CS-5820 Project Group # 10

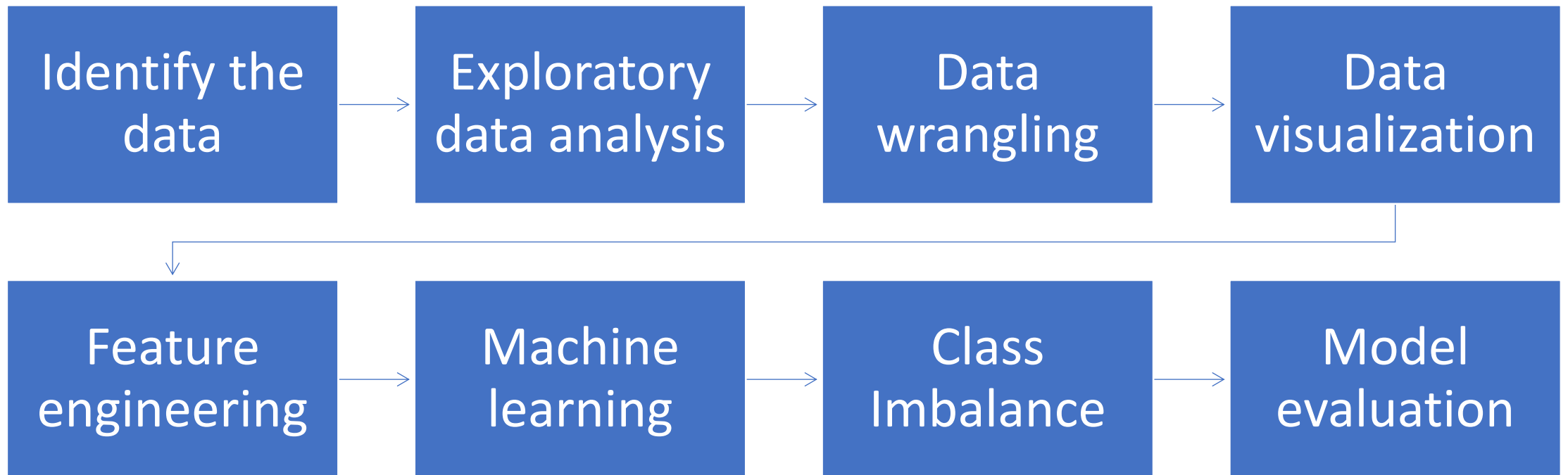# Credit Risk Evaluation using Machine Learning

By *Surya* Vaddhiparthy

Under the guidance of Professor Dr. Guan Yue Hong

Goal

Enable access to credit for people with limited credit history

Steps

Identify the data	Exploratory data analysis	Data wrangling	Data visualization
Feature engineering	Machine learning	Class Imbalance	Model evaluation

Click to access source code ➡ jupyter

Data

	SK_ID_CURR	TARGET	NAME_CONT	CODE_GEND	FLAG_OWN	FLAG_OWN	CNT_CHILDR	AMT_INCOM	AMT_CREDIT	AMT_ANNUI	AMT_GOODS	NAME_TYPE	NAME_INCO	NAME_EDUC	NAME_FAMI	NAME_HOUS	REGION_POI	DAYS_BIRTH	DAYS_EMPL	DAYS_REGIS	DAYS_ID_PU
2	100002	1	Cash loans	M	N	Y	0	202500	406597.5	24700.5	351000	Unaccompar	Working	Secondary /	Single / not	House / apar	0.018801	-9461	-637	-3648	-2120
3	100003	0	Cash loans	F	N	N	0	270000	1293502.5	35698.5	1129500	Family	State servan	Higher educa	Married	House / apar	0.003541	-16765	-1188	-1186	-291
4	100004	0	Revolving loa	M	Y	Y	0	67500	135000	6750	135000	Unaccompar	Working	Secondary /	Single / not	House / apar	0.010032	-19046	-225	-4260	-2531
5	100006	0	Cash loans	F	N	Y	0	135000	312682.5	29686.5	297000	Unaccompar	Working	Secondary /	Civil marriag	House / apar	0.008019	-19005	-3039	-9833	-2437
6	100007	0	Cash loans	M	N	Y	0	121500	513000	21865.5	513000	Unaccompar	Working	Secondary /	Single / not	House / apar	0.028663	-19932	-3038	-4311	-3458
7	100008	0	Cash loans	M	N	Y	0	99000	490495.5	27517.5	454500	Spouse, part	State servan	Secondary /	Married	House / apar	0.035792	-16941	-1588	-4970	-477
8	100009	0	Cash loans	F	Y	Y	1	171000	1560726	41301	1395000	Unaccompar	Commercial	Higher educa	Married	House / apar	0.035792	-13778	-3130	-1213	-619
9	100010	0	Cash loans	M	Y	Y	0	360000	1530000	42075	1530000	Unaccompar	State servan	Higher educa	Married	House / apar	0.003122	-18850	-449	-4597	-2379
10	100011	0	Cash loans	F	N	Y	0	112500	1019610	33826.5	913500	Children	Pensioner	Secondary /	Married	House / apar	0.018634	-20099	365243	-7427	-3514
11	100012	0	Revolving loa	M	N	Y	0	135000	405000	20250	405000	Unaccompar	Working	Secondary /	Single / not	House / apar	0.019689	-14469	-2019	-14437	-3992
12	100014	0	Cash loans	F	N	Y	1	112500	652500	21177	652500	Unaccompar	Working	Higher educa	Married	House / apar	0.0228	-10197	-679	-4427	-738
13	100015	0	Cash loans	F	N	Y	0	38419.155	148365	10678.5	135000	Children	Pensioner	Secondary /	Married	House / apar	0.015221	-20417	365243	-5246	-2512
14	100016	0	Cash loans	F	N	Y	0	67500	80865	5881.5	67500	Unaccompar	Working	Secondary /	Married	House / apar	0.031329	-13439	-2717	-311	-3227
15	100017	0	Cash loans	M	Y	N	1	225000	918468	28966.5	697500	Unaccompar	Working	Secondary /	Married	House / apar	0.016612	-14086	-3028	-643	-4911
16	100018	0	Cash loans	F	N	Y	0	189000	773680.5	32778	679500	Unaccompar	Working	Secondary /	Married	House / apar	0.010006	-14583	-203	-615	-2056
17	100019	0	Cash loans	M	Y	Y	0	157500	299772	20160	247500	Family	Working	Secondary /	Single / not	Rented apar	0.020713	-8728	-1157	-3494	-1368
18	100020	0	Cash loans	M	N	N	0	108000	509602.5	26149.5	387000	Unaccompar	Working	Secondary /	Married	House / apar	0.018634	-12931	-1317	-6392	-3866
19	100021	0	Revolving loa	F	N	Y	1	81000	270000	13500	270000	Unaccompar	Working	Secondary /	Married	House / apar	0.010966	-9776	-191	-4143	-2427
20	100022	0	Revolving loa	F	N	Y	0	112500	157500	7875	157500	Other_A	Working	Secondary /	Widow	House / apar	0.04622	-17718	-7804	-8751	-1259
21	100023	0	Cash loans	F	N	Y	1	90000	544491	17563.5	454500	Unaccompar	State servan	Higher educa	Single / not	House / apar	0.015221	-11348	-2038	-1021	-3964
22	100024	0	Revolving loa	M	Y	Y	0	135000	427500	21375	427500	Unaccompar	Working	Secondary /	Married	House / apar	0.015221	-18252	-4286	-298	-1800
23	100025	0	Cash loans	F	Y	Y	1	202500	1132573.5	37561.5	927000	Unaccompar	Commercial	Secondary /	Married	House / apar	0.025164	-14815	-1652	-2299	-2299
24	100026	0	Cash loans	F	N	N	1	450000	497520	32521.5	450000	Unaccompar	Working	Secondary /	Married	Rented apar	0.020713	-11146	-4306	-114	-2518
25	100027	0	Cash loans	F	N	Y	0	83250	239850	23850	225000	Unaccompar	Pensioner	Secondary /	Married	House / apar	0.006296	-24827	365243	-9012	-3684
26	100029	0	Cash loans	M	Y	N	2	135000	247500	12703.5	247500	Unaccompar	Working	Secondary /	Married	House / apar	0.026392	-11286	-746	-108	-3729
27	100030	0	Cash loans	F	N	Y	0	90000	225000	11074.5	225000	Unaccompar	Working	Secondary /	Married	House / apar	0.028663	-19334	-3494	-2419	-2893
28	100031	1	Cash loans	F	N	Y	0	112500	979992	27076.5	702000	Unaccompar	Working	Secondary /	Widow	House / apar	0.018029	-18724	-2628	-6573	-1827
29	100032	0	Cash loans	M	N	Y	1	112500	327024	23827.5	270000	Family	Working	Secondary /	Married	House / apar	0.019101	-15948	-1234	-5782	-3153
30	100033	0	Cash loans	M	Y	Y	0	270000	790830	57676.5	675000	Unaccompar	State servan	Higher educa	Single / not	House / apar	0.04622	-9994	-1796	-4668	-2661
31	100034	0	Revolving loa	M	N	Y	0	90000	180000	9000	180000	Unaccompar	Working	Higher educa	Single / not	With parents	0.030755	-10341	-1010	-4799	-3015
32	100035	0	Cash loans	F	N	Y	0	292500	665892	24592.5	477000	Unaccompar	Commercial	Secondary /	Civil marriag	House / apar	0.025164	-15280	-2668	-5266	-3787
33	100036	0	Cash loans	F	N	Y	0	112500	512064	25033.5	360000	Family	Working	Secondary /	Civil marriag	House / apar	0.008575	-11144	-1104	-7846	-2904
34	100037	0	Cash loans	F	N	N	0	90000	199008	20893.5	180000	Unaccompar	Working	Secondary /	Civil marriag	House / apar	0.010032	-12974	-4404	-7123	-4464
35	100039	0	Cash loans	M	Y	N	1	360000	733315.5	39069	679500	Unaccompar	Commercial	Secondary /	Married	House / apar	0.015221	-11694	-2060	-3557	-3557
36	100040	0	Cash loans	F	N	Y	0	135000	1125000	32895	1125000	Unaccompar	State servan	Higher educa	Married	House / apar	0.019689	-15997	-4585	-5735	-4067
37	100041	0	Cash loans	F	N	N	0	112500	450000	44509.5	450000	Unaccompar	Working	Higher educa	Married	House / apar	0.008575	-12158	-1275	-6265	-2009
38	100043	0	Cash loans	F	N	Y	2	198000	641173.5	23157	553500	Unaccompar	Commercial	Secondary /	Married	House / apar	0.01885	-17199	-768	-63	-735
39	100044	0	Cash loans	M	N	Y	0	121500	454500	15151.5	454500	Unaccompar	Working	Secondary /	Married	House / apar	0.030755	-21077	-1288	-5474	-4270
40	100045	0	Cash loans	F	N	Y	0	99000	247275	17338.5	225000	Unaccompar	Pensioner	Secondary /	Married	House / apar	0.006207	-23920	365243	-9817	-4969
41	100046	0	Revolving loa	M	Y	Y	0	180000	540000	27000	540000	Unaccompar	Working	Higher educa	Married	House / apar	0.025164	-16126	-1761	-8236	-4292
42	100047	1	Cash loans	M	N	Y	0	202500	1193580	35028	855000	Unaccompar	Commercial	Secondary /	Married	House / apar	0.025164	-17482	-1262	-1182	-1029
43	100048	0	Cash loans	F	N	Y	0	202500	604152	29196	540000	Unaccompar	Working	Secondary /	Married	House / apar	0.009175	-16971	-475	-3148	-513
44	100049	1	Cash loans	F	N	Y	0	135000	288873	16258.5	238500	Unaccompar	Working	Secondary /	Civil marriag	House / apar	0.007305	-13384	-3597	-45	-4409
45	100050	0	Cash loans	F	N	Y	0	108000	746280	42970.5	675000	Unaccompar	Pensioner	Higher educa	Single / not	House / apar	0.010966	-23548	365243	-5745	-4576
46	100051	0	Cash loans	M	N	Y	0	202500	661702.5	48280.5	598500	Unaccompar	Working	Secondary /	Civil marriag	House / apar	0.007114	-9827	-758	-4166	-2380
47	100052	0	Revolving loa	F	N	Y	1	90000	180000	9000	180000	Family	Working	Secondary /	Civil marriag	House / apar	0.009334	-7974	-895	-6591	-642
48	100053	0	Cash loans	F	N	Y	0	202500	305221.5	17649	252000	Family	Pensioner	Secondary /	Single / not	House / apar	0.019689	-20373	365243	-7760	-1993
49	100054	0	Cash loans	F	N	Y	0	99000	260640	26838	225000	Unaccompar	Working	Secondary /	Married	House / apar	0.022625	-20121	-5332	-9197	-3283
50	100055	0	Cash loans	F	N	N	0	130500	1350000	37255.5	1350000	Unaccompar	State servan	Secondary /	Married	House / apar	0.020713	-19727	-4411	-6225	-3199
51	100056	0	Cash loans	M	Y	Y	0	360000	1506816	49927.5	1350000	Unaccompar	Working	Secondary /	Married	House / apar	0.010006	-18264	-9523	-5504	-1797
52	100058	0	Revolving loa	F	N	Y	0	54000	135000	6750	135000	Unaccompar	State servan	Higher educa	Married	House / apar	0.01452	-12406	-1176	-1149	-2576

Exploratory Data Analysis and Data Wrangling
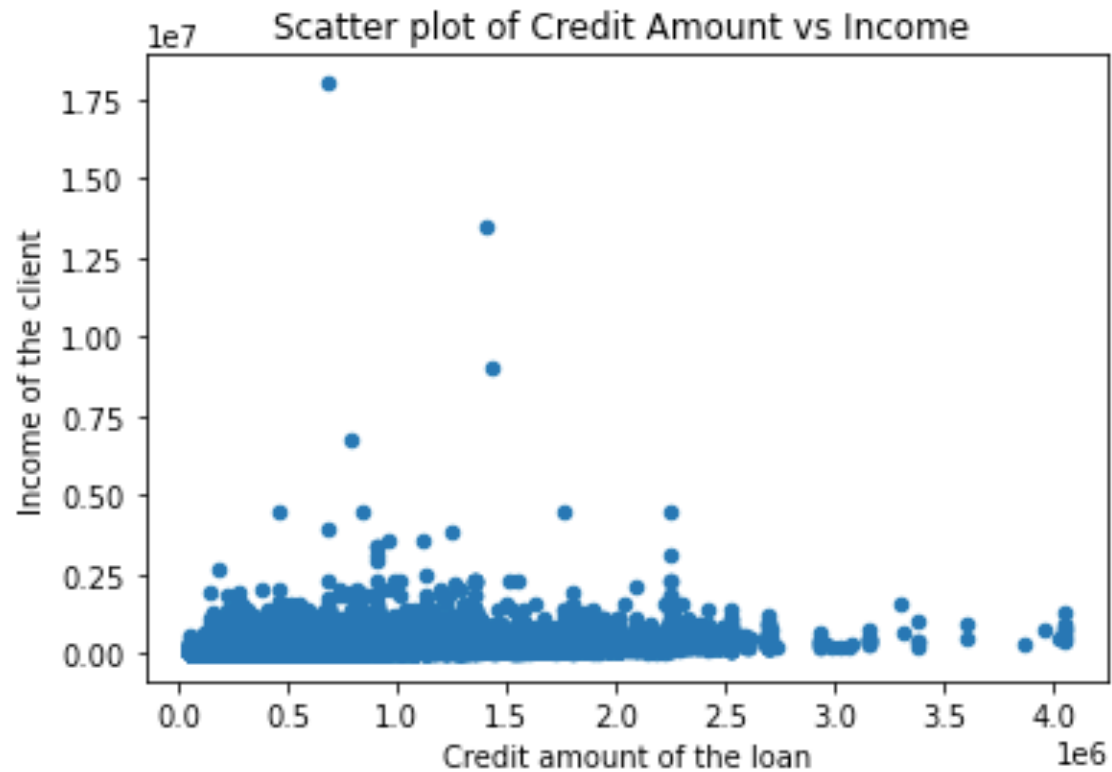
307,511 x 122

Target [1,0]

Int64, float64 and object

100,000 missing values

Imputation

Outliers

Data Visualization

Scatter plot of Credit Amount vs Income

df.plot(kind='scatter',x==' ',y='')
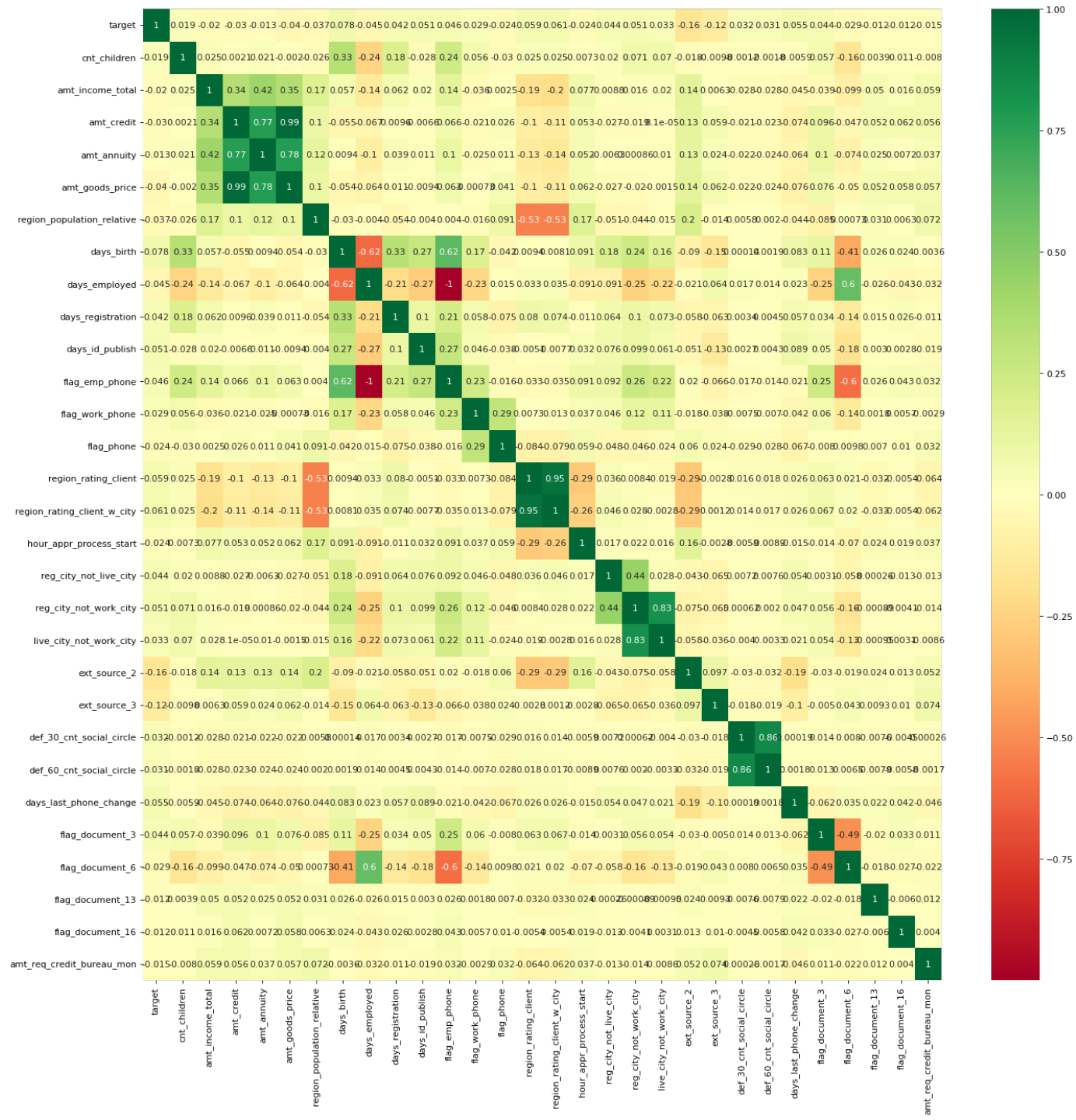
df.plot(kind='pie',legend=True,figsize=(12,12)).

df.plot(kind='bar')

plt.boxplot

Feature Engineering

Machine Learning (Logistic Regression Model)

```
For the model: LogisticRegression(multi_class='ovr', random_state=0) , the training accuracy
is 0.9191205164059706 and the test accracy is 0.9196936684985854
[[56560      5]
 [ 4934      3]]
```

	precision	recall	f1-score	support
0	0.92	1.00	0.96	56565
1	0.38	0.00	0.00	4937
accuracy			0.92	61502
macro avg	0.65	0.50	0.48	61502
weighted avg	0.88	0.92	0.88	61502

Class Imbalance

Proportion of borrowers that defaulted

Count

0-No Default, 1-Default
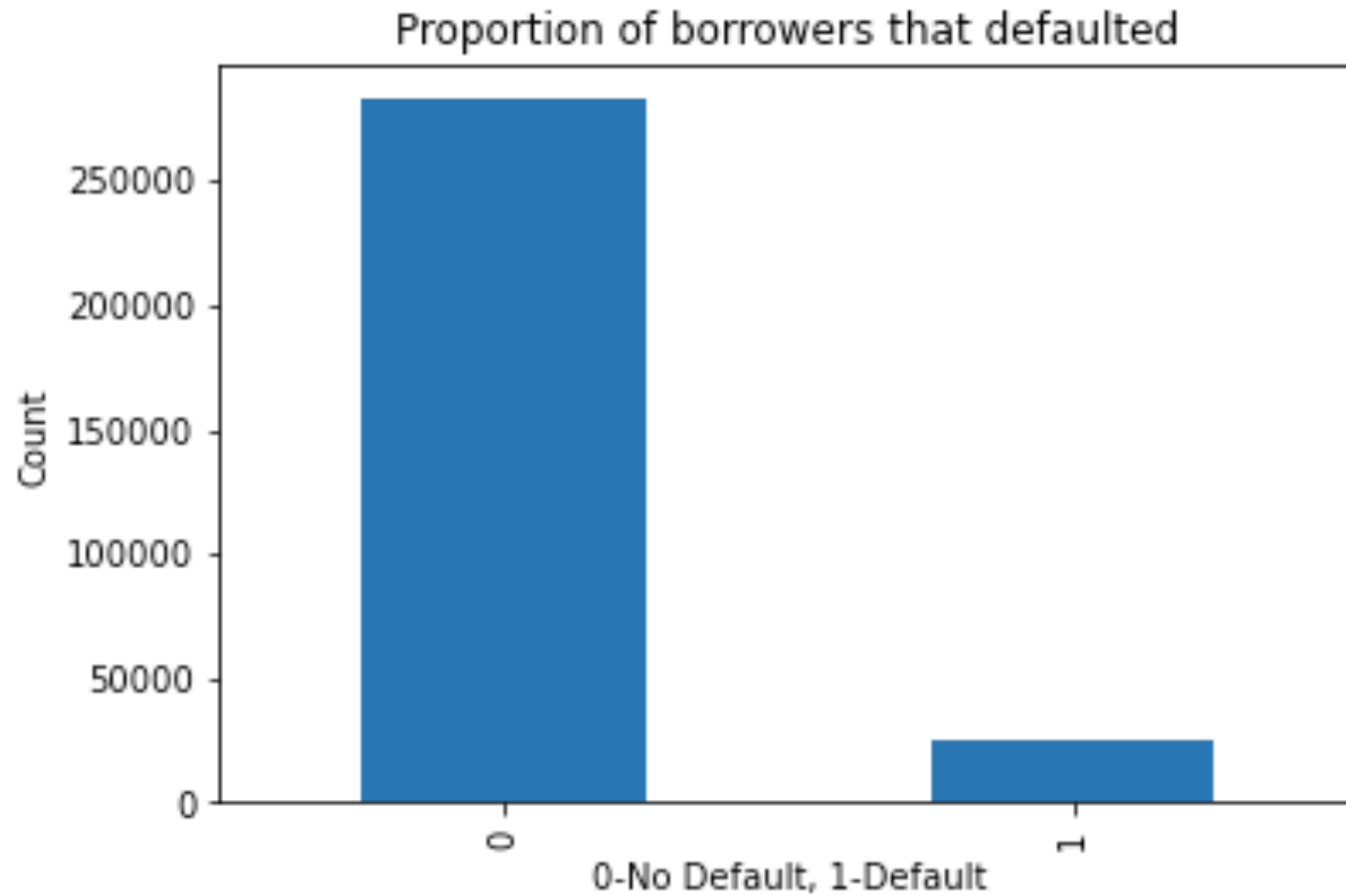
Extreme gradient boosting with synthetic minority over sampling

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric='logloss',
              feature_types=None, gamma=None, gpu_id=None, grow_policy=None,
              importance_type=None, interaction_constraints=None,
              learning_rate=None, max_bin=None, max_cat_threshold=None,
              max_cat_to_onehot=None, max_delta_step=None, max_depth=None,
              max_leaves=None, min_child_weight=None, missing=nan,
              monotone_constraints=None, n_estimators=100, n_jobs=None,
              num_parallel_tree=None, predictor=None, random_state=None, ...)
Mean cross-validation score: 0.95
K-fold CV average score: 0.95
[[41833   515]
 [ 4122 38336]]
```
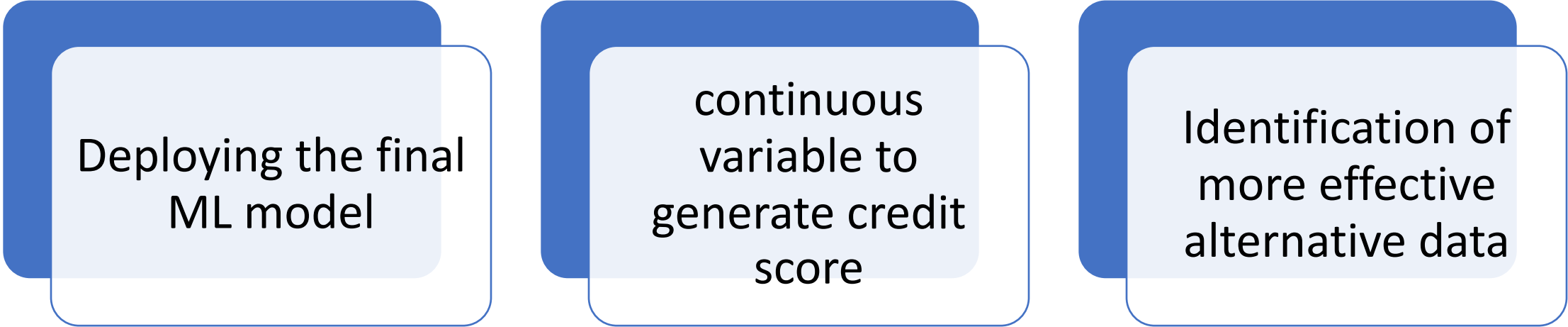
	precision	recall	f1-score	support
0	0.91	0.99	0.95	42348
1	0.99	0.90	0.94	42458
accuracy			0.95	84806
macro avg	0.95	0.95	0.95	84806
weighted avg	0.95	0.95	0.95	84806

Machine Learning Model Evaluation

Model	Precision	Recall	F1-score	Imbalanced
Logistic Regression	0.38	0.00	0.00	Yes
Stochastic Gradient Descent	0.05	0.00	0.00	Yes
XGBoost	0.55	0.02	0.04	Yes
Stochastic Gradient Descent	0.50	0.00	0.00	No
XGBoost	0.98	0.91	0.94	No
Random Forest	0.94	0.89	0.92	No

Future scope

Deploying the final ML model

continuous variable to generate credit score

Identification of more effective alternative data

Thank You!