ANALYSIS OF RATE SPREAD FOR HOME LOANS

By Jorge LeonFrausto

DATA SOURCE

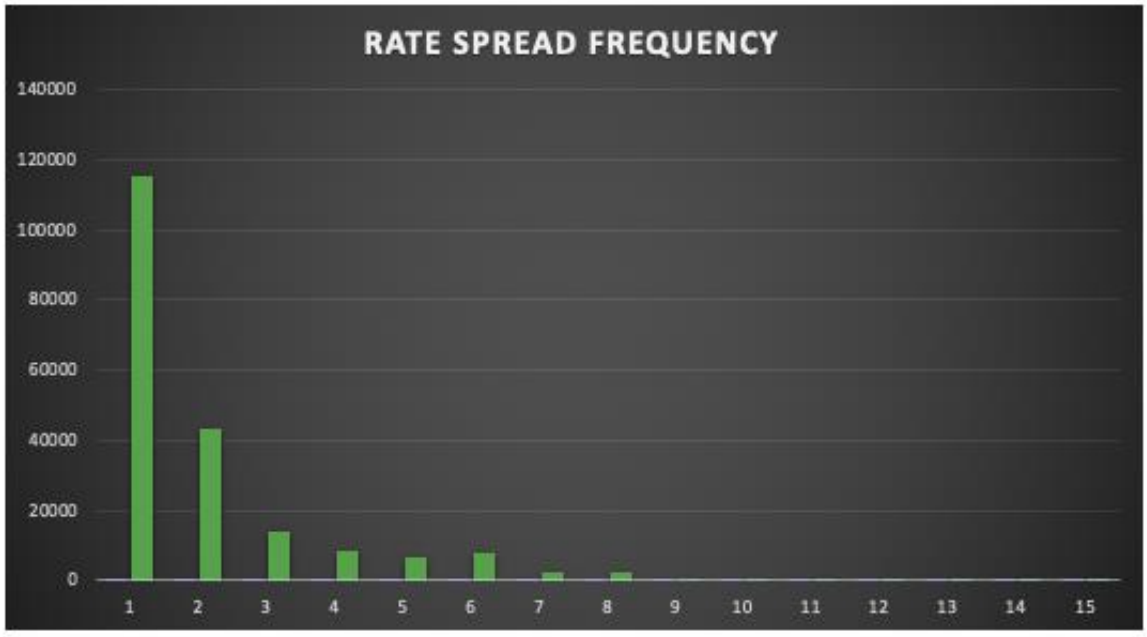
The document represents an analysis of data concerning home loan applications and the associated loan rate spread. The analysis is based on 200,000 observations, each containing specific variables and characteristics associated with a home loan application as well as its associated loan rate spread. Data were obtained from 2015 Home Mortgage Disclosure Act (HMDA) Loan / Application Register.

EXECUTIVE SUMMARY OF FINDINGS

- More than half the points have a rate spread of 1.
- Applicant income is inversely related to average spread rate.
- Loan amount is inversely related to average spread rate.
- Average spread rate is 1.9774
- Within categorical relationships, Loan Type 1 and Property Type 2 combination has the highest average rate spread observed at 4.4270.
- The lowest cross-categorical variable combination is Loan type 4 and Property type 2 at 1.0417.
- Generally speaking, there are four clusters for average rate spread when taking into consideration loan amount and income relationships.

STATISTICAL DISTRIBUTIONS BY VARIABLE

The primary variable of interest is the loan rate spread. It is a right-tailed distribution. The value 1 is the most common with more than 55% of entries. The average is 1.9774 for all values.



CATEGORICAL AVERAGES

LOAN TYPE	Average Rate Spread	LOAN PURPOSE	Average Rate Spread
1	2.7329	1	1.9145
2	1.3542	2	3.2723
3	1.3742	3	1.8504
4	1.1213		

PROPERTY TYPE	Average Rate Spread	CO-APPLICANT	Average Rate Spread
1	1.6127	FALSE	1.9208
2	3.9912	TRUE	2.0684
3	3.2899		

CATEGORICAL VARIABLES EXPLANATIONS

Loan Type: 1) Conventional; 2) FHA-Insured; 3) VA-guaranteed; 4) Farm Service/Rural Housing
Loan Purpose: 1) Home purchase; 2) Home improvement; 3) Refinancing
Property Type: 1) One-to-four family; 2) Manufactured housing; 3) multi-family
Co-Applicant: False) No co-applicant listed; True) Co-applicant listed.

CATEGORIAL RELATIONSHIPS

For this initial analysis, we will review how the four categorial variables affect the observed average rate spread. The effort is to analyze the general trends, and see if any particular combinations deviate from these trends.

		LOAN TYPE				
		1	2	3	4	TOTAL
PROPERTY TYPE	1	2.0578	1.3403	1.3964	1.1223	**1.6127**
	2	4.4270	1.6582	1.1398	1.0417	**3.9912**
	3	3.3493	1.7778			**3.2899**
LOAN PURPOSE	1	2.9379	1.2570	1.5987	1.1208	**1.9145**
	2	3.0385	4.1958	1.0476		**3.2723**
	3	2.1451	1.4586	1.0994	1.1471	**1.8504**
CO-APPLICANT	FALSE	2.7905	1.3313	1.4679	1.1120	**1.9208**
	TRUE	2.6623	1.4008	1.2612	1.1546	**2.0684**
		2.7329	**1.3542**	**1.3743**	**1.1213**	**1.9774**

The two outlier combinations are Property Type 2 with Loan Type 1; and, Loan Purpose 2 with Loan Type 2. This is consistent with the overall average rate for both Property Type 2 and Loan Purpose 2 which are above 3.

		PROPERTY TYPE			
		1	2	3	TOTAL
LOAN TYPE	1	2.0578	4.4270	3.3493	**2.7329**
	2	1.3403	1.6582	1.7778	**1.3542**
	3	1.3964	1.1398		**1.3743**
	4	1.1223	1.0417		**1.1213**
LOAN PURPOSE	1	1.4100	4.2355	2.8442	**1.9145**
	2	3.2991	2.9264	1.9231	**3.2723**
	3	1.7833	2.4849	3.6419	**1.8504**
CO-APPLICANT	FALSE	1.5607	4.0027	3.3911	**1.9208**
	TRUE	1.6978	3.9746	2.7222	**2.0684**
		1.6127	**3.9912**	**3.2899**	**1.9774**

The remainder three categories compared against Property Type do not show a deviation from what their overall average rates reflect. Property Type 2 consistently has the highest average rate spread. And, Loan Types 2, 3, and 4 carry a lower rate spread average than the average rate spread for the population (1.9774).

		LOAN PURPOSE			
		1	2	3	TOTAL
LOAN TYPE	1	2.9379	3.0385	2.1451	**2.7329**
	2	1.2570	4.1958	1.4586	**1.3542**
	3	1.5987	1.0476	1.0994	**1.3743**
	4	1.1208		1.1471	**1.1213**

		1	2	3	TOTAL
PROPERTY TYPE	1	1.4100	3.2991	1.7833	**1.6127**
	2	4.2355	2.9264	2.4849	**3.9912**
	3	2.8442	1.9231	3.6419	**3.2899**

CO-APPLICANT	FALSE	1.8620	3.3076	1.8180	**1.9208**
	TRUE	2.0079	3.2356	1.8925	**2.0684**
		1.9145	**3.2723**	**1.8504**	**1.9774**

Loan Purpose broken down by association shows that Loan Type 2 and Loan Purpose 2, along with Property Type 1 and Loan Purpose 1 carry the highest average rate spreads. Overall, the loan type association table shows lower than global average rate spreads. The Property Type table shows mostly higher than global average rate spreads, with the exception of Property Type 1 crossed with Loan Purpose 1 and 3. The co-applicant relationship seems to affect little the average rate spread compared to Loan Purpose averages by category with no cross association.

		CO APPLICANT		
		FALSE	TRUE	TOTAL
LOAN TYPE	1	2.7905	2.6623	**2.7329**
	2	1.3313	1.4008	**1.3542**
	3	1.4679	1.2612	**1.3743**
	4	1.1120	1.1546	**1.1213**

		CO APPLICANT		
		FALSE	TRUE	TOTAL
PROPERTY TYPE	1	1.5607	1.6978	**1.6127**
	2	4.0027	3.9746	**3.9912**
	3	3.3911	2.7222	**3.2899**
LOAN PURPOSE	1	1.8620	2.0079	**1.9145**
	2	3.3076	3.2356	**3.2723**
	3	1.8180	1.8925	**1.8504**
		1.9208	**2.0684**	**1.9774**

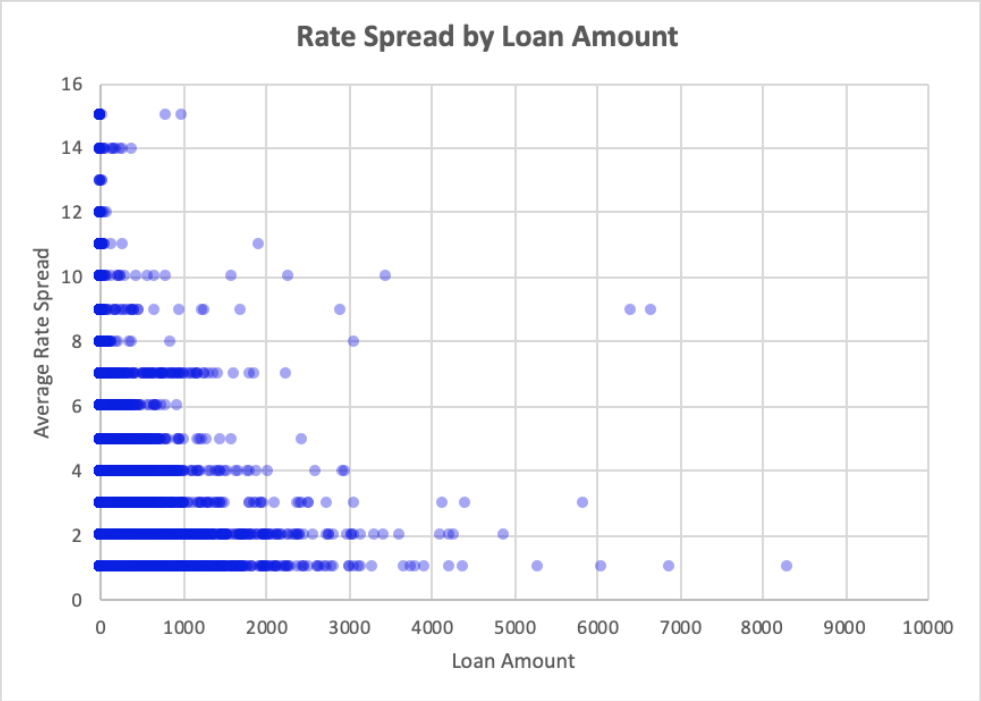
Co-Applicant over all does not seem to affect the average rate spread much. The highest deviations from the other categories averages are for Property Type 3 associated with value FALSE for Co-Applicant around 3% higher; and, for TRUE co-applicant with Property Type 1 with a 5% higher than rate average value.

NUMERICAL FEATURES ANALYSIS

With the completion of the categorical variables, we now look at the numerical values present in the data.

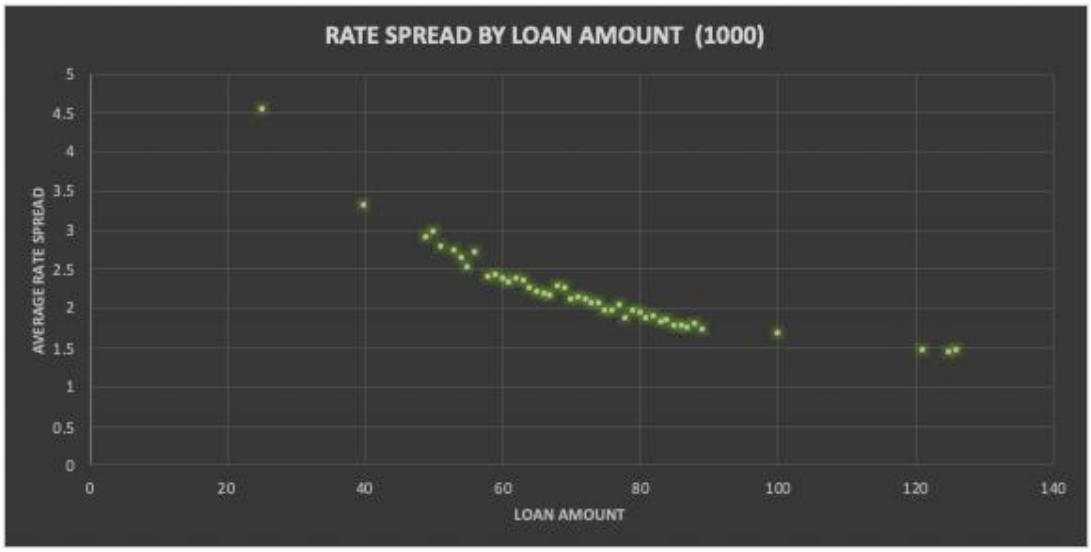
Relationship between Loan Amount and Rate Spread

In the next steps, reviewed is the relationship between the loan amount and the rate spread in the data.

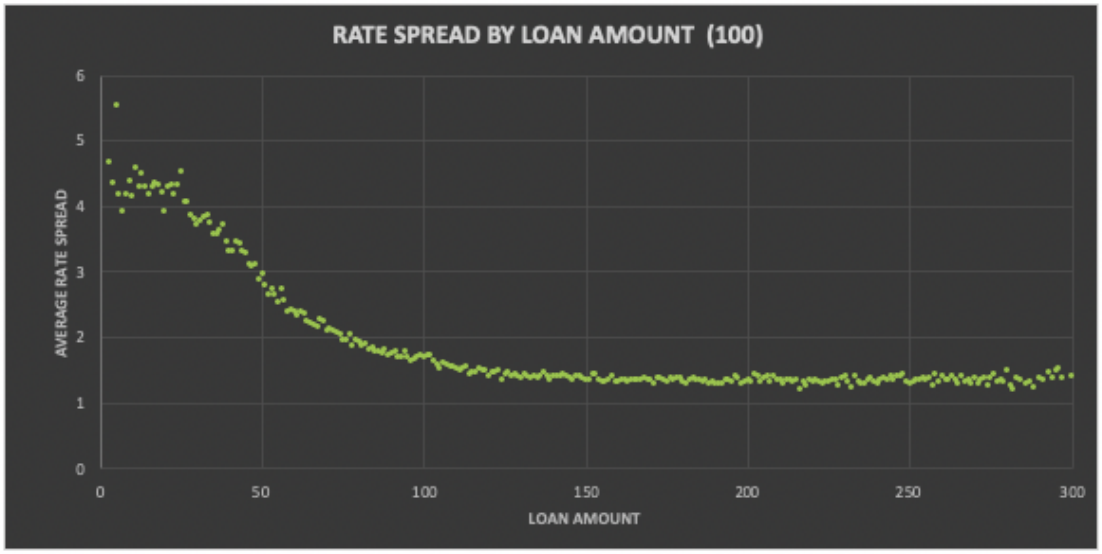


This graph represents the loan amounts and rate spread reported. The plotted points are mostly clustered under 3000 loan amount and rate spread under 7. The lighter points reflect instances with fewer observations in the data.

Below, is the graph representing the data reported with the average rate spread. There is a distinct inverse relationship between loan amount and average reported rate spread. From a business perspective, it makes sense given the underlying circumstances. We must take into consideration that these are approved loans. A high amount loan, even when paired with a lower interest rate will yield more profit than a lower amount loan paired with a high(er) interest. As we will see on most analysis throughout this report, analyzing the averages will yield relatively compact graphs – but the individual points do not hold a high correlation coefficient. In particular, this relationship has a week R value of -0.2179.

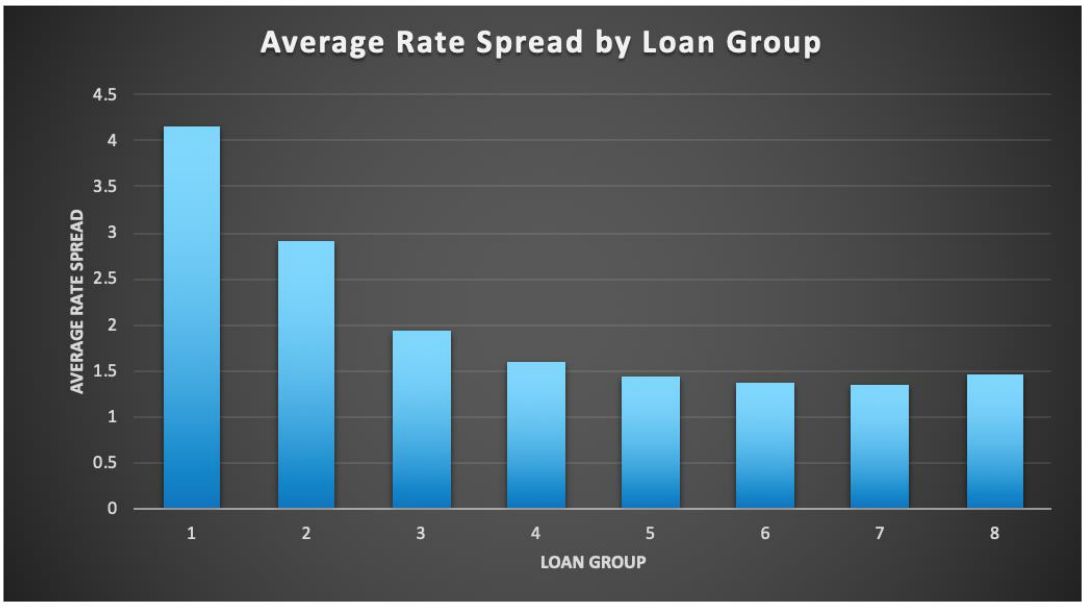


Graph with only loan amounts with at least 1000 entries.



Graph with all loan amounts with at least 100 entries.

Although, there is a generally closely packed line, the pattern breaks on either end of the graph. Very low loan amounts have a higher variance, as do higher-end loans. The second graph shows that the average rate spread begins to have a higher variance after the $250,000 loan amount.

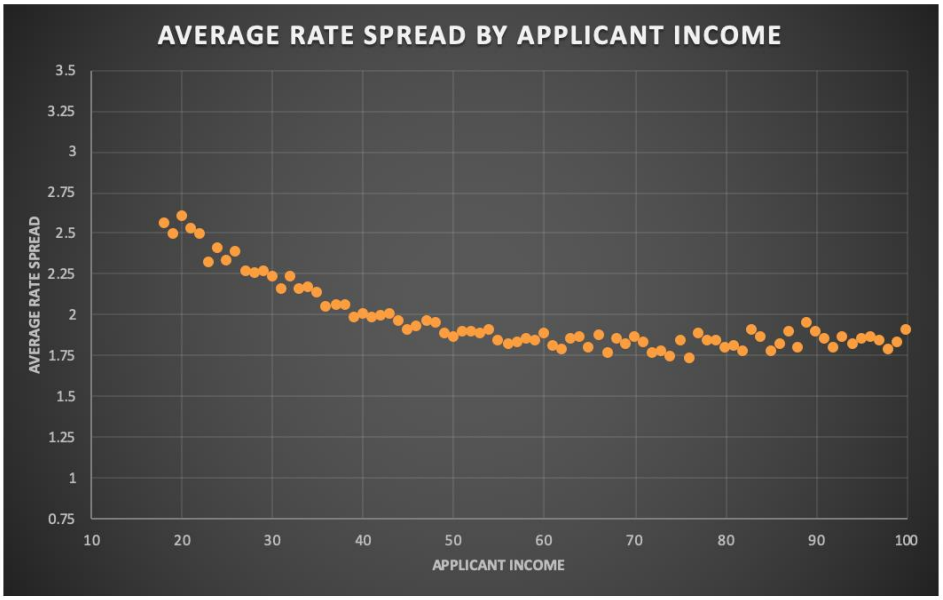


In anticipation of prediction modeling, the loans were classified as groups.

LOAN GROUP	DESCRIPTION	AVERAGE SPREAD RATE	OBSERVATION COUNT
1	Loan amounts 1 to 32	4.1471	16991
2	Amounts 33 to 65	2.9035	31900
3	Amounts 66 to 95.	1.9332	32237
4	Amounts 96 to 115	1.5990	18318
5	Amounts 116 to 145	1.4298	26750
6	Amounts 146 to 180	1.3612	24551
7	Amounts 181 to 230	1.3377	20918
8	Amounts 231 or higher	1.4519	28335

The trend continues to reflect the data. Note that there is a rise in Loan Group 8 having a higher average rate spread than group 7. In the data when all points are plotted, the highest loan amounts begin to separate and have a higher spread.

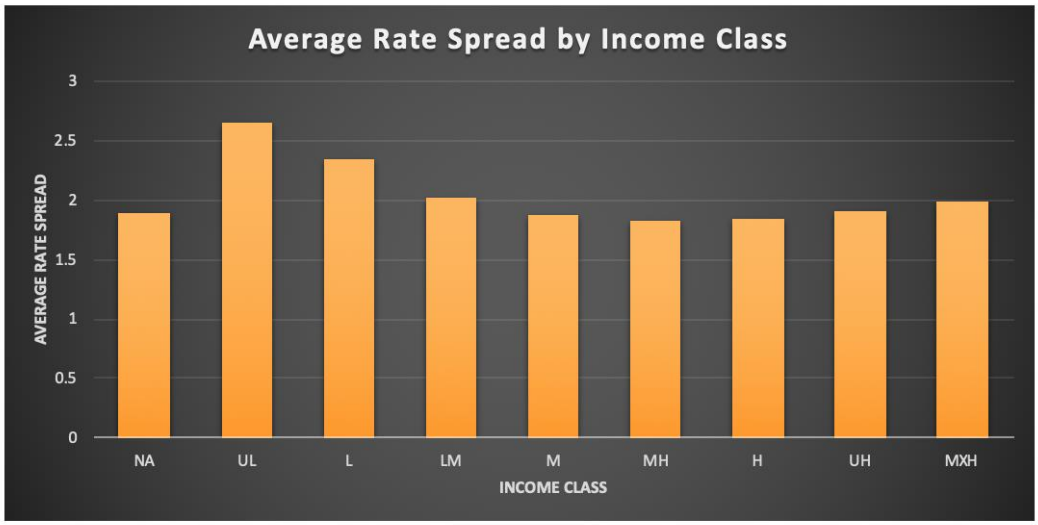
Income and Average Rate Spread



A similar trend to the loan amount – rate spread relationship appears when income is analyzed. Note that the visual line concaves and begins to draw upward for values higher than 70.

For categorical analysis, applicant income was grouped in a similar fashion to loan amount. The breakdown is below

INCOME GROUPS	RANGE	avg rate	OBSERVATIONS
NA (“Not Reported”)	non-listed	1.8423	11275
UL (“Ultra Low”)	1 to 15	2.3441	26207
L (“Low”)	16 to 32	2.0165	48977
LM (“Low Mid”)	33 to 49	1.8702	25640
M (“Mid”)	50 to 59	1.8296	35873
MH (“Mid High”)	60 to 79	1.9953	11607
H (“High”)	80 to 89	1.8990	10708
UH (“Ultra High”)	90 to 149	1.8997	28395
MXH (Maximum High”)	150 or more	2.6517	1318



When graphed, the pattern re-appears. Note that group "NA"

Rate spread by Income as affected by categorical variables

In the following sections, we will analyze how each categorical variable, in relation to the income classes created, affects the observed average rate spread.

LOAN GROUP	NA	UL	L	LM	M	MH	H	UH	MXH	Grand Total
1	2.9305	3.3434	4.1696	4.3039	4.2794	4.1475	4.2503	4.2214	4.1958	**4.1471**
2	2.2229	2.1714	2.8273	3.0900	2.9764	2.9501	2.8210	2.8183	2.8500	**2.9035**
3	1.5482	1.7881	1.6181	1.9004	2.1219	2.2514	2.2127	2.1603	2.4058	**1.9332**
4	1.5667	2.2632	1.3427	1.4695	1.6394	1.7596	1.8165	1.9716	2.0884	**1.5990**
5	1.5282	2.2121	1.2699	1.3189	1.4033	1.4677	1.5971	1.6622	2.0141	**1.4298**
6	1.6291	1.7727	1.3242	1.2447	1.2742	1.3359	1.4016	1.5417	1.8685	**1.3612**
7	1.7552	2.0588	1.8738	1.1980	1.2280	1.2604	1.3058	1.4163	1.8407	**1.3377**
8	2.4683	1.8286	2.3271	1.4943	1.2096	1.2179	1.2753	1.3508	1.6009	**1.4519**
Grand Total	**1.8990**	**2.6517**	**2.3441**	**2.0165**	**1.8702**	**1.8296**	**1.8423**	**1.8997**	**1.9953**	**1.9774**

When compared against the LOAN_GROUP categorical variable, which is dependent on the loan amount reported, we expected a similar trend where the higher the income class and higher loans will drive the average rate spread down. Thus, the lowest expected match is LOAN GROUP 8 with Income class MXH – reported at 1.6009. The highest would be the class UL with Loan Group 1 – reported at 2.9305.

However, the concentration of highest average rate is associated with Loan Group 1, and all income groups, except UL. The highest rate spread average is for income class L with Loan Group 1 combination.

For all income groups except UL and L, the average rate decreases as we move up in the Loan Group classification. Consistent with the analysis of loan amount and rate spread relationship.

LOAN PURPOSE	NA	UL	L	LM	M	MH	H	UH	MXH	Grand Total
1	3.1594	2.8022	2.3737	2.0022	1.8106	1.7174	1.6676	1.6616	1.7816	**1.9145**
2	2.2717	3.0751	2.7724	2.8837	3.1735	3.3364	3.5666	3.6477	3.3594	**3.2723**
3	1.6881	2.3854	2.0928	1.8792	1.8276	1.8318	1.8225	1.8374	1.9427	**1.8504**
Grand Total	**1.8990**	**2.6517**	**2.3441**	**2.0165**	**1.8702**	**1.8296**	**1.8423**	**1.8997**	**1.9953**	**1.9774**

In relation to Loan Purpose type, we can see that loan purpose type 2 carries the highest average rate, consistent with its unbroken distribution. Only for income classes UL and L does loan purpose type 1 carry a higher rate spread average than loan purpose type 3.

LOAN TYPE	NA	UL	L	LM	M	MH	H	UH	MXH	Grand Total
1	3.2235	2.9229	3.4704	3.1385	2.7846	2.5742	2.3789	2.2369	2.0876	**2.7329**
2	1.5748	1.7348	1.3454	1.3043	1.2988	1.3073	1.3580	1.4346	1.5692	**1.3542**
3	1.3415	1.0000	1.4677	1.4098	1.3248	1.4521	1.4024	1.2922	1.2308	**1.3743**
4	1.0000	1.0909	1.1046	1.1289	1.1297	1.1271	1.0000	1.2000	2.0000	**1.1213**
Grand Total	**1.8990**	**2.6517**	**2.3441**	**2.0165**	**1.8702**	**1.8296**	**1.8423**	**1.8997**	**1.9953**	**1.9774**

Broken by loan type, except for income class MXH, all others have the lowest average rate spread in the group with loan type 4. For income class MXH, loan type 4 is the second highest – compared to the lowest over all average rate spread among loan types. Consistently across all income classes is loan type 1 having the highest average rate spread.

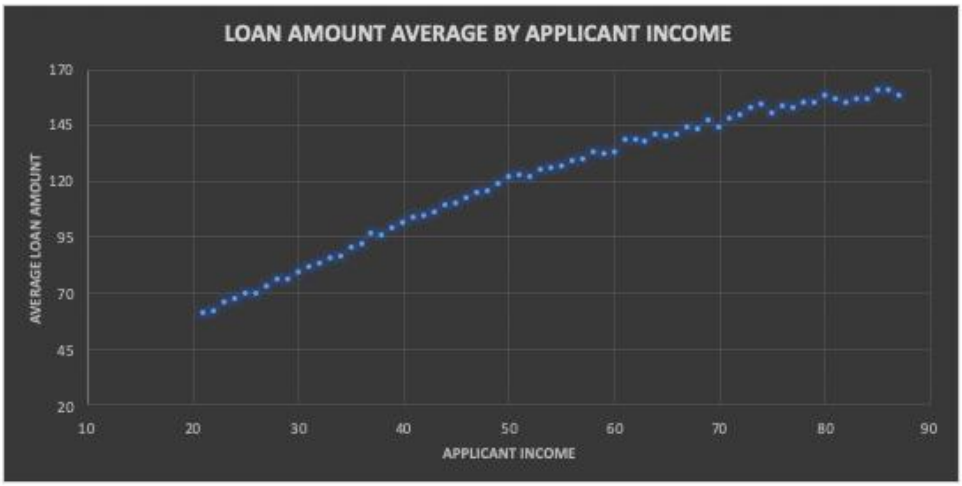
PROPERTY TYPE	NA	UL	L	LM	M	MH	H	UH	MXH	Grand Total
1	1.7945	2.1851	1.5830	1.4847	1.4948	1.5467	1.6308	1.7655	1.9494	**1.6127**
2	2.9246	4.1818	4.5094	4.2018	3.8960	3.6690	3.5218	3.2705	2.9566	**3.9912**
3	3.2899									**3.2899**
Grand Total	**1.8990**	**2.6517**	**2.3441**	**2.0165**	**1.8702**	**1.8296**	**1.8423**	**1.8997**	**1.9953**	**1.9774**

It is worth noting that Property type 3 is only listed with entries with no applicant income reported. Property type 2 has a higher average rate spread across all income classes than property type 1.

CO-APPLICANT	NA	UL	L	LM	M	MH	H	UH	MXH	Grand Total
FALSE	1.9804	2.6687	2.2153	1.8841	1.7445	1.7527	1.7980	1.9094	2.0265	**1.9208**
TRUE	1.7619	2.5538	3.0663	2.4449	2.1056	1.9172	1.8767	1.8939	1.9760	**2.0684**
Grand Total	**1.8990**	**2.6517**	**2.3441**	**2.0165**	**1.8702**	**1.8296**	**1.8423**	**1.8997**	**1.9953**	**1.9774**

Prior to analyzing the breakdown, it was assumed that having a co-applicant would lead to a lower spread rate. It was expected that it would directly affect the applicant income which is inversely related to rate spread. However, looking at the data, the not having a co-applicant leads to a lower average spread rate for all classes. The classes where the assumption holds true are for those in income classes UL and UH.

Average Loan Amount by Income



While there is a strong correlation for averages, 0.9854, the same does not hold true for actual values individually analyzed with a correlation coefficient of only 0.4463.

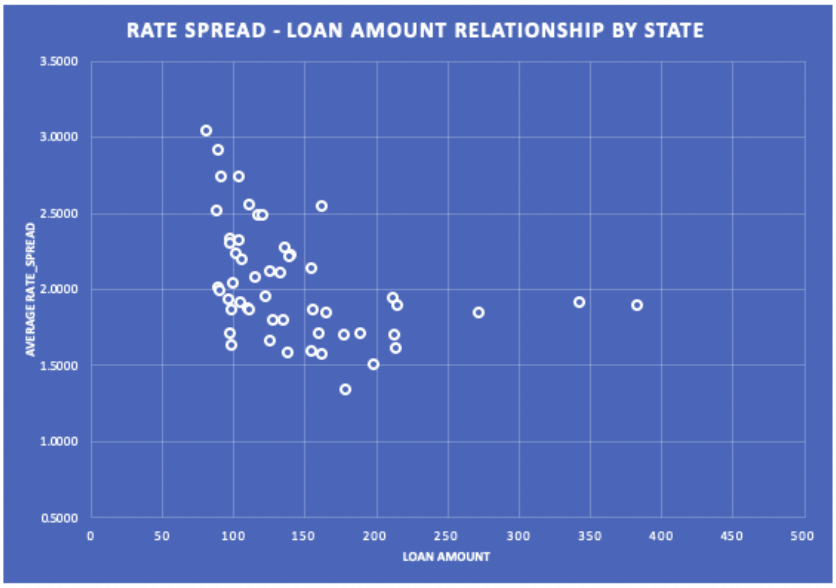
STATE CODES ANALYSIS

States – Highest vs Lowest

State Code	Rate Spread Average	Average Loan Amount
36	3.0414	80.8462
35	2.9163	88.7516
13	2.7477	91.2907
2	2.7436	103.0360
34	2.5614	110.7372
49	2.5494	161.0322
21	2.5188	88.2981
40	2.4914	117.2132
ALL DATA	**1.97911**	**142.5749**
12	1.6654	124.6578
25	1.6335	98.6959
19	1.6154	212.8940
33	1.6008	154.3174
15	1.5891	137.4645
47	1.5751	161.9017
46	1.5158	198.0789
24	1.3485	178.0951

Reviewing the ranked data and the overall loan amount to average rate spread values having an inversely related relationship, we see most states follow this trend. It appears state 21 is an outlier. On the lower end, state 25 appears to have a lower loan amount than associated with the average rate.

Plotting all average loan amounts against average rate spread by State renders the following:



Although, the state codes with the fewest data points (282) have less than 10% the number of entries compared to the state with the most data points (25596), this does not affect the overall distribution. Graphing only the 20 states with the most states renders the following:

Note that the primary missing data points are the ones with highest average loan amounts higher than 300. All others appear to be overlapping in the center cluster.

RATE SPREAD CLUSTER / CATEGORY ASSOCIATIONS

In anticipation of the data being used to make prediction models, categories were created for the reported rate spread. The goal of this initial classification was to aid in developing nodes and clusters. The assumption was that once categorized, a different model could be applied to each cluster for numerical scoring.

RATE SPREAD CLASSIFICATION

The value rate spread was broken into five categories, all values inclusive, as follows:

LOW	1
ML	2 to 4
MD	5 to 6
MH	7 to 9
HIGH	10 or higher

Note that the data reports the value only as integers. Maximum reported value is 99.

RATE CLASS DISTRIBUTION BY CATEGORIES

The following section looks at the distribution within each category for the rate spread class.

		LOW	ML	MD	MH	HIGH	Grand Total
PRE-APPROVAL	1	65.19%	33.14%	1.34%	0.26%	0.08%	100.00%
	2	74.03%	24.28%	1.34%	0.34%	0.01%	100.00%
	3	52.50%	35.03%	9.05%	3.33%	0.08%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

		LOW	ML	MD	MH	HIGH	Grand Total
Loan Type	1	36.32%	44.46%	14.08%	5.00%	0.15%	100.00%
	2	74.86%	23.21%	1.35%	0.57%	0.00%	100.00%
	3	78.37%	21.35%	0.28%	0.00%	0.00%	100.00%
	4	90.08%	9.76%	0.10%	0.00%	0.05%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

The primary stand-out value is the Loan Type 4 with over 90% of its entries in the LOW rate spread class. This loan type – Farm or Rural – has an average rate spread of 1.1215.

		LOW	ML	MD	MH	HIGH	Grand Total
Prop Type	1	64.82%	31.58%	2.87%	0.67%	0.06%	100.00%
	2	17.47%	38.88%	30.41%	13.12%	0.13%	100.00%
	3	22.27%	47.06%	30.67%	0.00%	0.00%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

Property type 2 and 3 is the stand-out variable in this table with each having about 30% of its entries in the MID rate spread class, compared to 7.10% for the whole data. Both of these types also have a lower LOW class concentration compared to the whole data.

		LOW	ML	MD	MH	HIGH	Grand Total
Loan Purpose	1	60.66%	29.44%	7.13%	2.73%	0.03%	100.00%
	2	23.32%	52.05%	17.19%	7.02%	0.42%	100.00%
	3	55.91%	38.81%	4.33%	0.84%	0.12%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

Most of the breakdown by loan purpose is consistent with the whole data values. The exception being Loan Purpose 2 – Home Improvement – carrying most of its distribution under “ML” rate spread class at 52% compared to 32% for the whole data. In turn, its “LOW” class has less than half the concentration (23%) compared to the whole data (57%).

		LOW	ML	MD	MH	HIGH	Grand Total
Loan Group	1	10.53%	50.24%	21.78%	16.89%	0.56%	100.00%
	2	33.51%	43.71%	16.28%	6.45%	0.05%	100.00%
	3	54.67%	36.01%	9.17%	0.14%	0.02%	100.00%
	4	64.38%	31.54%	3.96%	0.11%	0.00%	100.00%
	5	70.45%	27.51%	1.96%	0.07%	0.00%	100.00%
	6	73.93%	24.70%	1.31%	0.05%	0.01%	100.00%
	7	76.13%	22.46%	1.31%	0.09%	0.01%	100.00%
	8	71.55%	26.22%	1.80%	0.36%	0.07%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

Loan group, derived from loan amount, shows no major deviations. Stand out value is the Loan Group 1 with “ML” rate class at 50% compared to 32% for the whole data. In viewing the whole table the two findings when

analyzing loan amounts and average spreads hold with the highest amounts carrying the lowest rate spread averages.

		LOW	ML	MD	MH	HIGH	Grand Total
Ethnicity	1	64.50%	28.30%	5.33%	1.82%	0.05%	100.00%
	2	57.54%	34.29%	5.95%	2.14%	0.07%	100.00%
	3	42.75%	27.69%	21.44%	8.06%	0.05%	100.00%
	4	60.22%	36.27%	1.26%	1.54%	0.70%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

Under Ethnicity, class 3 – information not provided – has the highest concentration under MD rate class at 21% compared to 7% for the whole data.

		LOW	ML	MD	MH	HIGH	Grand Total
race	1	46.21%	33.41%	13.74%	6.58%	0.06%	100.00%
	2	60.94%	34.19%	4.08%	0.78%	0.00%	100.00%
	3	60.46%	29.39%	7.26%	2.86%	0.03%	100.00%
	4	66.23%	29.09%	3.07%	1.61%	0.00%	100.00%
	5	57.42%	33.49%	6.60%	2.42%	0.07%	100.00%
	6	54.34%	28.25%	13.25%	4.06%	0.10%	100.00%
	7	59.19%	38.01%	0.47%	1.56%	0.78%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

The race breakdown shows no major deviations from the averages as a whole. Race 6 – not reported – and 1 – native American – have a higher than average MD rate spread class concentration around 13% compared to 7.1% for the whole data.

		LOW	ML	MD	MH	HIGH	Grand Total
minority percentage group	1	49.98%	38.47%	8.15%	3.30%	0.10%	100.00%
	2	57.33%	32.64%	7.24%	2.70%	0.09%	100.00%
	3	58.06%	31.92%	7.33%	2.61%	0.07%	100.00%
	4	60.39%	30.17%	7.07%	2.32%	0.06%	100.00%
	5	62.58%	29.54%	5.90%	1.94%	0.04%	100.00%
		57.55%	**32.71%**	**7.10%**	**2.57%**	**0.07%**	**100.00%**

Final analysis is a categorial breakdown of reported minority percentage population within a particular area. Most values fall closely to the averages found across the whole data.

CONCLUSION

The analysis has shown a strong relationship between average rate spread and the two primary numerical values in the data: applicant income; and, loan amount. As one increases the observed rate spread average becomes lower. However, for the largest values the trend reverses and the rate spread is less consistent with data points showing a higher separation of rate spread averages.