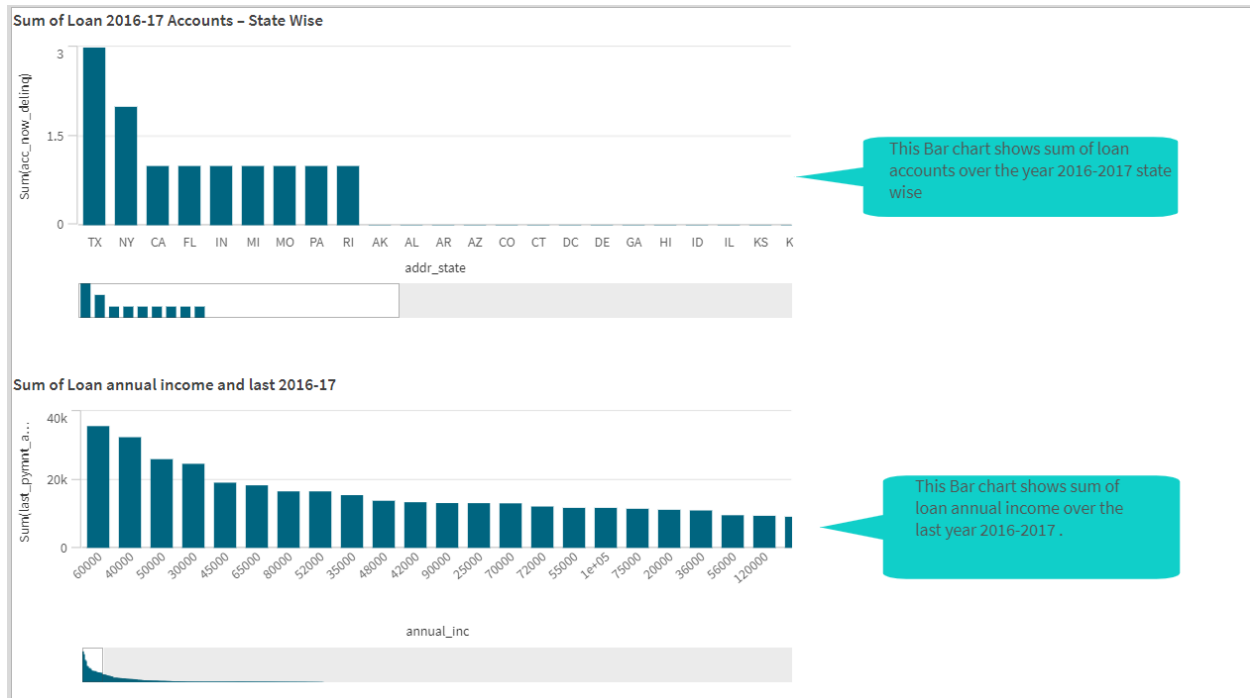
This Bar chart shows sum of loan accounts over the year 2016-2017 state wise

Sum of Loan annual income and last 2016-17

This Bar chart shows sum of loan annual income over the last year 2016-2017 .

Conclusion:-

Summary :-

This report provides a detailed analysis of Lending Club issued loans, uncovering key factors influencing loan performance and offering actionable insights for improving lending strategies.

Insights from Dashboards:-

Interest Rate Impact: Higher interest rates correlate with higher default rates.

Geographic Trends: Certain regions have higher concentrations of high-risk loans. Credit Score and Employment: Borrowers with higher credit scores and longer employment histories tend to have better loan performance. Recommendations Implement advanced credit scoring models to enhance risk assessment. Focus on borrower education to improve creditworthiness and loan performance. Continuously monitor and adapt lending strategies based on data-driven insights.

Submitted BY: *Project Guide:*

Anurag Jethi   *Mr. Revanth Reddy Gunapat*