Credit EDA Assignment

By Suraj Atkare

Batch ID: 1970

Program: upGrad & IIITB | Data Science Program - October 2021

Problem Statement

- Identify patterns which indicate if a client has difficulty paying their installments.

- Identify Key Parameters

Available Data

- application_data.csv' contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

- 'previous_application.csv' contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

- 'columns_description.csv' is data dictionary which describes the meaning of the variables.

Analysis

- Clients have defaulted in two loans category i.e Cash loans and Revolving loans.

- Clients who had taken cash loans are the one with higher payment difficulties.

- From the given data set it can be deduced that 59% of the client with payment difficulties are married.

```
In [35]: inp0_defaulter['NAME_CONTRACT_TYPE'].value_counts()

Out[35]: Cash loans          23221
         Revolving loans      1604
         Name: NAME_CONTRACT_TYPE, dtype: int64
```

From above we can see that clinets who had payment difcculties are divided in to two categories Cash loans and Revovling loans. Out of these two categories we can easily establish that number of clients with Cash loans have more payment difficulties as compared to Revovling loans.

```
In [36]: inp0_defaulter['NAME_FAMILY_STATUS'].value_counts(normalize=True)

Out[36]: Married               0.598187
         Single / not married  0.179537
         Civil marriage        0.119275
         Separated             0.065257
         Widow                 0.037744
         Name: NAME_FAMILY_STATUS, dtype: float64
```

From above data we can see that out of total number of clients who had payment difficulties approximately 59% of these clients are Married.
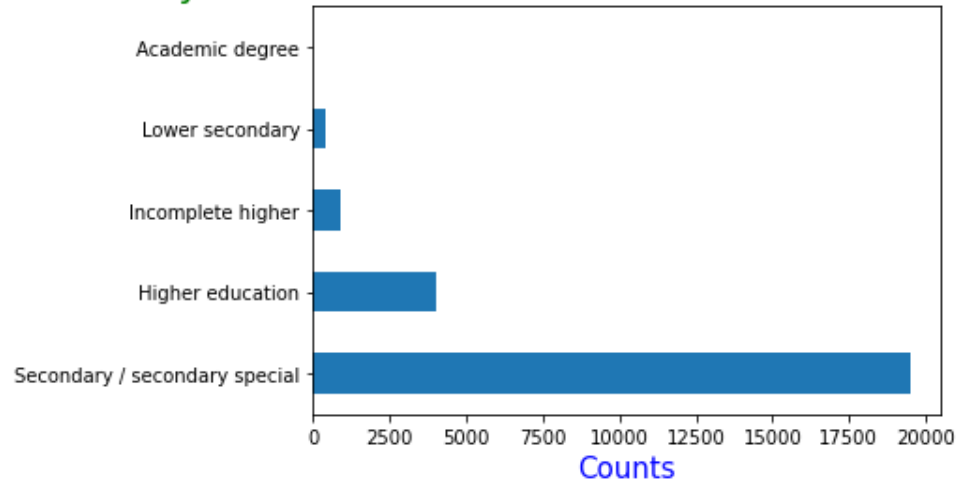
Analysis

Payment Difficulties Across Different Family Status

Payment Difficulties Across Different Education Type
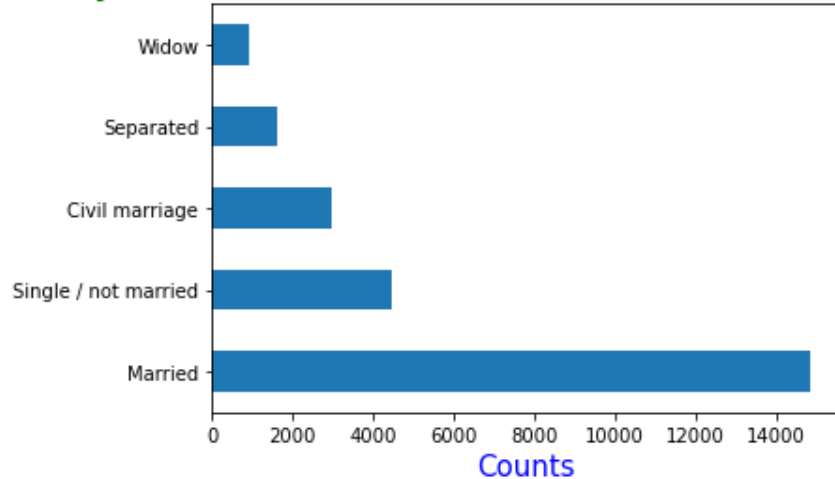
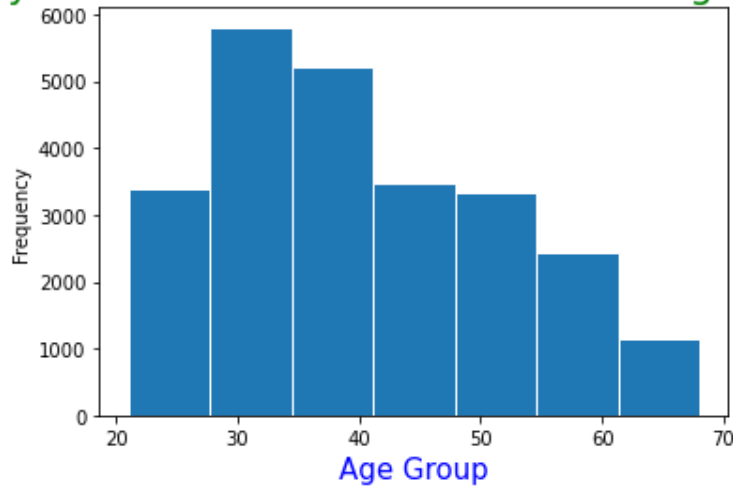- ➧ It can be observed in the bar chart that client who are married have higher payment difficulties.

- ➧ From Bar Chart it can be seen that people with Secondary/secondary special education category are the one's with higher payment difficulties for their loans.
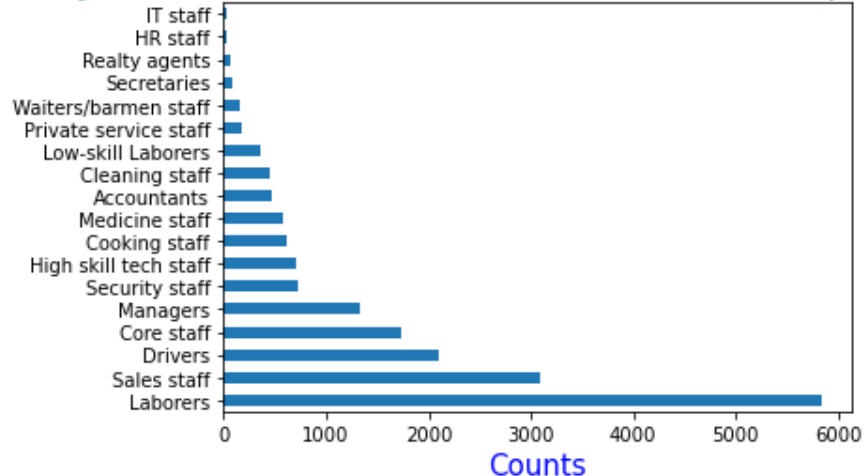
Analysis

Payment Difficulties Across Different Age Group

## Payment Difficulties Across Different Occupation Type

- This histogram chart clearly indicates that clients in age group of 28 to 41 at the time of loan application are the one's with higher payment difficulties against the loans.

- It can be observed that clients who had payment difficulties against their loans 57% of these clients were Female.

- It can be observed that clients working as IT professional are less likey to default as compared to clients working as Laborers.

- More interesting observation that can be drawn from the bar graph is that client working as Managers, Core Staff and Sales staff are among the top occupation category with payment difficulties.

```
Out[43]:  Female     0.570796
          Male       0.429204
          Name: CODE_GENDER, dtype: float64
```
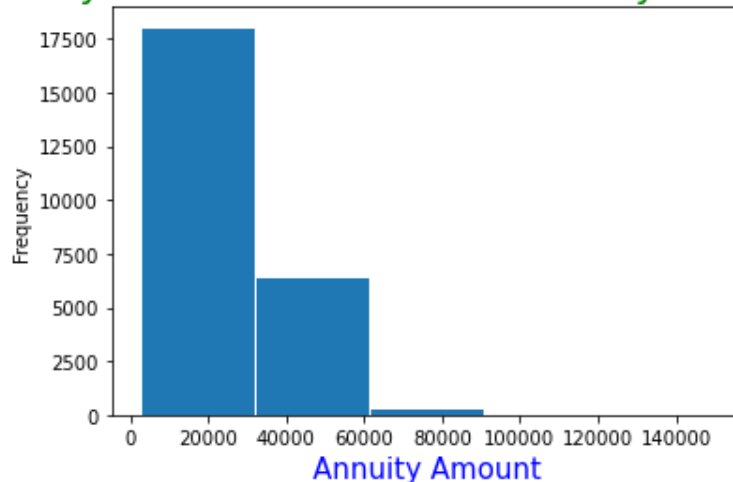
Analysis

Payment Difficulties As Per Region Rating w.r.t City
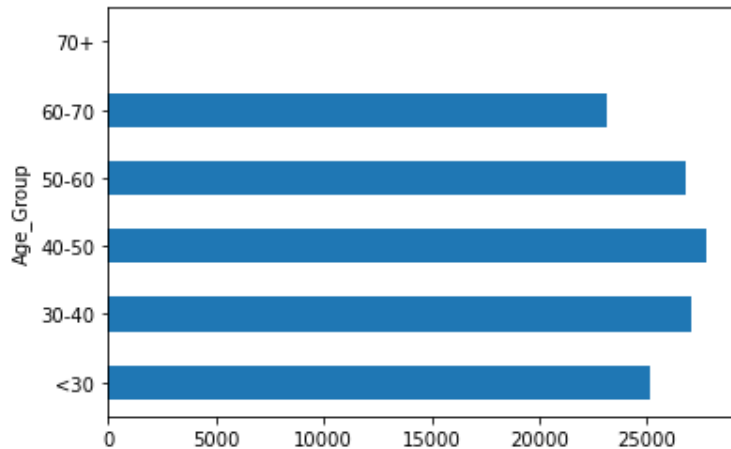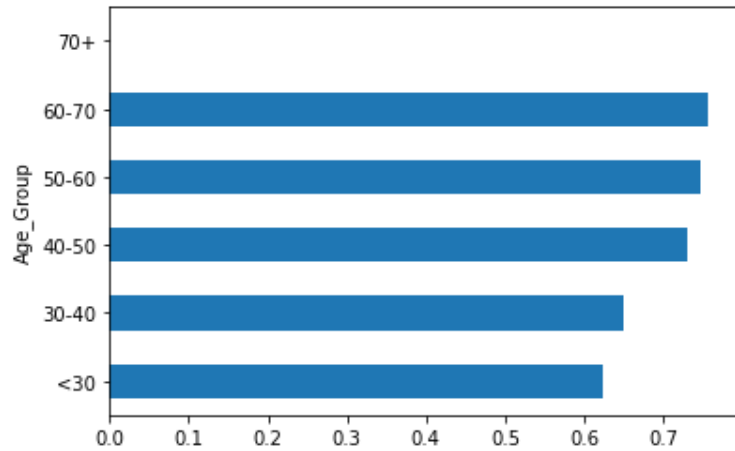

Payment Difficulties w.r.t Annuity Amount
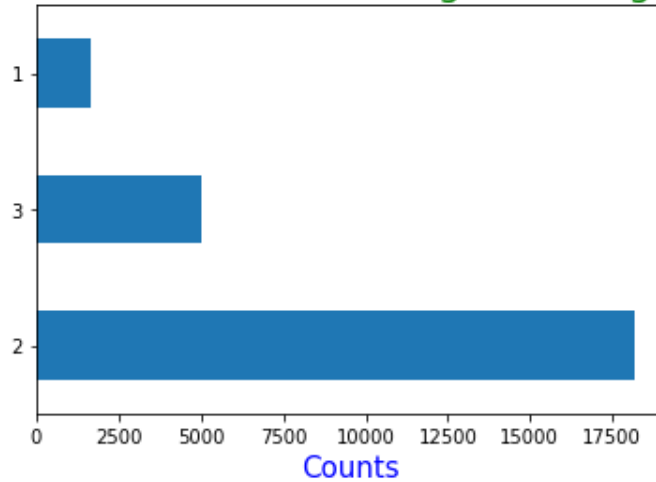
➡ In this bar graph it can be said that payment difficulties for client residing in 2 rated region is higher compared to 1 & 3 even after considering city in which client lives in.

➡ It can be observed that higher annuity amount has low payment difficulties it does not mean that annuity amount should be high in order to decrease deafult in loans but could suggest towards the income of the clients that clients with higher income has less probability of payment difficulties.
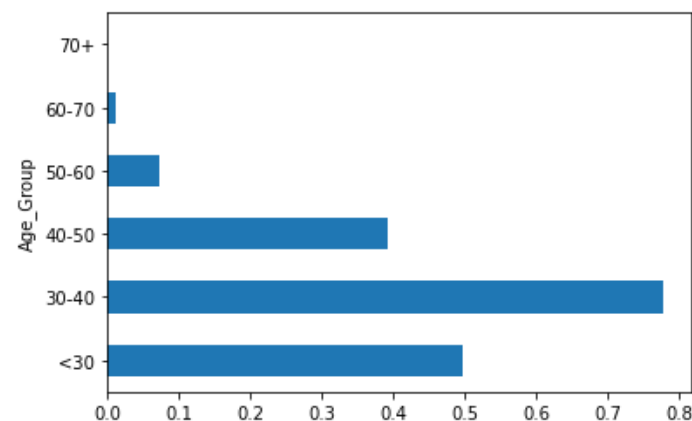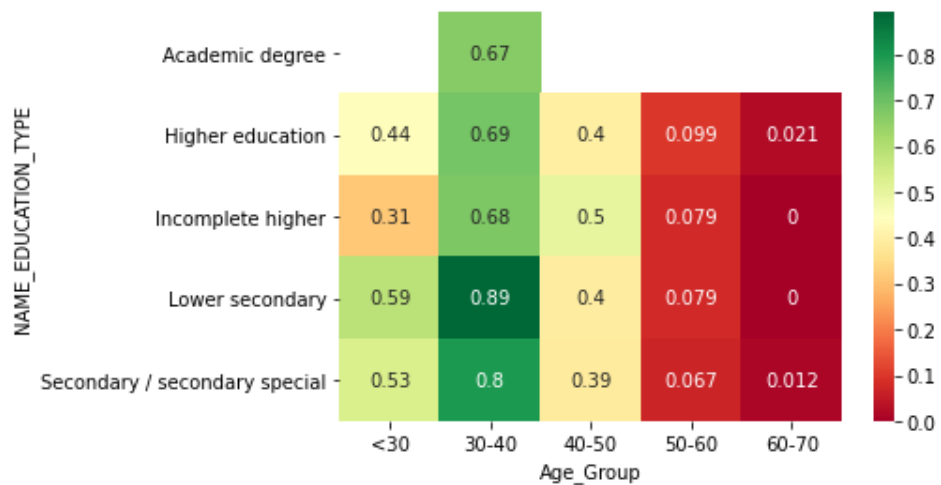
Analysis

- This Bar graph is plotted between age group and real estate owned by the clients who had difficulties with payment.

- It can be observed that age group of 60-70 is marginally higher as compared to age group of 40-50 and 50-60.

- In second bar graph average annuity amount vs age group is plotted. It can be observed that age group of 30-50 has higher payment difficulties when average annuity amount is higher than 25,000.

- It is also observed that clients who has children and fall in the age group of 30-40 have more difficulties with payment.

Analysis
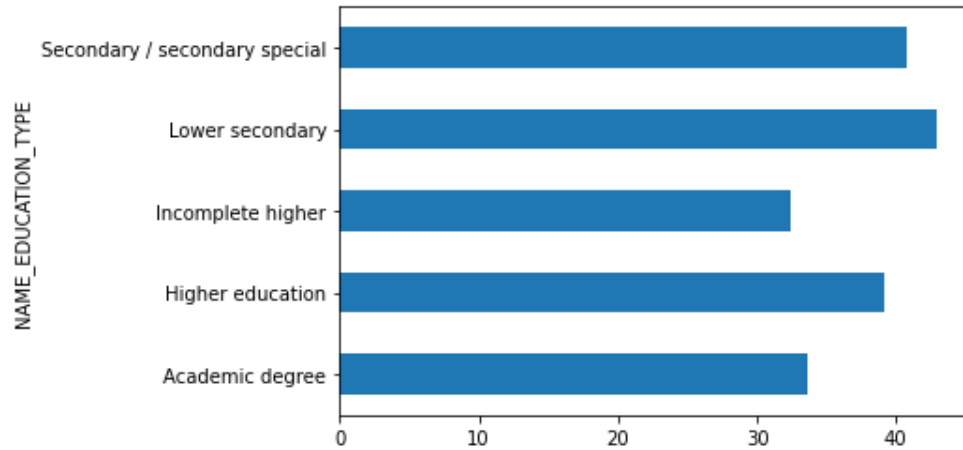
- First Bar Graph is plotted in between education category and Age.

- Clients in the age bracket of 30-40 with lower secondary, secondary, higher education education have higher payment difficulties.

- In the heat map it can clearly be observed that the age bracket of 30-40 with lower secondary education education have higher payment difficulties.

Conclusion / Suggestions

- The Analysis indicated that client in the age group of 30-50 with lower education and children are likely to default.

- It was also observed that client with previous loan has higher payment difficulties.

- To handle such clients bank should do a due diligence if client had any previous loans and payment history.

- Clients who do not have payment history for previous loans their loans can be approved if annuity amount against their loans is below INR 25,000.

- Clients who are single and have higher academic education can be target for higher loan amounts.