

# Collect The Dataset

There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.

In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to download the dataset.

Link: <https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset>

As the dataset is downloaded. Let us read and understand the data properly with the help of some visualisation techniques and some analysing techniques.

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount
LP001015	Male	Yes	0	Graduate	No	5720	0	11
LP001022	Male	Yes	1	Graduate	No	3076	1500	12
LP001031	Male	Yes	2	Graduate	No	5000	1800	20
LP001035	Male	Yes	2	Graduate	No	2340	2546	10
LP001051	Male	No	0	Not Graduate	No	3276	0	7
LP001054	Male	Yes	0	Not Graduate	Yes	2165	3422	15
LP001055	Female	No	1	Graduate	No	2226	0	5
LP001056	Male	Yes	2	Not Graduate	No	3881	0	14

LP001059	Male	Yes		2	e Graduate		13633	0	28
LP001067	Male	No		0	e Not Graduate	No	2400	2400	12
LP001078	Male	No		0	e Not Graduate	No	3091	0	9
LP001082	Male	Yes		1	e Graduate		2185	1516	16
LP001083	Male	No	3+		e Graduate	No	4166	0	4
LP001094	Male	Yes		2	e Graduate		12173	0	16
LP001096	Female	No		0	e Graduate	No	4666	0	12
LP001099	Male	No		1	e Graduate	No	5667	0	13
LP001105	Male	Yes		2	e Graduate	No	4583	2916	20
LP001107	Male	Yes	3+		e Graduate	No	3786	333	12
LP001108	Male	Yes		0	e Graduate	No	9226	7916	30
LP001115	Male	No		0	e Graduate	No	1300	3470	10
LP001121	Male	Yes		1	e Not Graduate	No	1888	1620	4
LP001124	Female	No	3+		e Not Graduate	No	2083	0	2
LP001128		No		0	e Not Graduate	No	3909	0	10
LP001135	Female	No		0	e Not Graduate	No	3765	0	12
LP001149	Male	Yes		0	e Graduate	No	5400	4380	29
LP001153	Male	No		0	e Graduate	No	0	24000	14
LP001163	Male	Yes		2	e Graduate	No	4363	1250	14
LP001169	Male	Yes		0	e Graduate	No	7500	3750	27
LP001174	Male	Yes		0	e Graduate	No	3772	833	5
LP001176	Male	No		0	e Graduate	No	2942	2382	12

LP001 177	Female	No	0	Not Graduate	No	2478	0	7
LP001 183	Male	Yes	2	Graduate	No	6250	820	19
LP001 185	Male	No	0	Graduate	No	3268	1683	15
LP001 187	Male	Yes	0	Graduate	No	2783	2708	15
LP001 190	Male	Yes	0	Graduate	No	2740	1541	10
LP001 203	Male	No	0	Graduate	No	3150	0	17
LP001 208	Male	Yes	2	Graduate		7350	4029	18
LP001 210	Male	Yes	0	Graduate	Yes	2267	2792	9
LP001 211	Male	No	0	Graduate	Yes	5833	0	11
LP001 219	Male	No	0	Graduate	No	3643	1963	13
LP001 220	Male	Yes	0	Graduate	No	5629	818	10
LP001 221	Female	No	0	Graduate	No	3644	0	11
LP001 226	Male	Yes	0	Not Graduate	No	1750	2024	9
LP001 230	Male	No	0	Graduate	No	6500	2600	20
LP001 231	Female	No	0	Graduate	No	3666	0	8
LP001 232	Male	Yes	0	Graduate	No	4260	3900	18
LP001 237	Male	Yes		Not Graduate	No	4163	1475	16
LP001 242	Male	No	0	Not Graduate	No	2356	1902	10
LP001 268	Male	No	0	Graduate	No	6792	3338	18
LP001 270	Male	Yes	3+	Not Graduate	Yes	8000	250	18
LP001 284	Male	Yes	1	Graduate	No	2419	1707	12
LP001 287		Yes	3+	Not Graduate	No	3500	833	12

LP001 291	Male	Yes	1	Graduate	No	3500	3077	16
LP001 298	Male	Yes	2	Graduate	No	4116	1000	3
LP001 312	Male	Yes	0	Not Graduate	Yes	5293	0	9
LP001 313	Male	No	0	Graduate	No	2750	0	13
LP001 317	Female	No	0	Not Graduate	No	4402	0	13
LP001 321	Male	Yes	2	Graduate	No	3613	3539	13
LP001 323	Female	Yes	2	Graduate	No	2779	3664	17
LP001 324	Male	Yes	3+	Graduate	No	4720	0	9
LP001 332	Male	Yes	0	Not Graduate	No	2415	1721	11
LP001 335	Male	Yes	0	Graduate	Yes	7016	292	12
LP001 338	Female	No	2	Graduate	No	4968	0	18
LP001 347	Female	No	0	Graduate	No	2101	1500	10
LP001 348	Male	Yes	3+	Not Graduate	No	4490	0	12
LP001 351	Male	Yes	0	Graduate	No	2917	3583	13
LP001 352	Male	Yes	0	Not Graduate	No	4700	0	13
LP001 358	Male	Yes	0	Graduate	No	3445	0	13
LP001 359	Male	Yes	0	Graduate	No	7666	0	18
LP001 361	Male	Yes	0	Graduate	No	2458	5105	18
LP001 366	Female	No		Graduate	No	3250	0	9
LP001 368	Male	No	0	Graduate	No	4463	0	6
LP001 375	Male	Yes	1	Graduate		4083	1775	13
LP001 380	Male	Yes	0	Graduate	Yes	3900	2094	23
LP001	Male	Yes	0	Not	No	4750	3583	14

386				Graduate				
LP001				Graduate				
400	Male	No		0	No	3583	3435	15
LP001				0				
407	Male	Yes		0	No	3189	2367	18
LP001				0				
413	Male	No		0	Yes	6356	0	5
LP001				0				
415	Male	Yes		1	No	3413	4053	
LP001				0				
419	Female	Yes		0	No	7950	0	18
LP001				0				
420	Male	Yes	3+		No	3829	1103	16
LP001								
428	Male	Yes	3+		No	72529	0	36
LP001								
445	Male	Yes		2	No	4136	0	14
LP001				0				
446	Male	Yes		0	No	8449	0	25
LP001				0				
450	Male	Yes		0	No	4456	0	13
LP001				0				
452	Male	Yes		2	No	4635	8000	10
LP001				0				
455	Male	Yes		0	No	3571	1917	13
LP001				0				
466	Male	No		0	No	3066	0	9
LP001								
471	Male	No		2	No	3235	2015	7
LP001				0				
472	Female	No		0		5058	0	20
LP001				0				
475	Male	Yes		0	Yes	3188	2286	13
LP001				0				
483	Male	Yes	3+		No	13518	0	39
LP001								
486	Male	Yes		1	No	4364	2500	18
LP001								
490	Male	Yes		2	No	4766	1646	10
LP001				1				
496	Male	Yes		1	No	4609	2333	12
LP001								
499	Female	Yes	3+		No	6260	0	11
LP001								
500	Male	Yes		1	No	3333	4200	25
LP001	Male	Yes		0	No	3500	3250	14

501				e					
LP001				Graduat					
517	Male	Yes	3+	e	No	9719	0	6	
LP001				Graduat					
527	Male	Yes	3+	e	No	6835	0	18	
LP001				Graduat					
534	Male	No		0 e	No	4452	0	13	
LP001	Fema			Graduat					
542	le	Yes		0 e	No	2262	0		
LP001				Graduat					
547	Male	Yes		1 e	No	3901	0	11	
				Not					
LP001				Graduat					
548	Male	Yes		2 e	No	2687	0	5	
LP001				Graduat					
558	Male	No		0 e	No	2243	2233	10	
LP001	Fema			Graduat					
561	le	Yes		0 e	No	3417	1287	20	
LP001				Graduat					
563		No		0 e	No	1596	1760	11	
LP001				Graduat					
567	Male	Yes	3+	e	No	4513	0	12	
LP001				Graduat					
568	Male	Yes		0 e	No	4500	0	14	
				Not					
LP001				Graduat					
573	Male	Yes		0 e	No	4523	1350	16	
LP001	Fema			Graduat					
584	le	No		0 e	Yes	4742	0	10	
LP001				Graduat					
587	Male	Yes		e	No	4082	0	9	
LP001	Fema			Graduat					
589	le	No		0 e	No	3417	0	10	
LP001	Fema			Graduat					
591	le	Yes		2 e	No	2922	3396	12	
LP001				Graduat					
599	Male	Yes		0 e	No	4167	4754	16	
LP001				Graduat					
601	Male	No	3+	e	No	4243	4123	15	
				Not					
LP001	Fema			Graduat					
607	le	No		0 e	No	0	1760	18	
LP001				Graduat					
611	Male	Yes		1 e	No	1516	2900	8	
LP001	Fema			Graduat					
613	le	No		0 e	No	1762	2666	10	
LP001				Graduat					
622	Male	Yes		2 e	No	724	3510	21	
LP001				Graduat					
627	Male	No		0 e	No	3125	0	6	

LP001 650	Male	Yes		0	Graduate	No	2333	3803	14
LP001 651	Male	Yes	3+		Graduate	No	3350	1560	13
LP001 652	Male	No		0	Graduate	No	2500	6414	18
LP001 655	Female	No		0	Graduate	No	12500	0	30
LP001 660	Male	No		0	Graduate	No	4667	0	12
LP001 662	Male	No		0	Graduate	No	6500	0	7
LP001 663	Male	Yes		2	Graduate	No	7500	0	22
LP001 667	Male	No		0	Graduate	No	3073	0	7
LP001 695	Male	Yes		1	Not Graduate	No	3321	2088	7
LP001 703	Male	Yes		0	Graduate	No	3333	1270	12
LP001 718	Male	No		0	Graduate	No	3391	0	13
LP001 728	Male	Yes		1	Graduate	Yes	3343	1517	10
LP001 735	Female	No		1	Graduate	No	3620	0	9
LP001 737	Male	No		0	Graduate	No	4000	0	8
LP001 739	Male	Yes		0	Graduate	No	4258	0	12
LP001 742	Male	Yes		2	Graduate	No	4500	0	14
LP001 757	Male	Yes		1	Graduate	No	2014	2925	12
LP001 769		No			Graduate	No	3333	1250	11
LP001 771	Female	No	3+		Graduate	No	4083	0	10
LP001 785	Male	No		0	Graduate	No	4727	0	15
LP001 787	Male	Yes	3+		Graduate	No	3089	2999	10
LP001 789	Male	Yes	3+		Not Graduate		6794	528	13
LP001 791	Male	Yes		0	Graduate	Yes	32000	0	55
LP001 794	Male	Yes		2	Graduate	Yes	10890	0	26
LP001	Female	No		0	Graduate	No	12941	0	15

797	le			e				
LP001				Not				
815	Male	No	0	Graduat	No	3276	0	9
LP001				e				
817	Male	No	0	Not	Yes	8703	0	19
LP001				Graduat				
818	Male	Yes	1	e	No	4742	717	13
LP001				Graduat				
822	Male	No	0	e	No	5900	0	15
LP001				Graduat				
827	Male	No	0	e	No	3071	4309	18
LP001				Graduat				
831	Male	Yes	0	e	No	2783	1456	11
LP001				Graduat				
842	Male	No	0	e	No	5000	0	14
LP001				Not				
853	Male	Yes	1	Graduat	No	2463	2360	11
LP001				e				
855	Male	Yes	2	Graduat	No	4855	0	7
LP001				e				
857	Male	No	0	Not	Yes	1599	2474	12
LP001				Graduat				
862	Male	Yes	2	e	Yes	4246	4246	21
LP001				Graduat				
867	Male	Yes	0	e	No	4333	2291	13
LP001				Graduat				
878	Male	No	1	e	No	5823	2529	18
LP001				Not				
881	Male	Yes	0	Graduat	No	7895	0	14
LP001				e				
886	Male	No	0	Graduat	No	4150	4256	20
LP001				e				
906	Male	No	0	Graduat		2964	0	8
LP001				e				
909	Male	No	0	Graduat	No	5583	0	11
LP001				e				
911	Fema	No	0	Graduat	No	2708	0	6
LP001	le			e				
921	Male	No	1	Graduat	No	3180	2370	8
LP001				Not				
923	Male	No	0	Graduat	No	2268	0	17
LP001				e				
933	Male	No	2	Not	No	1141	2017	12
				Graduat				



LP001 943	Male	Yes	3+	0	Graduate	No	3042	3167	13
LP001 950	Female	Yes			Graduate		1750	2935	9
LP001 959	Female	Yes		1	Graduate	No	3564	0	7
LP001 961	Female	No		0	Graduate	No	3958	0	11
LP001 973	Male	Yes		2	Not Graduate	No	4483	0	13
LP001 975	Male	Yes		0	Graduate	No	5225	0	14
LP001 979	Male	No		0	Graduate	No	3017	2845	15
LP001 995	Male	Yes		0	Not Graduate	No	2431	1820	11
LP001 999	Male	Yes		2	Graduate		4912	4614	16
LP002 007	Male	Yes		2	Not Graduate	No	2500	3333	13
LP002 009	Female	No		0	Graduate	No	2918	0	6
LP002 016	Male	Yes		2	Graduate	No	5128	0	14
LP002 017	Male	Yes	3+		Graduate	No	15312	0	18
LP002 018	Male	Yes		2	Graduate	No	3958	2632	16
LP002 027	Male	Yes		0	Graduate	No	4334	2945	16
LP002 028	Male	Yes		2	Graduate	No	4358	0	11
LP002 042	Female	Yes		1	Graduate	No	4000	3917	17
LP002 045	Male	Yes	3+		Graduate	No	10166	750	15
LP002 046	Male	Yes		0	Not Graduate	No	4483	0	13
LP002 047	Male	Yes		2	Not Graduate	No	4521	1184	15
LP002 056	Male	Yes		2	Graduate	No	9167	0	23
LP002 057	Male	Yes		0	Not Graduate	No	13083	0	

LP002059	Male	Yes	2	Graduate	No	7874	3967	33
LP002062	Female	Yes	1	Graduate	No	4333	0	13
LP002064	Male	No	0	Graduate	No	4083	0	9
LP002069	Male	Yes	2	Not Graduate		3785	2912	18
LP002070	Male	Yes	3+	Not Graduate	No	2654	1998	12
LP002077	Male	Yes	1	Graduate	No	10000	2690	41
LP002083	Male	No	0	Graduate	Yes	5833	0	11
LP002090	Male	Yes	1	Graduate	No	4796	0	11
LP002096	Male	Yes	0	Not Graduate	No	2000	1600	11
LP002099	Male	Yes	2	Graduate	No	2540	700	10
LP002102	Male	Yes	0	Graduate	Yes	1900	1442	8
LP002105	Male	Yes	0	Graduate	Yes	8706	0	10
LP002107	Male	Yes	3+	Not Graduate	No	2855	542	9
LP002111	Male	Yes		Graduate	No	3016	1300	10
LP002117	Female	Yes	0	Graduate	No	3159	2374	10
LP002118	Female	No	0	Graduate	No	1937	1152	7
LP002123	Male	Yes	0	Graduate	No	2613	2417	12
LP002125	Male	Yes	1	Graduate	No	4960	2600	18
LP002148	Male	Yes	1	Graduate	No	3074	1083	14
LP002152	Female	No	0	Graduate	No	4213	0	8
LP002165		No	1	Not Graduate	No	2038	4027	10
LP002167	Female	No	0	Graduate	No	2362	0	5
LP002168	Male	No	0	Graduate	No	5333	2400	20

LP002 172	Male	Yes	3+	Graduat e	Yes	5384	0	15
LP002 176	Male	No		Graduat e	No	5708	0	15
LP002 183	Male	Yes		Not Graduat e	No	3754	3719	11
LP002 184	Male	Yes		Not Graduat e	No	2914	2130	15
LP002 186	Male	Yes		Graduat e	No	2747	2458	11
LP002 192	Male	Yes		Graduat e	No	7830	2183	21
LP002 195	Male	Yes		Graduat e	Yes	3507	3148	21
LP002 208	Male	Yes		Graduat e	No	3747	2139	12
LP002 212	Male	Yes		Graduat e	No	2166	2166	10
LP002 240	Male	Yes		Not Graduat e	No	3500	2168	14
LP002 245	Male	Yes		Not Graduat e	No	2896	0	8
LP002 253	Fema le	No		Graduat e	No	5062	0	15
LP002 256	Fema le	No		Graduat e	Yes	5184	0	18
LP002 257	Fema le	No		Graduat e	No	2545	0	7
LP002 264	Male	Yes		Graduat e	No	2553	1768	10
LP002 270	Male	Yes		Graduat e	No	3436	3809	10
LP002 279	Male	No		Graduat e	No	2412	2755	13
LP002 286	Male	Yes	3+	Not Graduat e	No	5180	0	12
LP002 294	Male	No		Graduat e	No	14911	14507	13
LP002 298		No		Graduat e	Yes	2860	2988	13
LP002 306	Male	Yes		Graduat e	No	1173	1594	2
LP002 310	Fema le	No		Graduat e	No	7600	0	9

LP002 311	Fema le	Yes	0	Graduat e	No	2157	1788	10
LP002 316	Male	No	0	Graduat e	No	2231	2774	17
LP002 321	Fema le	No	0	Graduat e	No	2274	5211	11
LP002 325	Male	Yes	2	Not Graduat e	No	6166	13983	10
LP002 326	Male	Yes	2	Not Graduat e	No	2513	1110	10
LP002 329	Male	No	0	Graduat e	No	4333	0	6
LP002 333	Male	No	0	Not Graduat e	No	3844	0	10
LP002 339	Male	Yes	0	Graduat e	No	3887	1517	10
LP002 344	Male	Yes	0	Graduat e	No	3510	828	10
LP002 346	Male	Yes	0	Graduat e		2539	1704	12
LP002 354	Fema le	No	0	Not Graduat e	No	2107	0	6
LP002 355		Yes	0	Graduat e	No	3186	3145	15
LP002 358	Male	Yes	2	Graduat e	Yes	5000	2166	15
LP002 360	Male	Yes		Graduat e	No	10000	0	
LP002 375	Male	Yes	0	Not Graduat e	Yes	3943	0	6
LP002 376	Male	No	0	Graduat e	No	2925	0	4
LP002 383	Male	Yes	3+	Graduat e	No	3242	437	14
LP002 385	Male	Yes		Graduat e	No	3863	0	7
LP002 389	Fema le	No	1	Graduat e	No	4028	0	13
LP002 394	Male	Yes	2	Graduat e	No	4010	1025	12
LP002 397	Fema le	Yes	1	Graduat e	No	3719	1585	11
LP002 399	Male	No	0	Graduat e		2858	0	12
LP002	Fema	Yes	0	Graduat	No	3833	0	9

400	le				e				
LP002					Graduat				
402	Male	Yes		0	e	No	3333	4288	16
LP002					Graduat				
412	Male	Yes		0	e	No	3007	3725	15
LP002	Fema				Graduat				
415	le	No		1	e		1850	4583	8
					Not				
LP002					Graduat				
417	Male	Yes	3+		e	No	2792	2619	17
LP002					Graduat				
420	Male	Yes		0	e	No	2982	1550	11
LP002					Graduat				
425	Male	No		0	e	No	3417	738	10
LP002					Graduat				
433	Male	Yes		1	e	No	18840	0	23
LP002					Graduat				
440	Male	Yes		2	e	No	2995	1120	18
LP002					Graduat				
441	Male	No			e	No	3579	3308	13
					Not				
LP002	Fema				Graduat				
442	le	Yes		1	e	No	3835	1400	11
					Not				
LP002	Fema				Graduat				
445	le	No		1	e	No	3854	3575	11
LP002					Graduat				
450	Male	Yes		2	e	No	5833	750	4
LP002					Graduat				
471	Male	No		0	e	No	3508	0	9
					Not				
LP002	Fema				Graduat				
476	le	Yes	3+		e	No	1635	2444	9
LP002	Fema				Graduat				
482	le	No		0	e	Yes	3333	3916	21
LP002					Graduat				
485	Male	No		1	e	No	24797	0	24
LP002					Graduat				
495	Male	Yes		2	e	No	5667	440	13
LP002	Fema				Graduat				
496	le	No		0	e	No	3500	0	9
LP002					Graduat				
523	Male	Yes	3+		e	No	2773	1497	10
LP002					Graduat				
542	Male	Yes		0	e		6500	0	14
LP002	Fema				Graduat				
550	le	No		0	e	No	5769	0	11
LP002									
551	Male	Yes	3+		Not Graduate		3634	910	17
LP002		No		0	Graduat	No	29167	0	18

553				e					
LP002				Graduat					
554	Male	No	0	e	No	2166	2057	12	
LP002				Graduat					
561	Male	Yes	0	e	No	5000	0	12	
LP002	Fema			Graduat					
566	le	No	0	e	No	5530	0	13	
				Not					
LP002				Graduat					
568	Male	No	0	e	No	9000	0	12	
LP002	Fema			Graduat					
570	le	Yes	2	e	No	10000	11666	46	
LP002				Graduat					
572	Male	Yes	1	e		8750	0	29	
				Not					
LP002				Graduat					
581	Male	Yes	0	e	No	2157	2730	14	
LP002				Graduat					
584	Male	No	0	e		1972	4347	10	
LP002				Graduat					
592	Male	No	0	e	No	4983	0	14	
LP002				Graduat					
593	Male	Yes	1	e	No	8333	4000		
LP002				Graduat					
599	Male	Yes	0	e	No	3667	2000	17	
LP002				Graduat					
604	Male	Yes	2	e	No	3166	2833	14	
				Not					
LP002				Graduat					
605	Male	No	0	e	No	3271	0	9	
LP002	Fema			Graduat					
609	le	Yes	0	e	No	2241	2000	8	
LP002									
610	Male	Yes	1	Not Graduate		1792	2565	12	
LP002	Fema			Graduat					
612	le	Yes	0	e	No	2666	0	8	
LP002				Graduat					
614		No	0	e	No	6478	0	10	
LP002									
630	Male	No	0	Not Graduate		3808	0	8	
				Not					
LP002	Fema			Graduat					
635	le	Yes	2	e	No	3729	0	11	
LP002				Graduat					
639	Male	Yes	2	e	No	4120	0	12	
LP002				Graduat					
644	Male	Yes	1	e	Yes	7500	0	7	
LP002				Graduat					
651	Male	Yes	1	e		6300	0	12	
LP002	Fema	No		Graduat	Yes	14987	0	17	

654	le				e Not Graduat				
LP002 657		Yes		1	e	Yes	570	2125	6
LP002 711	Male	Yes		0	Graduat	No	2600	700	9
LP002 712	Male	No		2	e Not Graduat	No	2733	1083	18
LP002 721	Male	Yes		2	e	Yes	7500	0	18
LP002 735	Male	Yes		2	Not Graduat	No	3859	0	12
LP002 744	Male	Yes		1	Graduat	No	6825	0	16
LP002 745	Male	Yes		0	Graduat	No	3708	4700	13
LP002 746	Male	No		0	e	No	5314	0	14
LP002 747	Fema le	No	3+		Graduat	No	2366	5272	15
LP002 754	Male	No			e	No	2066	2108	10
LP002 759	Male	Yes		2	Graduat	No	5000	0	14
LP002 760	Fema le	No		0	Graduat	No	3767	0	13
LP002 766	Fema le	Yes		0	e	No	7859	879	16
LP002 769	Fema le	Yes		0	Graduat	No	4283	0	12
LP002 774	Male	Yes		0	Not Graduat	No	1700	2900	6
LP002 775		No		0	e Not Graduat	No	4768	0	12
LP002 781	Male	No		0	Graduat	No	3083	2738	12
LP002 782	Male	Yes		1	e	No	2667	1542	14
LP002 786	Fema le	Yes		0	Not Graduat	No	1647	1762	18
LP002 790	Male	Yes	3+		Graduat	No	3400	0	8
LP002 791	Male	No		1	e		16000	5000	4
LP002	Male	Yes		0	Graduat	No	5333	0	9

793					e				
LP002					Graduat				
802	Male	No		0	e	No	2875	2416	9
LP002									
803	Male	Yes		1	Not Graduate		2600	618	12
LP002					Graduat				
805	Male	Yes		2	e	No	5041	700	15
LP002					Graduat				
806	Male	Yes	3+		e	Yes	6958	1411	15
LP002					Graduat				
816	Male	Yes		1	e	No	3500	1658	10
LP002					Graduat				
823	Male	Yes		0	e	No	5509	0	14
LP002					Graduat				
825	Male	Yes	3+		e	No	9699	0	30
					Not				
LP002					Graduat				
826	Fema	Yes		1	e	No	3621	2717	17
LP002					Graduat				
843	Fema	Yes		0	e	No	4709	0	11
LP002					Graduat				
849	Male	Yes		0	e	No	1516	1951	3
LP002					Graduat				
850	Male	No		2	e	No	2400	0	4
					Not				
LP002					Graduat				
853	Fema	No		0	e	No	3015	2000	14
LP002					Graduat				
856	Male	Yes		0	e	No	2292	1558	11
LP002					Graduat				
857	Male	Yes		1	e	Yes	2360	3355	8
LP002					Graduat				
858	Fema	No		0	e	No	4333	2333	16
LP002					Graduat				
860	Male	Yes		0	e	Yes	2623	4831	12
LP002					Graduat				
867	Male	No		0	e	Yes	3972	4275	18
					Not				
LP002					Graduat				
869	Male	Yes	3+		e	No	3522	0	8
LP002					Graduat				
870	Male	Yes		1	e	No	4700	0	8
LP002					Graduat				
876	Male	No		0	e	No	6858	0	17
LP002					Graduat				
878	Male	Yes	3+		e	No	8334	0	26
LP002					Graduat				
879	Male	Yes		0	e	No	3391	1966	13
LP002									
885	Male	No		0	Not	No	2868	0	7
					Graduat				



LP002 890	Male	Yes	2	e Not Graduat e	No	3418	1380	13
LP002 891	Male	Yes	0	e Graduat e	Yes	2500	296	13
LP002 899	Male	Yes	2	e Graduat e	No	8667	0	25
LP002 901	Male	No	0	e Graduat e	No	2283	15000	10
LP002 907	Male	Yes	0	e Graduat e	No	5817	910	10
LP002 920	Male	Yes	0	e Graduat e	No	5119	3769	12
LP002 921	Male	Yes	3+	Not Graduat e	No	5316	187	15
LP002 932	Male	Yes	3+	Graduat e	No	7603	1213	19
LP002 935	Male	Yes	1	Graduat e	No	3791	1936	8
LP002 952	Male	No	0	Graduat e	No	2500	0	6
LP002 954	Male	Yes	2	Not Graduat e	No	3132	0	7
LP002 962	Male	No	0	Graduat e	No	4000	2667	15
LP002 965	Fema le	Yes	0	Graduat e	No	8550	4255	9
LP002 969	Male	Yes	1	Graduat e	No	2269	2167	9
LP002 971	Male	Yes	3+	Not Graduat e	Yes	4009	1777	11
LP002 975	Male	Yes	0	Graduat e	No	4158	709	11
LP002 980	Male	No	0	Graduat e	No	3250	1993	12
LP002 986	Male	Yes	0	Graduat e	No	5000	2393	15
LP002 989	Male	No	0	Graduat e	Yes	9200	0	9

## Importing The Libraries:

Import the necessary libraries as shown in the image. (optional)  
Here we have used visualisation style as fivethirtyeight.

```
import pandas as pd
import numpy as np
import pickle
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import sklearn
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier, RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import RandomizedSearchCV
import imblearn
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, f1_score
```

## Read The Dataset:

Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.

In pandas we have a function called read\_csv() to read the dataset. As a parameter we have to give the directory of the csv file.

```
In [99]: data = pd.read_csv("data set.csv")
data
```

```
Out[99]:
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
0	LP001015	Male	Yes	0	Graduate	No	5720	0	110	360	1
1	LP001022	Male	Yes	1	Graduate	No	3076	1500	126	360	1
2	LP001031	Male	Yes	2	Graduate	No	5000	1800	208	360	1
3	LP001035	Male	Yes	2	Graduate	No	2340	2546	100	360	1
4	LP001051	Male	No	0	Not Graduate	No	3276	0	78	360	1
...	...	...	...	...	...	...	...	...	...	...	...
362	LP002971	Male	Yes	3+	Not Graduate	Yes	4009	1777	113	360	1
363	LP002975	Male	Yes	0	Graduate	No	4158	709	115	360	1
364	LP002980	Male	No	0	Graduate	No	3250	1993	126	360	0
365	LP002986	Male	Yes	0	Graduate	No	5000	2393	158	360	1
366	LP002989	Male	No	0	Graduate	Yes	9200	0	98	180	1

367 rows x 12 columns

## Data Preparation:

As we have understood how the data is, let's pre-process the collected data.

The download data set is not suitable for training the machine learning model as it might have so much randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

- Handling missing values
- Handling categorical data
- Handling Imbalance Data

Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

```
data.info()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...
362	False	False	False	False	False	False	False	False	False	False	False
363	False	False	False	False	False	False	False	False	False	False	False
364	False	False	False	False	False	False	False	False	False	False	False
365	False	False	False	False	False	False	False	False	False	False	False
366	False	False	False	False	False	False	False	False	False	False	False

```
data.isnull()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 367 entries, 0 to 366
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   Loan_ID               367 non-null   object  
 1   Gender                367 non-null   object  
 2   Married               367 non-null   object  
 3   Dependents            367 non-null   object  
 4   Education              367 non-null   object  
 5   Self_Employed         367 non-null   object  
 6   ApplicantIncome       367 non-null   int64   
 7   CoapplicantIncome     367 non-null   int64   
 8   LoanAmount            367 non-null   int64   
 9   Loan_Amount_Term      367 non-null   int64   
10   Credit_History         367 non-null   int64   
11   Property_Area          367 non-null   object  
dtypes: int64(5), object(7)
memory usage: 34.5+ KB
```

## Handling Missing Values:

- Let's find the shape of our dataset first. To find the shape of our data, the `df.shape` method is used. To find the data type, `df.info()` function is used.
- For checking the null values, `df.isnull()` function is used. To sum those null values we use `.sum()` function. From the below image we found that there are no null values present in our dataset. So we can skip handling the missing values step.
- From the above code of analysis, we can infer that columns such as gender, married, dependents, self employed, loan amount, loan amount term and credit history are having the missing values, we need to treat them in a required way.
- We will fill in the missing values in the numeric data type using the mean value of that particular column and categorical data type using the most repeated value.

## **Handling Categorical Values:**

As we can see our dataset has categorical data we must convert the categorical data to integer encoding or binary encoding.

To convert the categorical features into numerical features we use encoding techniques.

There are several techniques but in our project we are using manual encoding with the help of list

comprehension.

In our project, Gender ,married,dependents,self-employed,co-applicants income,loan amount ,loan amount term, credit history With list comprehension encoding is done.

```
data['Gender']=data['Gender'].astype('object')
data['Married']=data['Married'].astype('object')
data['Dependents']=data['Dependents'].astype('object')
data['Self_Employed']=data['Self_Employed'].astype('object')
data['CoapplicantIncome']=data['CoapplicantIncome'].astype('int64')
data['LoanAmount']=data['LoanAmount'].astype('int64')
data['Loan_Amount_Term']=data['Loan_Amount_Term'].astype('int64')
data['Credit_History']=data['Credit_History'].astype('int64')
```

## **Handling Imbalance Data:**

Data Balancing is one of the most important step, which need to be performed for classification models, because when we train our model on imbalanced dataset ,we will get biassed results, which means our model is able to predict only one class element

For Balancing the data we are using the SMOTE Method.

SMOTE: Synthetic minority over sampling technique, which will create new synthetic data points for under class as per the requirements given by us using KNN method.

```
from imblearn.combine import SMOTETomek
```