

Gramener Case Study (Lending loan)

Exploratory Data Analysis (EDA)

SUBMISSION

Group Name:

1. Nyamath Shaik
2. Kapil Sharma
3. Manoj Tiwari
4. Naga Prasad

- Introduction
- Business Understanding & Objectives
- Methodology
- Data Cleaning & Manipulation
- Descriptive Analysis
 - Key Variables (Univariate Analysis)
 - Loan Amount Vs Interest Rate (Bivariate Analysis)
 - Loan Purpose (Univariate Analysis)
 - Grade Vs Interest Rate (Bivariate Analysis)
- Hypothesis Testing
- Insights
- Conclusions

- The prediction of defaulting loan in future by consumer is challenging task by financial institutions
- Techniques used to perform EDA: CRISP-DM Framework
- Analysis Performed on data provided:
 - Univariate Analysis
 - Segmented Univariate Analysis
 - Bivariate Analysis
- This case study aims to minimize business loss by finding out important variables which are strong indicators of default.



Business Understanding & Objective

Business Understanding

Consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

- 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

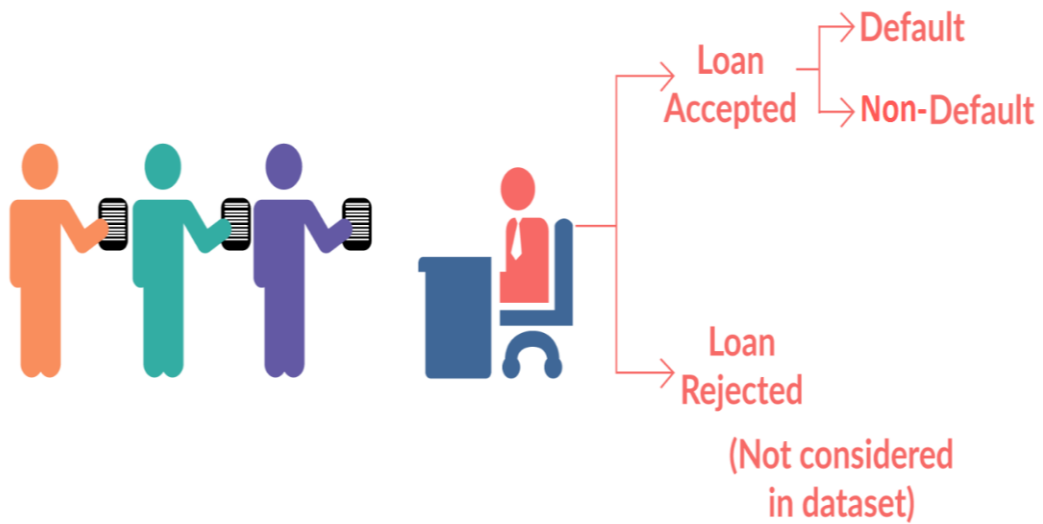
Business objective

The company is the largest online loan marketplace, giving different types of loans like personal, business, educational, medical procedure etc. Avoid lending loans to risky applicants to avoid financial loss. Company wants to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default.

Business Decision Points

- Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
 - **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)

Loan Dataset

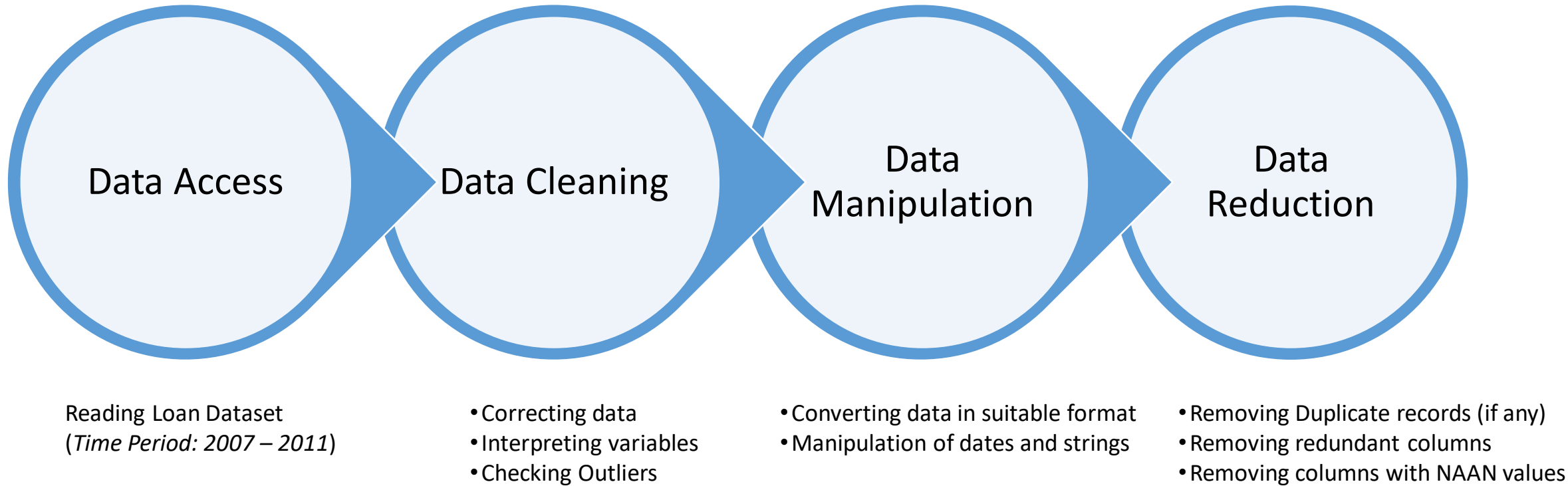


CRISP-DM FRAMEWORK



Tools used

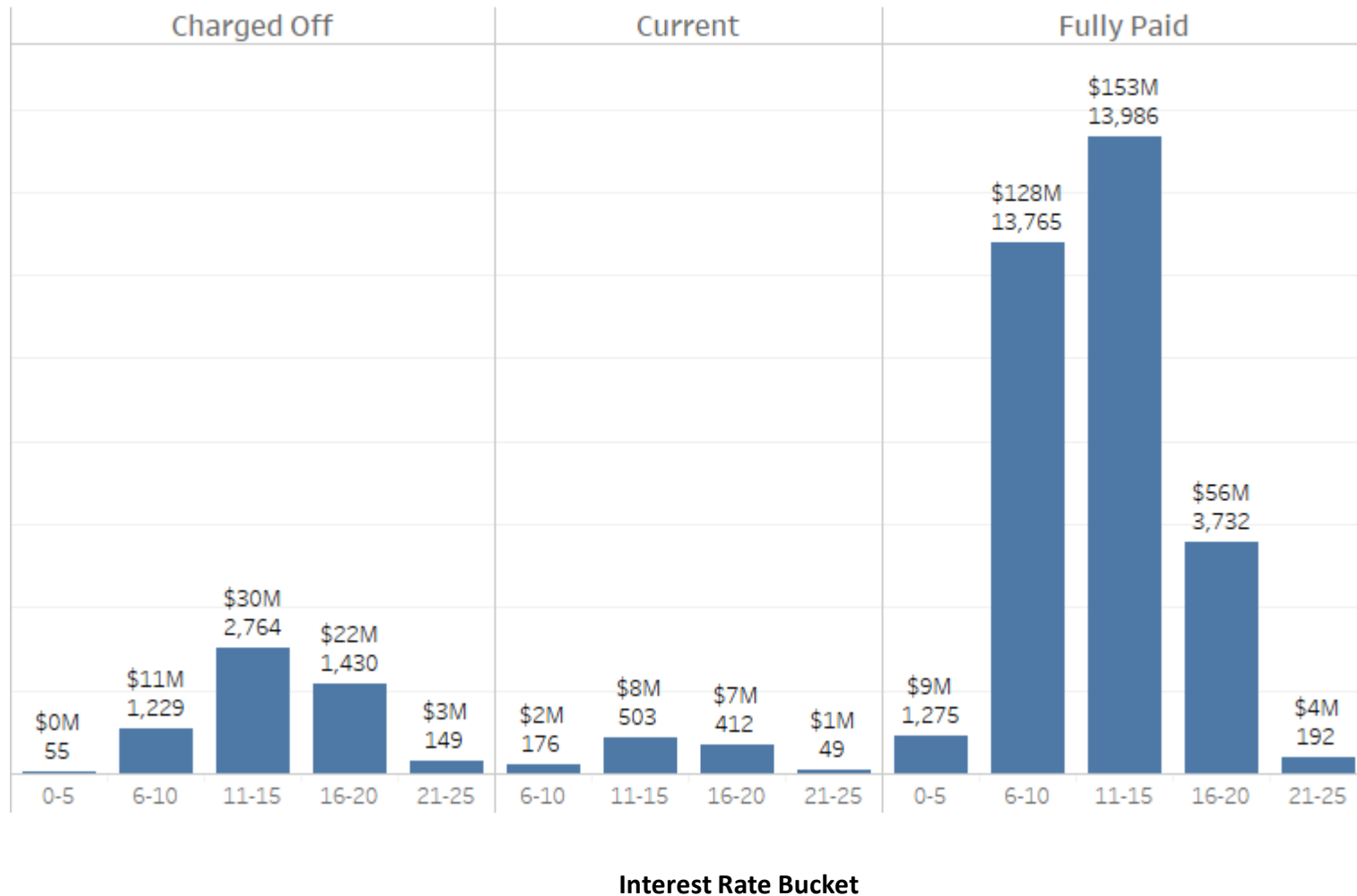




- Loan Amount Vs Interest Rate
- Loan Status
- Loan Purpose
- Interest
- Term



Descriptive Analysis – Loan Amount Vs Interest Rate



Finding 1: Loan Status

- 83.2% people are paid the loan
- 14% People are Rejected the Loan amount
- 2.7% people are paying the loan

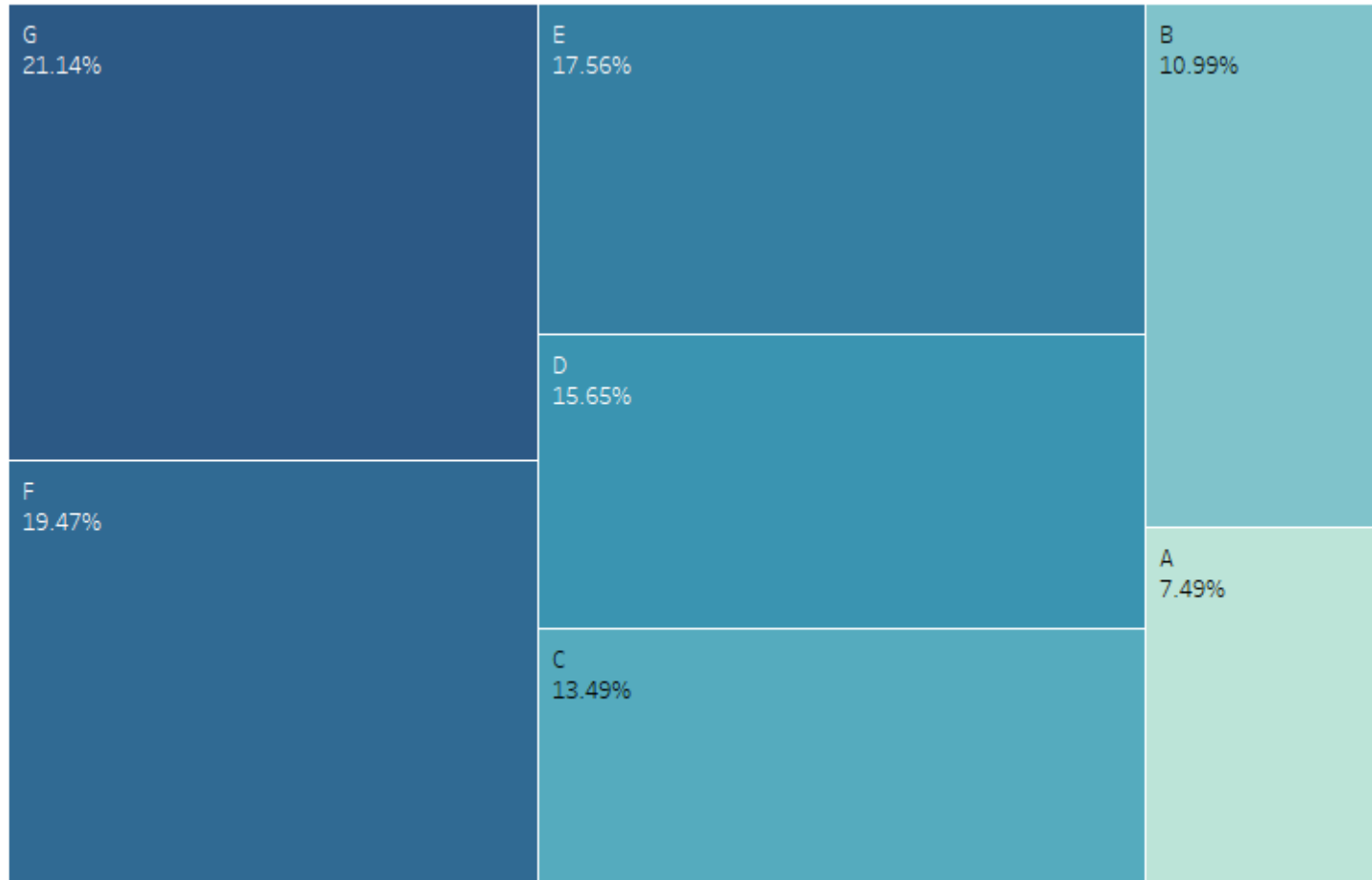


Descriptive Analysis – Loan Purpose (Word Cloud)



Finding 2: Top 5 Loan Purpose (Excludes Other)

- Debt Consolidation
- Credit Card
- Home Improvement
- Small Business
- Major Purchase



Finding 3: Interest Rate Findings

- Interest rate is lower near A
- Interest rate is higher near G
- Higher the grade the higher was the interest rate



Hypothesis Testing – Code and Output

```
# Critical Value Method
# finding critical values
import math
population_mean=df.loan_amnt.mean()
print("mean=",population_mean)
population_std_deviation=df.loan_amnt.std()
print("standard_deviation=",population_std_deviation)
numer_samples=df.shape[0]
print("number_samples=",numer_samples)
Standard_Error=(population_std_deviation/math.sqrt(numer_samples))
print("SE=",Standard_Error)
```

```
mean= 10741.796149081149
standard_deviation= 6757.325936240309
number_samples= 38744
SE= 34.32990798558645
```

```
#95% Confidence Level
Zc=1.645
UCV=population_mean+(Zc*Standard_Error)
LCV=population_mean-(Zc*Standard_Error)
print("Estimated Range in which population mean is:")
print("UCV=",UCV)
print("LCV=",LCV)
```

```
Estimated Range in which population mean is:
UCV= 10798.268847717438
LCV= 10685.32345044486
```

```
# t-test : testing the value of a population mean
#
mean_value=df.loan_amnt.mean()
stats.ttest_1samp(df.loan_amnt,round(df.loan_amnt.mean(),3))
```

```
Ttest_1sampResult(statistic=4.342602617169559e-06, pvalue=0.9999965351267762)
```

```
# sample t-test : testing for difference across populations
stats.ttest_ind(df.loan_amnt,df.annual_inc)
```

```
Ttest_indResult(statistic=-295.668439095428, pvalue=0.0)
```

```
# Paired tests : repeated measurement on the same individuals
stats.ttest_rel(df.loan_amnt,df.annual_inc)
```

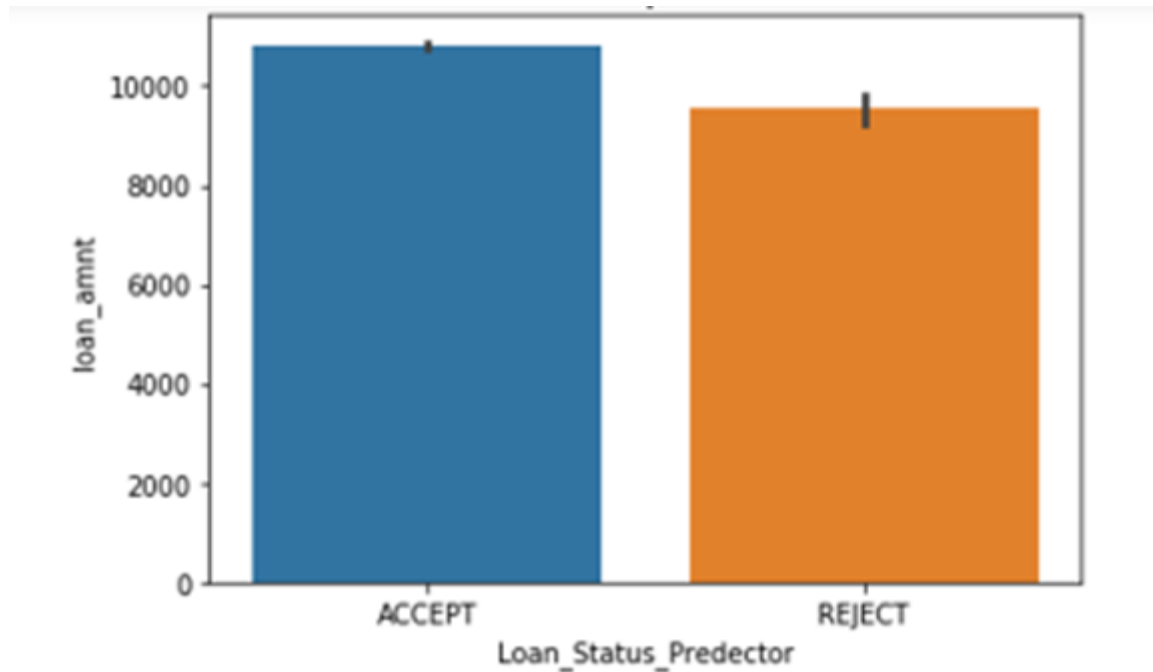
```
Ttest_relResult(statistic=-319.17714136289146, pvalue=0.0)
```

```
stats.zscore(df.loan_amnt)
```

```
array([-0.84972522, -1.21969883, -1.23449777, ..., -0.84972522,
       -0.84972522, -0.47975161])
```

- The higher grade indicates high risk rate and also high interest rate
- Higher interest rates leads to higher risk on lending loan
- Loan term of 60 month is around 3 times risk as 36 months loan term
- Top purpose of loan is Debt Consolidation
- Derived Metrics:
 - Total Credit
 - Revolving Balance
 - Required Credit
 - Total Credit

- If the Required credit is less than Total Credit then the loan can be Accepted else it's ok to reject



Total loan Acceptance Status

