MORTGAGE ANALYSIS

This case is about a finance company that specialises in giving various types of loans to customers. In this case analysis we will try to throw some light on the factors to know the basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

The data given contains information about past loan applicants and whether they 'defaulted' or not. **The aim is to identify factors which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (too risky applicants) at a higher interest rate, etc.**

Let us study the data and identify some conditions of Loan Accepting and Loan Rejecting

When a person applies for a loan, Either the bank accepts the loan application, or it gets rejected. In the data if we see the loan status variable, we can clearly see that-

Loan Accepted has also 3 criteria:

- Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
- Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset).

Starting with the analysis we will see first the

1. **geographical distribution of Loan amount** distributed.

Field used in this operation is –

- Addr_state from dimensions
- loan amount from measures

After analysing the map, it is identified that California, Texas, New York and Florida have the largest amount loans.

2. **Now we will see the average interest rate distribution among states.**

Field used in this operation is –

- Addr_state from dimensions
- Average Interest Rate from measures

From the analysis, it is clear that Applicants from Idaho and Iowa, and Maine experienced relatively much lower rates on average than the others.

3. Now I tried to analyse **what is the purpose for borrowing for the applicants.**

Field used in this operation is –

- Purpose from dimensions
- Total account from measures
  We applied packed bubble chart for the analysis

We came to know that Debt consolidation is the most common reason for borrowing. The advantage of peer to peer lending is the lower interest rate. So, most customer choose to consolidate debt. And credit card, house and small business are the other most popular reasons for borrowing.

4. **Interest Rate by Grade:**

Field used in this operation is –

- Grade from dimensions
- Number of records and Interest rate from measures
  We applied packed Box and whisker plot for the analysis

After analysing we get to know that the interest rates are being increased from A to G grade, as we move further.

5. **Then I tried to analyse the employee length and loan status which is fully paid by house ownership**
   For this I created a Calculated field named as Fully Paid having formula-

   [loan_status]= "Fully Paid" and [home_ownership]="OWN"

Field used in this operation is –

- Emp length from dimensions on colour shelf
- Total Accounts from measures
- Fully Paid calculated field on filter shelf and filtered the true records.
  From this analysis I came to know that Employment length is variability in 'fully paid' loan status, except people who work for 10 or more years, they fully paid their loan more.

6. DASHBOARD:

Then using all the sheets that we analysed, I created a visual dashboard showing every analysis on a single pane.

The dashboard is set to be filtered according to be Geographical Distribution of loan amount and it will show all the other graphs updated for the selected location.

As soon as u leave the selection, it comes to the original state.

Attached is the screenshot for the Dashboard.
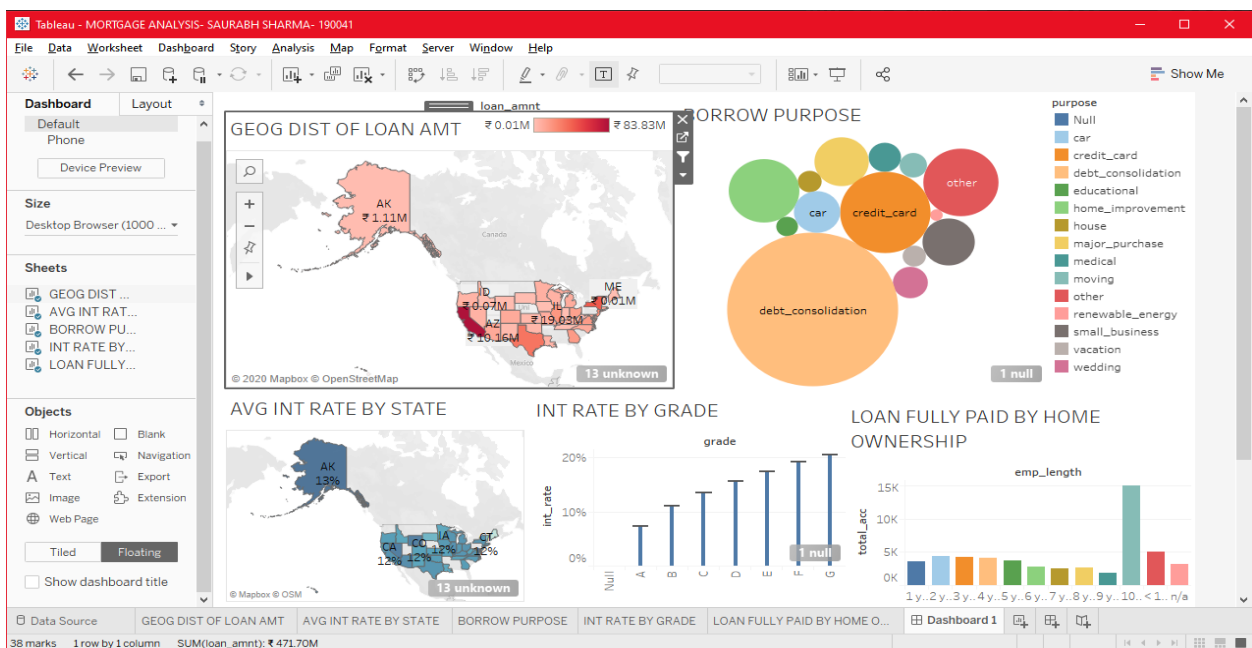
Conclusion:

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

From the geographical distribution we can see that

- **California, Texas, New York and Florida have the largest amount loans.**
- **Applicants from Idaho and Iowa, and Maine experienced relatively much lower rates on average than the others**
- **Debt consolidation is the most common reason for borrowing.**
- **the interest rates are being increased from A to G grade, as we move further.**
- **Employment length is variability in 'fully paid' loan status, except people who work for 10 or more years, they fully paid their loan more.**