Loan Approval (Report)

GOAL : To decide to grant a loan to a particular small business? Why or why not?

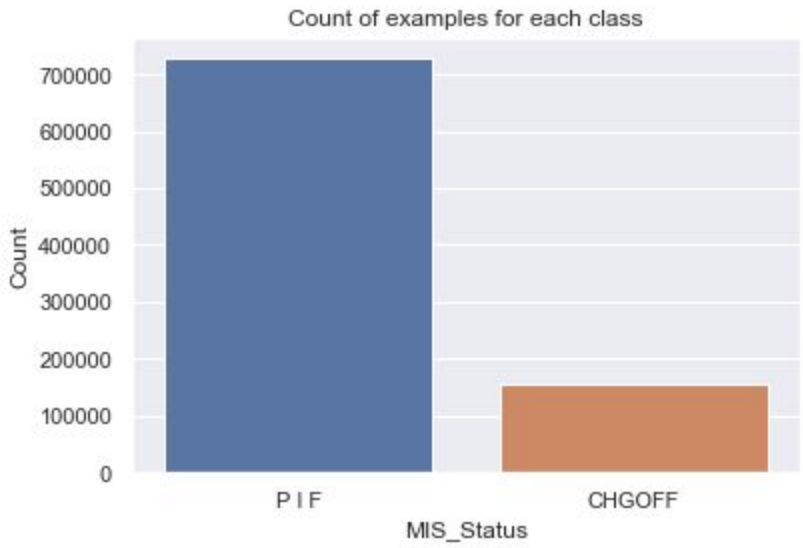
Brief of Model

- The model applied to predict the target variable is a Neural Network with 3 hidden layers with 'ReLU' activation. There is only one output unit with 'Sigmoid' activation.
- **Optimizer** : RMSprop optimizer is used, with learning rate (alpha) = 0.001, rho = 0.9, epsilon = 1e-08 and zero decay.
- **Loss function** : Binary_crossentropy loss is used.
- Also learning rate reduction is used to achieve better results.

Results

	precision	recall	f1-score	support
0	0.97	0.99	0.98	23380
1	1.00	0.99	1.00	109123
accuracy			0.99	132503
macro avg	0.98	0.99	0.99	132503
weighted avg	0.99	0.99	0.99	132503

- The overall accuracy of the model is 99%.
- The model is able to handle the positive class examples really well with 100% precision and 99% recall, but precision is low for negative class.
- **Class Distribution** : As seen below, the data is 'highly skewed'.

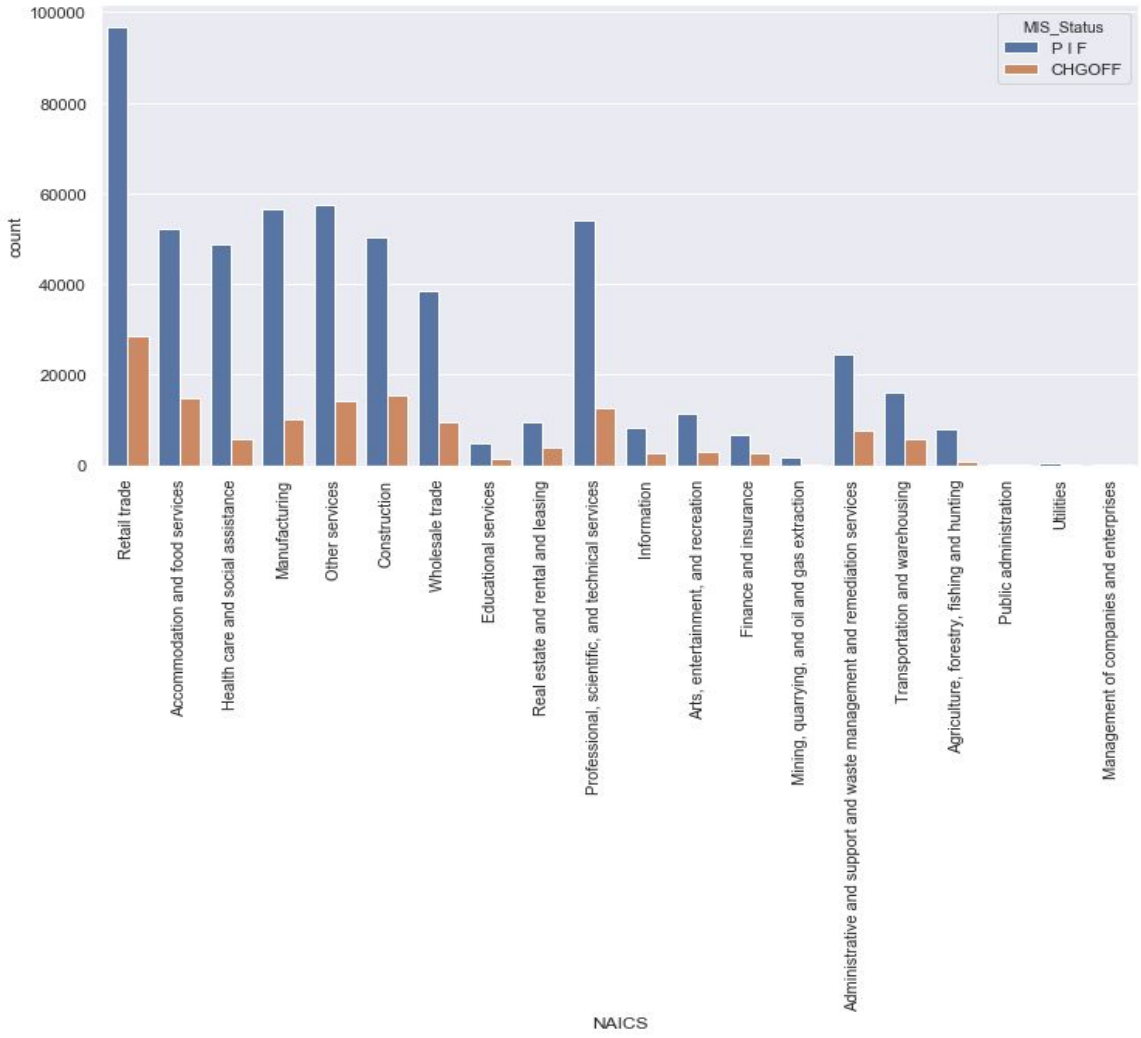


- Due to less examples of negative class (‘CHGOFF’) we observe lower precision for it.

Risk Indicators

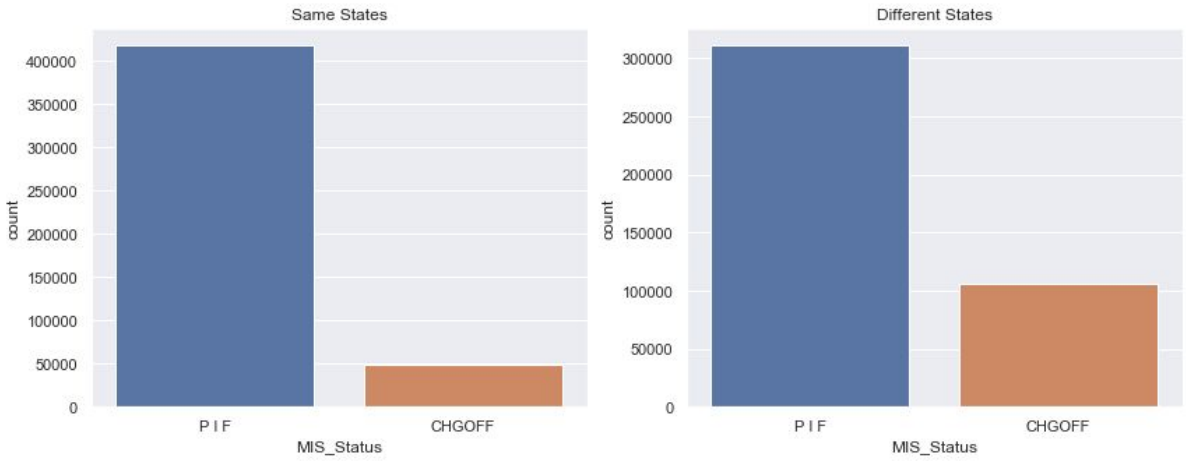
- **NAICS** (North American Industry Classification System) :

Description	Default rate (%)
Mining, quarrying, and oil and gas extraction	8
Agriculture, forestry, fishing and hunting	9
Management of companies and enterprises	10
Health care and social assistance	10
Utilities	14
Public administration	15
Professional, scientific, and technical services	19
Wholesale trade	19
Manufacturing	19, 16, 14
Other services (except public administration)	20
Arts, entertainment, and recreation	21
Accommodation and food services	22
Retail trade	22, 23
Construction	23
Administrative/support & waste management/remediation Service	24
Educational services	24
Information	25
Transportation and warehousing	27, 23
Finance and insurance	28
Real estate and rental and leasing	29

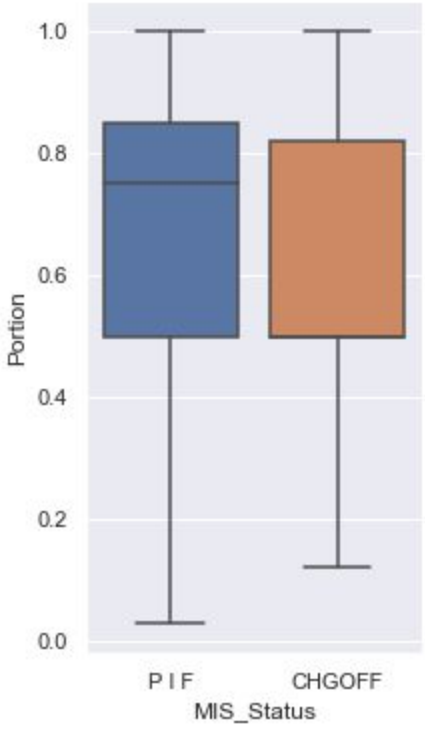


From the above graph and table, it is found that the industry with lowest default rates belong mostly to ‘primary’ sector industries such as agriculture and fishing. Also the healthcare industry is another safe option.
The most risky ones include the industries where profit and loss situations are extreme, like real estate and finance.

- **Borrower’s and Bank’s state** :
  The charts below suggest that if the borrower and the bank have the same state (‘location’) then it is a less risky application and is probable to be paid in full, that is it should be approved. Whereas if they have different states then there's a high probability that the loan will end up as a default and hence should not be approved.



- **Portion of SBA approved** :
  It is the ratio of the amount of loan SBA guarantees and the gross amount approved by the bank.

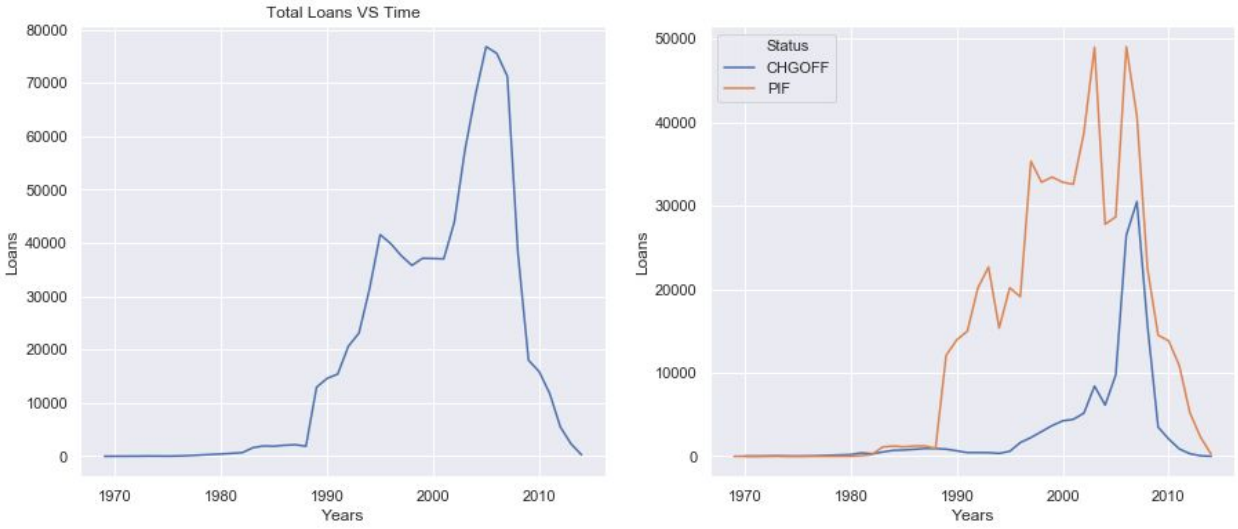


These boxplots show that typically loans that are paid in full have a slightly higher SBA - guaranteed portion.

- **Real Estate** :
  Whether a loan is backed by real estate is another risk indicator. Generally loans that have a term greater than 20 years are backed by real estate thus assuming this, it is

found that loans backed by real estate have a very high chance to be paid in full. While the ones not backed by real estate have no such significant assurance.

- Economic Recession :



The default rate (CHGOFF in the graph) suddenly peaked at around 2008. It is due to the Great Recession which lasted from Dec 2007 to Jun 2009, thus affecting the loans active during that interval.

Conclusion

Characteristic	Approved	Declined
Industry	Primary sector	Extreme situational
State	Same as bank	Different from that of bank
SBA portion	Greater than 85%	Less than 85%
Backed by RealEstate	yes	no
Active during Great Recession	no	yes