EDA Assignment

Imran Totti Shaik

Business and Data Understanding

- **Business Problem**
  - The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. When the company receives a loan application, the company must decide 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 Solution**
  - EDA Techniques on both Current and Previous loan applications data to gain patterns present in the data.

- **Data Problem**
  1. Current application data: The current data contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:
     - The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
     - All other cases: All other cases when the payment is paid on time.

  2. Previous loans data: When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

     - Approved: The Company has approved loan Application
     - Cancelled: The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
     - Refused: The company had rejected the loan (because the client does not meet their requirements etc.).
     - Unused offer: Loan has been cancelled by the client but on different stages of the process.

- **Data Solution**
  1. Identify the clients who has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
  2. Identify the clients capable of repaying the loan are not rejected.

EDA Techniques

- Using below mentioned EDA steps are followed to explore the data and find key insights:

   1. **Data Sourcing and Data Profiling:**

      1. Read the data and do data profiling to understand the data structure, shape, columns, rows and irregularities.

   2. **Data Cleaning:**

      1. Improve the data quality by fixing the rows – address incorrect, summary and extraneous rows.

      2. Fix the columns – rename, delete, add, split, merge and align columns if required.

      3. Deal with the missing values in the data sources. Impute or delete the missing data.

      4. Standardize the data values, fix invalid values, identify and decide on outliers, and filter data.

   3. **Univariate Analysis:**

      1. Analyze the categorical unordered variables.

      2. Explore the categorical Ordered variables.

      3. Explorer the numerical variables.

   4. **Bivariate analysis:**

      1. Analyse numerical versus numerical variables

      2. Analyse numerical versus categorical variables

      3. Analyse categorical versus categorical various variables

   5. **Multivariate analysis:**

      1. Analyse multiple variables

   6. Merge the data sources and investigate the relationships and patterns on combined set.
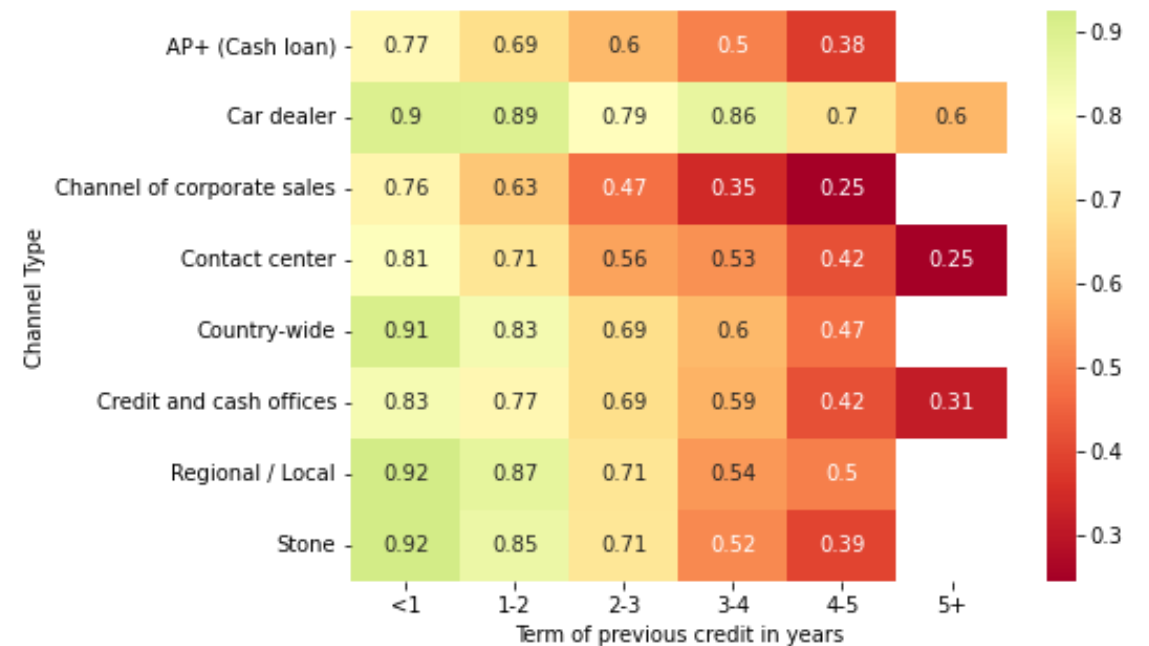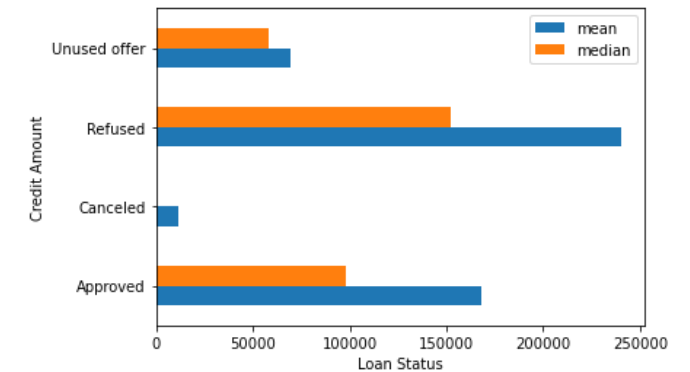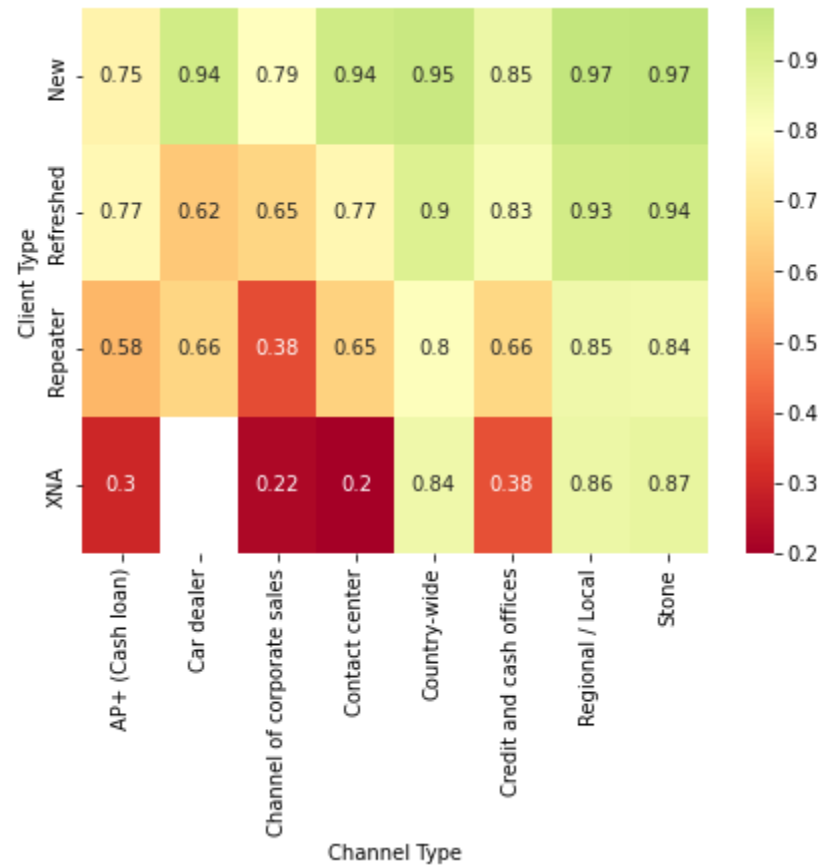
Insights – Current Loan Applications Data

1. Clients with payment difficulties are:

   • Clients with secondary education, working as laborers and on maternity leave.

   • Unemployed clients with secondary education.

2. Clients capable of repaying the loan are:

   • Businessman and Students with academic degrees.
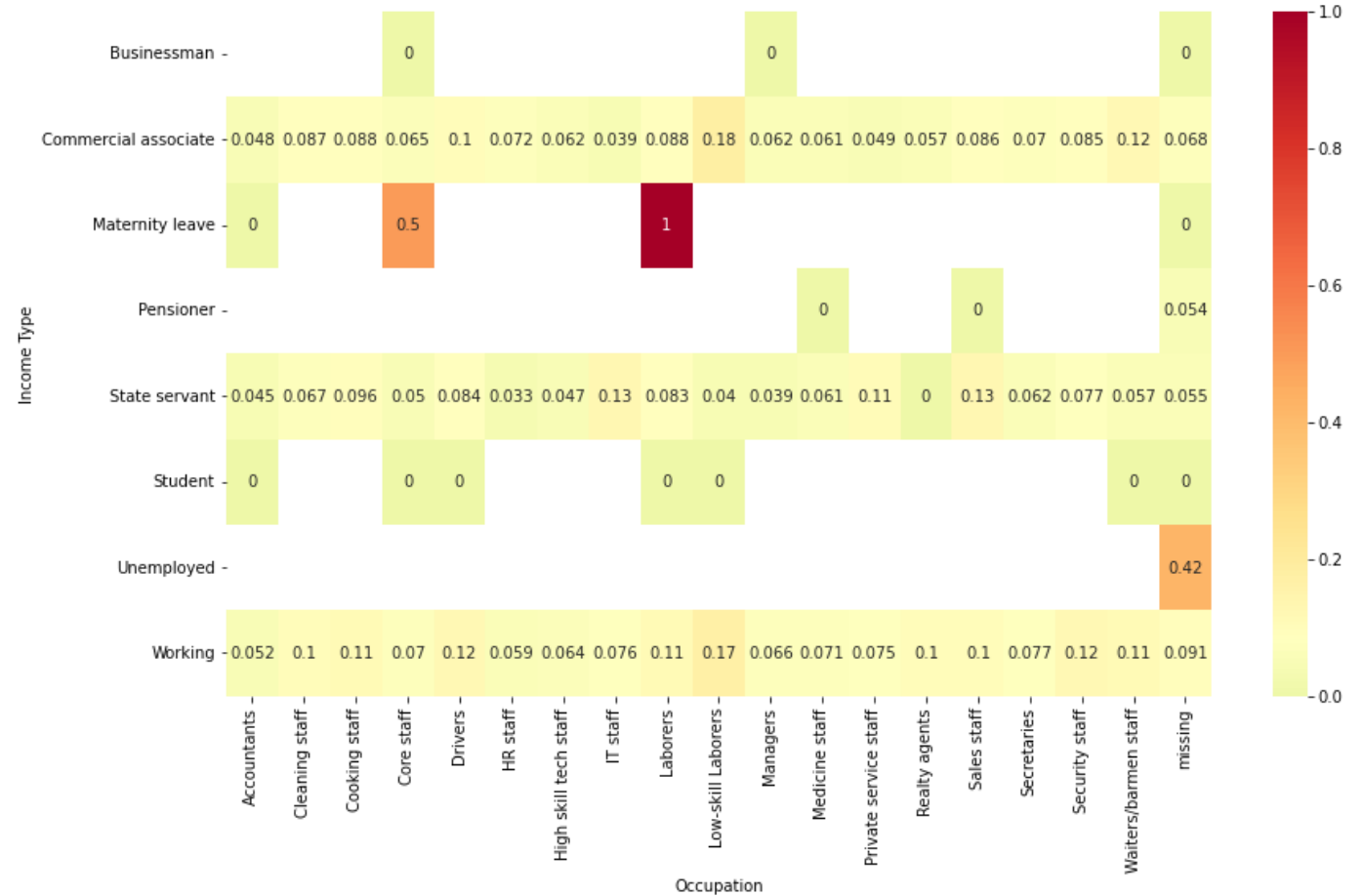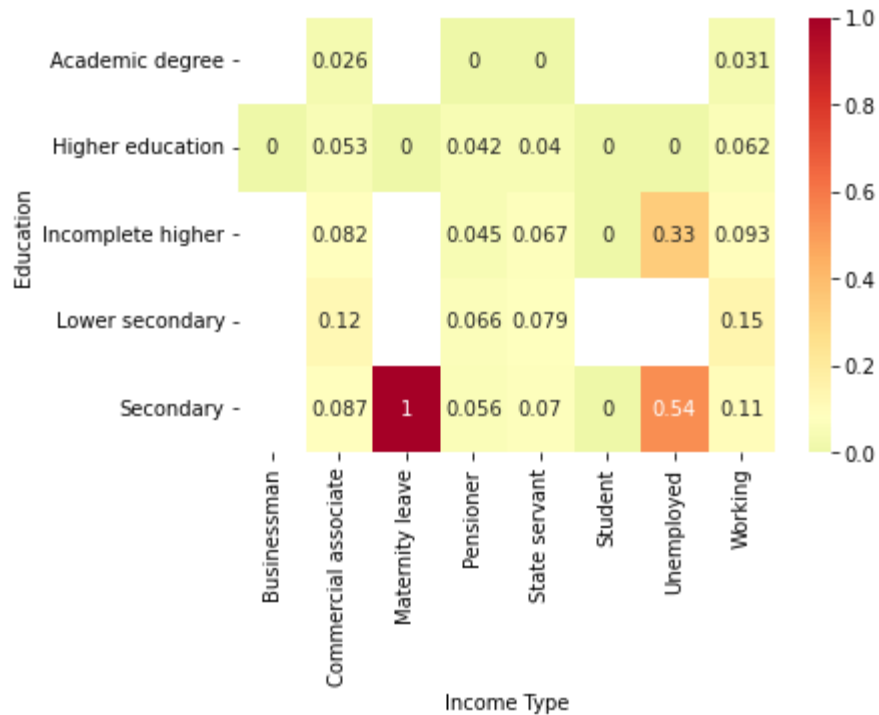
Insights – Previous Loan Applications Data

- Clients applied for large credit amount most likely to be refused.

- New clients whose loans applied through the Stone and Regional/local channel types are most likely to be approved. Repeater clients are being refused but still their loan will be approved when it is applied through stone and local channels.

- Clients whose term of previous credit at application of the previous application less than 2 years most likely to get approval.

Recommendations

- Here is the summary of key insights that we gained from the two data sets. And suggested here are the actions that can be taken on clients' loan application:

- **Clients with payment difficulties:**

  - Laborers with secondary education, who are unemployed or on maternity leave most likely to face payment difficulties.

- Action:

  - Approve when the credit amount is less, these are new or applied for the first time, and applying through the regional or local channels.

  - Approve when these clients are not applying for the first time, and their term of previous credit at application of the previous application is less than 2 years.

- **Clients without payment difficulties:**

  - Businessman and Students with academic degrees.

- Action:

  - Approve as these clients are least likely to default on loan.