

Prepare the Data

3 Types of Data

- **Numerical**

Discrete:	10, 15, IQ, SAT
Continuous:	.005, 3.1415, GPA

- **Categorical** Gender, Nationality

- **Ordinal** 1st, 2nd, star rating

STUDENT LOAN APPLICATION

Personal Information

(Last)

(First)

(City)

(Middle Initial)

Home Telephone
() -

Other Telephone
() -

(Zip)



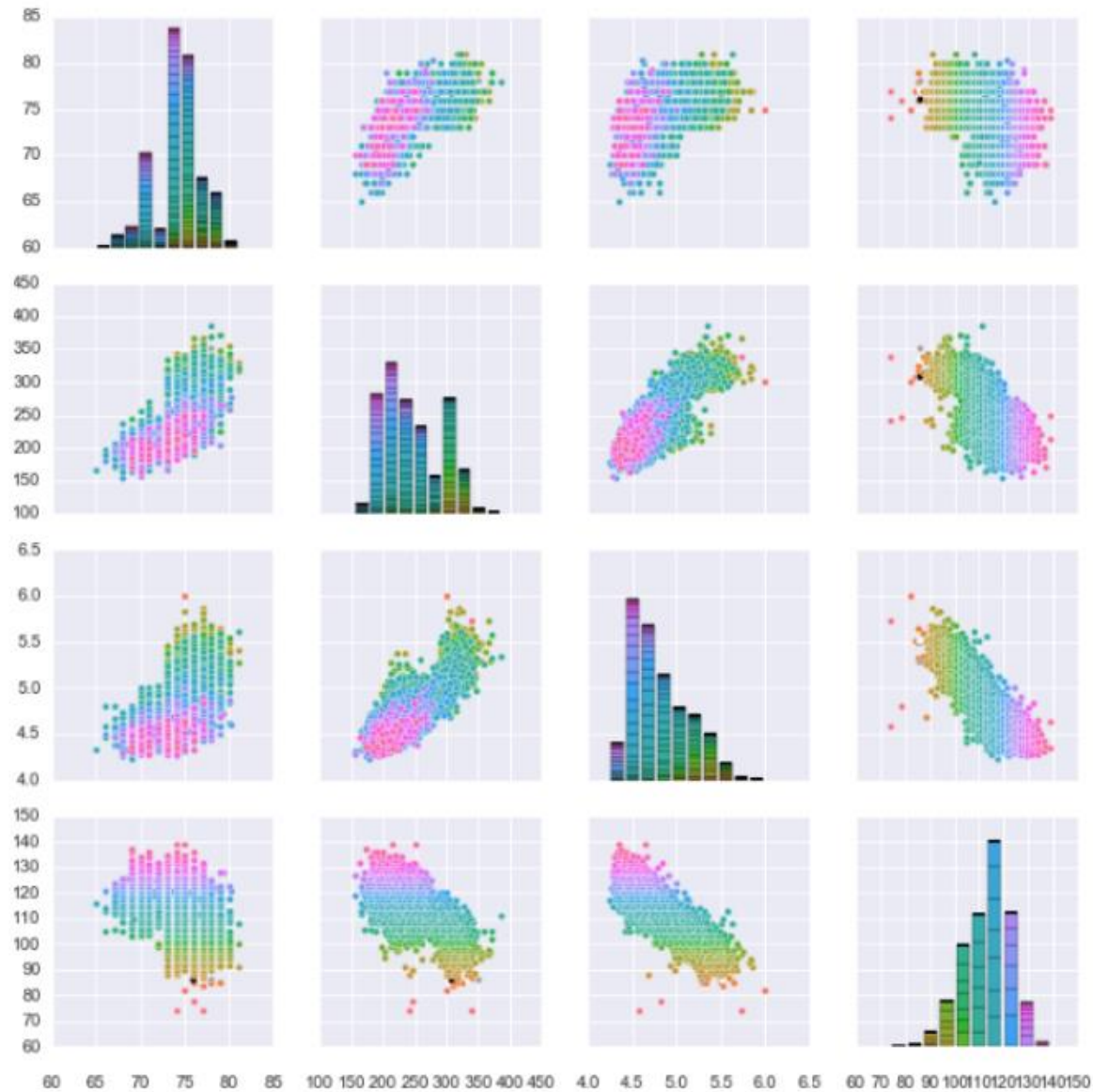
Student Loan Data

Age	Years Since Graduation	Income	Loan Balance	Next Payment	Defaulted
22	0	39247	14947.63	123	N
37	9	57981	27172.43	365	N
26	5	49031	18102.54	273	Y
42	16	63989	93429.79	491	Y
26	1	41250	13456.54	215	N
29	7	56482	29048.47	316	N
43	9	71202	26279.13	385	N
30	11	27649	6826.10	55	N

Student Loan Data Summary

	Age	Years Since Graduation	Income	Loan Balance	Next Payment
count	4947	4947	4947	4947	4933
mean	32.17	6.33	57981.98	37172.43	376.51
std	6.12	4.21	9031.62	18102.54	137.09
min	2	0.00	0.00	93.00	5.00
25%	26	3	41250.00	13456.00	112.00
50%	29	7	56482.00	39084.00	381.00
75%	43	9	71202.00	66212.00	485.00
max	67	31	276498.00	269910.00	755.00

Pair Plot



Clean the Data

Student Loan Data

Loan ID#	Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
10148975	Yes	1	31567	327	0	N
19773966	Yes	3	27909	-	3	N
25220947	Yes	2	11,463	243	25	Y
17090812	No	2	29801	255	15	Y
23956341	Yes	3	18755	173	0	N
12680900	Yes	1	16,211	122	7	N
23435111	No	1	5064	84	0	N

Student Loan Data

Name	Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
Ken	Yes	1	31567	327	0	N
Mary	Yes	3	27909	-	3	N
Reuben	Yes	2	11,463	243	25	Y
Amulya	No	2	29801	255	15	Y
Diane	Yes	3	18755	173	0	N
Timothy	Yes	1	16,211	122	7	N
Michelle	No	1	5064	84	0	N

Student Loan Data

Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
Yes	1	31567	327	0	N
Yes	3	27909	-	3	N
Yes	2	11,463	243	25	Y
No	2	29801	255	15	Y
Yes	3	18755	173	0	N
Yes	1	16,211	122	7	N
No	1	5064	84	0	N

Student Loan Data

Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
Yes	1	31567	327	0	N
Yes	3	27909	-	3	N
Yes	2	11463	243	25	Y
No	2	29801	255	15	Y
Yes	3	18755	173	0	N
Yes	1	16211	122	7	N
No	1	5064	84	0	N

Student Loan Data

Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
Yes	1	31567	327	0	N
Yes	3	27909	200	3	N
Yes	2	11463	243	25	Y
No	2	29801	255	15	Y
Yes	3	18755	173	0	N
Yes	1	16211	122	7	N
No	1	5064	84	0	N

Transform the Data

Student Loan Data

Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
Yes	1	31567	327	0	N
Yes	3	27909	200	3	N
Yes	2	11463	243	25	Y
No	2	29801	255	15	Y
Yes	3	18755	173	0	N
Yes	1	16211	122	7	N
No	1	5064	84	0	N

Student Loan Data

Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	1	31567	327	0	N
1	3	27909	200	3	N
1	2	11463	243	25	Y
0	2	29801	255	15	Y
1	3	18755	173	0	N
1	1	16211	122	7	N
0	1	5064	84	0	N

One-hot Encoding

Student Loan Data

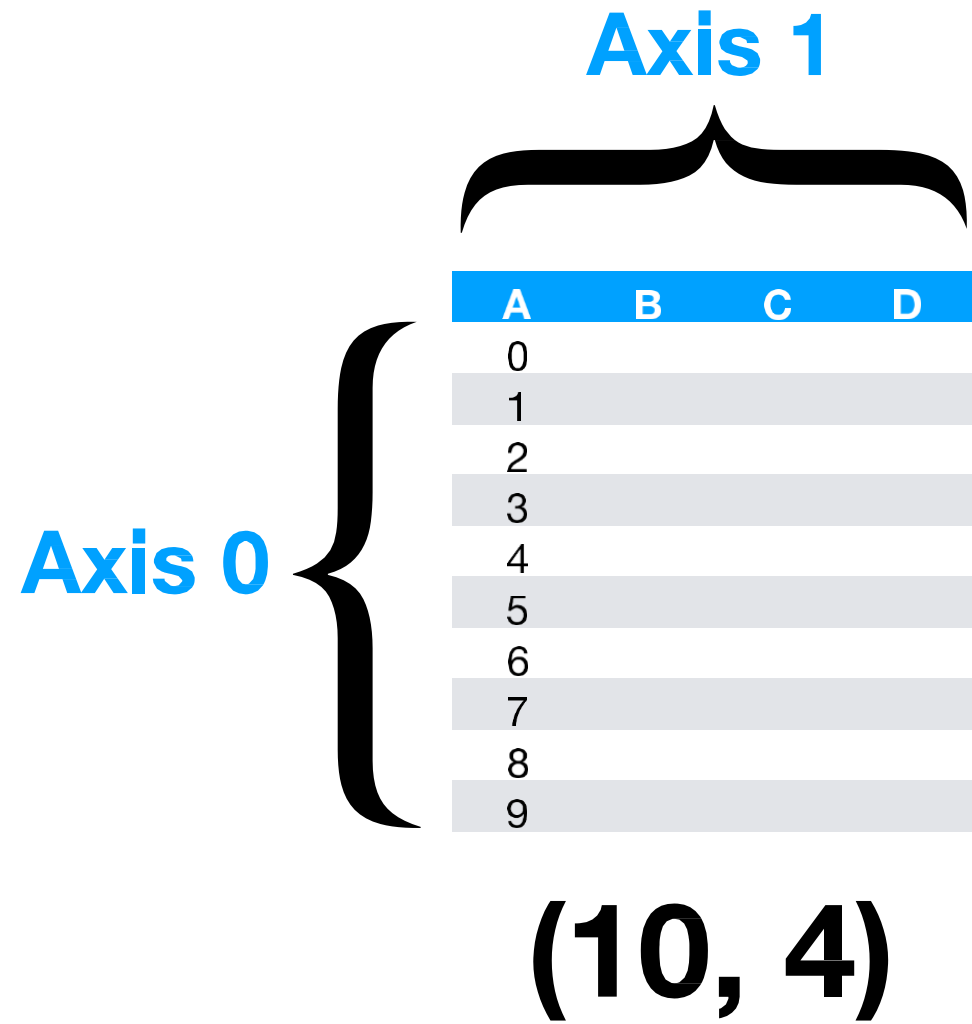
Graduated	Loan Type	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	1	31567	327	0	N
1	3	27909	200	3	N
1	2	11463	243	25	Y
0	2	29801	255	15	Y
1	3	18755	173	0	N
1	1	16211	122	7	N
0	1	5064	84	0	N

Student Loan Data

Graduated	Loan 1	Loan 2	Loan 3	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	1	0	0	31567	327	0	N
1	0	0	1	27909	200	3	N
1	0	1	0	11463	243	25	Y
0	0	1	0	29801	255	15	Y
1	0	0	1	18755	173	0	N
1	1	0	0	16211	122	7	N
0	1	0	0	5064	84	0	N

Student Loan Data

Graduated	Loan 1	Loan 2	Loan 3	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	1	0	0	31567	327	0	0
1	0	0	1	27909	200	3	0
1	0	1	0	11463	243	25	1
0	0	1	0	29801	255	15	1
1	0	0	1	18755	173	0	0
1	1	0	0	16211	122	7	0
0	1	0	0	5064	84	0	0



DataFrame

Student Loan Data

Graduated	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	31567	327	0	0
1	27909	200	3	0
1	11463	243	25	1
0	29801	255	15	1
1	18755	173	0	0
1	16211	122	7	0
0	5064	84	0	0
0	17198	154	0	0
1	21309	201	0	0
0	14693	193	4	1

Student Loan Data

features →
(X)

Graduated	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	31567	327	0	0
1	27909	200	3	0
1	11463	243	25	1
0	29801	255	15	1
1	18755	173	0	0
1	16211	122	7	0
0	5064	84	0	0
0	17198	154	0	0
1	21309	201	0	0
0	14693	193	4	1

Student Loan Data

Graduated	Loan Balance	Next Payment	Months Delinquent	Defaulted	← target (y)
1	31567	327	0	0	
1	27909	200	3	0	
1	11463	243	25	1	
0	29801	255	15	1	
1	18755	173	0	0	
1	16211	122	7	0	
0	5064	84	0	0	
0	17198	154	0	0	
1	21309	201	0	0	
0	14693	193	4	1	

Student Loan Data

	Graduated	Loan Balance	Next Payment	Months Delinquent	Defaulted
	1	31567	327	0	0
	1	27909	200	3	0
	1	11463	243	25	1
sample →	0	29801	255	15	1
	1	18755	173	0	0
	1	16211	122	7	0
	0	5064	84	0	0
	0	17198	154	0	0
	1	21309	201	0	0
	0	14693	193	4	1

Student Loan Data

Graduated	Loan Balance	Next Payment	Months Delinquent	Defaulted
1	31567	327	0	0
1	27909	200	3	0
1	11463	243	25	1
0	29801	255	15	1
1	18755	173	0	0
1	16211	122	7	0
0	5064	84	0	0
0	17198	154	0	0
1	21309	201	0	0
0	14693	193	4	1

Student Loan Data

1	31567	327	0	0
1	27909	200	3	0
1	11463	243	25	1
0	29801	255	15	1
1	18755	173	0	0
1	16211	122	7	0
0	5064	84	0	0
0	17198	154	0	0
1	21309	201	0	0
0	14693	193	4	1

Student Loan Data

X = features	1	31567	327	0	0
	1	27909	200	3	0
	1	11463	243	25	1
	0	29801	255	15	1
	1	18755	173	0	0
	1	16211	122	7	0
	0	5064	84	0	0
	0	17198	154	0	0
	1	21309	201	0	0
	0	14693	193	4	1
					y = target