LENDING CLUB

DATA ANALYSING

PREREQUISITES

- Import the necessary libraries to the Jupyter Notebook
- Load the CSV file from the path
- Get the Jupyter ready for Data Analysis

VERIFYING THE DATA

- check the missing values in the dataframe
- Find the columns where all the values are null
- Check the null columns

DATA CLEANING

- Drop the null columns (containing all the values as null)
- Get the shape of the Loan Dataframe after dropping the null columns
- Drop the remaining unwanted columns
- Get the shape of the Loan Dataframe after dropping the unwanted columns

- Get the emp_length column with only numbers
- Remove the % symbol from Int_rate column
- Check the numeric values for loan_amnt
- Get the percentage values for the status of the Loan

DATA ANALYSIS

- Analyze the purpose of loan in total loans issued
- Derive separate columns for the necessary variables
- Get the description of the important Data such Loan Amount, Interest Rate, Loan Purpose and Loan Status.

UNIVARIATE ANALYSIS

- Loan Amount in Boxplot for univariate Analysis
- Description about total payment
- Remove the outliers for Annual Income as few of them are relevant for analyzing the data
- Analyzing for Interest Rate

CORRELATION METRICS

- As we have already converted the necessary string data, analyse it with Heatmap to get the exact correlation.
- Histogram Analysis of the variable loan_amnt , funded_amnt and Funded Amount Inv as these three are highly correlated

UNIVARIATE ANALYSIS

- Analyse the variables of Loan Amount , Loan Status , Interest Rate and Loan Purpose Individually

BIVARIATE ANALYSIS

- Here the Loan Status plays the important role as the revenue of the company is highly dependent on it.
- So, Higher the charged off, Higher the loss to the company.
- Therefore analysing the important variable with the Loan Status as "Charged Off" would give us the better result.
- So, Analysing the Loan Status with the important variables such as Loan Amount, Loan Purpose , Interest Rate , total principle received , Location – Address State, Revolving Line utilization and DTI (Debt To Income)

OMITTED COLUMNS

- As many columns were not having the relevant data to analyse properly , omitting those columns would give the smooth data for analysis.

MULTIVARIATE ANALYSIS

- Considering multiple variables Loan status , Loan Amount and Loan Purpose as the variables can help us in a 3 dimensional data analysis.
- Some of the plots used for multi-variate data analysis for the Lending Club is histplot and pairgrid

CONCLUSION

From the analysis, the conclusion/recommendation is:

1. Loan Amount provided is in higher range for the higher interest rate and the charged off is also high (Rec: Providing the appropriate Interest rate would avoid the charged Off)
2. Low Annual Income and High Interest rate has also lead to Charged Off (Rec: Providing low interest rate to the person with low income can decrease the “charged off”)
3. Less Loan Amount is provided when Annual Income is low. Loan status is Charged Off when high Loan Amount is provided for Low Annual Income (Rec: Providing less loan amount to low incomed person can avoid the “charged off”)
4. Charged Off is high when the Interest Rate is provided in higher range for the person with Low Annual income. (Rec : Low Interest to Low incomed person can avoid the “Charged Off”)

The above recommendation can help the company to avoid the loss to the revenue of the company.