

Data Quality Report - Initial Findings

- Descriptive statistics for all the continuous features:

	count	mean	std	min	25%	50%	75%	max
ExternalRiskEstimate	956.0	72.637029	10.530387	-9.0	65.0	73.0	81.00	94.0
MSinceOldestTradeOpen	956.0	191.313808	104.487229	-9.0	122.0	179.0	248.00	803.0
MSinceMostRecentTradeOpen	956.0	9.904812	13.686556	-9.0	3.0	6.0	12.00	152.0
AverageMInFile	956.0	78.495816	34.420391	-9.0	55.0	76.0	97.25	245.0
NumSatisfactoryTrades	956.0	21.172594	12.028262	-9.0	12.0	20.0	28.00	78.0
NumTrades60Ever2DerogPubRec	956.0	0.528243	1.236545	-9.0	0.0	0.0	1.00	11.0
NumTrades90Ever2DerogPubRec	956.0	0.336820	1.010699	-9.0	0.0	0.0	0.00	10.0
PercentTradesNeverDelq	956.0	92.755230	12.452327	-9.0	90.0	99.5	100.00	100.0
MSinceMostRecentDelq	956.0	8.028243	21.777414	-9.0	-7.0	-7.0	15.00	83.0
NumTotalTrades	956.0	22.496862	13.675713	-9.0	13.0	21.0	30.00	100.0
NumTradesOpeninLast12M	956.0	1.959205	2.082942	-9.0	0.0	1.5	3.00	19.0
PercentInstallTrades	956.0	34.169456	18.254872	-9.0	20.0	33.0	45.00	100.0
MSinceMostRecentInqexcl7days	956.0	-0.171548	5.802677	-9.0	-7.0	0.0	1.00	24.0
NumInqLast6M	956.0	1.471757	2.330817	-9.0	0.0	1.0	2.00	29.0
NumInqLast6Mexcl7days	956.0	1.413180	2.293125	-9.0	0.0	1.0	2.00	29.0
NetFractionRevolvingBurden	956.0	32.882845	28.844825	-9.0	6.0	27.0	55.00	115.0
NetFractionInstallBurden	956.0	39.214435	41.709981	-9.0	-8.0	46.0	79.00	132.0
NumRevolvingTradesWBalance	956.0	3.889121	3.470481	-9.0	2.0	3.0	5.00	26.0
NumInstallTradesWBalance	956.0	1.550209	3.416711	-9.0	1.0	2.0	3.00	19.0
NumBank2NatlTradesWHighUtilization	956.0	0.478033	2.686507	-9.0	0.0	0.0	1.00	12.0
PercentTradesWBalance	956.0	66.132845	22.226538	-9.0	50.0	67.0	83.00	100.0

All of the continuous features' minimum Numbers that are -9, which means No Bureau Record or No Investigation. This shows that there is a large part of data that is not recorded by the credit scoring company.

The average external risk estimate is between 25% and 50%. Months Since Oldest Trade Open has a maximum of around 800 and a standard deviation of 104, the number of it from 25% till 75%'s difference is only around 200, but from the last 25% to max the difference is around 600, so I think the max value is an invalid value, but the value still indicating that most customers don't Trade for years.

From the most recent month in delinquency , up to 50% of customers' values are -7, which means not Met (e.g. No entries, No refinements), so these values should set to null in order to have influence to standard and mean. In addition, the Max number of indicating someone delinquent up to 80 months. More than 25% of the customers never delinquent.

There are very few people with more than 60 trades, as many as half of them, and 75% of them don't have more than 90 trades. The number of trade open in the past 12 months, the inquiries in last 6 months, the inquiries Last 6 Months excluding 7days and the Number Bank/National Trades with high utilization ratio have a lot of zero values, as many as 25% of the data. Their maximum values are also around 10 to 20, which seems like outliers. The number of installment trades and the trade open in the past 12 month shows in the last year, few people used credit cards, and even when they did, nearly half of them paid by installments. Even though the account had been open for a long time, there had been very little use of credit cards in the last 12 months, with an average of 1.96.

From the number satisfactory trades, percentage trade never delinquent, shows than most recent customers have paid their bills on time, which means they have good customer service credit.

- **Descriptive statistics for all the categorical features:**

	count	mean	std	min	25%	50%	75%	max
MaxDelq2PublicRecLast12M	956.0	5.867155	1.689552	-9.0	6.0	6.0	7.0	7.0
MaxDelqEver	956.0	6.414226	1.992122	-9.0	6.0	7.0	8.0	8.0

The feature of Max Delinquency/Public Records Last 12 Months shows more than 25% people is 7, which means current and never delinquent in the past 12 months. In addition, "MaxDelqEver" ranges from 75% to a maximum of 8, this proves that more than 25% people current or never delinquent.

- **Histogram for continuous features[plots attached at end of file]**

The consolidated version of risk markers shows to be normally distributed with the center about 65.

The months since oldest trade open, most recent inquiries excluding 7 days, most recent trade open, and the number of bank/national trades with high utilization ratio, inquiries last 6 months, inquiries of last 6 months excluding 7 days, install trades with balance, revolving trades with balance, trades more than 60 and 90 ever, trades open in last 12 months are all appear to be exponentially decreasing.

The average months in File, the months since most recent trade open, the number satisfactory trades, the total trades (total number of credit accounts), the percent install trades percent trades that never delinquent appear skewed right distribution.

Net Fraction Install Burden shows linear increase tendency. This is install balance divided by original loan amount.

Net Fraction Revolving Burden have a linear decline trend. This is revolving balance divided by credit limit.

- **Box Plots for Continuous Features [plots attached at end of file]**

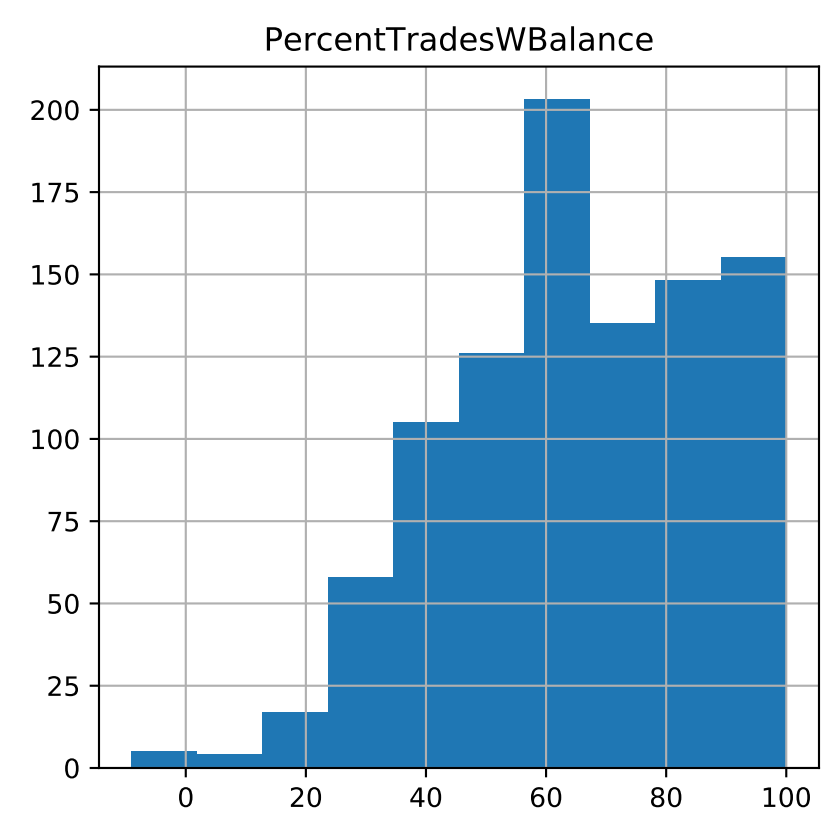
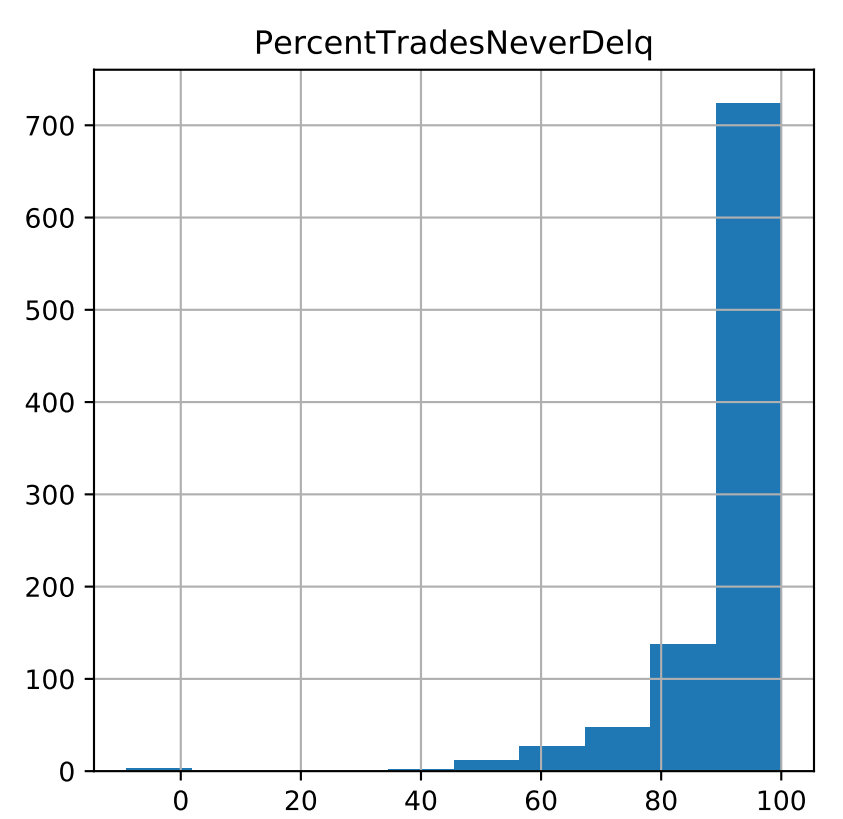
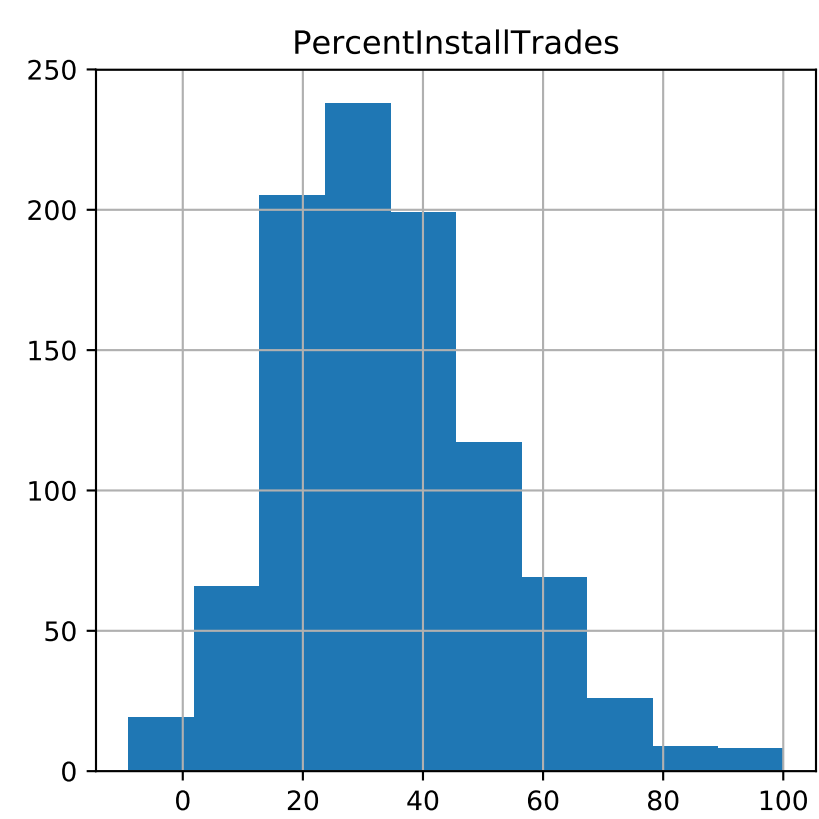
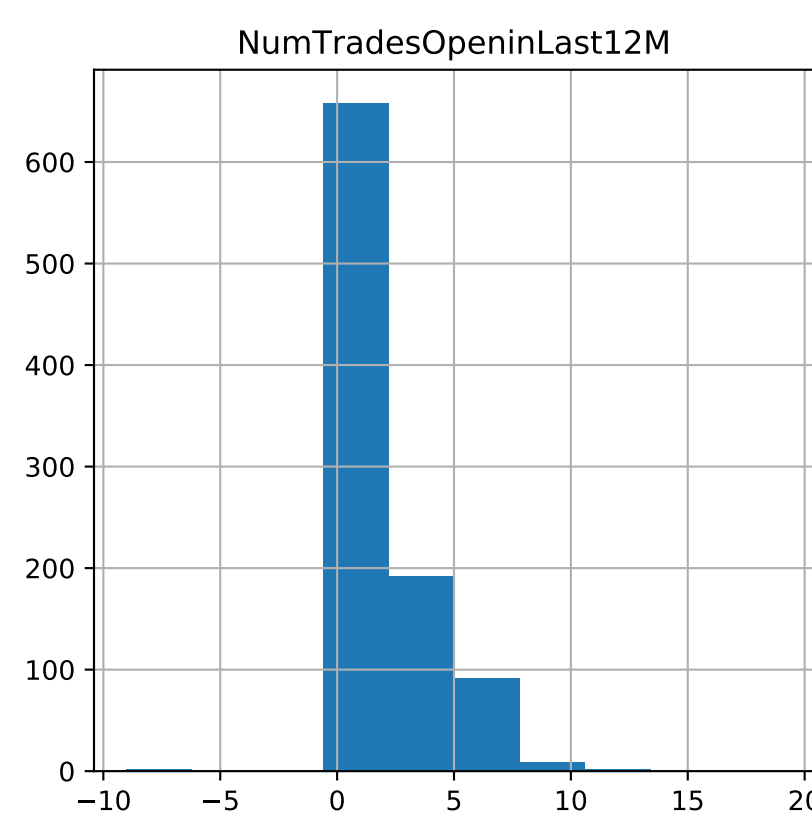
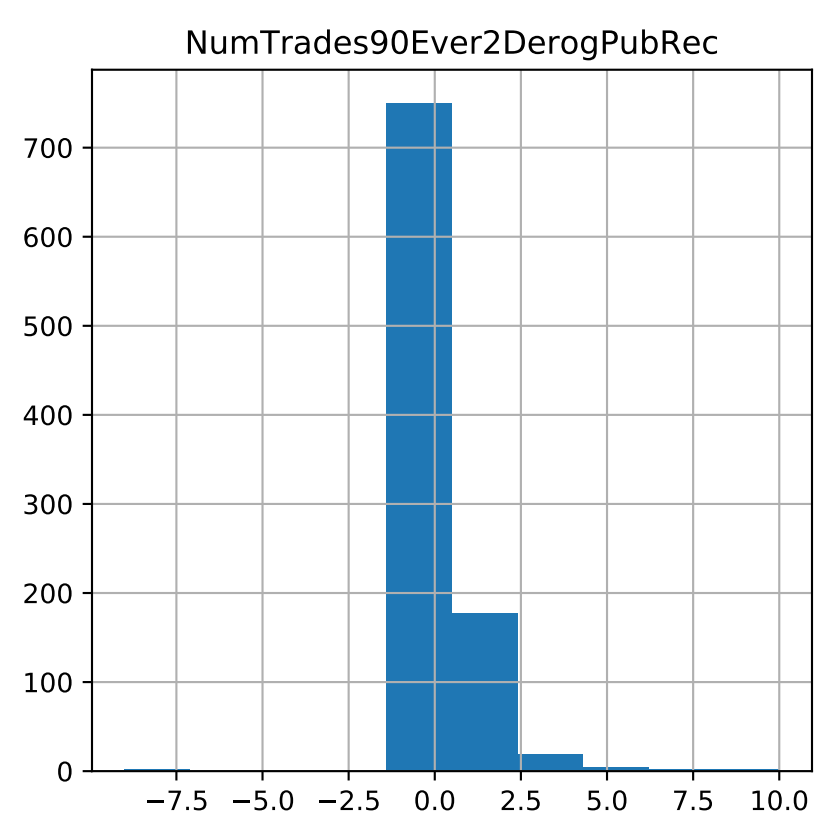
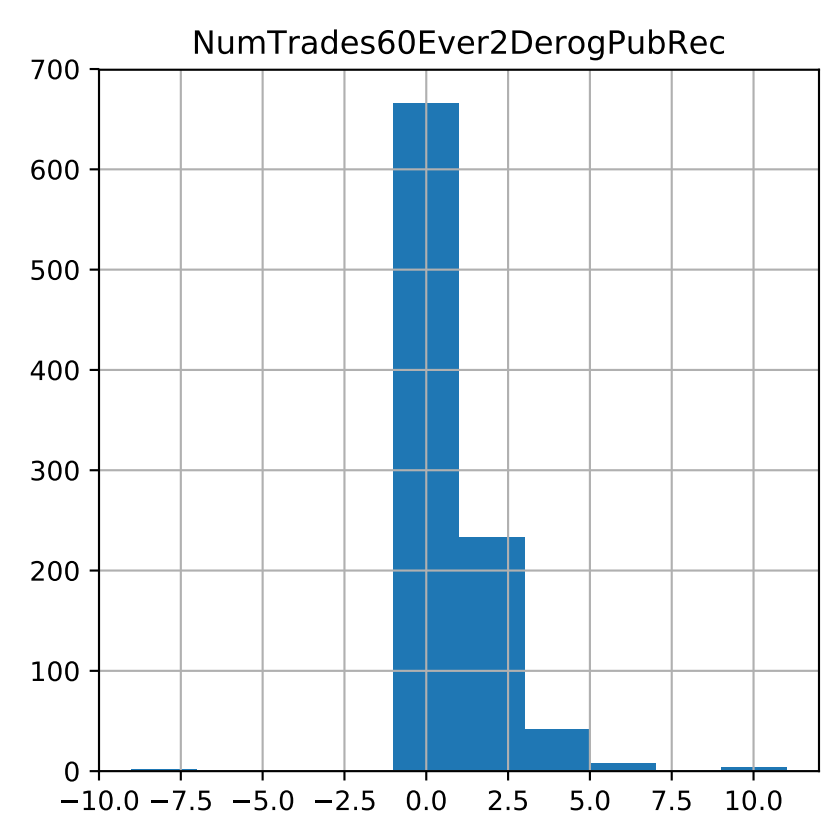
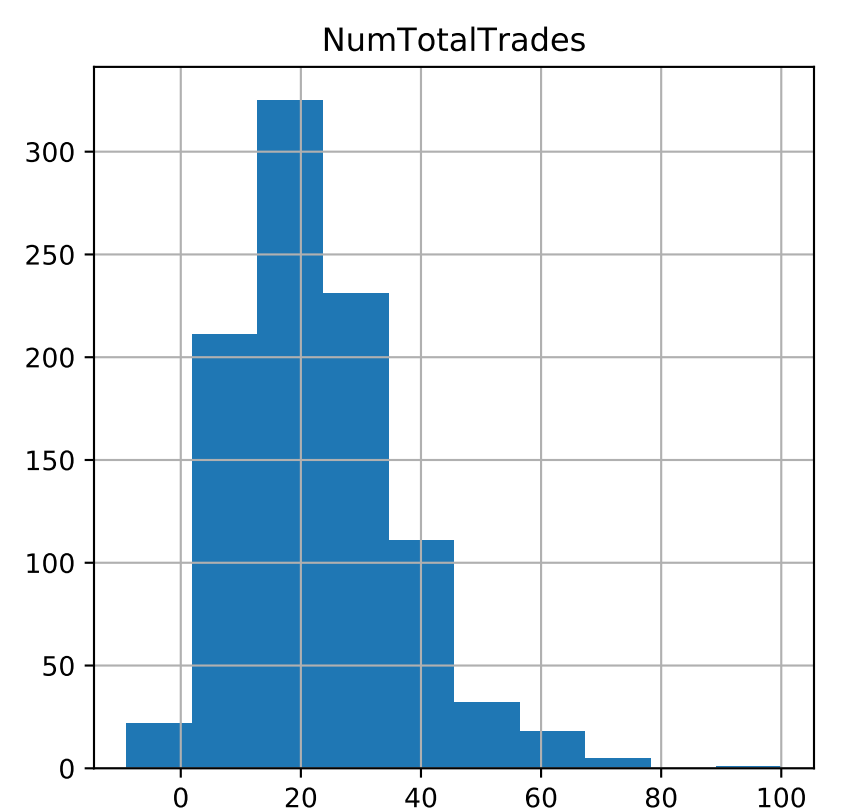
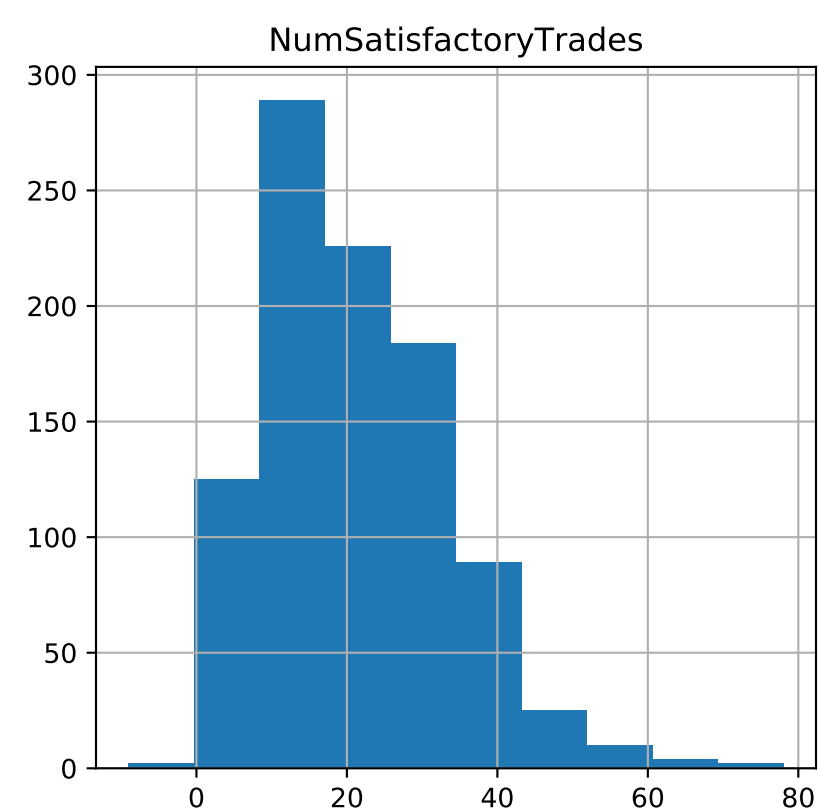
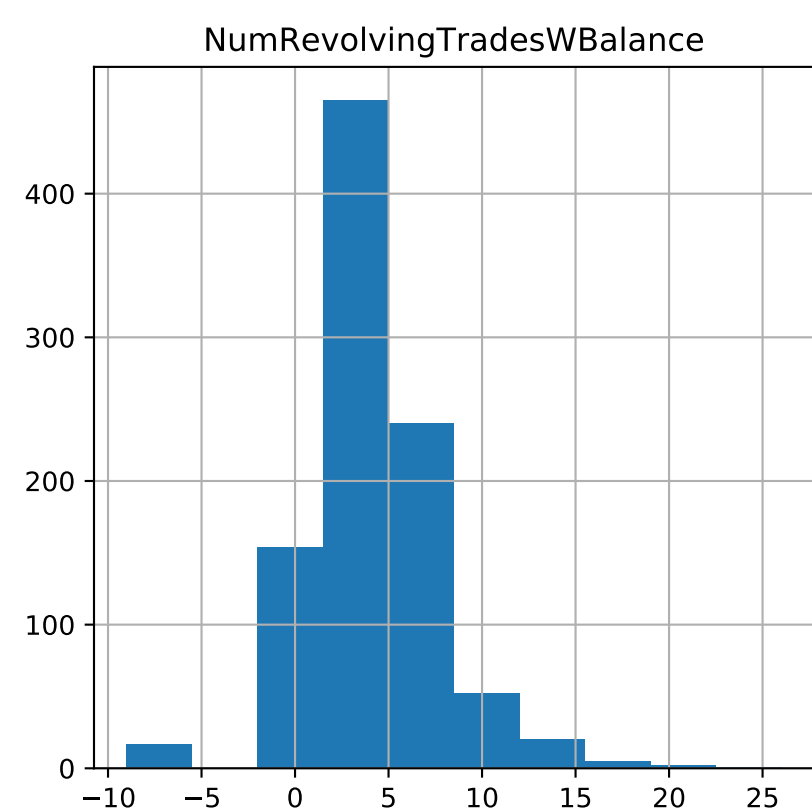
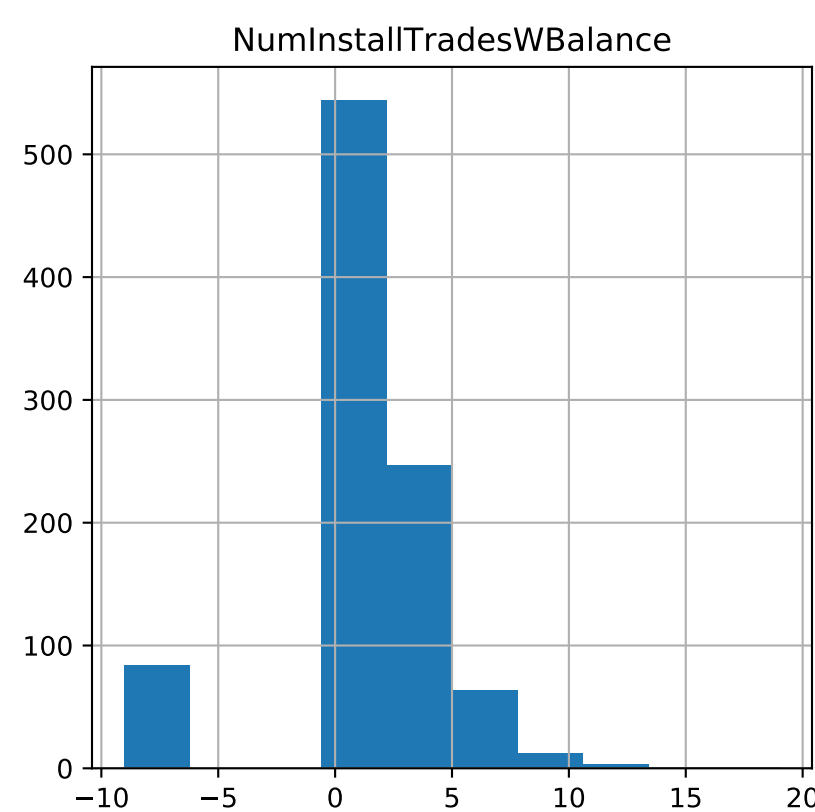
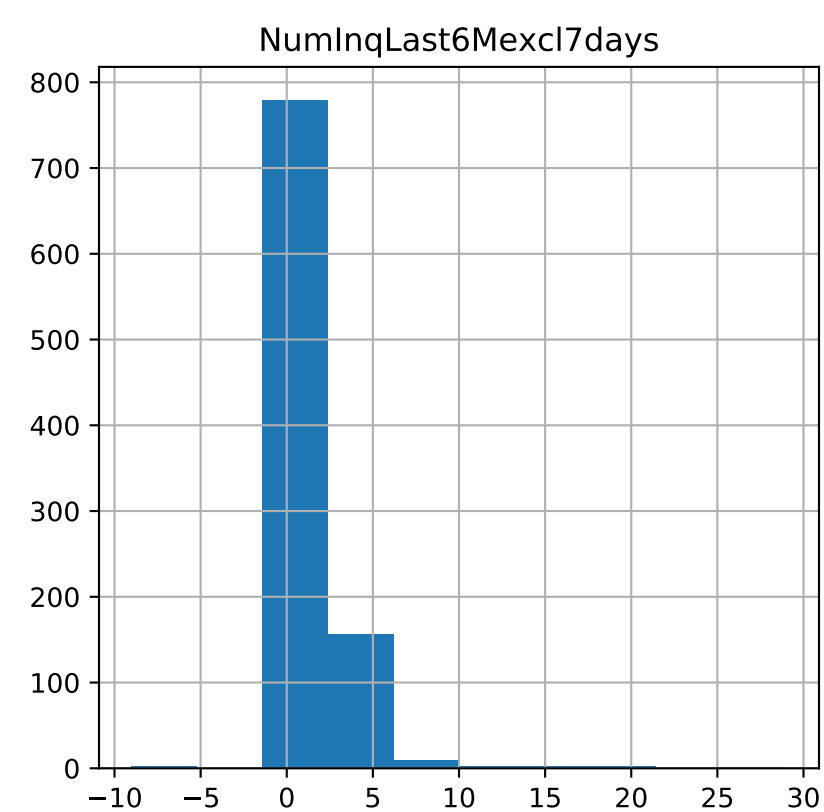
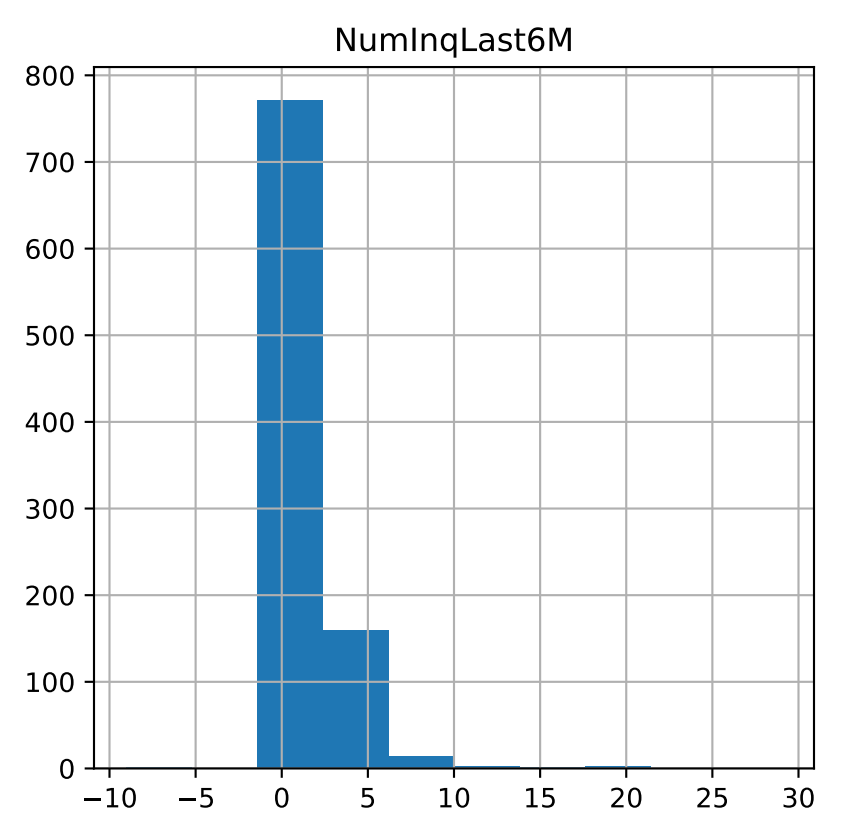
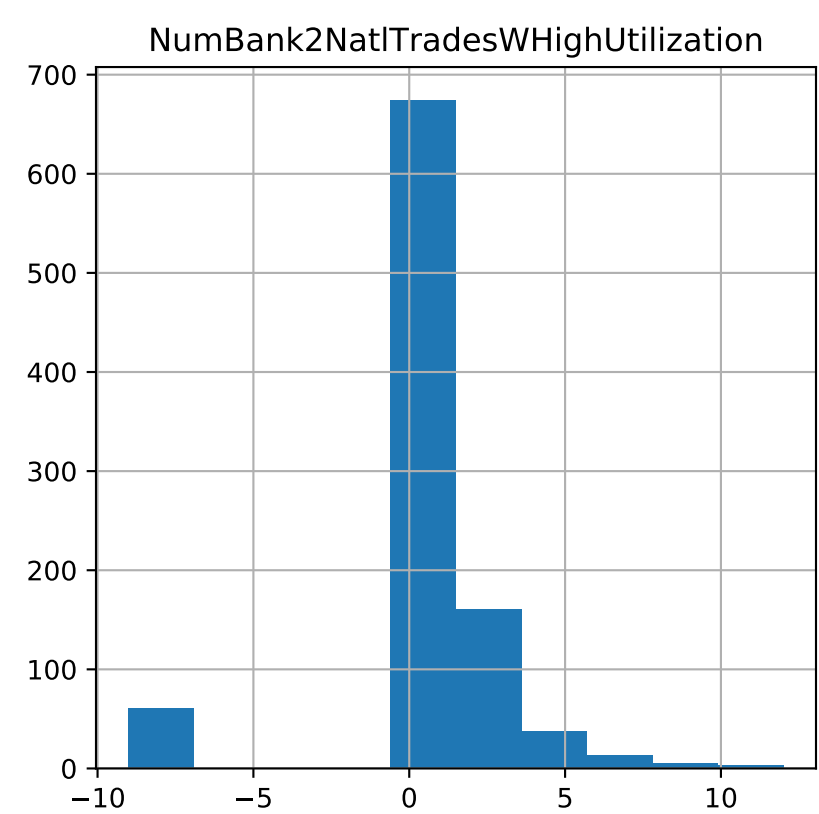
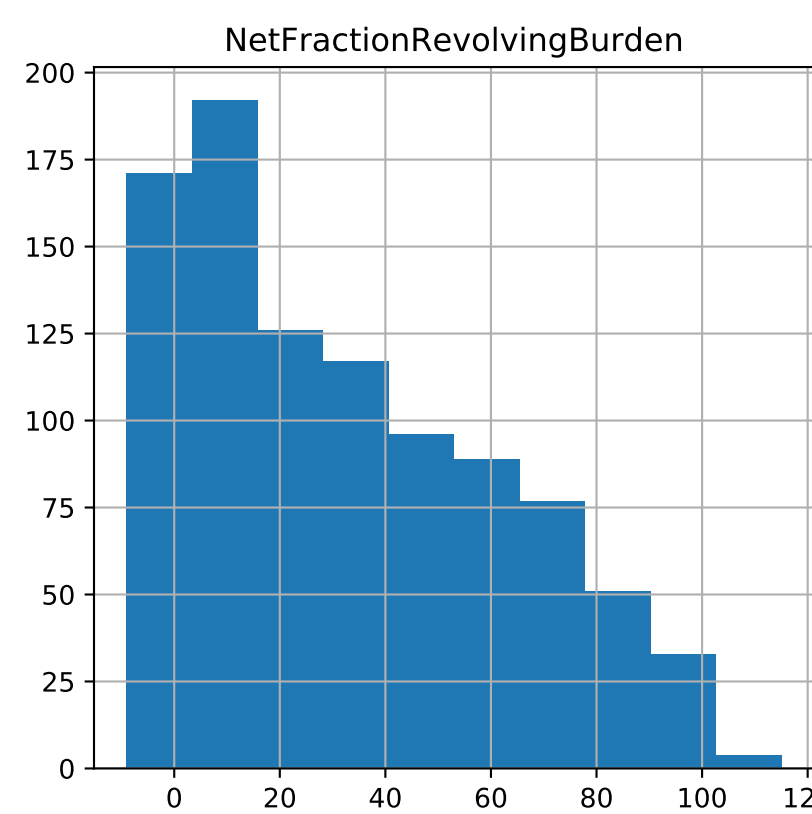
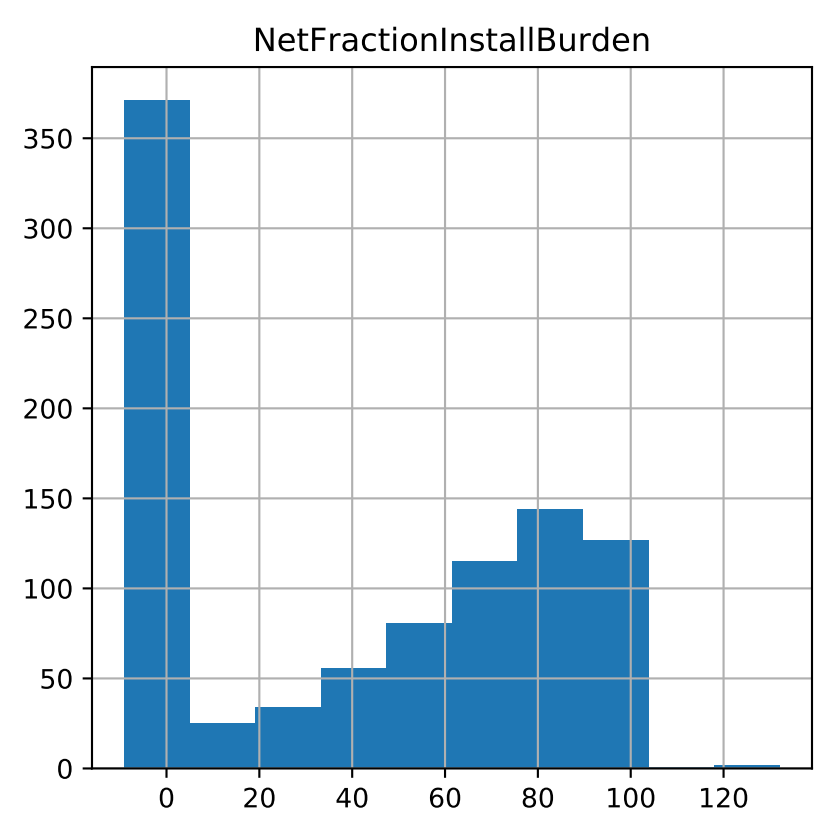
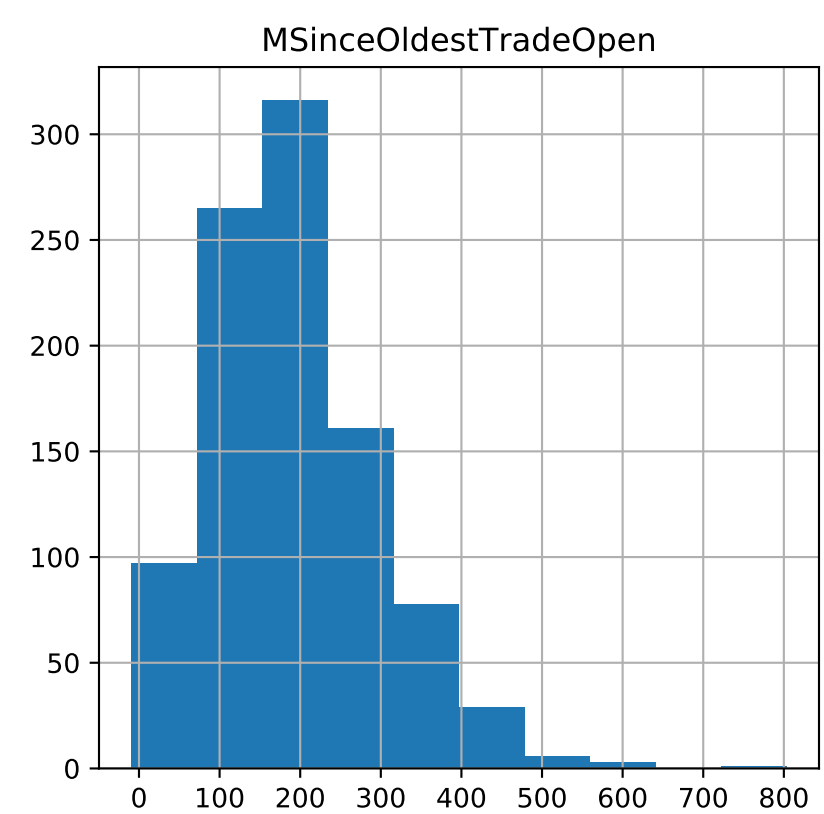
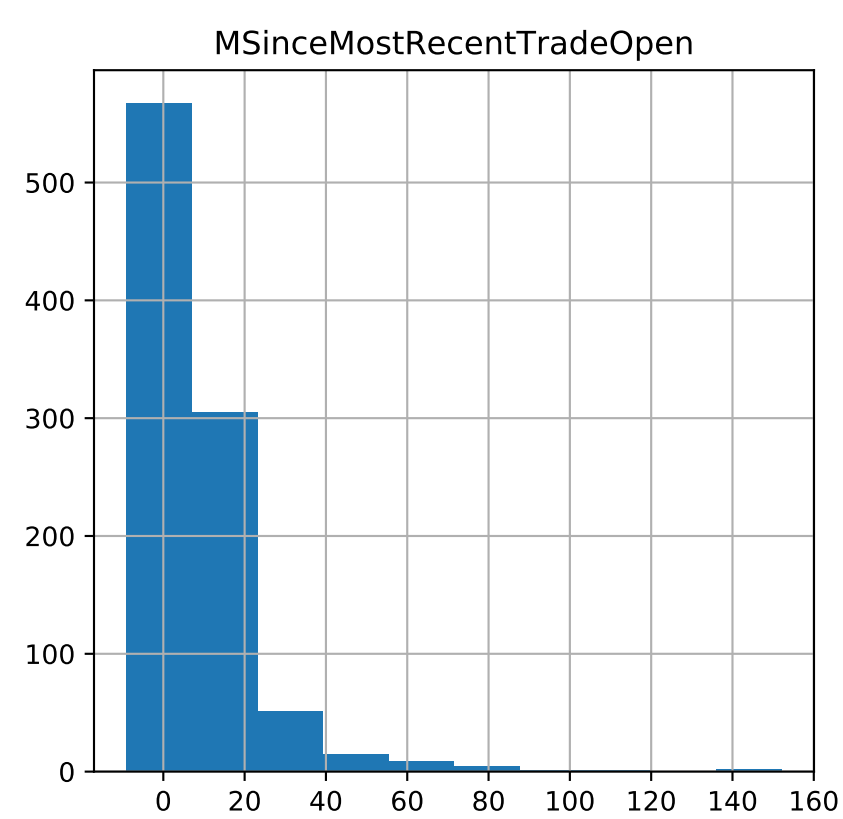
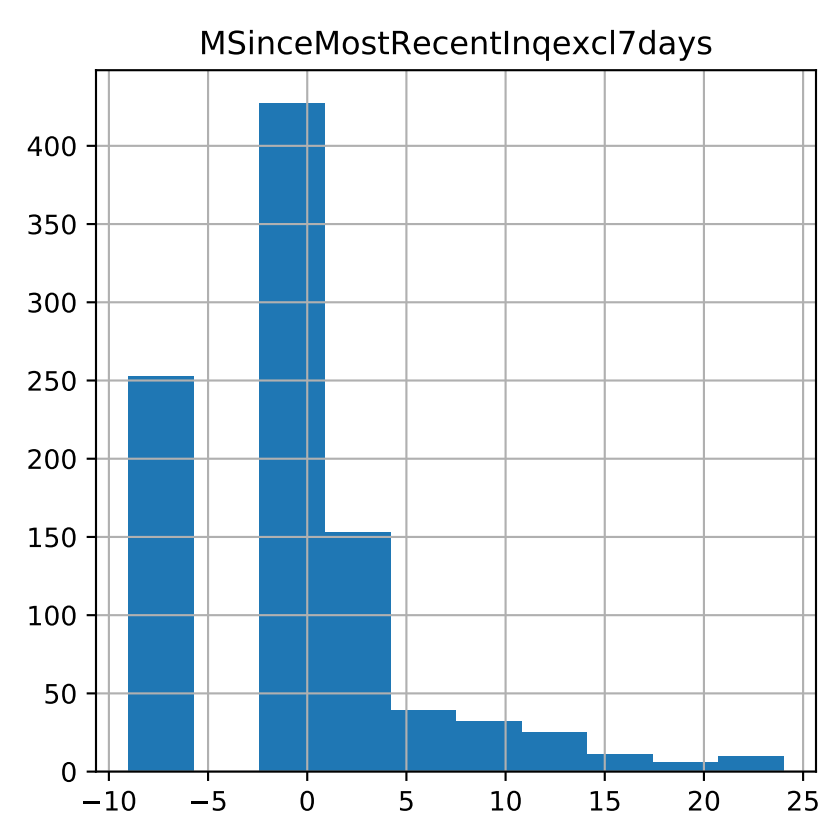
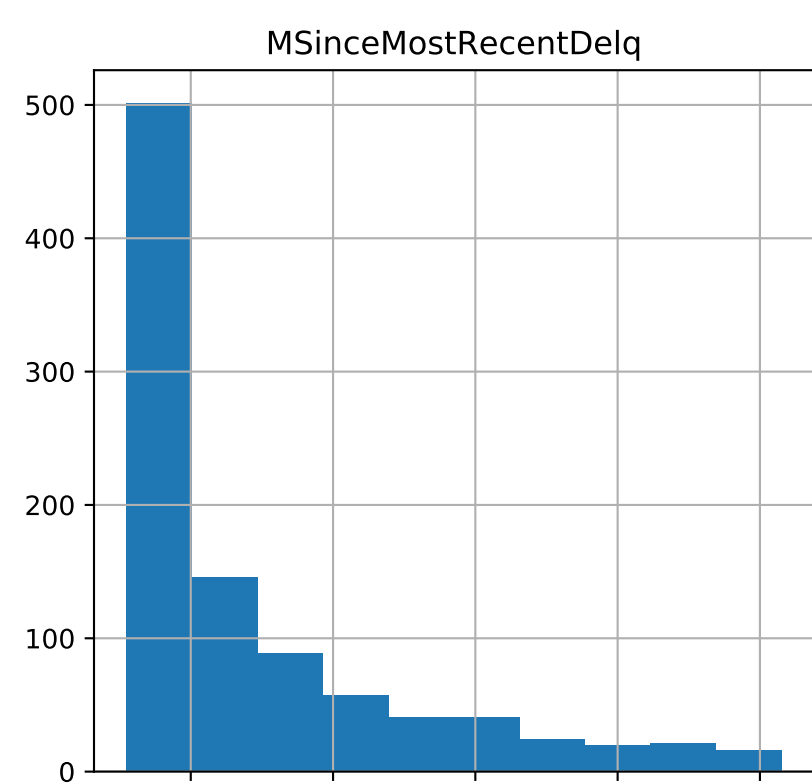
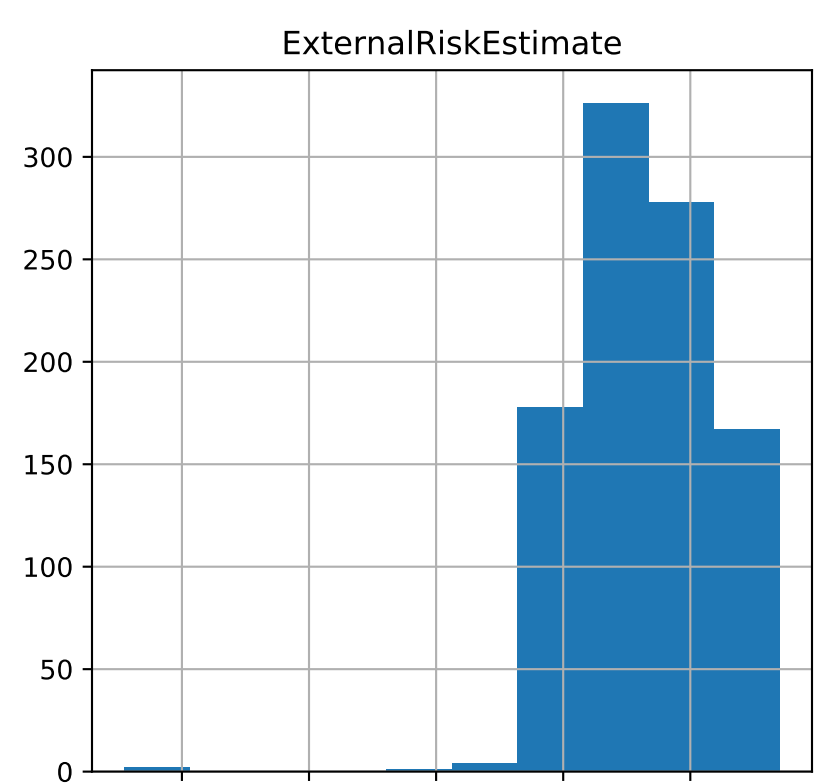
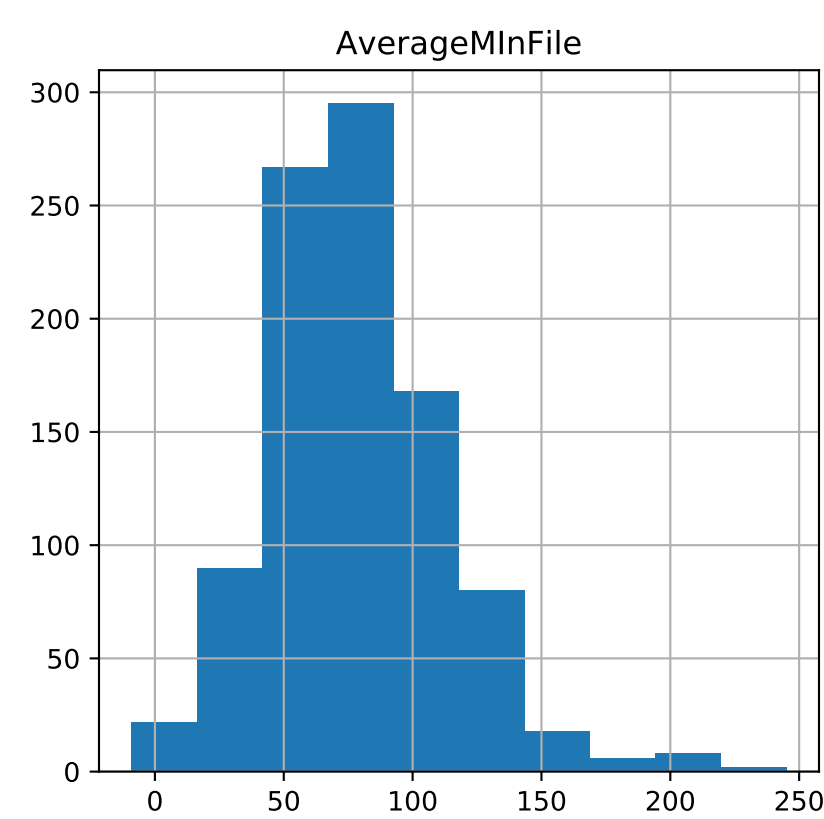
There are a lot of features have outliers, External Risk Estimate, Month Since Oldest Trade Open, Month Since Most Recent Trade Open, Average Months in File, Number Satisfactory Trades, Number Trades 60+ Ever, Number Trades 90+ Ever, Months Since Most Recent Delinquency, Number of Total Trades (total number of credit accounts), Number of Trades Open in Last 12 Months, Percent Install trades, Months Since Most Recent Inquiries excluding 7 days, Number of Inquiries Last 6 Months, Number of Inquiries Last 6 Months excluding 7 days, Number Revolving Trades with Balance, Number Install Trades with Balance, Number Bank/Natl Trades with high utilization ratio, Percent Trades Never Delinquent, Percent Trades with Balance, Net Fraction Revolving Burden, Net Fraction Install Burden.

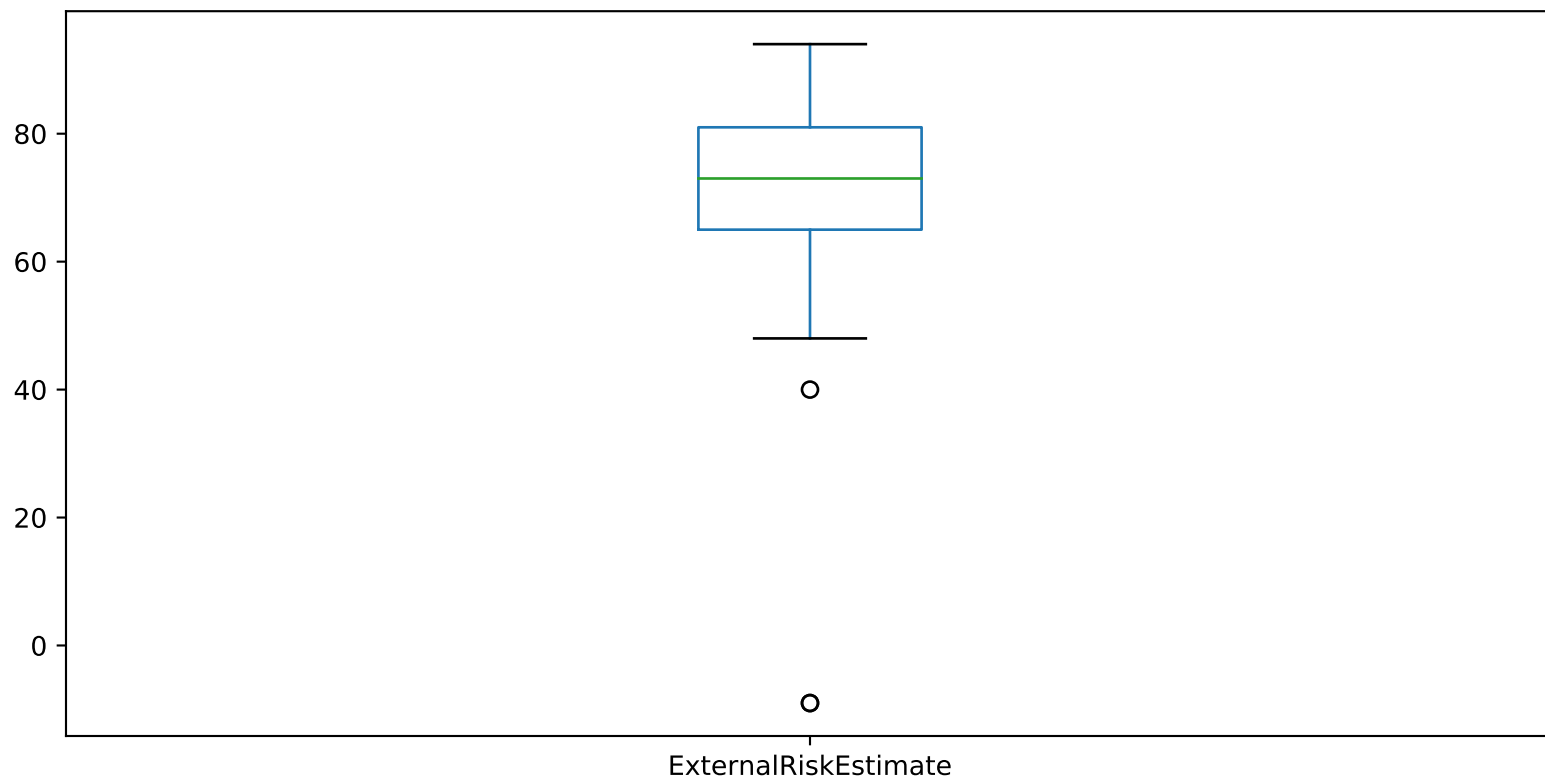
- **Bar Plots for categorical Features [plots attached at end of file]**

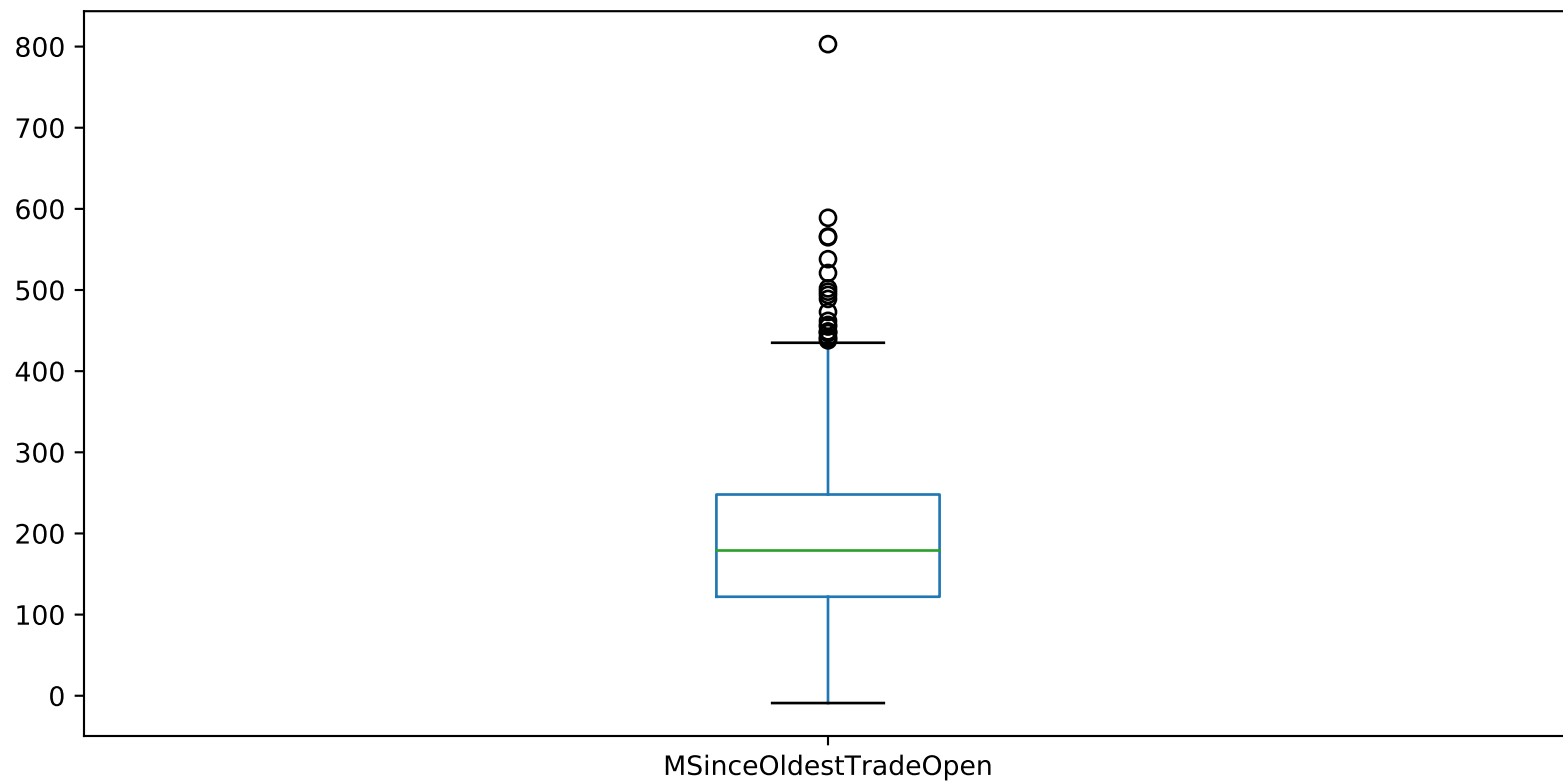
The risk performance shows that the percentage of good performance is almost the same of bad performance, but we can see bad performance customer is still a little more than good performance. Therefore, this credit company need to take risk precautions when loan to customers.

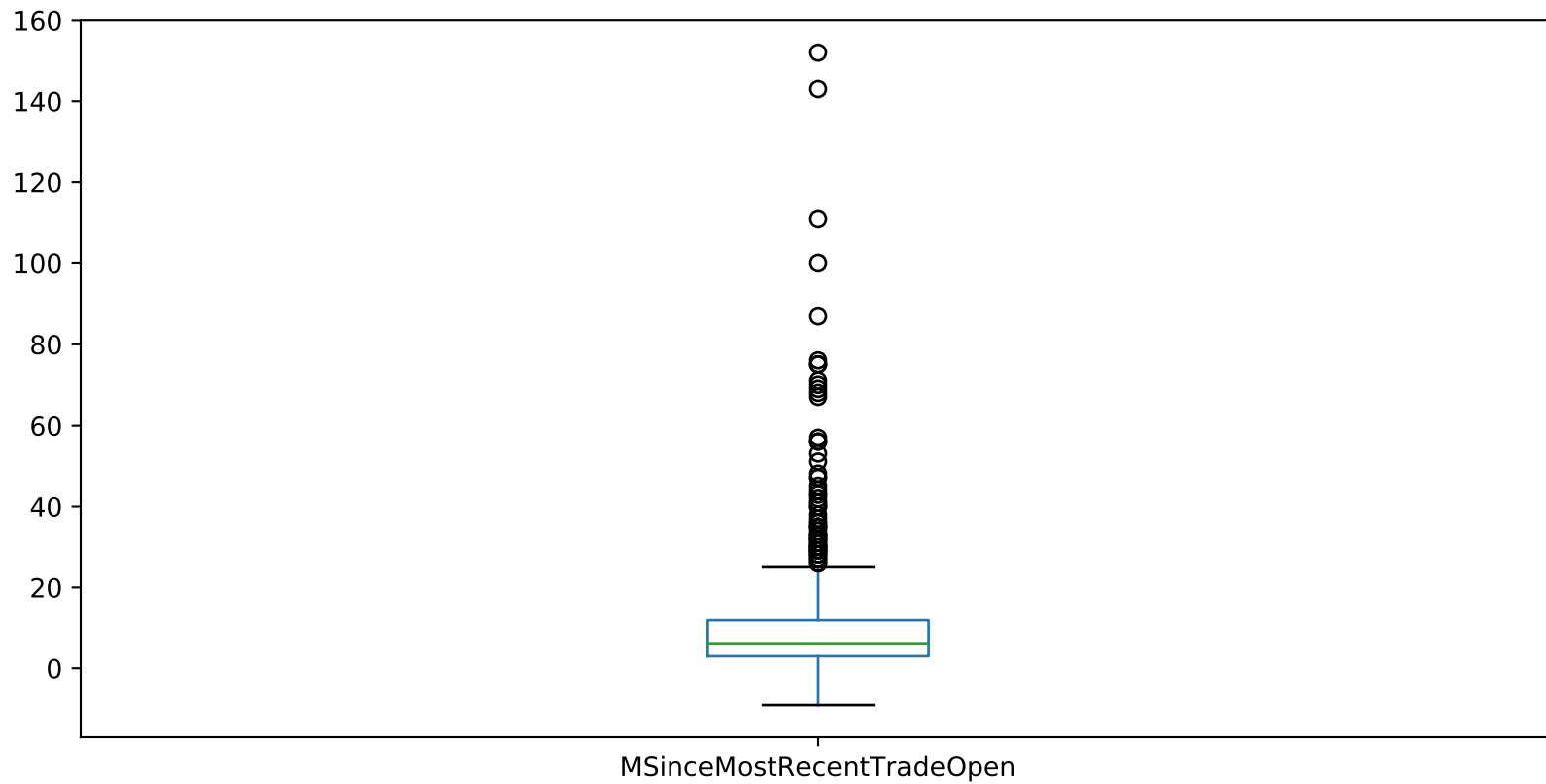
The most value of Max delinquent/Public Records Last 12 Month and Max Delinquency Ever current and never delinquent, this tell us the people never delinquent account for the

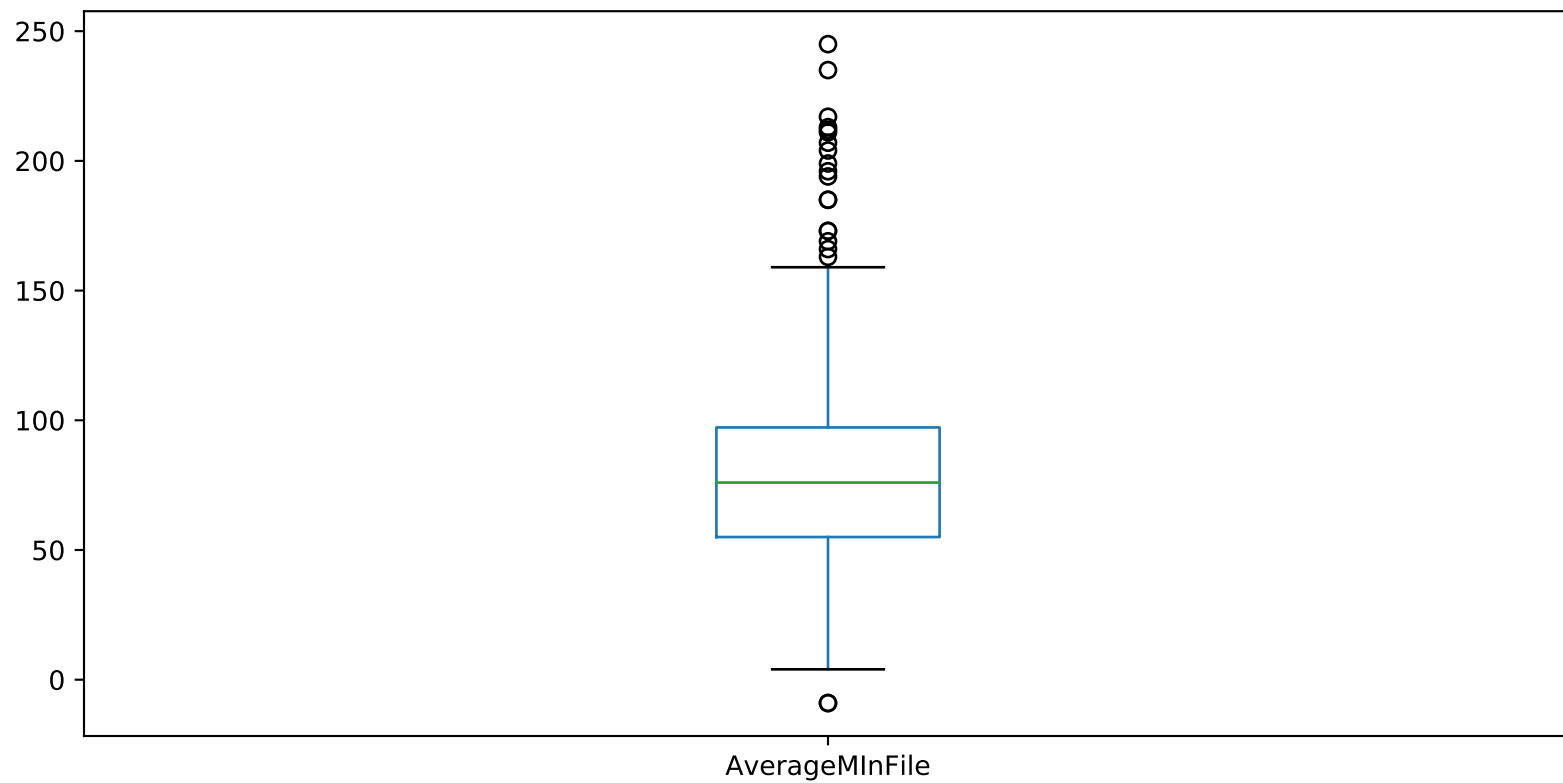
majority. The customers that unknown the time of delinquent is the second most. In addition, the Max delinquent and Public Records Last 12 Month bar plots show that the more days of delinquent, the less number of people.

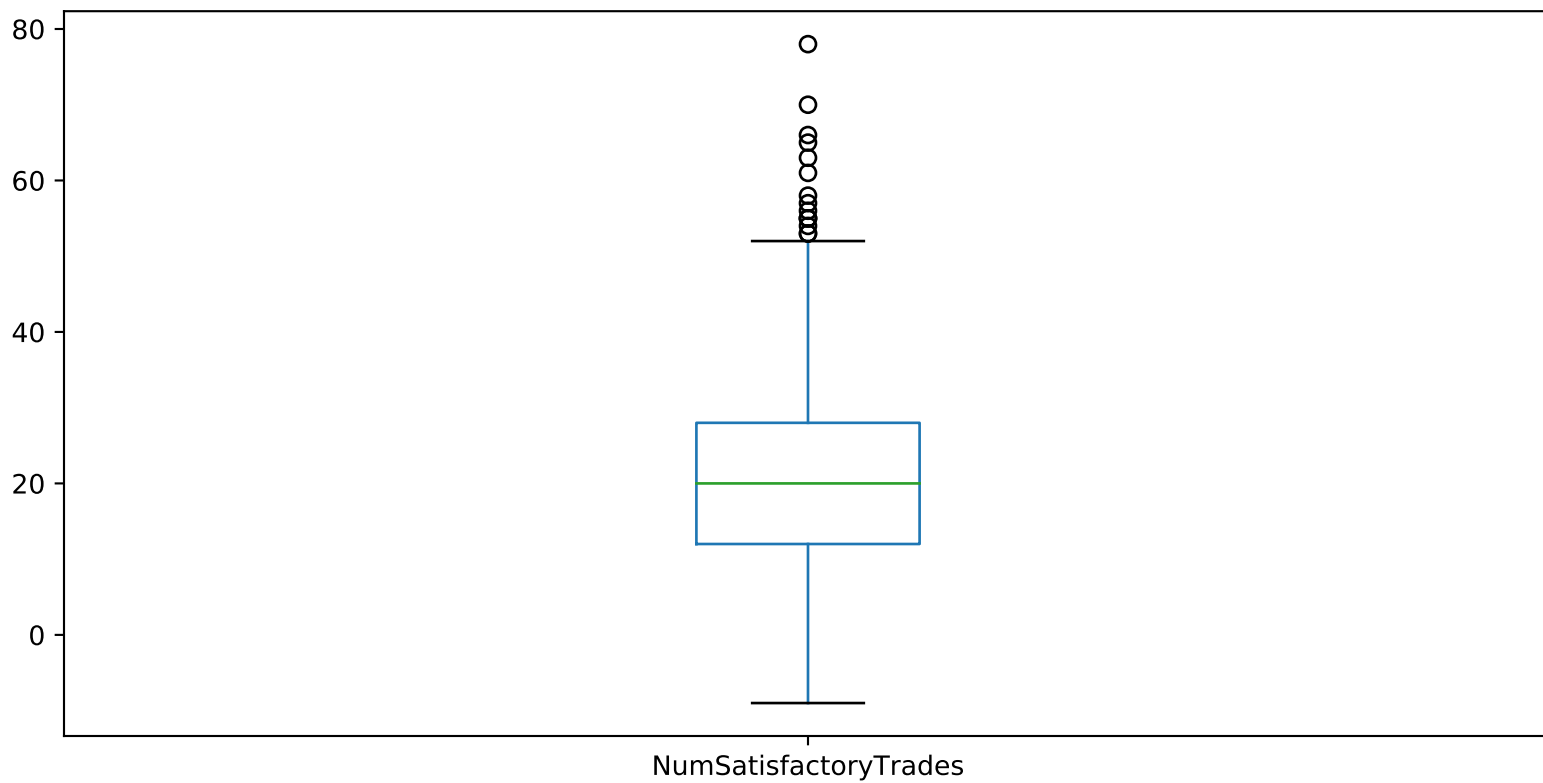


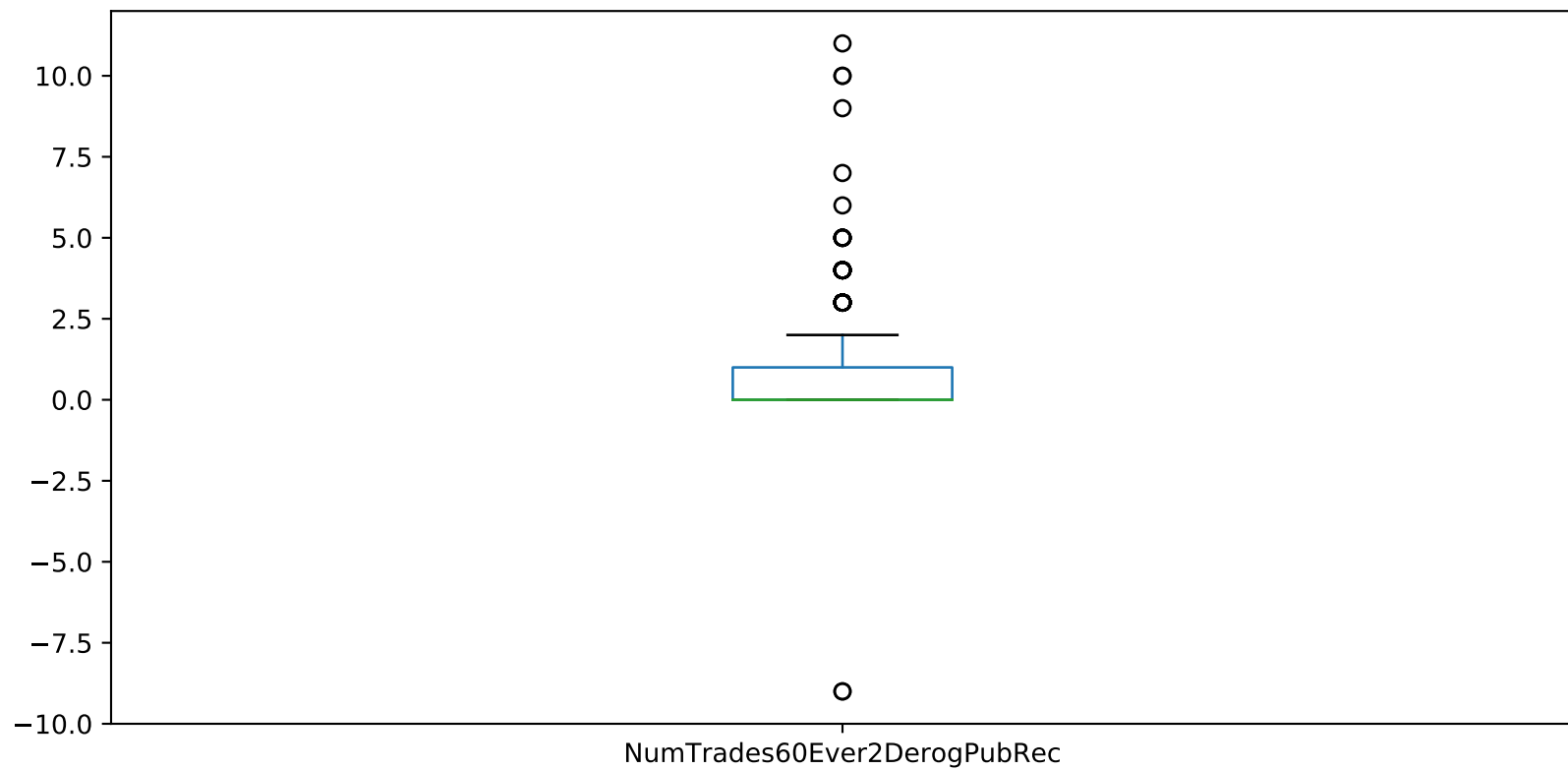


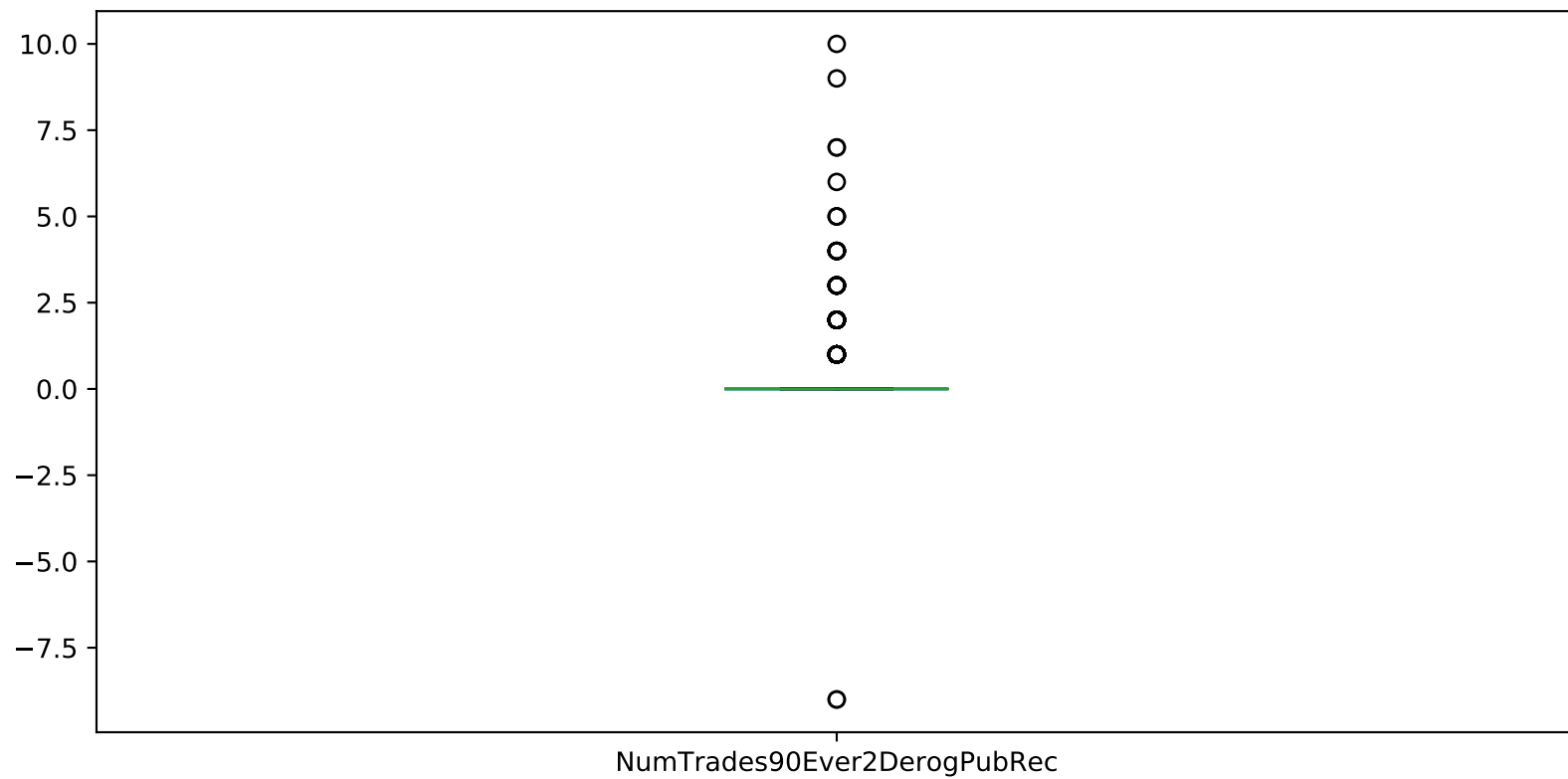


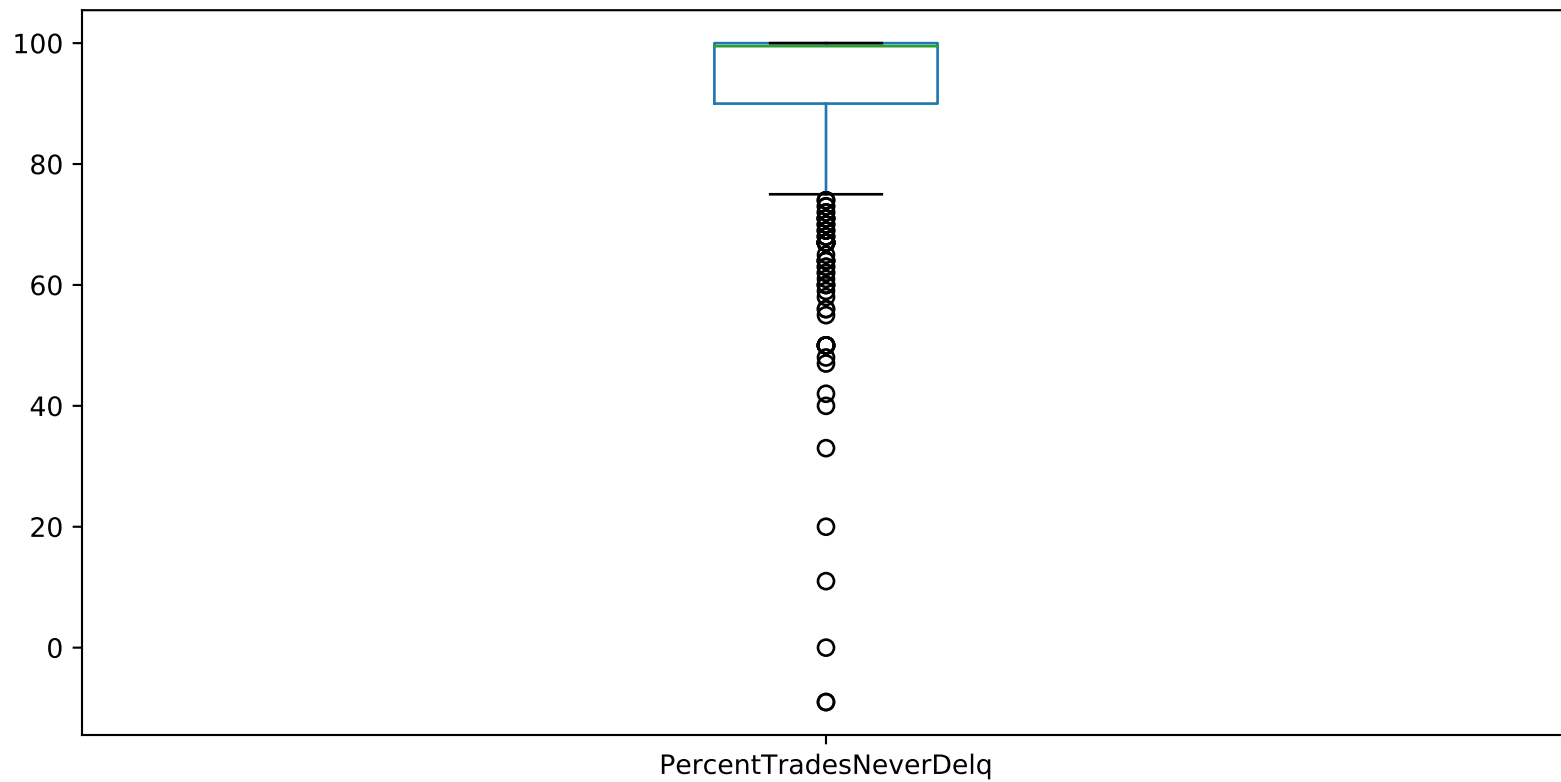


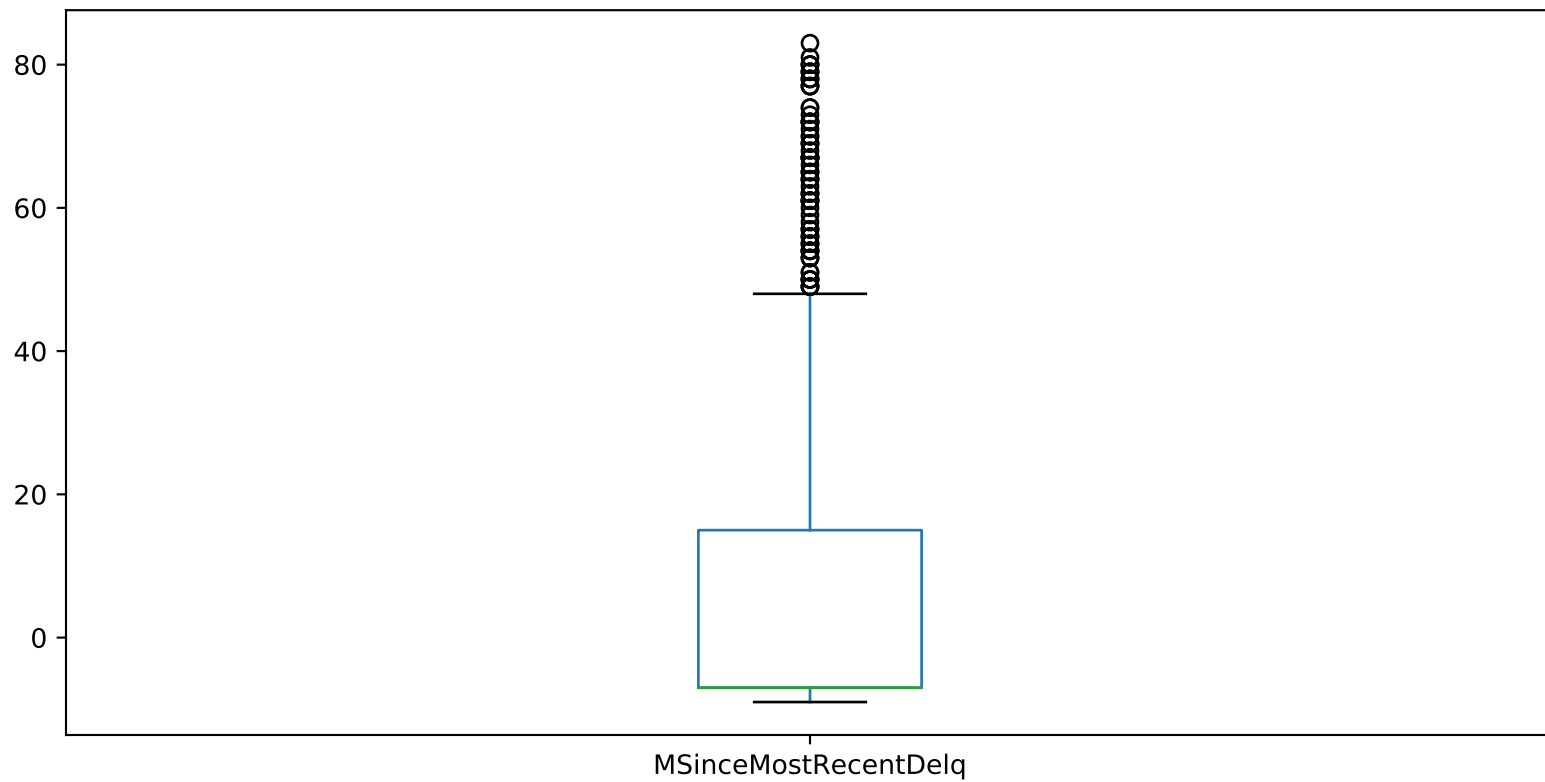


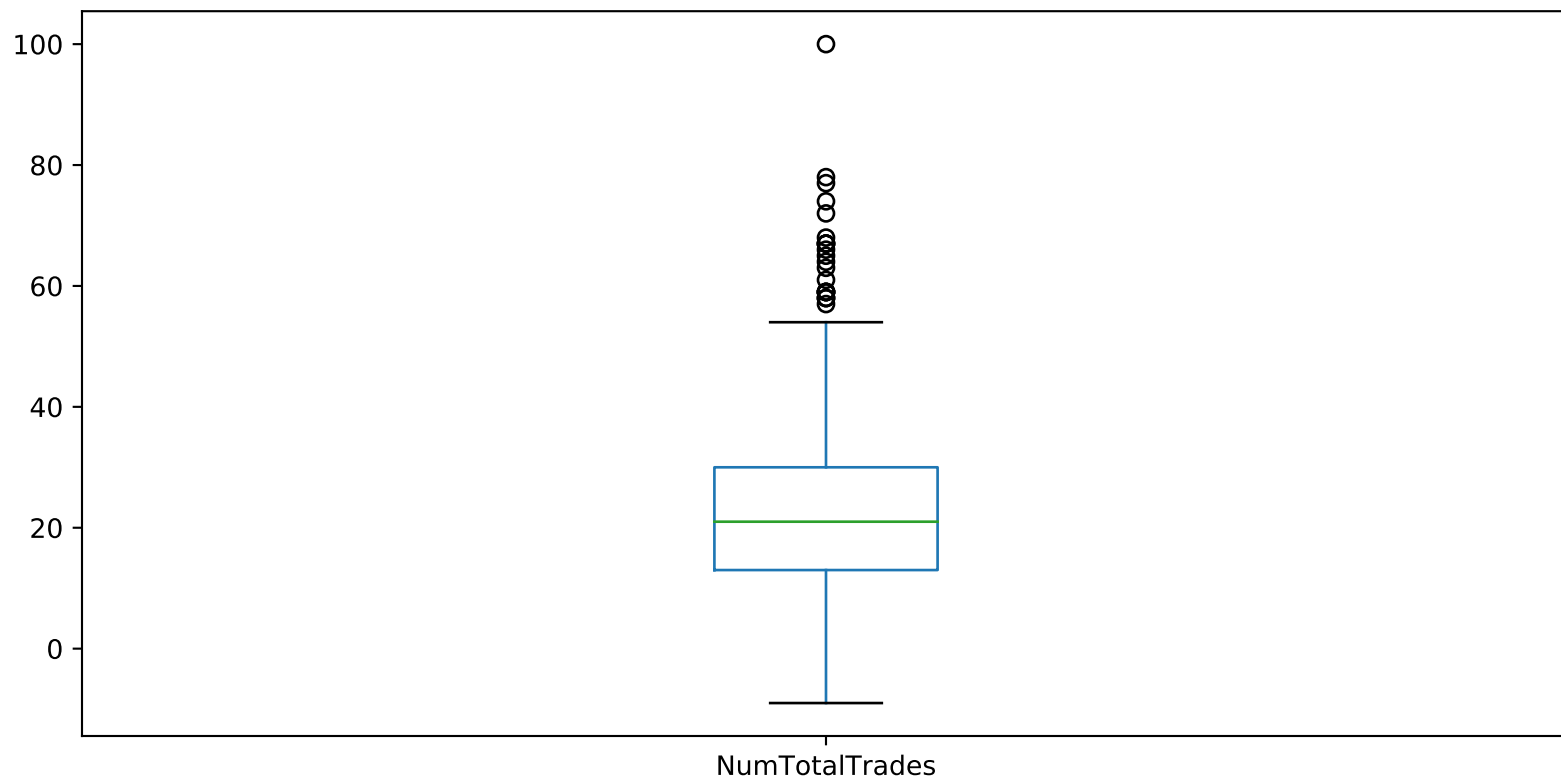


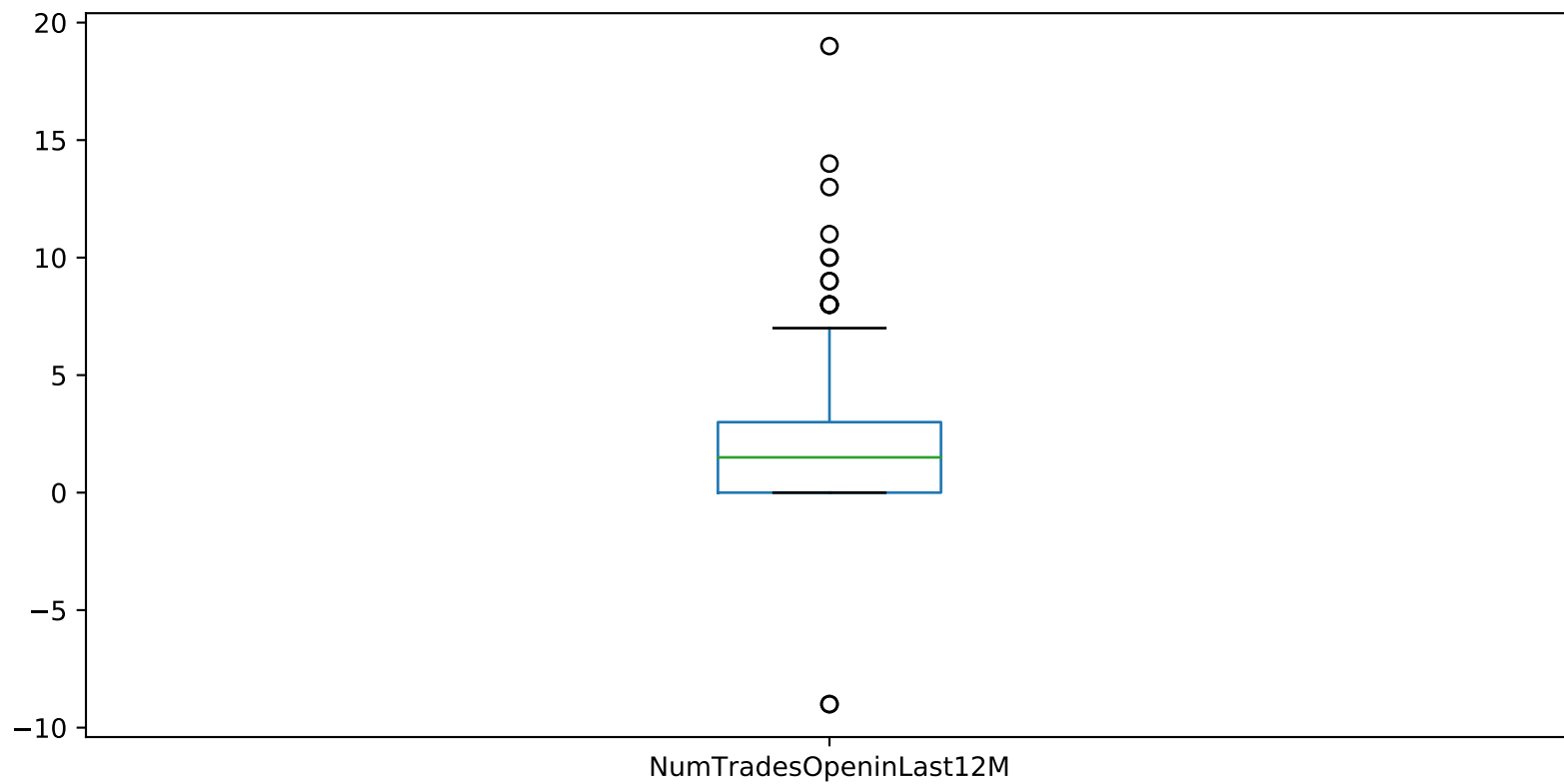


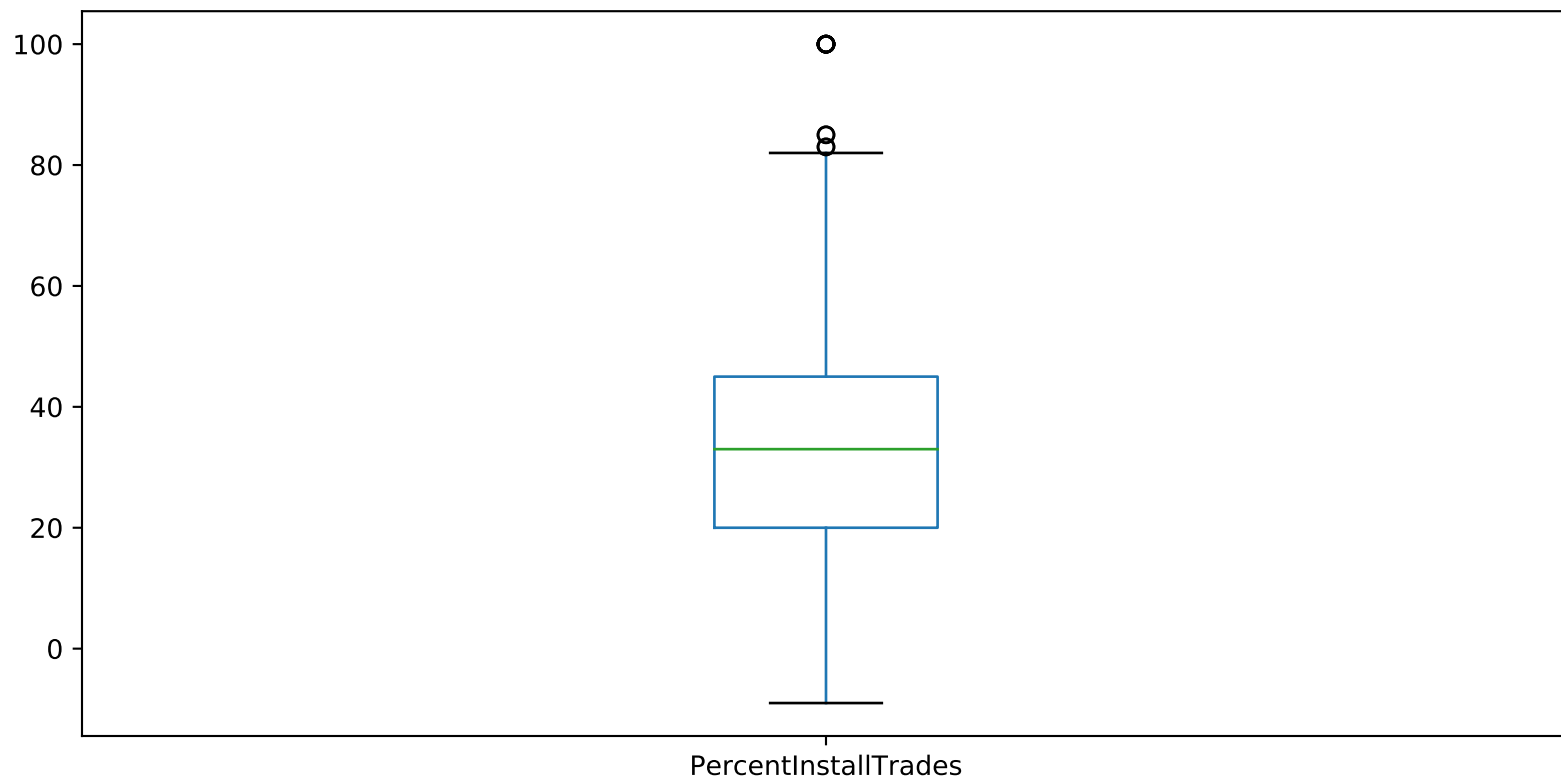


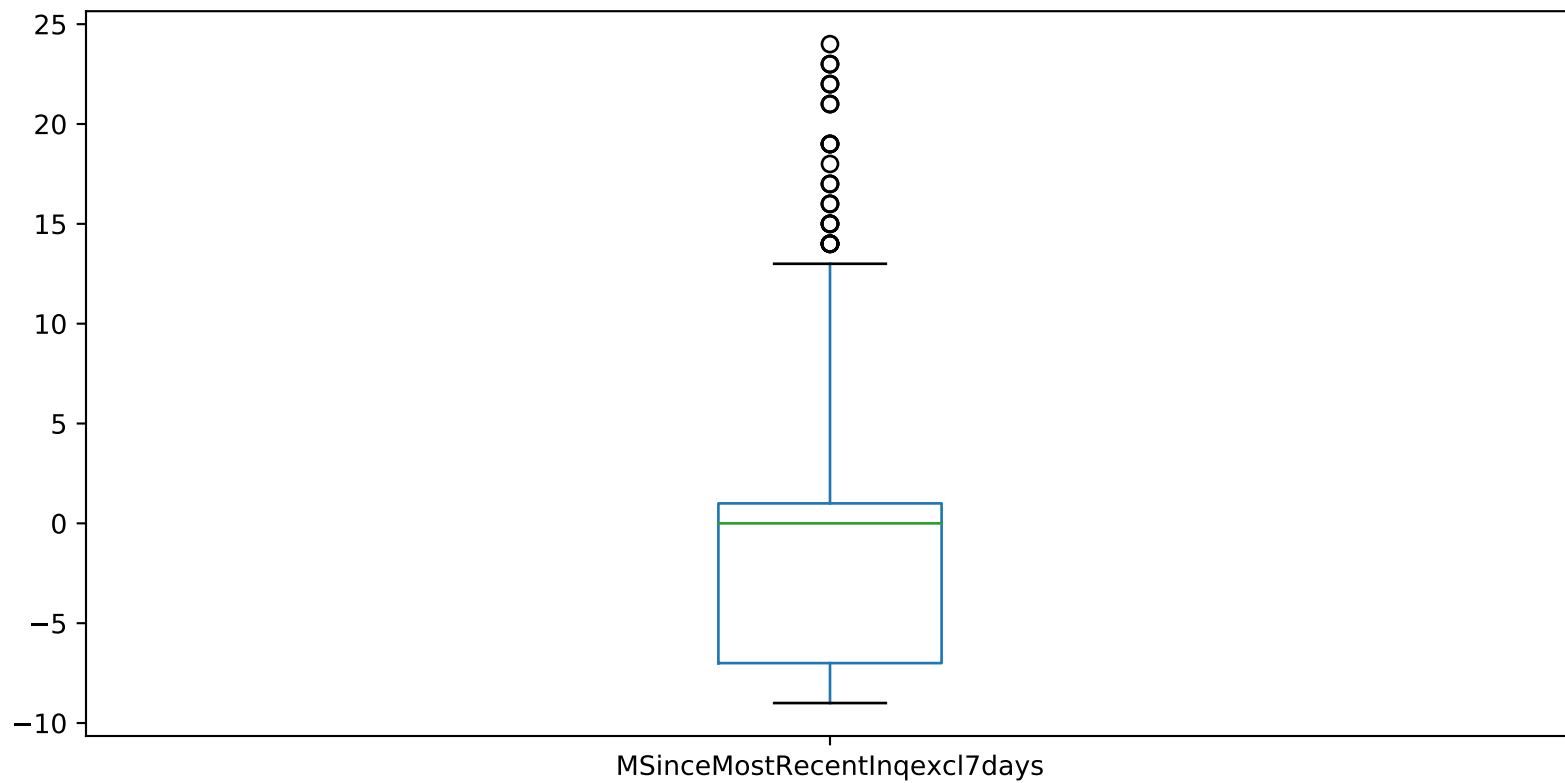


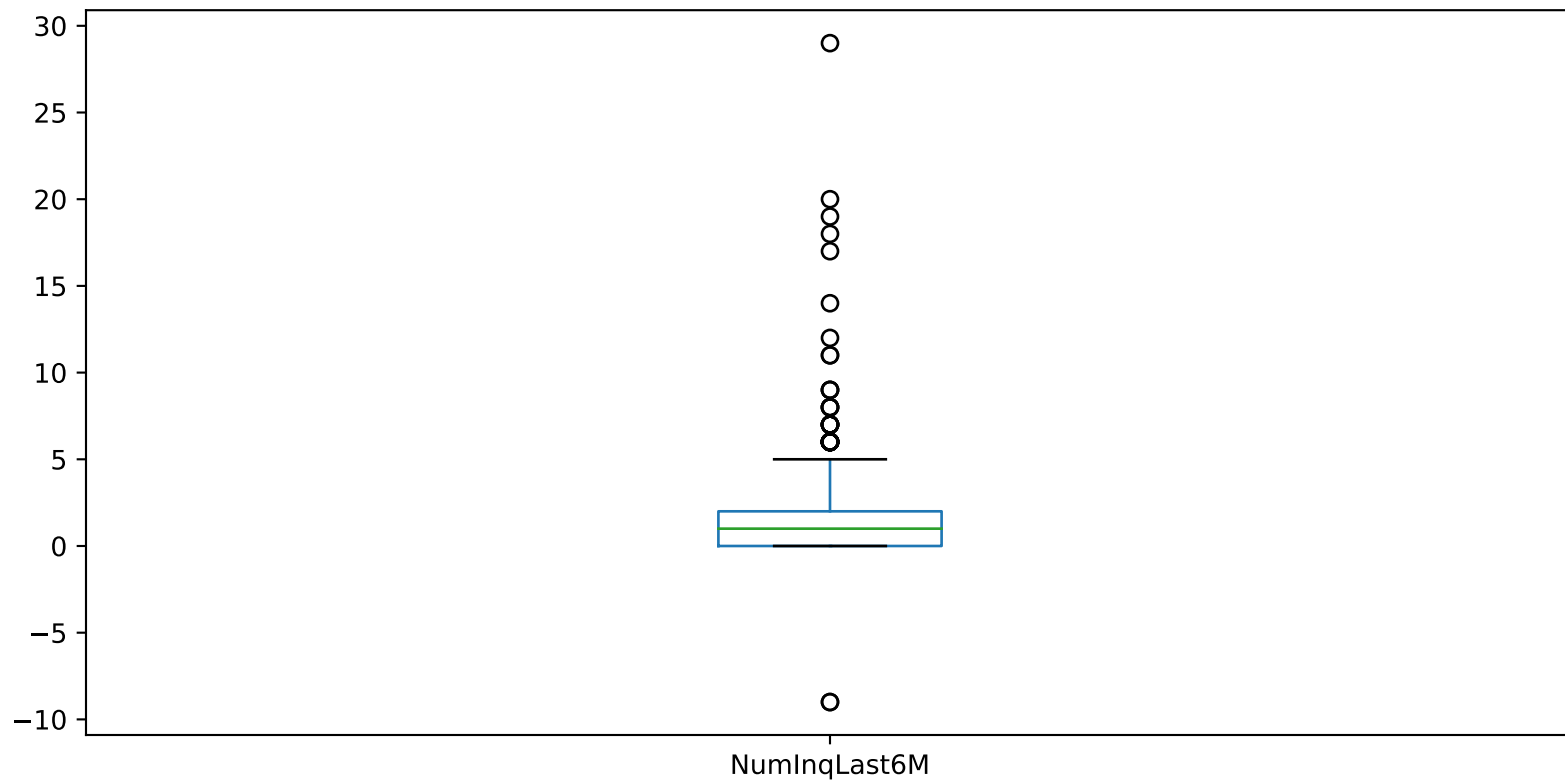


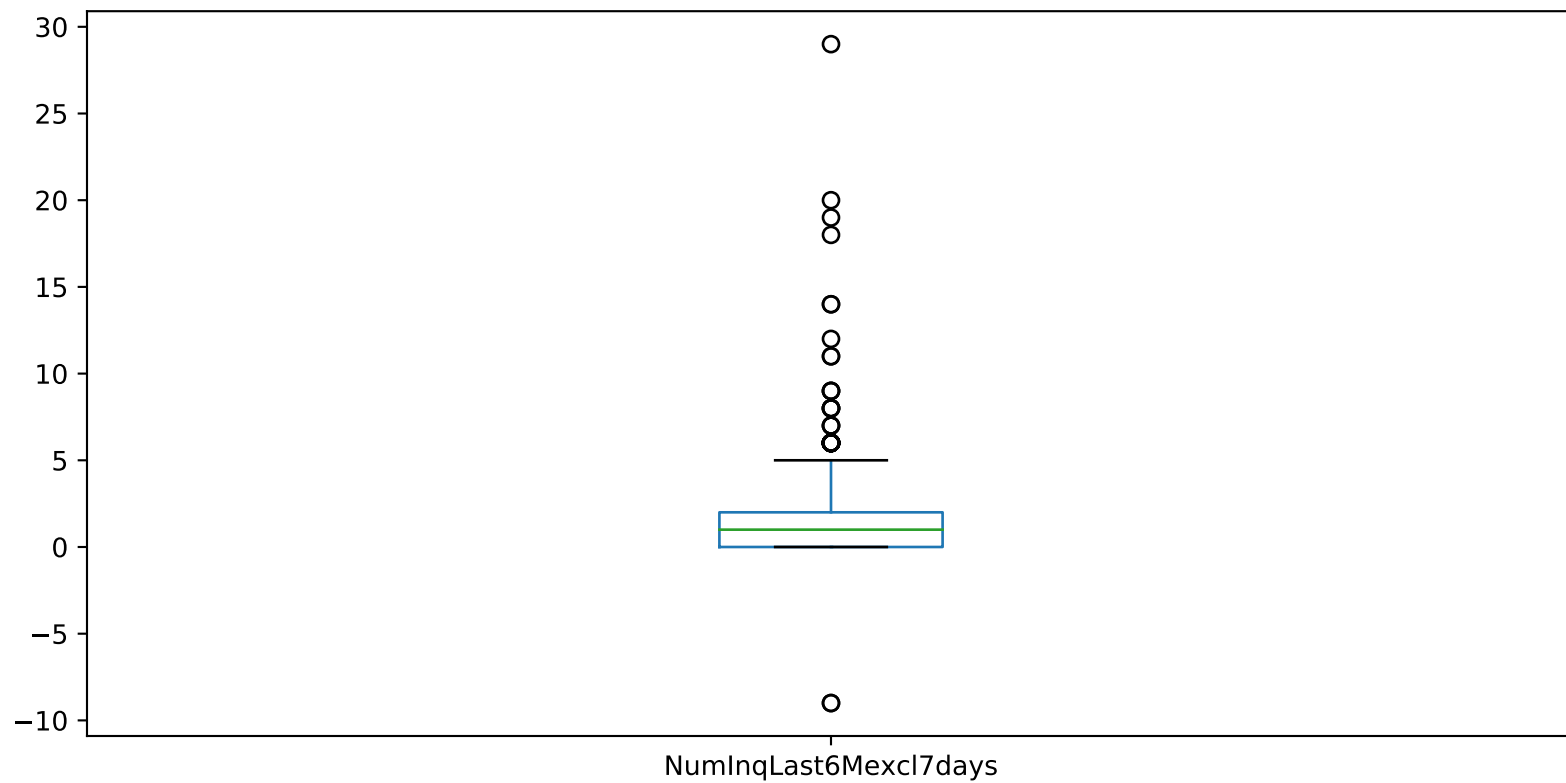


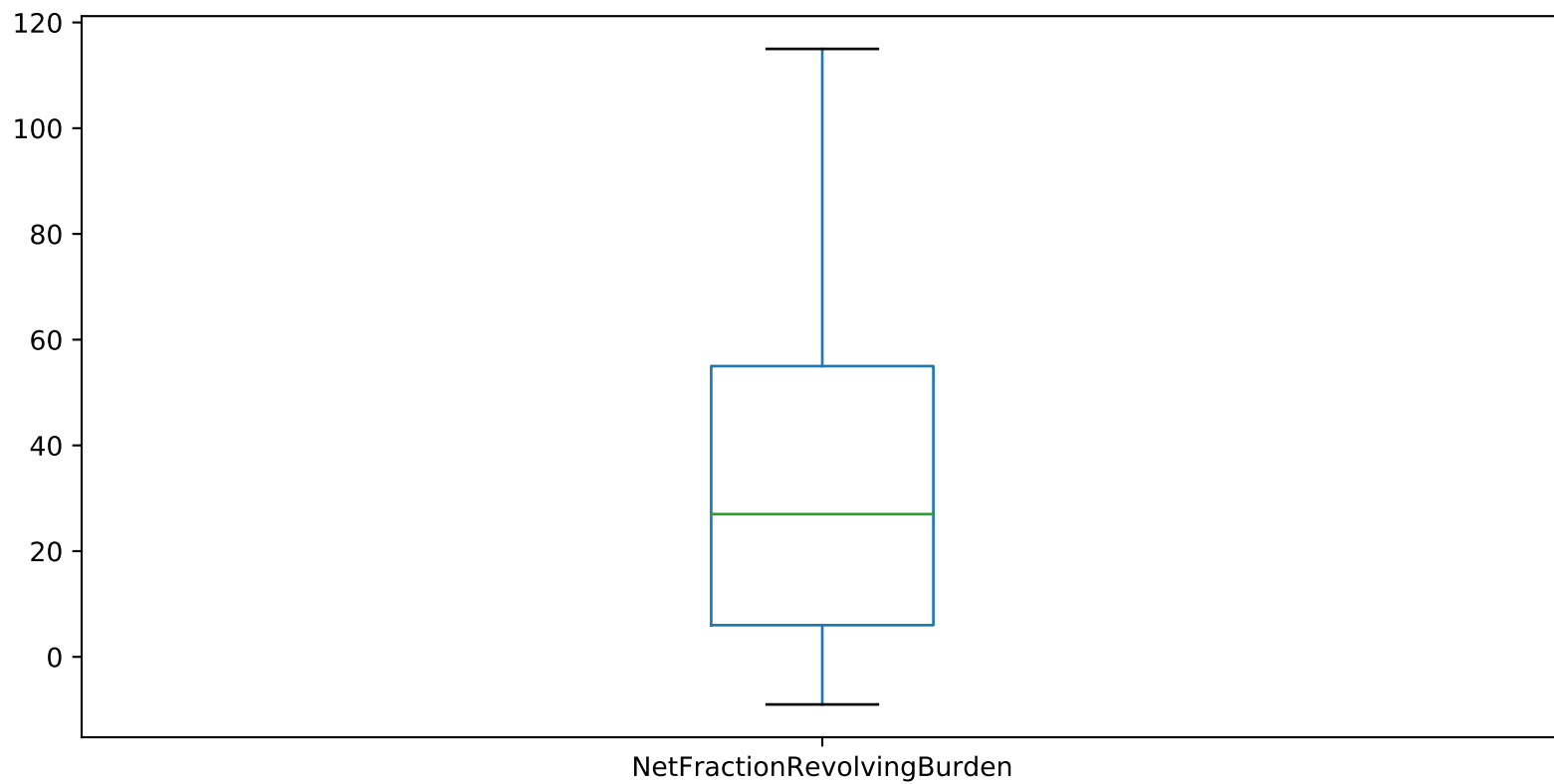


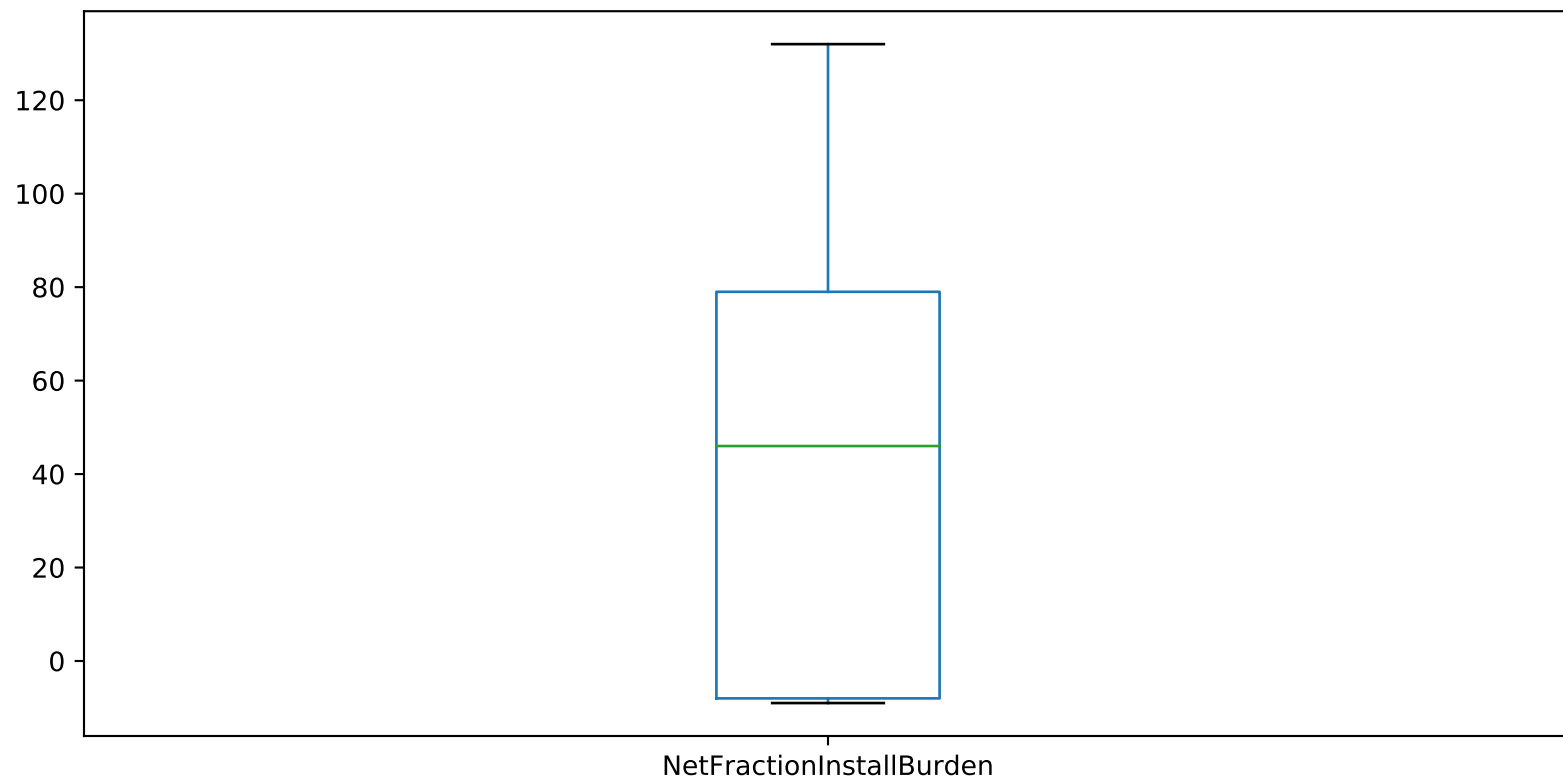


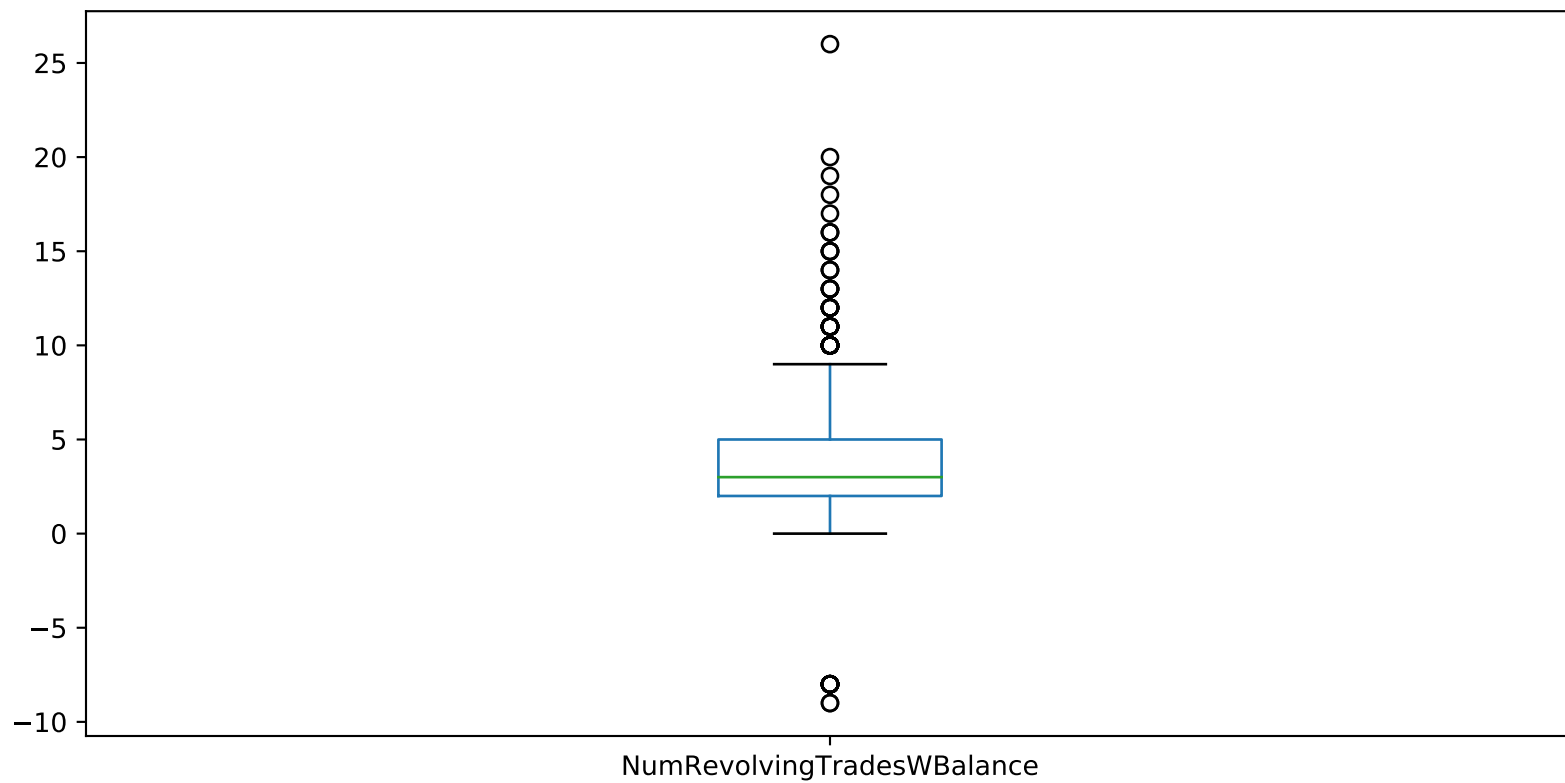


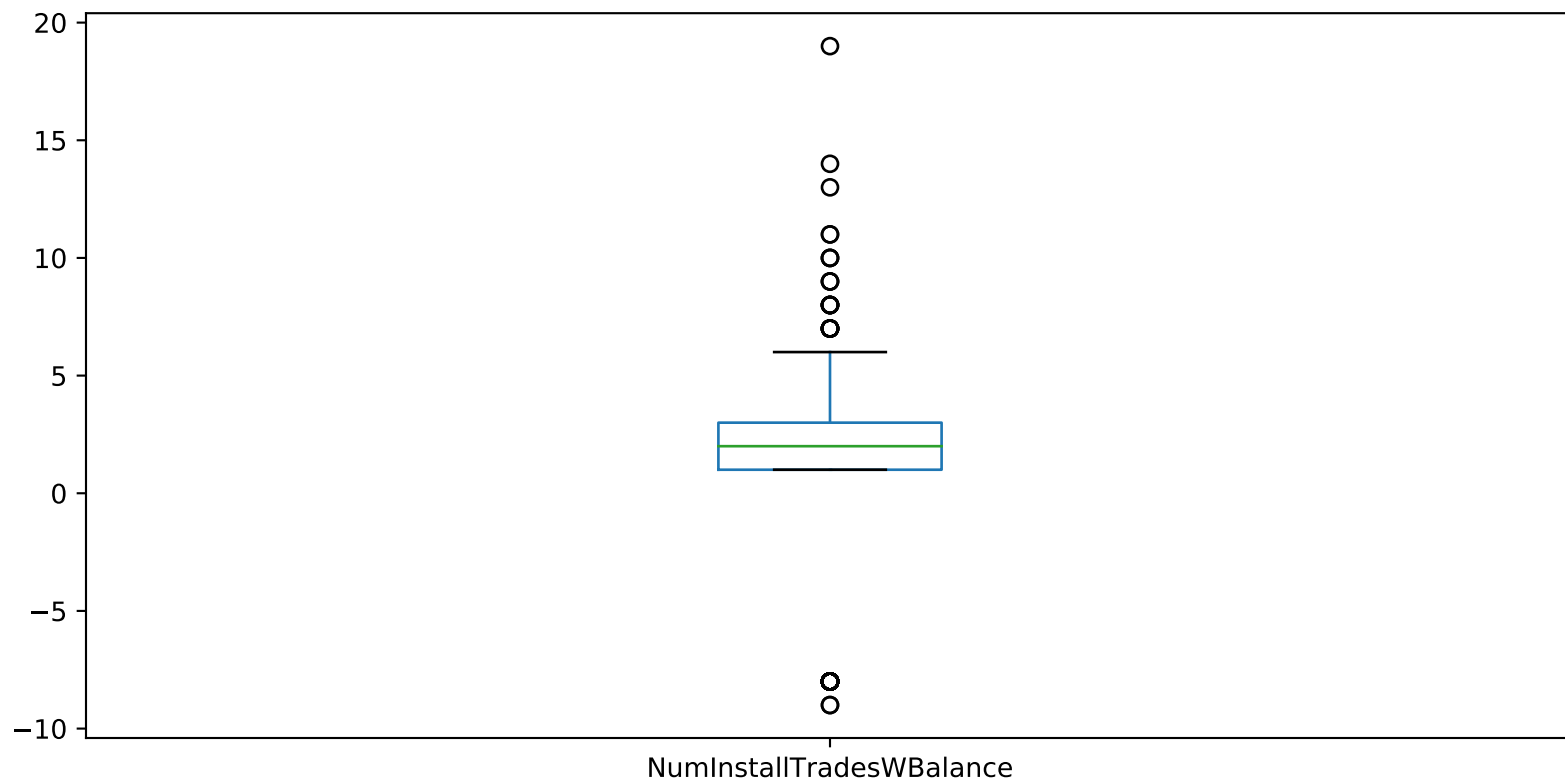


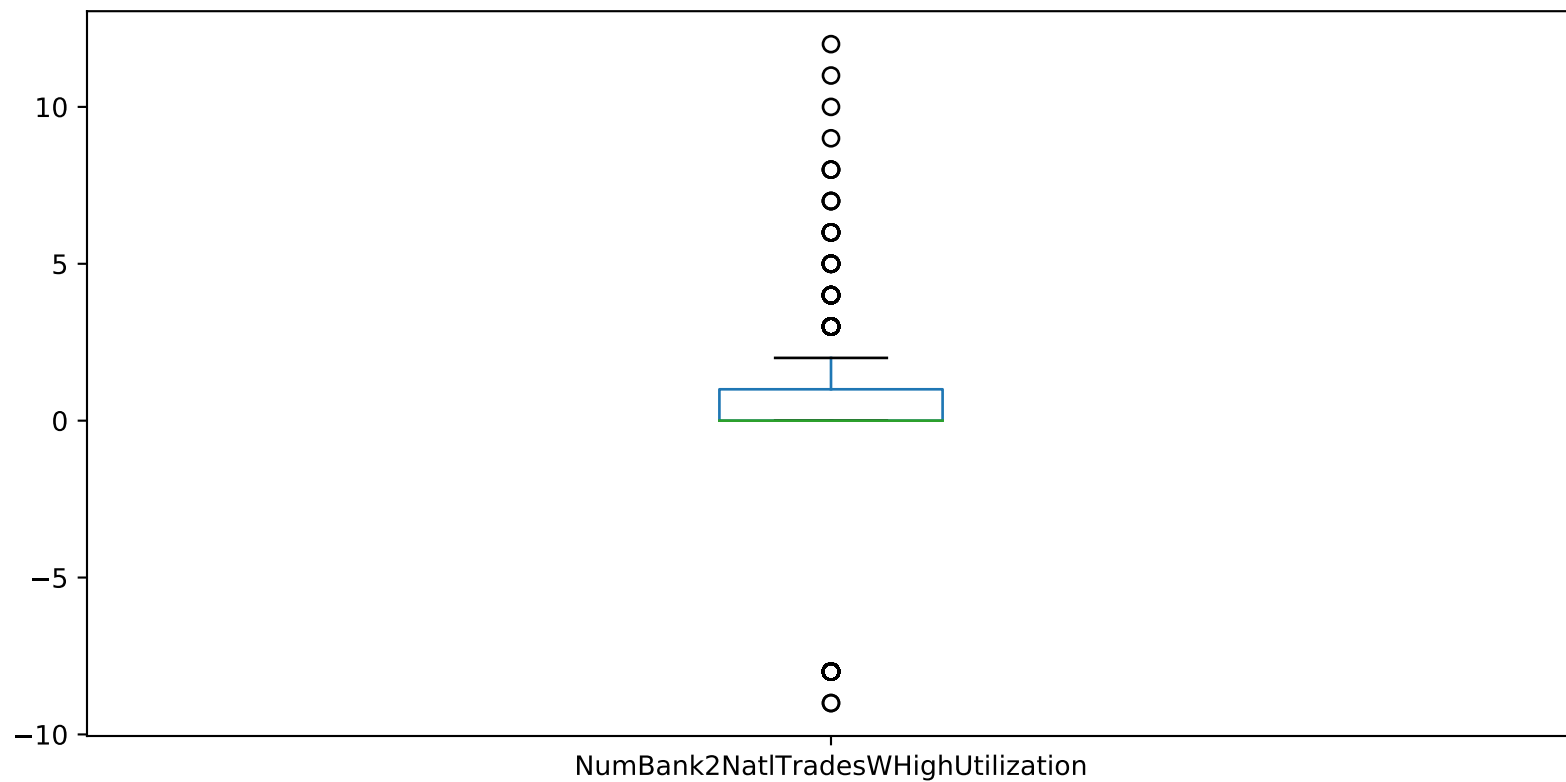


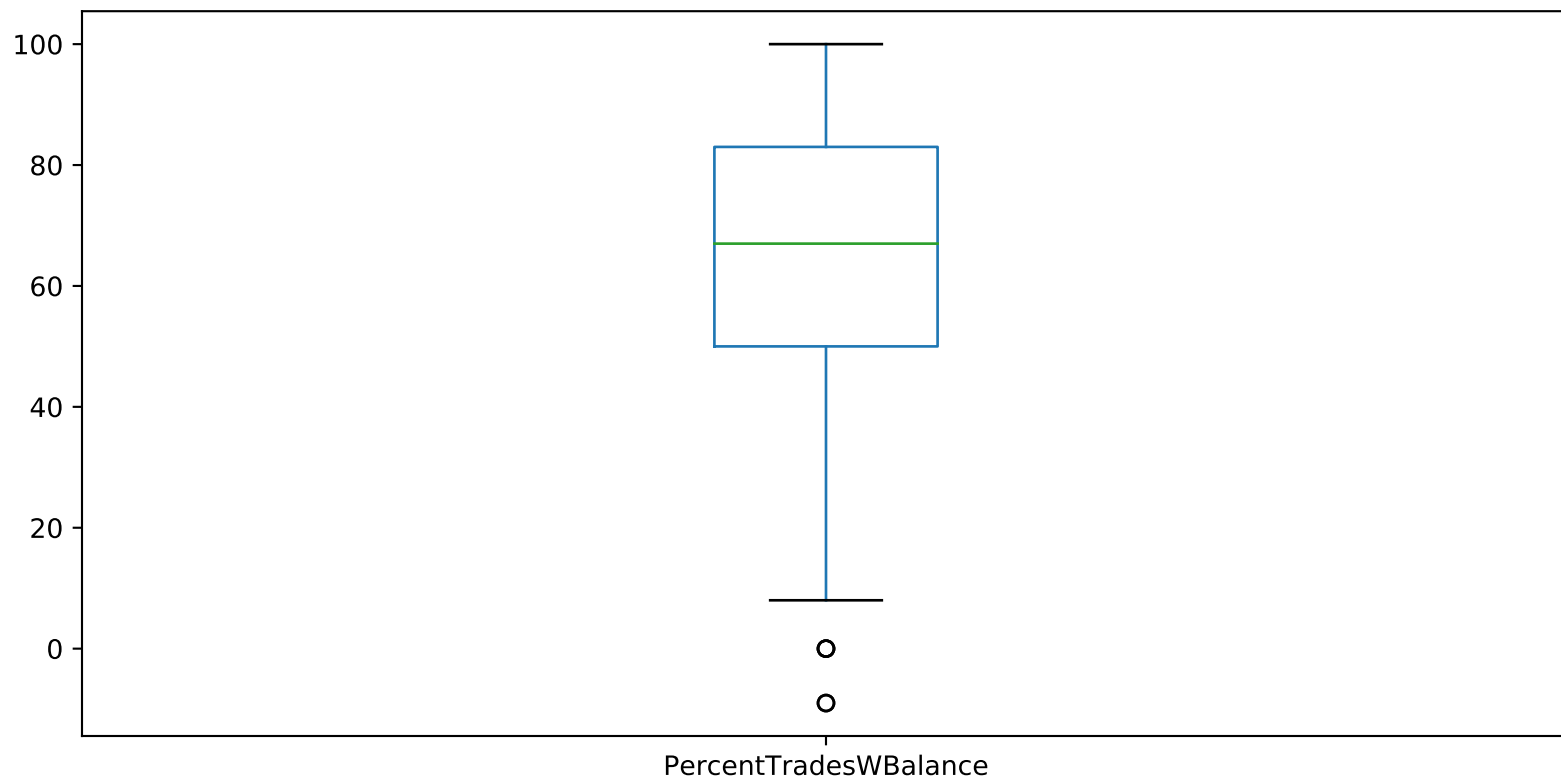




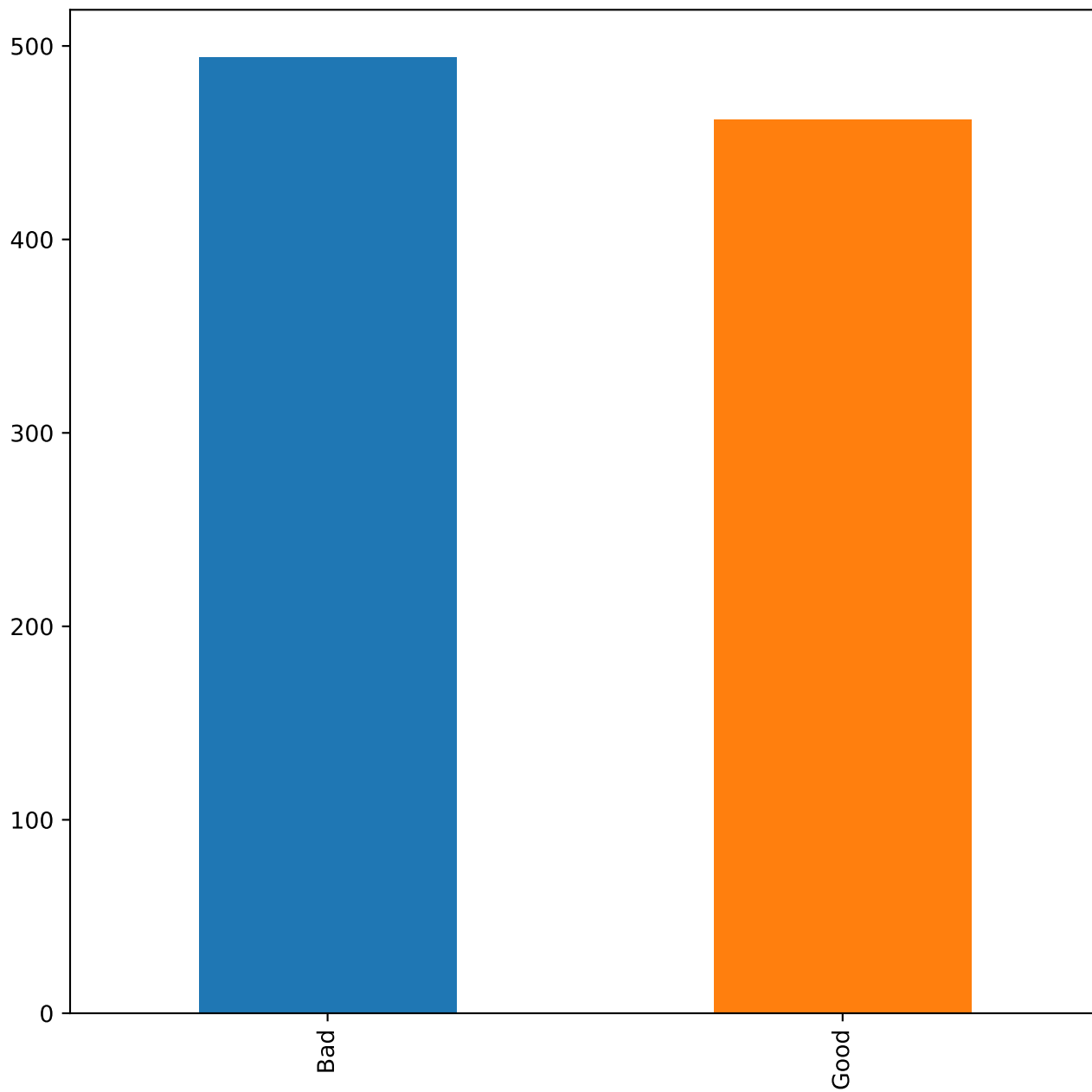




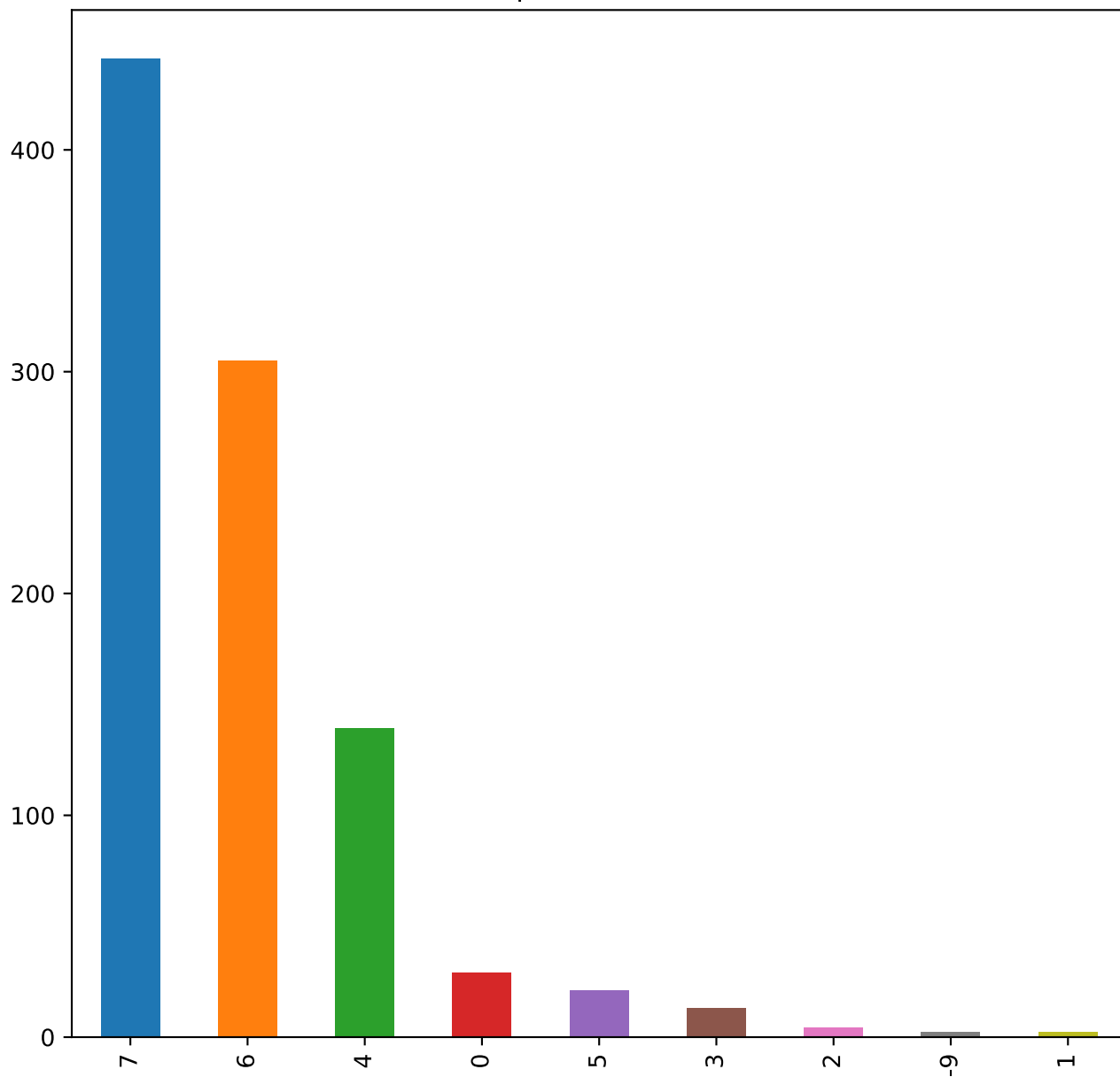




RiskPerformance



MaxDelq2PublicRecLast12M



MaxDelqEver

