

Credit Analytics Case Study

Note: - All the analysis is done in a Jupyter notebook in Python.

Observations: -

Data Understanding: -

Origination data – Provides characteristics of the students.

Performance data – Provides Early Risk Score of the students. This is the target variable.

After merging two datasets there are 958 records and 13 columns.

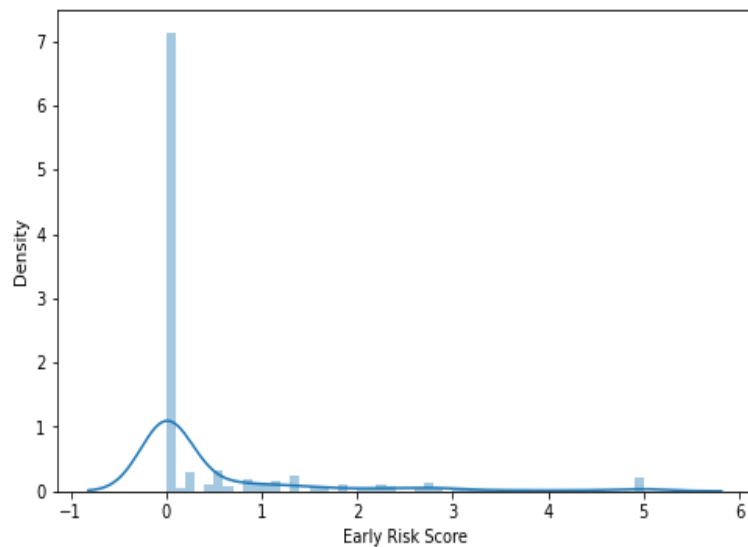
Missing Values: -

Columns	Percentage of missing values	Statistics	Treatment of missing values
Tell Us About You	30.48%		Filled the missing values with “None” considering that a student didn’t mention anything about her/himself.
Credit Score 1	54.28%	count 438.000000 mean 678.831050 std 51.168352 min 528.000000 25% 639.000000 50% 677.500000 75% 721.750000 max 796.000000	It seems to be fairly well distributed variable, where mean is closer to median. We have imputed the missing values with mean.
Credit Score 2	34.03%	count 632.000000 mean 79.083861 std 223.989538 min 0.000000 25% 0.000000 50% 0.000000 75% 0.000000 max 828.000000	The distribution is right skewed with huge deviation. Hence, we didn’t impute the missing values here.

Exploratory Data Analysis

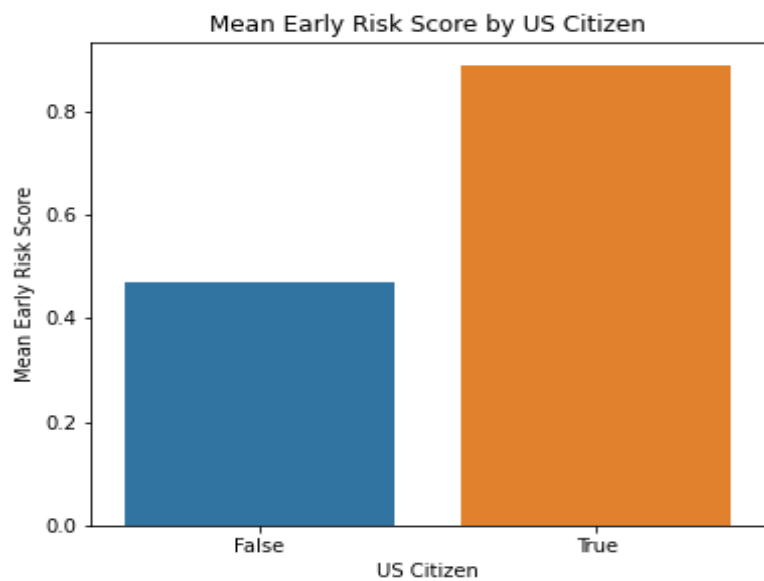
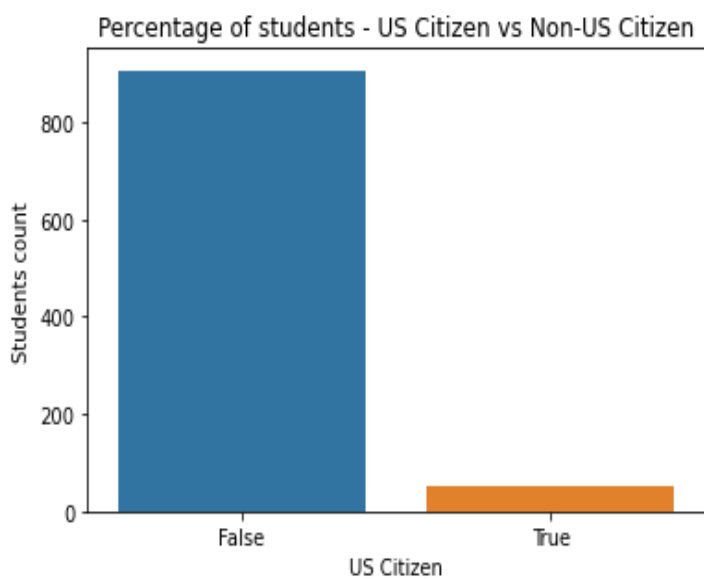
Univariate Analysis

Target Variable – Early Risk Score



For most of the students, the Early Risk Score is zero and the Risk Score is high for very less number of students.

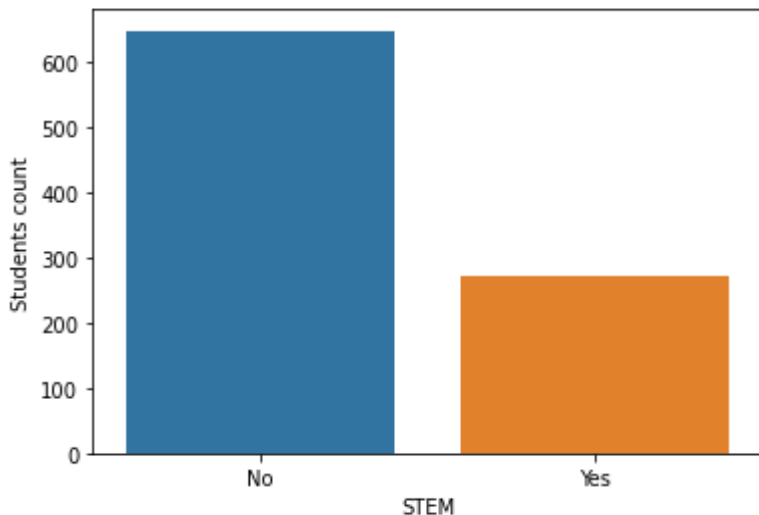
US Citizen



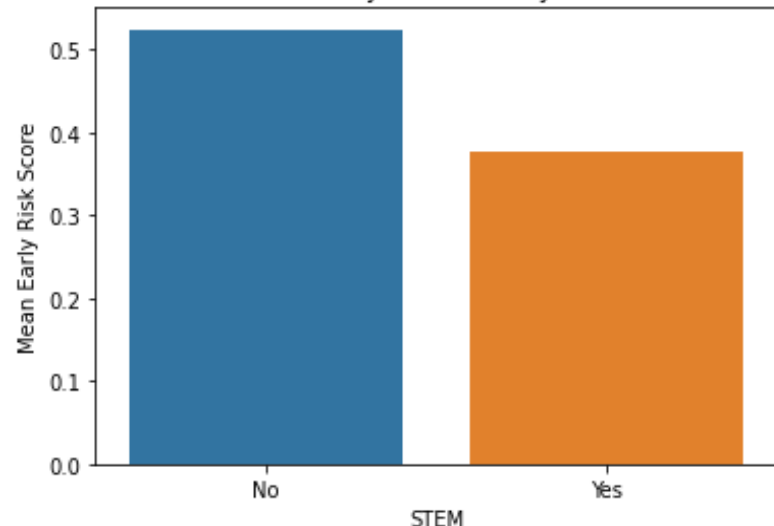
1. The count of US Citizen students is very less compared to Non-US Citizen. There is a huge data imbalance here.
2. The mean Early Risk Score for US citizen is 0.89, which is almost twice as Non-US Citizen (0.47).

STEM

Count of students - STEM vs Non-STEM



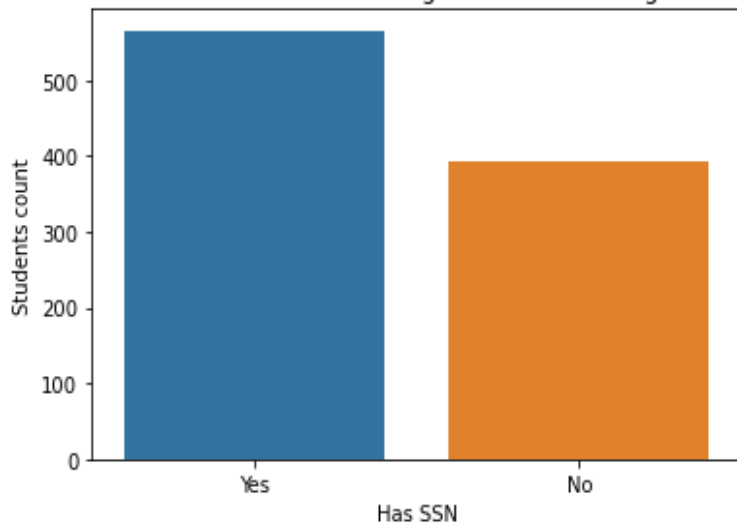
Mean Early Risk Score by STEM



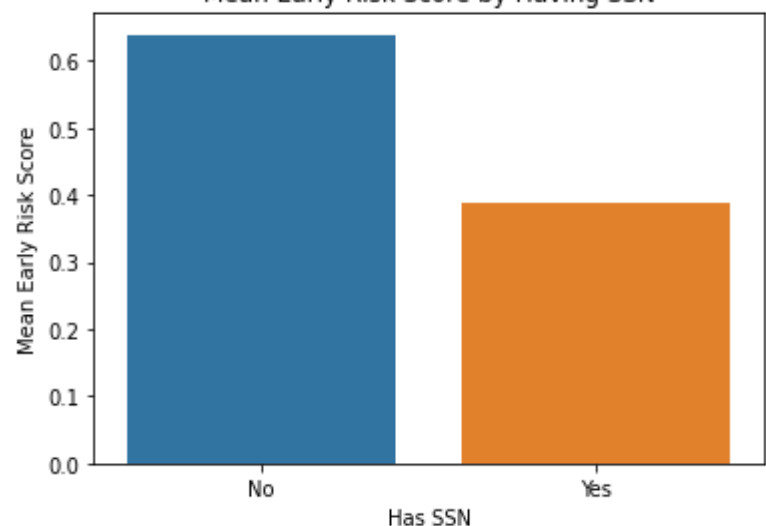
1. The students enrolled with STEM course is very less in comparison with the students enrolled with other courses.
2. The Risk score for STEM students is 0.37, whereas for other students it is 0.52. Meaning the STEM students are less risky than the others.

Having SSN

Count of students - Having SSN vs Not having SSN

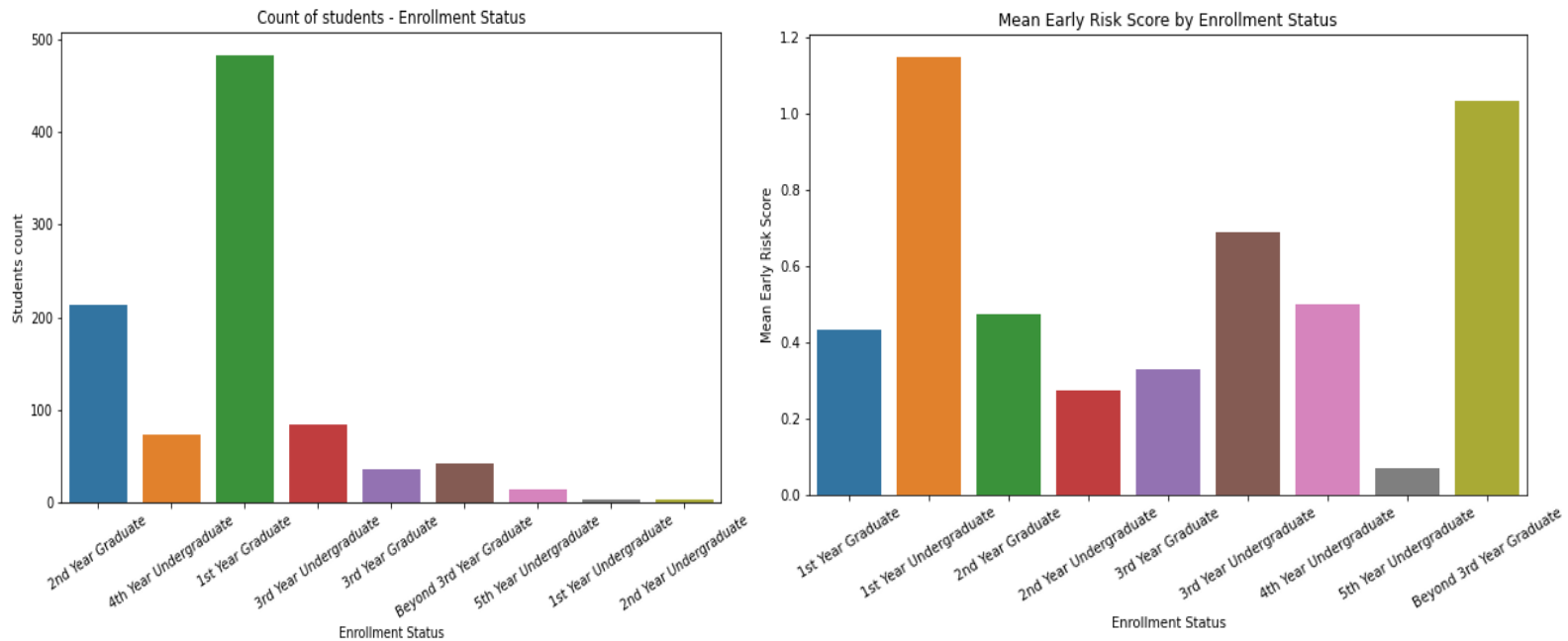


Mean Early Risk Score by Having SSN



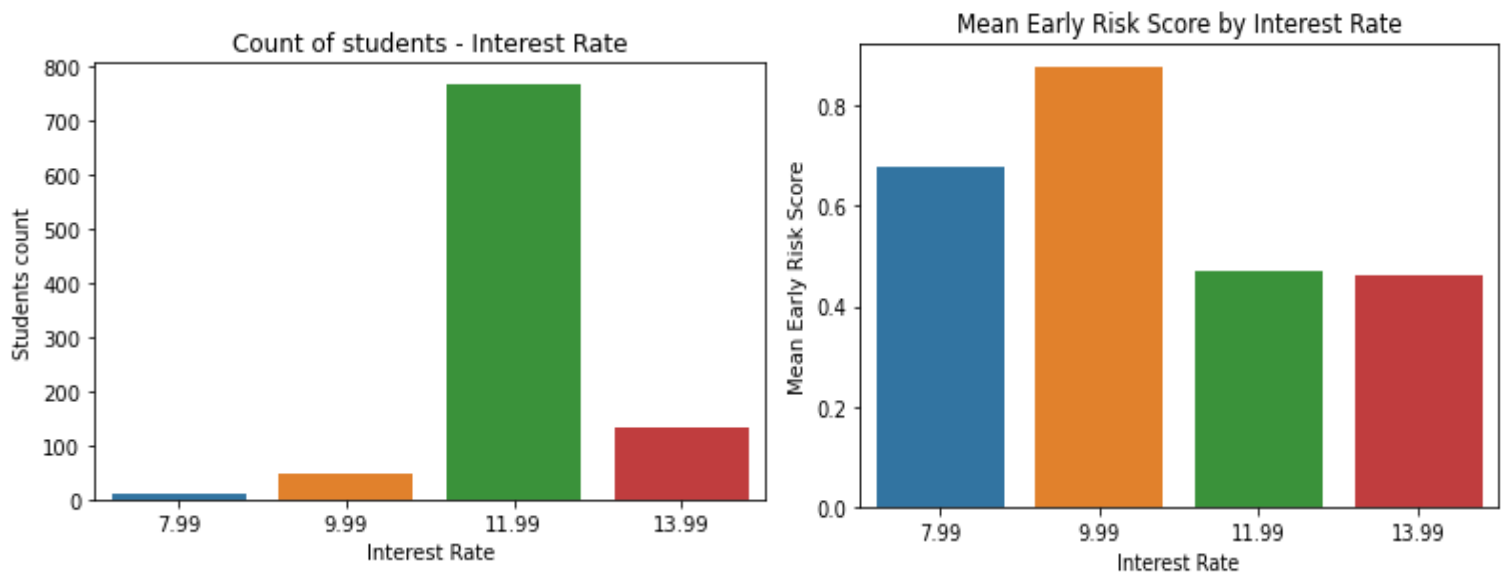
1. The counts are almost comparable for both the groups.
2. The Risk Score (0.63) is quite higher for the students not having SSN than the students having SSN (0.38).

Enrolment Status



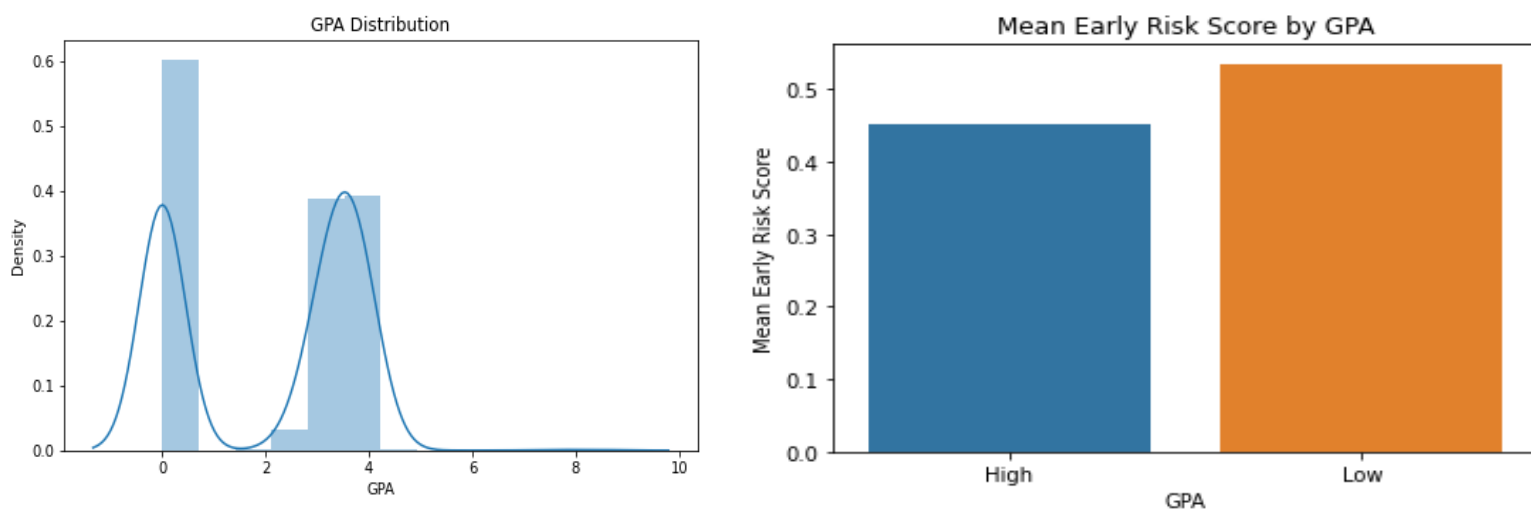
1. There is a data imbalance for few of the groups such as 1st Year Undergraduate, 2nd Year Undergraduate. There are only few records in these groups. Hence, we can not make any conclusion on basis of count of students of these groups.
2. High risky groups are - Beyond 3rd Year Graduate, 3rd Year Undergraduate, 4th Year Undergraduate.
Low risky groups are - 1st Year Graduate, 2nd Year Graduate.

Interest Rate



1. Most of the students were offered the loan with the higher interest rate at 11.99 whereas, very few students were offered the loan with 7.99 and 9.99 interest rate.
2. The count of the students is very less for 7.99 and 9.99 interest rate. Hence, any conclusion for these two groups will be biased. But looks like the risk score doesn't vary for the interest rate of 11.99 and 13.99 and it is lower than the interest rate of 7.99 and 9.99.

GPA

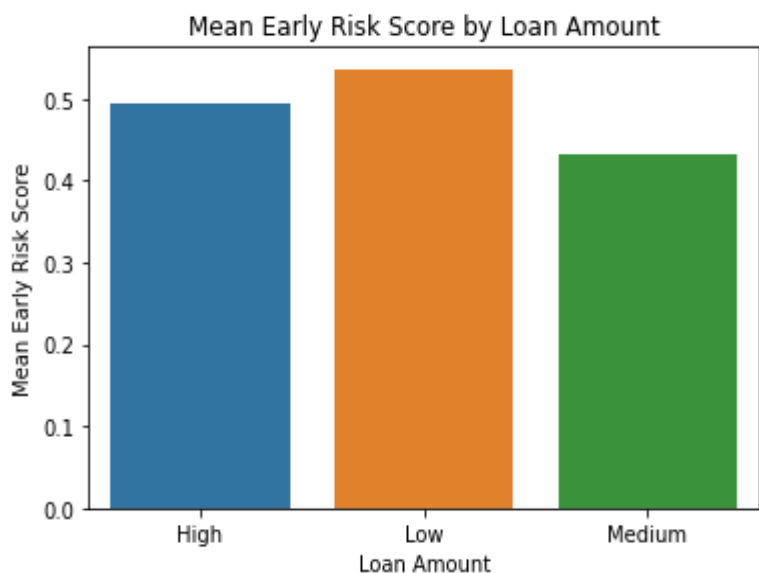
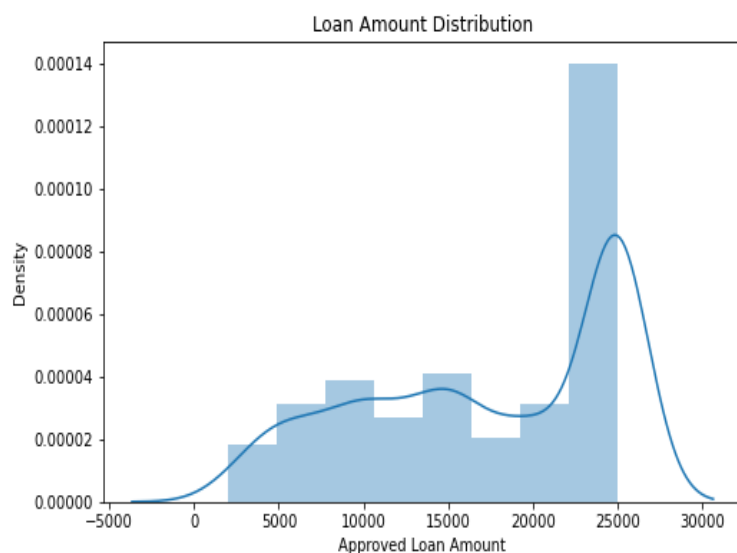


We can see that for most of the students GPAs are zero. Assuming that the GPA is not provided for these students. Approximately 50% of students have GPA below 3.07. We are creating bins as following: -.

- Below 3.0 is considered as Low GPA
- Above 3.0 is considered as High GPA

As expected, the risk score is slightly higher for lower GPA students. However, the difference is not much considerable. Hence, we can conclude that GPA is not a strong indicator for Risk score.

Approved Loan Amount



```
count      958.000000
mean      17397.002985
std       7463.420160
min       2001.000000
25%      10500.000000
50%      18751.000000
75%      25000.000000
max       25000.000000
```

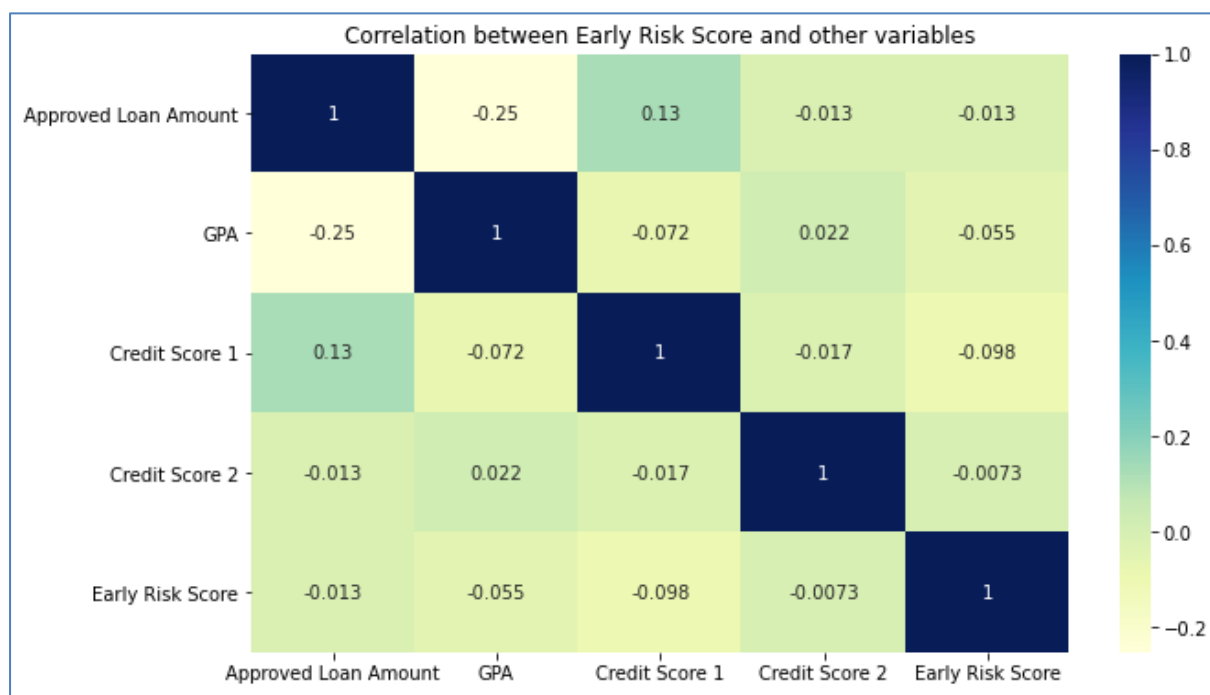
The distribution for loan amount is slightly right skewed. We have created three buckets for approved loan amount as following: -

- < 14000: - Low loan amount
- <= 14000 < 23000: - Medium loan amount
- > 23000: - High loan amount

Surprisingly, there is not much difference in the risk score for these three groups. Just low loan amount students are slightly riskier than the other groups.

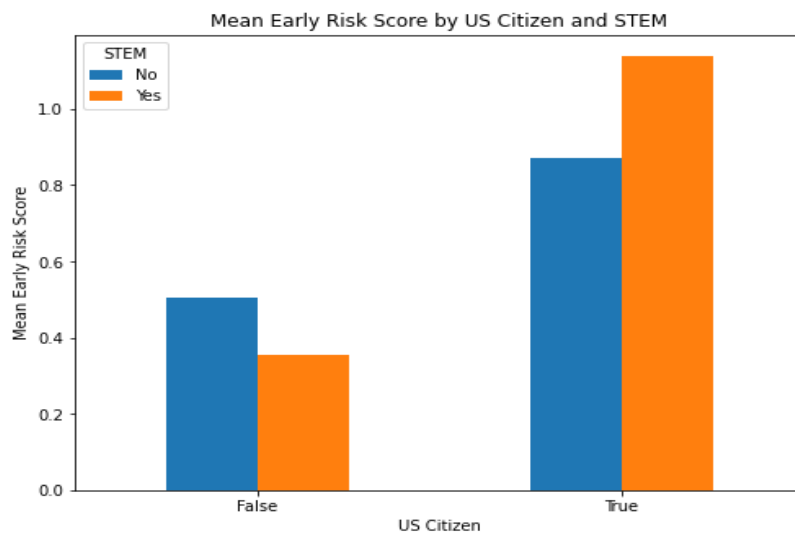
Bivariate Analysis

Correlation between Early Risk Score and continuous variables



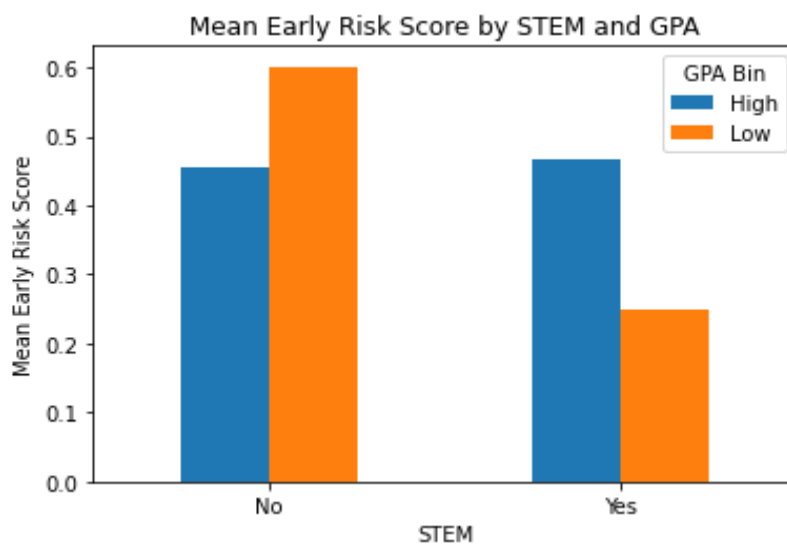
There is very less correlation between Early Risk Score and other variables. Though this observation we have already seen, while binning the Loan amount and GPA variables. These variables have no significance for predicting the risk score.

Mean Early Risk Score by - US Citizen and STEM



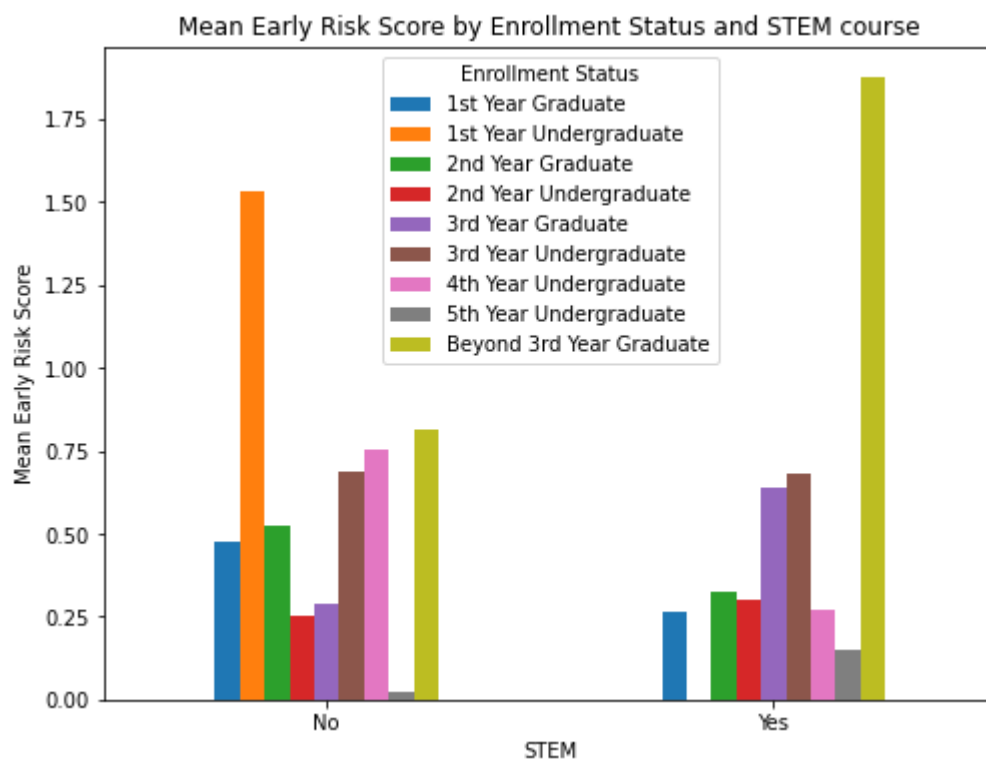
Non-US Citizen with STEM course students are the least risky group, whereas US Citizen with STEM course students are the riskiest group.

Mean Early Risk Score by - GPA and STEM



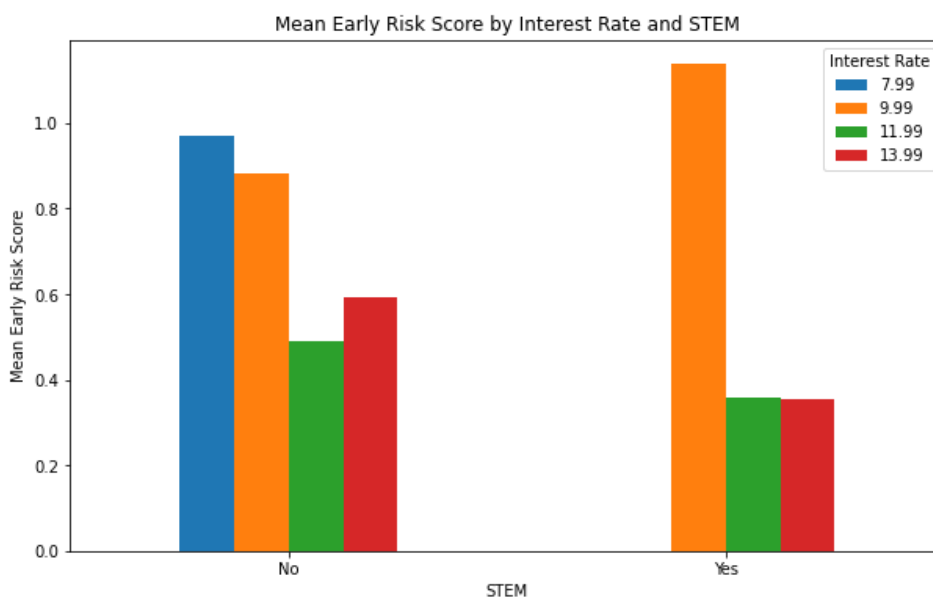
Students other than STEM course with low GPA are the riskiest group.

Mean Early Risk Score by - Enrolment Status and STEM



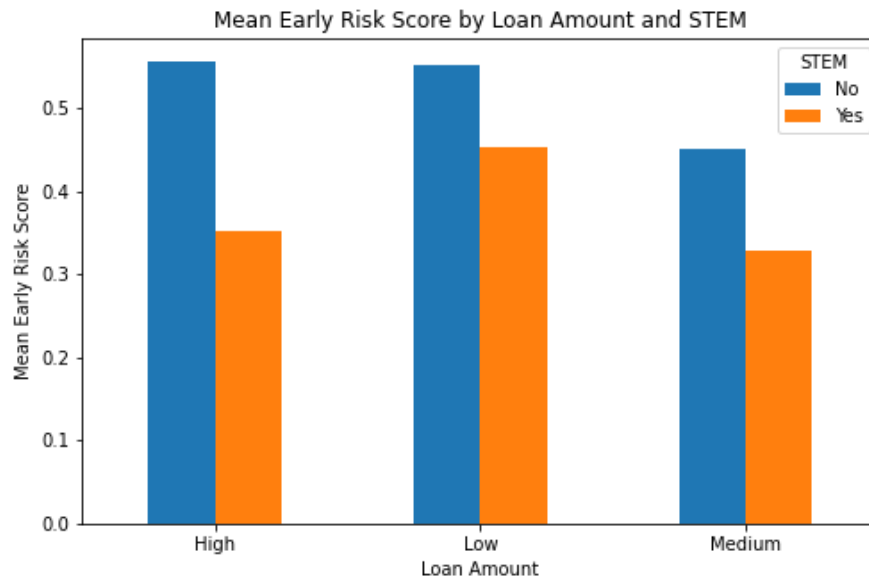
Beyond 3rd year graduates with STEM course is the riskiest group and 1st year graduates with STEM course is the least risky group.

Mean Early Risk Score by - Interest Rate and STEM



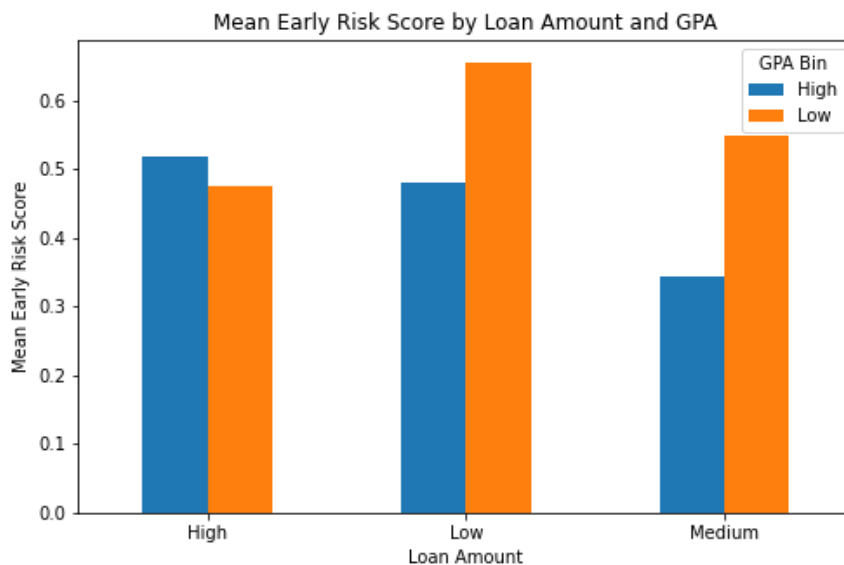
Students with STEM course and offered loan with higher interest rate have lower risk score. Students with non-STEM course with lower interest rate have higher risk score.

Mean Early Risk Score by - Loan Amount and STEM



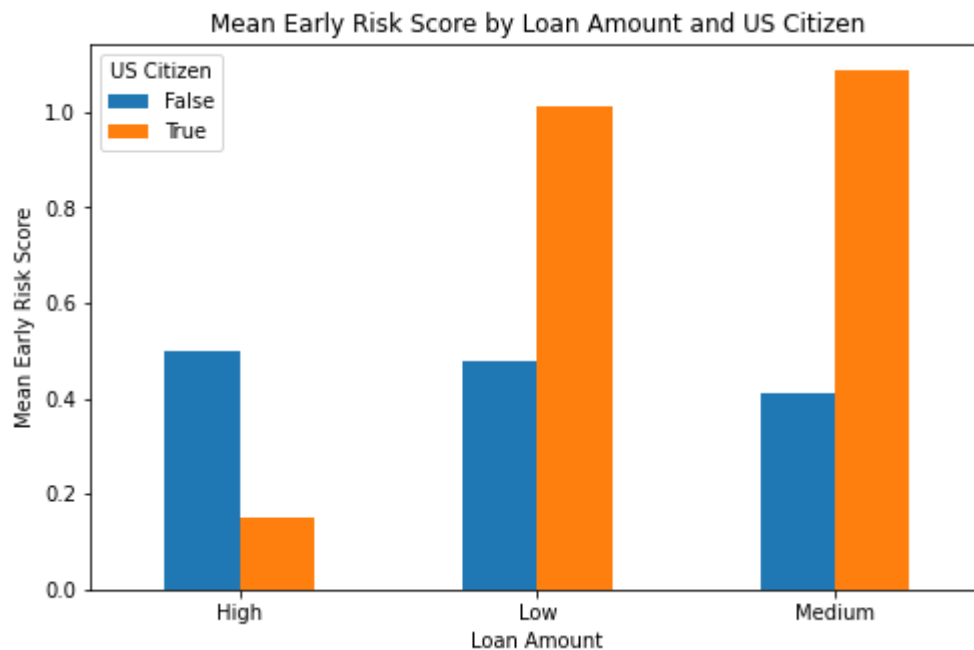
Students enrolled with STEM course and offered higher loan amount are less likely to be riskier than the non-STEM students with high loan amount.

Mean Early Risk Score by - Loan Amount and GPA



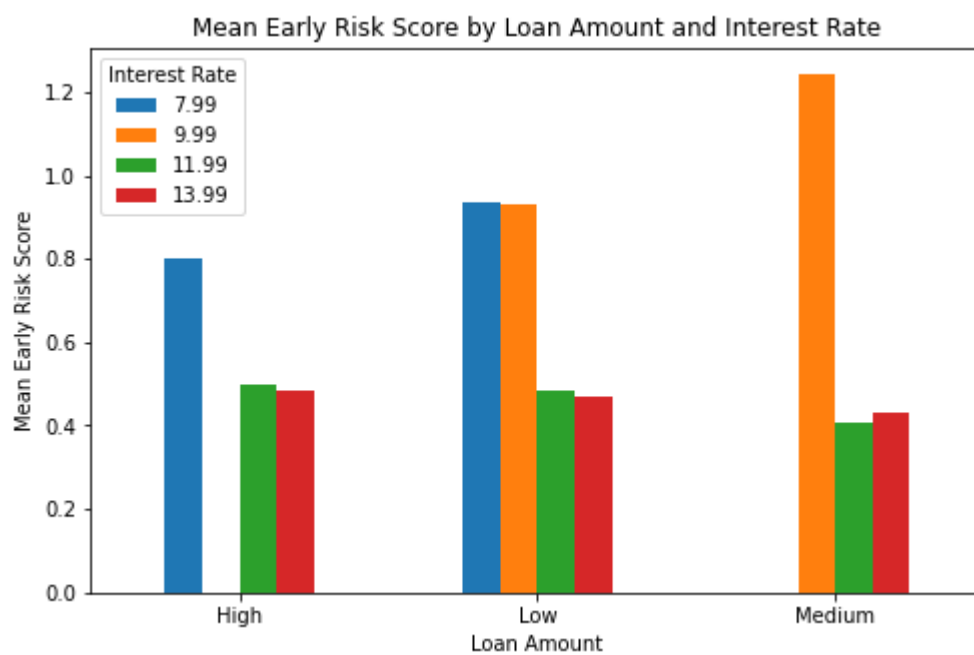
Students with low GPA and offered with low loan amount are slightly riskier than the other groups.

Mean Early Risk Score by - Loan Amount and US Citizen



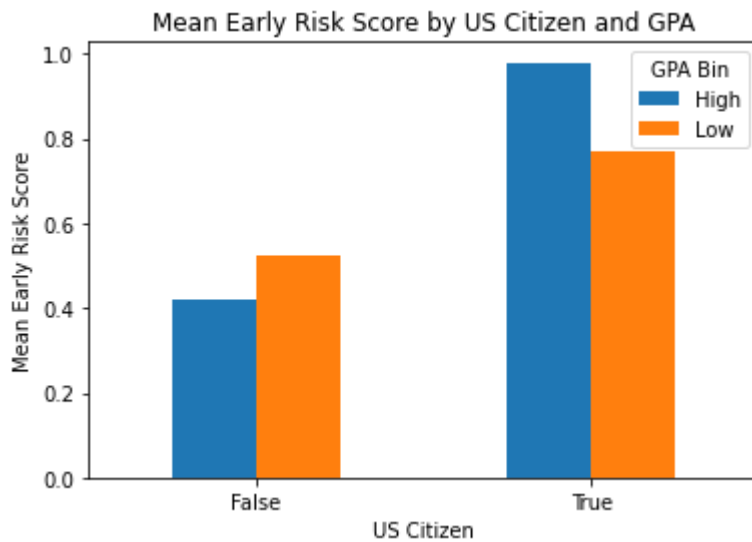
US Citizen students with higher loan amount are less risky than US citizen students with lower loan amount.

Mean Early Risk Score by - Loan Amount and Interest Rate



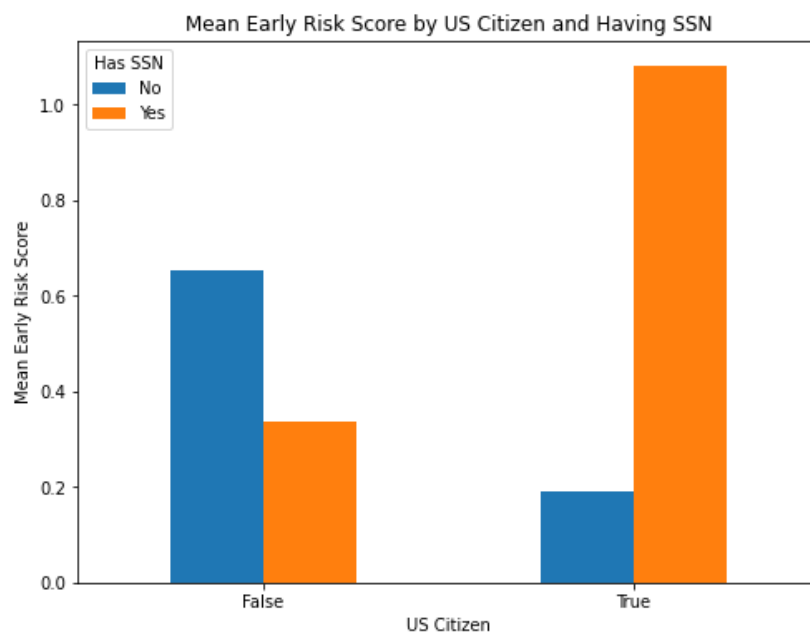
Irrespective of the loan amount the lower interest rate loans are more risky than higher interest rate.

Mean Early Risk Score by - GPA and US Citizen



Non-US citizen with higher GPA are the least risky students and US citizen with high and low GPA are the riskiest students.

Mean Early Risk Score by - US Citizen and SSN



US Citizen students not having SSN are the least risky and US citizen students having SSN are most risky.

Business Recommendations

From the detailed analysis of the data, we can categorise the students in two groups such as high risky and low risky students.

Low Risky Students

1. Students enrolled with STEM course.
2. 1st year graduate and 2nd year graduate students.
3. Students offered loan with higher interest rate.
4. STEM course students offered with higher loan amount.
5. Students, who are not US citizen and not having SSN.
6. Students, who are Non-US citizen and enrolled with STEM courses.
7. Students enrolled with STEM course and offered loan with higher interest rate.
8. Non-US citizen with higher GPA are less risky than US citizen.

High Risky Students

1. Students enrolled with Non-STEM courses.
2. Students beyond 3rd Year Graduate, 3rd Year Undergraduate and 4th Year Undergraduate.
3. Students offered loan with lower interest rate.
4. Students, who are US citizen and enrolled with STEM courses.
5. Students enrolled with Non-STEM courses and low GPA. That means the risk score is high, if a Non-STEM student performs poor.
6. Students enrolled with a Non-STEM course and offered loan with lower interest rate.
7. Non-STEM students offered with higher loan amount.

Note: -

There is very little correlation between Early Risk Score and approved loan amount, credit score and GPA. Hence, these variables are not good indicators for risk score.

Conclusion: -

It is recommended to grant loan with the lowest risk for the students, who are enrolled with STEM courses, offered loan with higher interest rate. More preferably, 1st year and 2nd year graduate students, who are not US citizen and not having SSN.

On the other hand, it is riskier to provide loans to the students, who are not enrolled with STEM courses, having low GPA, offered loan with lower interest rate, 3rd Year Undergraduate, 4th Year Undergraduate, beyond 3rd year graduate.