**LLOYDS BANKING GROUP DIG DATA CHALLENGE**

**Task A –** Data Strategy: Understand and summarise the different behaviours or attributes between customers who paid back their loan and customers who did not

**Submitted by: Nincy Alisa Jacob**

**Email: ninciealisa23@gmail.com**

**Overview of the dataset**

This dataset includes loan information for 18,324 clients who have previously held a loan, together with the loan's status (whether the consumer has repaid the loan or not):

Customers who repaid their debt are classified as "Fully Paid," while those who defaulted on it are classified as "Charged-off." Additional credit and product information that can be utilised to evaluate a customer's credit or financial behaviour also exists.

The dataset consists of 31 columns out of which the most informative 15 attributes are chosen, and the remaining are removed.

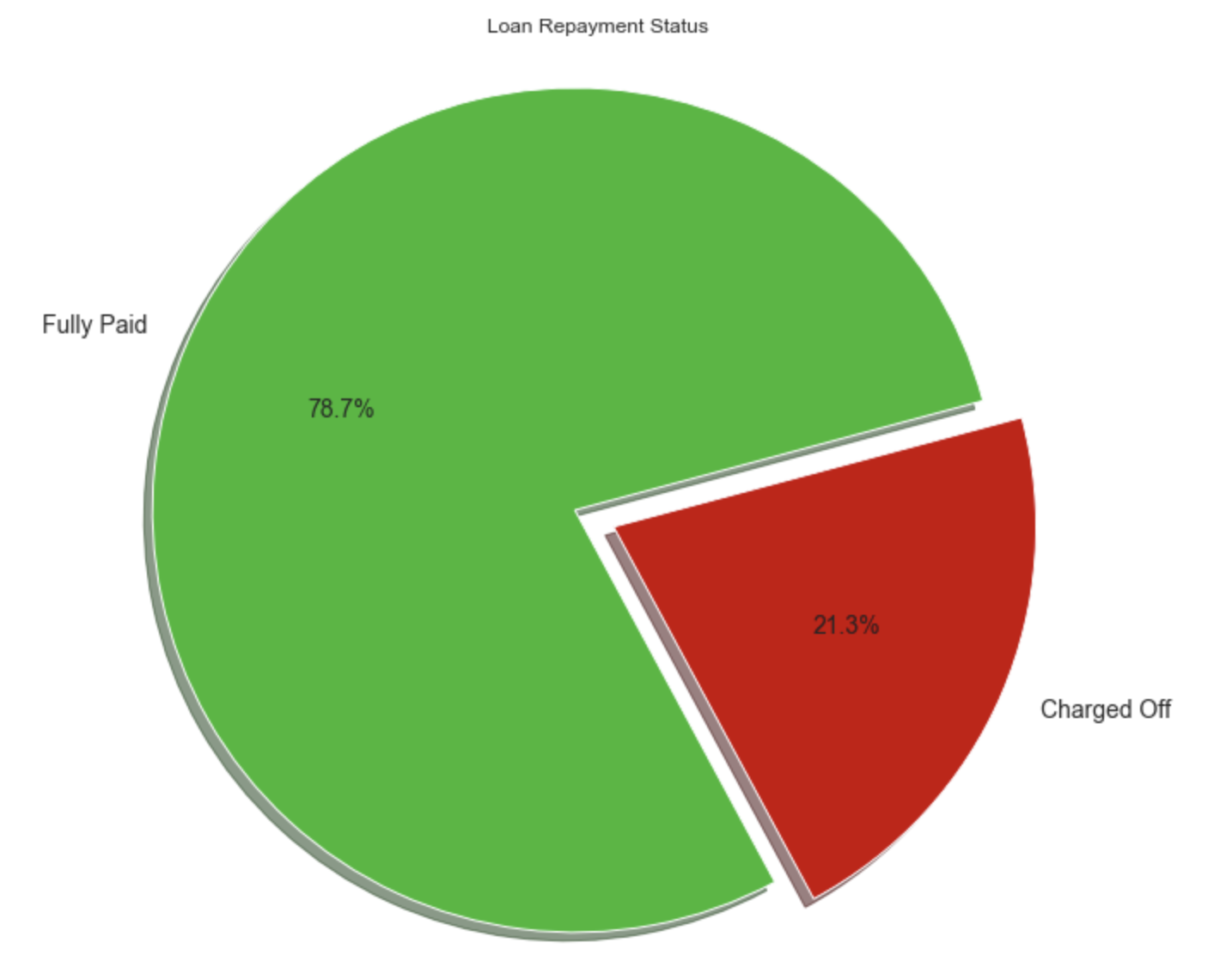
**Data Cleaning**

Diagram

Description automatically generated with medium confidence

The above figure gives an overview of the number of missing values in the dataset. The attributes with highest proportion of missing values are dropped from the dataset. Remaining missing values are imputed with NA, mode or mean appropriately.

**DATA VISUALISATION**

****

The above pie chart depicts that around 21% of the total loan customers are defaulters and their loans are charged off. Hence, the bank must consider more factors precisely before disbursing a loan.

**Loan Repayment Status based on Purpose of Loan**

**Graphical user interface

Description automatically generated with low confidence**

**Observation:**

1. Debt consolidation is the category with highest no of loans and a significant proportion of these loans are charged off.
2. Credit card loans and small business loans have high proportion of loan defaulters as compared to other purposes.

**Chart, histogram

Description automatically generated**

**Observation:**

1. Customers with instalment amount greater than 700 end up in defaulting the loan payments.

**Chart, bar chart

Description automatically generated**

**Observation:**

1. As the loan amount increases, the chances of not repaying the loan also increases.
2. Loans less than 15000 have higher chances of getting repaid.

Chart, histogram

Description automatically generated

**Observations:**

1. Loan defaulting chances are higher when monthly instalment exceeds 10% of the in-hand salary.

**Loan Repayment status based on Loan Term**

**Chart, bar chart

Description automatically generated**

**Observations:**

1. Most people opt for 36 months or shorter term
2. People opting longer term are less likely to repay loan as compared to others.

**Bivariate Analysis**

The status of loan is visualised against some important attributes which might have contributed to the charged-off loans. For this, the proportion of charged off loans is calculated by grouping the loans according to different variables and loan status. This is divided by the total no of loans. Higher the proportion, more chances of defaulting loans.

Chart, bar chart

Description automatically generated

**Observations:**

1. Income range 80000+ has less chances of getting charged off.
2. Income range 20000-40000 has high chances of getting charged off.
3. Income ranges of 0-20000, 40000-60000 and 60000-80000 have almost equal chances of getting charged off

Chart, histogram

Description automatically generated

Chart, bar chart

Description automatically generated

**Observations:**

1. Debt consolidation is the purpose for highest no of loans. At the same time, this category has a significant proportion of loan defaulters.
2. Educational Loans have highest charged off proportion. However, the no of educational loans was very less to make any conclusion out of this.
3. Home improvement loans have higher chances of getting charged off.
4. Car Loans and Wedding Loans are safer as compared to other categories

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

**Observations:**

1. The states IA, ID, AK, DC, ME, MT, ND, SD, WY has very less no of loan applications. Hence no conclusions can be derived using the charged off proportion.
2. AR, CA, FL, OK, TX, VA, MD has higher no of charged-off loans.

Chart, bar chart

Description automatically generated

**Observations:**

1. Employment length is not contributing much as a factor in deciding whether loan will be repaid or not. Loan defaulters are almost evenly distributed in all range of employment experiences
2. Employees with less than 1year experience also have almost same chances of repaying loan as compared to employees with longer experience.

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

**Observations:**

1. Customers with rented home are less likely to repay the loan

Chart

Description automatically generated

**Chart, bar chart

Description automatically generated**

**Observations:**

1. Customers with no bankruptcy record have higher chances of repaying the loan.
2. Customers with more than 1 bankruptcy record are very much less likely to repay the loan.

**KEY INFERENCES**

1. Loan Amount and Monthly instalment amount should be calculated based on the customer’s salary and the instalment amount should ideally be 10% or below of the in-hand salary
2. Higher Loan Amounts should be approved more carefully after additional checking and refinements
3. Customers with rented home ownership should only be granted loans after assessing their credit lines and income
4. Employment length should not be given more weightage while assessing the chances of approving a loan
5. Borrowers profile must be thoroughly checked before granting loans if the total no of credit lines accounts exceeds 20
6. Customers with high annual income have high chances of loan repayment.
7. Loans for home improvement and debt consolidation must be granted after careful consideration
8. Customers with more than 1 public bankruptcy record should be granted loan after careful consideration.
9. More background checks should be carried out for customers from the states: AR, CA, FL, OK, TX, VA, MD has higher no of charged-off loans.