

Current Application Data

Importing data

- Shape – 307511 x 122

Preparing for Analysis - I

- Dropping Columns with missing percentage greater than 50 % - 41 columns
- 6 columns had around 13 % missing values and were not imputed for analysis as they were critical values and imputing them with 0(the mean, mode and median) would give a very wrong input
- Converting to right datatypes – Days column, Family Members and Region Rating W City

Preparing for Analysis - II

- Binning of continuous numerical values like Income total, emi, credit, days birth
- Dividing the dataset into two based on TARGET variable
- Taking random sample of 5000 for two dataset for the analysis and now data is balanced for analysis

Analysis

- Tried to analysis using Univariate and Bivariate approach using numerical and categorical columns to establish any pattern or trend with a specific feature

Contents 🔄 ⚙️

- ▶ 1 Structure of data (Normal Routine Check)
- ▼ 2 Data Quality Check
 - 2.1 Missing percentage of each column
 - ▼ 2.2 Dropping columns having missing pe
 - 2.2.1 Column Name with missing perce
 - 2.2.2 Shape of 'missing_value_more50
 - 2.2.3 Adding column names to 'drop_c
 - 2.2.4 Dropping the 'drop_column_nam
 - 2.3 Reporting the Values for missing perc
 - ▼ 2.4 Column Datatypes Correction
 - 2.4.1 Printing Column Name, Datatype
 - 2.4.2 Days column as absolute values
 - 2.4.3 Regional Rating to convert as flo
 - 2.4.4 Family Member
 - ▼ 2.5 Outlier Identification
 - 2.5.1 CNT_CHILDREN
 - 2.5.2 AMT_INCOME_TOTAL
 - 2.5.3 AMT_CREDIT
 - 2.5.4 AMT_ANNUNITY
 - 2.5.5 CNT_FAM_MEMBERS
 - 2.5.6 DAYS_BIRTH
 - 2.5.7 AMT_GOODS_PRICE
 - ▼ 2.6 Binning
 - 2.6.1 AMT_INCOME_TOTAL
 - 2.6.2 AMT_CREDIT
 - 2.6.3 AMT_ANNUNITY
 - 2.6.4 DAYS_BIRTH
- ▼ 3 Analysis :
 - 3.1 Checking the imbalance percentage
 - 3.2 Divide the dataset into two part

Preparing Data for Merge

- Cleaning the Previous Data – Dropping columns with missing greater than 50 %, converted the days column to absolute
- Grouped Columns with respect to **SK_ID_CURR**
- Selected the columns those are required by performing the median/mean operation on respect columns
- In columns having categorical values, used `pd.crosstab` and extracted their value counts as new numeric columns
- Merged the current and processed previous dataframe using inner join on **SK_ID_CURR**
- Shape - 291057 rows × 94 columns

Preparing for Analysis - I

- Binning process to create categorical columns for better analysis
- Split the dataset into two dataset based on **TARGET** value
- Took random sample of 5000 for analysis

Analysis

- Tried to analysis using Univariate and Bivariate approach using numerical and categorical columns to establish any pattern or trend with a specific feature of applicants previous loan application history to current loan difficulties in paying EMI

Previous Application Data

Contents 🔄 ⚙️

- ▼ 4 Previous Application Data
 - ▶ 4.1 Data Overview - Previous Application
 - ▼ 4.2 Data Cleaning
 - 4.2.1 Checking null value count
 - 4.2.2 Dropping Column with Null Value
 - 4.2.3 Converting the **DAYS_DECISION**
 - 4.2.4 Printing Previous Post Cleaning
 - ▼ 4.3 Preparing the data for analysis
 - 4.3.1 Pivot - **SK_ID_CURR**
 - 4.3.2 Scrapping the required details fr
 - 4.3.3 Creating the dataframe from prev
 - 4.3.4 Resultant Dataframe from Previc
 - 4.4 Merging the Current & Previous Data
 - ▼ 4.5 Adding some Categorical Columns
 - 4.5.1 **AMT_APPLICATION** - Binning
 - 4.5.2 **AMT_CREDIT_previous** - Binning
 - 4.5.3 **DAYS_DECISION** - Binning
 - 4.5.4 **CONTRACT_TYPE_PREVIOUS**
 - 4.5.5 **EMI PAYMENT TYPE PREVIOUS**
 - ▼ 4.6 Univariate Analysis
 - 4.6.1 Checking for data imbalance
 - 4.6.2 Splitting the dataframe with **TARGET**
 - ▼ 4.7 Univariate Analysis for Combined data
 - ▼ 4.7.1 Univariate Categorical
 - 4.7.1.1 Previous Application Amount
 - 4.7.1.2 Previous Credit Median Cate
 - 4.7.1.3 Last Application Result Days
 - 4.7.1.4 Previous Applicant Contract T
 - 4.7.1.5 **BIN_PREVIOUS_PAYMENT**
 - ▼ 4.7.2 Univariate Continous
 - 4.7.2.1 Last application Median Valu
 - 4.7.2.2 Last Application Credit Media
 - 4.7.2.3 Last Day since application R