CLASSIFICATION DATASET

Follow the below steps for the 2 datasets below:

1. Create all necessary plots using Matplotlib and seaborn library answering necessary questions
2. Create a one pager explanation of the case study along with the codebook. Add graphical plots in the case study - It could be a PDF or a PPT.
3. You can experiment with different cleaning and scaling techniques.
4. Create your own ML model on the dataset below
5. Comment each code block

Health Insurance Lead Prediction

Link to the dataset:

📗 Health Insurance Lead Prediction

Link to the session's colab Notebook for reference:

∞ Health_insurance_final.ipynb

Loan Delinquency Prediction

Loan default prediction is one of the most critical and crucial problem faced by financial institutions and organizations as it has a noteworthy effect on the profitability of these institutions. In recent years, there is a tremendous increase in the volume of non – performing loans which results in a jeopardizing effect on the growth of these institutions. Therefore, to maintain a healthy portfolio, the banks put stringent monitoring and evaluation measures in place to ensure timely repayment of loans by borrowers. Despite these measures, a major proportion of loans become delinquent. Delinquency occurs when a borrower misses a payment against his/her loan. Given the information like mortgage details, borrowers related details and payment details, our objective is to identify the delinquency status of loans for the next month given the delinquency status for the previous 12 months (in number of months).

UNDERSTANDING BANKING TERMS

- •   **- Debt to income ratio**:

> measures the amount of income a person or organization generates in order to service a debt. A low DTI ratio indicates sufficient income relative to debt servicing, and makes a borrower more attractive.

- •   **- Loan to Value ratio :**

> A loan-to-value (LTV) ratio compares the amount of a loan you're hoping to borrow against the appraised value of the property you want to buy. Lenders use LTVs to determine how risky a loan is and whether they'll approve or deny it. A higher LTV ratio suggests more risk because there's a higher chance of default. The more money a lender gives you, the higher your LTV ratio and the more risk they're taking. If you're considered a higher risk for the lender, this usually means that:

  1. It's harder to get approved for loans.
  2. You might have to pay a higher interest rate.
  3. You might have to pay additional costs, such as mortgage insurance.

- •   **- Insurance percent**

>a provision in a health insurance contract stipulating that the insurer and insured will share covered losses in agreed proportions. For example, the insurer may be required to pay 80 percent of the insured's hospital costs with the insured responsible for the remainder.

Data Dictionary of the above Dataset:

Variable	Description
loan_id	Unique loan ID
source	Loan origination channel
financial_institution	Name of the bank
interest_rate	Loan interest rate
unpaid $principal$ bal	Loan unpaid principal balance
loan_term	Loan term (in days)
origination_date	Loan origination date (YYYY-MM-DD)
first $payment$ date	First instalment payment date
loan to value	Loan to value ratio
number of borrowers	Number of borrowers
debt to income_ratio	Debt-to-income ratio
borrower $credit$ score	Borrower credit score
loan_purpose	Loan purpose
insurance_percent	Loan Amount percent covered by insurance
co-borrower $credit$ score	Co-borrower credit score
insurance_type	0 - Premium paid by borrower, 1 - Premium paid by Lender
m1 to m12	Month-wise loan performance (deliquency in months)
m13	target, loan deliquency status (0 = non deliquent, 1 = deliquent)