



Which Kiva loans expire before receiving funding?

Spring 2019
BUAN 6337: Predictive Analytics using SAS



Under the Guidance of
Prof. Shervin Shahrokhi
Tehrani

Team Members : Vinay Singh | Megan Malisani | Pragati Mishra | Vaibhav Srivastav | Ishan Jain | Erhao Liang

WHAT IS KIVA?

Kiva is a peer to peer microloan website.

Kiva's partners create loans and post them to the site.

Kiva's users then donate money to fund specific loans.

The person receiving the loan then pays the loan back over a period of months (or years).

Loans are typically in poorer areas, however there are also plenty of loans to Americans.

kiva

Lend ▾



Borrow

About ▾

Sign in

Kiva Website Layout



66% funded

Only 19 hours left!

\$250 to go

Total loan: \$750

Powered by 16 lenders

José Agustín



Barranquilla, Colombia / Services

\$25 ▾

Lend now

A loan of \$750 helps to buy paints and solvents so he can have the supplies he needs to deliver orders on time.

WHAT IF A KIVA LOAN DOESN'T GET FUNDED?

After 30 days, if the Kiva loan is not fully funded, it is removed from the site.

Kiva's partner assumes the risk on the loan, which makes them less likely to work with Kiva in the future.

Kiva NEEDS to fully fund its loans.

SUMMARY STATISTICS

Out of 1,016,535 loans, 46,541 did not get funded, or 4.6%

Philippines, Peru and Kenya borrowed the highest total amount

Average loan size was \$832

“Agriculture and Food” and “Farming and General Store” were the most popular sectors. “Wholesale” was the least.

SUMMARY STATISTICS :Qualitative Variables

Inference

- 5% of the loan come under expired loan out of total 1 million
- 63% of total loans are taken by the female
- Majority of Loans are taken in Africa and Asia

Summary Statistics of categorical variables

The FREQ Procedure

Expired	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	969993	95.42	969993	95.42
1	46541	4.58	1016534	100.00

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Group	148732	14.63	148732	14.63
female	646770	63.63	795502	78.26
male	221032	21.74	1016534	100.00

Continent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Africa	269424	26.50	269424	26.50
Americas	273137	26.87	542561	53.37
Asia	452733	44.54	995294	97.91
Europe	8554	0.84	1003848	98.75
Oceania	12686	1.25	1016534	100.00

Variable	N	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
loan_amount	1016534	832.14	1079.43	25.00	300.00	500.00	1000.00	100000.00
Month	1016534	13.00	7.52	1.00	8.00	12.00	14.00	156.00
num_lenders_total	1016534	22.22	27.81	0.00	8.00	15.00	27.00	2986.00

SUMMARY STATISTICS: Quantitative Variables

Inference

- Average Loan Amount in \$832
- Customers return their loan on average of 13 months
- 22 lender comprises of Lending Amount on average

SUMMARY STATISTICS: Expired Loan

Inference

- People expect for a higher loan when the loan is expired

The MEANS Procedure

Expired=0

Variable	N	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
loan_amount	969993	800.23	1048.42	25.00	275.00	500.00	950.00	100000.00
Month	969993	12.74	7.39	1.00	8.00	12.00	14.00	156.00
num_lenders_total	969993	22.45	28.17	1.00	8.00	15.00	27.00	2986.00

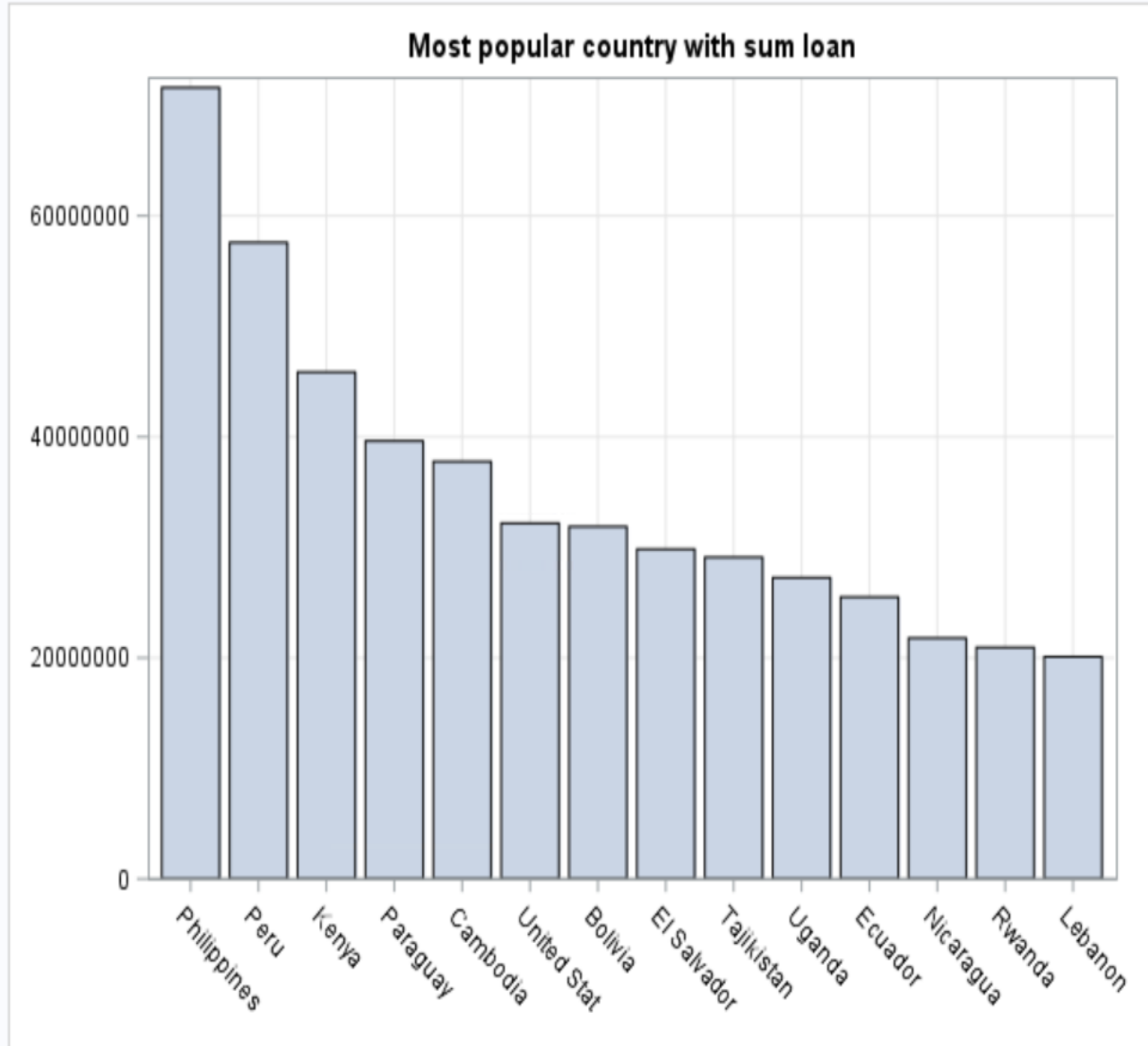
Expired=1

Variable	N	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
loan_amount	46541	1497.20	1441.18	25.00	700.00	1050.00	1650.00	50000.00
Month	46541	18.37	8.11	3.00	14.00	15.00	21.00	145.00
num_lenders_total	46541	17.40	18.16	0.00	7.00	12.00	22.00	905.00

LOAN AMOUNT BY COUNTRIES

Inference

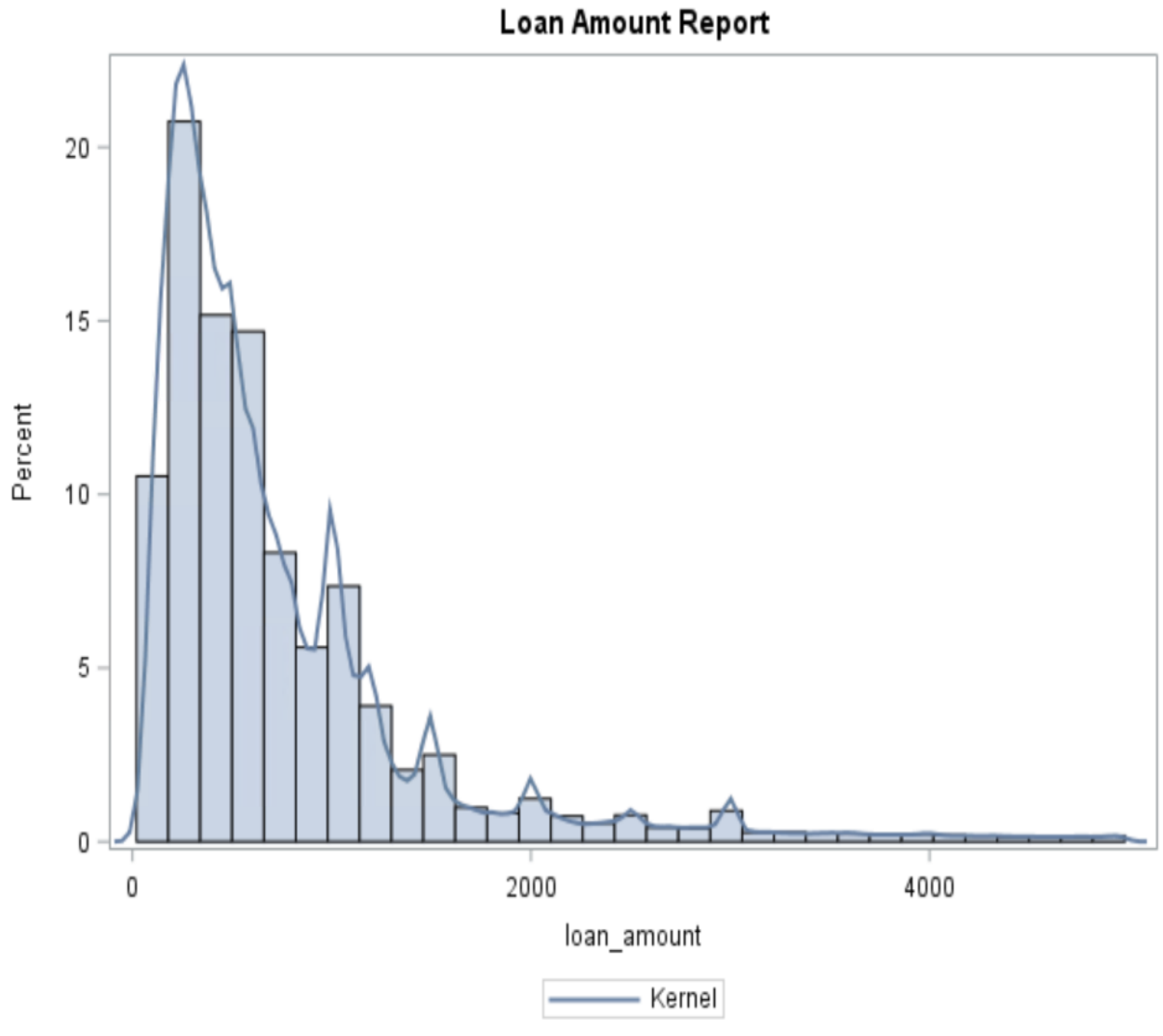
- Philippines and Peru takes the most amount of loan from Kiva
- We can infer from this that most of them are African and Asian countries



KIVA LOAN AMOUNT SIZE

Inference

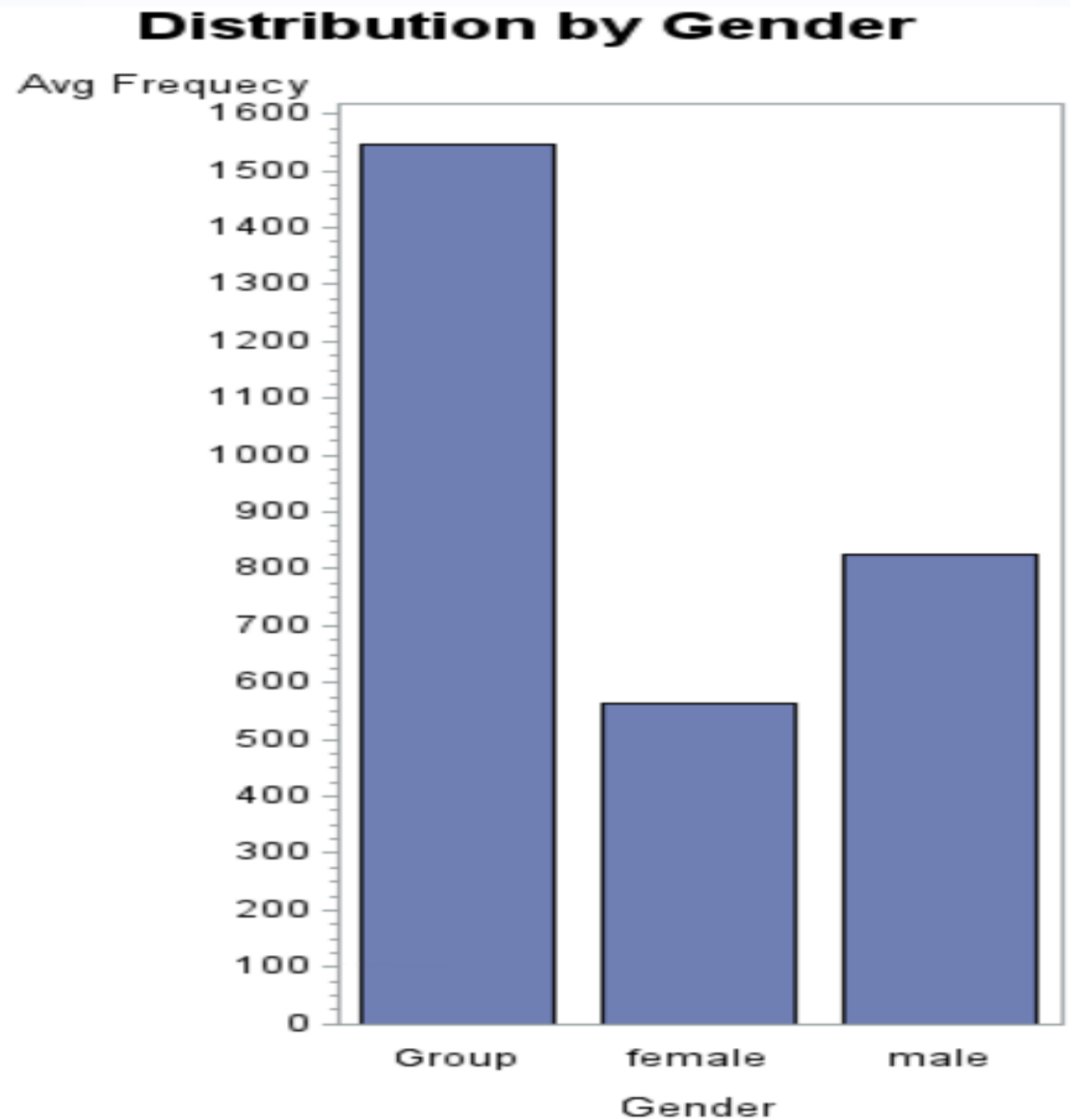
- Avg Loan size is \$832
- Majority of Population takes loan in a range of \$500 - \$1000



DISTRIBUTION OF LOAN BY GENDER

Inference

- Males ask for more money than Females on Kiva
- People forming a group are asking more money than individuals

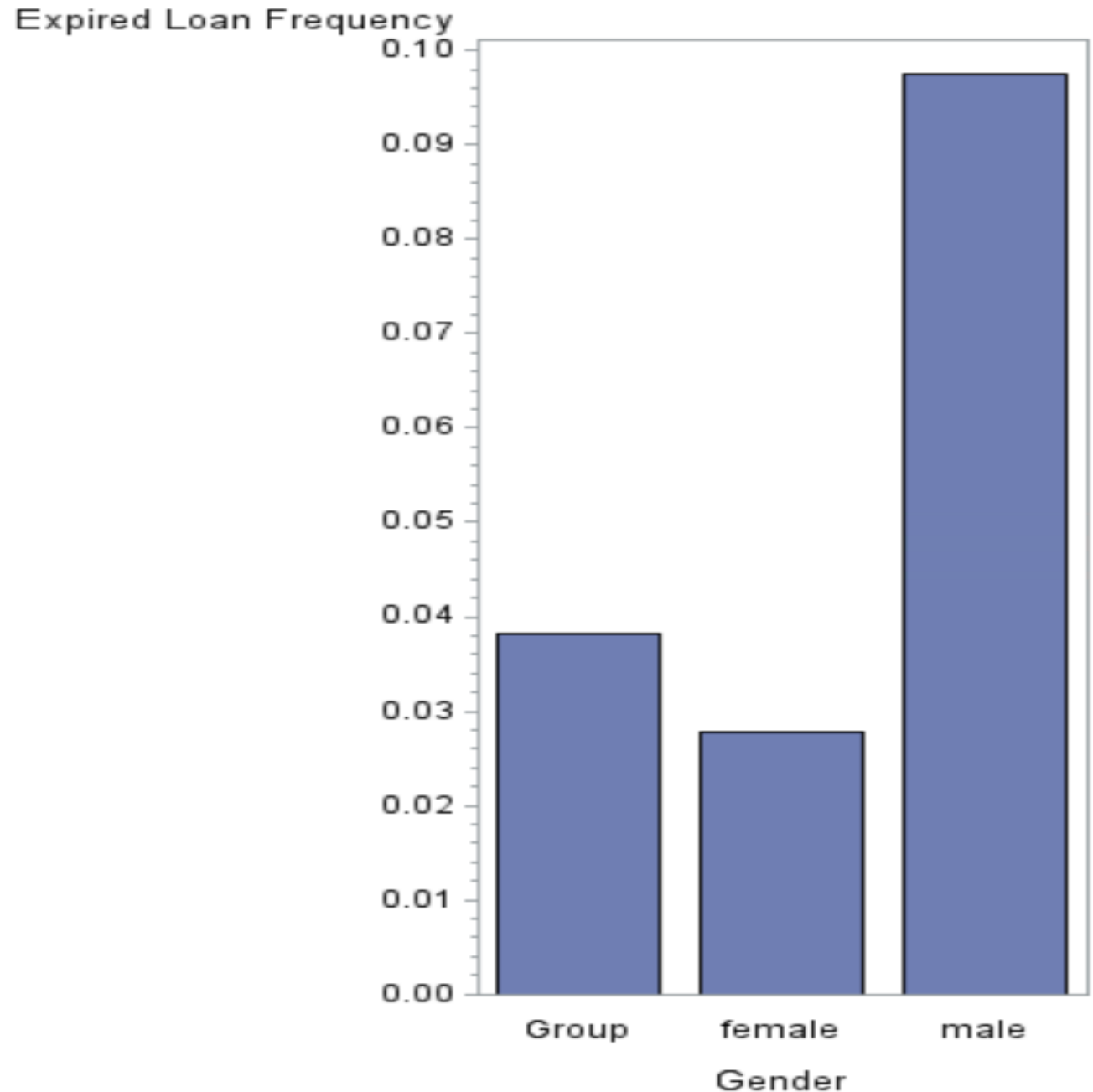


EXPIRED LOAN: DISTRIBUTION BY GENDER

Inference

- Males has the highest number Expired Loan
- Contrary to the Avg Loan Amount, Female has the lowest rate of Expired Loan
- Rate of Failure to get a Loan is 3 Times higher for Male than Female

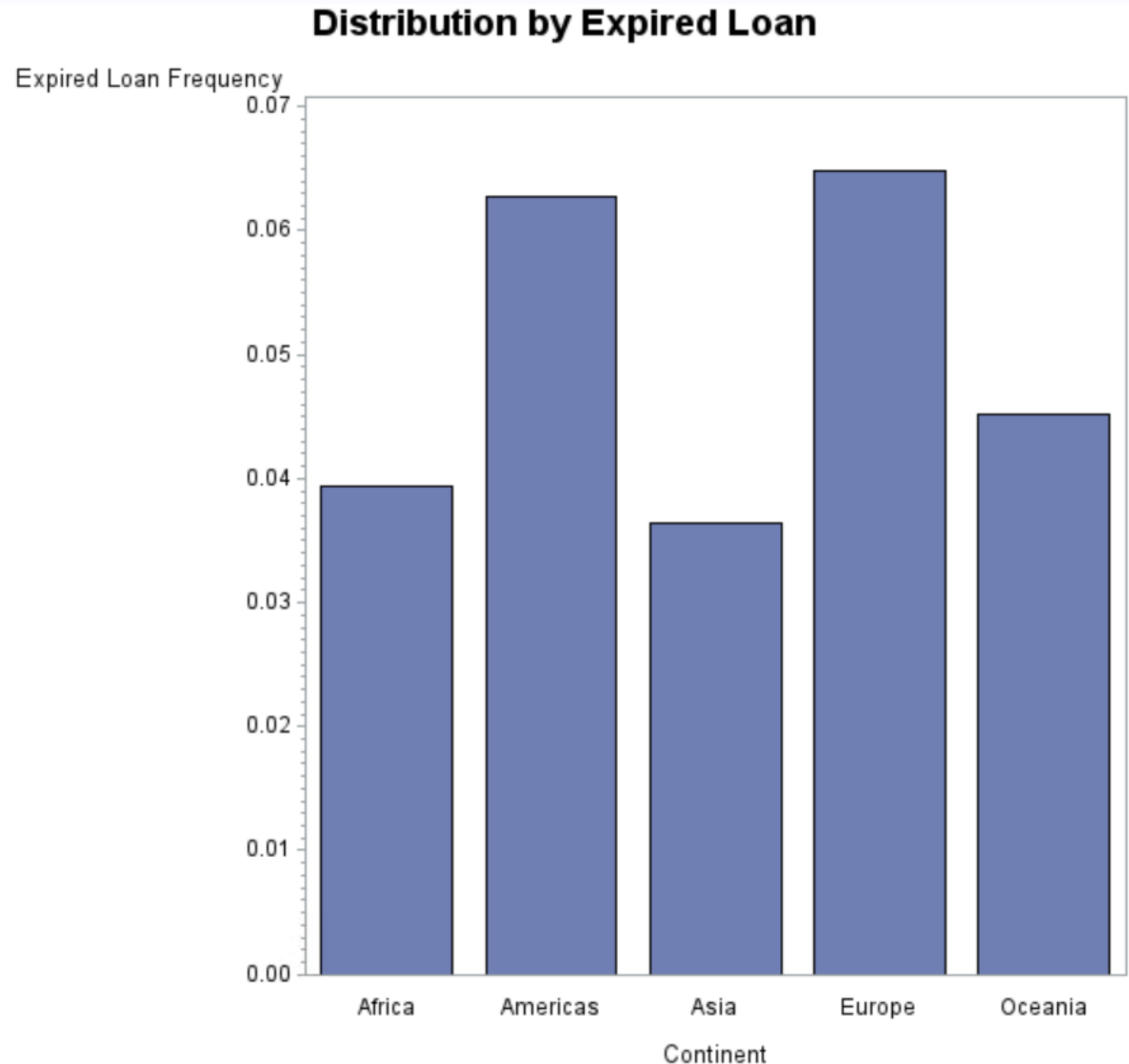
Distribution by Expired Loan




EXPIRED LOAN: DISTRIBUTION BY CONTINENT

Inference

- Africa and Asia has the lowest number of Expired Loan (contrary to the Loan Amount by Continent)
- Developed Countries people have a high chance of Loan not getting approved



MODEL BUILDING



LINEAR REGRESSION MODEL

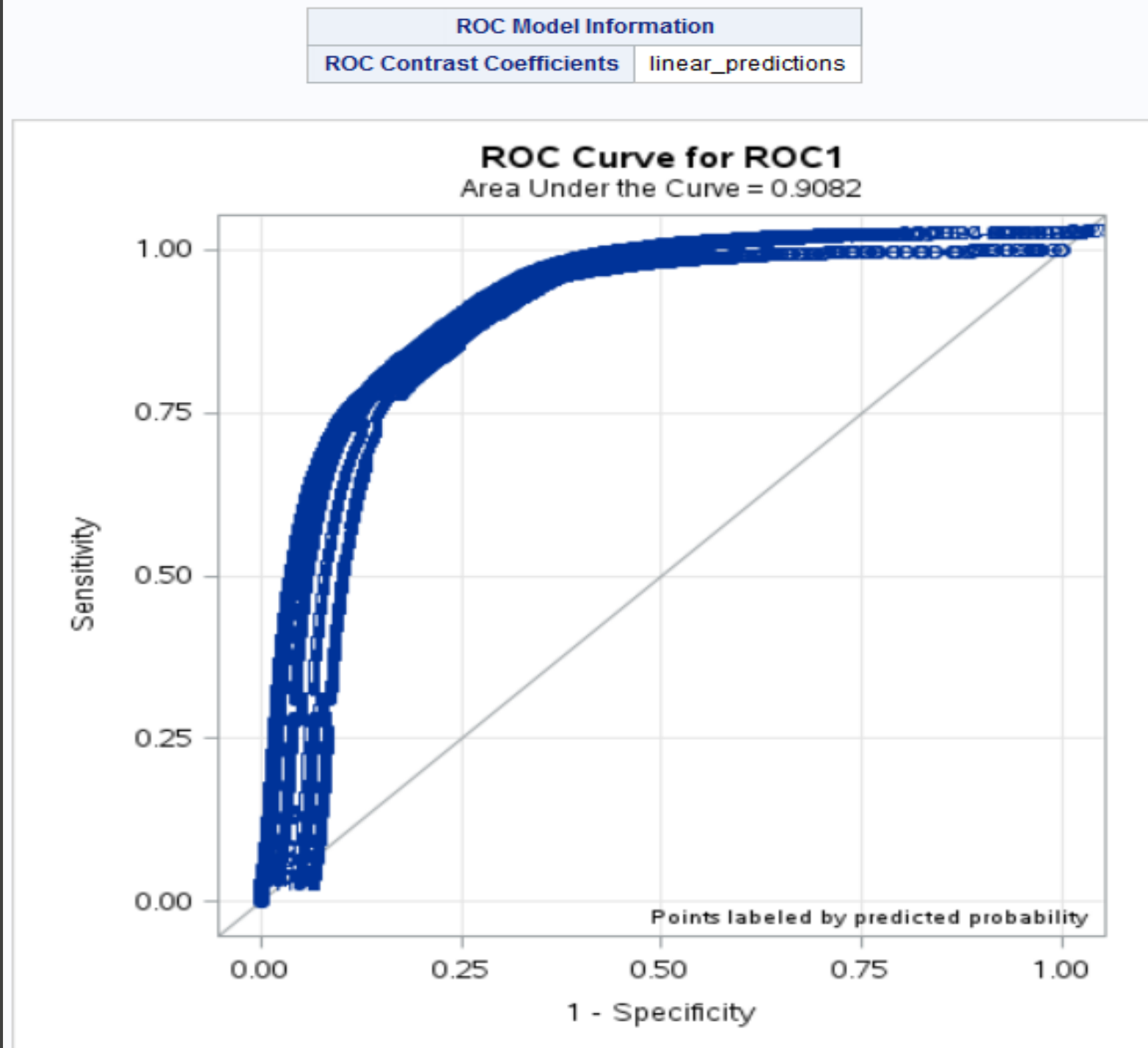
Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	0.0399372331	B	0.00078487	50.88	<.0001
loan_amount	0.0001088676		0.00000041	265.76	<.0001
Gender Group	-.0612741697	B	0.00083861	-73.07	<.0001
Gender female	-.0556619249	B	0.00058897	-94.51	<.0001
Gender male	0.0000000000	B	.	.	.
num_lenders_total	-.0041043263		0.00001545	-265.69	<.0001
Month	0.0040920735		0.00003256	125.69	<.0001
Continent Africa	0.0003863586	B	0.00064120	0.60	0.5468
Continent Asia	-.0063660416	B	0.00057857	-11.00	<.0001
Continent Europe	0.0101980707	B	0.00257308	3.96	<.0001
Continent Oceania	0.0106539864	B	0.00212652	5.01	<.0001
Continent Americas	0.0000000000	B	.	.	.

In the linear model result, it provides evidence for previous inferences: Compare to Male, Female and Group have lower possibility issued expired loan; compare to America, Asia has lower possibility issued expired loan.

Africa has not been approved by linear model

ROC CURVE: Linear Model

This is the ROC
curve of our Linear
Model
 $AUC=0.9082$





LOGISTIC REGRESSION MODEL

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.7315	0.0216	29845.6515	<.0001
loan_amount		1	0.00128	8.88E-6	20709.9108	<.0001
Gender	Group	1	-0.8194	0.0187	1924.2727	<.0001
Gender	female	1	-0.2067	0.0118	305.0749	<.0001
num_lenders_total		1	-0.0759	0.000559	18395.1891	<.0001
Month		1	0.0675	0.000604	12490.0599	<.0001
Continent	Africa	1	-0.2133	0.0195	120.0226	<.0001
Continent	Asia	1	-0.3001	0.0184	266.3412	<.0001
Continent	Europe	1	0.2693	0.0499	29.1210	<.0001
Continent	Oceania	1	0.3790	0.0444	72.8239	<.0001

In the Logistic Regression model result (TRAINING) it provides perfect evidence for previous inferences: Compare to Male, Female and Group have lower possibility issued expired loan; compare to America, Asia and **Africa** has lower possibility issued expired loan.

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.7315	0.0216	29845.6515	<.0001
loan_amount		1	0.00128	8.88E-6	20709.9108	<.0001
Gender	Group	1	-0.8194	0.0187	1924.2727	<.0001
Gender	female	1	-0.2067	0.0118	305.0749	<.0001
num_lenders_total		1	-0.0759	0.000559	18395.1891	<.0001
Month		1	0.0675	0.000604	12490.0599	<.0001
Continent	Africa	1	-0.2133	0.0195	120.0226	<.0001
Continent	Asia	1	-0.3001	0.0184	266.3412	<.0001
Continent	Europe	1	0.2693	0.0499	29.1210	<.0001
Continent	Oceania	1	0.3790	0.0444	72.8239	<.0001

This is the Logistic Regression model result by TEST data. it provides same evidences.

Probability of Factors for getting loans



Months: One more month = 7% lower probability of getting loan



Group: 56% higher probability of getting loan relative to male



Female: 19% higher probability of getting loan relative to male



Asia: 26% higher probability of getting loan relative to USA



Africa: 19% higher probability of getting loan relative to USA



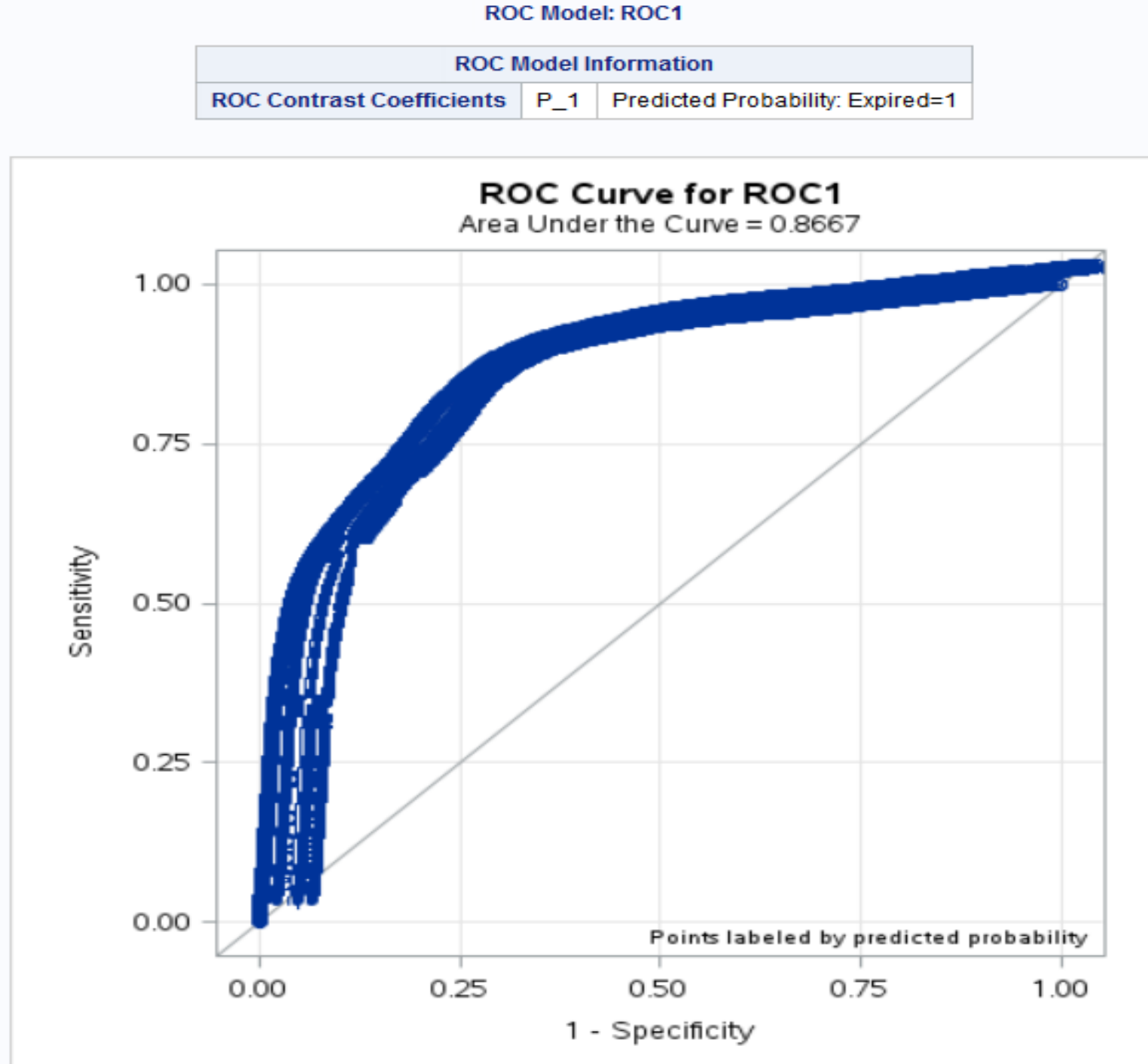
Europe: 31% lower probability of getting loan relative to USA



Oceania: 46% lower probability of getting loan relative to USA

ROC CURVE: Logistic Model

This is the ROC
curve of our
Logistic Model
 $AUC=0.8667$






PROBIT REGRESSION MODEL

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.7315	0.0216	29845.6515	<.0001
loan_amount		1	0.00128	8.88E-6	20709.9108	<.0001
Gender	Group	1	-0.8194	0.0187	1924.2727	<.0001
Gender	female	1	-0.2067	0.0118	305.0749	<.0001
num_lenders_total		1	-0.0759	0.000559	18395.1891	<.0001
Month		1	0.0675	0.000604	12490.0599	<.0001
Continent	Africa	1	-0.2133	0.0195	120.0226	<.0001
Continent	Asia	1	-0.3001	0.0184	266.3412	<.0001
Continent	Europe	1	0.2693	0.0499	29.1210	<.0001
Continent	Oceania	1	0.3790	0.0444	72.8239	<.0001

Result of Probit Regression model
Also has a same evidences for our inferences



CONCLUSIONS

- Female Person taking a Loan has a higher chance of Loan getting funded
 - If a person is from developing countries like Asia and Africa, chances of Loan getting approved is much higher
- 

REFERENCES

- <https://nycdatascience.com/blog/student-works/kiva-loans/>
- <https://www.kiva.org/borrow>
- [https://en.wikipedia.org/wiki/Kiva_\(organization\)](https://en.wikipedia.org/wiki/Kiva_(organization))
- <https://www.kiva.org/blog/what-factors-affect-loan-funding-times>

ANY QUESTIONS?

APPENDIX

LOAN DISTRIBUTION BY COUNTRIES

Countries	Count of loan_id	Sum of loan_amount
Philippines	208599	71607925
Kenya	98721	45862550
Peru	61977	57584850
Cambodia	58431	37766575
El Salvador	45675	29837000
Uganda	32593	27256950
Pakistan	32493	17438425
Tajikistan	32428	29101700
Nicaragua	30541	21784575
Ecuador	23992	25504900
Colombia	23146	16339725
Bolivia	18271	31876225
Paraguay	18117	39623450
Vietnam	16193	18234200
Ghana	15068	10608250
Lebanon	14711	20099225
Nigeria	14333	5481575
Mexico	13809	17785525
Samoa	12047	8981025

LOAN DISTRIBUTION BY SECTOR

Sector	Count of loan_id	Sum of loan_amount
Agriculture	247965	197953975
Food	230545	190334475
Retail	205789	164061775
Services	73407	73726825
Clothing	58058	60163025
Housing	43208	31783125
Personal Use	36724	16606775
Education	32992	32620975
Transportation	28119	21017125
Arts	20217	18801425
Construction	13374	12712675
Health	11874	12474775
Manufacturing	11312	9712475
Entertainment	1478	1948425
Wholesale	1472	1977150

Calculation of Probability of getting loan

	Estimates	Probability of expire loan	getting loan
Loan amount	0.00128	1.001	-0.001
group	-0.8194	0.441	0.56
female	-0.2067	0.813	0.19
num lender	-0.0759	0.927	0.07
month	0.0675	1.070	-0.07
africa	-0.2133	0.808	0.19
asia	-0.3001	0.741	0.26
europe	0.2693	1.309	-0.31
oceania	0.379	1.461	-0.46