**Milestone 1: Market Research Phase**

**Problem Statement:**

**Develop a machine learning solution for the banking domain that automates and accelerates the loan approval process by effectively filtering and identifying suitable candidates from a high volume of loan applications. The current manual review process is time-consuming and inefficient, hindering the productivity of bank employees. By leveraging the available banking data, the goal is to create a model that accurately predicts the approval status of loan applications, enabling the bank to streamline operations, reduce processing time, and make more informed decisions regarding loan approvals.**

**Objective:**

**The objective of this machine learning project in the banking domain is to build a predictive model using historical banking data. The model aims to assist bank employees in efficiently and accurately determining suitable candidates for loan approval.**

**Data Dictionary:**

**(Currently gathered features)**

* **Loan ID: Unique Loan ID issued on every loan for an applicant**
* **Gender: Gender of an applicant whether male or female**
* **Married: Marital status of an applicant i.e., Yes for married and NO for single**
* **Dependents: Number of individuals who are financially dependent on applicant**
* **Education: Highest Education of applicant, (Bachelor, Post Graduation etc)**
* **Self-employed: Whether the applicant is self-employed or not (Yes for self-employed or else NO)**
* **Applicant Income: Income of the applicant**
* **Co-Applicant Income: Applicant have to put one nominee name that is called Co-Applicant. So, it is column related to co-applicant income**
* **Loan Amount: Amount of loan applicant wants to issue from the bank.**
* **Loan Amount Term: The amount of time the lender gives you to repay your whole loan**
* **Credit History: The credit done in the past by the applicant**
* **Property Area: This feature tells us about the applicant property is in which area i.e., Rural or Urban**
* **Loan status: It is a target variable column which tells about whether the applicant application for loan approval is passed or not**

**Independent Feature Impact Analysis on the Target Feature**

* **Loan ID:**

**The loan ID is a unique identifier assigned to each loan application. Since each loan ID is different and does not carry any inherent pattern or relationship with the target variable, it does not contribute to the predictive power of the model. Thus, would not hold importance as a contributing feature.**

* **Gender:**

**(Gender of an applicant whether male or female)**

**Gender may have a potential impact on loan approval as it could be used as a factor in assessing risk or determining eligibility criteria. For example, in certain regions or cultures, there may be specific loan programs or considerations based on gender. Analysing the gender distribution and its relationship with the loan approval status could provide insights into any potential biases or patterns.**

* **Married:**

**(Marital status of an applicant i.e., Yes for married and NO for single)**

**Income and debt of both applicants may be considered while assessing credibility. The marital status of the applicant could impact loan approval. For instance, some lenders may consider married applicants to be more financially stable or responsible. Analysing the loan approval status based on marital status can help identify any potential preferences or biases in the loan approval process.**

* **Dependents:**

**(Number of individuals who are financially dependent on applicant)**

**The number of dependents an applicant has may influence their ability to repay the loan. More dependents could indicate a higher financial burden, which might affect loan approval. Understanding the relationship between the number of dependents and loan approval status can provide insights into the risk assessment process.**

* **Education:**

**Highest Education of applicant, (Bachelor, Post Graduation etc)**

**The educational background might impact loan approval. Some lenders may prioritize applicants with higher education as they might be perceived to have better job prospects or financial stability. Analysing the loan approval status based on education level can help identify any correlations or biases.**

* **Self-employed:**

**Whether the applicant is self-employed or not**

**Being self-employed could potentially affect loan approval, as it may involve more irregular income or higher perceived risk as compared to individuals with jobs. Analysing the loan approval status for self-employed individuals can help determine if there are any specific considerations or challenges in loan assessment for this group.**

* **Applicant Income:**

**Income of the applicant**

**The income level of the applicant is a crucial factor in loan approval. Higher income levels generally indicate a greater ability to repay the loan. Analysing the relationship between applicant income and loan approval status can provide insights into the income thresholds or criteria used by the bank for loan approval.**

* **Co-Applicant Income:**

**Applicant have to put one nominee name that is called Co-Applicant. So, it is column related to co-applicant income**

**The income of the co-applicant, if applicable, can also impact loan approval. Combining the income of both the applicant and co-applicant can enhance the overall repayment capacity. Analysing the loan approval status based on co-applicant income can help identify the significance of this factor in loan assessment.**

* **Loan Amount:**

**Amount of loan applicant wants to issue from the bank.**

**The loan amount requested by the applicant is an essential factor in loan approval. Higher loan amounts may involve greater risk or stricter assessment criteria. Analysing the relationship between loan amount and loan approval status can provide insights into the loan size preferences or policies of the bank.**

* **Loan Amount Term:**

**The amount of time the lender gives you to repay your whole loan**

**The loan amount term represents the duration of the loan. Longer-term loans may have different approval criteria or considerations compared to shorter-term loans. Analysing the loan approval status based on loan amount term can help identify any preferences or biases related to loan duration.**

* **Credit History:**

**The credit done in the past by the applicant**

**The Credit history is a critical factor in loan approval. It indicates the applicant's past repayment behaviour and creditworthiness. Analysing the loan approval status based on credit history can provide insights into the importance of creditworthiness in the bank's decision-making process.**

* **Property Area: This feature tells us about the applicant property is in which area i.e., Rural or Urban**

**The property area represents the location of the property for which the loan is being sought. Different property areas may have varying market conditions, risks, or loan programs. Analysing the loan approval status based on property area can help identify any geographical considerations or preferences in the loan approval process.**

**Suggest ways in which the organisation can benefit as a result of analysing the data**

* **Improved Loan Approval Process**
* **Operational Efficiency and Cost Savings**
* **Compliance and Fraud Detection**
* **Business Strategy and Decision Making**

**Suggesting missing features that can help with the analysis based on business logic**

* **Type of Property loan: What category the said loan falls under.**

**The specific type of property loan, such as a home loan, commercial property loan, construction loan or land loans can influence the loan approval process and the criteria used by the bank or lender to assess the borrower's eligibility. Since every type of loan has varying risk profiles, requirements and approval processes. It could be an important feature we could analyse.**

* **Occupation / Employment Sector**

**Capturing the sector in which the borrower is employed, such as government, private, or self-employed, can help understand the impact of employment sector on loan approval status. Different sectors may have varying income stability and risk profiles.**

* **Employment Stability Measure**

**The length of employment, number of job changes etc.**

* **Interest Rate:**

**This could be useful for the company. In the sense of assessing what rate is most preferred and applied for by their customers.**

* **Loan-to-Income Ratio:**

**Calculating the ratio of the loan amount to the borrower's income provides insights into their debt relative to their earning capacity. This feature helps assess affordability and risk associated with the loan.**

* **Debt to income ratio**

**In some cases, the applicant might have a previous loan. This could affect their eligibility in repaying a new loan taken. Calculating the ratio of a borrower's total debt obligations to their income can provide a measure of their overall debt burden. This feature helps assess the borrower's ability to manage additional loan repayments and can impact loan approval decisions**

**What is the best way to collect data for the suggested features?**

* **Internal Data**

**Working within a banking institution, leverage the internal data available within the organization.**

* **Surveys and Questionnaires:**

**Conducting surveys or questionnaires targeted at loan applicants or existing customers can provide valuable data for certain features.**

* **Publicly Available Data:**

**Explore publicly available data sources that provide information on occupation, income ranges, industry statistics, and economic indicators.**

* **Data Scraping:**

**In some cases, data can be collected through web scraping techniques. This method involves extracting information from websites or online platforms that provide occupation-related data, income ranges, or industry-specific details. However, it's important to respect website terms of use and legal considerations when using this method.**