slideshare
a Scribd company

- Home
- Explore

Search Submit Search

- Upload
- s
  You

- ssuser74138d

- My Clipboards
- My Uploads
- My Comments
- Analytics
- Account Settings
- Account Settings

- Support
- Logout

s
You ssuser74138d

- Home
- Explore

- My Clipboards
- Account Settings

- Support
- Logout

Search
×Oops! Something went wrong. Please try again later.
Successfully reported this slideshow.

Credit EDA Case Study : Exploratory Data Analysis on Bank Loan Data

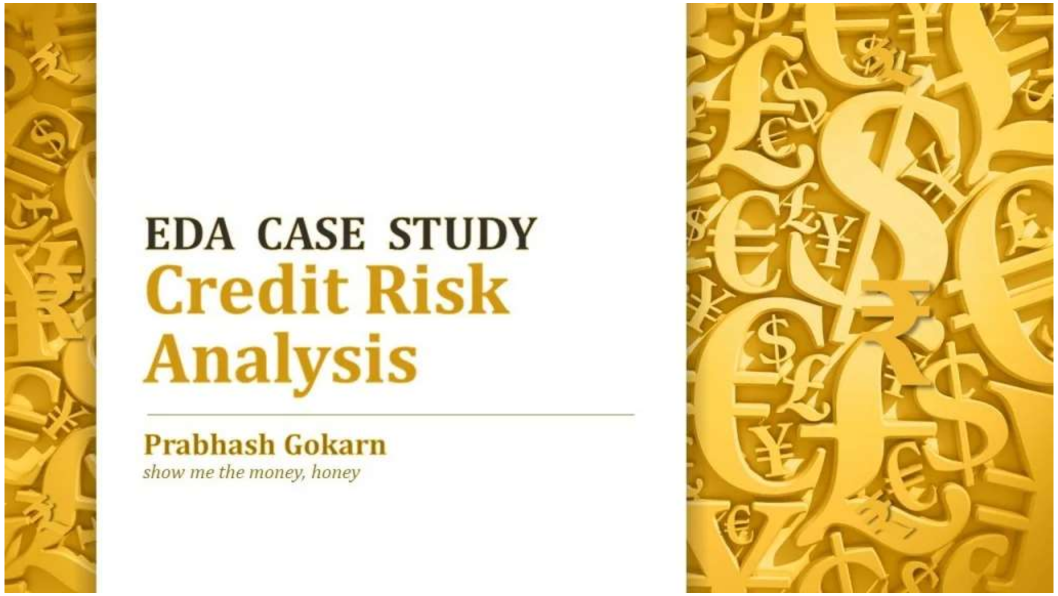
2

Share

PRABHASH GOKARN
•
Nov. 04, 2020
• 2 likes • 3,286 views
Download Now Download

•

1



Problem Statement

There are two types of risks associated with any loan request :

- **H0** If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- **H1** 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.

Analysis of the data set has been done in Python on a Jupyter Notebook.

2

Analysis Done

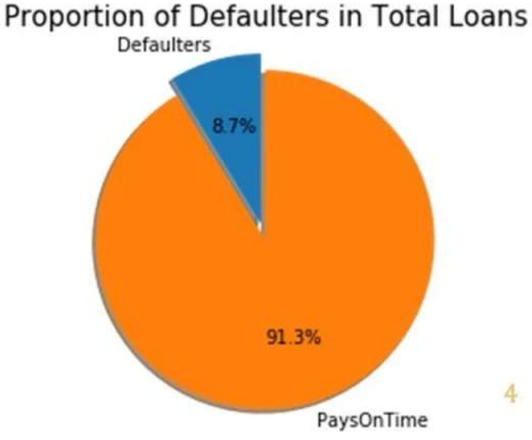
Steps :

- Check Missing Values a. which to handle, how to handle
- Check Outliers; check data imbalance; ratio of imbalance
- Top 10 correlation for the Client with payment difficulties other variables within Application DF & Previous App DF
- Which correlation is most relevant

3

Proportion of Defaulters in the Data Set

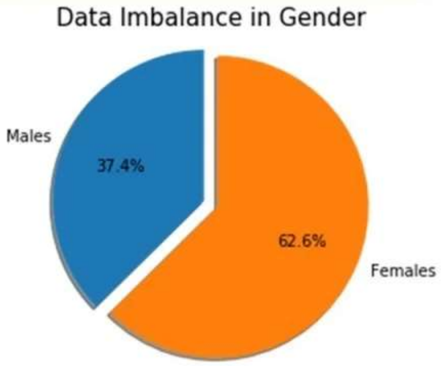
- Defaulters are 8.7% of total



4

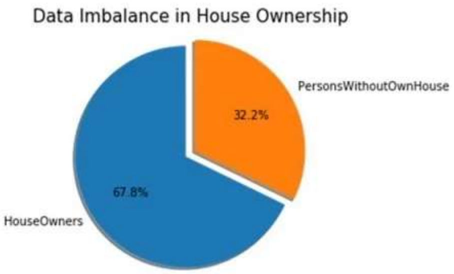
Gender Imbalance

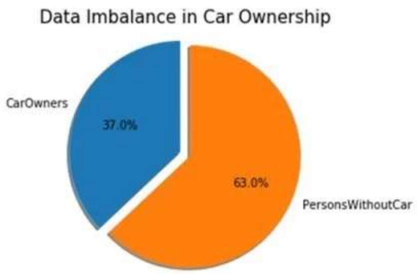
- 63% Females



5

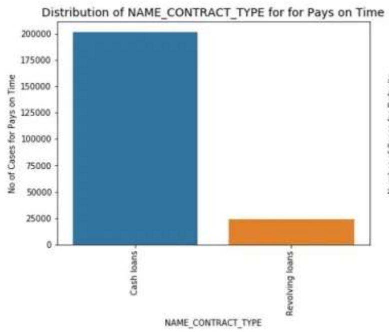
House and Car Ownership (i.e. Gaadi & Baadi)

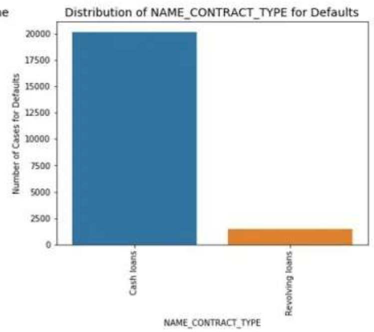




6

Proportion of Defaulters in the Data Set





- Less defaulters for Revolving Loans

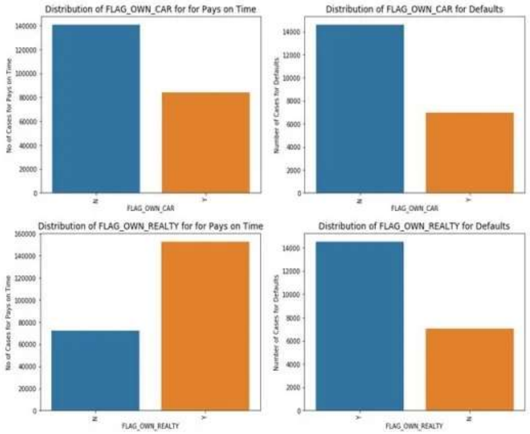
7

Proportion of Defaulters By Gender

- Females in defaults is higher

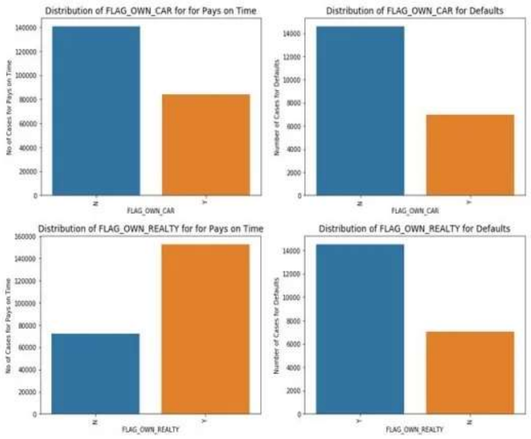
Proportion of Defaulters By Gender

- Own Houses default similar to people without houses, but own car defaults less



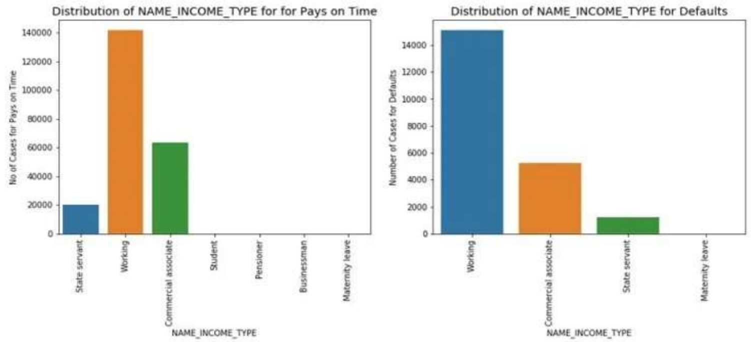
Proportion of Defaulters By Gaadi / Baadi

- Own Houses default similar to people without houses, but own car defaults less



10

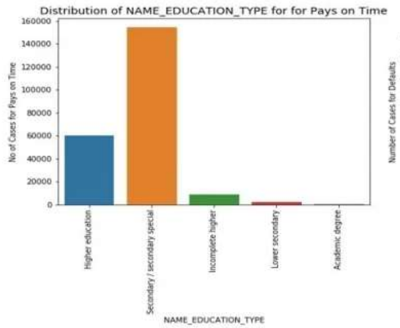
Proportion of Defaulters By Profession

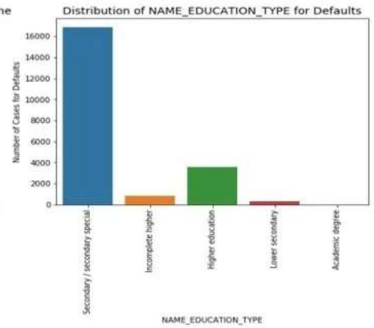


- Proportion of Working Population in defaults is higher and State Servants is lower in Defaults

11

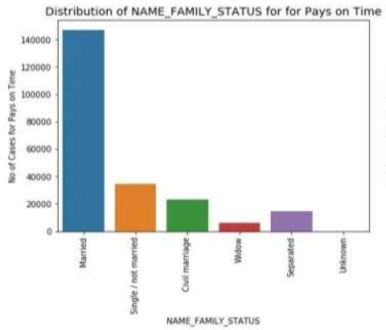
Proportion of Defaulters By Education

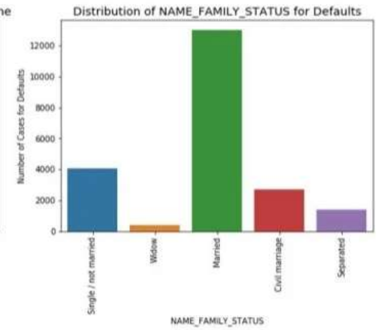




- Higher Education default less

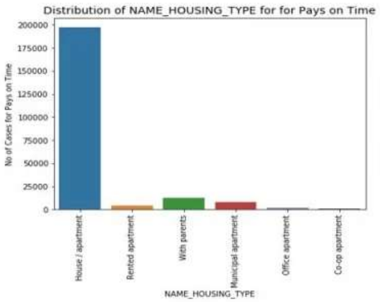
Proportion of Defaulters By Family Status

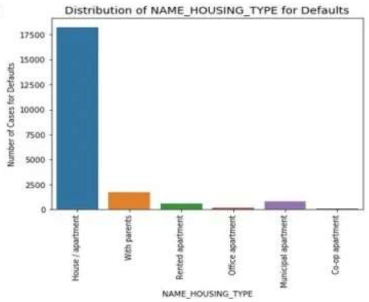




- Married people default less

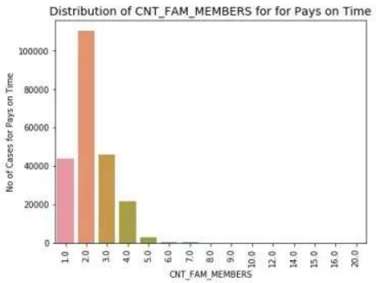
Proportion of Defaulters By House Status

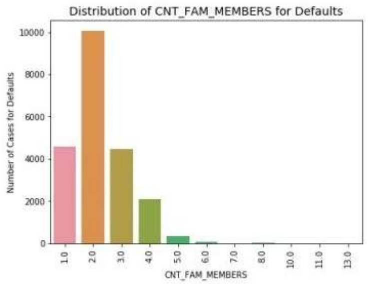




- People staying in own house default less

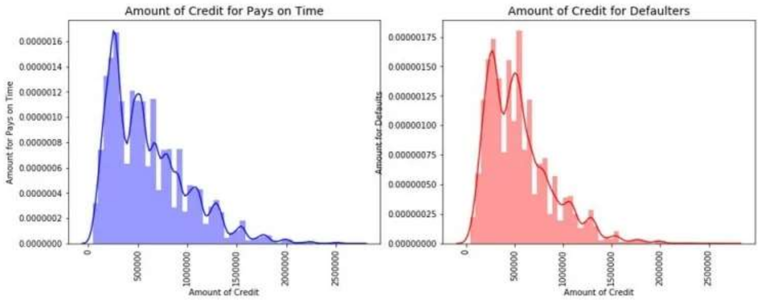
Proportion of Defaulters By No of Kids





- No difference in Count of Family Members

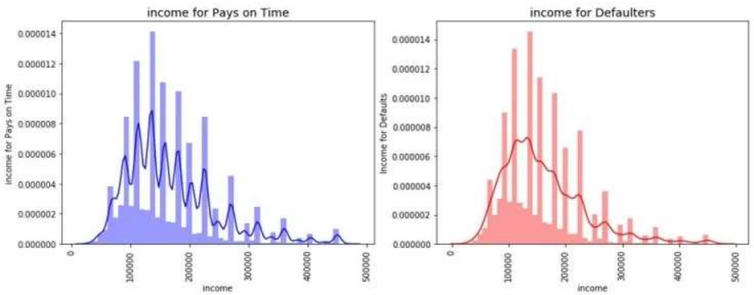
Proportion of Defaulters By Credit Amount



- Better return of loans as loan size increases

16

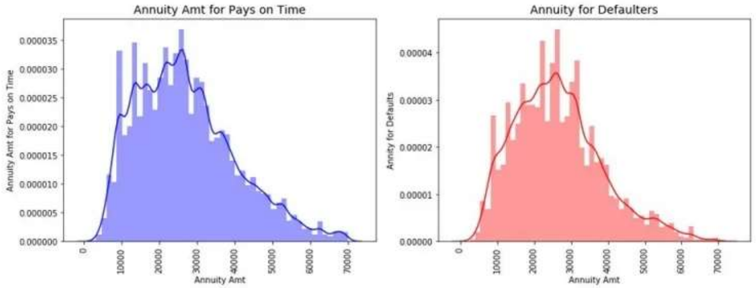
Proportion of Defaulters By Income Level



- Better return of loans as Income increases
- Very Unlike in India where Middle Class Returns Loans and Rich like Mallya flee

17

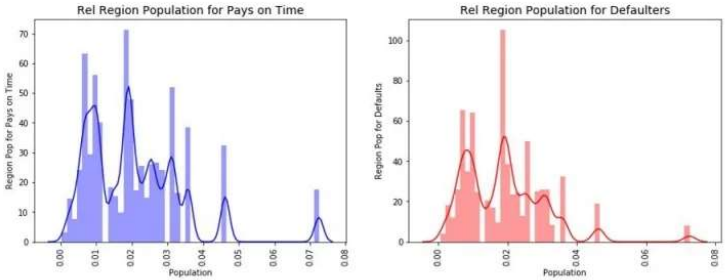
Proportion of Defaulters By Annuity Amt



- Better return of loans as Amt of Annuity Increases.

18

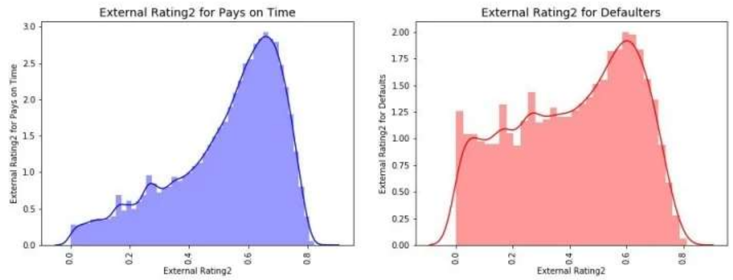
Proportion of Defaulters By Rel Region



- Better return of loans for higher Relative Region Population

19

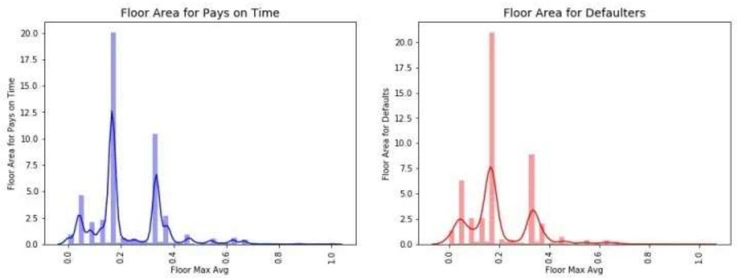
Proportion of Defaulters By External Rating



- Better Rating – Better Return

20

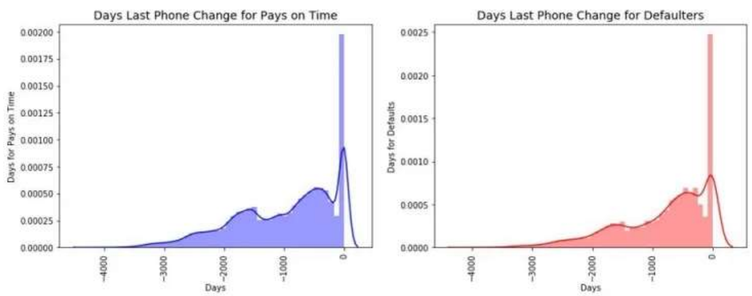
Proportion of Defaulters By No of Kids



- Lower Floor area – More default

21

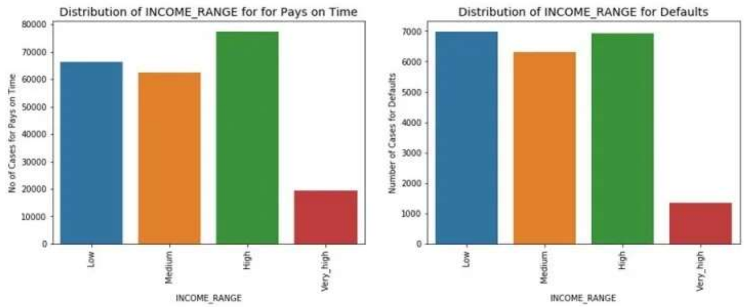
Proportion of Defaulters By No of Kids



- More defaults if phone changed more recently

22

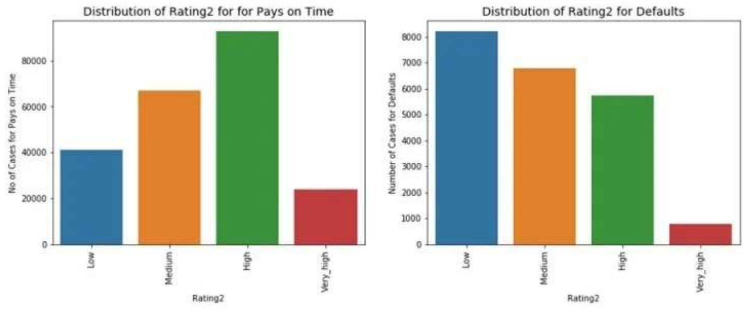
Proportion of Defaulters By Income



- Fewer Defaults when Income Range is higher

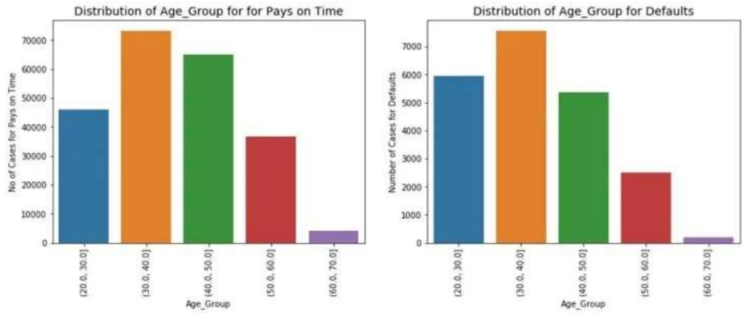
23

Proportion of Defaulters By Rating



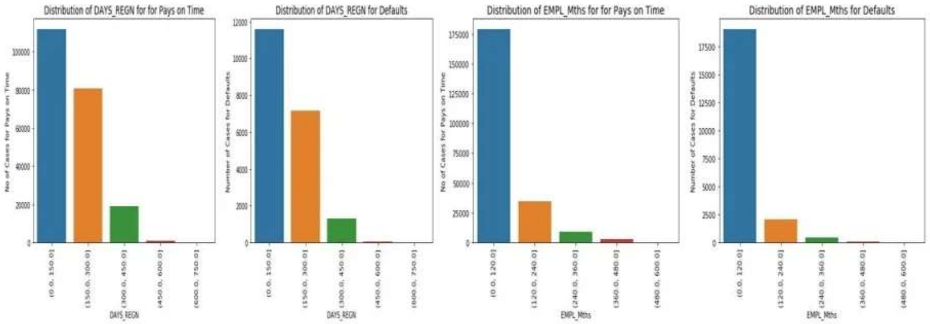
- Better Rating Better Return

Proportion of Defaulters By Age



- Older People – more likely to return

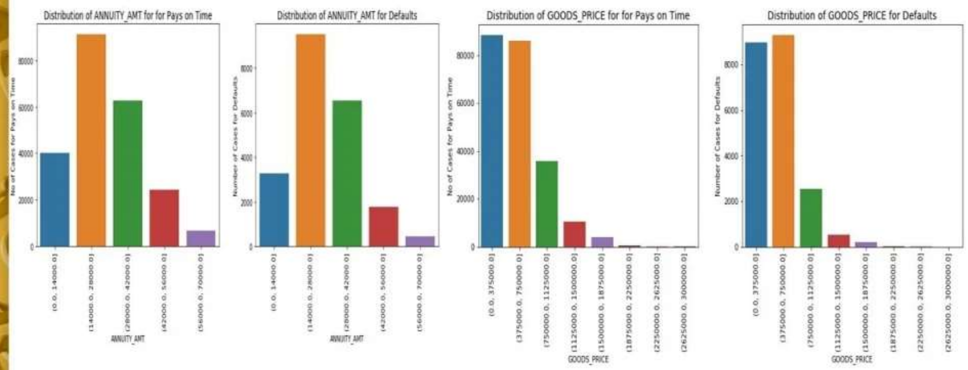
Proportion of Defaulters By Days Regn, Emp



- Recent Employment or Recent Change in Regn – Defaults more

26

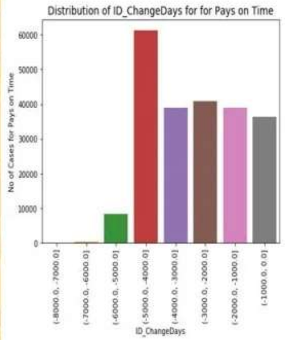
Proportion of Defaulters By Annuity Amt, Goods Price

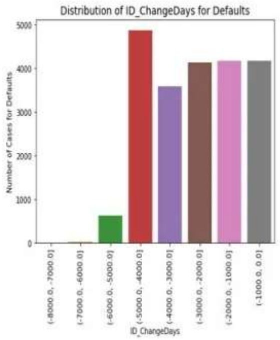


- Annuity Amount, Goods Price lower more at risk for default

27

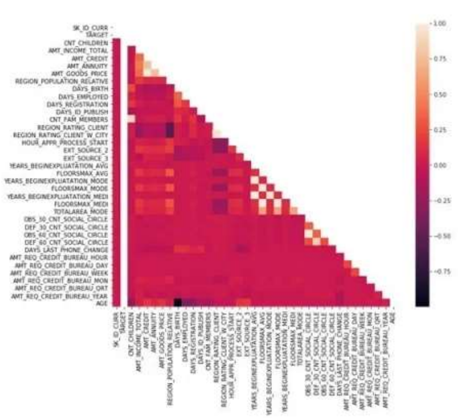
Proportion of Defaulters ID Change, Region

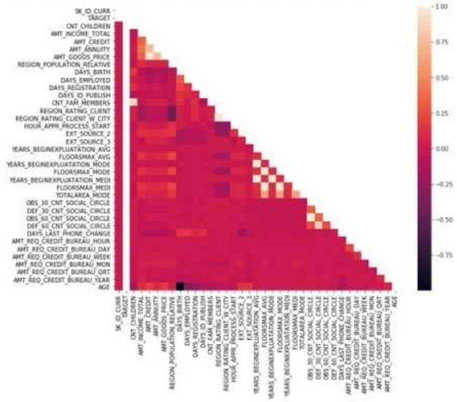




- Recent ID Change, Region Rating of City 2 to 2.5 hogher default

Correlations - Defaults and Pays on Time





Both Correlations look similar

Top Co Relations For Non Defaulters & Defaulters

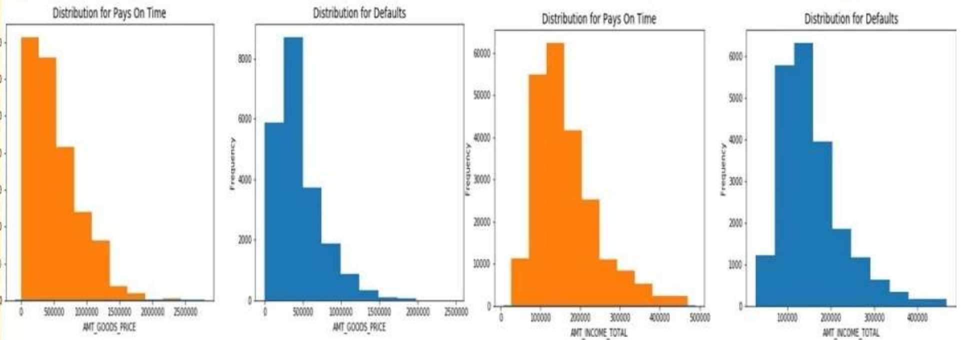
```
AMT_REQ_CREDIT_BUREAU_YEAR   AMT_REQ_CREDIT_BUREAU_YEAR    1.000000
OBS_30_CNT_SOCIAL_CIRCLE     OBS_60_CNT_SOCIAL_CIRCLE      0.998495
FLOORSMAX_MEDI               FLOORSMAX_AVG                 0.997089
YEARS_BEGINEXPLUATATION_MEDI YEARS_BEGINEXPLUATATION_AVG   0.993061
FLOORSMAX_MEDI               FLOORSMAX_MODE                0.988055
                                                             ...
AGE                          DAYS_EMPLOYED                -0.352339
REGION_RATING_CLIENT_W_CITY  REGION_POPULATION_RELATIVE   -0.529862
REGION_POPULATION_RELATIVE   REGION_RATING_CLIENT         -0.532110
DAYS_BIRTH                   AGE                          -0.999593
SK_ID_CURR                   TARGET                             NaN
Length: 597, dtype: float64
```

```
]: AMT_REQ_CREDIT_BUREAU_YEAR   AMT_REQ_CREDIT_BUREAU_YEAR     1.000000
   OBS_30_CNT_SOCIAL_CIRCLE     OBS_60_CNT_SOCIAL_CIRCLE       0.998279
   FLOORSMAX_AVG                FLOORSMAX_MEDI                 0.997492
   YEARS_BEGINEXPLUATATION_AVG  YEARS_BEGINEXPLUATATION_MEDI   0.995884
   FLOORSMAX_MEDI               FLOORSMAX_MODE                 0.989253
                                                                 ...
   DAYS_EMPLOYED                AGE                           -0.306965
   REGION_POPULATION_RELATIVE   REGION_RATING_CLIENT          -0.437613
   REGION_RATING_CLIENT_W_CITY  REGION_POPULATION_RELATIVE    -0.441615
   AGE                          DAYS_BIRTH                    -0.999566
   SK_ID_CURR                   TARGET                              NaN
   Length: 597, dtype: float64
```

30

Top ten CoRelations in Both Population who Pays and Population who Defaults Same

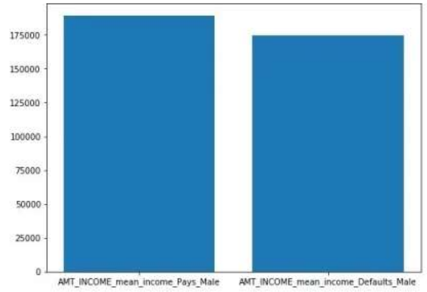
Proportion of Defaulters

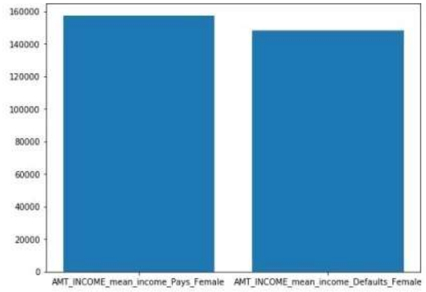


- Defaults higher for loans < 500,000
- Defaults are higher when Income lower

31

Proportion of Defaulters

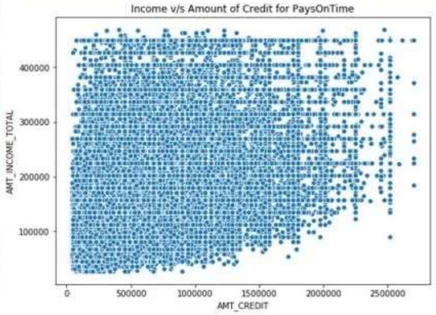


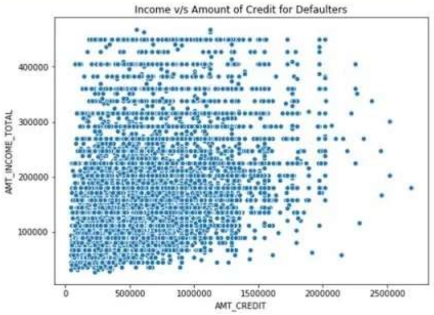


- Mean Income generally lower for defaulters Male Defaulters have lowest mean income

32

Defaulters Density

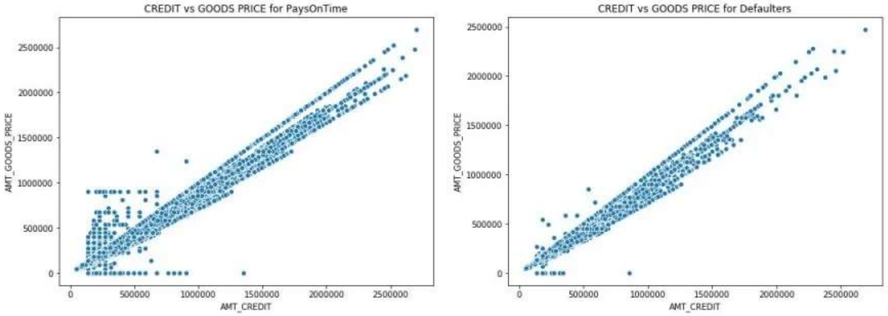




- Default Density is lower where Amount of Credit > 150,000 or Income > 300,000

33

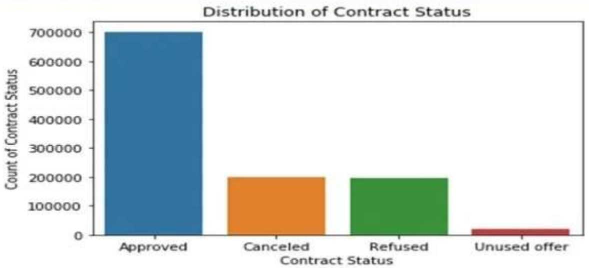
Defaulters Correlations

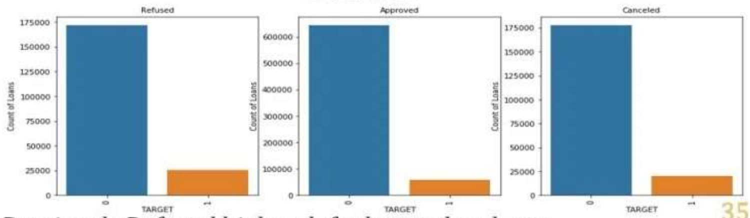


- Strong Correlation between Credit Amount and Price of Goods. Defaults are lower where Amount of Credit and the Price of Goods is > 150,000

34

Contract Status

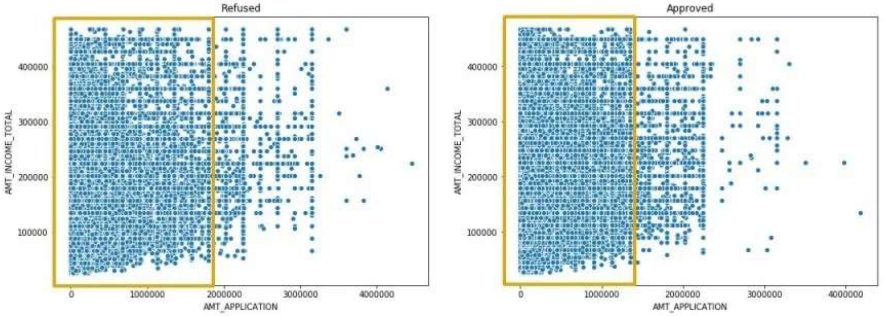




35

- Loans Previously Refused higher default rate than loans

Loan Rejection Rate



- Loans higher than 200,000 have a higher rejection rate
- Rejection rate is lower if income higher than 500,000

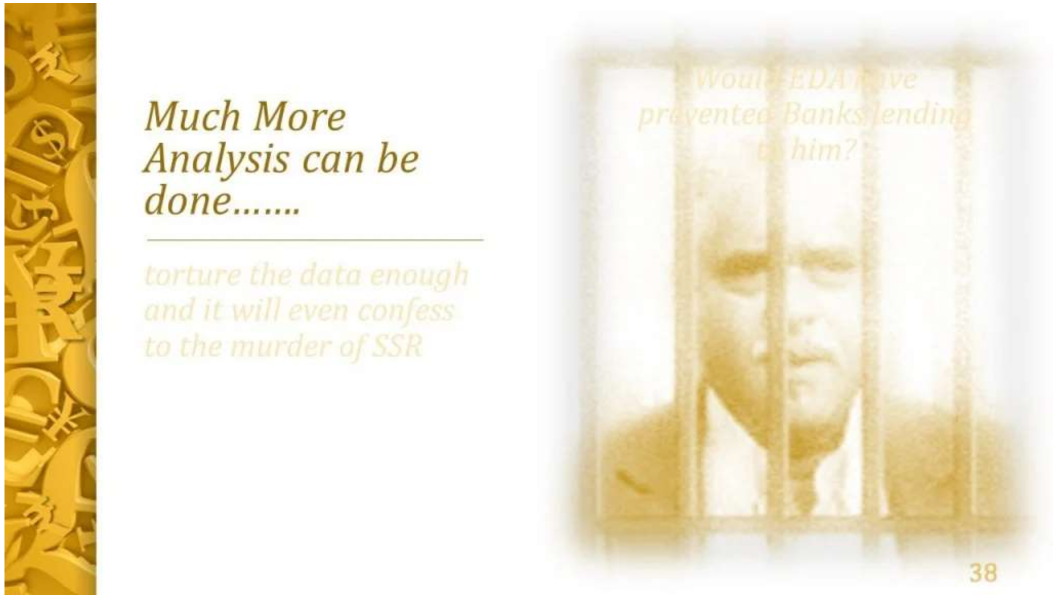
36

Conclusions

EDA for the banking data set revealed that :

- The proportion of defaulters is 8.7%
- The bank lends more to females
- More Cash Loans go into default : bank should give more Revolving Loans.
- Proportion of Working defaults more / State Servants less
- People with Higher Education; Older People default less
- Single People Default more : giving loans to married safer
- Higher Amount loans; Higher Income – less defaults
- Longer employment history, longer registration days less default
- Loans Previously Refused or Cancelled - higher default rate

37



[Next SlideShares]
Upcoming SlideShare

[Data Science Intermediate Program - Gold Certificate]
Loading in …3
×

6 of 38
1

Credit EDA Case Study : Exploratory Data Analysis on Bank Loan Data

Nov. 04, 2020
• 2 likes • 3,286 views

2

Share

Download Now Download

Download to read offline

Data & Analytics

Find patterns in the data that are predictors of loan defaults using Exploratory Data Analysis on Bank Data.

Read more

PRABHASH GOKARN

Following Follow

Business Technologist, B2B Marketing & Projects Professional | Speaker, Author, Researcher

-

Find patterns in the data that are predictors of loan defaults using Exploratory Data Analysis on Bank Data.

Read more

Data & Analytics

Recommended

- 
  Data Science Intermediate Program - Gold Certificate
  PRABHASH GOKARN
- 

  Exploratory Data Analysis Example - Credit Risk Analysis (Second Attempt)
  PRABHASH GOKARN
- 

  Exploratory Data Analysis on Baltimore Police Arrest Data
  PRABHASH GOKARN
- 

  Exploratory Data Analysis Example - Telecom Churn
  PRABHASH GOKARN
-

Magnetic Separation of Metallics from Ferrochrome Slag
PRABHASH GOKARN

- 

  60 not out - sixty successful years of continuous ferro alloy making at joda
  PRABHASH GOKARN

- 

  FERRO ALLOY PLANT JODA - 60 not out
  PRABHASH GOKARN

- 

  SEWAGE AND ITS TREATMENT - Experience from setting up Sewage Treatment Plants
  PRABHASH GOKARN

- 

  Sewage and its treatment - experience from setting up STPs
  PRABHASH GOKARN

-

[FAMD OUTSOURCING EXPERIENCE - CCT Meeting - Gokarn1](#)
[PRABHASH GOKARN](#)

Recommended

- 
  [Data Science Intermediate Program - Gold Certificate](#)
  [PRABHASH GOKARN](#)
- 

  [Exploratory Data Analysis Example - Credit Risk Analysis (Second Attempt)](#)
  [PRABHASH GOKARN](#)
- 

  [Exploratory Data Analysis on Baltimore Police Arrest Data](#)
  [PRABHASH GOKARN](#)
- 

  [Exploratory Data Analysis Example - Telecom Churn](#)
  [PRABHASH GOKARN](#)
- 
  [Magnetic Separation of Metallics from Ferrochrome Slag](#)
  [PRABHASH GOKARN](#)
-

[60 not out - sixty successful years of continuous ferro alloy making at joda](#)
[PRABHASH GOKARN](#)

- 

  [FERRO ALLOY PLANT JODA - 60 not out](#)
  [PRABHASH GOKARN](#)
- 

  [SEWAGE AND ITS TREATMENT - Experience from setting up Sewage Treatment Plants](#)
  [PRABHASH GOKARN](#)
- 

  [Sewage and its treatment - experience from setting up STPs](#)
  [PRABHASH GOKARN](#)
- 

  [FAMD OUTSOURCING EXPERIENCE - CCT Meeting - Gokarn1](#)
  [PRABHASH GOKARN](#)

More Related Content

You Might Also Like

Carousel previous

Carousel next

- 1 2 3 4 5 6 7 8

Related Books

Free with a 30 day trial from Scribd

See all

Carousel previous

Carousel next

- 1 2 3

1. **Clipboards keep everything organized**

   You can create clipboards around different topics, and share your great finds.

1. **A particular slide catching your eye?**

   Clipping is a handy way to collect important slides you want to go back to later.

Report as inappropriate

×

Select your reason for reporting this presentation as inappropriate.

There was an error while reporting this slideshow. Please try again.

None Flag

2 likes

×

NamiaMohamedali Jul. 30, 2021

[Joseph Babu Apr. 05, 2021]

Views

×
Total views
3,286
On SlideShare
0
From Embeds
0
Number of Embeds
0

You have now unlocked unlimited access to 20M+ documents!

×

Unlimited Reading

Learn faster and smarter from top experts

Unlimited Downloading

Download to take your learnings offline and on the go

You also get free access to Scribd!

Instant access to millions of ebooks, audiobooks, magazines, podcasts and more.

Read and listen offline with any device.

Free access to premium services like Tuneln, Mubi and more.

Discover More On Scribd

- About
- Support
- Terms
- Privacy
- Copyright

English

© 2022 SlideShare from Scribd

- English

- Español
- Português
- Français
- Deutsch

×

Share Clipboard

×

- Facebook
- Twitter
- LinkedIn

Link

Public clipboards featuring this slide

×

No public clipboards found for this slide

Select another clipboard

×

Looks like you've clipped this slide to already.

Search for a clipboard

Create a clipboard

You just clipped your first slide!

Clipping is a handy way to collect important slides you want to go back to later. Now customize the name of a clipboard to store your clips.

Name* Best of Slides
Description Add a brief description so oth

Visibility

Others can see my Clipboard

Cancel Save

Special Offer to SlideShare Readers ×

slideshare

+

SCRIBD

Wait! Exclusive 60 day trial to the world's largest digital library.

The SlideShare family just got bigger. You now have unlimited* access to books, audiobooks, magazines, and more from Scribd.

Activate your free 60 day trial

Cancel anytime.