EDA Assignment Submission

Student Name: Tarriq Ferrose Khan

Batch : DSC 68

upGrad

upGrad & IIITB | Data Science Program - May 2024

Approach Note

Missing Values

- Exported null columns information , ordered by percentage of null descending.
- Checked the Distribution of Significant columns using value_counts(normalize=True)
- Handled the null values basis the type of null columns, for example Occupation Type fill na is done as 'NOT DECLARED' as per my observation it is MNAR.

Outlier Handling

- To Check and Remove outliers, implemented IQR calculation , created a new Outlier flag column with 1 if value is less the IQRmin or greater than IQR max and removed the records with Outlier flag column.
- Implemented this approach for AMT_INCOME_TOTAL, DAYS_BIRTH, CNT_CHILDREN, CNT_FAM_MEMBERS

Data Transformations

- Created new columns like YRS_BIRTH, by converting absolute(Days) into Years, and implemented the same logic for similar type of columns.
- Flag columns like FLAG_OWN_CAR, FLAG_OWN_REALTY are translated to 1 for Y and 0 for N, so that it can be used for Correlations.

Group Column Creation

- Wherever applicable, group columns are created for having proper dimension of data analysis, like AGE, INCOME and EXPERIENCE GROUPS are created.

File Segregation

- Segregated application data basis the Loan Type (NAME_CONTRACT_TYPE) and TARGET values (0,1) and merged with previous data file for necessary columns

Correlation Study

<u>Correlation Study - Insights</u>

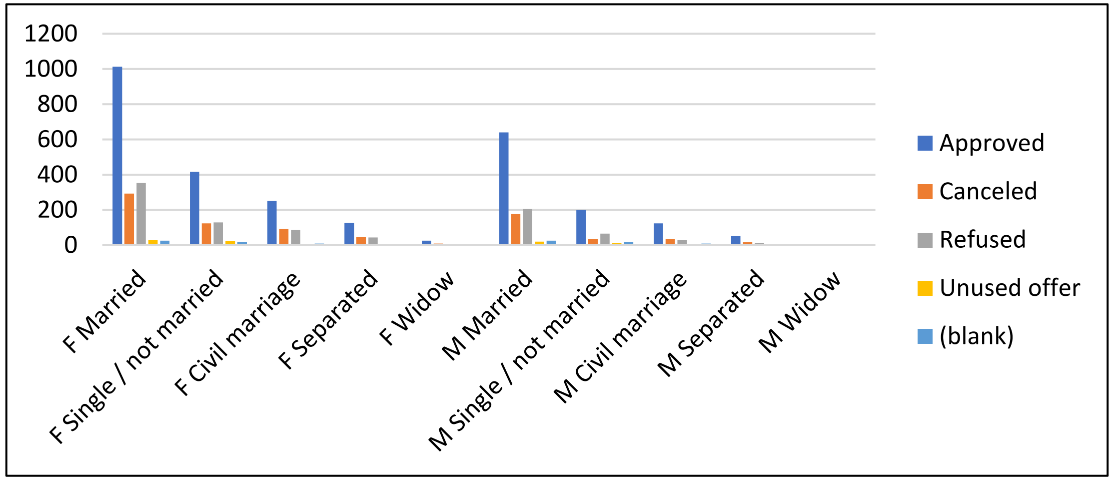
- Considered below attributes for further analysis
- GENDER | Education Type | Occupation Type | Income Group | Family Status | Housing Type | Owns Realty| Owns Car | #Children | #Family Members | Previous Application approval Status |
- **REASON FOR COLUMN SELECTION:** These columns are selected basis the events that will influence the income and expenditure of a house hould that could cause delay of the loan payment.
  - For Instance if OWN_REALTY is 1 , the client may be already paying an EMI
  - INCOME GROUP may is lesser range.
  - #Children and Family members may cause payment difficulties.
  - Prior Loan if approved and if payment term for that is due and conflicting with current one.
  - Housing Type is considered to understand the investments already made.
  - Occupation Type, Education Type, Family Status are all considered to understand which combination is having more difficulty so that , such categories can be avoided to award loan sanctions

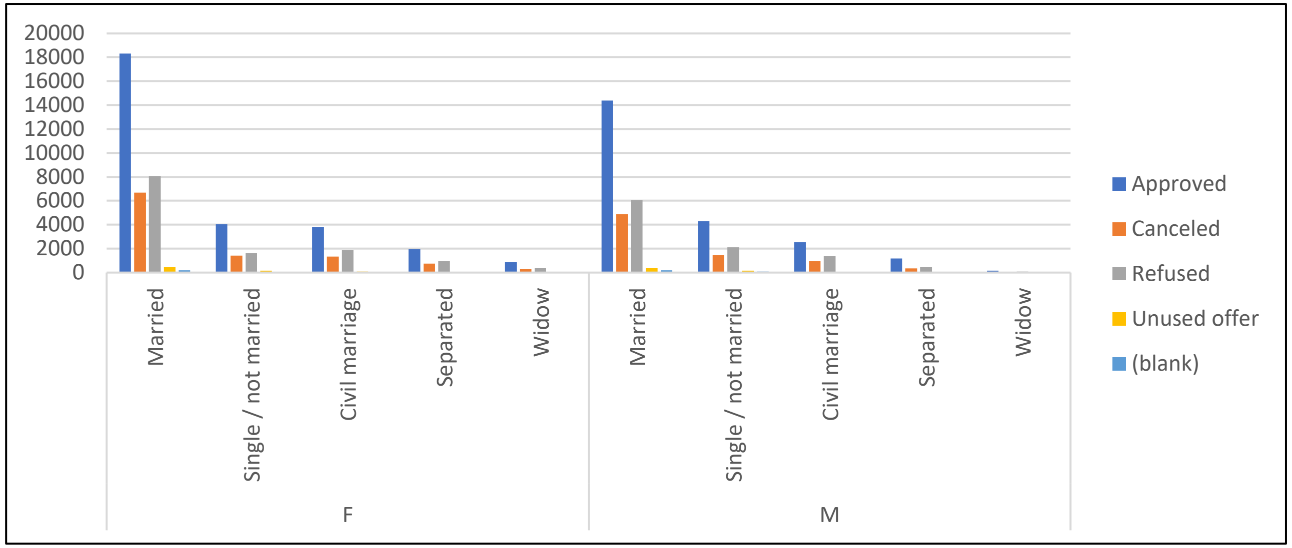
Payment Difficulty observations

- **Revolving Loans**
  - MALE has more payment difficulties than Female
  - MALE Married and prior application approved has more payment difficulties
  - Person who owns realty has more difficulties, regardless of GENDER
  - Owning car and realty does not have effect in payment difficulties
  - Owning car and realty has relatively less effect compared to Cash loans
  - Owning REALTY seems to have more effect than #Children regardless of GENDER
  - Occupation Type of Labourers with Exp bucket of 0-5 tends to have more difficulties but far less when compared to Cash loans of same Criteria
- **Cash Loans**
  - Female has more payment difficulties than Male.
  - MALE Married and prior application approved has more payment difficulties
  - Person who owns realty has more difficulties, regardless of GENDER
  - Person who Owns car and realty has more further more difficulties in payment
  - Owning car and realty has relatively high effect compared to Revolving loans
  - Owning REALTY seems to have more effect than #Children regardless of GENDER
  - Occupation Type of Labourers with Exp bucket of 0-5 tends to have more difficulties but far high when compared to Cash loans of same Criteria.
  - 0-5 Exp Group with 1.25 – 1.5 Lakh Income group tends to have more difficulties where REALTY is 1

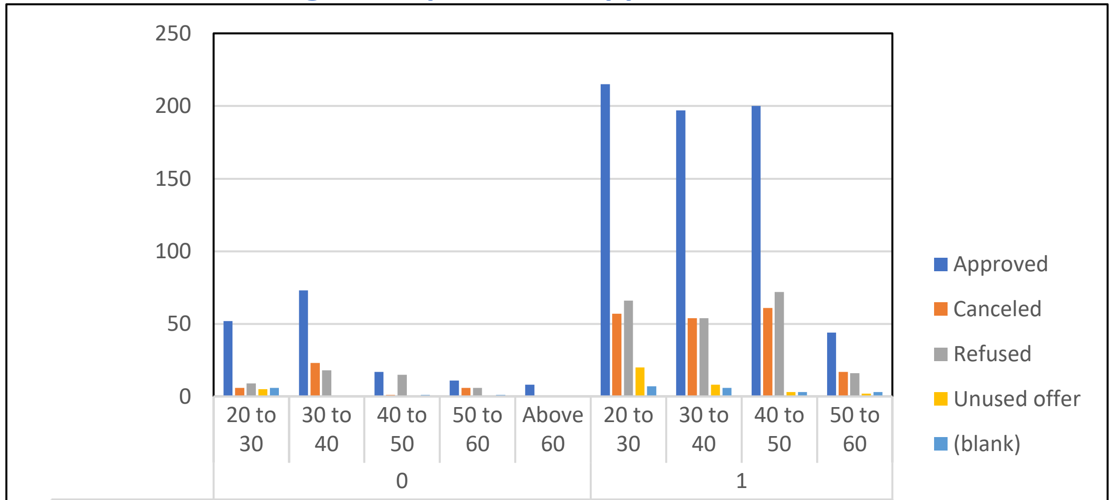
Revolving Loans (Left) vs Cash Loans (Right)

GENDER Vs Family Status Vs Prior Approved

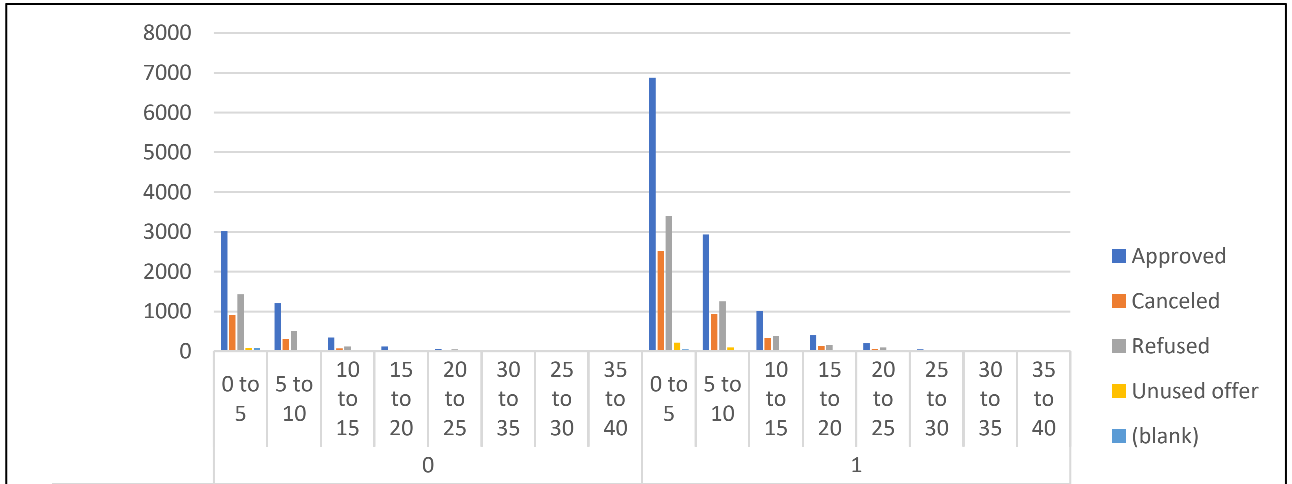




OWN REALTY Vs Age Group Vs Prior Approved



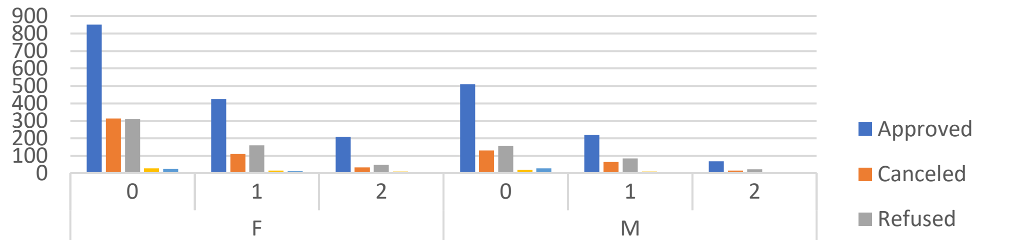
	0					1			
	20 to 30	30 to 40	40 to 50	50 to 60	Above 60	20 to 30	30 to 40	40 to 50	50 to 60
Approved	52	73	17	11	8	215	197	200	44
Canceled	6	23	1	6		57	54	61	17
Refused	9	18	15	6		66	54	72	16
Unused offer	5					20	8	3	2
(blank)	6		1	1		7	6	3	3



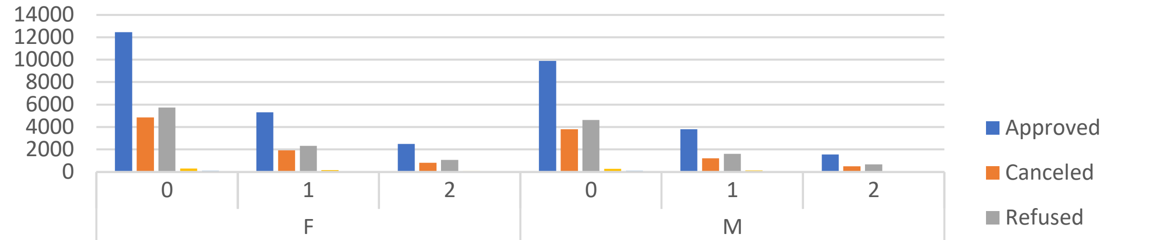
	0								1							
	0 to 5	5 to 10	10 to 15	15 to 20	20 to 25	30 to 35	25 to 30	35 to 40	0 to 5	5 to 10	10 to 15	15 to 20	20 to 25	25 to 30	30 to 35	35 to 40
Approved	3013	1204	345	121	59	6	8	2	6877	2937	1013	397	203	44	33	2
Canceled	916	316	73	34	18	5	4		2520	933	337	131	55	10	5	
Refused	1432	510	122	34	43	6	3	1	3397	1253	380	151	99	16	7	
Unused offer	86	32	12	5	3		1		214	99	29	12	5			
(blank)	87	20	6		1		1		46	10	7	3				

Revolving Loans (Left) vs Cash Loans (Right)

OWN_REALTY=1, vs GENDER, Vs #Children

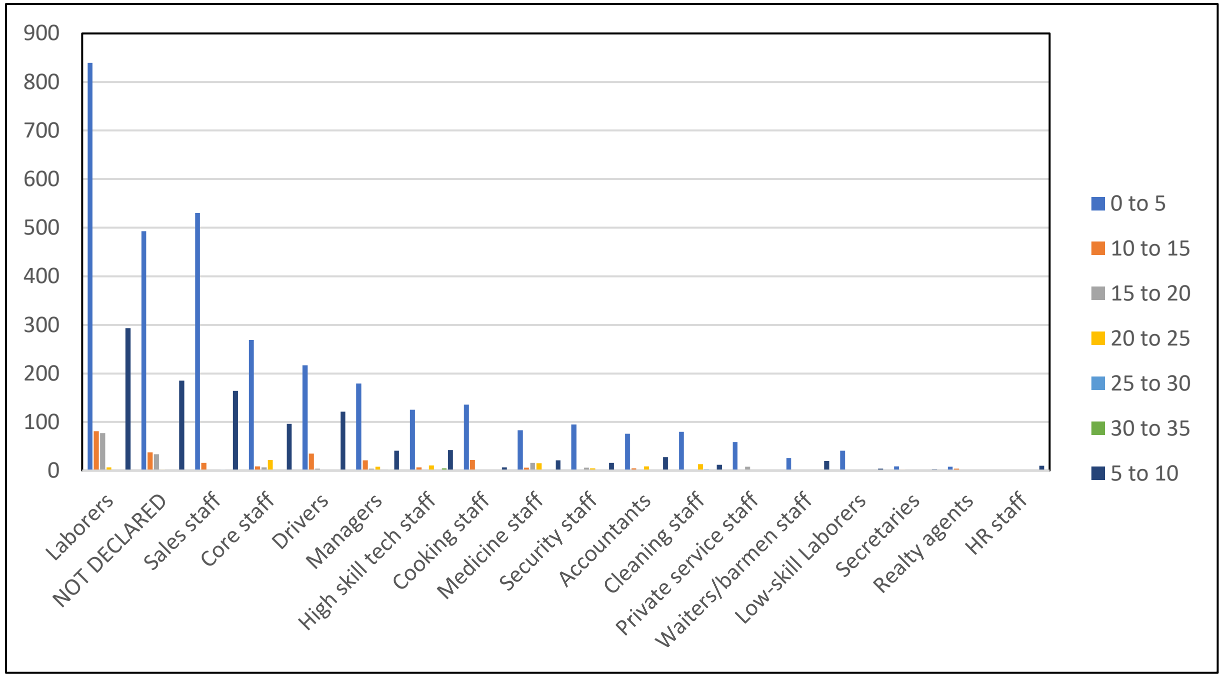


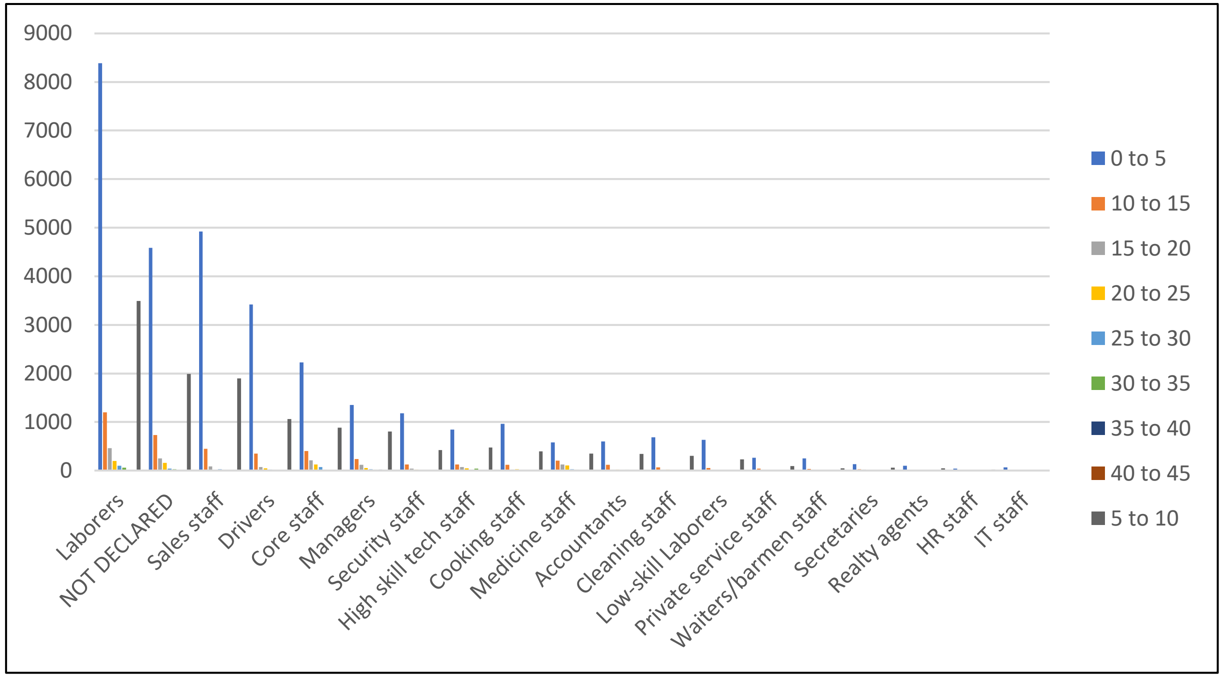
	F			M		
	0	1	2	0	1	2
Approved	851	426	209	510	221	68
Canceled	313	110	34	131	64	15
Refused	312	159	49	157	84	23
Unused offer	28	16	9	19	10	1
(blank)	25	12	5	28	7	3



	F			M		
	0	1	2	0	1	2
Approved	12448	5306	2484	9899	3788	1537
Canceled	4839	1920	803	3799	1204	501
Refused	5729	2329	1059	4625	1619	683
Unused offer	287	151	63	271	121	54
(blank)	101	36	12	107	22	4

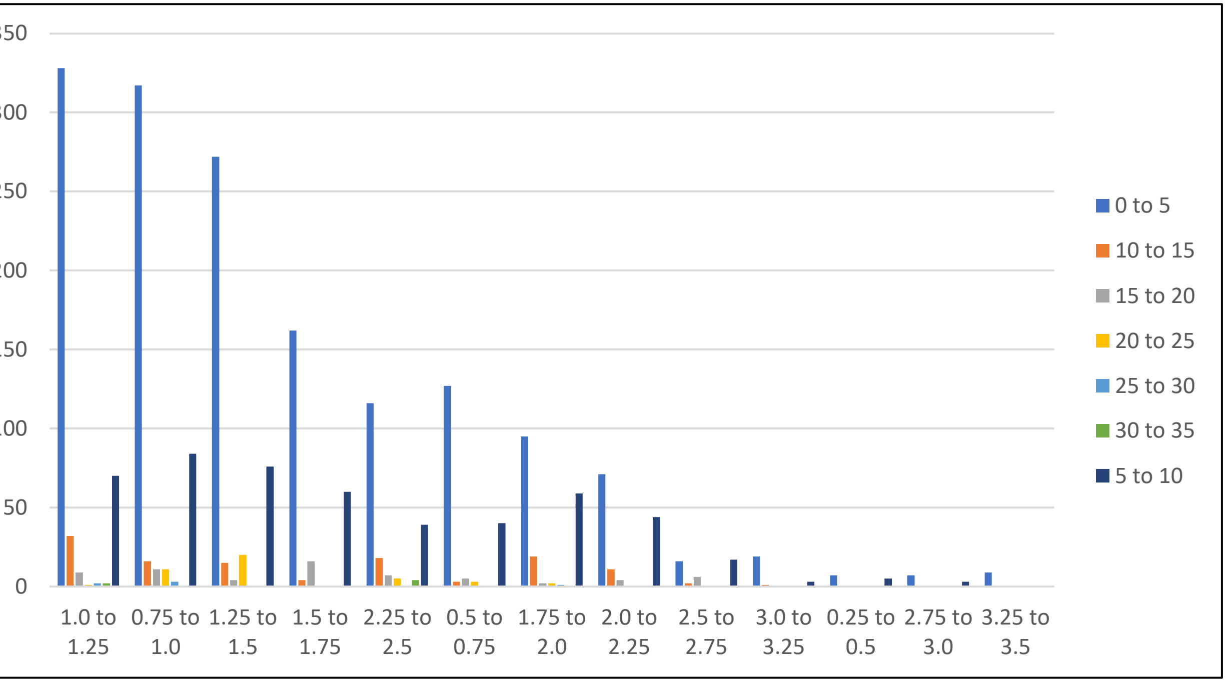
OCCUPATION TYPE vs EXPERIENCE GROUP vs PRIOR APPROVED

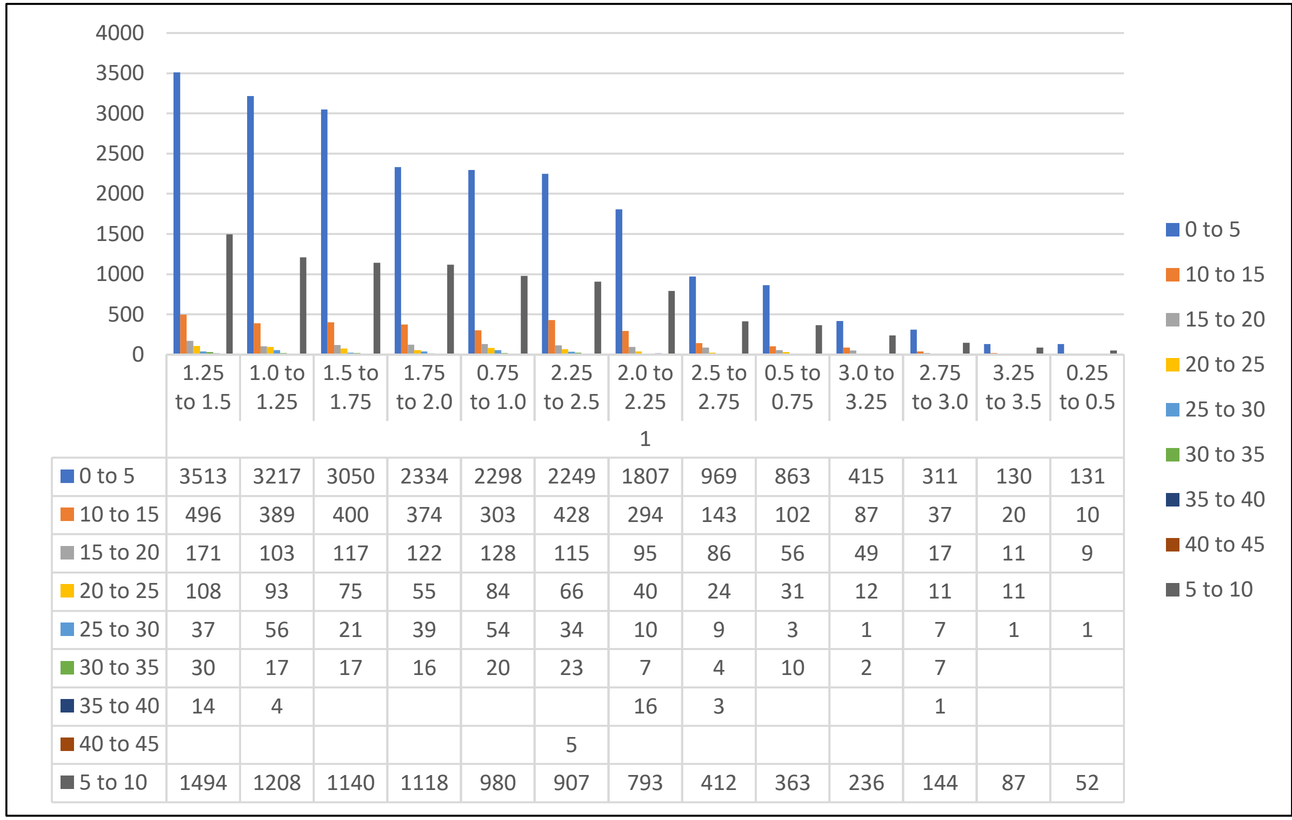




Revolving Loans (Left) vs Cash Loans (Right)

EXPERIENCE GROUP Vs INCOME_GROUP Vs PRIOR APPROVED





	1.25 to 1.5	1.0 to 1.25	1.5 to 1.75	1.75 to 2.0	0.75 to 1.0	2.25 to 2.5	2.0 to 2.25	2.5 to 2.75	0.5 to 0.75	3.0 to 3.25	2.75 to 3.0	3.25 to 3.5	0.25 to 0.5
0 to 5	3513	3217	3050	2334	2298	2249	1807	969	863	415	311	130	131
10 to 15	496	389	400	374	303	428	294	143	102	87	37	20	10
15 to 20	171	103	117	122	128	115	95	86	56	49	17	11	9
20 to 25	108	93	75	55	84	66	40	24	31	12	11	11	
25 to 30	37	56	21	39	54	34	10	9	3	1	7	1	1
30 to 35	30	17	17	16	20	23	7	4	10	2	7		
35 to 40	14	4					16	3			1		
40 to 45						5							
5 to 10	1494	1208	1140	1118	980	907	793	412	363	236	144	87	52

THANK YOU