1. Executive summary

The goal of this project is to build a fraud detection model to achieve a relatively high accuracy of the Card Transaction Data. The whole process goes through the steps of data description, data cleaning, variable creation, feature selection, model exploration, Final model performance displacement, and finally a financial curves and recommended cutoff is presented. After doing all the steps above, a final model of Random Forest is chosen, with top 20 features, n_estimator=100, criterion=gini, max_depth=15, min_samples_leaf=200, and max_deatures=log2. The model could achieve a Training accuracy of 0.759, a Testing accuracy rate of 0.726, and a OOT of 0.531. The model can capture 53.1% of all the fraud at the top 3 percent. After drawing the financial curves, we recommend a score cutoff at 3%, where we achieve a maximum overall savings at 19,332,000.

2. Description of the data

This data is about Card Transaction Data. It is a collection of card transaction records from a US government organization in year 2010. The data has 96,753 rows and 10 fields.

(1) Numerical Table

Field Name	% Populated	Min	Max	Mean	Stdev	% Zero
Date	100.00	01/01/2010	12/31/2010	/	/	0.00
Amount	100.00	0.01	3,102,045.53	427.89	10,006.14	0.00

Table 1

(2) Categorical Table

Field Name	% Population	# Unique Values	Most Common Value
Recnum	100.00	96,753	N/A
Cardnum	100.00	1,645	5142148452
Merchnum	96.51	13,092	930090121224
Merch description	100.00	13,126	GSA-FSS-ADV
Merch state	98.76	228	TN
Merch zip	95.19	4,568	38118
Transtype	100.00	4	P
Fraud	100.00	2	0

Table 2

3. Data cleaning

We first change date to datetime format. After excluding an extreme value which has an amount of $3,102,045.53 and including only 'P' in the Transgtype field, we filled in NAs with the imputation logic. The imputation logic is as follows.

Merchnum
- Mapped each merch Merch description with a Merchnum
- Filled the Na values with the most common Merchnum corresponding to that Merch Description
- Assigned 'unknown' Merchnum for the transaction whose Merch Description is 'Retail Credit Adjustment' or 'Retail Debit Adjustment'
- Filled the rest with current max Merch Number plus 1

Merch state
- Mapped each Merch zip, Merchnum, and Merch description with a Merch State
- Filled the Na values with the most common Merch state corresponding to that Merch zip
- Filled the Na values with the most common Merch state corresponding to that Merchnum
- Filled the Na values with the most common Merch state corresponding to that Merch description
- Changed Merch state to foreign if it has a value and the value (not 'unknown') is not within the 53 states
- Filled the rest with 'unknown' value

Merch zip
- Mapped each Merchnum and Merch description with a Merch Zip
- Filled the Na values with the most common Merch zip corresponding to that Merch num
- Filled the Na values with the most common Merch zip corresponding to that Merch description
- Assigned 'unknown' Merch zip for the transaction whose Merch Description is 'Ret ail Credit Adjustment' or 'Retail Debit Adjustment'
- Filled the rest with 'unknown' value

4. Variable creation

Description of variables	# Variables created
Original variables: Original fields from the dataset excluding 'Recnum' and 'Fraud'	8
Data of Week Variables: Date of week target encoded (average fraud percentage/fraud risk of that day)	2

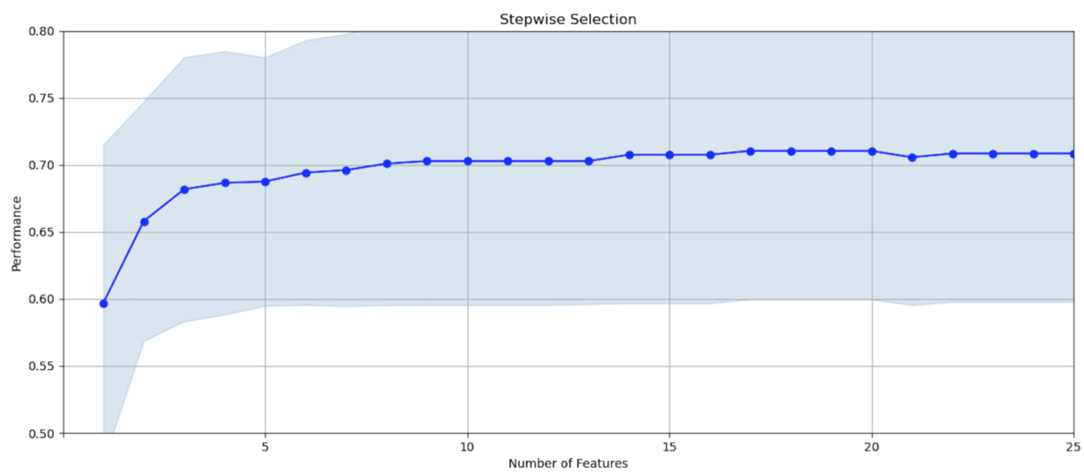
New entities: combining/concatenating/excerpting from different original fields	14
Days since Variables: # of days since the last transaction of that entity has been seen.	19
Frequency &Amount Variables: Count of each entity over the past {0,1,3,7,14,30,60} days; and Average, Maximum, Median, Total, Actual/Average, Actual/Maximum, Actual/Median, and Actual/Total amount for the same entity our the past {0,1,3,7,14,30,60} days	1197
Velocity Change Variables: # of transactions with one entity in the past {0,1} day divided by the average daily # of transitions with the same entity over the past {7,14,30,60} days across all the entities	152
Velocity Days-since Variables: For the past {0,1} day over past {7,14,30,60} days, velocity variables divided by day since variables across all the entities	152
Variability Variables: Average, Median, and Max amount difference between one record of one entity and the former records of the same entity over the past { 0,1,3,7,14,30,60} days across all the entities	399
Acceleration Variables: # of transactions with one entity in the past {0,1} day divided by the # of transitions with the same entity over the past {7,14,30,60} days over the power of days	152
Amount Bins: Divide Amount variable into 5 categories(1,2,3,4,5), where each category represents a quintile of amount	1
Original Total Variables	**2096**
Drop: Duplicated and frivolous variables such as ‘Date’, ‘Transtype’, and ‘Dow’	22
Total Variables	**2074**

Table 3

5. Feature selection

- The variable is then used for feature selection
- 6 models are tried:
  - Backward Selection (LGBM (n_estimators=10, num_leaves=4), num_filter=100, num_wrapper=25)
  - Forward Selection (Ramdom Foreast (n_estimators=5),num_filter=100, num_wrapper=25)
  - Forward Selection (LGBM (n_estimators=20, num_leaves=4), num_filter=100, num_wrapper=25)
  - Forward Selection (LGBM (n_estimators=20, num_leaves=4), num_filter=200, num_wrapper=25)
  - Forward Selection (LGBM (n_estimators=20, num_leaves=4), num_filter=300, num_wrapper=25)

- Forward Selection (LGBM (n_estimators=20, num_leaves=4), num_filter=400, num_wrapper=25)
- Comparing all the models I have tried, Forward LGBM with filter number of 300 and 400 both generate considerable high performance at about 0.71. I chose the one with 300 filters and 25 wrappers as the final feature selection model, because with less filter, it costs less time to achieve the same good result.



Graph 1

The list of final variables sorted by their univariate KS's score is attached below.

wrapper	variable	filter score
1	cardnum_merchnum_merchstate_total_14	0.630059
2	cardnum_zip3_max_14	0.629515
3	cardnum_merchnum_merchzip_avg_14	0.518122
4	cardnum_merchnum_merchdes_avg_7	0.519505
5	Merch_description_max_0	0.530588
6	cardnum_merchnum_avg_14	0.518386
7	cardnum_merchnum_merchdes_total_14	0.612649
8	Merch_zip_max_0	0.515098
9	cardnum_merchnum_avg_7	0.524281
10	cardnum_merchnum_merchstate_avg_7	0.524270
11	cardnum_merchnum_zip3_avg_14	0.518397
12	cardnum_merchnum_merchstate_avg_14	0.518365
13	cardnum_merchnum_merchdes_avg_14	0.515387
14	cardnum_merchnum_zip3_avg_7	0.524292
15	cardnum_merchnum_merchzip_avg_7	0.523337
16	cardnum_merchdes_avg_7	0.516608

17	cardnum_merchnum_avg_30	0.520958
18	cardnum_merchnum_merchzip_avg_30	0.521967
19	cardnum_merchnum_merchstate_avg_30	0.520947
20	cardnum_merchnum_zip3_avg_30	0.520926
21	merchnum_merchdes_max_0	0.530686
22	merchnum_zip3_max_0	0.533066
23	merchnum_merchstate_max_0	0.533034
24	Merchnum_max_0	0.533023
25	merchnum_merchzip_max_0	0.530032

* Note: LGBM forward selection, num_filter = 300, num_wrapper = 25

Table 4

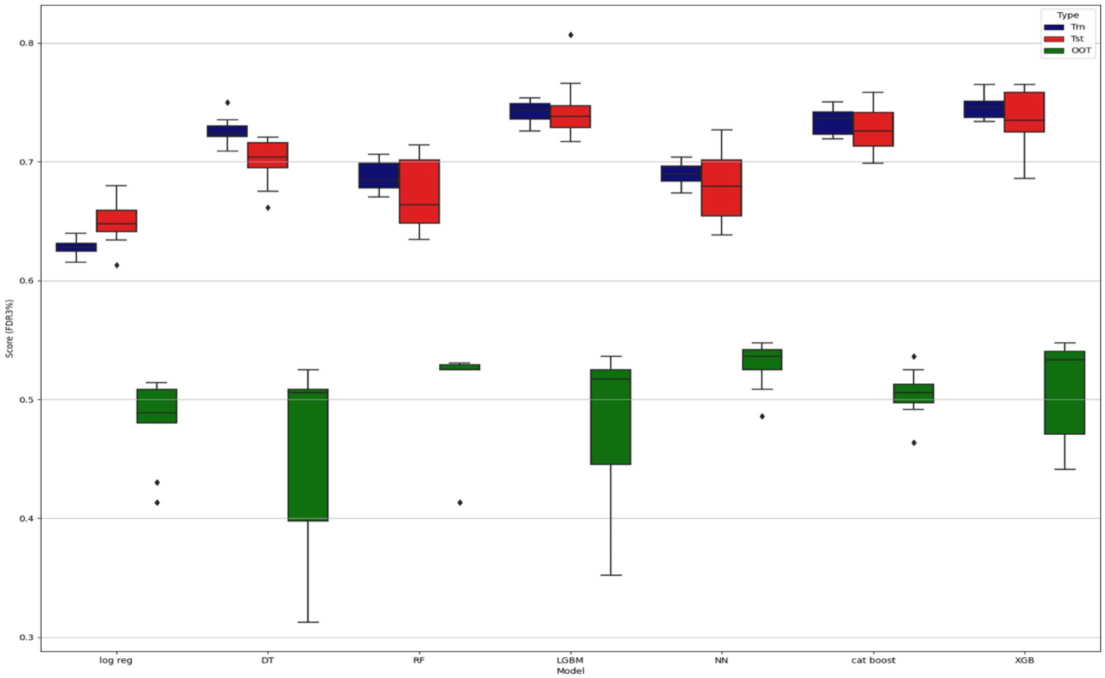
6. Preliminary model explores

After 25 best features were selected based on their ranked importance, multiple models including Logistic Regression, Decision Tree, Random Forest, Neural Network, and Boosted Tree (LGBM, CatBoost, XGBoost) were tested. Different combinations of hyperparameters associated with different models were tested, and the models' performances, which were set to be the average FDR@3%, were recorded as below.

Model			Parameter							Average FDR at 3%		
Logistic Regression	iteration	mber of Variab	penalty	C	solver		l1_ratio			Train	Test	OOT
	1	15	l2	1	lbfgs		0.4			0.640	0.626	0.495
	2	15	l1	1	liblinear		None			0.637	0.632	0.487
	3	15	l1	0.1	liblinear		None			0.632	0.644	0.452
	4	20	l1	1	saga		0.5			0.635	0.642	0.469
	5	20	l2	0.8	lbfgs		0.8			0.645	0.608	0.488
	6	20	l2	1	lbfgs		None			0.635	0.622	0.494
Decision Tree	iteration	mber of Variab	Criterion	max_features	min_samples_split	min_samples_leaf	splitter		max_depth	Train	Test	OOT
	1	15	gini	log2	100	20	random		5	0.338	0.334	0.185
	2	15	entropy	None	40	20	best		5	0.724	0.701	0.507
	3	15	entropy	None	40	5	best		10	0.852	0.766	0.322
	4	20	gini	None	40	35	best		5	0.708	0.678	0.502
	5	20	gini	None	50	25	best		5	0.695	0.674	0.461
	6	20	entropy	None	50	30	best		10	0.831	0.754	0.364
Random Forest	iteration	mber of Variab	n_estimators	criterion	max_depth	min_samples_split	min_samples_leaf		max_features	Train	Test	OOT
	1	15	300	gini	2	50	500		8	0.642	0.655	0.496
	2	15	200	gini	5	50	500		log2	0.644	0.642	0.453
	3	15	200	entropy	3	80	200		log2	0.687	0.692	0.527
	4	20	100	gini	15	20	200		log2	0.759	0.726	0.531
	5	20	200	entropy	15	50	400		None	0.666	0.659	0.441
	6	20	100	entropy	15	30	200		log2	0.755	0.737	0.531
LightGBM	iteration	mber of Variab	num_leaves	max_depth	learning_rate	boosting_type	n_estimators	min_child_samples	child_weight=0	Train	Test	OOT
	1	15	10	5	0.01	gbdt	50	10	0.001	0.744	0.734	0.474
	2	15	15	15	0.001	gbdt	50	20	0.001	0.734	0.707	0.356
	3	15	10	10	0.01	gbdt	50	20	0.002	0.740	0.759	0.521
	4	20	10	15	0.03	gbdt	50	10	0.001	0.801	0.769	0.515
	5	20	15	20	0.03	gbdt	200	10	0.001	0.880	0.792	0.378
	6	20	20	20	0.01	gbdt	80	25	0.001	0.872	0.777	0.432
Neural Network	iteration	mber of Variab	dden_layer_siz	activation	alpha	learning_rate	learning_rate_init		max_iter	Train	Test	OOT
	1	15	2	relu	0.0001	adaptive	0.001		200	0.579	0.571	0.405
	2	15	5,5	relu	0.0001	adaptive	0.001		200	0.680	0.670	0.510
	3	15	5,5,5	relu	0.001	constant	0.01		200	0.697	0.673	0.508
	4	20	5,5	relu	0.0001	adaptive	0.01		100	0.696	0.693	0.502
	5	20	5,5,5	identity	0.001	adaptive	0.001		200	0.630	0.617	0.407
	6	20	5,5	relu	0.001	constant	0.001		100	0.672	0.674	0.503
CatBoost	iteration	mber of Variab	bootstrap_type	max_depth	iterations	l2_leaf_reg	learning_rate		random_state	Train	Test	OOT
	1	15	Bayesian	6	1000	3	0.01		None	0.799	0.769	0.442
	2	15	Bayesian	6	500	12	0.01		None	0.737	0.727	0.510
	3	15	Bernoulli	6	500	3	0.01		None	0.750	0.714	0.474
	4	20	Bayesian	5	500	5	0.01		None	0.727	0.731	0.436
	5	20	MVS	7	500	12	0.02		4	0.863	0.793	0.425
	6	20	Bayesian	7	500	15	0.01		4	0.737	0.738	0.503
XGBoost	iteration	mber of Variab	booster	max_depth	tree_method	min_child_weight	colsample_bytree		n_estimator	Train	Test	OOT
	1	15	gbtree	6	approx	1	1		100	0.945	0.847	0.412
	2	15	gbtree	6	exact	100	1		80	0.743	0.726	0.508
	3	15	dart	6	auto	100	0.8		80	0.744	0.725	0.507
	4	20	gbtree	6	approx	100	1		100	0.733	0.703	0.484
	5	20	gbtree	7	auto	100	0.8		100	0.743	0.724	0.503
	6	20	dart	7	auto	200	1		300	0.653	0.652	0.385

* Note: The line in Yellow is the final model, which achieves a high OOT of 0.531

Table 5



Graph 2

7. Final model performance

The final model chosen is the Random Forest, with top 20 features, n_estimator=100, criterion=gini, max_depth=15, min_samples_leaf=200, and max_deatures=log2. The model could achieve a Training accuracy of 0.759, a Testing accuracy rate of 0.726, and a OOT of 0.531. Tables below shows the performance of training, testing, and OOT respectively. The model can capture 53.1% of all the fraud at the top 3 percent.

TRN	# Records	# Goods	# Bads	Fraud Rate						
	59,010	58,395	615	0.0104						
	Bin Statistics				Cumulative Statistcis					
Population Bin %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Cumulative Goods	% Cumulative Bads(FDR)	KS	FDR
1	590	278	312	47.12%	52.88%	590	278	312	0.48%	50.73%	50.26	0.89
2	590	495	95	83.90%	16.10%	1180	773	407	1.32%	66.18%	64.86	1.90
3	590	528	62	89.49%	10.51%	1770	1301	469	2.23%	76.26%	74.03	2.77
4	590	569	21	96.44%	3.56%	2360	1870	490	3.20%	79.67%	76.47	3.82
5	590	558	32	94.58%	5.42%	2950	2428	522	4.16%	84.88%	80.72	4.65
6	591	572	19	96.79%	3.21%	3541	3000	541	5.14%	87.97%	82.83	5.55
7	590	583	7	98.81%	1.19%	4131	3583	548	6.14%	89.11%	82.97	6.54
8	590	582	8	98.64%	1.36%	4721	4165	556	7.13%	90.41%	83.27	7.49
9	590	587	3	99.49%	0.51%	5311	4752	559	8.14%	90.89%	82.76	8.50
10	590	583	7	98.81%	1.19%	5901	5335	566	9.14%	92.03%	82.90	9.43
11	590	586	4	99.32%	0.68%	6491	5921	570	10.14%	92.68%	82.54	10.39
12	590	584	6	98.98%	1.02%	7081	6505	576	11.14%	93.66%	82.52	11.29
13	590	585	5	99.15%	0.85%	7671	7090	581	12.14%	94.47%	82.33	12.20
14	590	588	2	99.66%	0.34%	8261	7678	583	13.15%	94.80%	81.65	13.17
15	591	587	4	99.32%	0.68%	8852	8265	587	14.15%	95.45%	81.29	14.08
16	590	588	2	99.66%	0.34%	9442	8853	589	15.16%	95.77%	80.61	15.03
17	590	582	8	98.64%	1.36%	10032	9435	597	16.16%	97.07%	80.92	15.80
18	590	590	0	100.00%	0.00%	10622	10025	597	17.17%	97.07%	79.91	16.79
19	590	587	3	99.49%	0.51%	11212	10612	600	18.17%	97.56%	79.39	17.69
20	590	588	2	99.66%	0.34%	11802	11200	602	19.18%	97.89%	78.71	18.60

Table 6

TST	# Records	# Goods	# Bads	Fraud Rate								
	25,290	25,025	265	0.0105								
	Bin Statistics					Cumulative Statistcis						
Population Bin %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Cumulative Goods	% Cumulative Bads(FDR)	KS	FDR
1	253	121	132	47.83%	52.17%	253	121	132	0.48%	49.81%	49.33	0.92
2	253	217	36	85.77%	14.23%	506	338	168	1.35%	63.40%	62.05	2.01
3	253	226	27	89.33%	10.67%	759	564	195	2.25%	73.58%	71.33	2.89
4	253	242	11	95.65%	4.35%	1012	806	206	3.22%	77.74%	74.52	3.91
5	252	243	9	96.43%	3.57%	1264	1049	215	4.19%	81.13%	76.94	4.88
6	253	250	3	98.81%	1.19%	1517	1299	218	5.19%	82.26%	77.07	5.96
7	253	253	0	100.00%	0.00%	1770	1552	218	6.20%	82.26%	76.06	7.12
8	253	250	3	98.81%	1.19%	2023	1802	221	7.20%	83.40%	76.20	8.15
9	253	250	3	98.81%	1.19%	2276	2052	224	8.20%	84.53%	76.33	9.16
10	253	250	3	98.81%	1.19%	2529	2302	227	9.20%	85.66%	76.46	10.14
11	253	251	2	99.21%	0.79%	2782	2553	229	10.20%	86.42%	76.21	11.15
12	253	251	2	99.21%	0.79%	3035	2804	231	11.20%	87.17%	75.97	12.14
13	253	250	3	98.81%	1.19%	3288	3054	234	12.20%	88.30%	76.10	13.05
14	253	252	1	99.60%	0.40%	3541	3306	235	13.21%	88.68%	75.47	14.07
15	253	251	2	99.21%	0.79%	3794	3557	237	14.21%	89.43%	75.22	15.01
16	252	252	0	100.00%	0.00%	4046	3809	237	15.22%	89.43%	74.21	16.07
17	253	252	1	99.60%	0.40%	4299	4061	238	16.23%	89.81%	73.58	17.06
18	253	252	1	99.60%	0.40%	4552	4313	239	17.23%	90.19%	72.95	18.05
19	253	251	2	99.21%	0.79%	4805	4564	241	18.24%	90.94%	72.71	18.94
20	253	252	1	99.60%	0.40%	5058	4816	242	19.24%	91.32%	72.08	19.90

Table 7

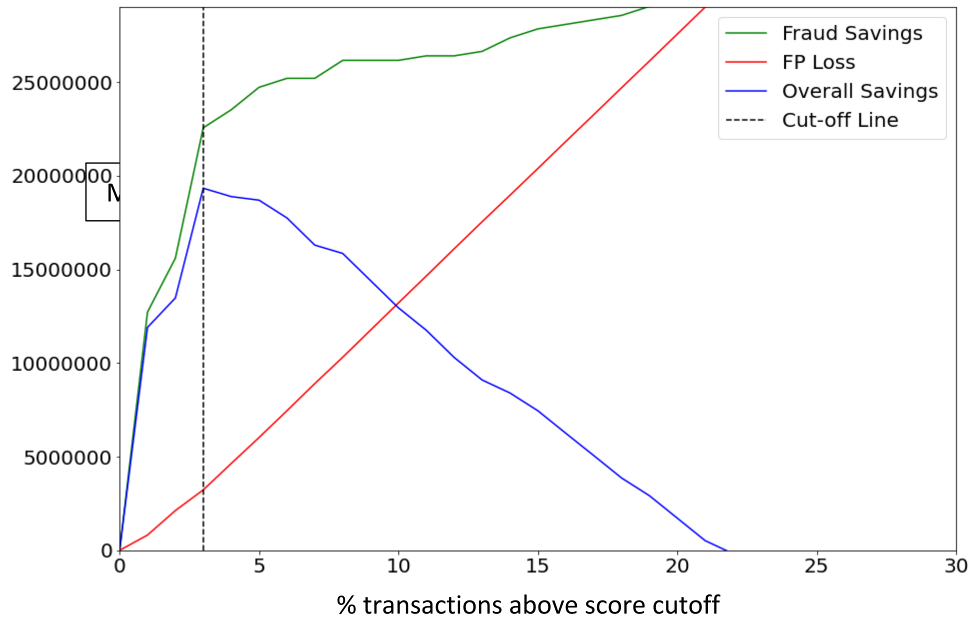
OOT	# Records	# Goods	# Bads	Fraud Rate								
	12,097	11,918	179	0.0148								
	Bin Statistics					Cumulative Statistcis						
Population Bin %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Cumulative Goods	% Cumulative Bads(FDR)	KS	FDR
1	121	68	53	56.20%	43.80%	121	68	53	0.57%	29.61%	29.04	1.28
2	121	109	12	90.08%	9.92%	242	177	65	1.49%	36.31%	34.83	2.72
3	121	92	29	76.03%	23.97%	363	269	94	2.26%	52.51%	50.26	2.86
4	121	117	4	96.69%	3.31%	484	386	98	3.24%	54.75%	51.51	3.94
5	121	116	5	95.87%	4.13%	605	502	103	4.21%	57.54%	53.33	4.87
6	121	119	2	98.35%	1.65%	726	621	105	5.21%	58.66%	53.45	5.91
7	121	121	0	100.00%	0.00%	847	742	105	6.23%	58.66%	52.43	7.07
8	121	117	4	96.69%	3.31%	968	859	109	7.21%	60.89%	53.69	7.88
9	121	121	0	100.00%	0.00%	1089	980	109	8.22%	60.89%	52.67	8.99
10	121	121	0	100.00%	0.00%	1210	1101	109	9.24%	60.89%	51.66	10.10
11	121	120	1	99.17%	0.83%	1331	1221	110	10.25%	61.45%	51.21	11.10
12	121	121	0	100.00%	0.00%	1452	1342	110	11.26%	61.45%	50.19	12.20
13	121	120	1	99.17%	0.83%	1573	1462	111	12.27%	62.01%	49.74	13.17
14	121	118	3	97.52%	2.48%	1694	1580	114	13.26%	63.69%	50.43	13.86
15	121	119	2	98.35%	1.65%	1815	1699	116	14.26%	64.80%	50.55	14.65
16	121	120	1	99.17%	0.83%	1936	1819	117	15.26%	65.36%	50.10	15.55
17	120	119	1	99.17%	0.83%	2056	1938	118	16.26%	65.92%	49.66	16.42
18	121	120	1	99.17%	0.83%	2177	2058	119	17.27%	66.48%	49.21	17.29
19	121	119	2	98.35%	1.65%	2298	2177	121	18.27%	67.60%	49.33	17.99
20	121	120	1	99.17%	0.83%	2419	2297	122	19.27%	68.16%	48.88	18.83

Table 8

8. Financial curves and recommended cutoff

- Assume \$400 gain for every fraud that's caught (green curve).
- Assume \$20 loss for every false positive (a good that's flagged as a bad) (red).
- Assume we got a sample of 100,000 records from a portfolio of 10 million accounts. Multiply the oot \$'s by (12/2)*(10,000,000/100,000).

- According to the curve, we recommend a score cutoff at 3%, where we achieve a maximum overall savings at 19,332,000.



Graph 3

9. Summary

The goal of this project is to build a fraud detection model to achieve a relatively high accuracy of the Card Transaction Data. The whole process goes through the steps of data description, data cleaning, variable creation, feature selection, model exploration, final model performance displacement, and finally a recommendation of cutoff based on some financial curves.

After doing all the steps above, a final model of Random Forest is chosen, with top 20 features, n_estimator=100, criterion=gini, max_depth=15, min_samples_leaf=200, and max_deatures=log2. The model could achieve a Training accuracy of 0.759, a Testing accuracy rate of 0.726, and a OOT of 0.531. The model can capture 53.1% of all the fraud at the top 3 percent. According to the financial curve, we recommend a score cutoff at 3%, where we achieve a maximum overall savings at 19,332,000.

Others we could do is to try more sophisticated models to improve the accuracy of fraud detection, during which, automatic algorisms might be implemented to get through all possible combinations of hyperparameters to find the best models. In addition, More variables could be built during the process to monitor the fraud, some of the examples of such variables are the frequency of the transactions happened in gas station or online.

10. Appendix

This data is about Card Transaction Data. It is a collection of card transaction records from a US government organization in year 2010. The data has 96,753 rows and 10 fields.

(1) Numerical Table

Field Name	% Populated	Min	Max	Mean	Stdev	% Zero
Date	100.00	01/01/2010	12/31/2010	/	/	0.00
Amount	100.00	0.01	3,102,045.53	427.89	10,006.14	0.00

Table 9

(2) Categorical Table

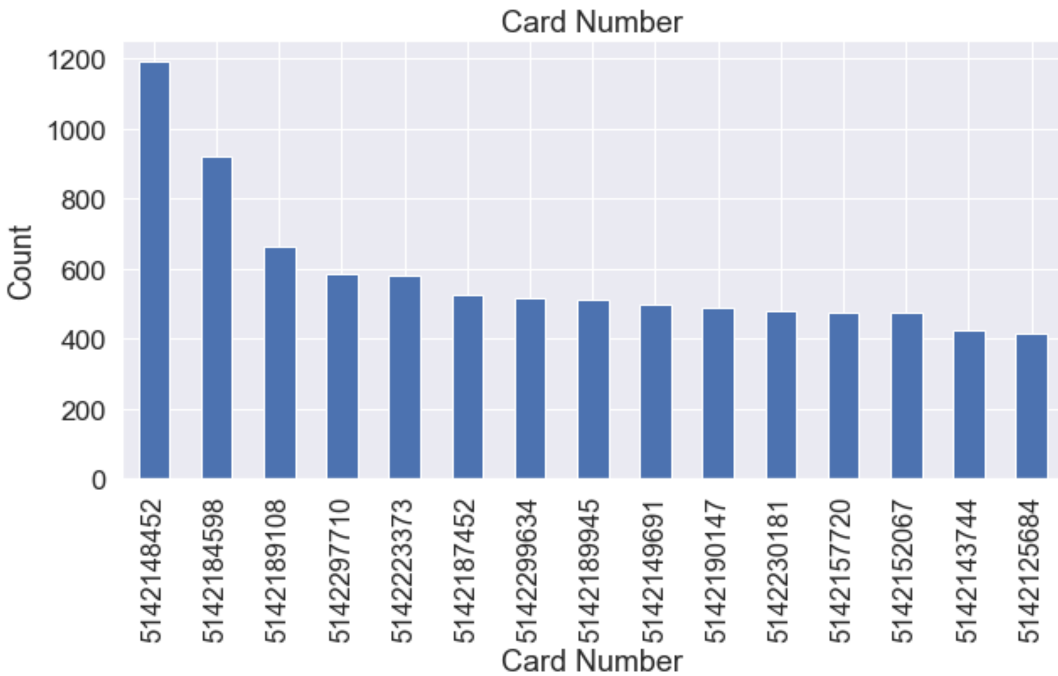
Field Name	% Population	# Unique Values	Most Common Value
Recnum	100.00	96,753	N/A
Cardnum	100.00	1,645	5142148452
Merchnum	96.51	13,092	930090121224
Merch description	100.00	13,126	GSA-FSS-ADV
Merch state	98.76	228	TN
Merch zip	95.19	4,568	38118
Transtype	100.00	4	P
Fraud	100.00	2	0

Table 10

(1) Field Name: Recnum
Description: Record Field. Ordinal unique positive integer for each transaction, from 1 to 96,753.

(2) Field Name: Cardnum

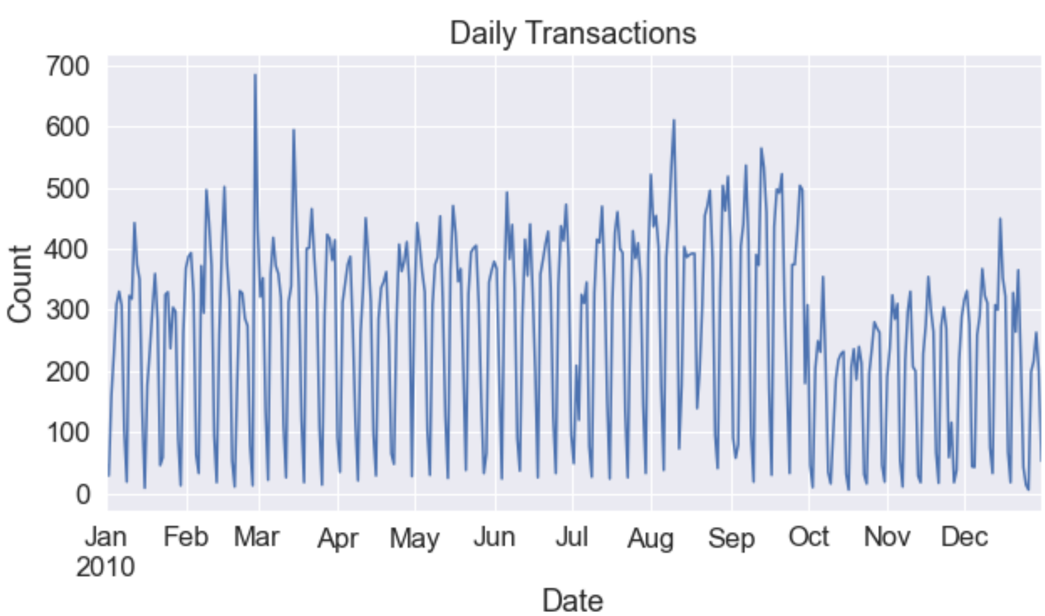
Description: Card Number Field. The card number for each transaction. The graph below indicates the count of top 15 card numbers. The most common card number is 5142148452, the count of which is 1,192.



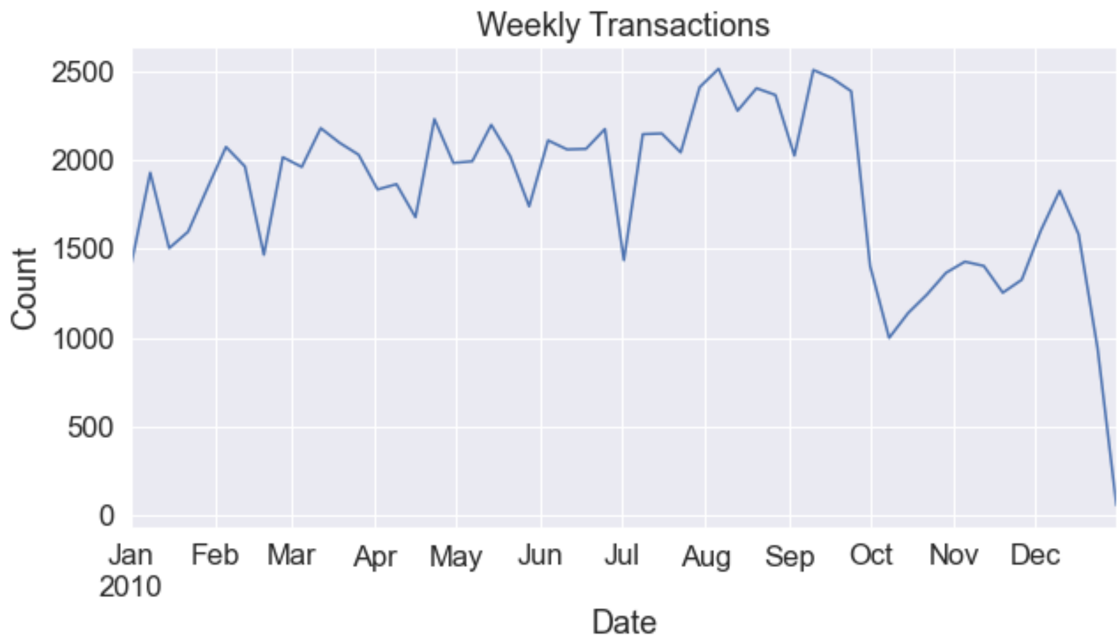
Graph 4

(3) Field Name: Date

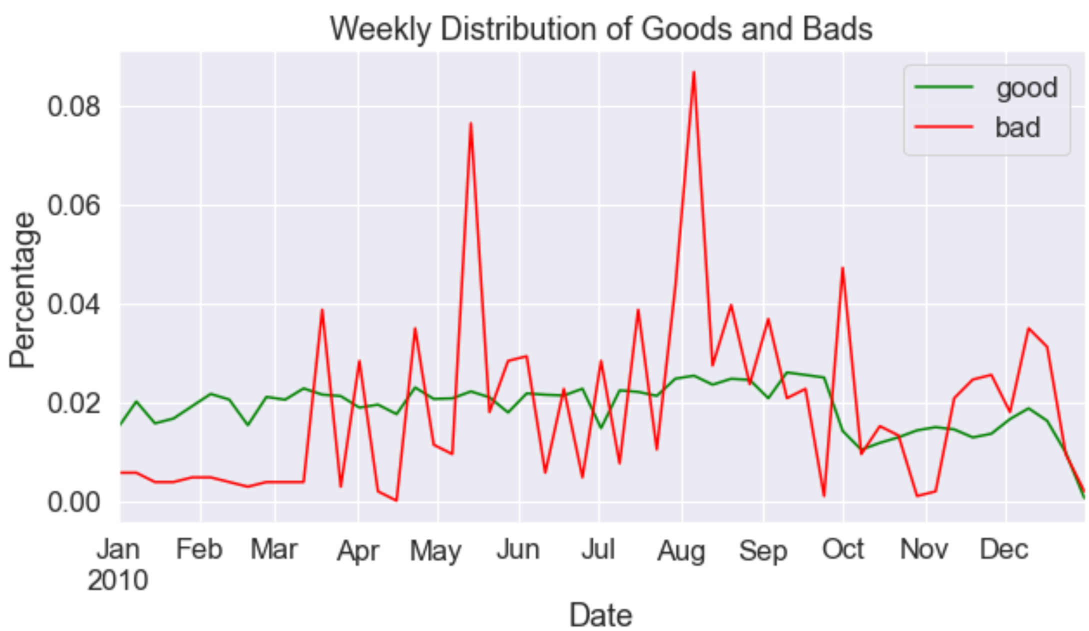
Description: Date Field. The graphs below display daily and weekly transaction distributions. For each week, the transaction peaked at weekend; From the year base, the transaction at first decreased tremendously in October and then exhibited another decline at the end of December. According to the graph showing weekly distribution of goods and bads, the number of goods was comparably stable, while the number of bads was fluctuated over weeks. The number of bads peaked in May and August respectively.



Graph 5



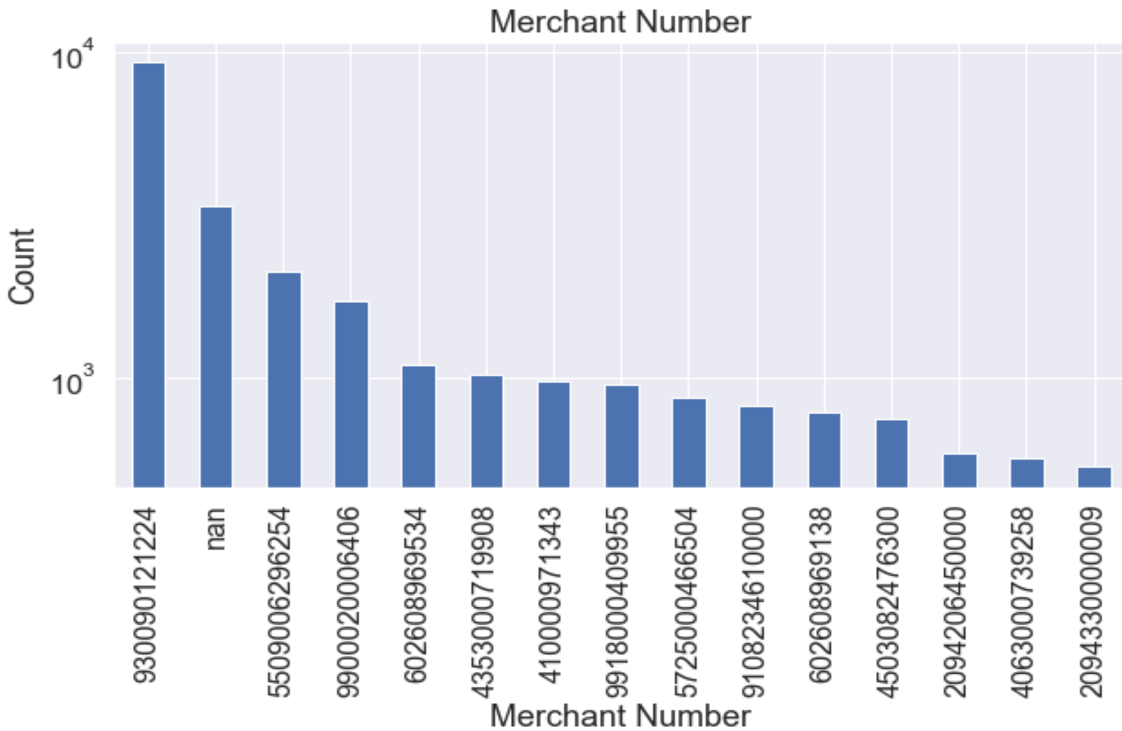
Graph 6



Graph 7

(4) Field Name: Merchnum

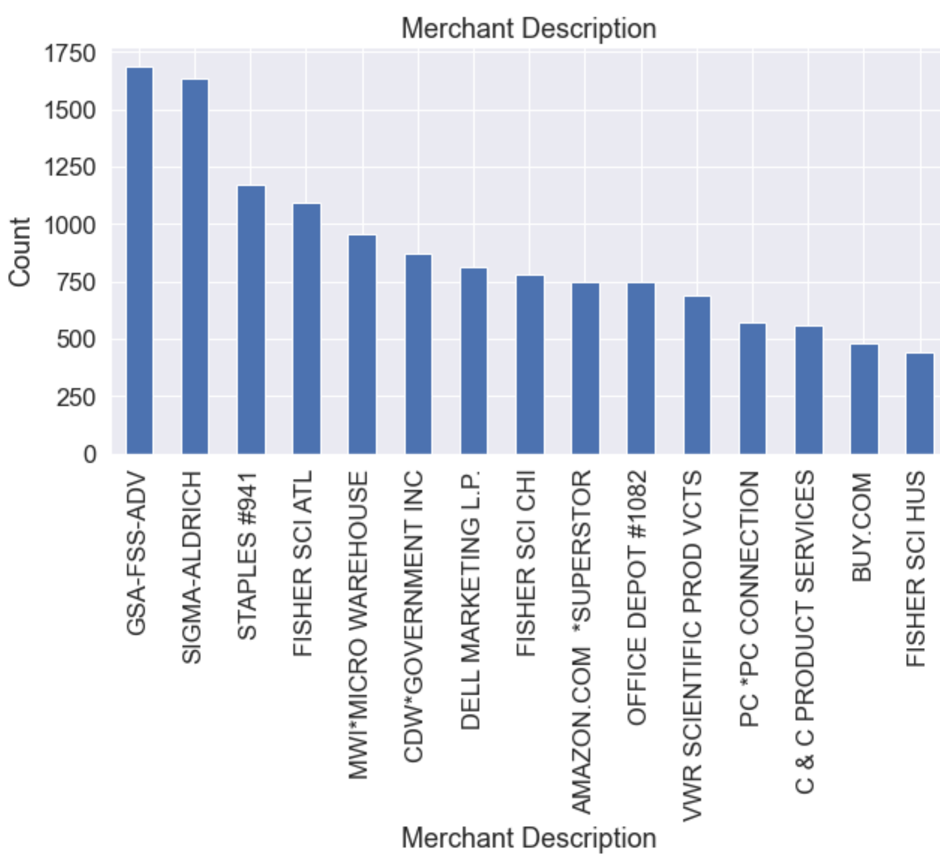
Description: Merchant Number Field. The graph below indicates the count of top 15 merchant numbers. The most common merchant number is 930090121224, the count of which is 9,310.



Graph 8

(5) Field Name: Merch description.

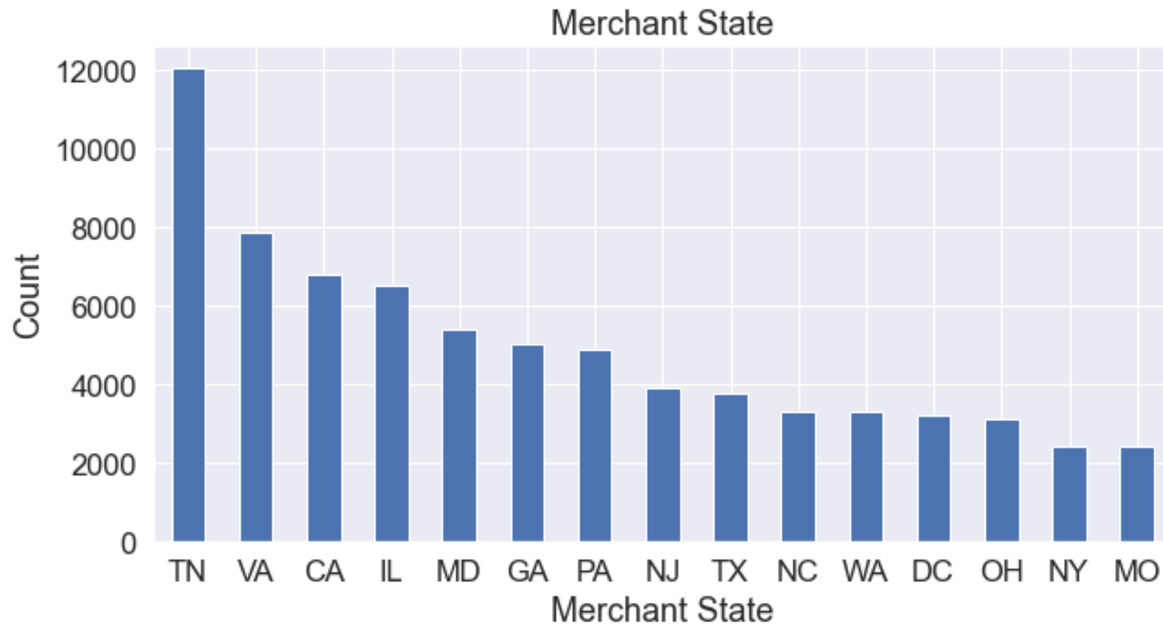
Description: Merchant Description Field. The graph below indicates the count of top 15 merchant descriptions. The most common merchant number is GSA-FSS-ADV, the count of which is 1,688.



Graph 9

(6) Field Name: Merch state

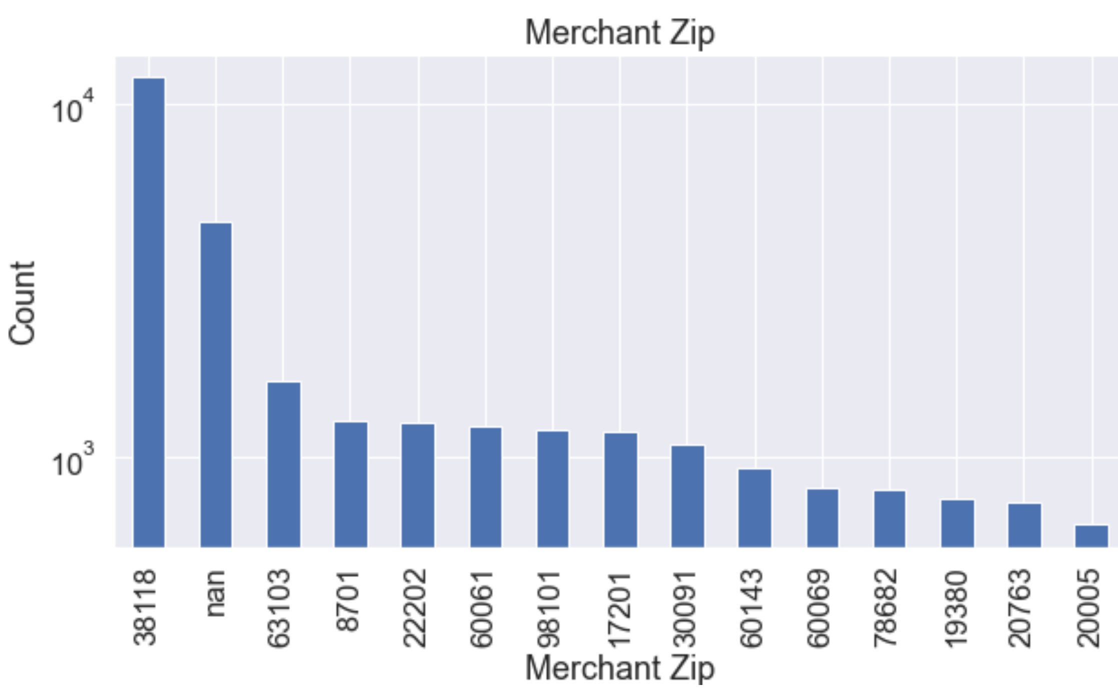
Description: Merchant Sate Field. The graph below indicates the count of top 15 merchant states. The most common merchant state is Tennessee(TN), the count of which is 12,035.



Graph 10

(7) Field Name: Merch zip

Description: Merchant Zip Code Field. The graph below indicates the count of top 15 merchant zip codes. The most common merchant zip code is 38118, the count of which is 11,868.



Graph 11

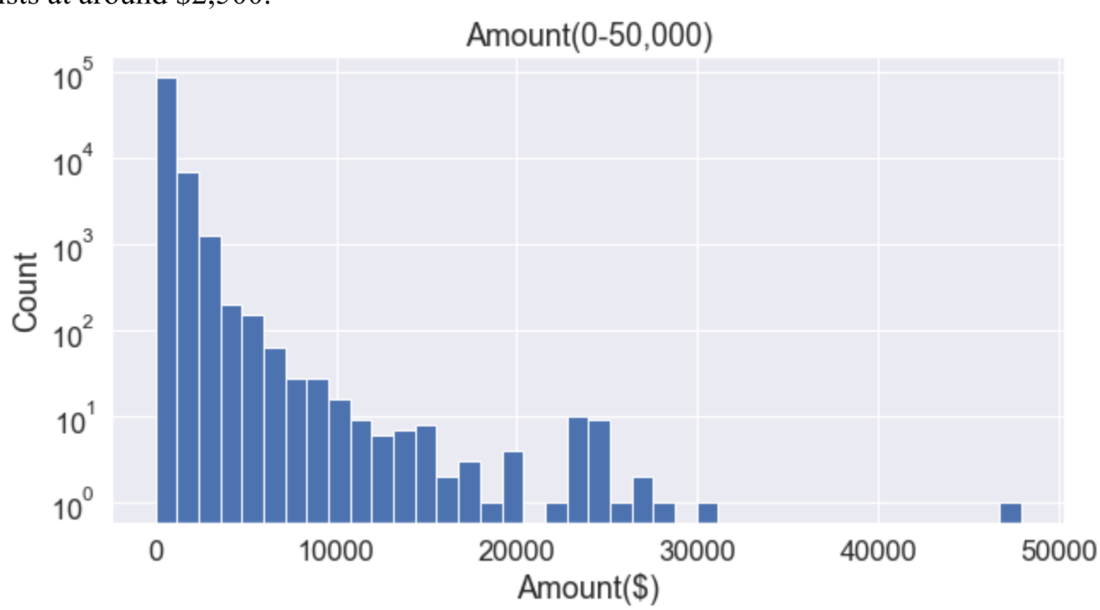
(8) Field Name: Transtype

Description: Transaction Type Field. The graph below indicates the count of each transition type, including P, A, D, and Y. The most common transaction type is P, standing for purchase, the count of which is 96,398.

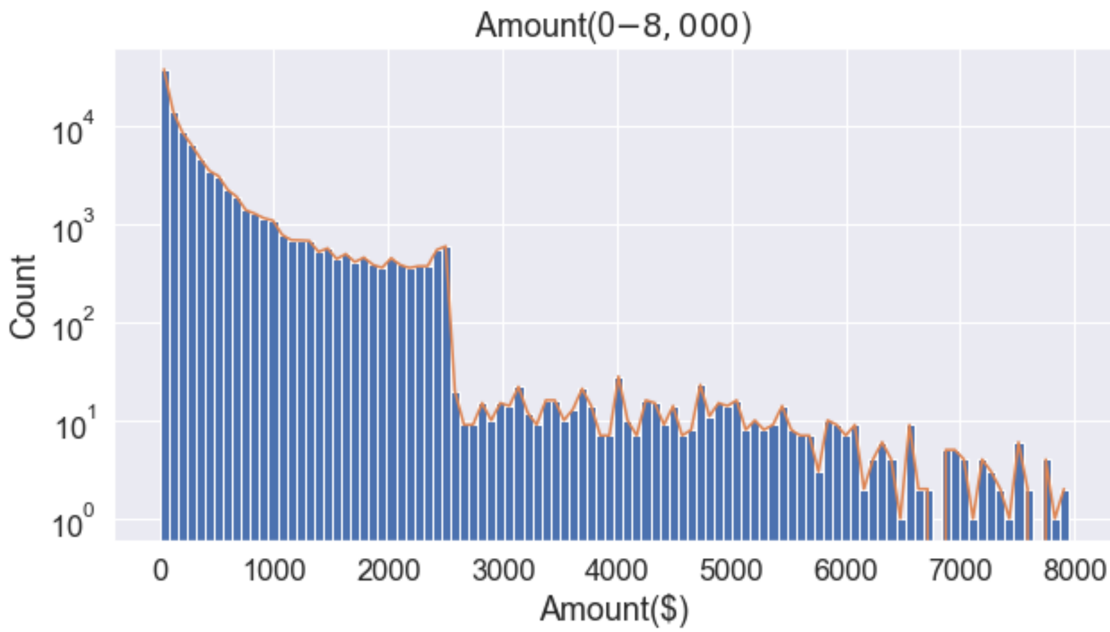
Graph 12

(9) Field Name: Amount

Description: Transaction Amount Field. It ranges from $0.01 to $3,102,045.53. There is a unique value of $3,102,045.53on 2010-07-15. After excluding the extreme value, the distribution of Amounts is showed in the histogram below. Most of the transaction amounts are between $0 and $30000. In general, the transaction frequency decreases as transaction amount increases, while there is a bump at around $25,000. Then, zoom in to the range between $0 and $8,000, a significant drop exists at around $2,500.



Graph 13



Graph 14

(10) Field Name: Fraud. Description: Fraud Label Field. A binary variable with Fraud =0(no fraud) and Fraud =1(fraud). The count of Fraud =0 is 95,6

Graph 15