1-Define Problem/Problem Understanding

1.1. Specify The Business Problem

The current lending strategy is not working well because it doesn't use enough information from LendingClub loan data. This makes it hard for the institution to understand borrower behavior and market trends. As a result, they face problems like not identifying risks correctly, having trouble predicting loan defaults, and not being able to adjust lending criteria as market conditions change.

1.2. Business Requirements

The business needs a strong data analytics system to get useful insights from LendingClub loan data. This system should help the financial institution understand borrower behavior, find high-risk groups, accurately predict default rates, and make real-time changes to lending criteria. It should also be scalable, adaptable, and able to work with existing systems for easy implementation.

1.3.Literature Survey

A comprehensive literature survey is crucial for understanding existing methodologies, tools, and best practices in optimizing lending strategies and managing risk on peer-to-peer lending platforms like LendingClub. The goal of this survey is to find relevant studies on data analytics in finance, trends in peer-to-peer lending, and similar analyses done by financial institutions to improve their decision-making processes.

2-Data Collection & Extraction From Database

Data collection is the process of gathering and measuring information in a structured way. This helps to answer research questions, test ideas, evaluate results, and gain insights from the data.

Step 1: Downloading the dateset from the given link
Dataset link - https://www.kaggle.com/datasets/husainsb/lendingclub-issued-loans

Step 2: Understand the data
The dataset includes comprehensive meta information about the columns present in the CSV files.

Here is a description of each column in the dataset:

- **member_id**: A unique identifier assigned to each member.
- **loan_amnt**: The amount of the loan taken by members.
- **term**: The duration of the loan.
- **int_rate**: The interest rate applied to the loan.
- **grade**: The grade assigned to the members.

3- Data Preparation

To prepare the data for visualization, we need to:

1. Remove any irrelevant or missing data.
2. Transform the data into a format that's easy to visualize.
3. Explore the data to find patterns and trends.
4. Filter the data to focus on specific subsets.
5. Prepare the data for visualization software.
6. Ensure the data is accurate and complete.

This process makes the data easy to understand and ready for creating visualizations that provide insights into performance and efficiency. Since the data is already cleaned, we can proceed to visualization.

Data Preparation link - https://drive.google.com/file/d/1Yhu87yAPfzpEJir-LVOWeFRMabaFz1w8/view?usp=sharing

4-Data Visualization
Now perform the various visualizations

- **Loan Amount Analysis:** This visualization shows the average loan amount taken by members for different loan terms, like 36 months or 60 months. Common visualizations for analyzing bank performance and efficiency include bar charts, line charts, heat maps, scatter plots, pie charts, and maps.
- **Average Loan amount For the Account_Type**
- **Total Loan_amount**
- **Total number of Loan_Account**
- **Average Loan Amount -State Wise**
- **Grade Wise- Count of Members**
- **Verification_Status**

5- Dashboard

A dashboard is a visual interface that presents information and data in a clear and organized way. It's often used for real-time data monitoring and analysis, tailored to specific purposes or tasks. Dashboards are used in many fields like business, finance, manufacturing, healthcare, and more. They help track key performance indicators (KPIs), monitor metrics, and display data as charts, graphs, and tables.

Dashboard 1: Average loan analysis
Explanation video link:
https://drive.google.com/file/d/1XIGMXB9JdEMlM2NEzFqrywGjMJ6gPBad/view?usp=sharing

Dashboard 2: Count analysis
Explanation video link :
https://drive.google.com/file/d/14WcP_VPGO5SGPK1HtYlwQqP9FfTNCbmB/view?usp=sharing

6- Story
A data story presents data and analysis in a narrative format to make the information more engaging and understandable. It usually includes:

- **Introduction**: Sets the stage and explains the context
- **Body**: Presents the data and analysis logically and systematically.
- **Conclusion**: Summarizes key findings and highlights their implications.

Data stories can be shared through reports, presentations, interactive visualizations, and videos.

Explanation video link for design of story:
Link 1:
https://drive.google.com/file/d/1Xrq6Q03zFz3K9QiD-F-50LZUc7qju2Hx/view?usp=sharing
Link 2:
https://drive.google.com/file/d/1lwR9mnFji4DKAKJeaSuPfpIvEAApf_K9/view?usp=sharing

7- Performance Testing

- **Amount of data loaded:** "Amount of Data Loaded" refers to how much data has been imported or loaded into a system, software, database, or other storage or processing environment. It measures the volume of data that has been successfully processed and is now available for analysis, manipulation, or use within the system.
- **Utilization of filters :** "Utilization of Filters" refers to using filters in a system, software, or data processing pipeline to selectively extract, manipulate, or analyze data based on

specific criteria. Filters help narrow down the data, focusing only on the relevant information that meets predefined conditions.

- **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