TURBO
TITLE LOAN

Utilizing Quant Model for Screening High Quality Customers

---- Based on the analysis of 2016 customer info datasets

Presented by Xiaoyue Sun

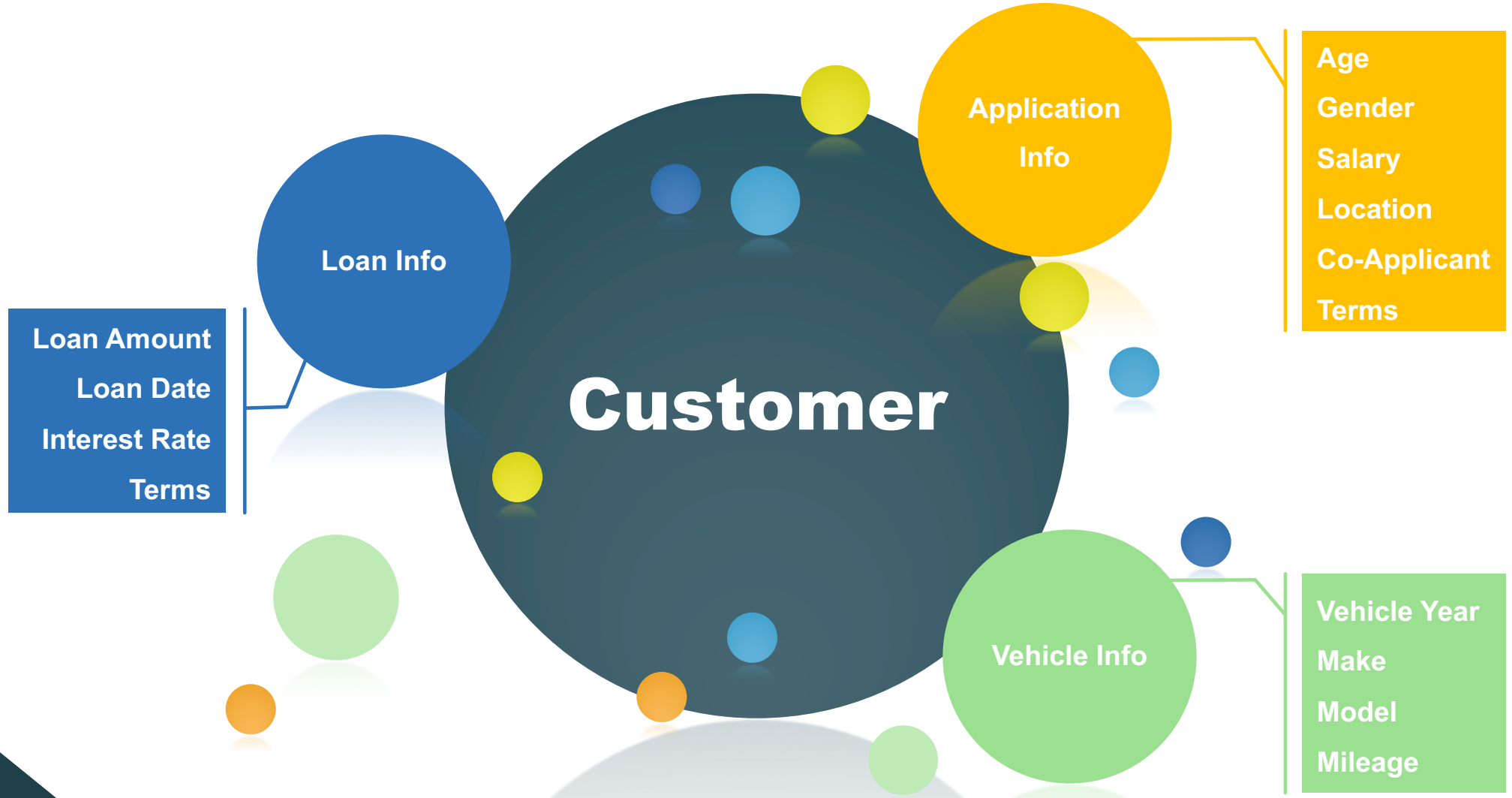
November 29, 2017

TURBO
TITLE LOAN

CONTENT

- One Dimensional Customer Feature Demo
- Two Dimensional Customer Feature Analysis
- Modeling for Customer Default Probability
- Summary

Dataset Info Demo

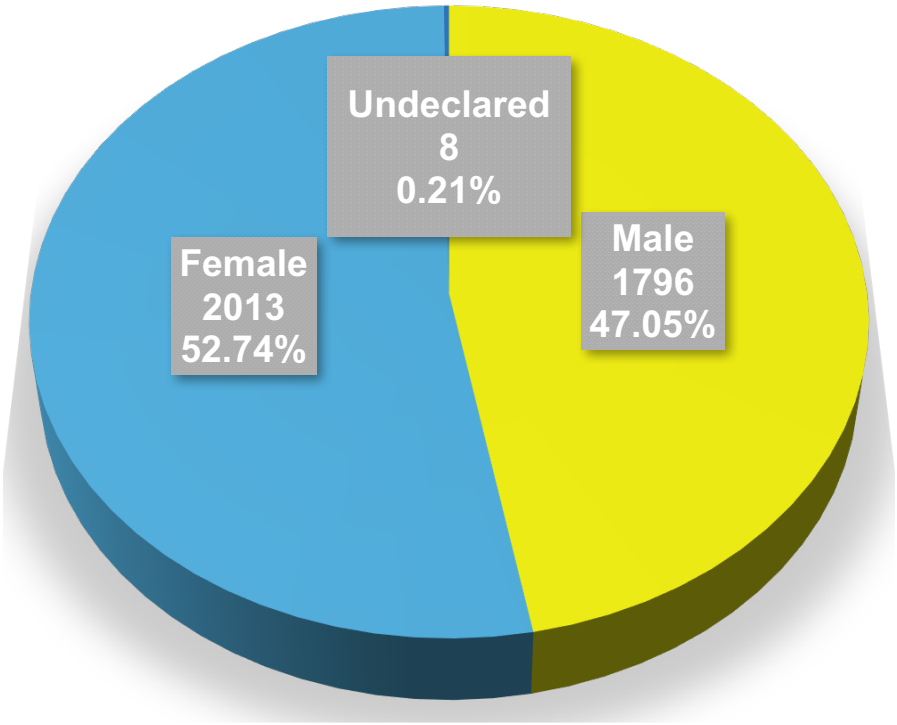


TURBO
TITLE LOAN

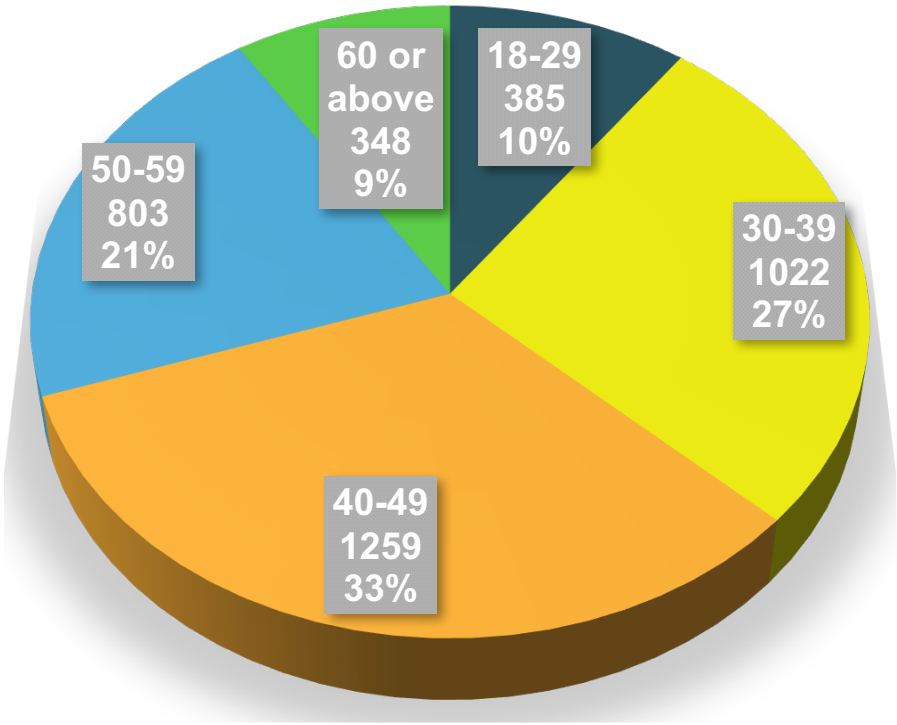
Part 1

One Dimensional Customer Feature Demo

Gender Distribution

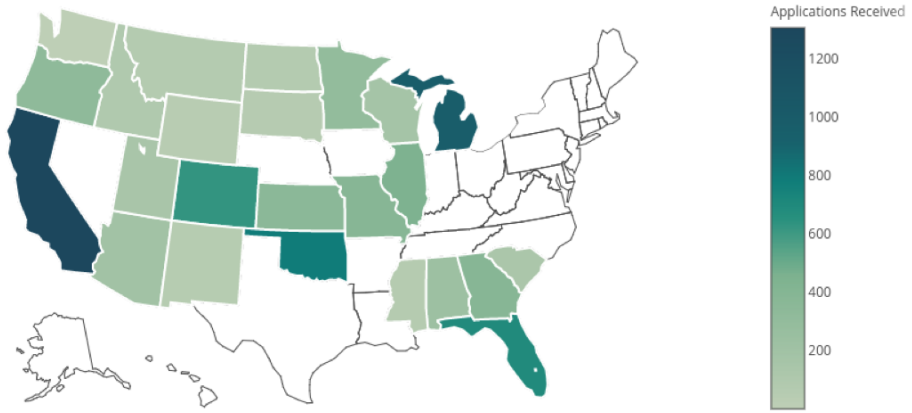


Age Distribution

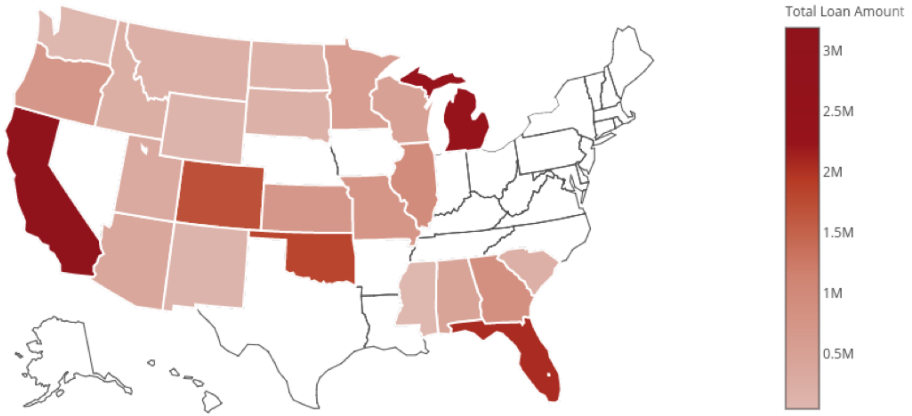


One Dimensional Customer Feature Demo

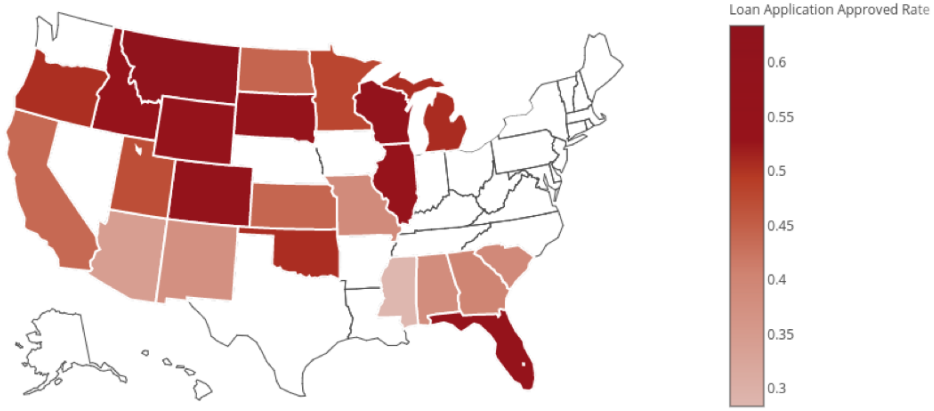
2016 Loan Applications Received by State



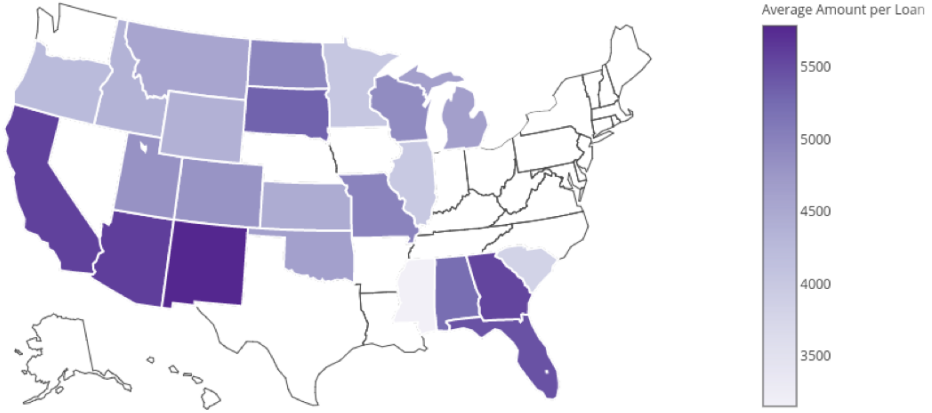
2016 US Loan Amount by State



2016 Loan Application Approved Rate by State Excluding WA

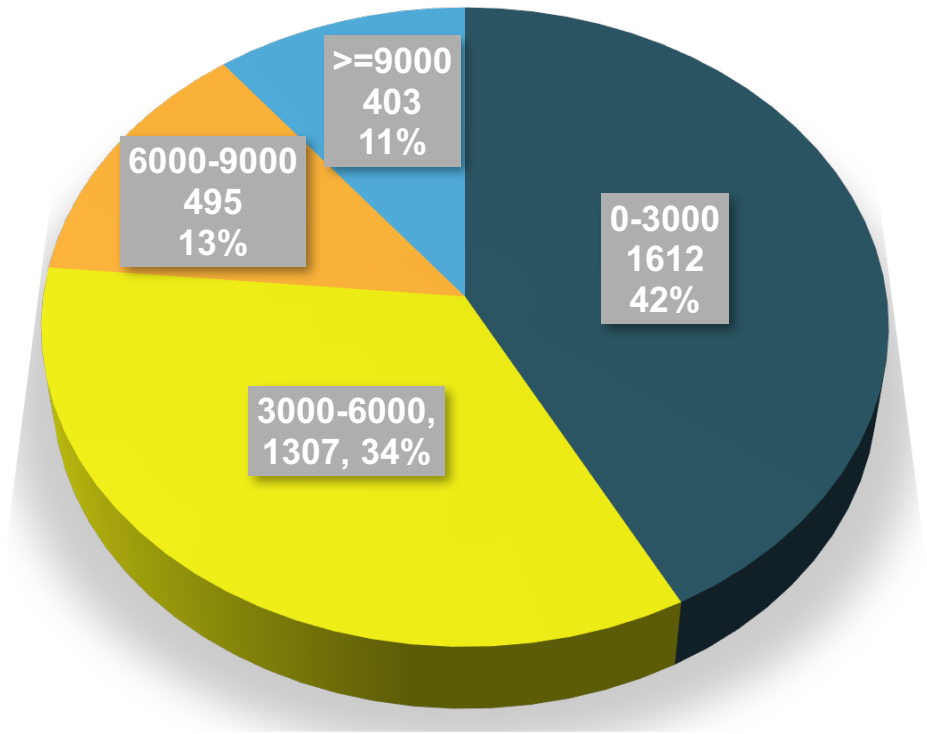


2016 US Average Amount per Loan by State Excluding WA

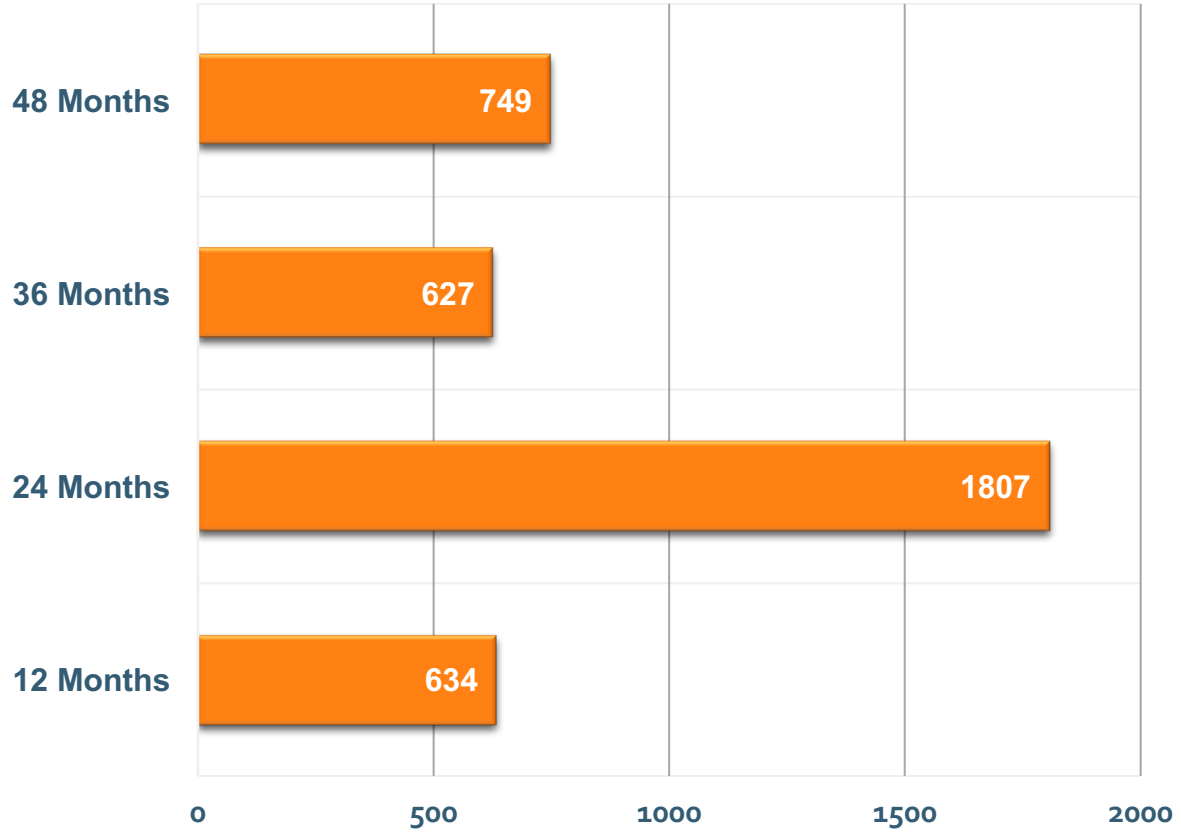


One Dimensional Customer Feature Demo

Loan Amount Distribution by Contract

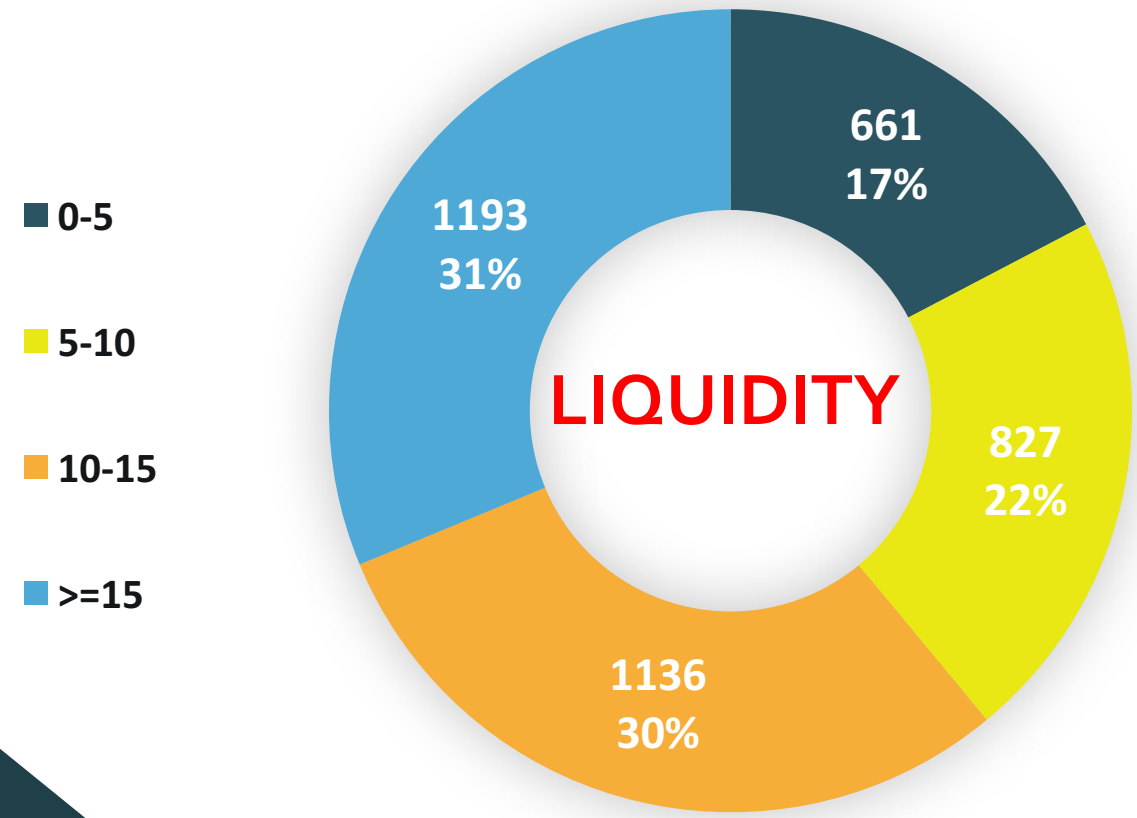


Terms by Contract Distribution



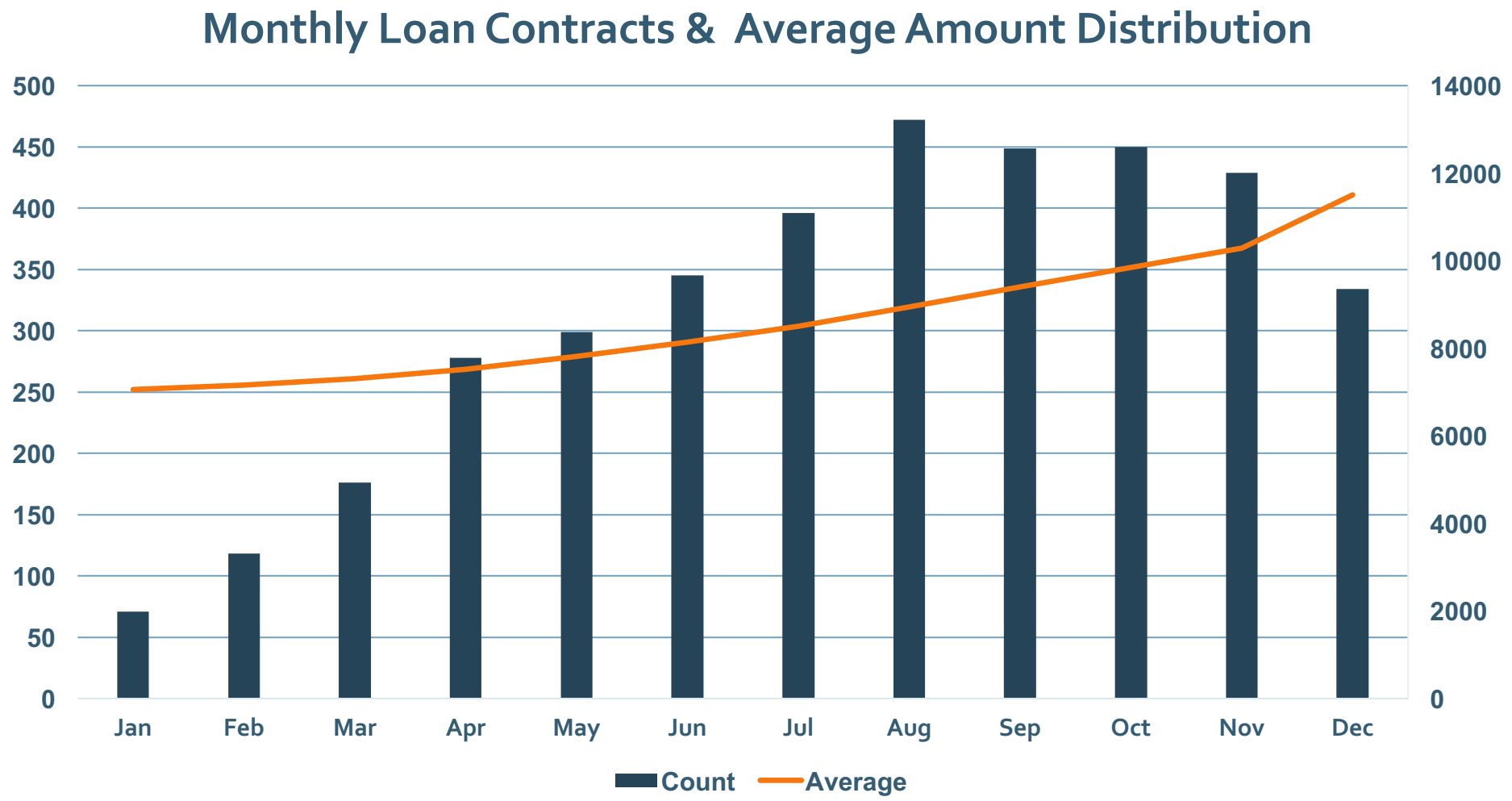
One Dimensional Customer Feature Demo

SP Ratio Distribution



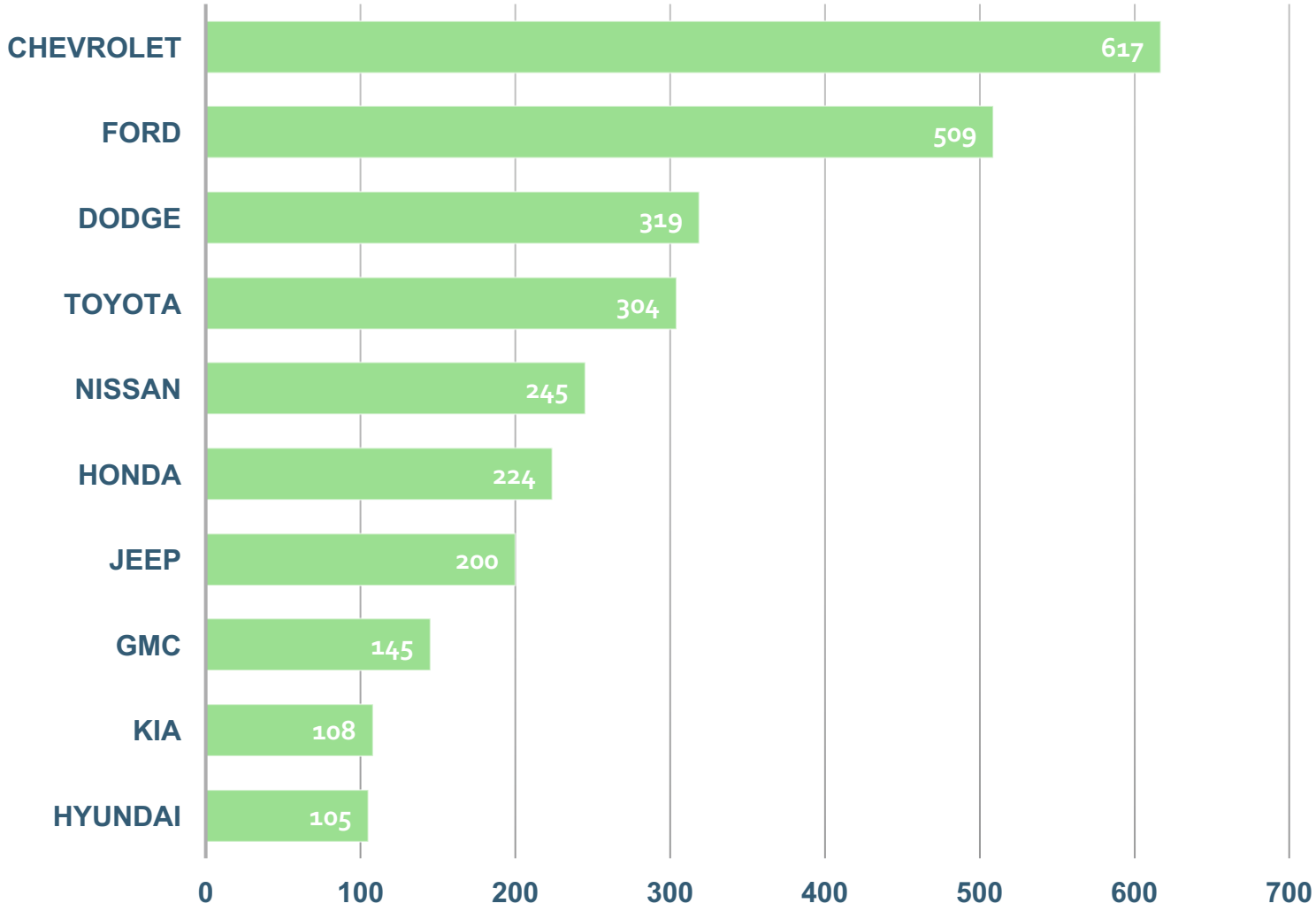
$$SP\ Ratio = \frac{Monthly\_Salary}{Monthly\_Payment}$$

One Dimensional Customer Feature Demo



One Dimensional Customer Feature Demo

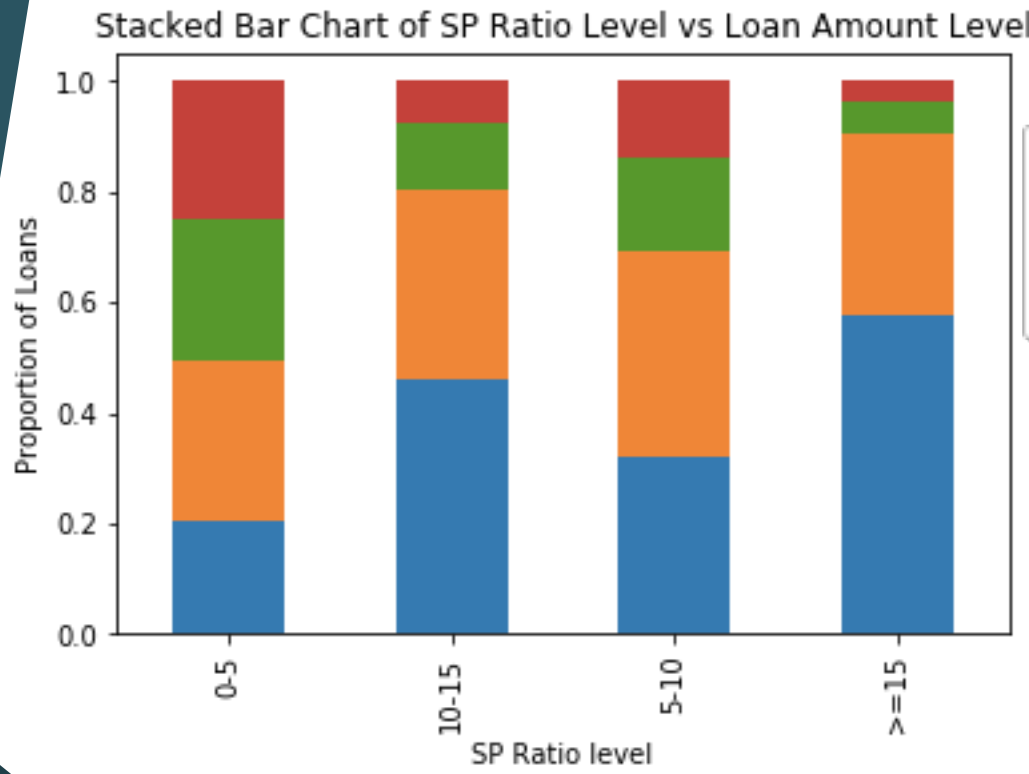
Vehicle Make Distribution

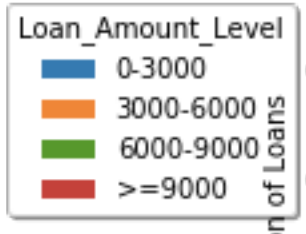


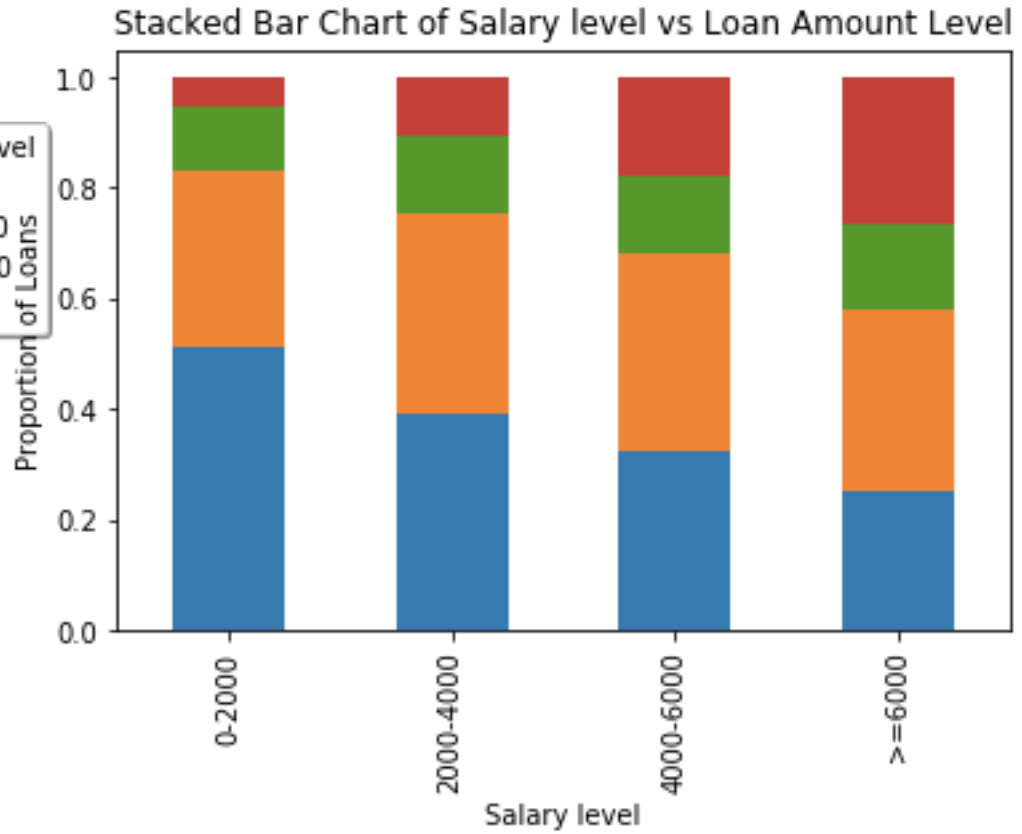
TURBO
TITLE LOAN

Part 2

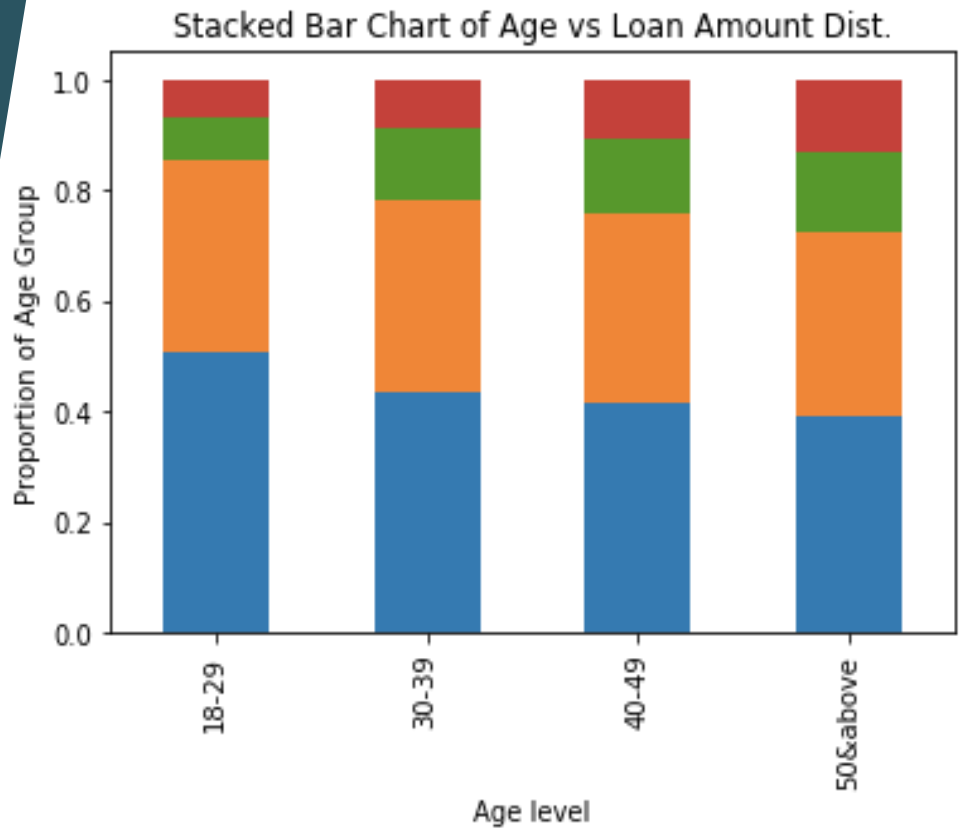
Two Dimensional Customer Feature Analysis

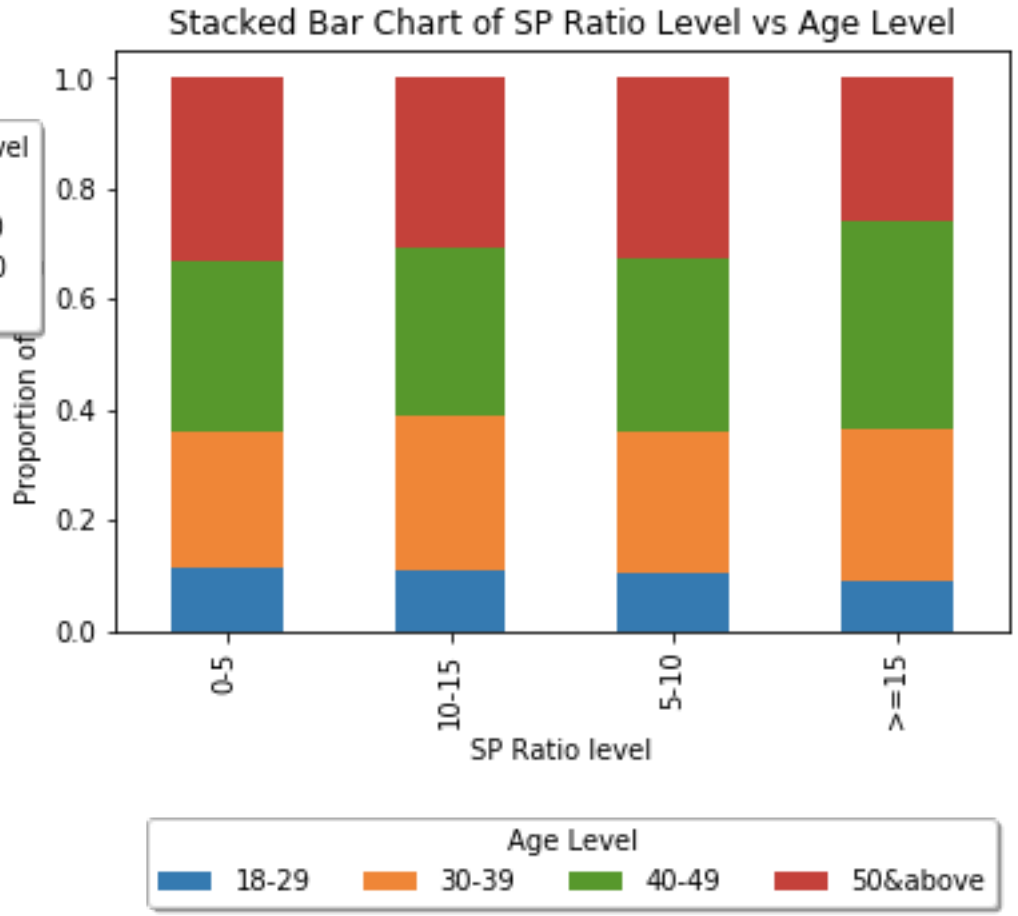






Two Dimensional Customer Feature Analysis





TURBO
TITLE LOAN

Part 3

Modeling for Customer Default Probability

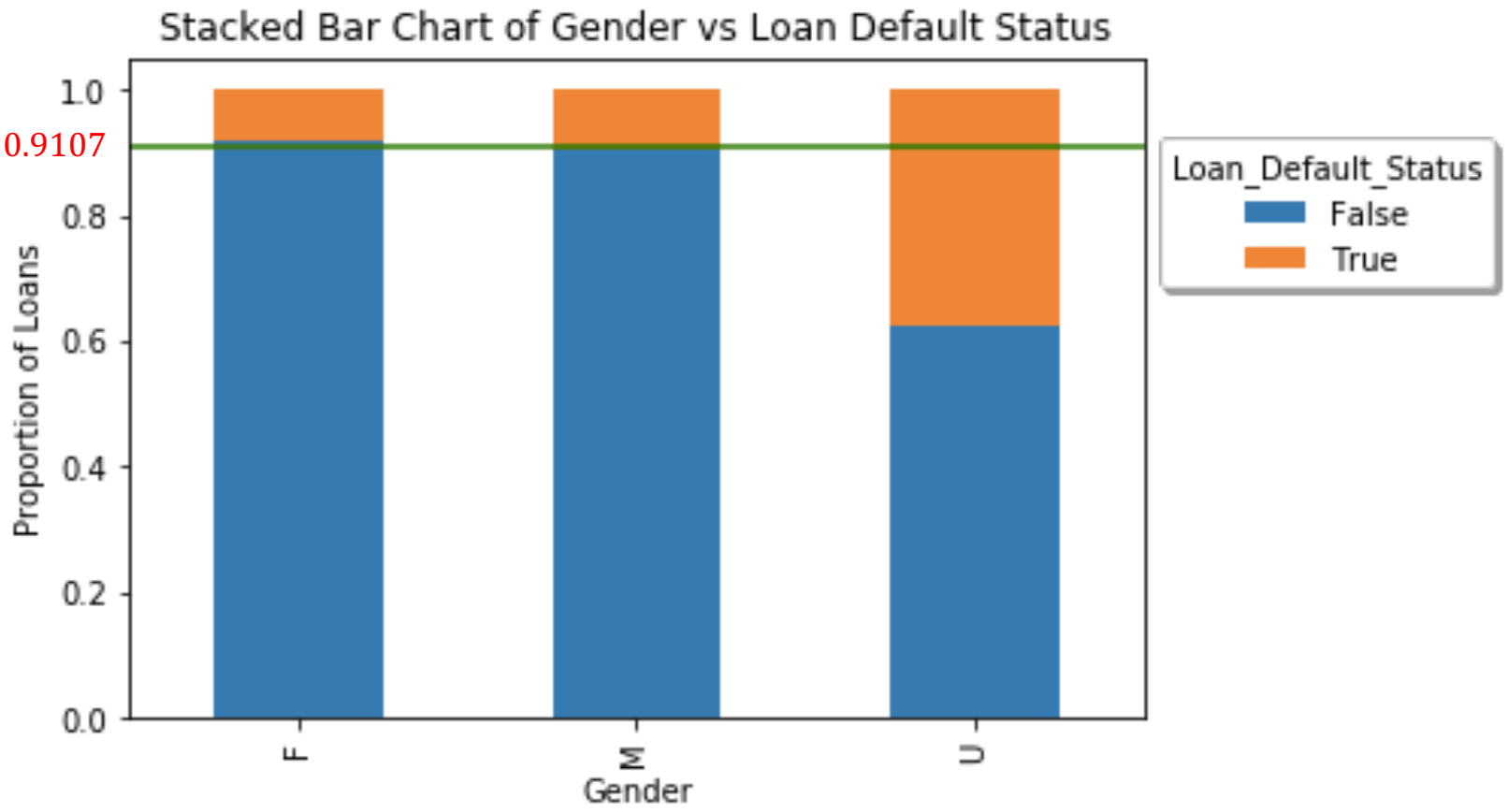
- Why Logistic Regression?
- Would there be some rules to help identify important factors for regression?

Modeling for Customer Default Probability

2D Demo against Loan Default Status

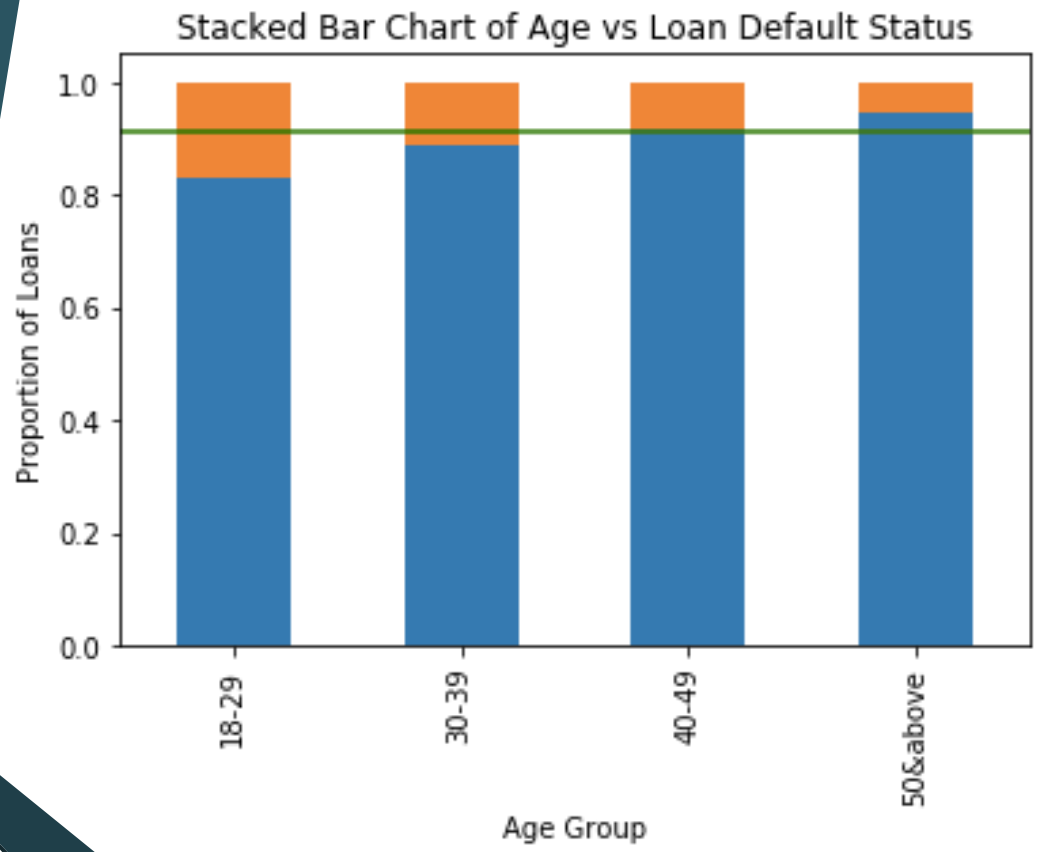
$$Default\ Rate = \frac{Number\ of\ Default\ contracts\ in\ 2016}{Total\ number\ of\ contracts\ in\ 2016} = \frac{341}{3817} = 0.0893$$

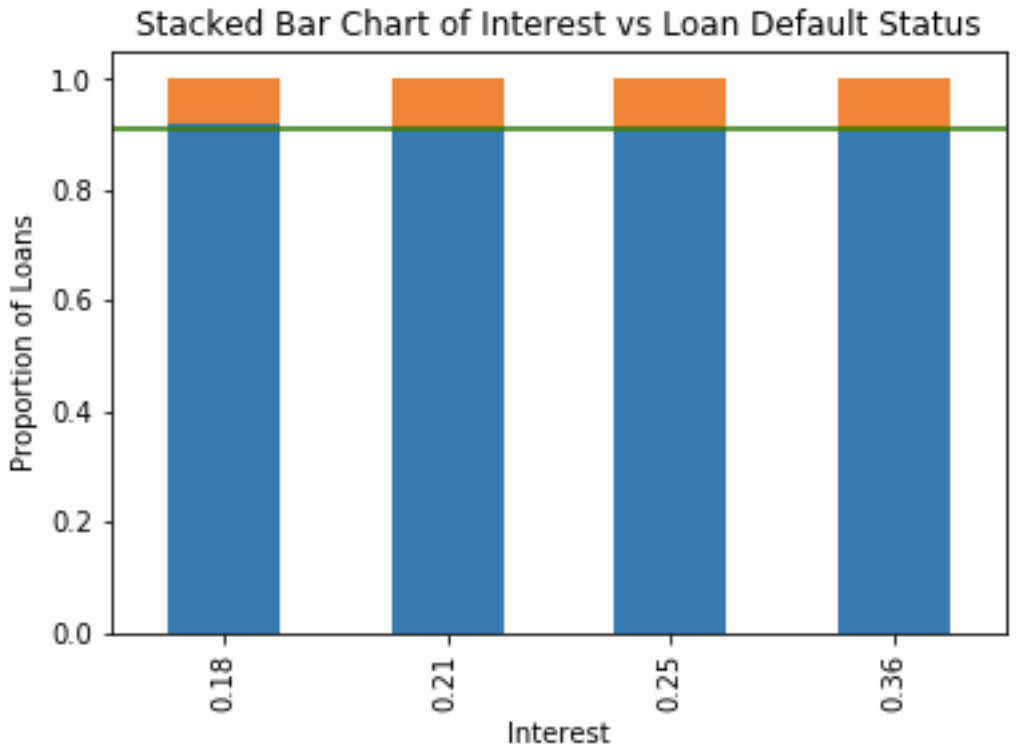
$1 - 0.0.893 = 0.9107$



Modeling for Customer Default Probability

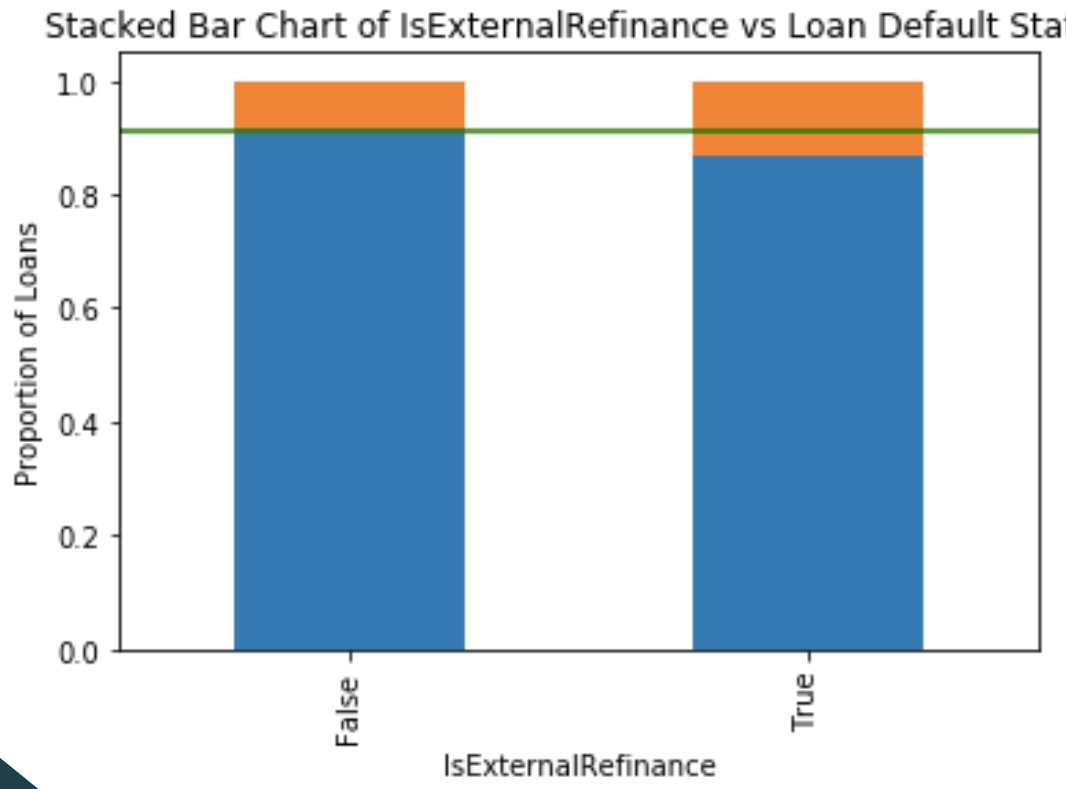
2D Demo against Loan Default Status

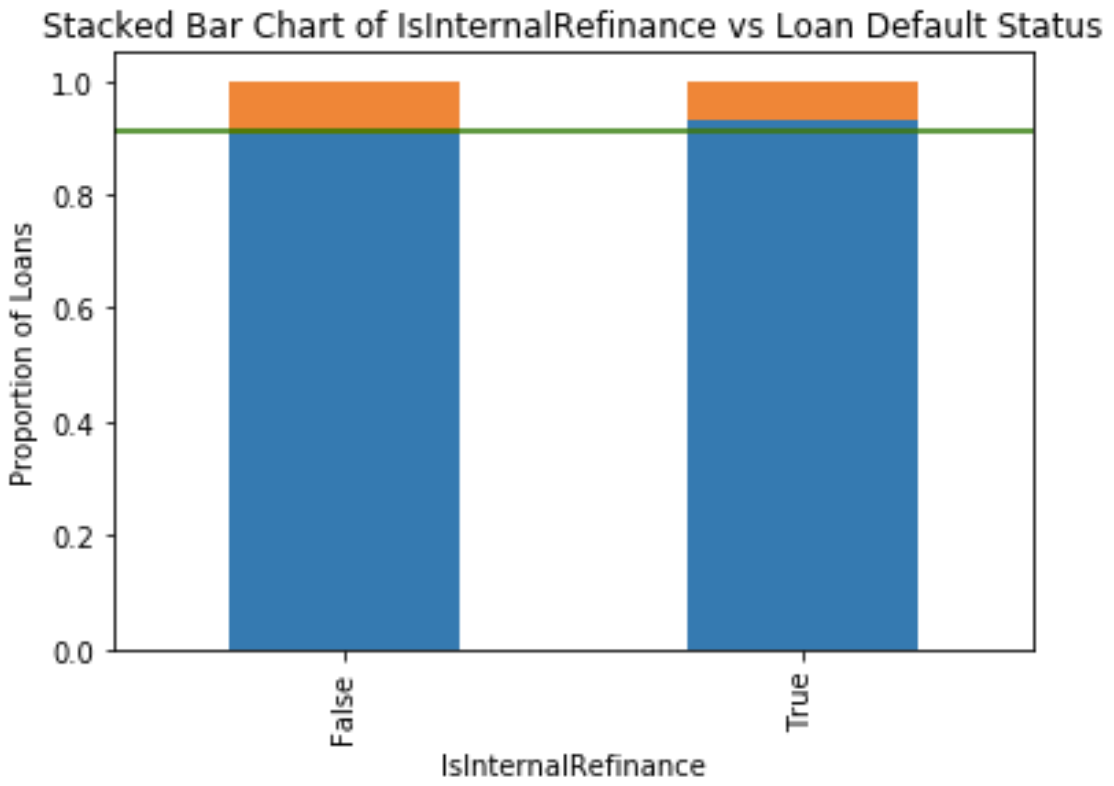




Modeling for Customer Default Probability

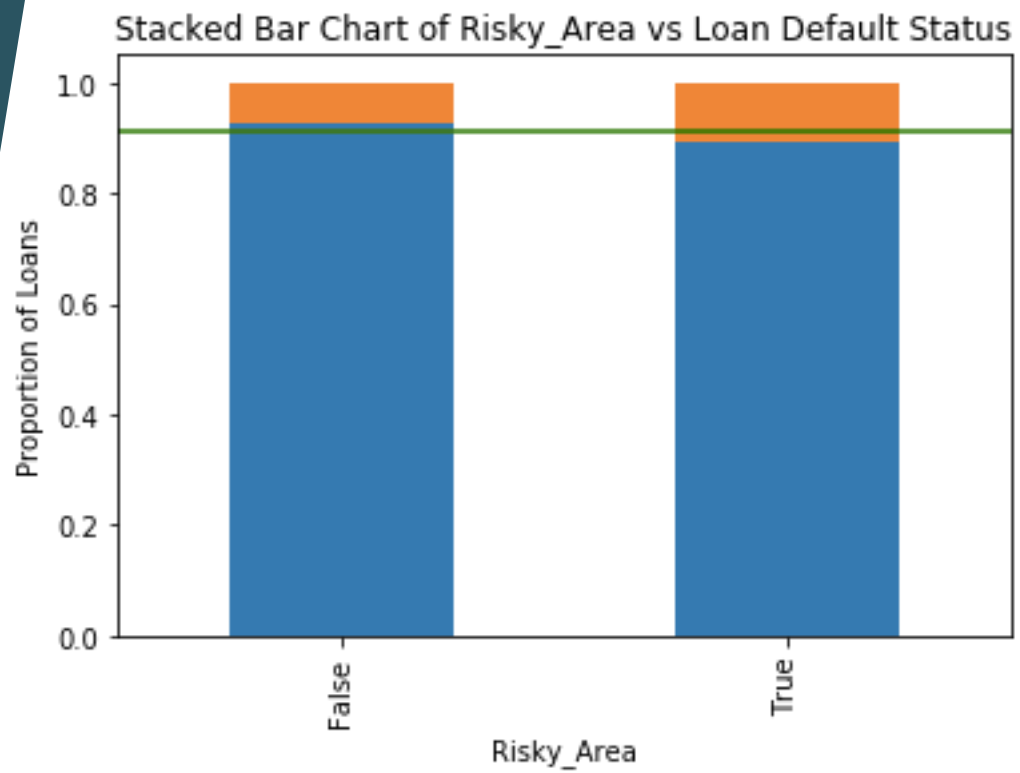
2D Demo against Loan Default Status

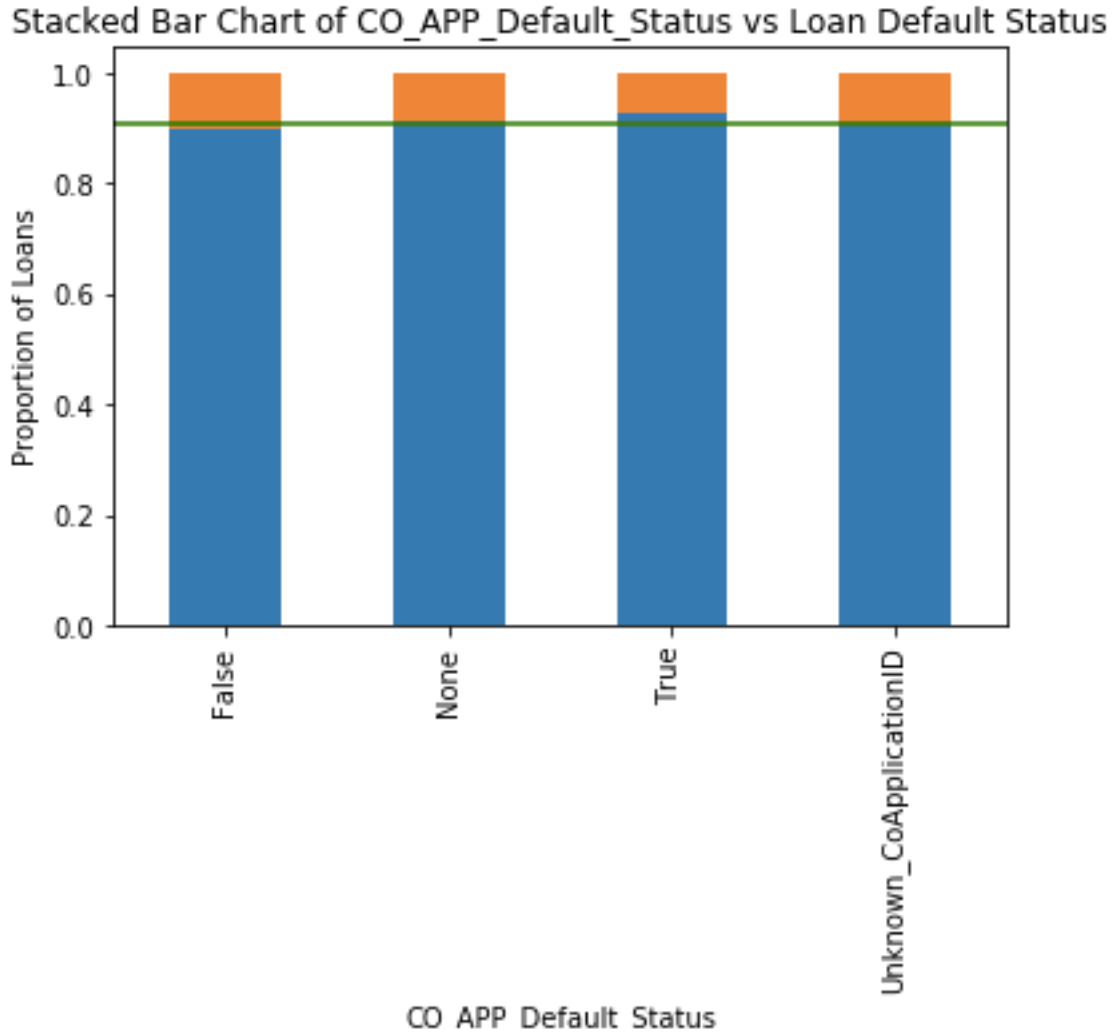




Modeling for Customer Default Probability

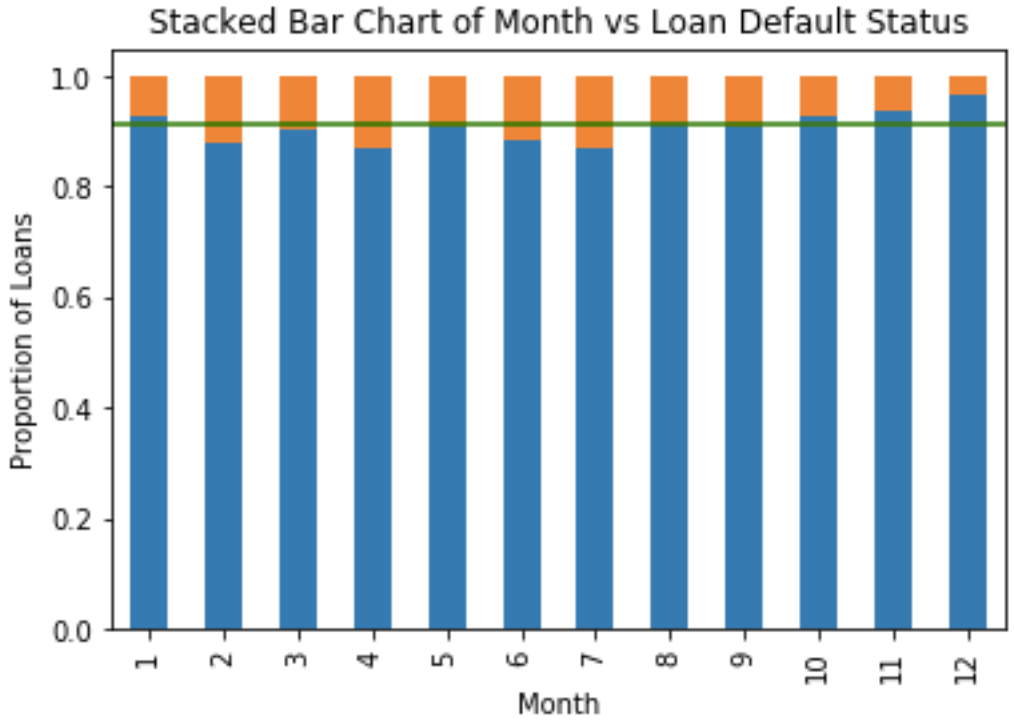
2D Demo against Loan Default Status

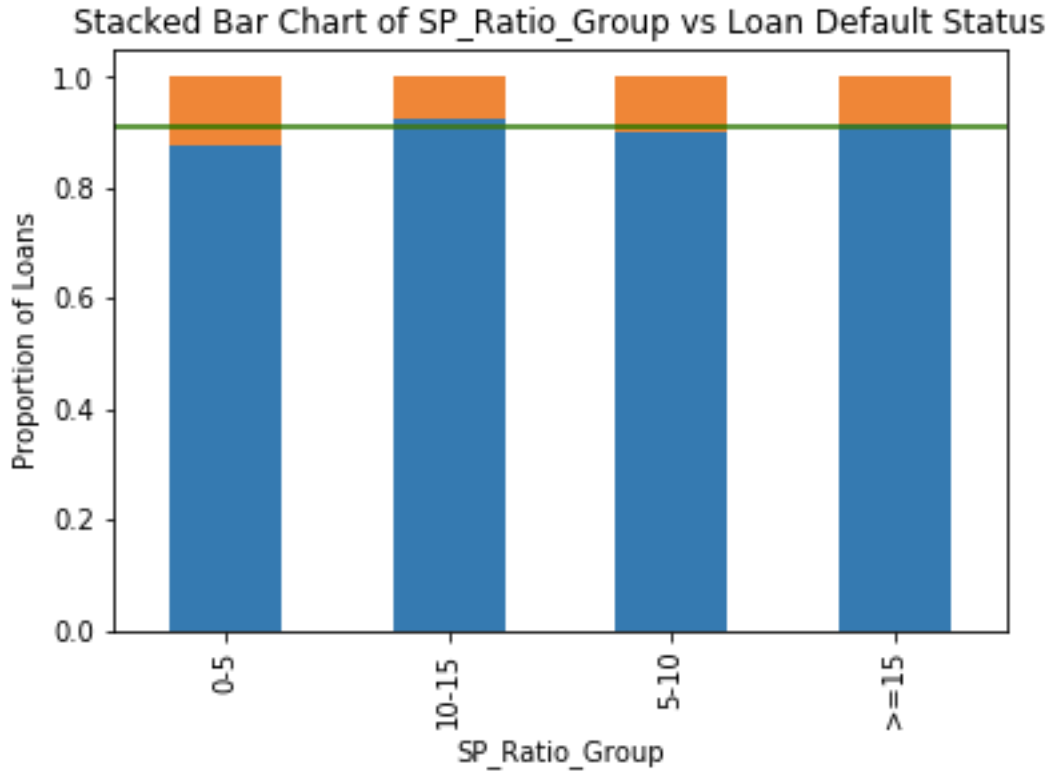




Modeling for Customer Default Probability

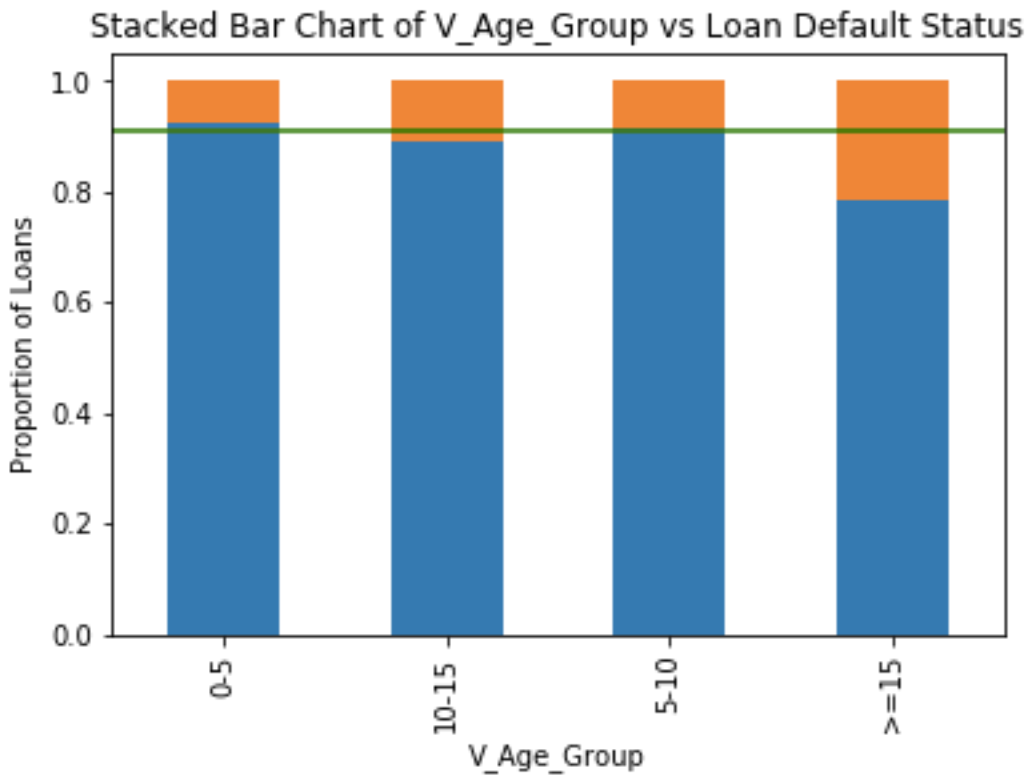
2D Demo against Loan Default Status

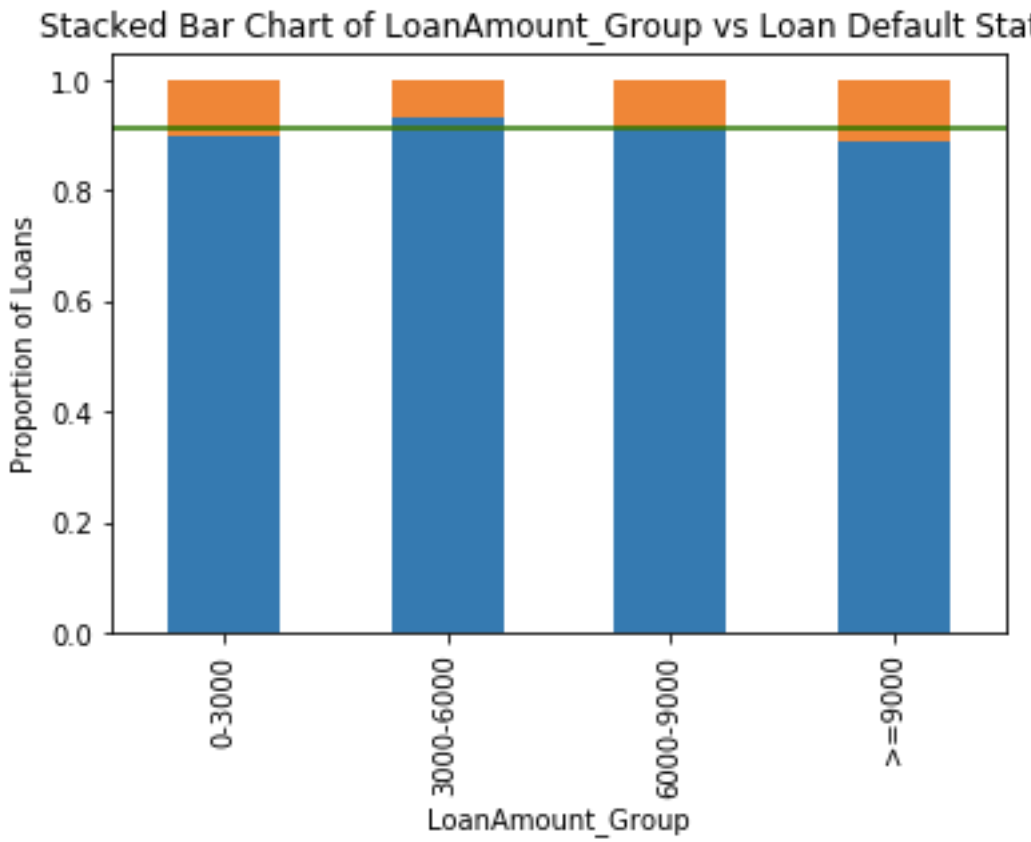




Modeling for Customer Default Probability

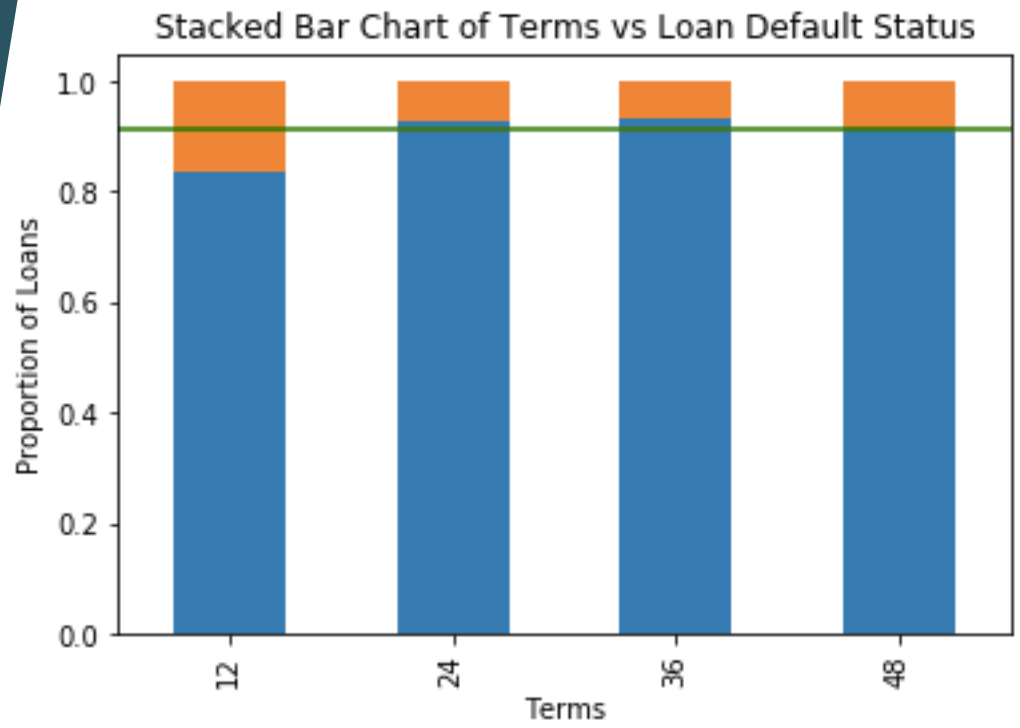
2D Demo against Loan Default Status

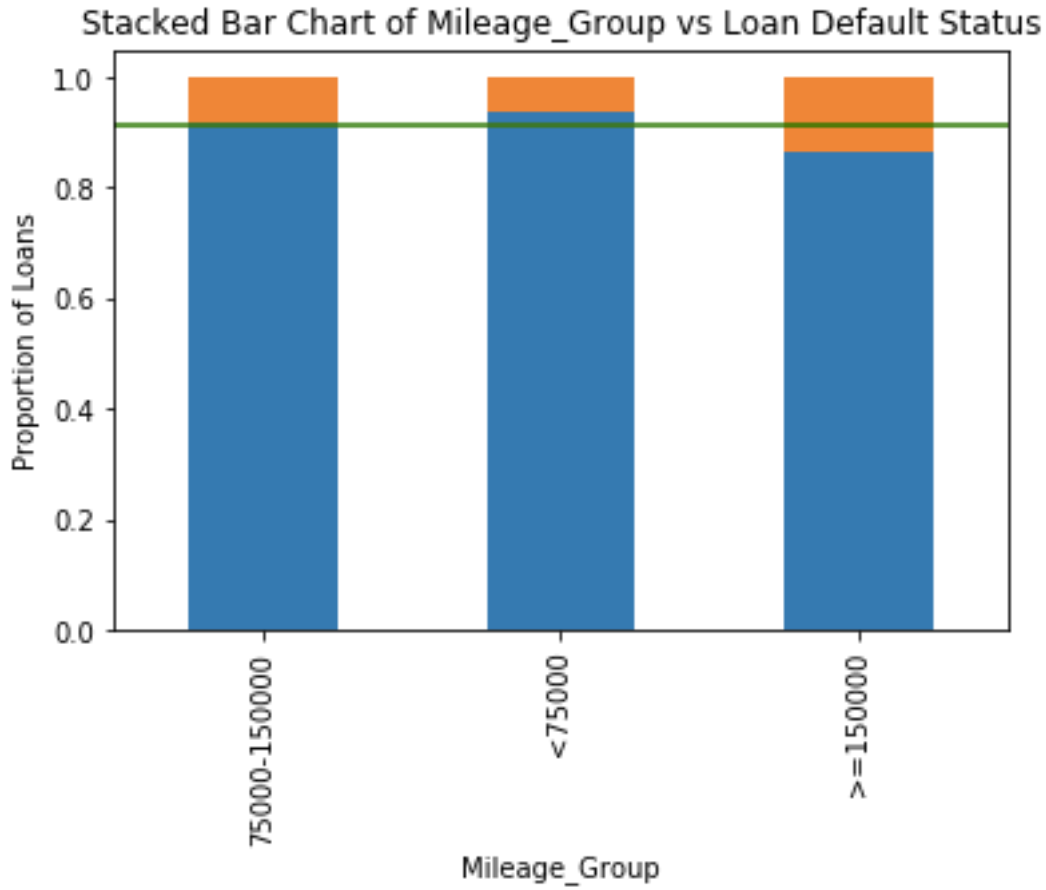




Modeling for Customer Default Probability

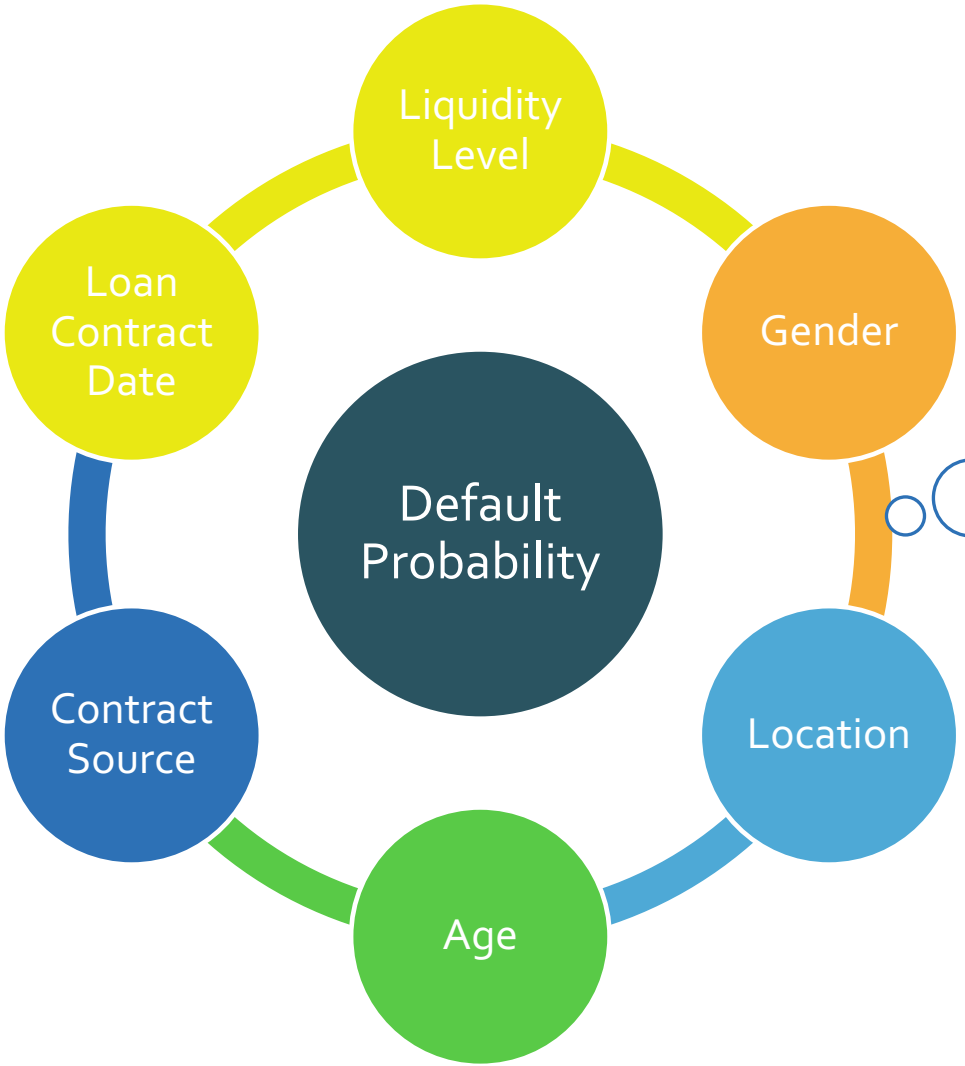
2D Demo against Loan Default Status





Modeling for Customer Default Probability

2D Demo against Loan Default Status



Risky States

AL	AZ
GA	IL
KS	MI
MN	ND
NM	OR
SC	UT
WI	WY

Modeling for Customer Default Probability

Logistic Regression Result

Optimization terminated successfully.
Current function value: 0.279154
Iterations 7

Logit Regression Results

Dep. Variable:	y	No. Observations:	3817
Model:	Logit	Df Residuals:	3800
Method:	MLE	Df Model:	16
Date:	Wed, 29 Nov 2017	Pseudo R-squ.:	0.07258
Time:	07:33:35	Log-Likelihood:	-1065.5
converged:	True	LL-Null:	-1148.9
		LLR p-value:	3.689e-27

	coef	std err	z	P>|z|	[0.025	0.975]
Month_2	0.2249	0.303	0.742	0.458	-0.369	0.819
Month_5	-0.3722	0.221	-1.681	0.093	-0.806	0.062
Month_7	0.3084	0.174	1.768	0.077	-0.033	0.650
Month_10	-0.3290	0.204	-1.616	0.106	-0.728	0.070
Month_11	-0.5913	0.221	-2.679	0.007	-1.024	-0.159
Month_12	-1.1620	0.320	-3.632	0.000	-1.789	-0.535
Terms_12	0.8146	0.140	5.819	0.000	0.540	1.089
IsExternalRefinance_False	-0.4484	0.191	-2.343	0.019	-0.823	-0.073
Gender_F	-1.5747	0.233	-6.754	0.000	-2.032	-1.118
Gender_M	-1.3533	0.236	-5.743	0.000	-1.815	-0.891
Risky_Area_False	-0.4080	0.118	-3.448	0.001	-0.640	-0.176
SP_Ratio_10-15	-0.2398	0.151	-1.589	0.112	-0.536	0.056
LoanAmount_3000-6000	-0.2306	0.138	-1.671	0.095	-0.501	0.040
LoanAmount_>=9000	0.4064	0.189	2.153	0.031	0.037	0.776
Age_18-29	0.5654	0.173	3.265	0.001	0.226	0.905
Age_40-49	-0.3303	0.146	-2.257	0.024	-0.617	-0.043
Age_50&above	-0.8139	0.169	-4.818	0.000	-1.145	-0.483

Modeling for Customer Default Probability

Logistic Regression Result

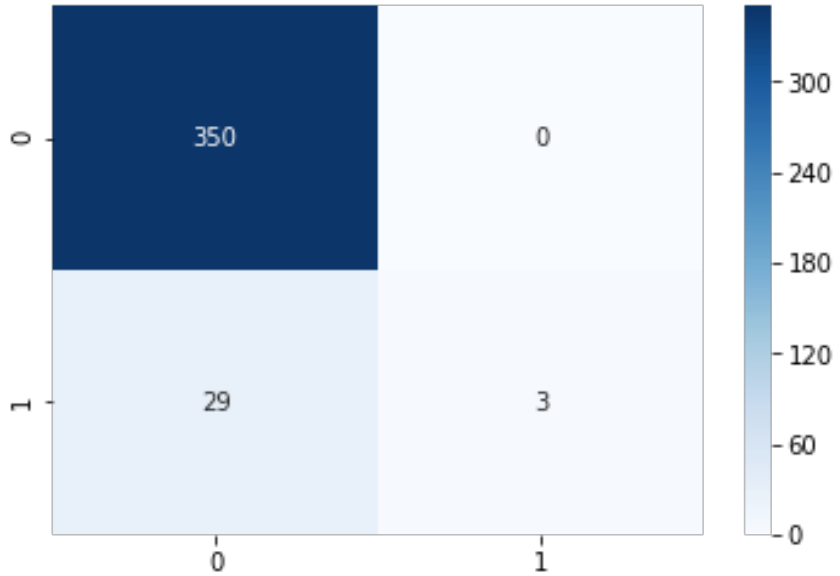
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=0)
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
logreg = LogisticRegression()
logreg.fit(X_train, y_train)
```

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
          intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
          penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
          verbose=0, warm_start=False)
```

```
y_pred = logreg.predict(X_test)
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_test)))
```

```
Accuracy of logistic regression classifier on test set: 0.92
```

Confusion Matrix:



Modeling for Customer Default Probability

Logistic Regression Model for estimating Default Probability:

$$\pi_i = \frac{\exp(\beta_0 + \beta_i X_i(1))}{1\_\exp(\beta_0 + \beta_i X_i(1))}$$

Modeling for Customer Default Probability

Example

	LoanID	Terms	IsInternalRef	IsExternalRef	Gender	State	SalaryFreque	VehicleYear	Make	ApplicationT	ApplicationD
38	7803	36	FALSE	FALSE	M	CO	1	2015	TOYOTA	24	5/12/16

MonthlyPayr	Month	Loan_Defaul	LoanAmount	Salary_Grou	SP_Ratio_Gr	V_Age_Grou	Interest	CO_APP_Def	Mileage_Gro	Risky_Area	Age_Group
495.427135	5	0	>=9000	2000-4000	5-10	0-5	0.21	None	<75000	FALSE	30-39

$$\pi = \frac{\exp(-0.3722+(-0.4484)+(-1.3533)+(-0.4080)+(-0.2398)+0.4064)}{1+exp(-0.3722+(-0.4484)+(-1.3533)+(-0.4080)+(-0.2398)+0.4064)} \approx 0.0820$$

This indicates that this client ‘s default probability would only be 8.2%.
And in fact, this contract didn’t default.

Summary

- 1D demo gives us features about main customer group, while 2D analysis tends to reveal some more interesting habits of current customers.
- Based on the accuracy score and example test, this logistic regression model built today would help with screening high quality customers.
- However, there are still a lot could be done to improve the model, i.e. , several other factors could introduced or other quant models could be tested.

TURBO
TITLE LOAN

Q&A