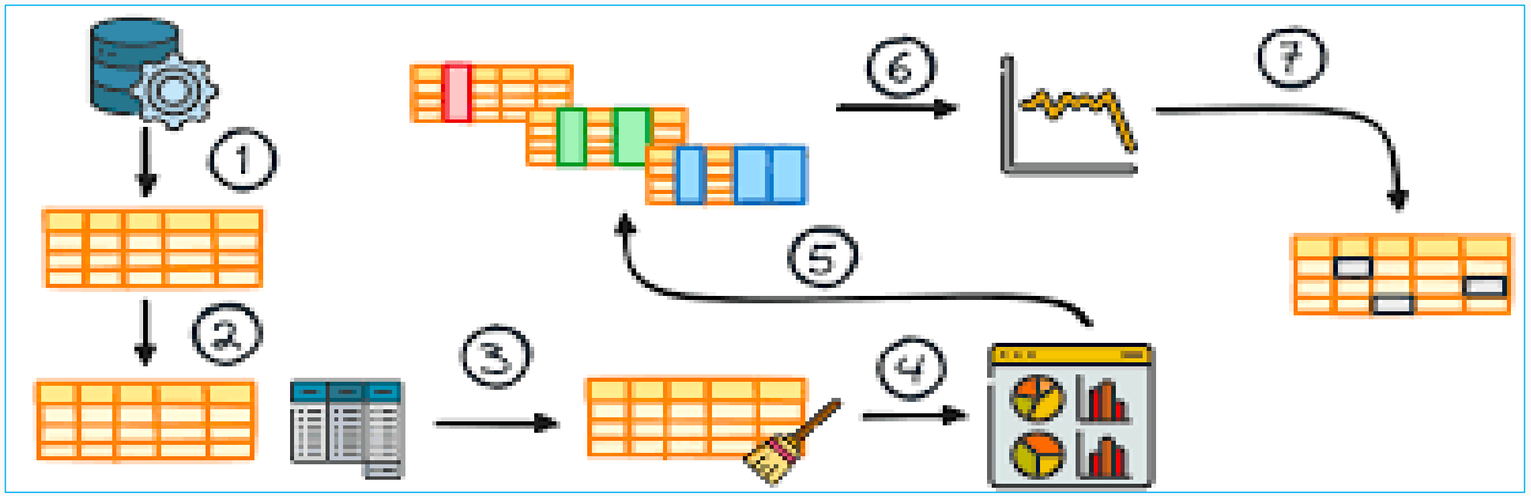
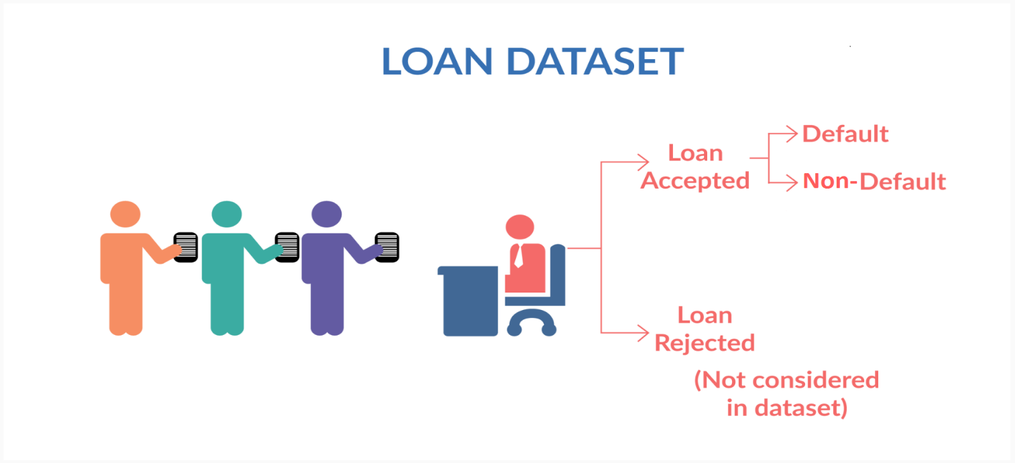


EDA for Credit Risk Analysis

A Case Study on Identifying Loan Defaulters and Their Factors

BIPUL RANJAN KUMAR
Learner: Data Analysis

ANKIT KUMAR
Learner: Data Analysis



Loan Default Insights

Introduction

A leading consumer finance company specializes in lending various types of loans to urban customers. They face the challenge of identifying potential defaulters before granting loans. This case study aims to use Exploratory Data Analysis (EDA) to understand the factors that influence loan default and develop a predictive model to mitigate risks.

01

Problem Statement

Identifying potential loan defaulters is vital for reducing financial risk for institutions.

02

Objective

Utilize Exploratory Data Analysis (EDA) to uncover factors influencing loan defaults.

03

Dataset Overview

Key variables include loan amount, tenure, Interest Rate, credit history, and income, which are critical for analysis.

Exploratory Data Analysis

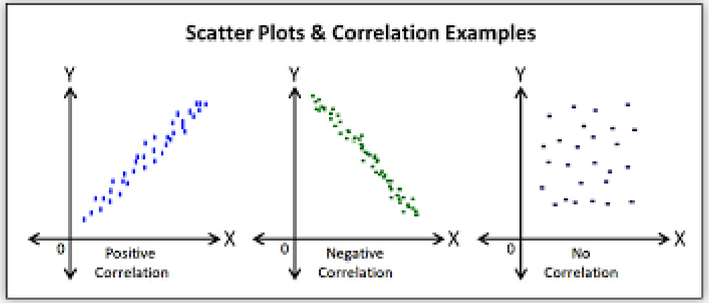
Exploratory Data Analysis

Key Steps and Insights in Credit Risk Assessment

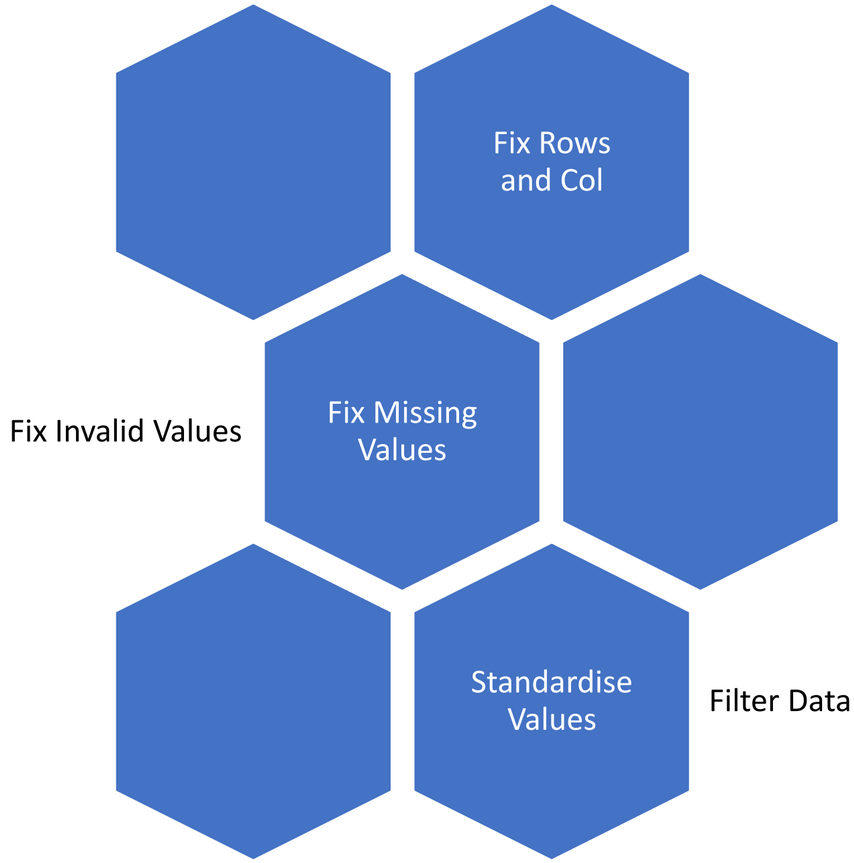
Data Cleaning and Preprocessing

Univariate Analysis

Bivariate Analysis



Data Cleaning and Preprocessing



Data Cleaning and Preprocessing

1

- ***Dimension of Dataframe = 39717 Rows and 111 Column***
  - Number of columns with all rows as NA = 54
    - All columns were dropped from dataframe
  - Three Columns have majority NA
    - mths_since_last_delinq [25682 rows null]
    - mths_since_last_record [36931 rows null]
    - next_pymnt_d [38577 rows null]
    - All three columns were dropped

2

- ***Dimension of Dataframe = 39717 Rows and 54 Column***
  - Dropped the column having only one values

pymnt_plan	policy_code
title	collections_12_mths_ex_med
initial_list_status	acc_now_delinq
application_type	chargeoff_within_12_mth
delinq_amnt	tax_liens

3

- ***Dimension of Dataframe = 39717 Rows and 44 Column***
  - Dropped the column not needed for analysis
    - URL
    - emp_title
    - desc

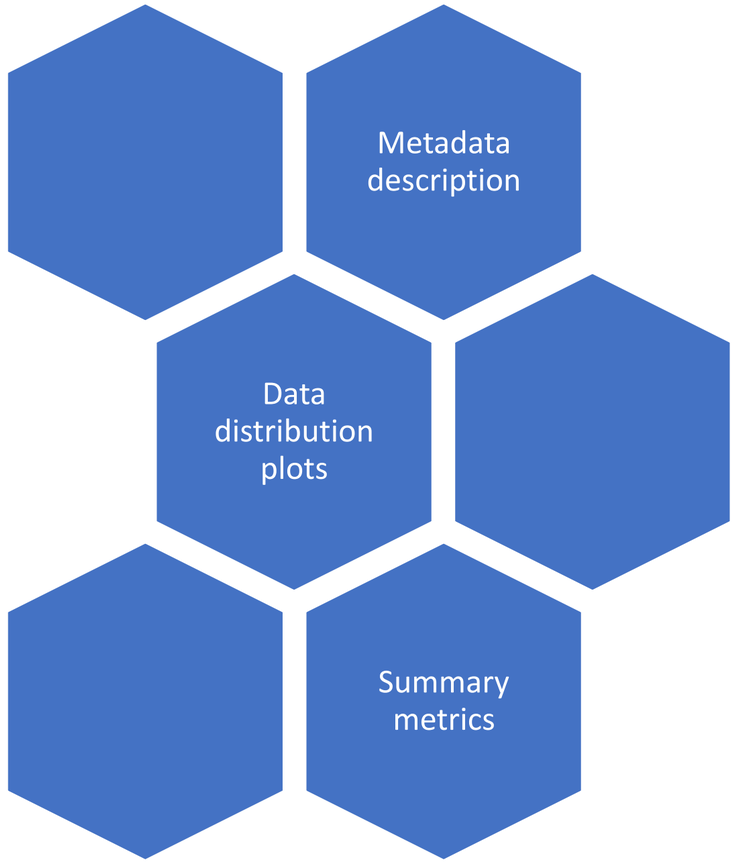
4

- ***Dimension of Dataframe = 39717 Rows and 41 Column***
  - Cleaning the dataframe
    - Removed 'months' suffix from term column
    - Removed '%' suffix from int_rate_percent column & revol_util_percent
    - Changed column issue_d, earliest_cr_line, last_pymnt_d and last_credit_pull_d to date[mmm-YY] format

Final Dimension of table

```
Dimension of dataframe (39717, 41)
Dataframe column having null values
 emp_length            1075
revol_util               50
last_pymnt_d             71
last_credit_pull_d        2
dtype: int64
```

Univariate Analysis

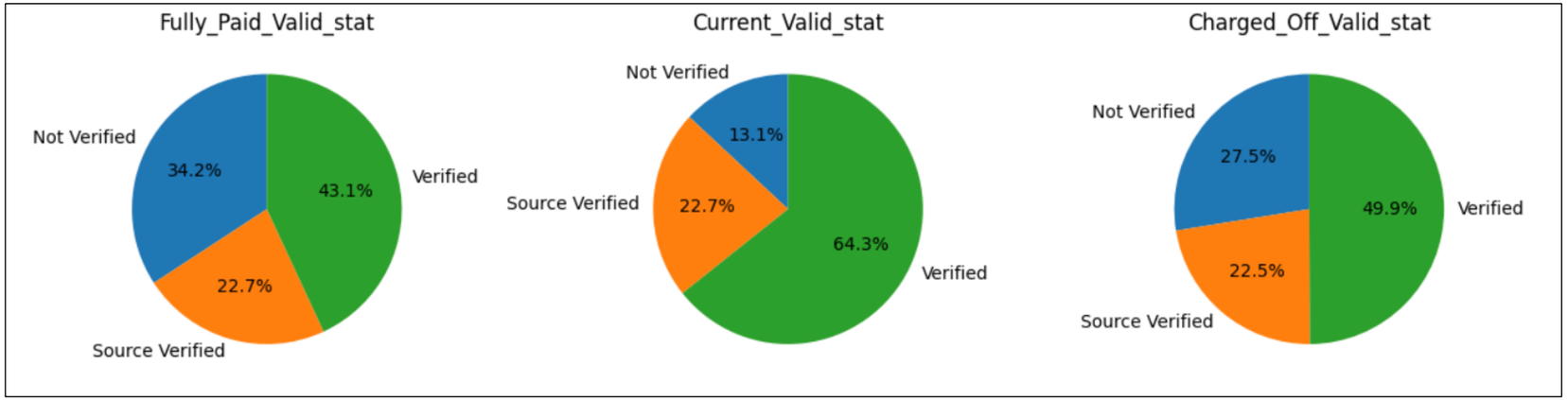


Univariate Analysis: Metadata description

#	Column Name	Categorical (Ord, Un-Ord)/ Numeric	Data Type	Description	Total-Number of rows	Missing	Mean	Std	Lower Whisker	25th Percentile	50th Percentile	75th Percentile	Upper Whisker
1	id	Numeric	int64	A unique LC assigned ID fo	39717	0	NA	NA	NA	NA	NA	NA	NA
2	member_id	Numeric	int64	A unique LC assigned Id fo	39717	0	NA	NA	NA	NA	NA	NA	NA
3	loan_amnt	Numeric	int64	The listed amount of the lo	39717	0	11219.44	7456.67	500.00	5500.00	10000.00	15000.00	35000.00
4	funded_amnt	Numeric	int64	The total amount committ	39717	0	10947.71	7187.24	500.00	5400.00	9600.00	15000.00	35000.00
5	funded_amnt_inv	Numeric	float64	The total amount committ	39717	0	10397.45	7128.45	0.00	5000.00	8975.00	14400.00	35000.00
6	installment	Numeric	float64	The monthly payment owe	39717	0	324.56	208.87	15.69	167.02	280.22	430.78	1305.19
7	grade	Ord-Caterogical	object	LC assigned loan grade	39717	0	NA	NA	NA	NA	NA	NA	NA
8	sub_grade	Ord-Caterogical	object	LC assigned loan subgrad	39717	0	NA	NA	NA	NA	NA	NA	NA
9	emp_length	Ord-Caterogical	object	Employment length in yea	38642	1075	NA	NA	NA	NA	NA	NA	NA
10	home_ownership	Un-Ord-Caterogical	object	The home ownership statu	39717	0	NA	NA	NA	NA	NA	NA	NA
11	annual_inc	Numeric	float64	The self-reported annual i	39717	0	68968.93	63793.77	4000.00	40404.00	59000.00	82300.00	6000000.00
12	verification_status	Un-Ord-Caterogical	object	Indicates if income was ve	39717	0	NA	NA	NA	NA	NA	NA	NA
13	issue_d	Ord-Caterogical	datetime64[ns]	The month which the loan	39717	0	NA	NA	NA	NA	NA	NA	NA
14	loan_status	Un-Ord-Caterogical	object	Current status of the loan	39717	0	NA	NA	NA	NA	NA	NA	NA
15	purpose	Un-Ord-Caterogical	object	A category provided by the	39717	0	NA	NA	NA	NA	NA	NA	NA
16	zip_code	Un-Ord-Caterogical	object	The first 3 numbers of the	39717	0	NA	NA	NA	NA	NA	NA	NA
17	addr_state	Un-Ord-Caterogical	object	The state provided by the b	39717	0	NA	NA	NA	NA	NA	NA	NA
18	dti	Numeric	float64	A ratio calculated using th	39717	0	13.32	6.68	0.00	8.17	13.40	18.60	29.99
19	delinq_2yrs	Numeric	int64	The number of 30+ days p	39717	0	0.15	0.49	0.00	0.00	0.00	0.00	11.00
20	earliest_cr_line	Ord-Caterogical	datetime64[ns]	The month the borrower's	39717	0	NA	NA	NA	NA	NA	NA	NA
21	inq_last_6mths	Numeric	int64	The number of inquiries in	39717	0	0.87	1.07	0.00	0.00	1.00	1.00	8.00
22	open_acc	Numeric	int64	The number of open credi	39717	0	9.29	4.40	2.00	6.00	9.00	12.00	44.00
23	pub_rec	Numeric	int64	Number of derogatory pub	39717	0	0.06	0.24	0.00	0.00	0.00	0.00	4.00
24	revol_bal	Numeric	int64	Total credit revolving bala	39717	0	13382.53	15885.02	0.00	3703.00	8850.00	17058.00	149588.00
25	total_acc	Numeric	int64	The total number of credit	39717	0	22.09	11.40	2.00	13.00	20.00	29.00	90.00
26	out_prncp	Numeric	float64	Remaining outstanding pr	39717	0	51.23	375.17	0.00	0.00	0.00	0.00	6311.47
27	out_prncp_inv	Numeric	float64	Remaining outstanding pr	39717	0	50.99	373.82	0.00	0.00	0.00	0.00	6307.37
28	total_pymnt	Numeric	float64	Payments received to date	39717	0	12153.60	9042.04	0.00	5576.93	9899.64	16534.43	58563.68
29	total_pymnt_inv	Numeric	float64	Payments received to date	39717	0	11567.15	8942.67	0.00	5112.31	9287.15	15798.81	58563.68
30	total_rec_prncp	Numeric	float64	Principal received to date	39717	0	9793.35	7065.52	0.00	4600.00	8000.00	13653.26	35000.02
31	total_rec_int	Numeric	float64	Interest received to date	39717	0	2263.66	2608.11	0.00	662.18	1348.91	2833.40	23563.68
32	total_rec_late_fee	Numeric	float64	Late fees received to date	39717	0	1.36	7.29	0.00	0.00	0.00	0.00	180.20
33	recoveries	Numeric	float64	post charge off gross reco	39717	0	95.22	688.74	0.00	0.00	0.00	0.00	29623.35
34	collection_recovery_fee	Numeric	float64	post charge off collection	39717	0	12.41	148.67	0.00	0.00	0.00	0.00	7002.19
35	last_pymnt_d	Ord-Caterogical	datetime64[ns]	Last month payment was r	39646	71	NA	NA	NA	NA	NA	NA	NA
36	last_pymnt_amnt	Numeric	float64	Last total payment amoun	39717	0	2678.83	4447.14	0.00	218.68	546.14	3293.16	36115.20
37	last_credit_pull_d	Ord-Caterogical	datetime64[ns]	The most recent month LC	39715	2	NA	NA	NA	NA	NA	NA	NA
38	pub_rec_bankruptcies	Numeric	float64	Number of public record b	39717	0	0.04	0.20	0.00	0.00	0.00	0.00	2.00
39	term_in_month	Numeric	int64	The number of payments o	39717	0	42.42	10.62	36.00	36.00	36.00	60.00	60.00
40	int_rate_percent	Numeric	float64	Interest Rate on the loan	39717	0	12.02	3.72	5.42	9.25	11.86	14.59	24.59
41	revol_util_percent	Numeric	float64	Revolving line utilization ra	39667	50	48.83	28.33	0.00	25.40	49.30	72.40	99.90

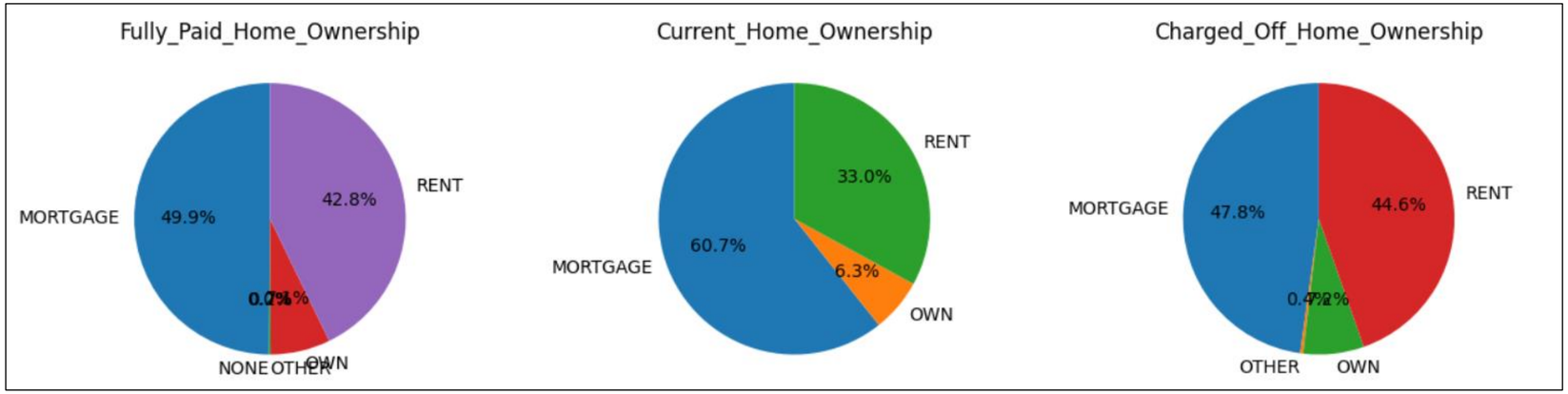
Univariate Analysis: Data Plots : Piecharts

Analysis of validation_status column



No Trend observed

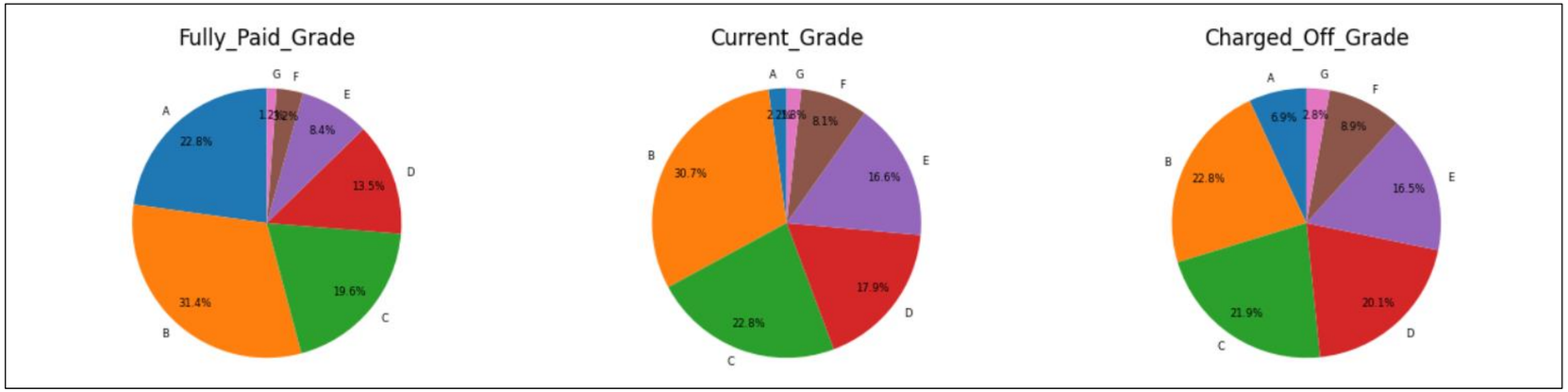
Analysis of home_Ownership column



No Trend observed

Univariate Analysis: Data Plots : Piecharts

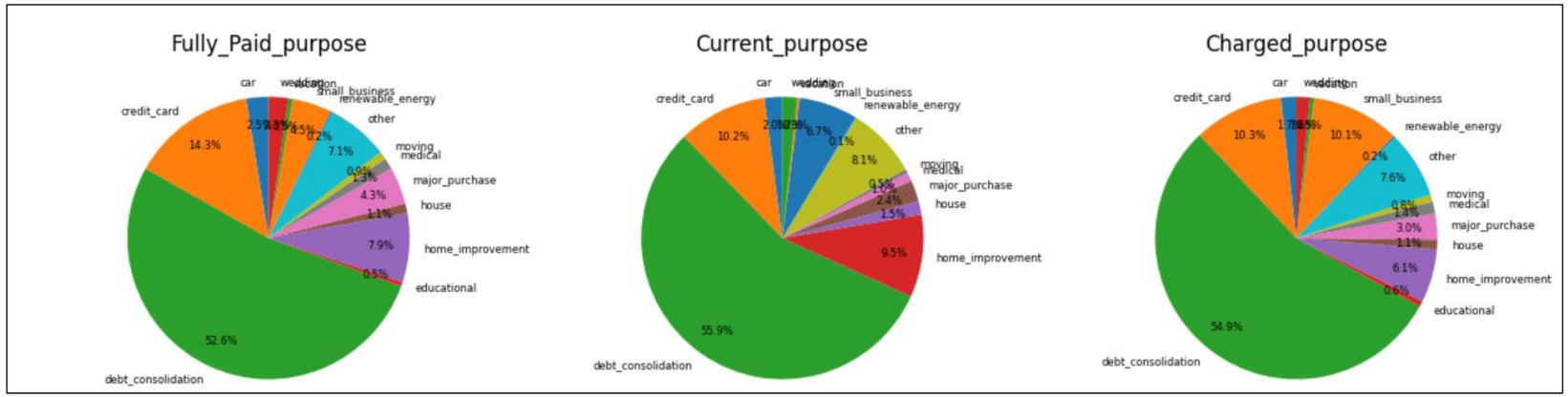
Analysis of grade column



Fully Paid cumulative % of A+B+C Grade is 73.8%

Charged off cumulative % of A+B+C Grade is **51.6%**

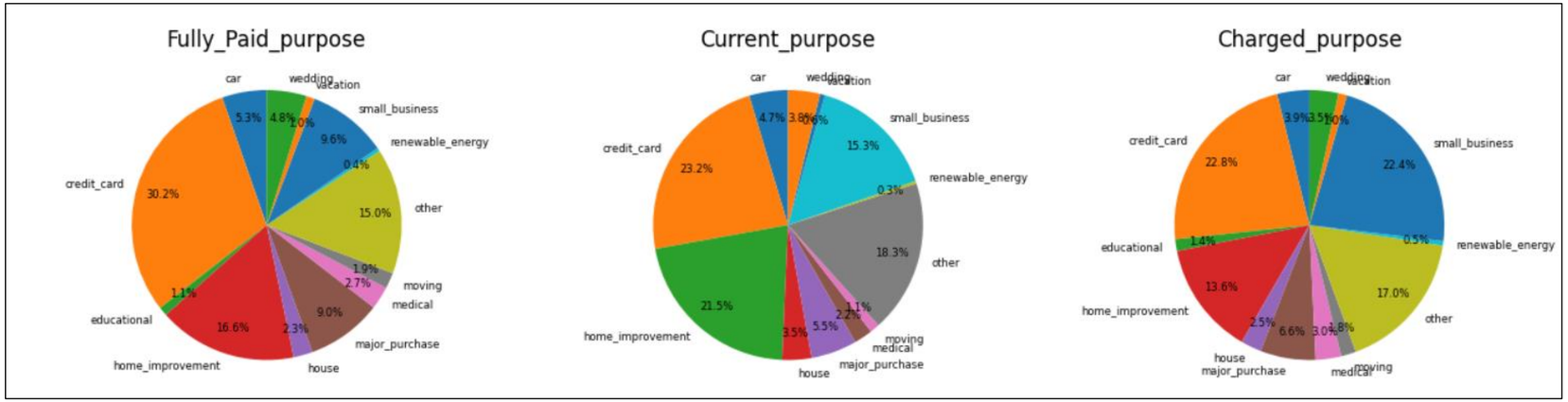
Analysis of Purpose column



Debt_Consolidation has the highest % [54.9%] for the Charged off

Univariate Analysis: Data Plots : Piecharts

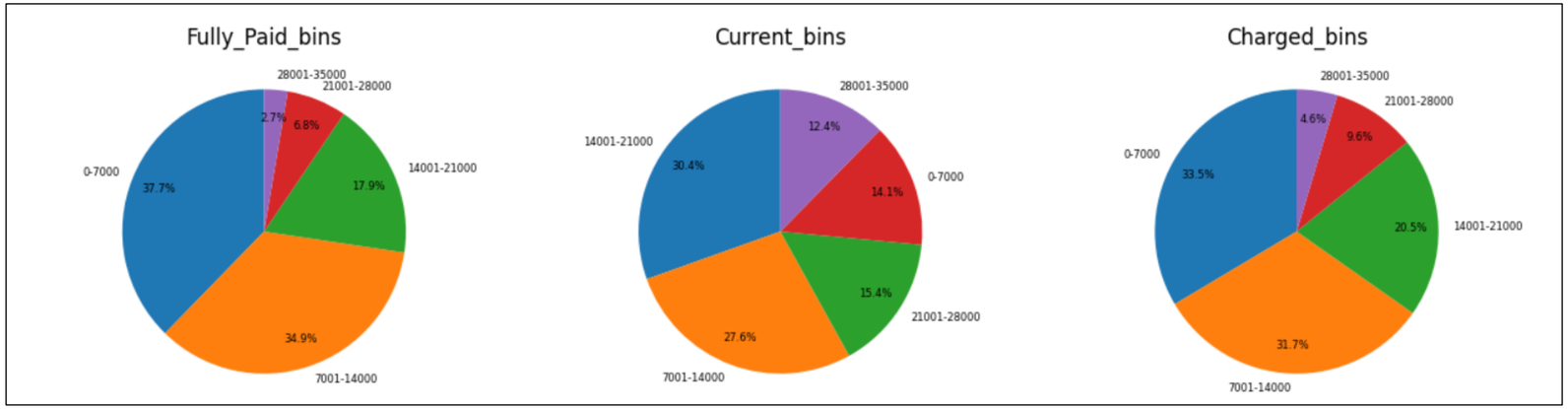
Analysis of Purpose column [without debt_consolidation purpose]



Fully Paid for small business is 9.6%

Charged off for small business is 22.4 %

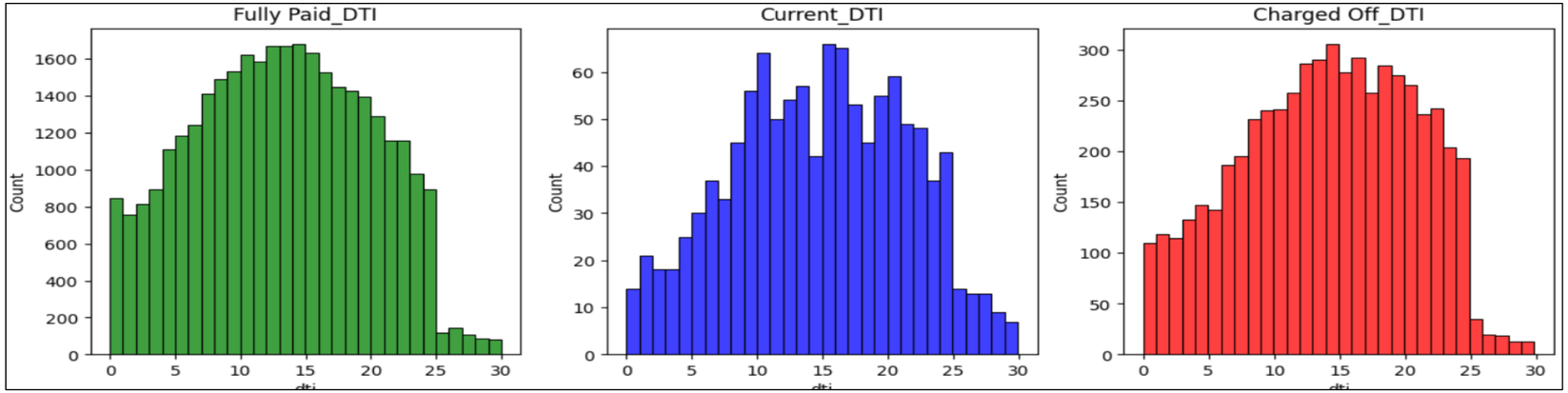
Analysis loan amount column



80% of defaulter is having loan amount less than 21000

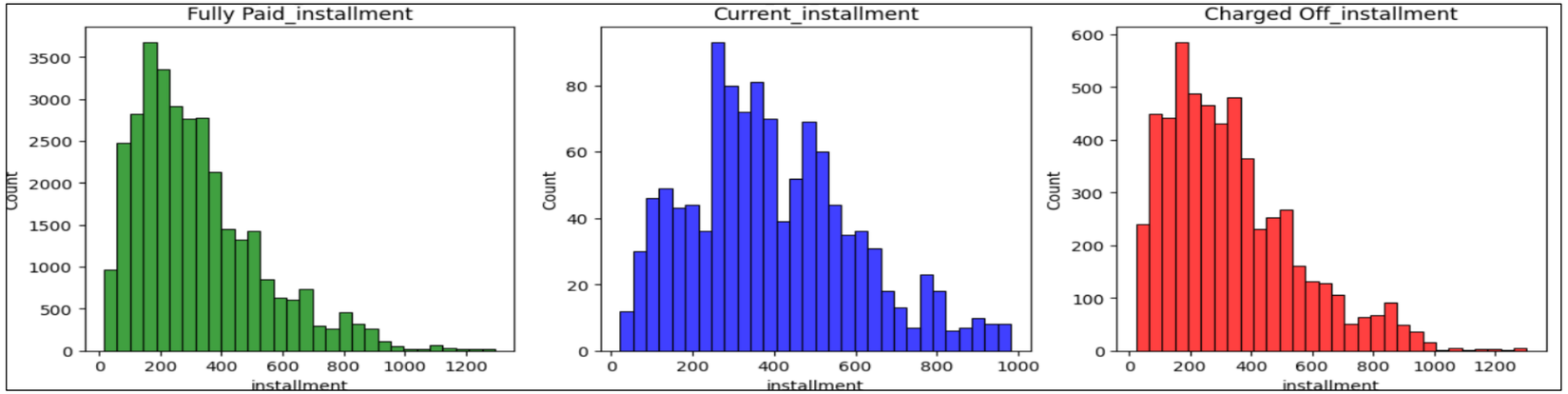
Univariate Analysis: Data Plots : Histogram

Analysis of DTI column



No Trend observed

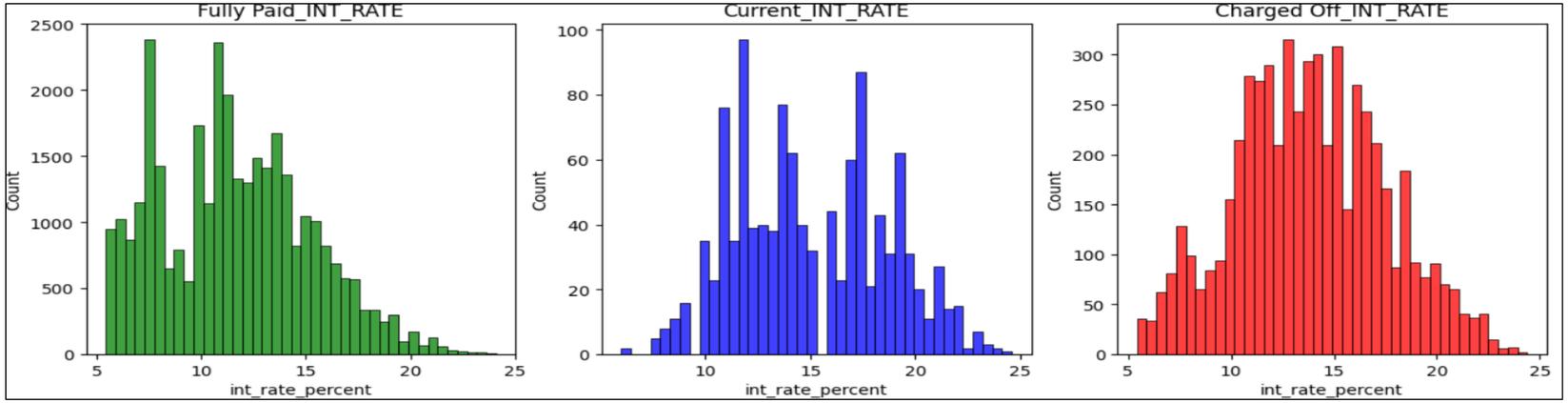
Analysis of Installment column



No Trend observed

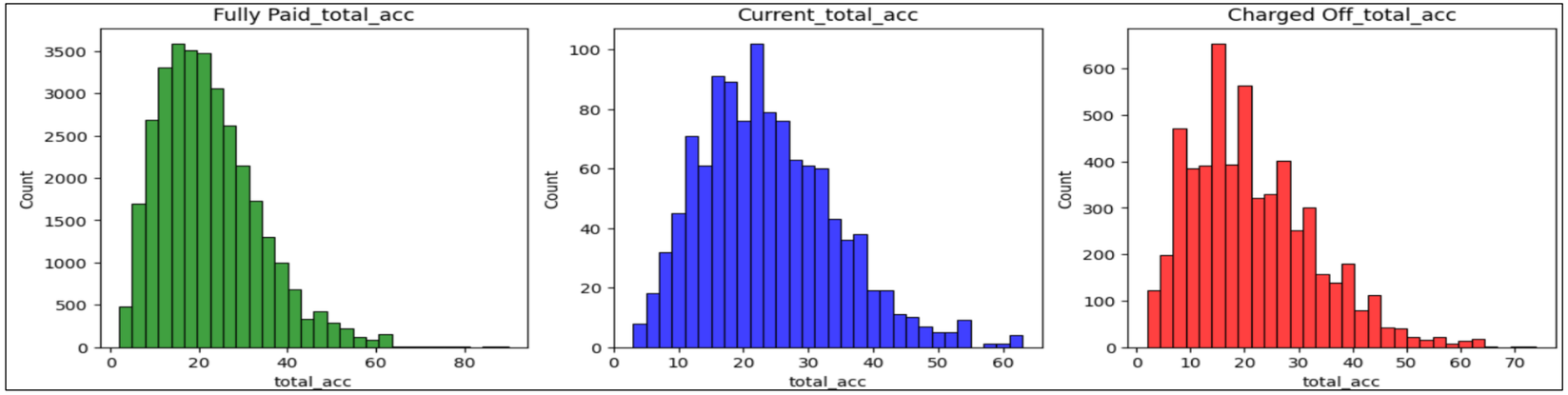
Univariate Analysis: Data Plots : Histogram

Analysis of Interest Rate column



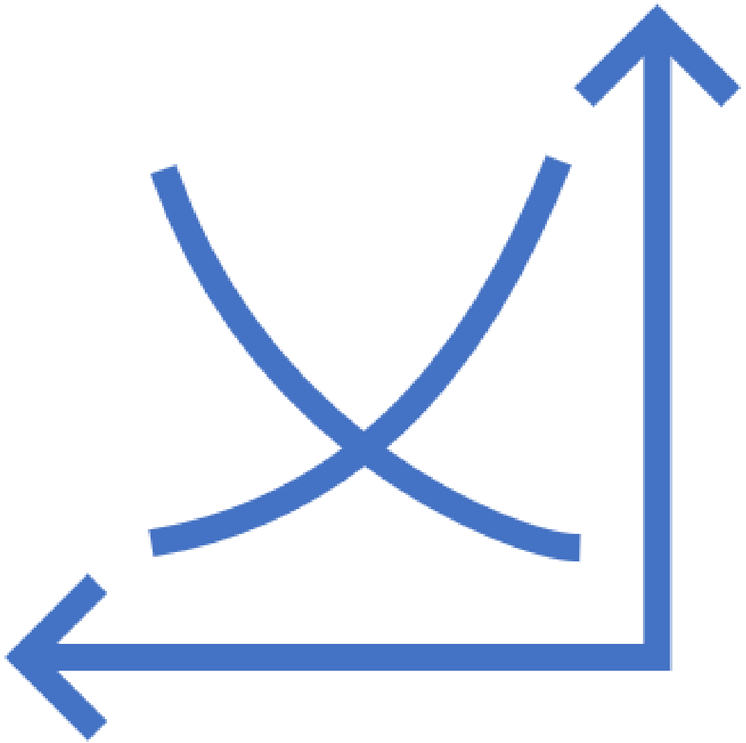
Between Interest rate **11-17%** is the highest defaulter

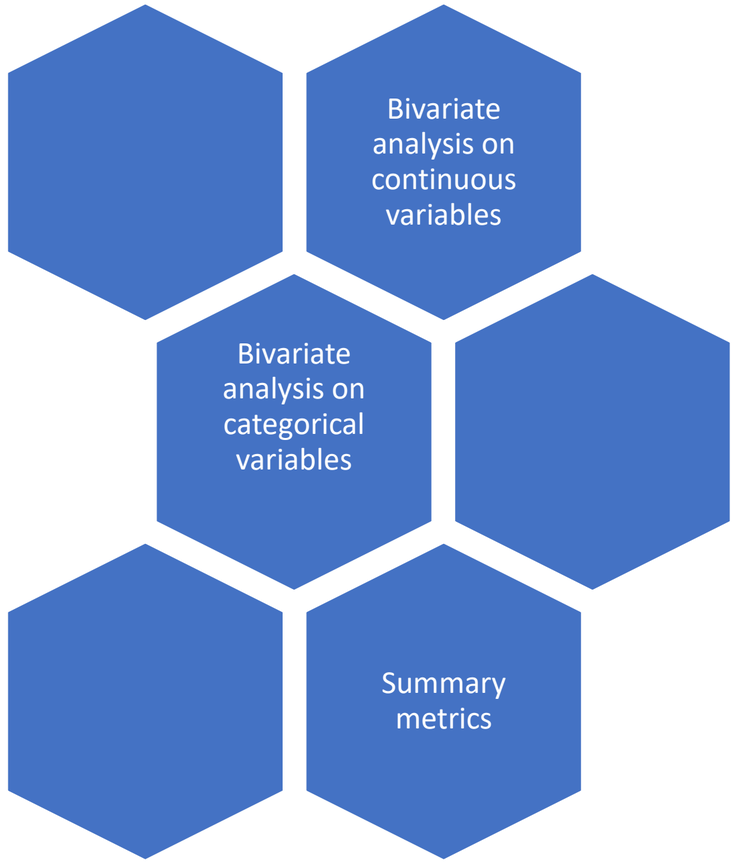
Analysis of total_acc column



No Trend observed

Bivariate Analysis

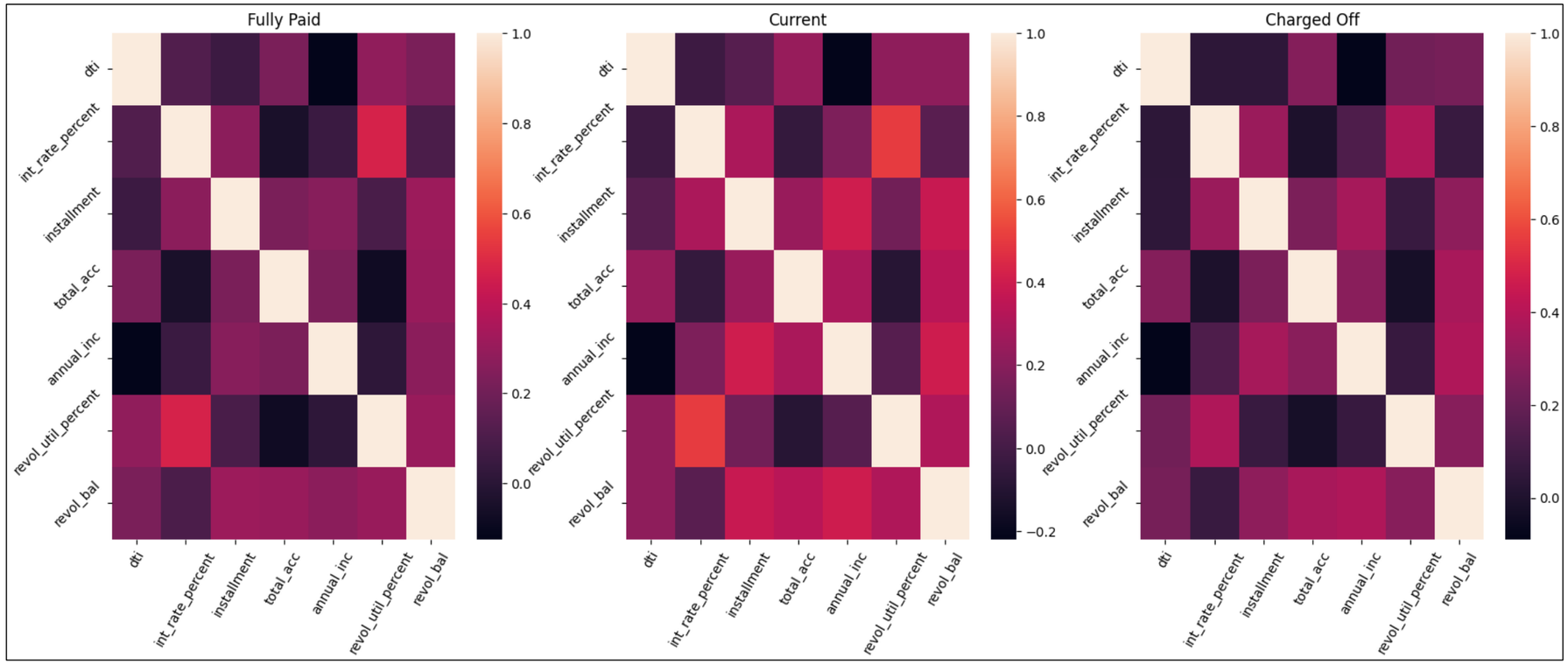




Bivariate Analysis: Data Plots : HeatMap

Analysis of correlation between column s

dti','int_rate_percent','installment','total_acc','annual_inc','revol_util_percent','revol_bal'



No Correlation observed

- Loan Grade D,E,F and G has higher % of Defaulters ~50% in comparision of ~28% for fully paid
- Loan taken for “Small Business” purpose has maximum defaulters ~22% [After removing Debt Consideration Type]
- Interest rate 11-17% has highest % of Defaulters

Call to Action

Implement Findings to Enhance Credit Risk Assessments Today!