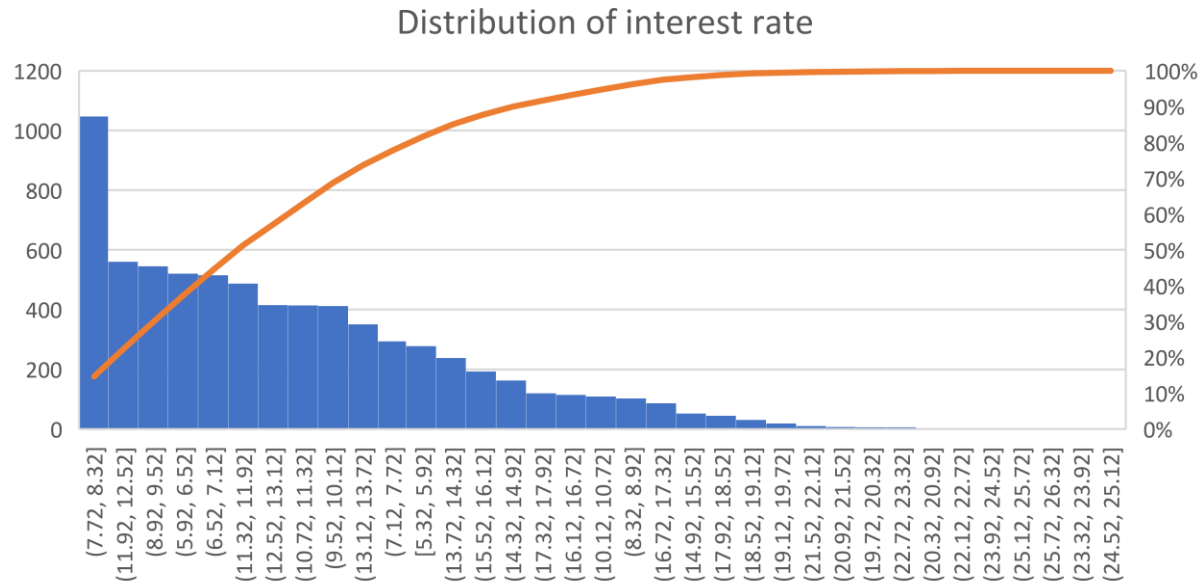
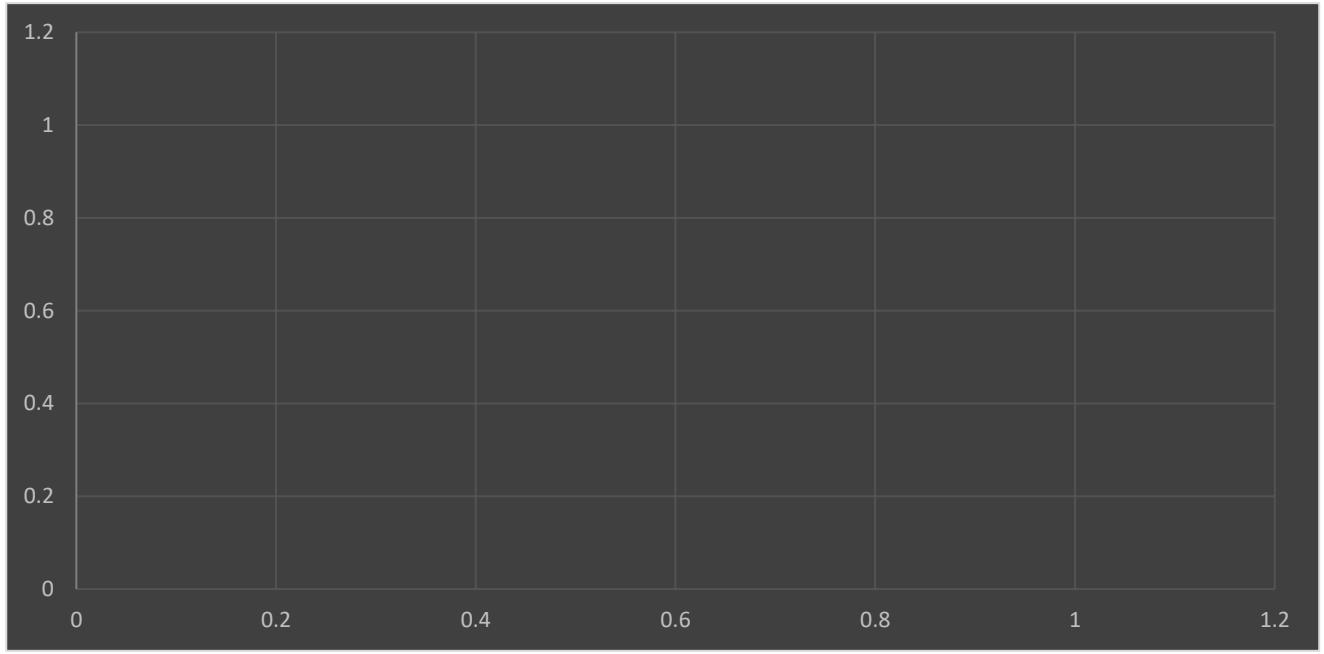
Presentation on investment allocation

Interest rate distribution



The first plot here shows distribution of interest rate. We can see here that the highest bin is between 7.72 and 8.32 and 80% of rate is from 5.32 to 13.72

Interest rate vs. average annual income

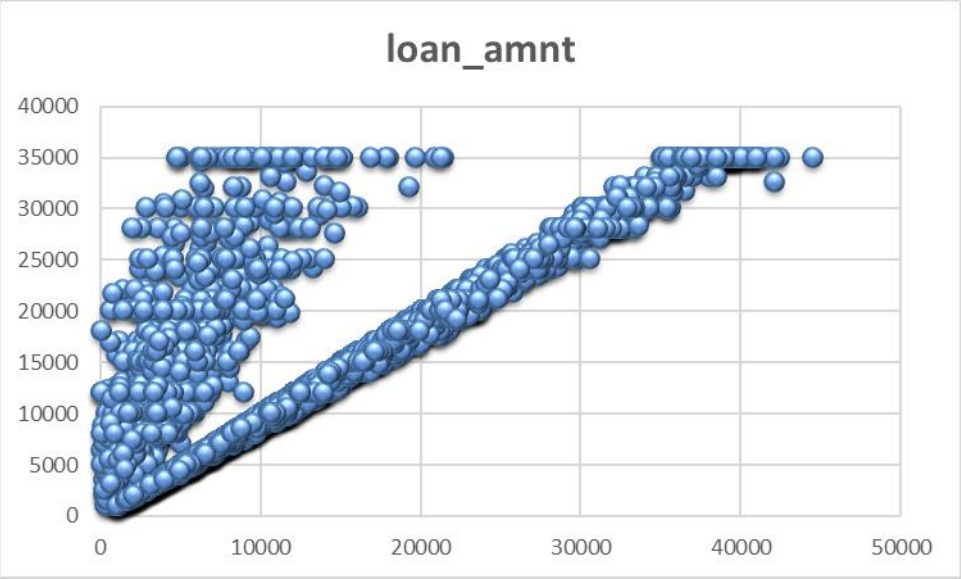


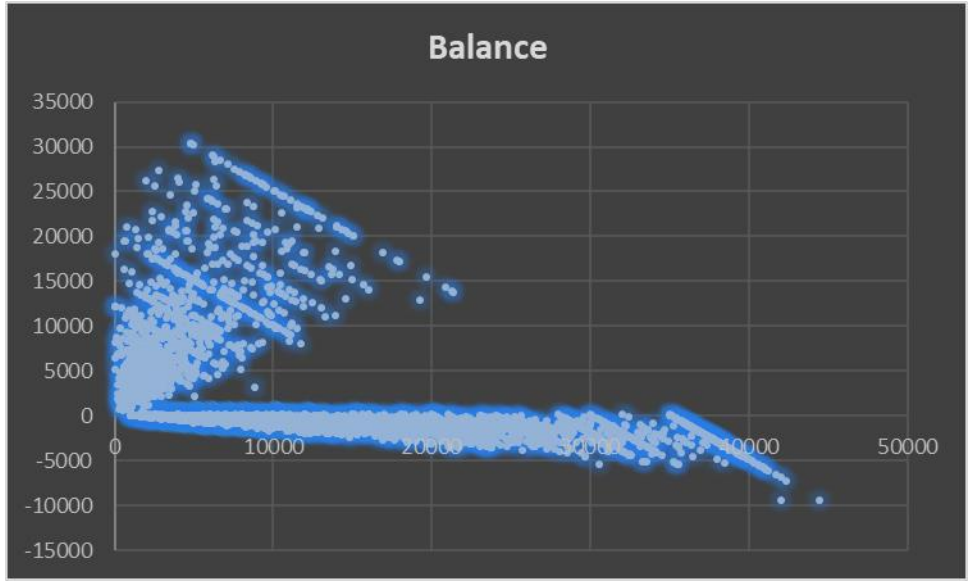
This plot is the relation between interest rate and average annual income. There is a moderate relationship between these two.

Total payment vs loan amount

In this Total Payment v Loan Amount scatterplot, we see that there are two clusters: one has a good linear relationship and another one is more scattered cloud like cluster. Further investing the matter, a second plot is created by taking balance(loan amount-total payment).

In this second plot it is clearly seen the linear relationship exists for the loan amount which have been paid fully with the interest charged for the loan. The cloud like cluster is for the loans which have not yet fully paid yet or have not matured.





Regression Model

A regression model is built to predict the the amount of losses, where **R-Squared** and the **RMS Error are** - 0.832889298 & - 2481.591206 respectively.

The features that predict if the loan status will be bad the most are **annual_inc, revol_bal,**

loan_amnt and **revol_util.**

Cluster analysis

Cluster	loan_amnt	int_rate	installment	annual_inc	revol_bal	total_acc	acc_open_past_24mths	total_pymnt
Cluster 1	13413	10	436	69957	16735	27	5	12629
Cluster 2	24167	9	776	173789	53786	33	5	23829
Cluster 3	8671	11	286	39368	9754	22	5	7756
Cluster 4	18843	10	606	109040	24204	29	5	18337
Cluster 5	27538	10	890	403833	94701	33	5	26454

Using k-means clustering with the normalized data option and k=5, the above clusters are explained below:

Cluster 1 - Loans are big, annual income high and payment is done.

Cluster 2 - Income is below average, smaller loans with a low-interest rate, and who opened their accounts recently.

Cluster 3 - Lowest income class with highest interest rate.

Cluster 4 - average income cluster, who pay most of their loans and opened account in the distance past.

Cluster 5 - Individuals with average income, who opened their accounts recently and pay most of their loans.

2 Million Invested in Each Cluster

Total return is 46178 and the probability is 1% to exceed half a million

Cluster	Int_Rate_Mean	Int_Rate_Dev	Interest%	Mean_Default_Rate	Default rate SDV	Default Rate%	Loss Alpha	Loss Beta	Loss Fraction	Investment	Return
1	10.78%	3.62%	11.59%	10.19%	0.77%	11.32%	2.19	1.15	0.98	2,000,000	(16,865.90)
2	9.32%	2.74%	10.32%	11.50%	0.58%	10.34%	1.79	0.84	0.94	2,000,000	(9,351.69)
3	11.54%	3.19%	7.86%	21.47%	0.88%	20.89%	0.75	0.35	0.38	2,000,000	(34,418.90)
4	8.24%	2.06%	7.44%	10.73%	0.51%	10.47%	1.95	0.88	0.87	2,000,000	(48,573.43)
5	12.79%	2.91%	18.95%	28.33%	0.73%	28.49%	0.48	0.22	0.20	2,000,000	155,388.56
									Total Return		46,178.63
									Mean Return	0.46%	9,235.73
									STD Return	3.717%	74343.88
									Sharpe ratio	-41.4%	

Optimal Investment

To optimize Sharpe ratio, we get lower return.

Cluster	Int_Rate_Mean	Int_Rate_Dev	Interest%	Mean_Default_Rate	Default rate SDV	Default Rate%	Loss Alpha	Loss Beta	Loss Fraction	Investment	Return
1	10.78%	3.62%	11.59%	10.19%	0.77%	11.32%	2.19	1.15	0.98	-	-
2	9.32%	2.74%	10.32%	11.50%	0.58%	10.34%	1.79	0.84	0.94	-	-
3	11.54%	3.19%	7.86%	21.47%	0.88%	20.89%	0.75	0.35	0.38	-	-
4	8.24%	2.06%	7.44%	10.73%	0.51%	10.47%	1.95	0.88	0.87	0	(0.00)
5	12.79%	2.91%	18.95%	28.33%	0.73%	28.49%	0.48	0.22	0.20	2,000,000	155,388.56
									Total Return		155,388.56
									Mean Return	1.55%	31,077.71
									STD Return	3.108%	62155.42
									Sharpe ratio	(0.14)	