

RandomForest_Part01

March 7, 2022

```
[1]: import pandas as pd
import seaborn as sns; sns.set()
import matplotlib.pyplot as plt
import numpy as np

pd.options.display.max_columns = None
pd.set_option('display.max_rows', 200)
pd.set_option('display.float_format', lambda x: '%.3f' % x)

#import dataset
df = pd.read_csv ('WinnipegDataset.txt',sep = ",")
```

0.1 Exploratory Data Analysis

```
[2]: # determine data types in dataset
df.dtypes
```

```
[2]: label      int64
f1         float64
f2         float64
f3         float64
f4         float64
f5         float64
f6         float64
f7         float64
f8         float64
f9         float64
f10        float64
f11        float64
f12        float64
f13        float64
f14        float64
f15        float64
f16        float64
f17        float64
f18        float64
f19        float64
```

f20	float64
f21	float64
f22	float64
f23	float64
f24	float64
f25	float64
f26	float64
f27	float64
f28	float64
f29	float64
f30	float64
f31	float64
f32	float64
f33	float64
f34	float64
f35	float64
f36	float64
f37	float64
f38	float64
f39	float64
f40	float64
f41	float64
f42	float64
f43	float64
f44	float64
f45	float64
f46	float64
f47	float64
f48	float64
f49	float64
f50	float64
f51	float64
f52	float64
f53	float64
f54	float64
f55	float64
f56	float64
f57	float64
f58	float64
f59	float64
f60	float64
f61	float64
f62	float64
f63	float64
f64	float64
f65	float64
f66	float64

f67	float64
f68	float64
f69	float64
f70	float64
f71	float64
f72	float64
f73	float64
f74	float64
f75	float64
f76	float64
f77	float64
f78	float64
f79	float64
f80	float64
f81	float64
f82	float64
f83	float64
f84	float64
f85	float64
f86	float64
f87	float64
f88	float64
f89	float64
f90	float64
f91	float64
f92	float64
f93	float64
f94	float64
f95	float64
f96	float64
f97	float64
f98	float64
f99	float64
f100	float64
f101	float64
f102	float64
f103	float64
f104	float64
f105	float64
f106	float64
f107	float64
f108	float64
f109	float64
f110	float64
f111	float64
f112	float64
f113	float64

f114	float64
f115	float64
f116	float64
f117	float64
f118	float64
f119	float64
f120	float64
f121	float64
f122	float64
f123	float64
f124	float64
f125	float64
f126	float64
f127	float64
f128	float64
f129	float64
f130	float64
f131	float64
f132	float64
f133	float64
f134	float64
f135	float64
f136	float64
f137	int64
f138	int64
f139	int64
f140	int64
f141	int64
f142	float64
f143	float64
f144	float64
f145	float64
f146	float64
f147	float64
f148	float64
f149	float64
f150	float64
f151	float64
f152	float64
f153	float64
f154	float64
f155	float64
f156	float64
f157	float64
f158	float64
f159	float64
f160	float64

```

f161    float64
f162    float64
f163    float64
f164    float64
f165    float64
f166    float64
f167    float64
f168    float64
f169    float64
f170    float64
f171    float64
f172    float64
f173    float64
f174    float64
dtype: object

```

```

[3]: #display attributes
display(df)

```

	label	f1	f2	f3	f4	f5	f6	f7	f8	\		
0	1	-13.559	-21.407	-11.404	-15.248	-11.923	-15.291	-2.155	-7.847			
1	1	-12.802	-20.335	-10.399	-14.132	-11.096	-14.361	-2.404	-7.533			
2	1	-12.431	-19.902	-10.074	-13.598	-10.829	-14.048	-2.357	-7.472			
3	1	-12.689	-19.529	-10.028	-13.350	-11.056	-14.014	-2.661	-6.840			
4	1	-12.686	-19.278	-9.819	-13.108	-10.932	-13.939	-2.868	-6.592			
...			
325829	7	2.482	-7.687	1.079	0.743	-0.941	0.905	1.403	-10.169			
325830	7	2.523	-7.675	1.087	0.762	-0.912	0.937	1.437	-10.198			
325831	7	-1.927	-11.416	-2.435	-3.454	-4.151	-3.481	0.508	-9.489			
325832	7	0.125	-10.144	-0.622	-1.542	-2.310	-1.525	0.747	-10.269			
325833	7	0.201	-10.050	-0.599	-1.501	-2.251	-1.463	0.800	-10.251			
...			
0		f9	f10	f11	f12	f13	f14	f15	f16	f17	f18	\
0		-10.002	0.042	3.325	3.368	0.356	0.058	0.585	0.241	0.519	0.239	
1		-9.937	0.228	3.036	3.264	0.343	0.061	0.597	0.252	0.508	0.240	
2		-9.828	0.450	2.769	3.219	0.345	0.062	0.593	0.264	0.499	0.238	
3		-9.501	0.664	2.294	2.958	0.328	0.068	0.605	0.281	0.477	0.241	
4		-9.459	0.831	2.176	3.007	0.317	0.069	0.614	0.288	0.475	0.238	
...	
325829		-8.767	-0.162	-1.684	-1.846	0.549	0.053	0.398	0.368	0.250	0.382	
325830		-8.761	-0.175	-1.674	-1.848	0.551	0.053	0.396	0.367	0.250	0.383	
325831		-8.980	0.027	-0.698	-0.670	0.499	0.056	0.444	0.351	0.299	0.349	
325832		-9.522	-0.017	-0.768	-0.785	0.517	0.049	0.435	0.352	0.295	0.353	
325833		-9.451	-0.038	-0.749	-0.787	0.519	0.049	0.432	0.351	0.295	0.354	
...	
0		f19	f20	f21	f22	f23	f24	f25	f26	f27	f28	\
0		-0.624	-0.815	-0.708	-0.656	-0.196	-0.632	0.092	0.027	0.005	0.620	

1	-0.572	-0.749	-0.648	-0.628	-0.188	-0.554	0.116	0.031	0.006	0.604
2	-0.533	-0.703	-0.588	-0.618	-0.178	-0.486	0.127	0.032	0.006	0.587
3	-0.504	-0.670	-0.525	-0.612	-0.175	-0.436	0.127	0.031	0.006	0.578
4	-0.503	-0.682	-0.507	-0.591	-0.175	-0.452	0.132	0.032	0.006	0.579
...
325829	-1.581	-1.683	-0.174	-1.650	-0.040	-1.534	2.421	0.755	0.047	0.561
325830	-1.558	-1.654	-0.170	-1.624	-0.039	-1.513	2.439	0.758	0.046	0.560
325831	-2.196	-2.467	-0.167	-2.232	-0.044	-2.194	0.889	0.376	0.019	0.617
325832	-2.000	-2.368	-0.218	-2.068	-0.047	-1.974	1.391	0.570	0.032	0.615
325833	-1.952	-2.295	-0.220	-2.023	-0.048	-1.921	1.408	0.575	0.033	0.616

	f29	f30	f31	f32	f33	f34	f35	f36	f37	f38	\
0	0.684	44.369	0.424	0.196	0.260	0.120	0.055	0.162	-11.923	-12.823	
1	0.675	44.992	0.407	0.196	0.268	0.129	0.052	0.158	-11.096	-11.805	
2	0.680	45.466	0.399	0.188	0.281	0.132	0.048	0.149	-10.829	-11.377	
3	0.675	46.654	0.390	0.188	0.285	0.137	0.047	0.147	-11.056	-11.263	
4	0.665	46.828	0.385	0.194	0.280	0.141	0.049	0.152	-10.932	-11.109	
...	
325829	0.882	63.745	0.496	0.066	0.387	0.052	0.019	0.059	-0.941	3.518	
325830	0.885	63.669	0.495	0.064	0.390	0.051	0.019	0.057	-0.912	3.544	
325831	0.903	60.734	0.557	0.060	0.346	0.037	0.022	0.060	-4.151	-0.820	
325832	0.892	60.757	0.549	0.066	0.343	0.041	0.023	0.065	-2.310	1.167	
325833	0.890	60.623	0.548	0.068	0.342	0.042	0.024	0.066	-2.251	1.215	

	f39	f40	f41	f42	f43	f44	f45	f46	f47	f48	f49	\
0	-21.407	0.176	0.168	0.003	0.066	0.025	0.028	0.057	0.037	0.006	0.025	
1	-20.335	0.189	0.182	0.006	0.079	0.027	0.034	0.067	0.043	0.006	0.028	
2	-19.902	0.196	0.191	0.010	0.027	0.086	0.040	0.073	0.047	0.006	0.030	
3	-19.529	0.198	0.194	0.014	0.025	0.091	0.043	0.075	0.049	0.006	0.030	
4	-19.278	0.199	0.198	0.016	0.023	0.095	0.046	0.076	0.051	0.007	0.031	
...	
325829	-7.687	0.565	0.940	0.014	0.355	1.554	0.523	0.390	1.545	0.026	0.471	
325830	-7.675	0.606	1.014	0.016	0.412	1.827	0.587	0.457	1.815	0.034	0.520	
325831	-11.416	0.432	0.748	0.005	0.195	0.971	0.335	0.207	0.967	0.008	0.319	
325832	-10.144	0.496	0.834	0.009	0.275	1.222	0.401	0.296	1.216	0.015	0.371	
325833	-10.050	0.534	0.897	0.011	0.325	1.424	0.449	0.352	1.416	0.019	0.410	

	f50	f51	f52	f53	f54	f55	f56	f57	f58	\
0	-12.564	-23.436	-13.858	-18.209	-11.556	-17.934	1.294	-10.873	-9.578	
1	-11.559	-22.730	-12.844	-17.334	-10.500	-17.095	1.285	-11.172	-9.886	
2	-11.040	-22.566	-12.458	-16.903	-9.997	-16.873	1.418	-11.526	-10.108	
3	-10.910	-22.520	-12.126	-16.840	-9.699	-16.965	1.216	-11.610	-10.395	
4	-10.581	-22.387	-11.858	-16.623	-9.384	-16.690	1.277	-11.805	-10.529	
...	
325829	2.600	-14.506	-2.713	-1.836	0.133	-1.522	5.313	-17.106	-11.793	
325830	2.964	-14.247	-2.550	-1.510	0.414	-1.159	5.514	-17.211	-11.697	
325831	-0.670	-18.389	-6.103	-6.081	-2.222	-5.675	5.433	-17.718	-12.286	
325832	1.641	-16.098	-3.370	-2.929	-0.490	-2.667	5.011	-17.739	-12.728	

325833 2.070 -15.409 -3.201 -2.534 -0.169 -2.231 5.271 -17.479 -12.208

	f59	f60	f61	f62	f63	f64	f65	f66	f67	f68	f69	\
0	-0.275	6.652	6.378	0.548	0.045	0.407	0.149	0.691	0.159	-1.873	-0.785	
1	-0.239	6.834	6.595	0.549	0.042	0.409	0.145	0.701	0.154	-1.955	-0.753	
2	-0.031	6.907	6.876	0.558	0.039	0.403	0.145	0.710	0.146	-2.361	-0.724	
3	0.125	7.141	7.265	0.548	0.038	0.414	0.140	0.724	0.136	-2.684	-0.671	
4	0.067	7.239	7.306	0.552	0.036	0.411	0.137	0.727	0.135	-2.442	-0.662	

... ..

325829	-0.314	1.969	1.655	0.761	0.015	0.224	0.274	0.431	0.295	-0.957	-1.298	
325830	-0.351	1.924	1.573	0.769	0.015	0.216	0.275	0.428	0.298	-0.887	-1.322	
325831	-0.406	3.859	3.453	0.767	0.013	0.220	0.221	0.537	0.242	-0.689	-0.996	
325832	-0.262	2.438	2.176	0.751	0.013	0.237	0.262	0.460	0.278	-0.933	-1.263	
325833	-0.303	2.365	2.062	0.760	0.014	0.226	0.263	0.454	0.282	-0.909	-1.321	

	f70	f71	f72	f73	f74	f75	f76	f77	f78	f79	f80	\
0	-1.990	-2.997	-1.774	-1.733	0.071	0.025	0.004	0.664	0.705	42.797	0.468	
1	-1.978	-2.940	-1.828	-1.755	0.091	0.031	0.005	0.651	0.715	42.706	0.466	
2	-2.128	-3.689	-2.106	-1.668	0.102	0.034	0.005	0.638	0.721	42.195	0.460	
3	-2.298	-3.803	-2.348	-1.802	0.109	0.034	0.006	0.626	0.719	42.803	0.450	
4	-2.098	-3.860	-2.130	-1.765	0.117	0.036	0.006	0.619	0.727	42.535	0.450	

... ..

325829	-1.357	-0.989	-1.072	-0.540	1.880	0.482	0.029	0.515	0.886	26.197	0.456	
325830	-1.373	-0.913	-1.030	-0.521	2.039	0.503	0.030	0.506	0.887	25.437	0.449	
325831	-0.950	-0.703	-0.766	-0.461	0.905	0.202	0.010	0.475	0.906	26.574	0.430	
325832	-1.235	-0.982	-1.001	-0.561	1.514	0.410	0.020	0.518	0.908	27.211	0.470	
325833	-1.299	-0.946	-1.012	-0.540	1.662	0.432	0.023	0.513	0.898	26.204	0.461	

	f81	f82	f83	f84	f85	f86	f87	f88	f89	f90	\
0	0.196	0.237	0.099	0.062	0.174	-11.556	-15.740	-23.436	0.182	0.119	
1	0.185	0.249	0.099	0.057	0.163	-10.500	-14.859	-22.730	0.201	0.130	
2	0.178	0.261	0.101	0.053	0.154	-9.997	-14.509	-22.566	0.216	0.138	
3	0.176	0.269	0.105	0.051	0.150	-9.699	-14.532	-22.520	0.226	0.141	
4	0.169	0.277	0.104	0.049	0.143	-9.384	-14.269	-22.387	0.235	0.146	

... ..

325829	0.059	0.430	0.055	0.015	0.049	0.133	1.219	-14.506	0.669	0.717	
325830	0.057	0.438	0.056	0.015	0.047	0.414	1.567	-14.247	0.719	0.802	
325831	0.045	0.475	0.050	0.011	0.036	-2.222	-2.986	-18.389	0.552	0.559	
325832	0.048	0.437	0.044	0.013	0.041	-0.490	0.112	-16.098	0.614	0.655	
325833	0.052	0.437	0.049	0.014	0.044	-0.169	0.521	-15.409	0.664	0.736	

	f91	f92	f93	f94	f95	f96	f97	f98	f99	f100	f101	\
0	0.003	0.058	0.020	0.018	0.059	0.019	0.004	0.016	0.009	0.057	0.048	
1	0.002	0.072	0.023	0.020	0.073	0.023	0.004	0.019	0.001	0.044	0.037	
2	0.000	0.084	0.026	0.021	0.084	0.026	0.004	0.021	0.001	0.046	0.046	
3	0.001	0.093	0.028	0.022	0.093	0.028	0.004	0.021	0.018	0.071	0.074	
4	0.002	0.100	0.030	0.023	0.101	0.030	0.004	0.022	0.023	0.066	0.068	

... ..

325829	0.027	0.517	1.324	0.121	0.614	1.269	0.040	0.038	0.001	0.039	0.031
325830	0.031	0.597	1.636	0.137	0.709	1.577	0.050	0.033	0.046	0.088	0.069
325831	0.021	0.359	0.828	0.071	0.425	0.787	0.024	0.022	0.001	0.026	0.005
325832	0.024	0.456	1.097	0.093	0.533	1.052	0.031	0.028	0.001	0.041	0.024
325833	0.026	0.532	1.363	0.110	0.621	1.315	0.039	0.030	0.005	0.045	0.045
	f102	f103	f104	f105	f106	f107	f108	f109	f110	f111	f112 \
0	0.165	0.369	0.770	7.688	1.188	0.505	0.618	0.525	0.086	0.732	0.503
1	0.103	0.290	0.774	7.838	1.189	0.420	0.598	0.459	0.086	0.737	0.425
2	0.107	0.298	0.733	6.478	1.000	0.402	0.532	0.448	0.000	0.733	0.393
3	0.120	0.284	0.587	3.838	0.959	0.330	0.372	0.367	-0.021	0.600	0.293
4	0.131	0.264	0.590	3.882	0.971	0.327	0.401	0.353	-0.015	0.600	0.280
...
325829	0.113	0.309	0.818	9.968	1.258	0.467	0.670	0.496	0.114	0.776	0.482
325830	0.117	0.222	0.526	3.217	1.275	0.296	0.414	0.290	0.121	0.432	0.250
325831	0.096	0.408	0.976	81.600	5.200	0.704	0.957	0.662	0.677	0.880	0.914
325832	0.107	0.278	0.841	11.583	1.708	0.449	0.711	0.475	0.262	0.743	0.484
325833	0.097	0.229	0.672	5.089	1.000	0.315	0.459	0.357	0.000	0.672	0.289
	f113	f114	f115	f116	f117	f118	f119	f120	f121	f122 \	
0	0.382	2.236	-0.486	17.280	0.549	0.128	7.380	3.438	22.667	8.222	
1	0.476	2.816	-0.401	16.240	0.471	0.099	4.240	2.784	18.556	0.469	
2	0.472	2.785	-0.399	16.580	0.399	0.098	3.660	2.326	19.000	0.667	
3	0.406	2.367	-0.257	14.270	0.237	0.090	2.640	1.622	17.333	1.556	
4	0.337	2.015	-0.330	11.320	0.317	0.114	3.700	1.927	16.667	0.444	
...	
325829	0.464	2.735	-0.487	16.900	0.569	0.084	5.240	3.645	22.222	24.395	
325830	0.310	1.897	-0.141	9.160	0.258	0.115	3.640	1.696	15.444	25.136	
325831	0.619	4.250	-0.574	27.380	0.901	-0.533	6.300	19.200	23.111	3.877	
325832	0.444	2.598	-0.446	14.730	0.634	0.072	5.660	4.458	19.111	4.543	
325833	0.405	2.361	-0.366	11.360	0.366	0.089	3.120	2.156	14.667	10.222	
	f123	f124	f125	f126	f127	f128	f129	f130	f131	f132	f133 \
0	0.275	13.444	3.000	2.197	0.111	0.453	7.778	17.284	0.276	24.556	3.667
1	0.600	1.333	0.889	1.523	0.235	-0.132	3.889	0.321	0.722	0.556	0.556
2	0.578	2.444	1.111	1.677	0.210	0.317	3.889	0.321	0.667	0.667	0.667
3	0.400	2.000	1.333	1.523	0.235	0.081	6.222	1.506	0.533	2.000	1.111
4	0.611	0.778	0.778	1.273	0.309	0.811	6.556	0.914	0.600	1.333	0.889
...
325829	0.156	40.667	5.333	1.677	0.210	0.380	3.778	0.173	0.833	0.333	0.333
325830	0.116	32.000	4.444	1.889	0.160	0.015	6.556	6.469	0.242	16.667	3.333
325831	0.429	3.778	1.556	2.043	0.136	0.791	2.222	0.395	0.722	0.556	0.556
325832	0.320	7.667	2.333	1.889	0.160	0.039	3.222	0.617	0.667	0.667	0.667
325833	0.144	18.444	3.778	1.889	0.160	-0.300	6.111	11.210	0.199	29.667	4.556
	f134	f135	f136	f137	f138	f139	f140	f141	f142	f143	f144 \
0	2.043	0.136	0.651	6711	6143	4570	5064	8212	0.285	1.797	1.344
1	1.273	0.309	0.105	6274	5084	3297	3777	8214	0.427	2.491	1.542

2	1.273	0.309	0.614	6215	5035	3033	3837	8588	0.478	2.832	1.660
3	1.889	0.160	0.064	6836	5745	4212	4534	7691	0.292	1.826	1.364
4	1.581	0.210	0.658	6682	5883	4434	4627	7072	0.229	1.595	1.327
...
325829	0.849	0.506	-0.189	5960	5510	3023	4711	9327	0.510	3.085	1.823
325830	2.043	0.136	0.823	7649	6544	4690	4823	7104	0.205	1.515	1.395
325831	1.677	0.210	0.561	6004	5343	3444	4255	8071	0.402	2.344	1.551
325832	1.523	0.235	0.620	6234	5469	3442	4472	6766	0.326	1.966	1.589
325833	2.197	0.111	0.704	7980	6950	5297	5004	6458	0.099	1.219	1.312

	f145	f146	f147	f148	f149	f150	f151	f152	f153	f154	\
0	-0.619	0.543	0.427	0.147	0.144	0.586	0.237	1.622	0.237	294110.000	
1	-0.645	0.925	0.641	0.213	0.235	0.751	0.370	2.175	0.370	412400.000	
2	-0.700	1.035	0.717	0.248	0.261	0.809	0.382	2.238	0.382	439570.000	
3	-0.475	0.658	0.438	0.154	0.145	0.601	0.258	1.696	0.258	296240.000	
4	-0.401	0.528	0.344	0.140	0.092	0.536	0.209	1.528	0.209	232610.000	
...
325829	-0.914	0.982	0.766	0.291	0.257	0.868	0.329	1.980	0.329	423430.000	
325830	-0.273	0.608	0.307	0.165	0.041	0.560	0.191	1.473	0.191	222500.000	
325831	-0.710	0.803	0.603	0.216	0.203	0.745	0.310	1.897	0.310	354320.000	
325832	-0.430	0.825	0.488	0.227	0.106	0.734	0.204	1.513	0.204	216430.000	
325833	-0.134	0.424	0.148	0.135	-0.037	0.411	0.127	1.291	0.127	150320.000	

	f155	f156	f157	f158	f159	f160	f161	f162	f163	f164	\
0	0.051	2199.400	92560.000	1.108	48.444	3.136	0.578	2.444	1.111	1.831	
1	0.068	2338.400	100280.000	1.146	49.778	0.395	0.467	1.333	1.111	1.215	
2	0.117	3321.300	128320.000	1.265	49.444	0.247	0.611	0.778	0.778	1.369	
3	0.037	1748.100	80640.000	1.076	50.667	0.889	0.611	0.778	0.778	1.889	
4	0.021	1365.400	69540.000	1.044	51.222	0.395	0.567	1.667	1.000	1.735	
...
325829	0.218	5811.100	200760.000	1.558	47.000	9.556	0.367	10.556	2.556	2.043	
325830	0.014	1460.900	82140.000	1.028	49.667	2.222	0.371	4.889	1.778	2.043	
325831	0.105	3239.500	124620.000	1.236	48.778	6.617	0.341	9.000	2.333	2.043	
325832	0.130	3867.200	142880.000	1.299	51.556	0.247	0.778	0.444	0.444	1.215	
325833	-0.028	224.020	48540.000	0.945	51.333	1.333	0.495	3.111	1.333	2.197	

	f165	f166	f167	f168	f169	f170	f171	f172	f173	f174	
0	0.185	0.726	5.333	6.000	0.295	9.778	2.444	1.677	0.210	0.654	
1	0.333	-0.488	2.111	0.099	0.833	0.333	0.333	0.849	0.506	-0.189	
2	0.259	0.253	2.222	0.173	0.689	0.889	0.667	1.273	0.309	0.105	
3	0.160	0.438	4.111	0.321	0.833	0.333	0.333	1.149	0.383	0.416	
4	0.185	0.350	4.000	0.444	0.689	0.889	0.667	1.581	0.210	0.500	
...
325829	0.136	0.761	2.333	0.222	0.889	0.222	0.222	1.061	0.358	0.632	
325830	0.136	-0.384	6.000	10.444	0.127	19.556	4.000	2.197	0.111	0.812	
325831	0.136	0.755	2.333	0.222	0.833	0.333	0.333	1.273	0.309	0.316	
325832	0.333	0.060	3.000	0.000	0.944	0.111	0.111	0.349	0.802	0.189	
325833	0.111	0.168	6.667	11.111	0.356	14.333	3.000	2.043	0.136	0.727	

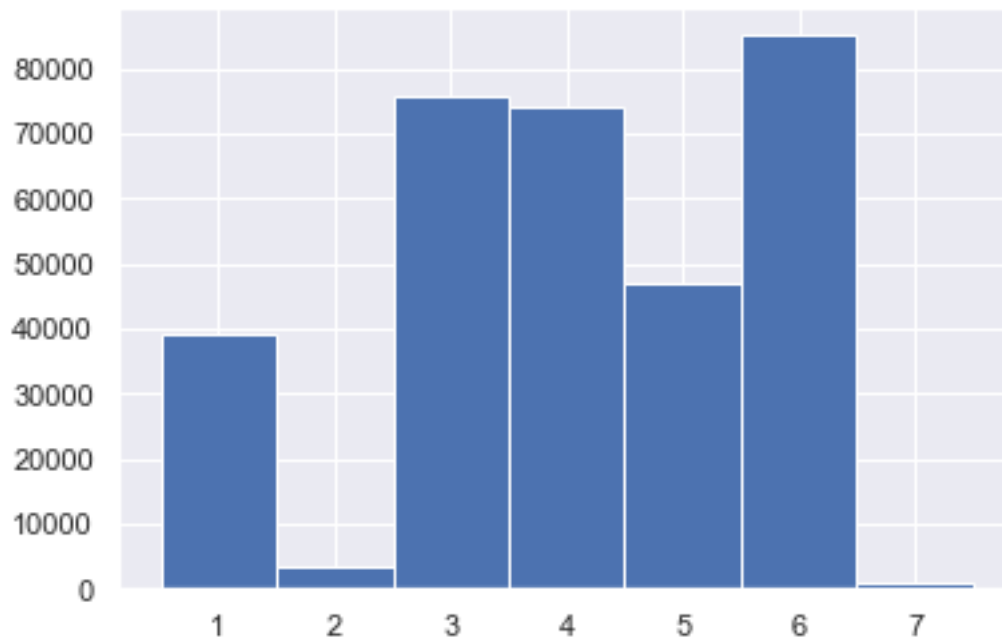
[325834 rows x 175 columns]

```
[4]: # identifies columns having null values as a list
df.columns[df.isna().any()].tolist()
```

[4]: []

```
[5]: # plot histogram of label column
plt.hist(df['label'], bins=np.arange(df['label'].min(), df['label'].max()+2)-0.
↪5)
```

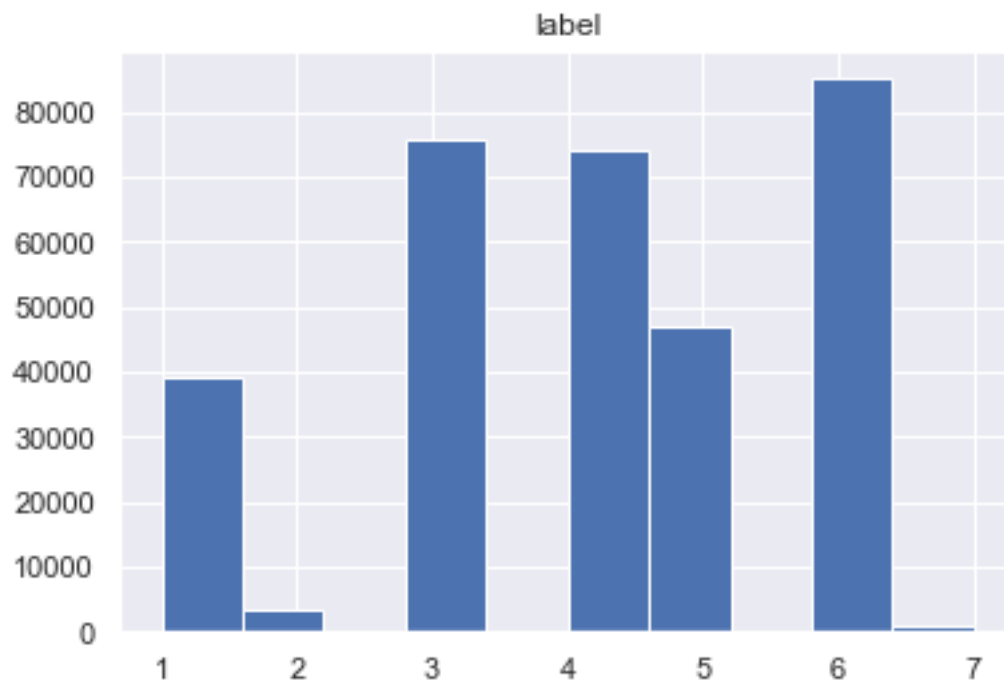
```
[5]: (array([39162., 3598., 75673., 74067., 47117., 85074., 1143.]),
array([0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5]),
<BarContainer object of 7 artists>)
```



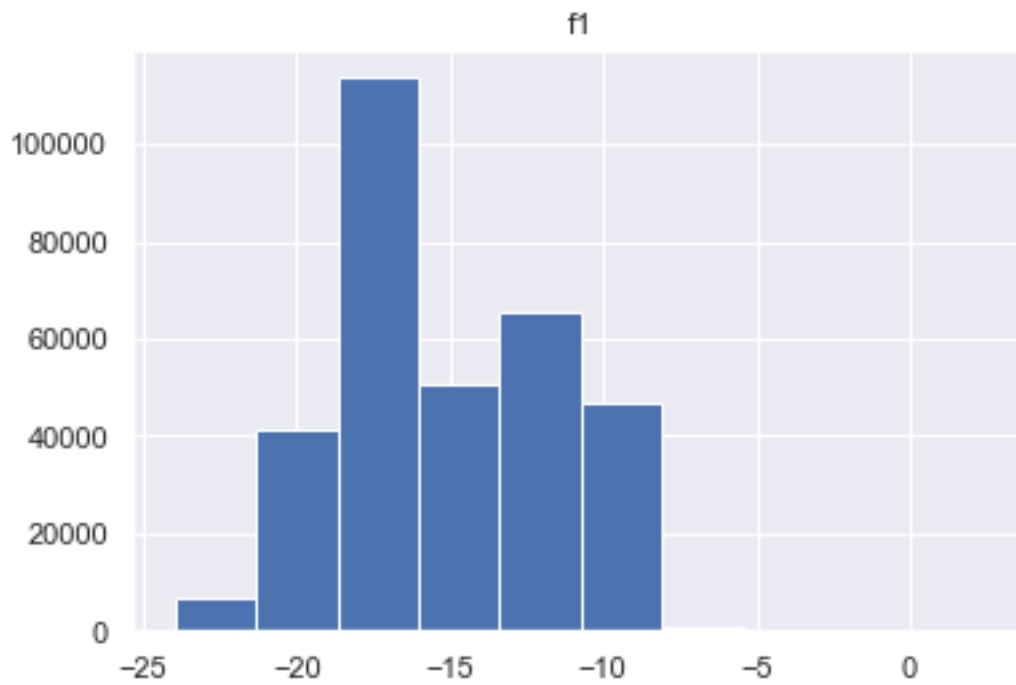
```
[6]: # histograms of dataset attributes
for column in df:
    plt.figure()
    df.hist([column])
```

<ipython-input-6-16d84b1e181f>:3: RuntimeWarning: More than 20 figures have been opened. Figures created through the pyplot interface (`matplotlib.pyplot.figure`) are retained until explicitly closed and may consume too much memory. (To control this warning, see the rcParam `figure.max_open_warning`).

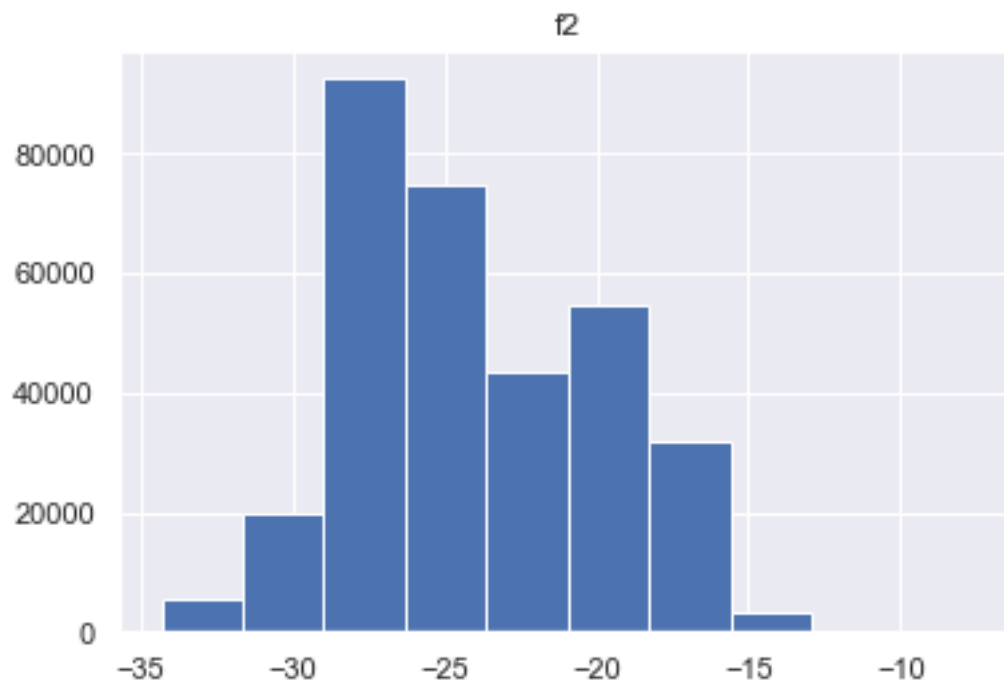
```
plt.figure()  
<Figure size 432x288 with 0 Axes>
```



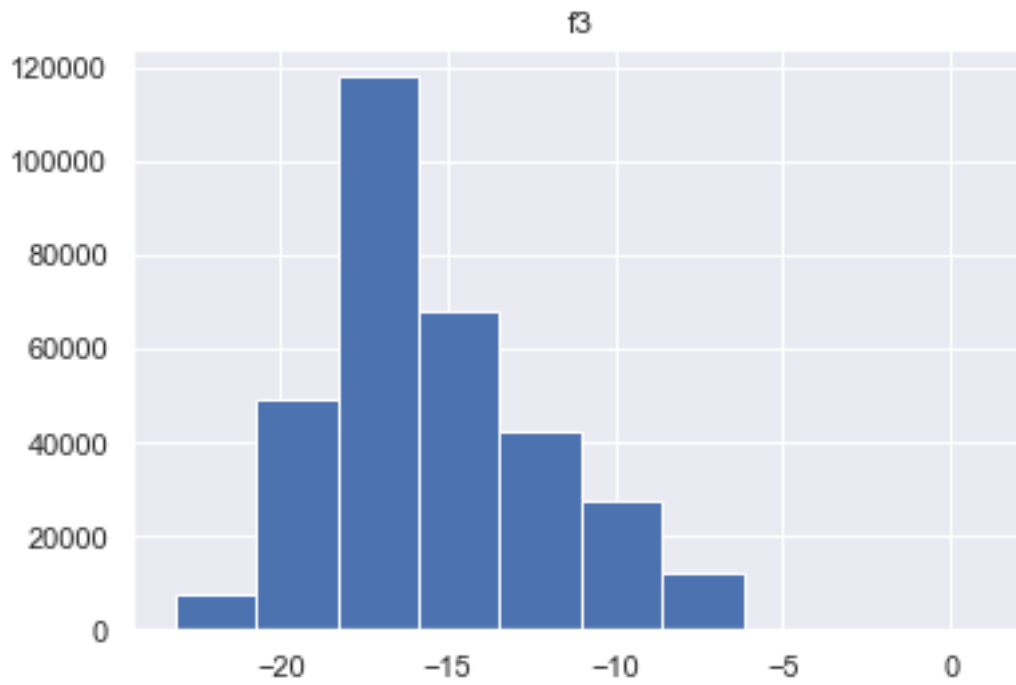
```
<Figure size 432x288 with 0 Axes>
```



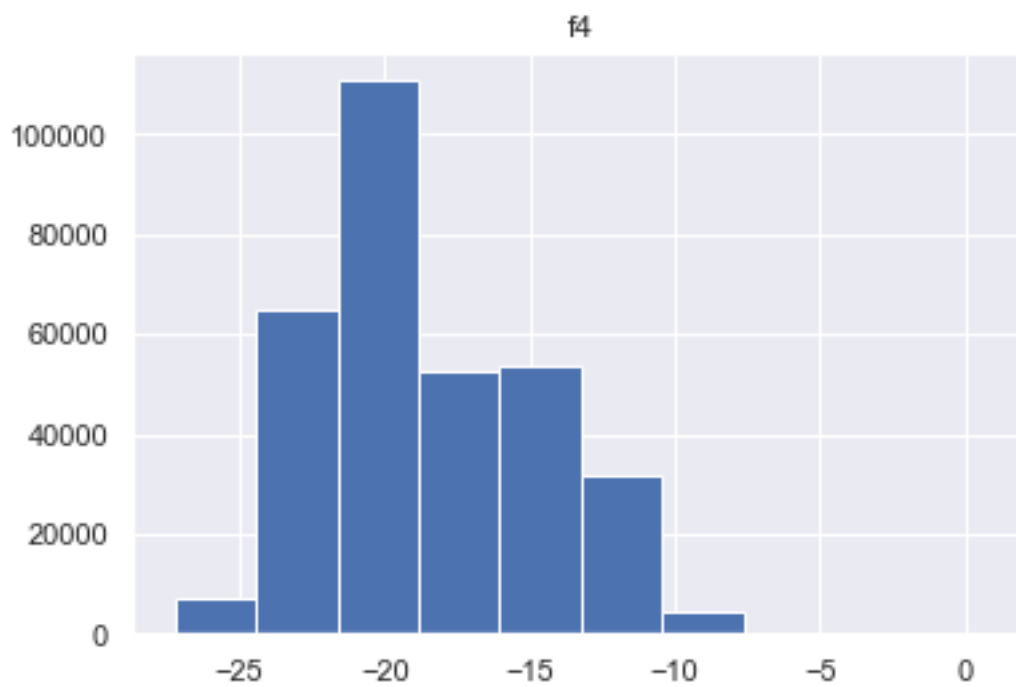
<Figure size 432x288 with 0 Axes>



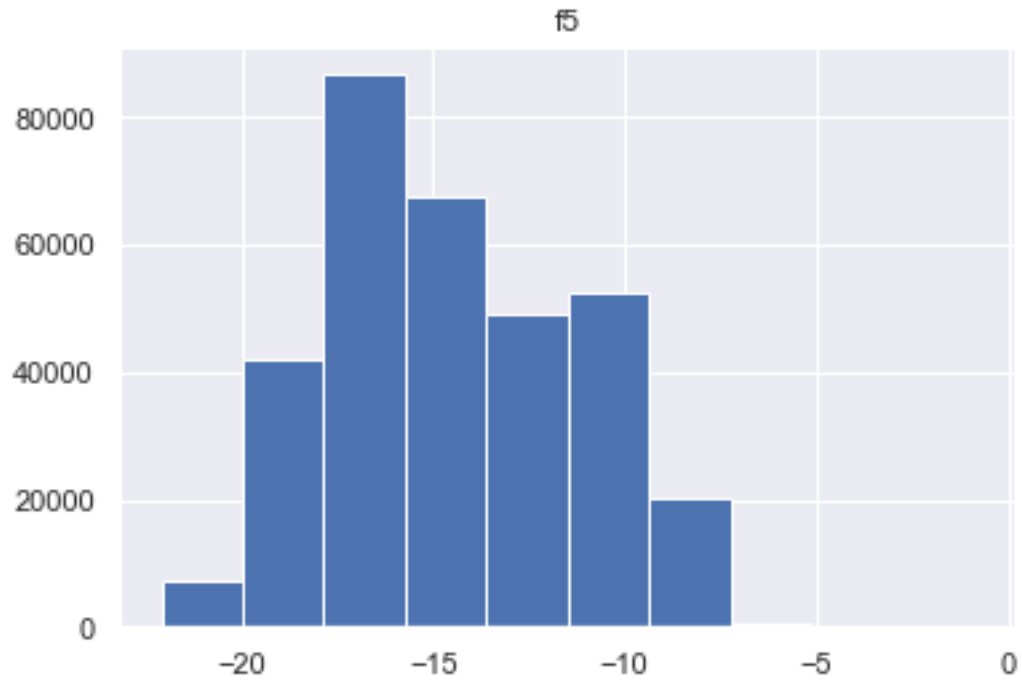
<Figure size 432x288 with 0 Axes>



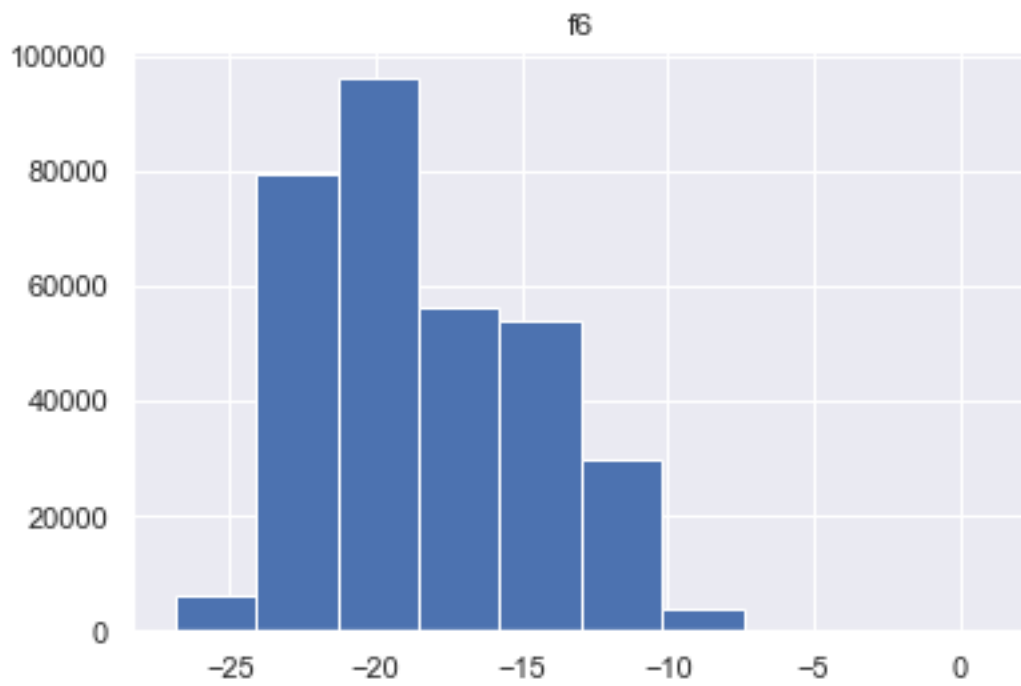
<Figure size 432x288 with 0 Axes>



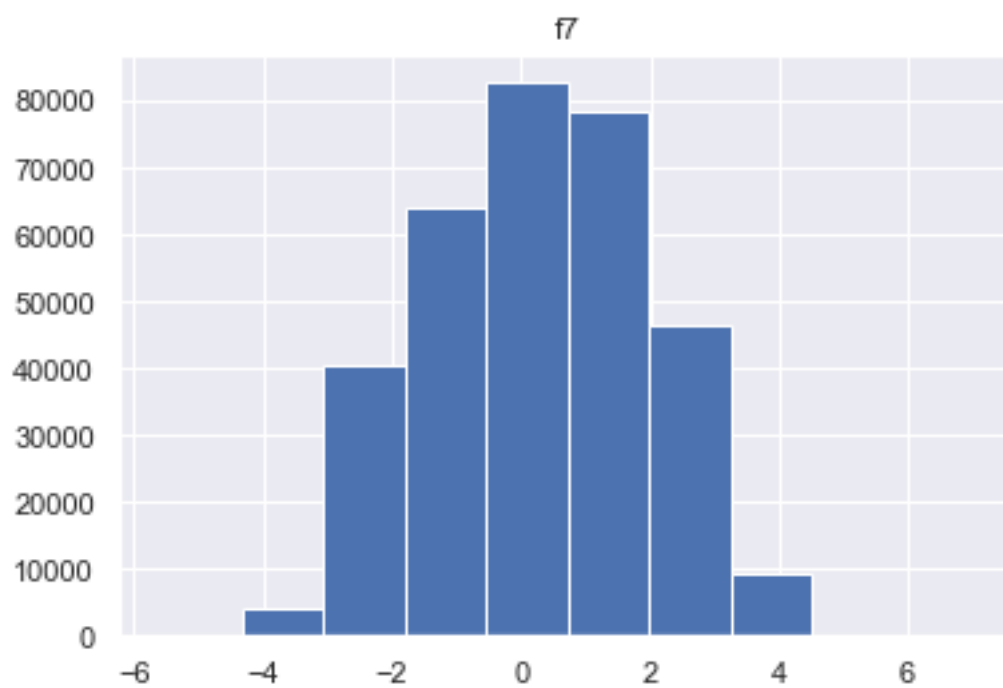
<Figure size 432x288 with 0 Axes>



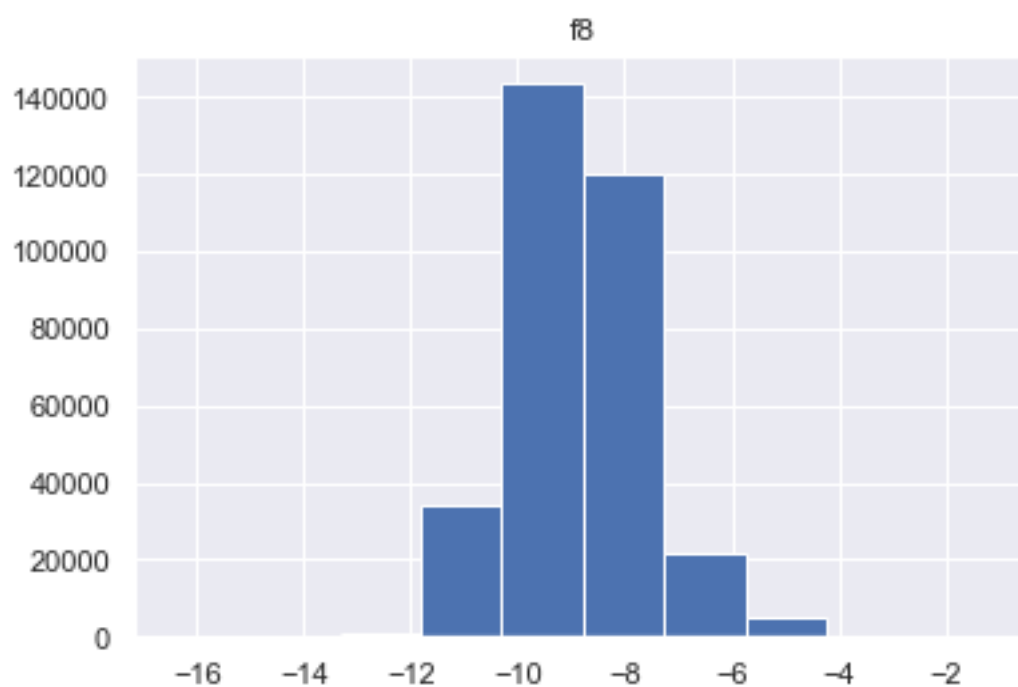
<Figure size 432x288 with 0 Axes>



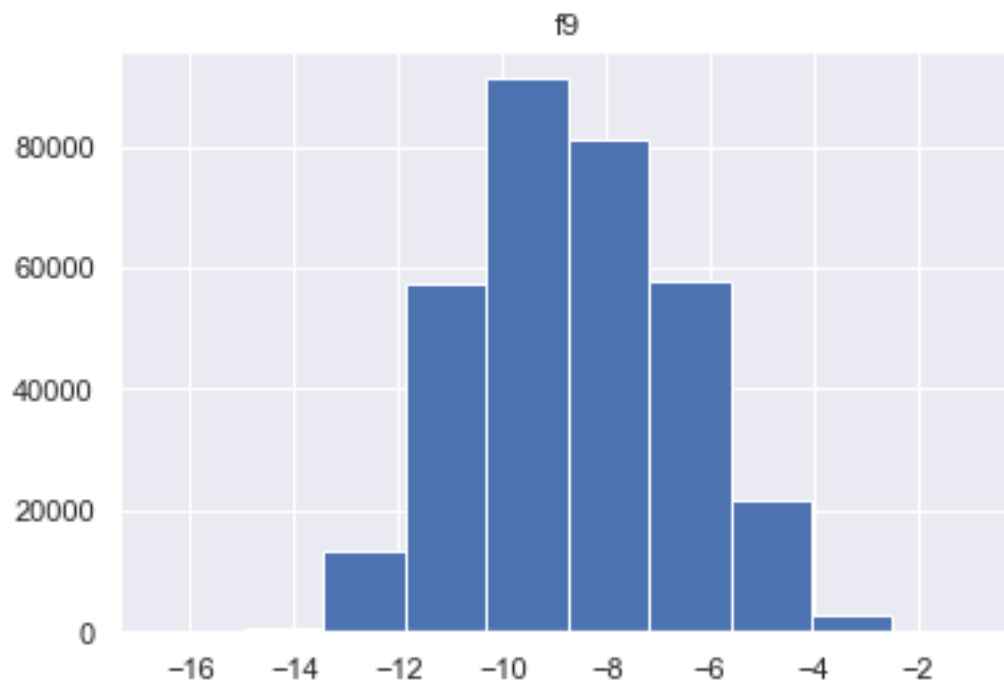
<Figure size 432x288 with 0 Axes>



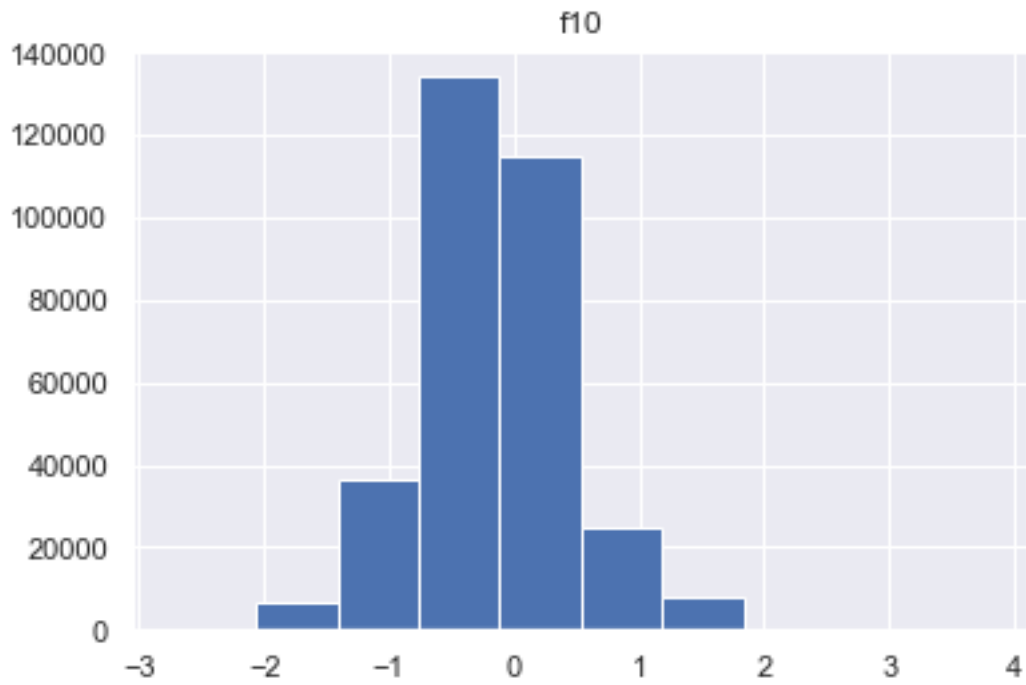
<Figure size 432x288 with 0 Axes>



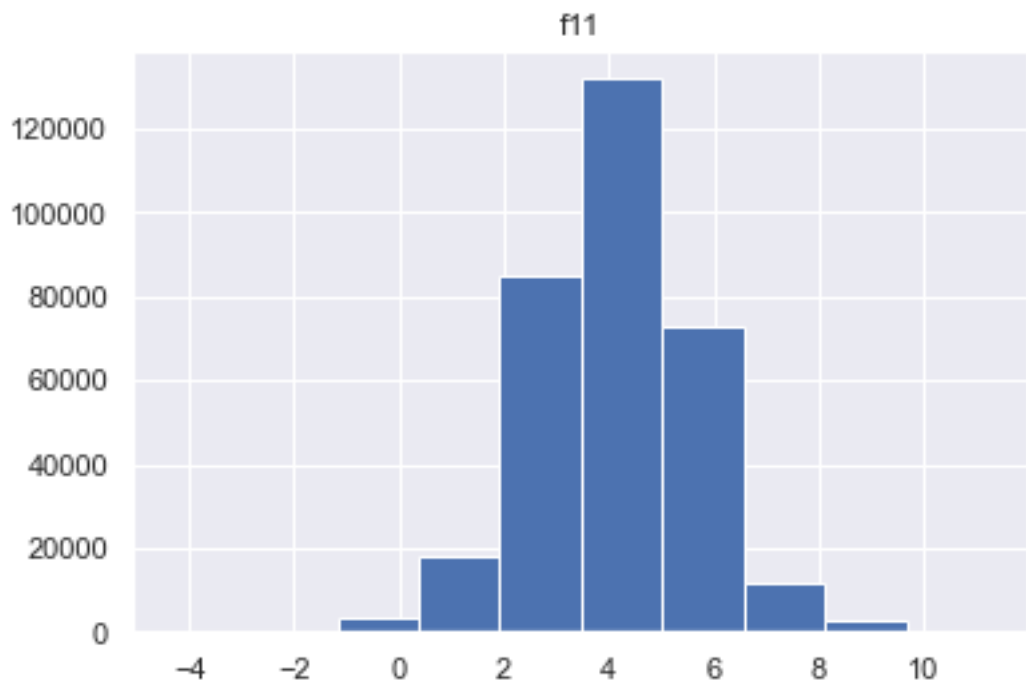
<Figure size 432x288 with 0 Axes>



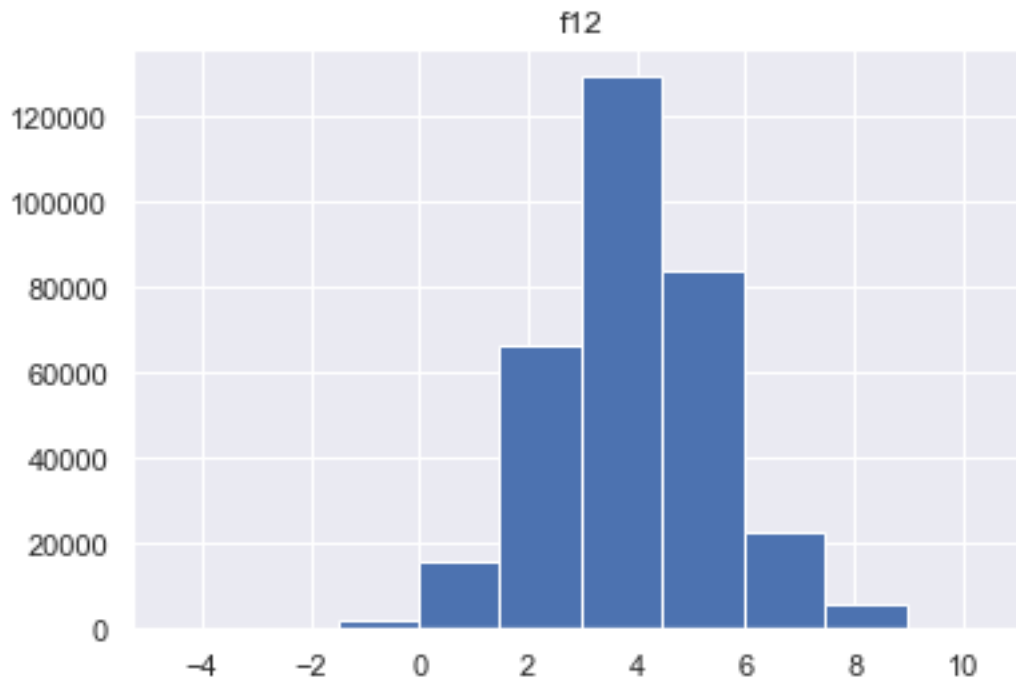
<Figure size 432x288 with 0 Axes>



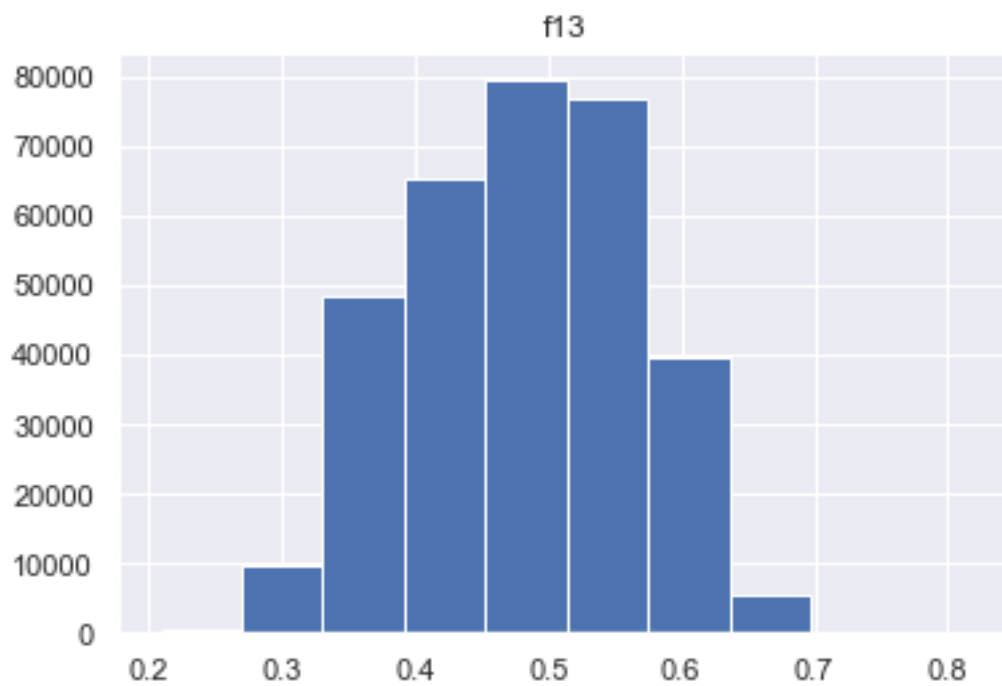
<Figure size 432x288 with 0 Axes>



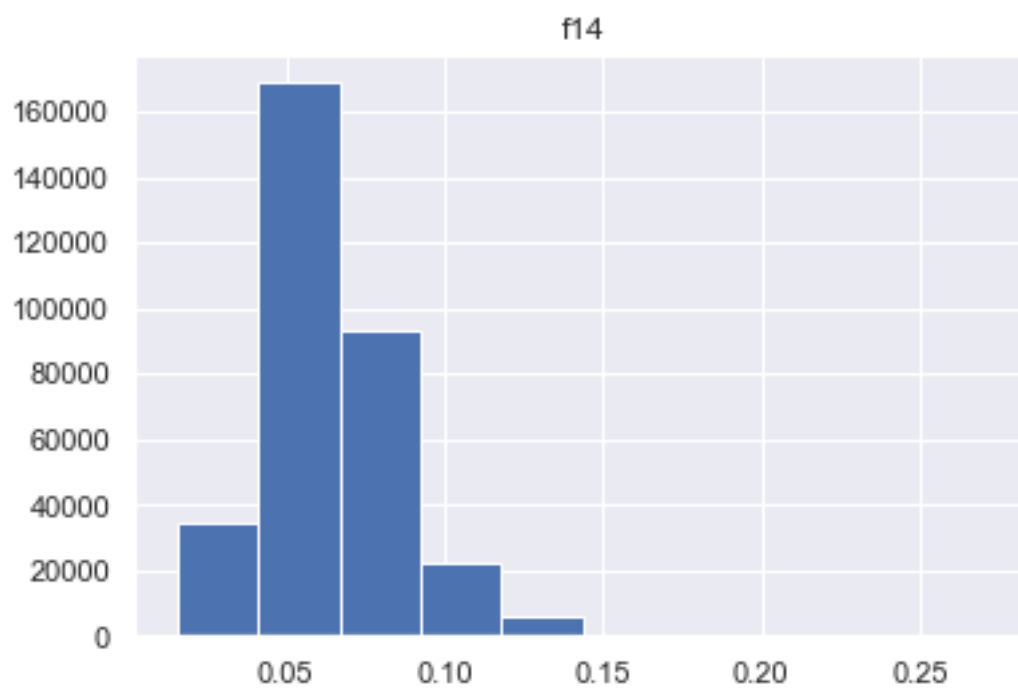
<Figure size 432x288 with 0 Axes>



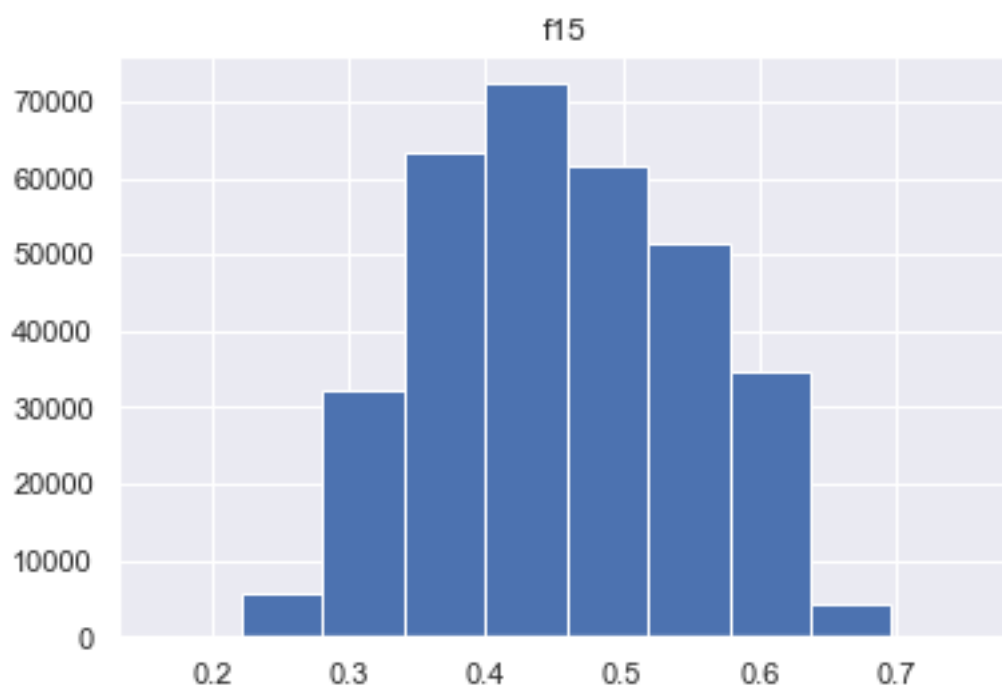
<Figure size 432x288 with 0 Axes>



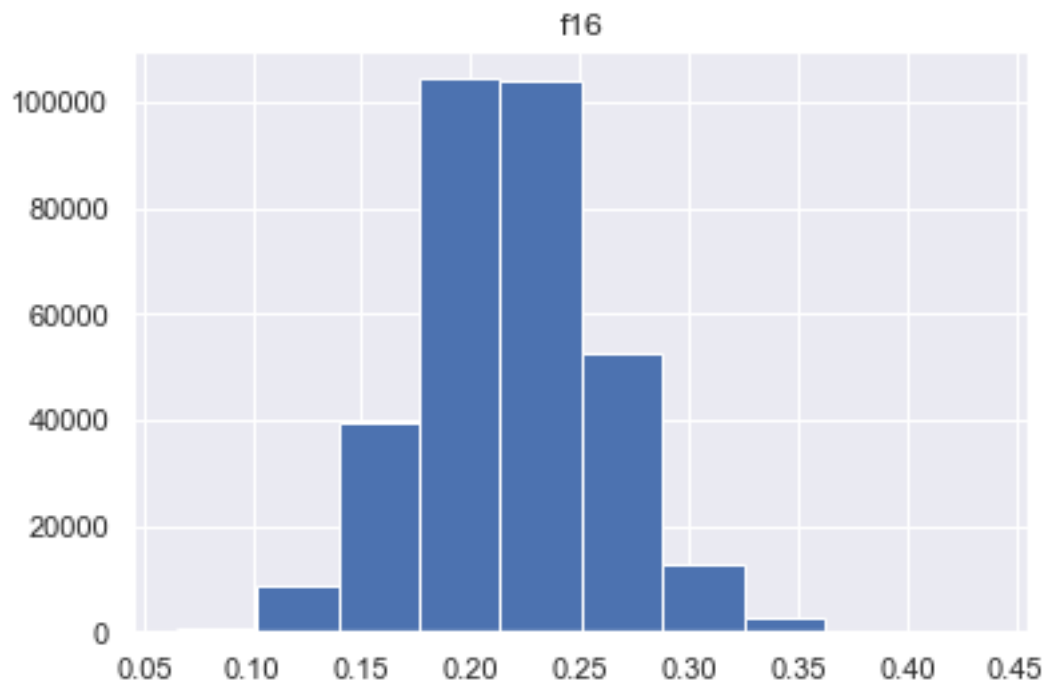
<Figure size 432x288 with 0 Axes>



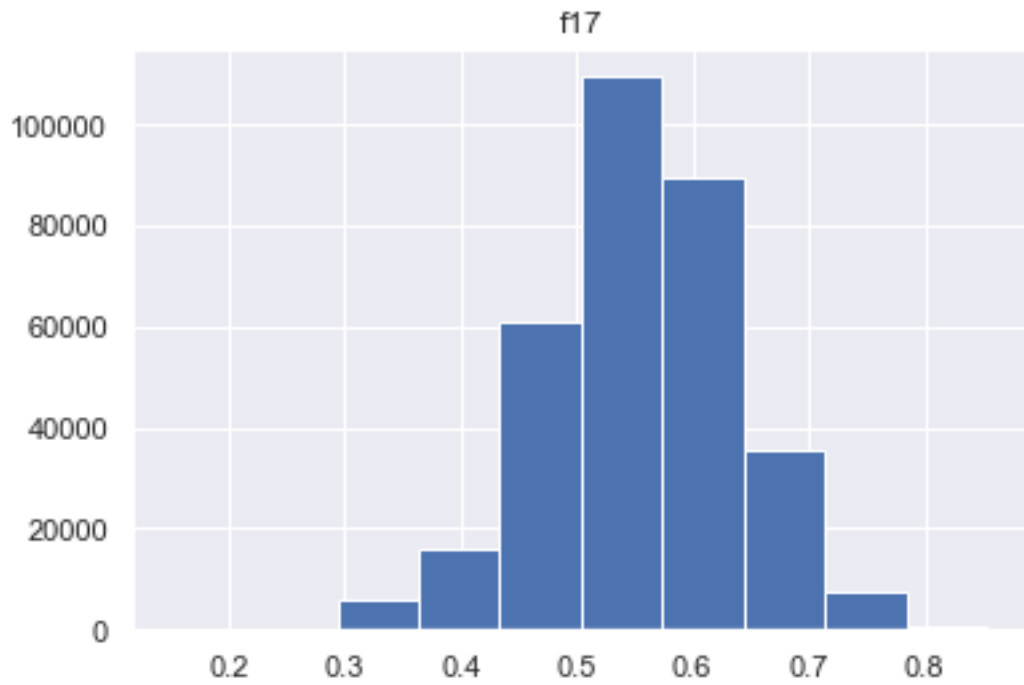
<Figure size 432x288 with 0 Axes>



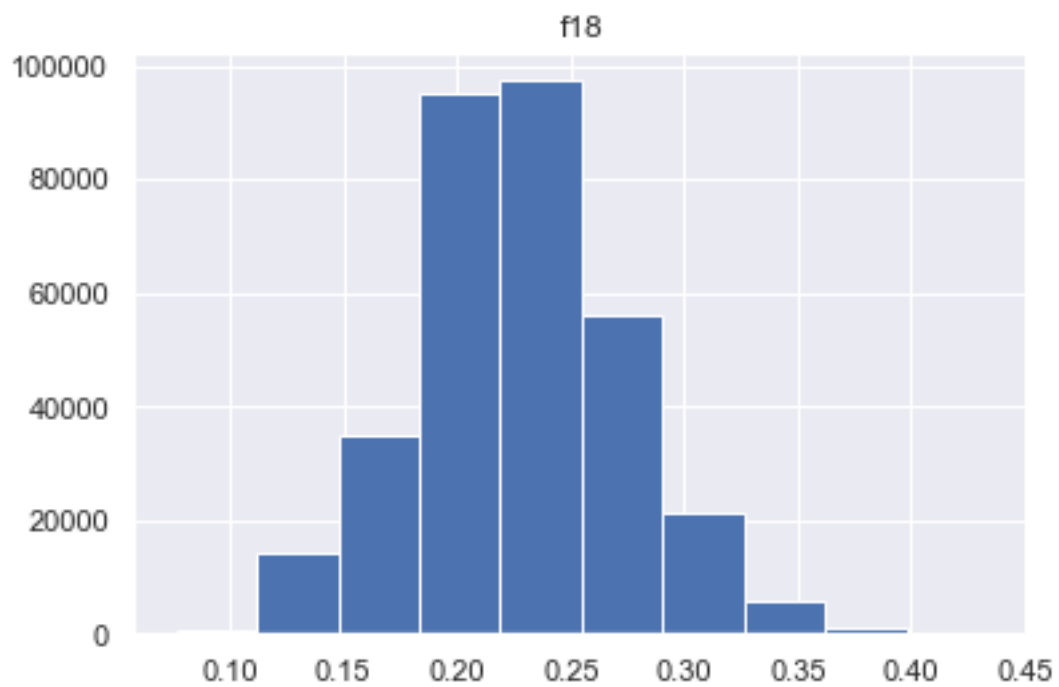
<Figure size 432x288 with 0 Axes>



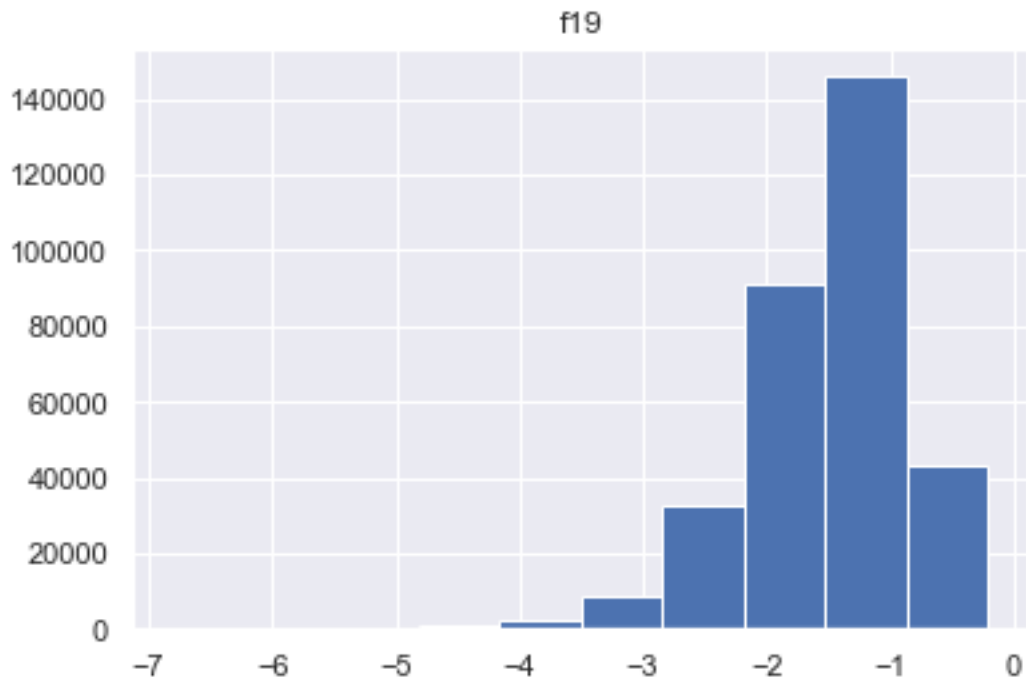
<Figure size 432x288 with 0 Axes>



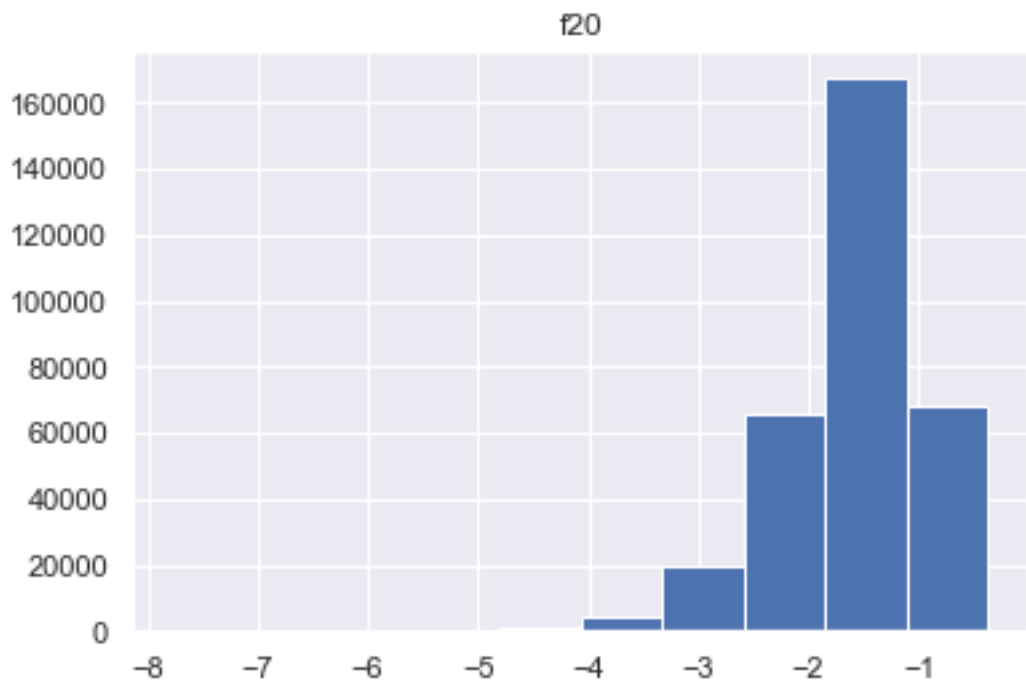
<Figure size 432x288 with 0 Axes>



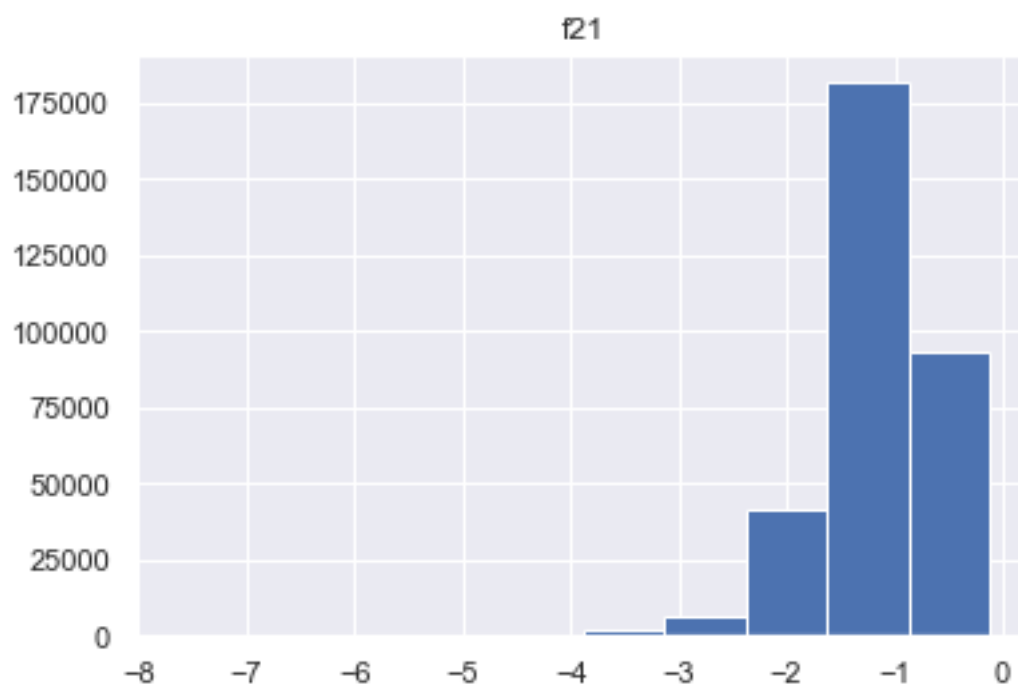
<Figure size 432x288 with 0 Axes>



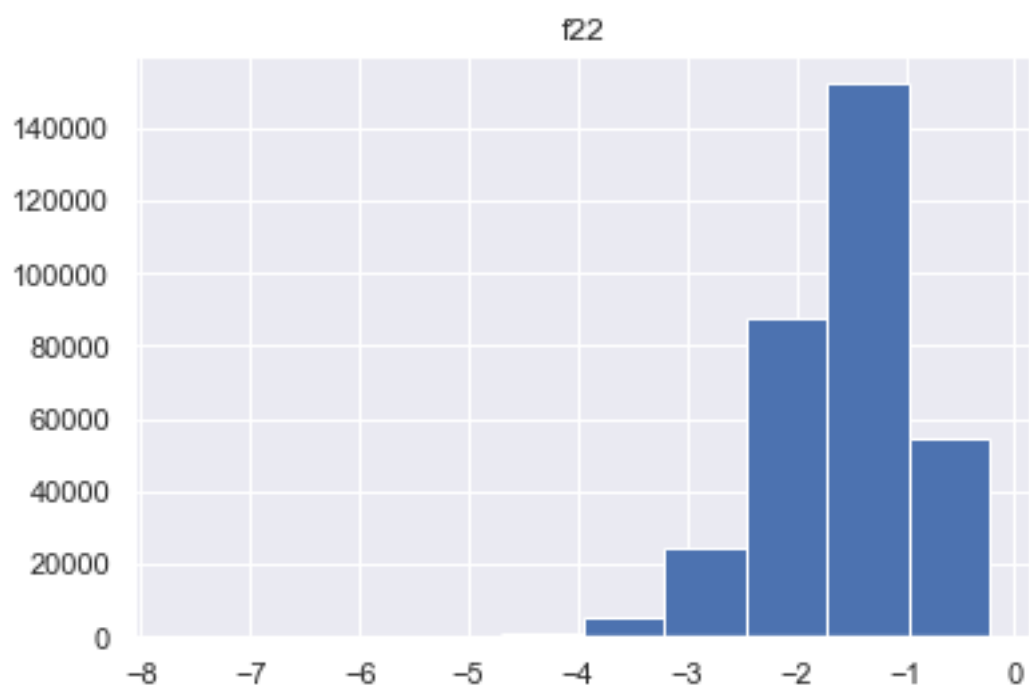
<Figure size 432x288 with 0 Axes>



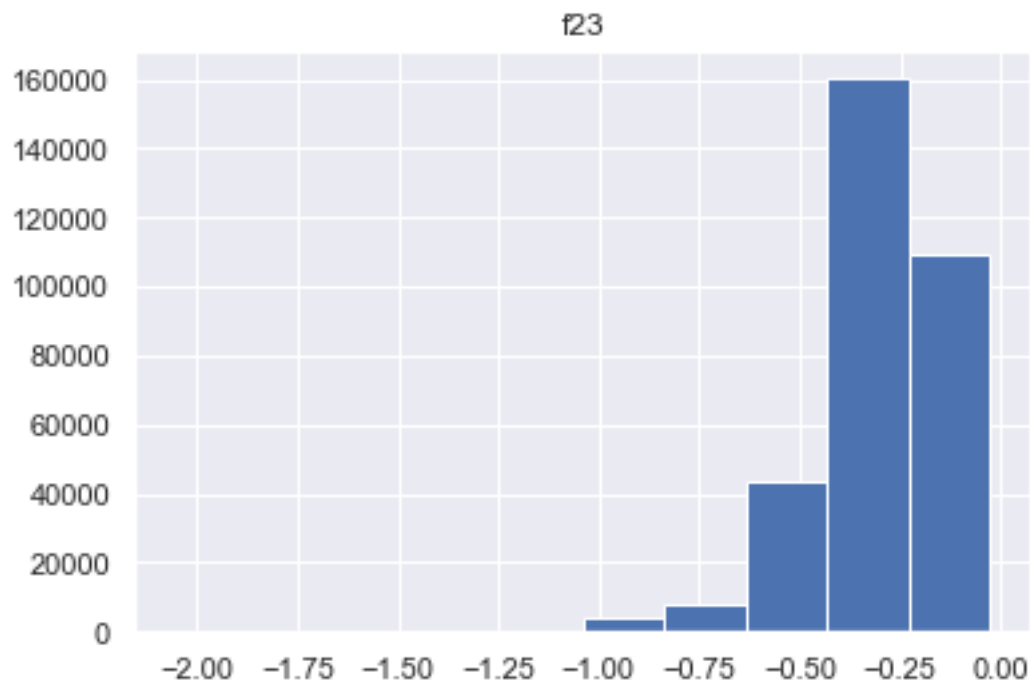
<Figure size 432x288 with 0 Axes>



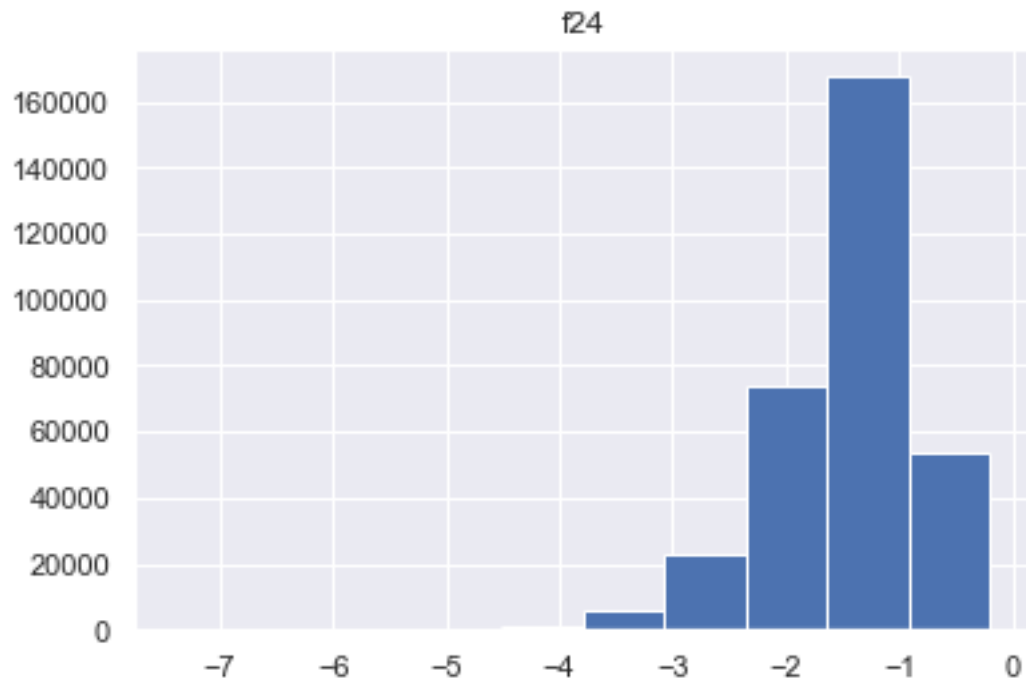
<Figure size 432x288 with 0 Axes>



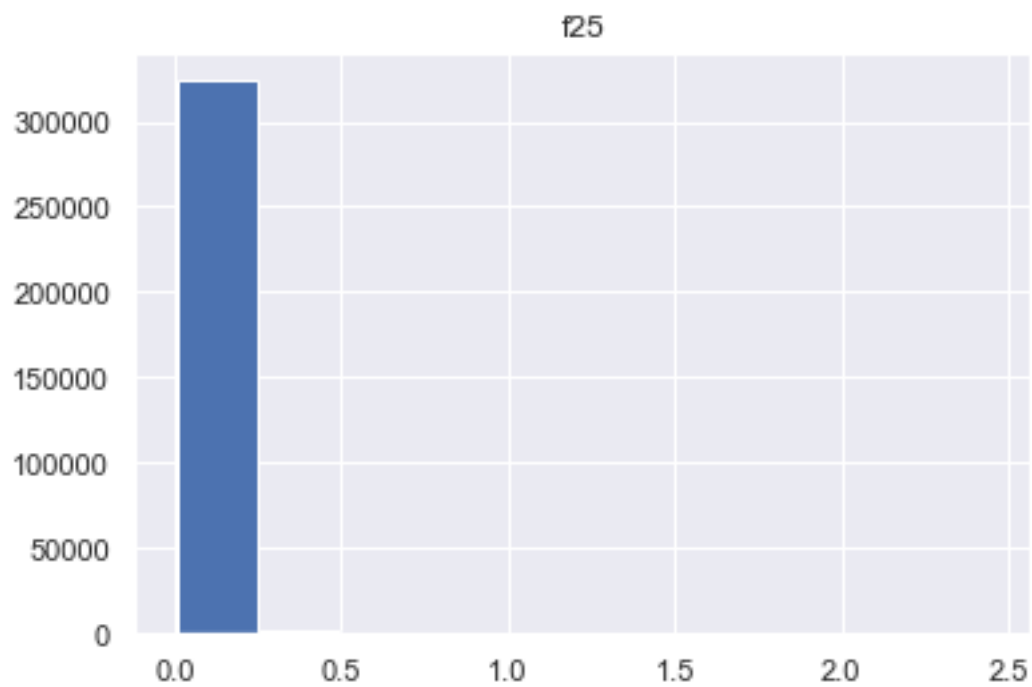
<Figure size 432x288 with 0 Axes>



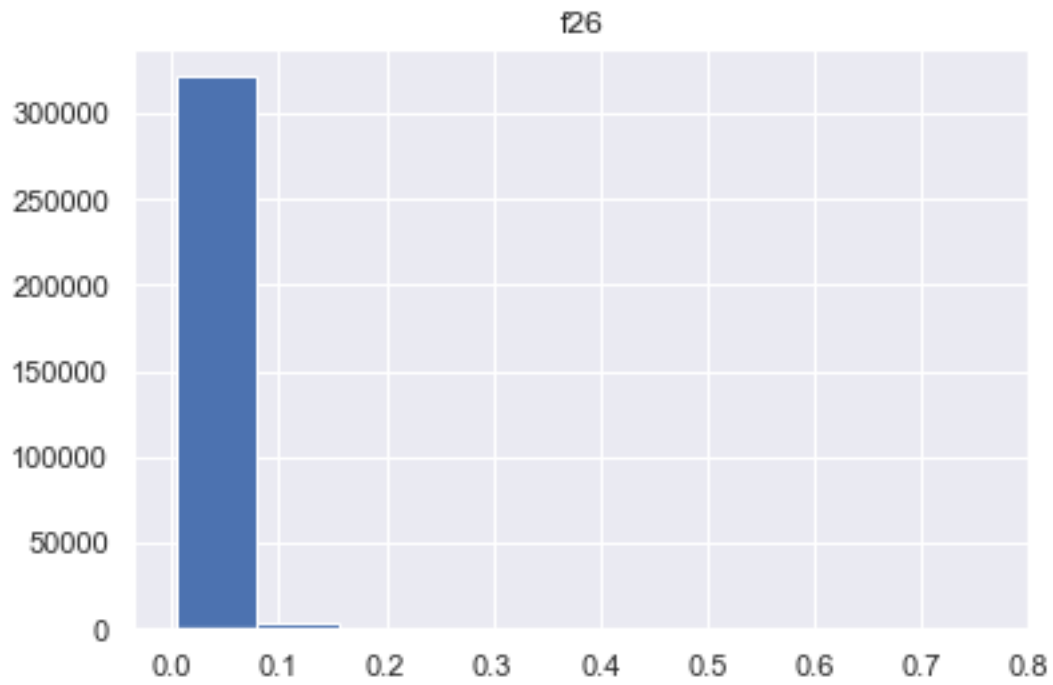
<Figure size 432x288 with 0 Axes>



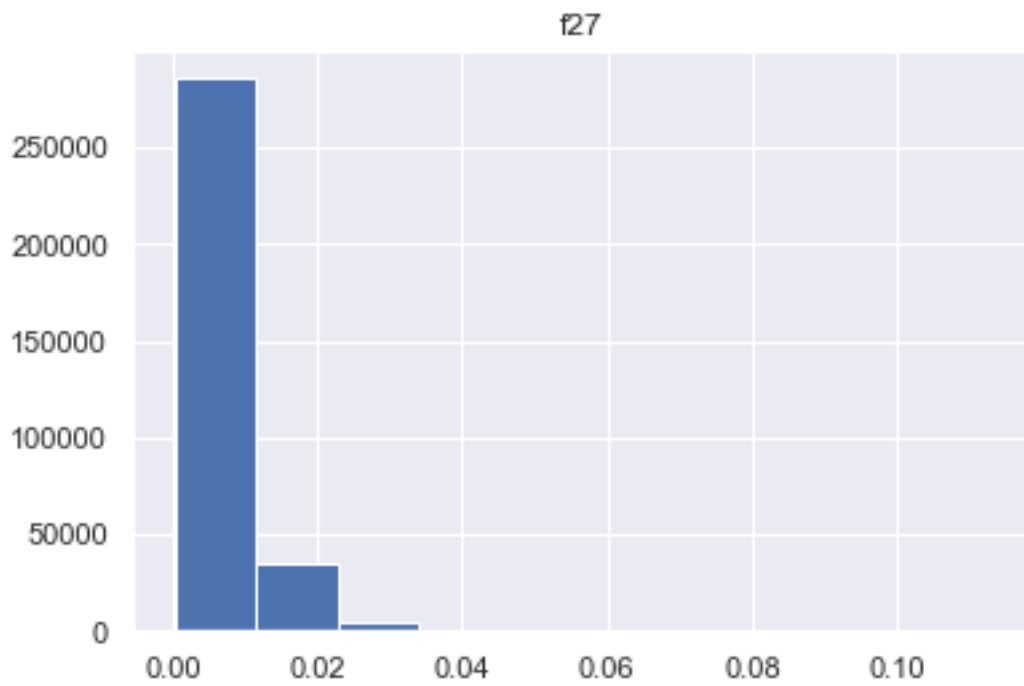
<Figure size 432x288 with 0 Axes>



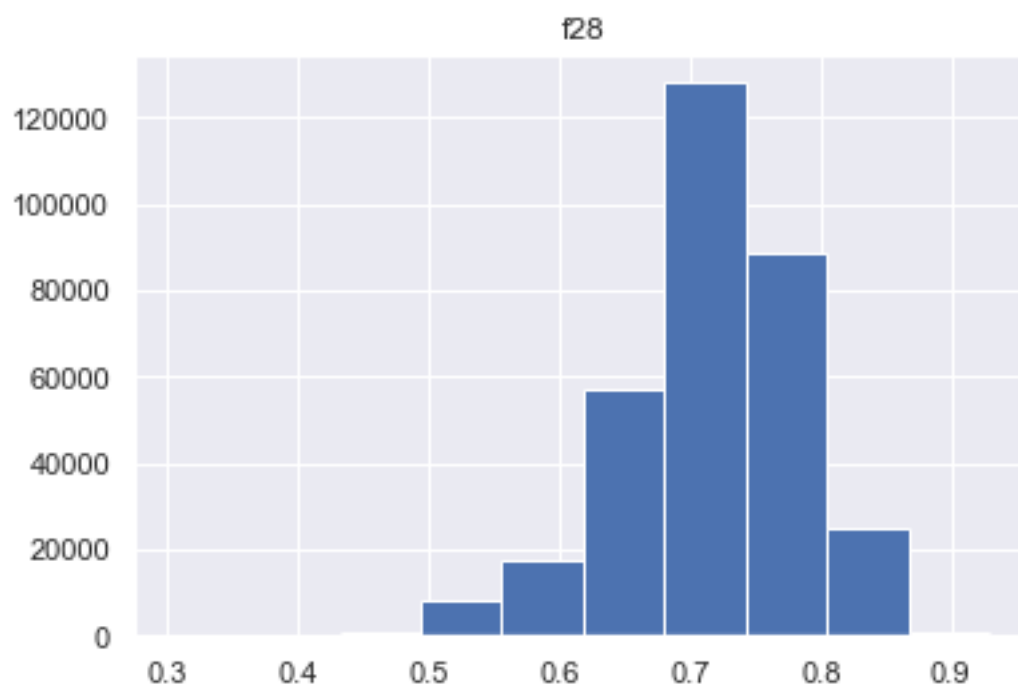
<Figure size 432x288 with 0 Axes>



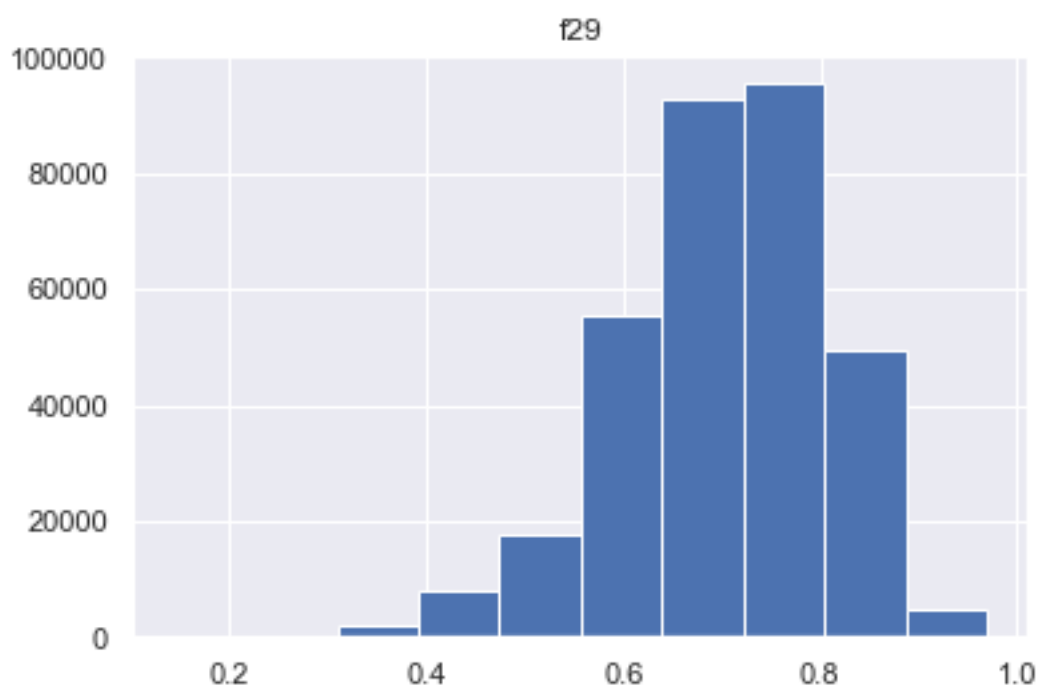
<Figure size 432x288 with 0 Axes>



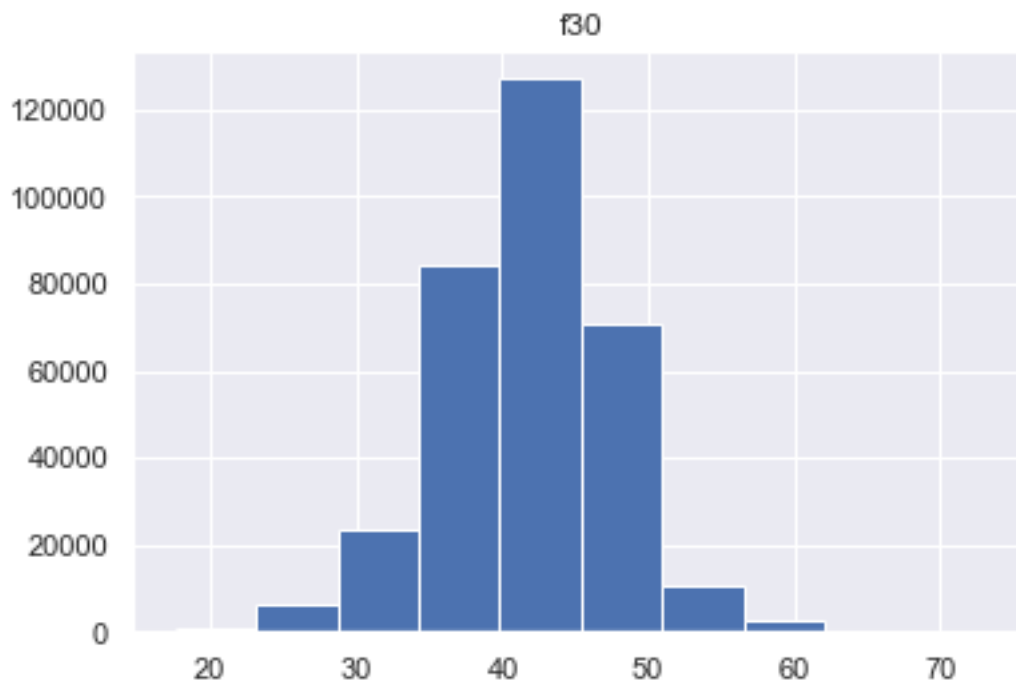
<Figure size 432x288 with 0 Axes>



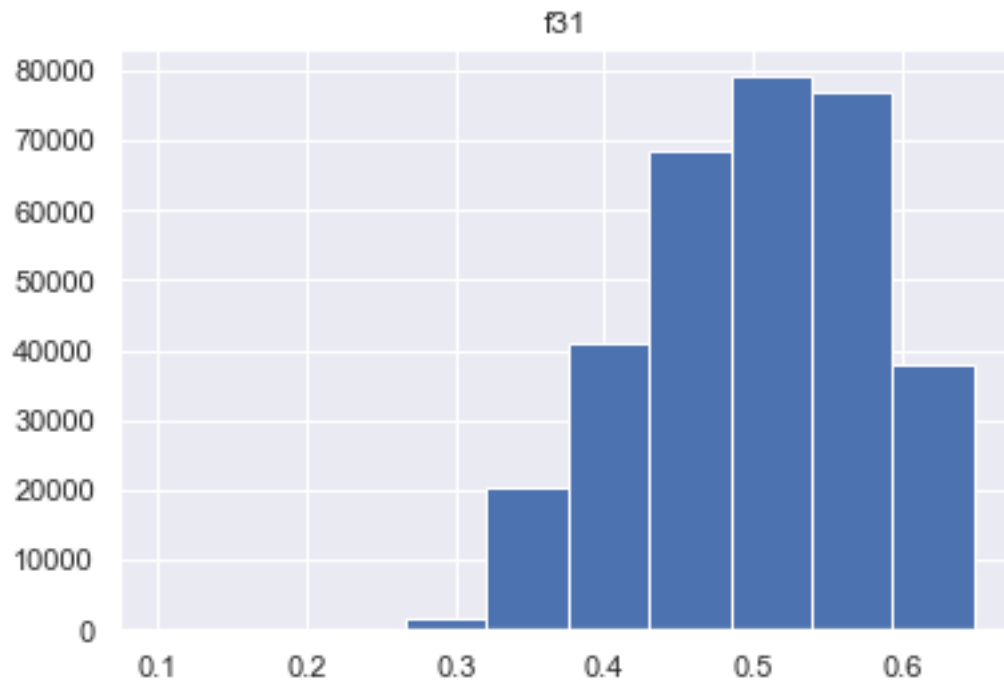
<Figure size 432x288 with 0 Axes>



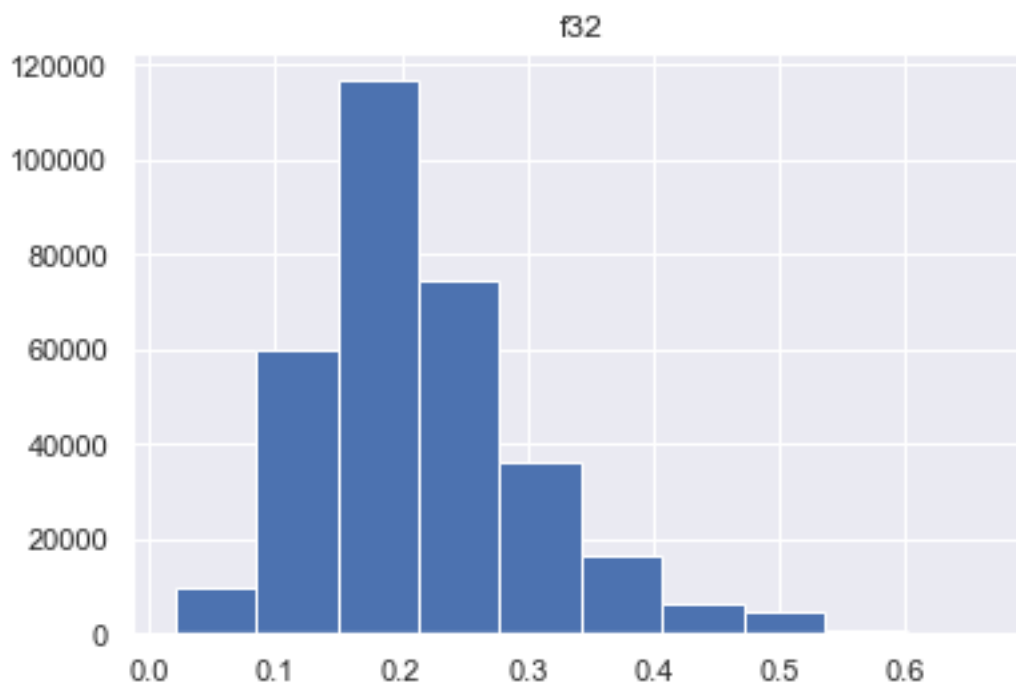
<Figure size 432x288 with 0 Axes>



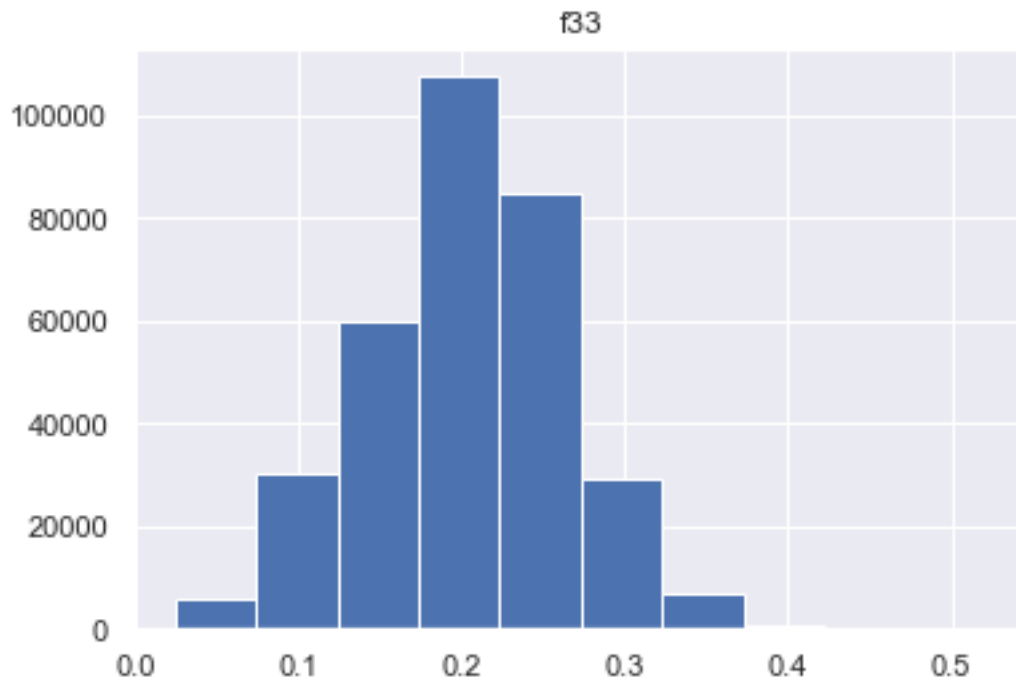
<Figure size 432x288 with 0 Axes>



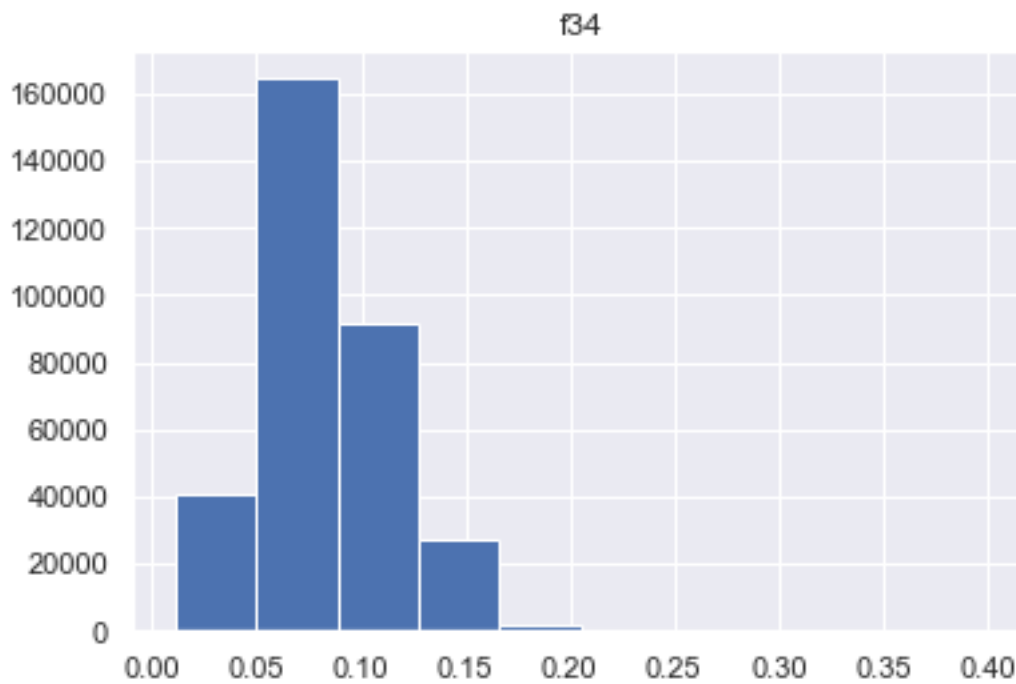
<Figure size 432x288 with 0 Axes>



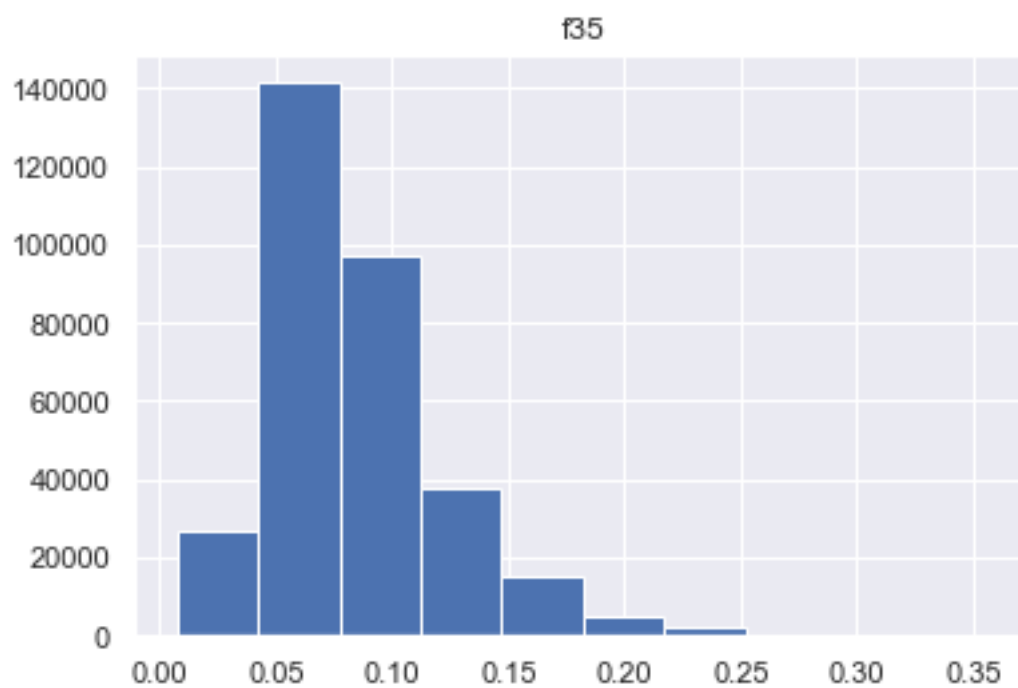
<Figure size 432x288 with 0 Axes>



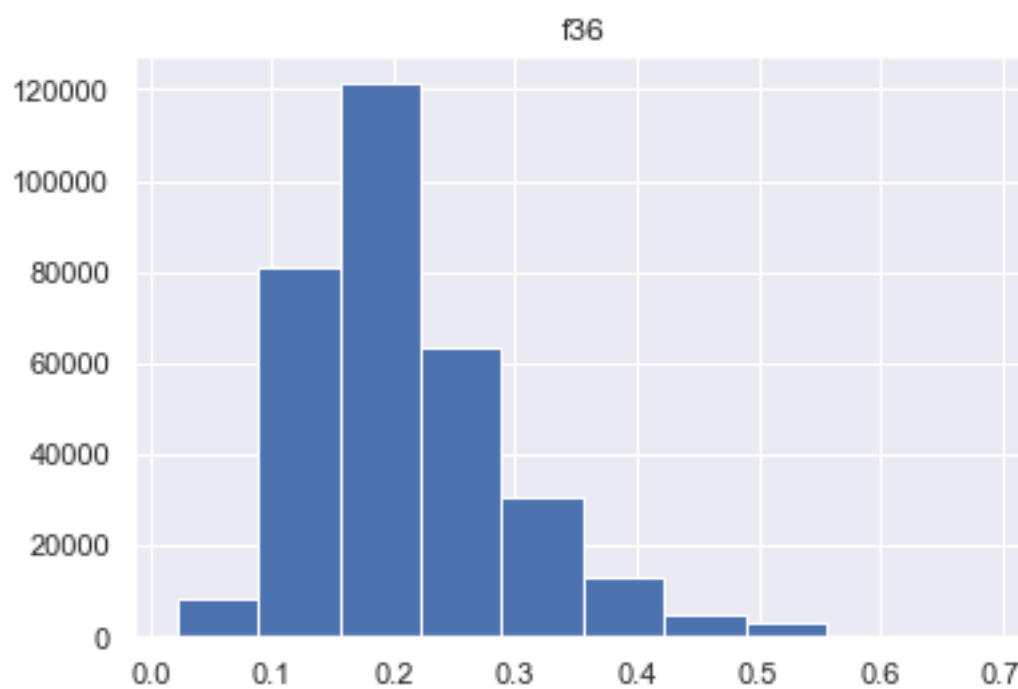
<Figure size 432x288 with 0 Axes>



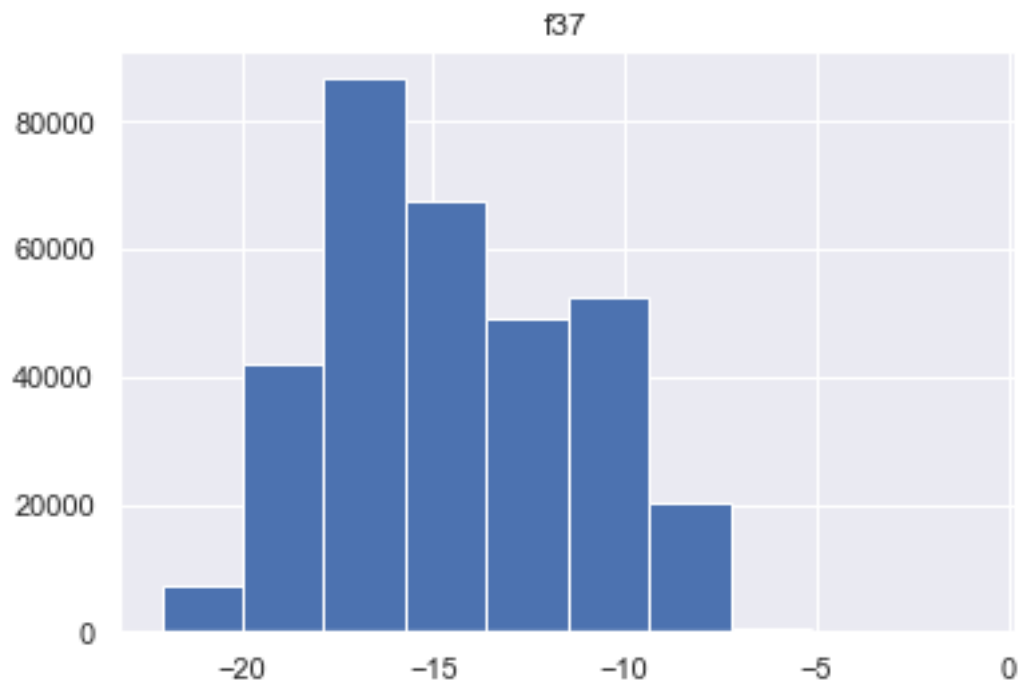
<Figure size 432x288 with 0 Axes>



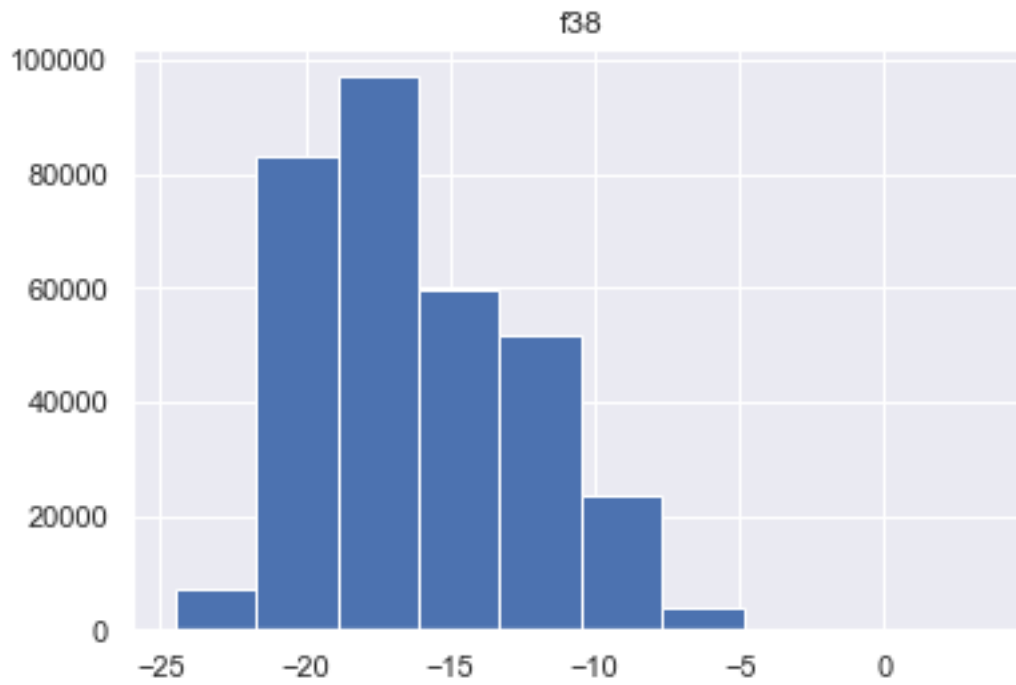
<Figure size 432x288 with 0 Axes>



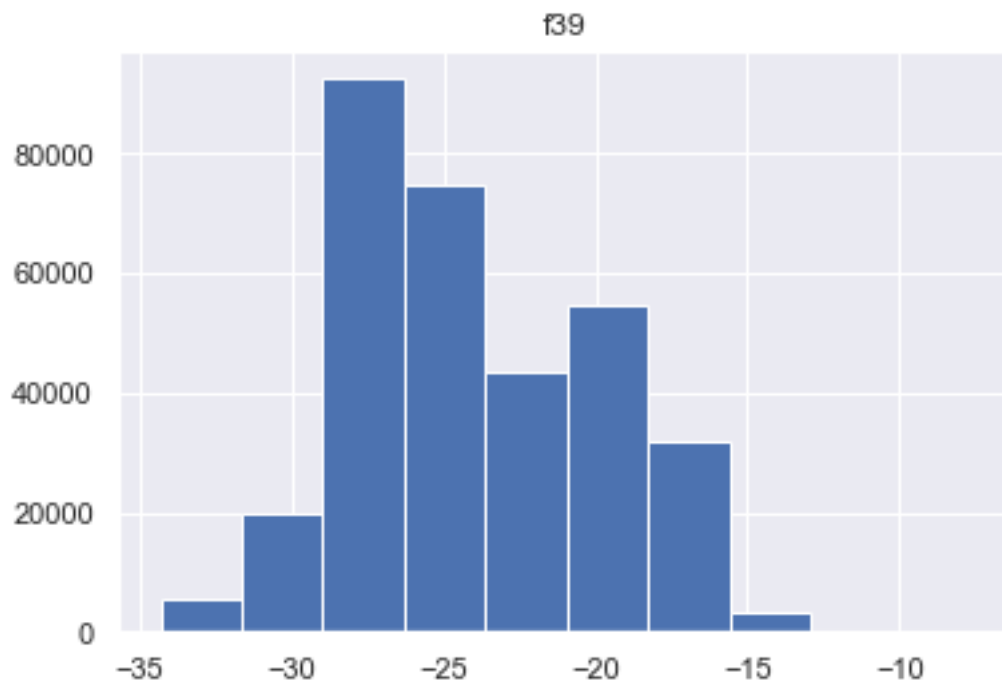
<Figure size 432x288 with 0 Axes>



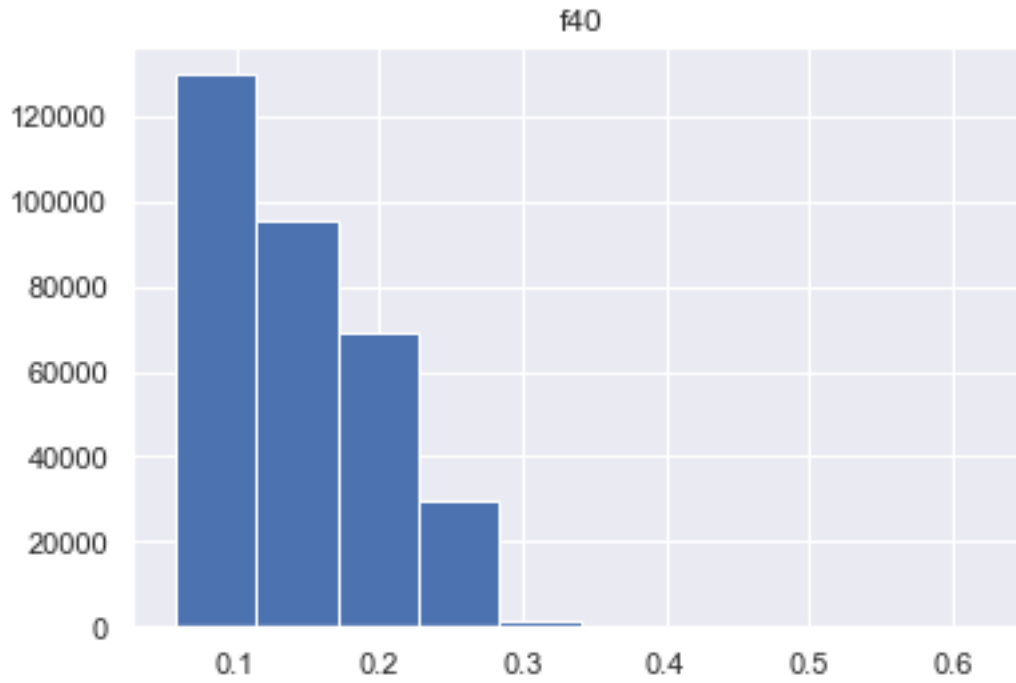
<Figure size 432x288 with 0 Axes>



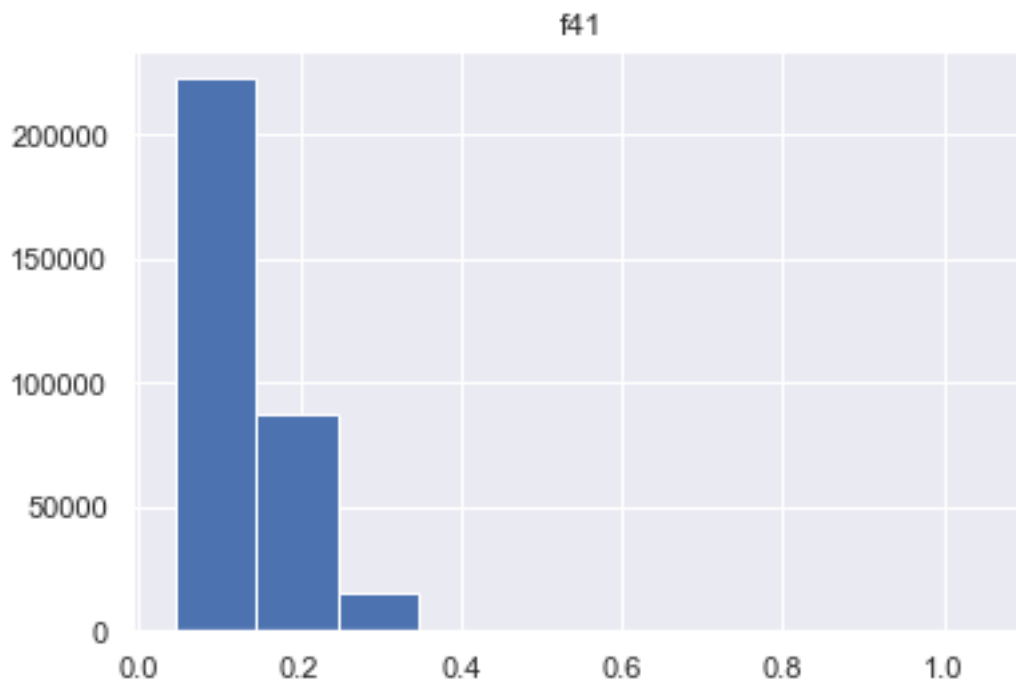
<Figure size 432x288 with 0 Axes>



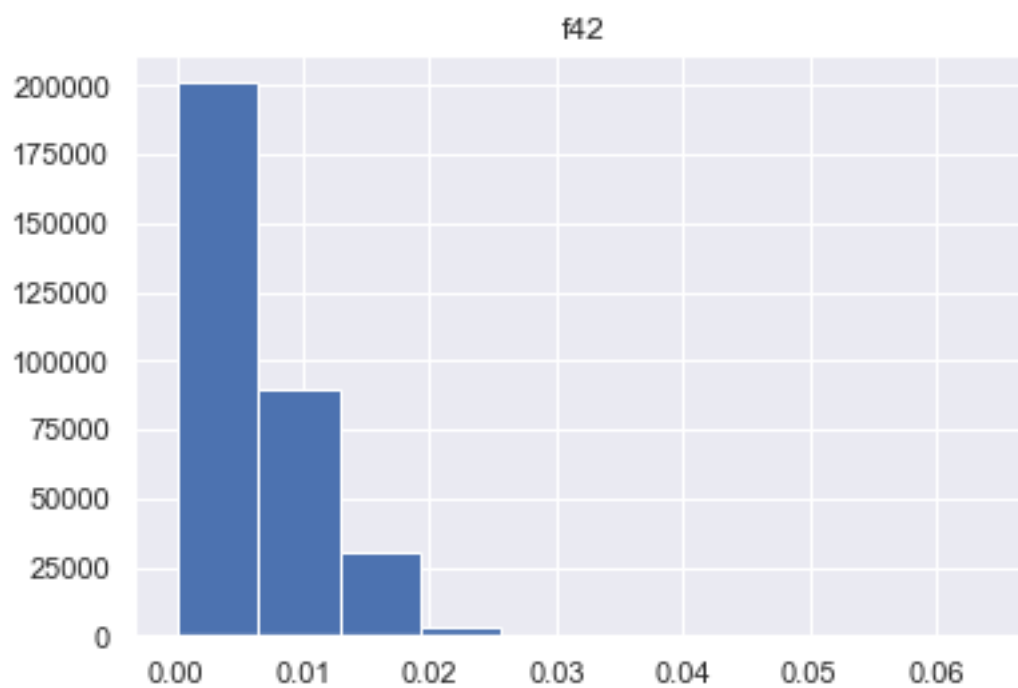
<Figure size 432x288 with 0 Axes>



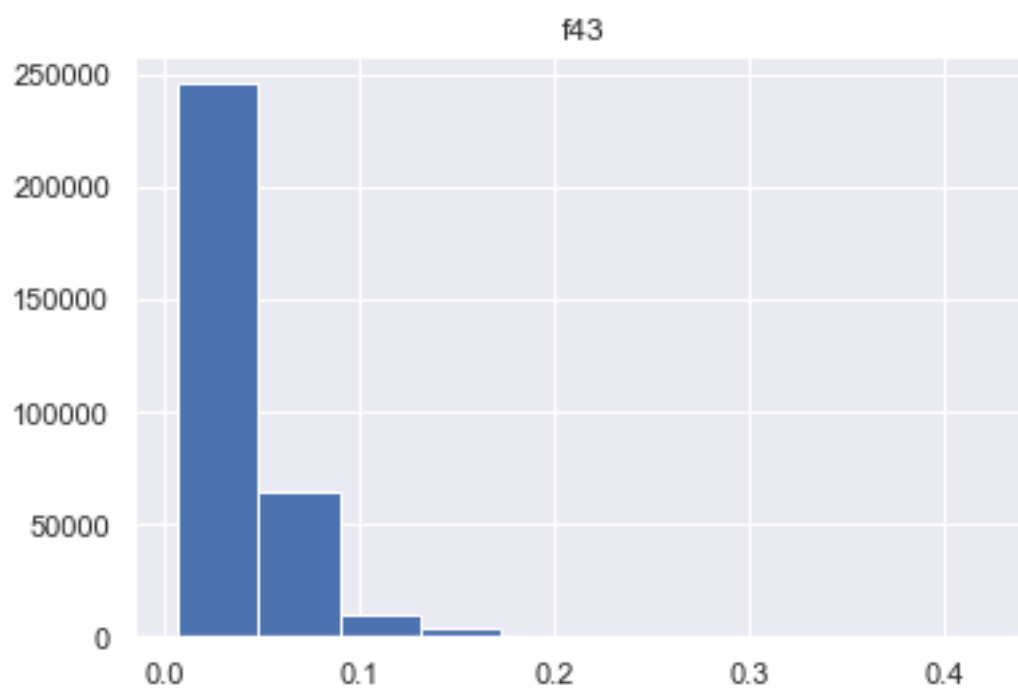
<Figure size 432x288 with 0 Axes>



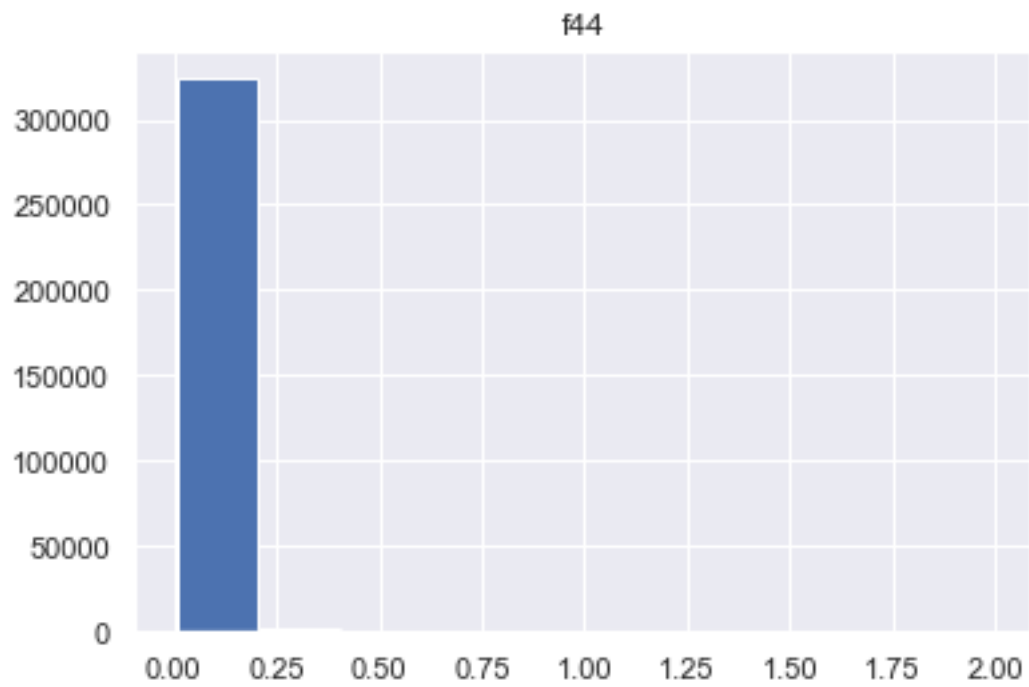
<Figure size 432x288 with 0 Axes>



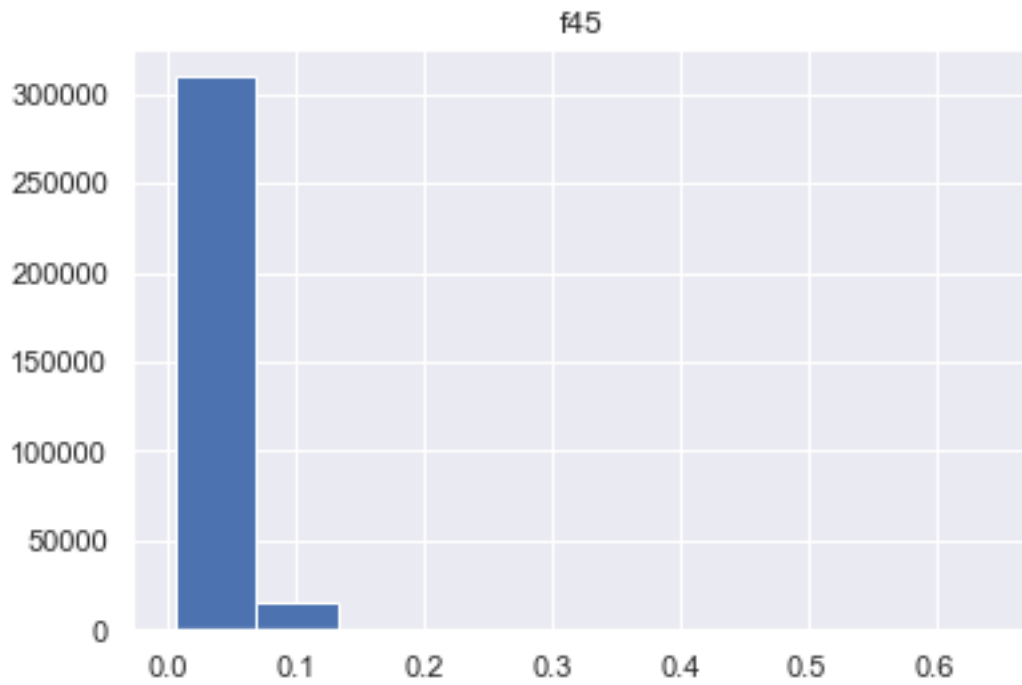
<Figure size 432x288 with 0 Axes>



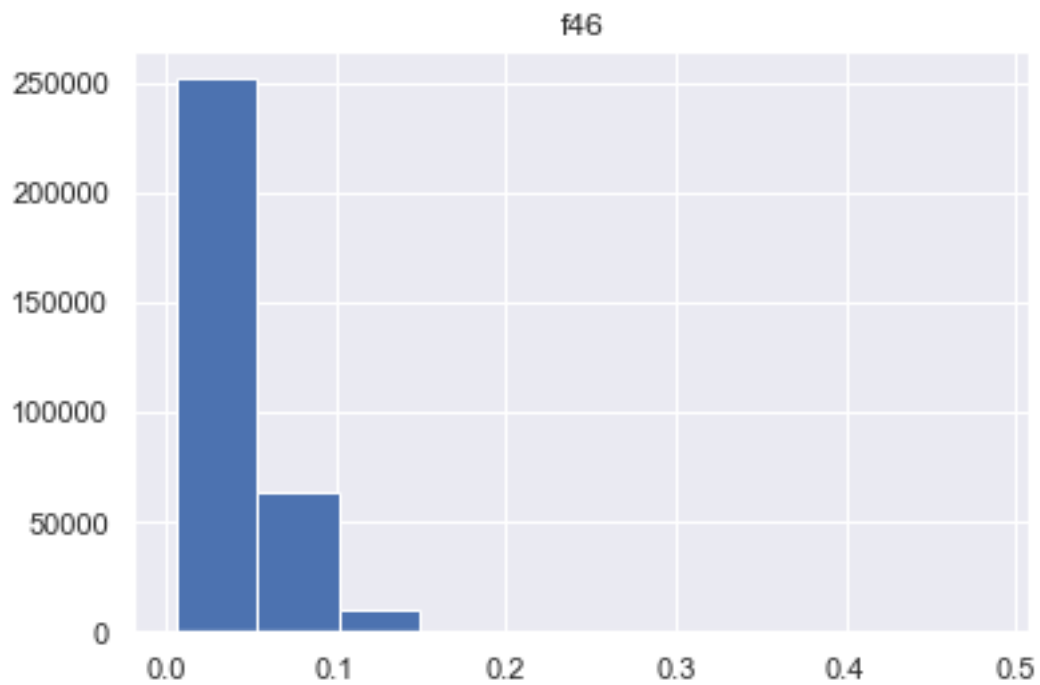
<Figure size 432x288 with 0 Axes>



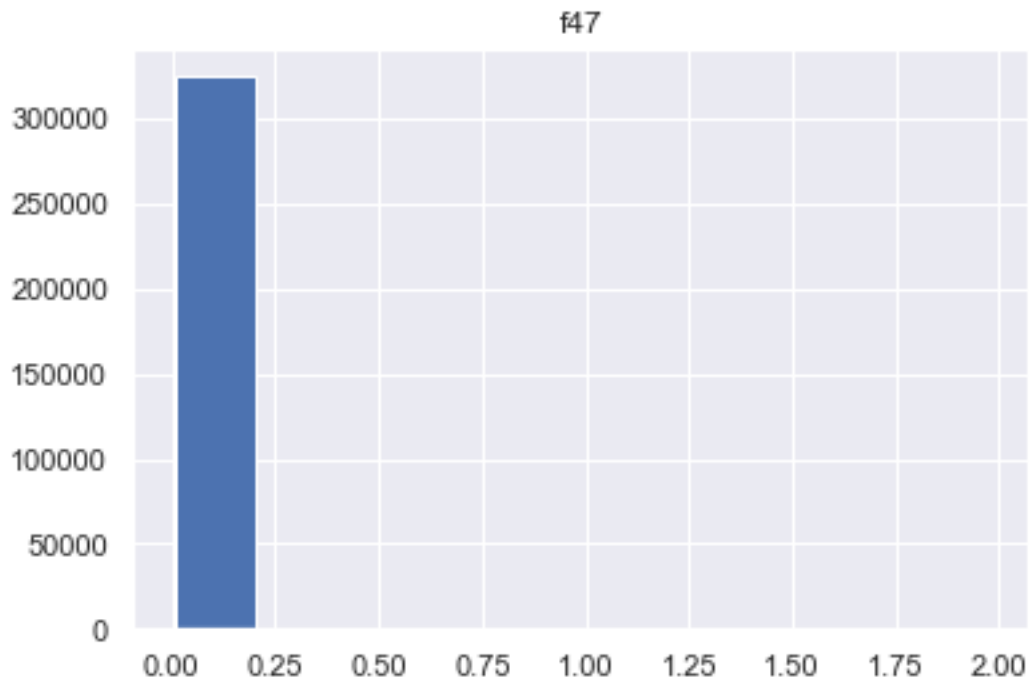
<Figure size 432x288 with 0 Axes>



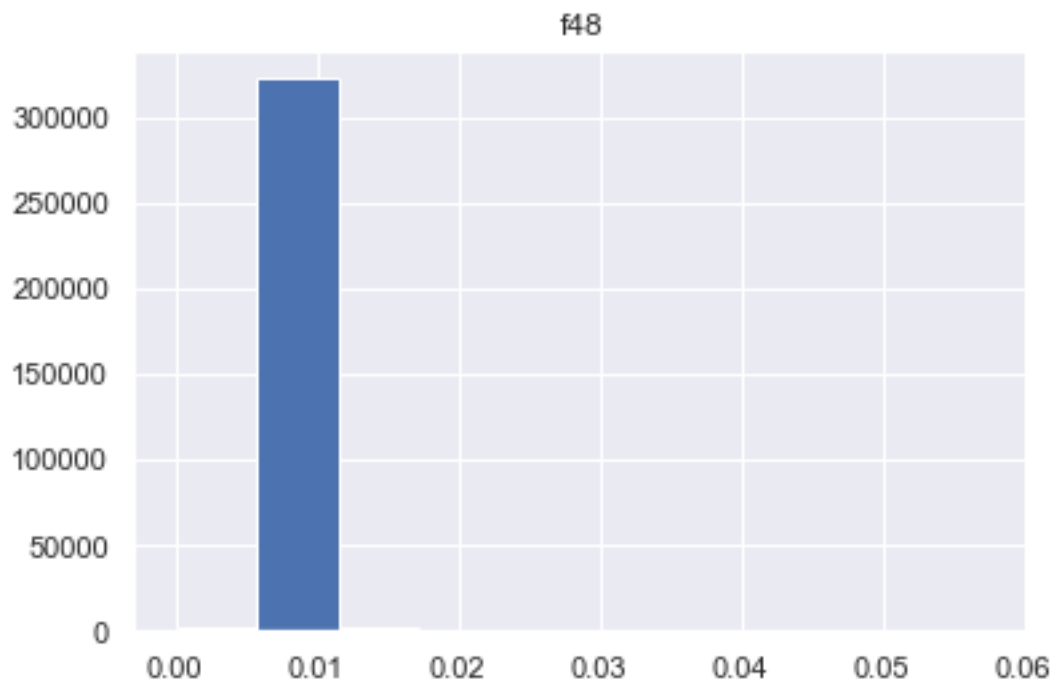
<Figure size 432x288 with 0 Axes>



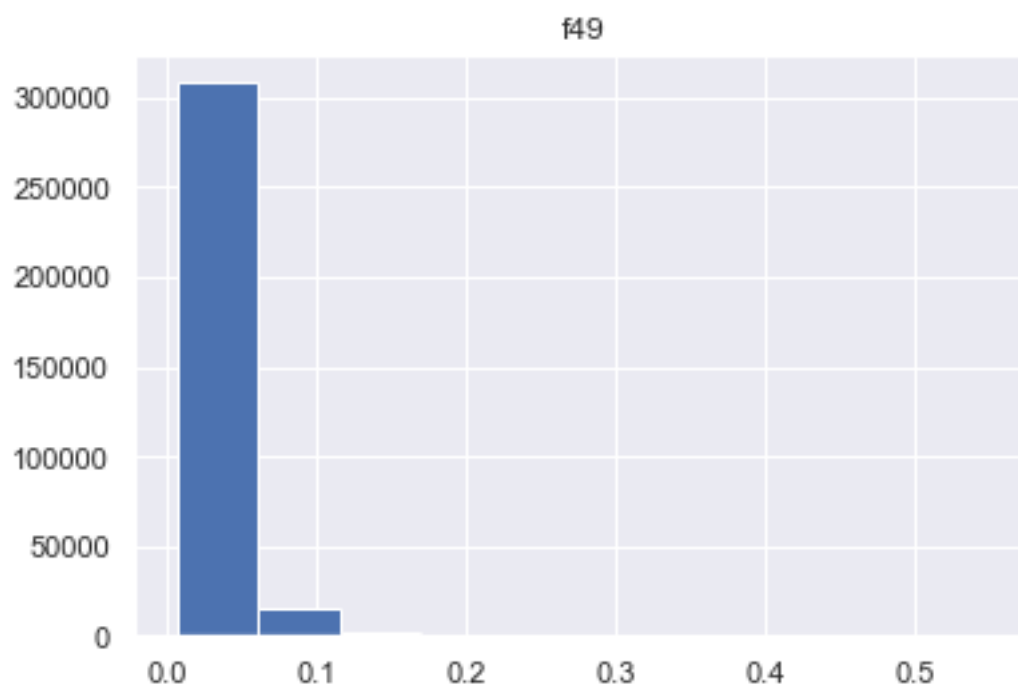
<Figure size 432x288 with 0 Axes>



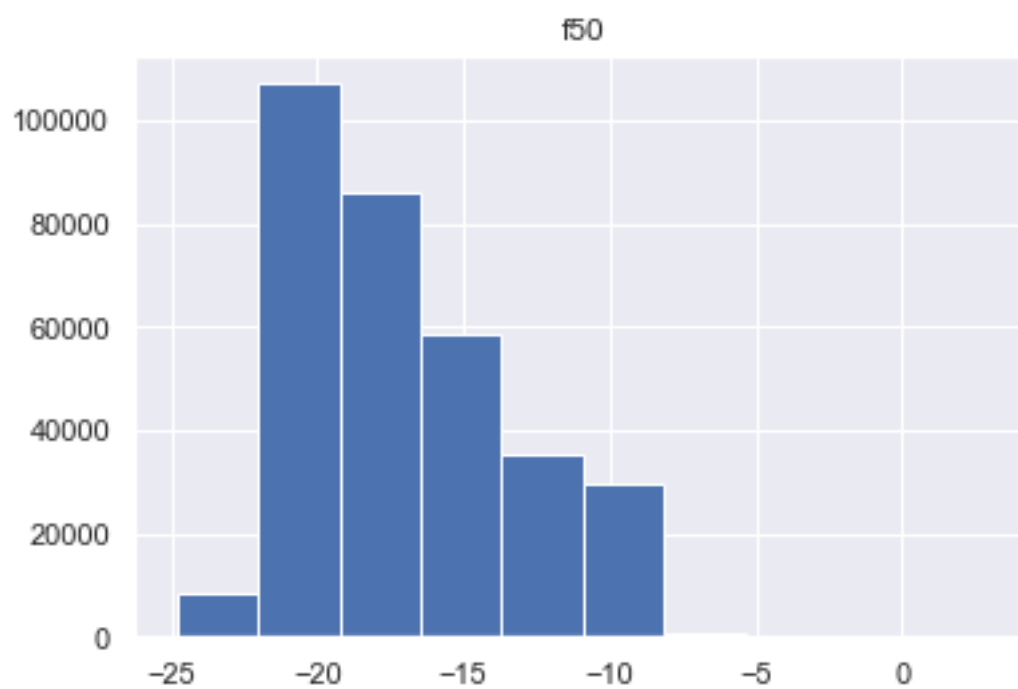
<Figure size 432x288 with 0 Axes>



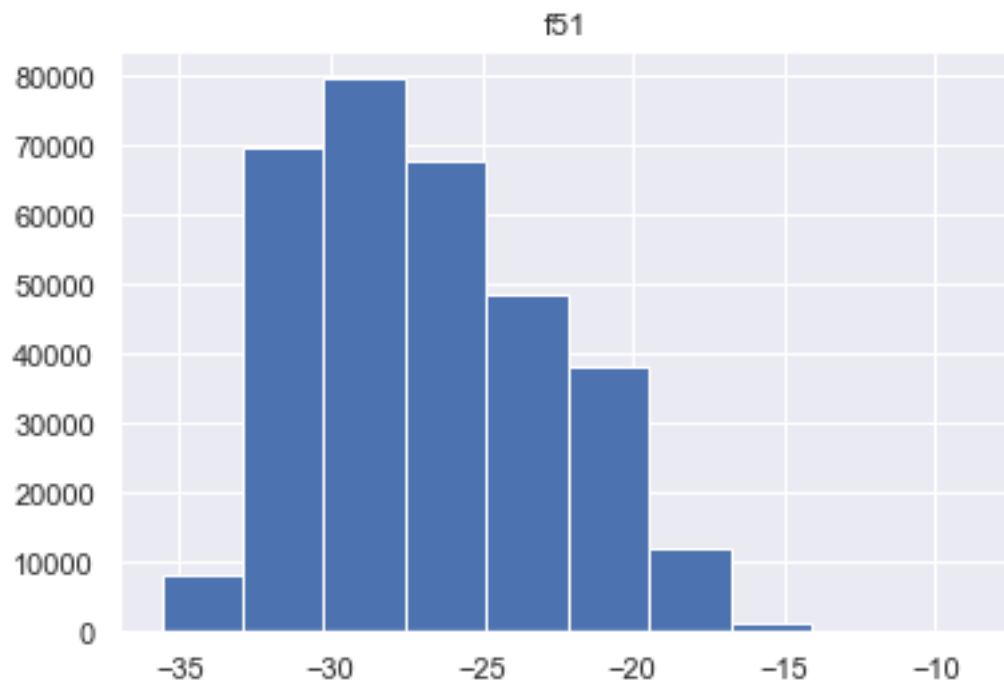
<Figure size 432x288 with 0 Axes>



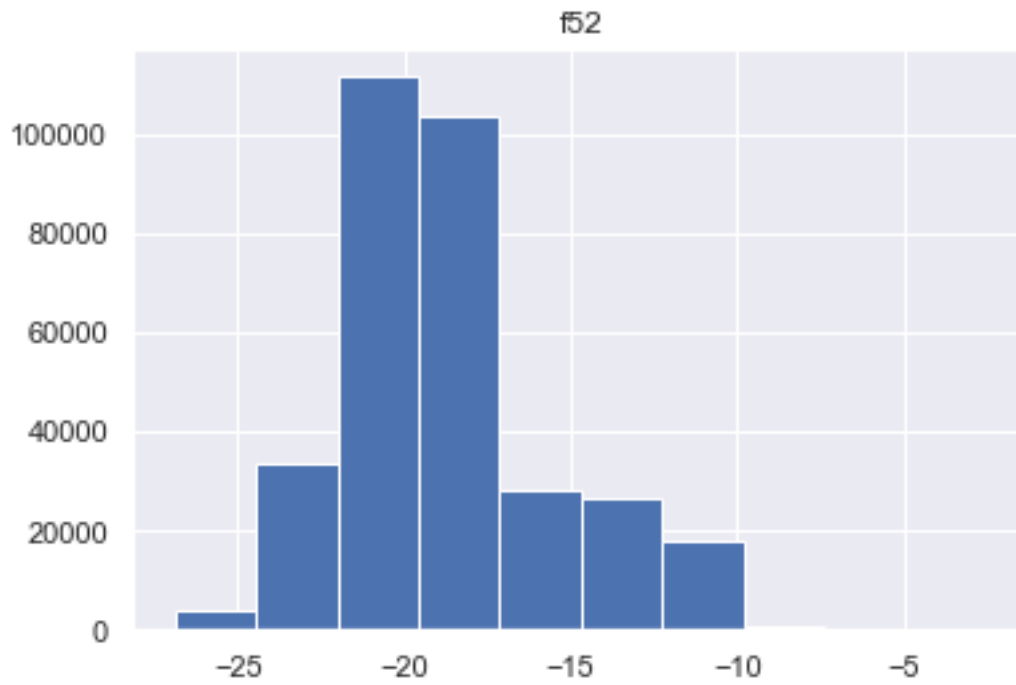
<Figure size 432x288 with 0 Axes>



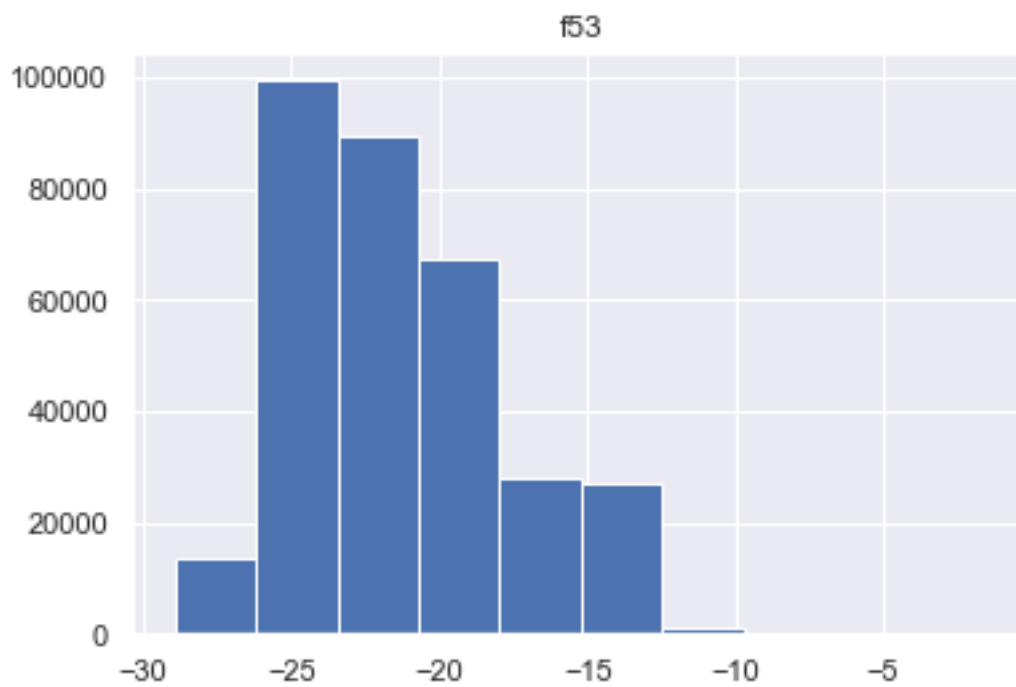
<Figure size 432x288 with 0 Axes>



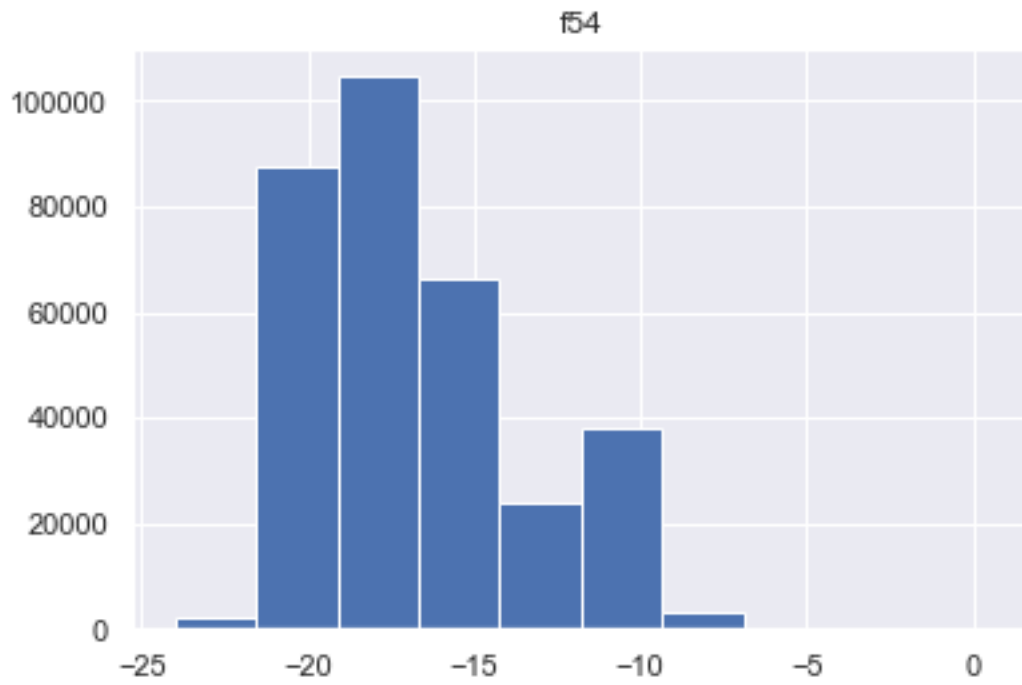
<Figure size 432x288 with 0 Axes>



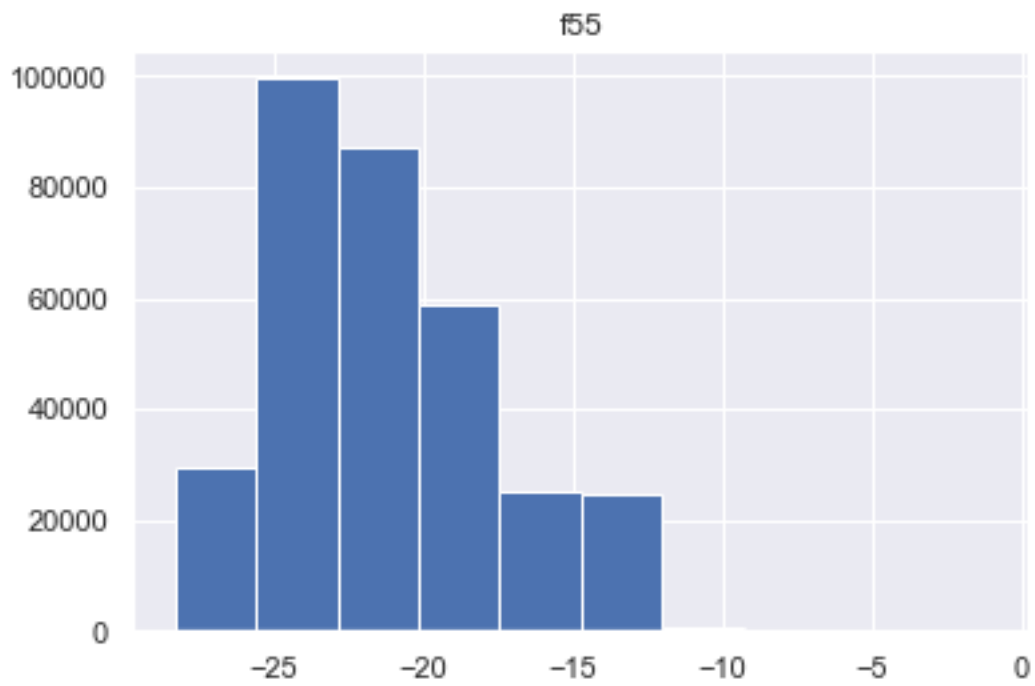
<Figure size 432x288 with 0 Axes>



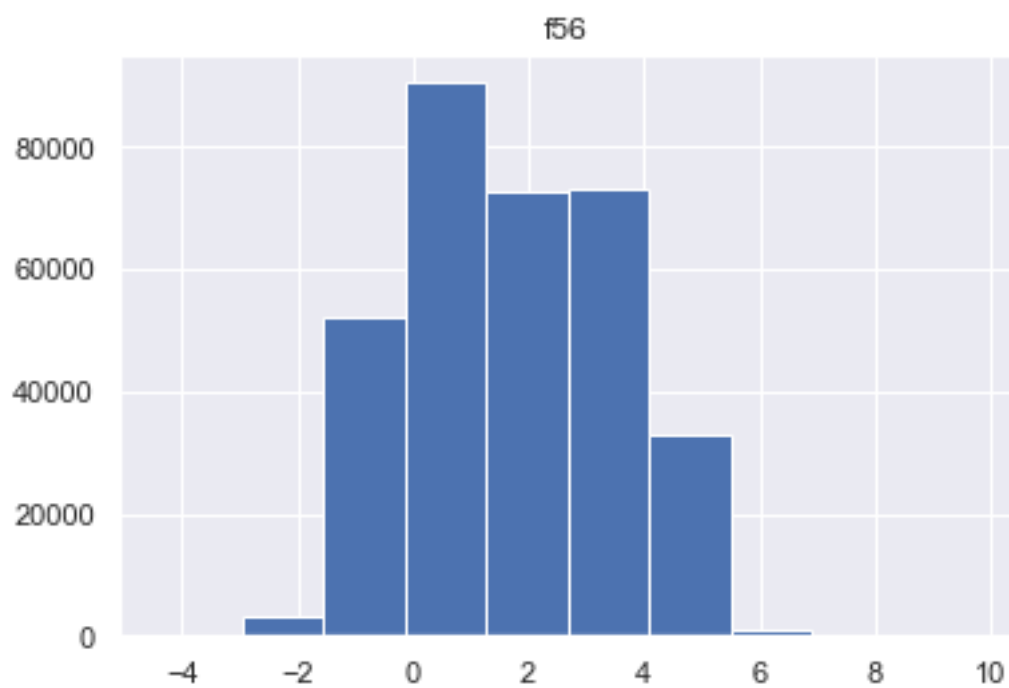
<Figure size 432x288 with 0 Axes>



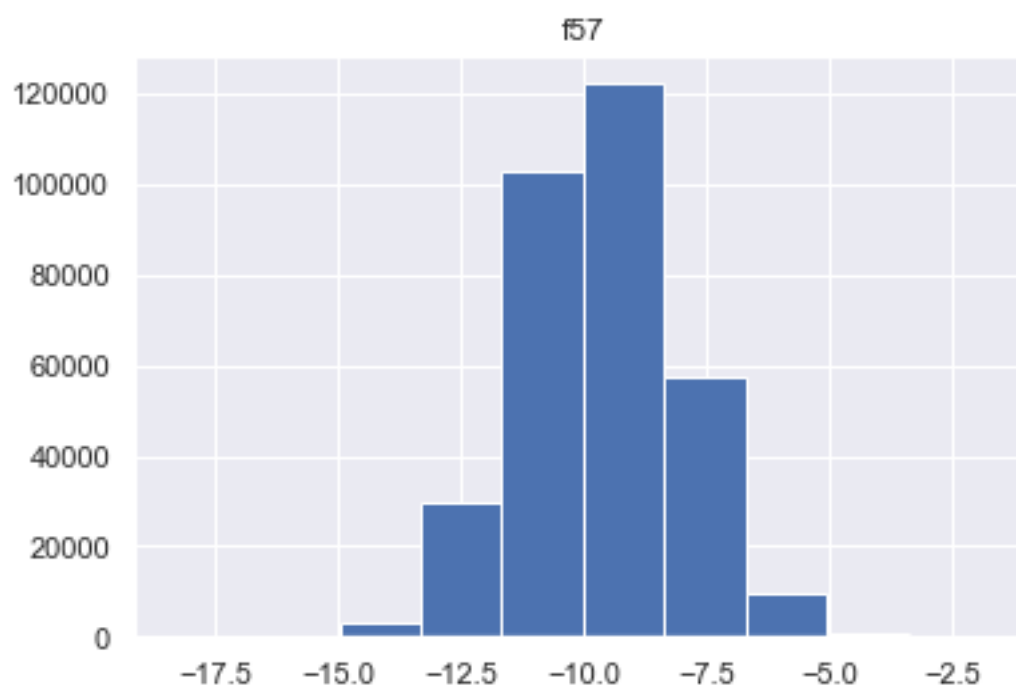
<Figure size 432x288 with 0 Axes>



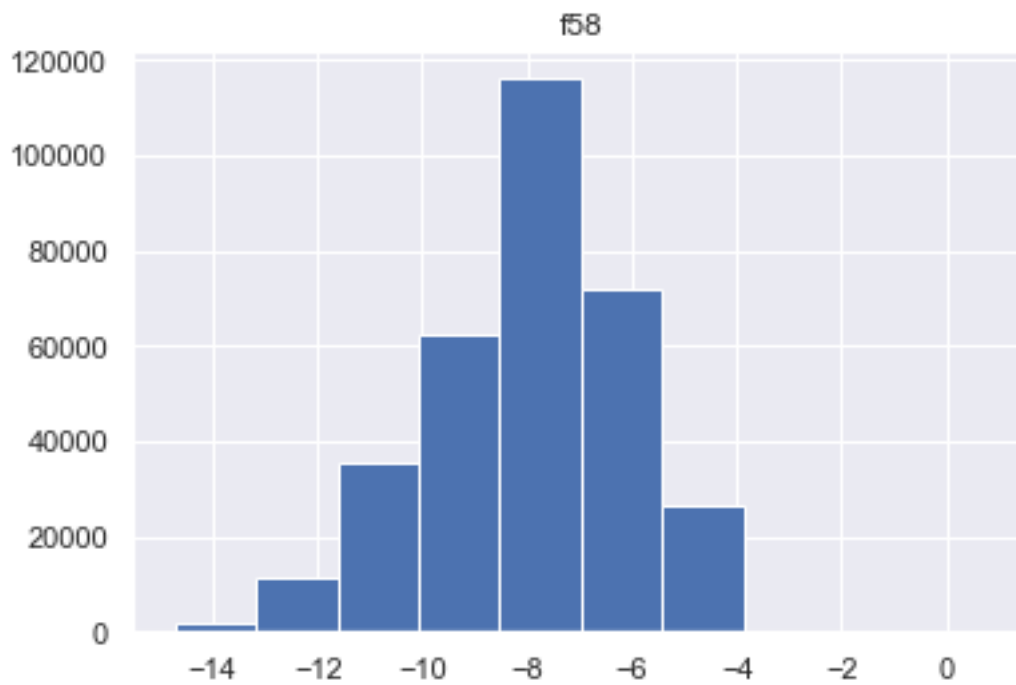
<Figure size 432x288 with 0 Axes>



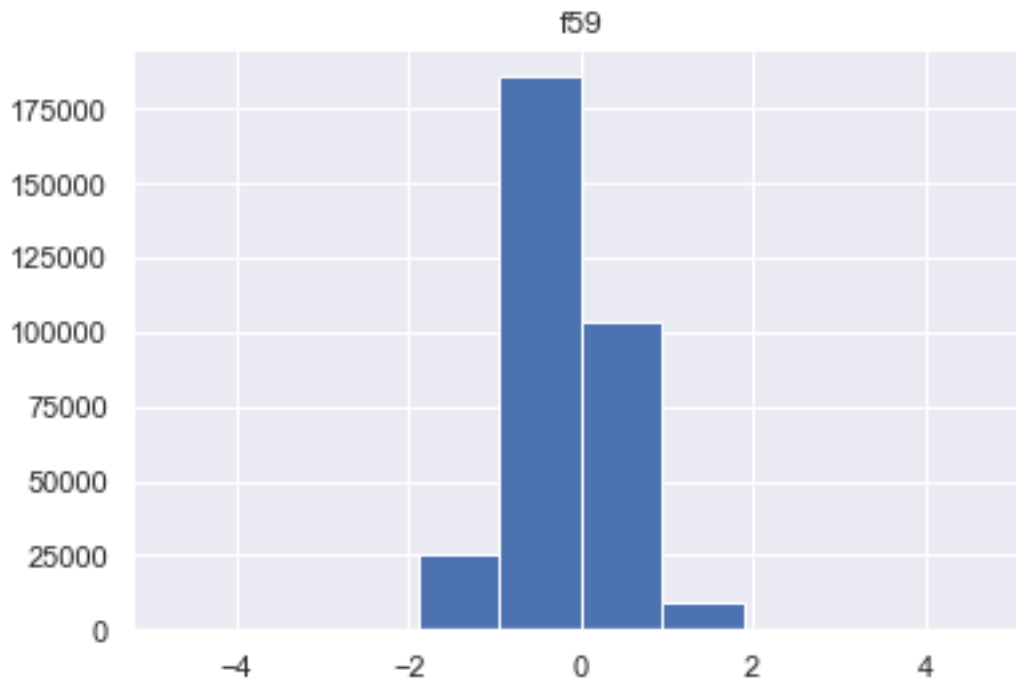
<Figure size 432x288 with 0 Axes>



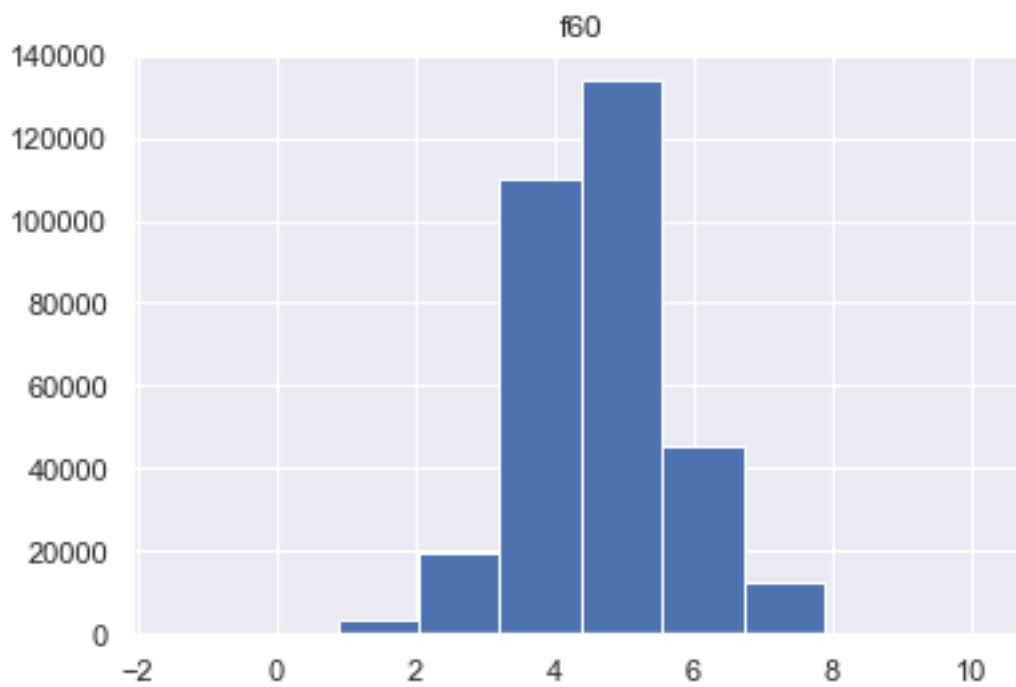
<Figure size 432x288 with 0 Axes>



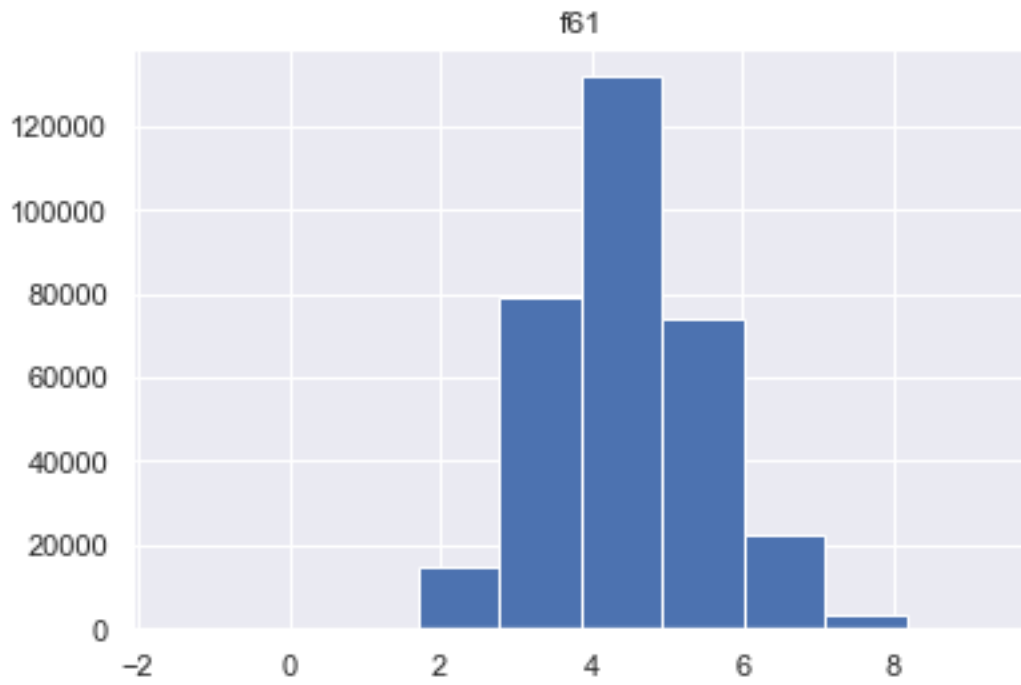
<Figure size 432x288 with 0 Axes>



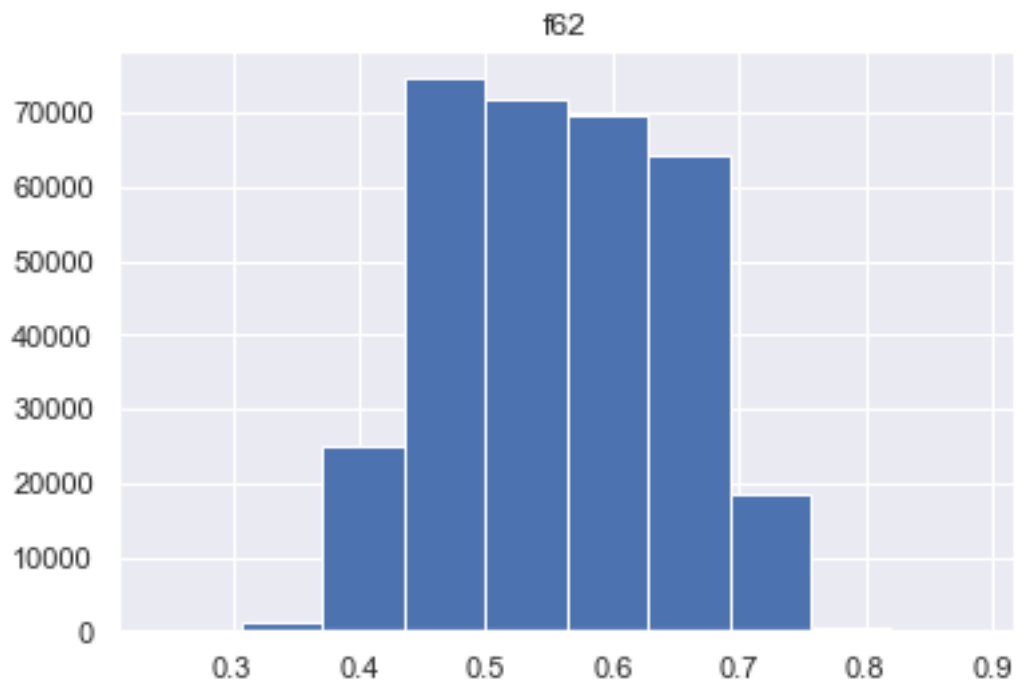
<Figure size 432x288 with 0 Axes>



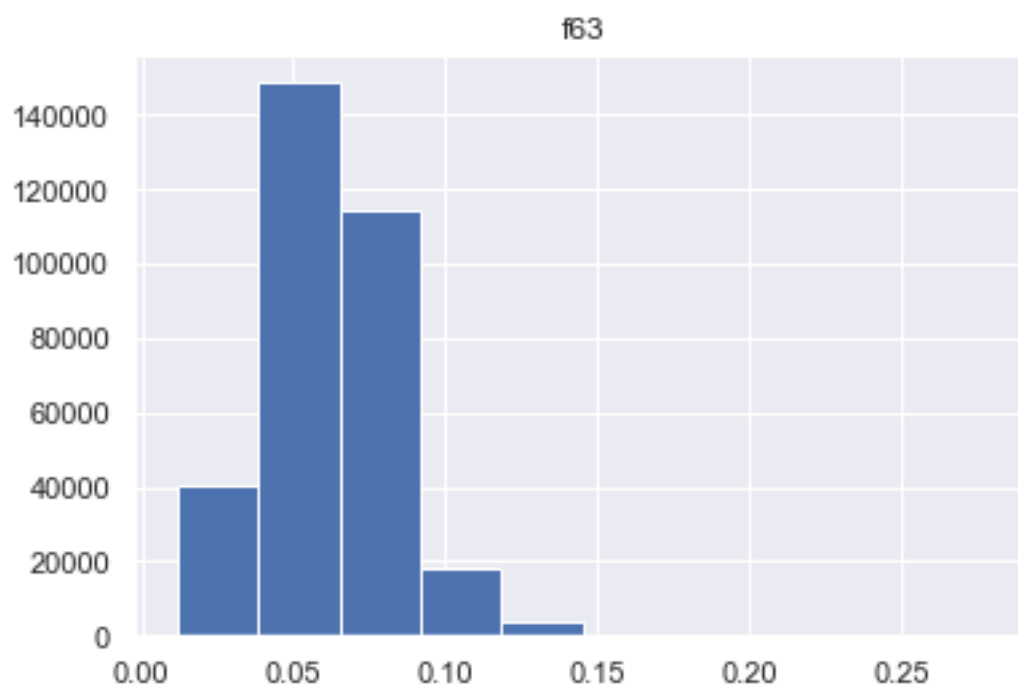
<Figure size 432x288 with 0 Axes>



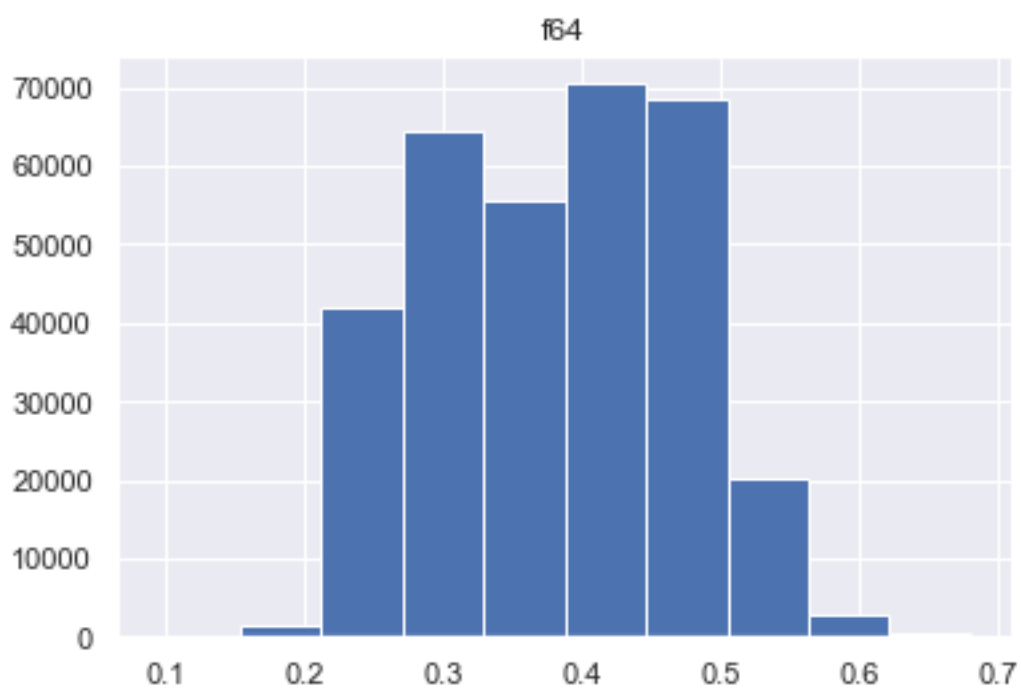
<Figure size 432x288 with 0 Axes>



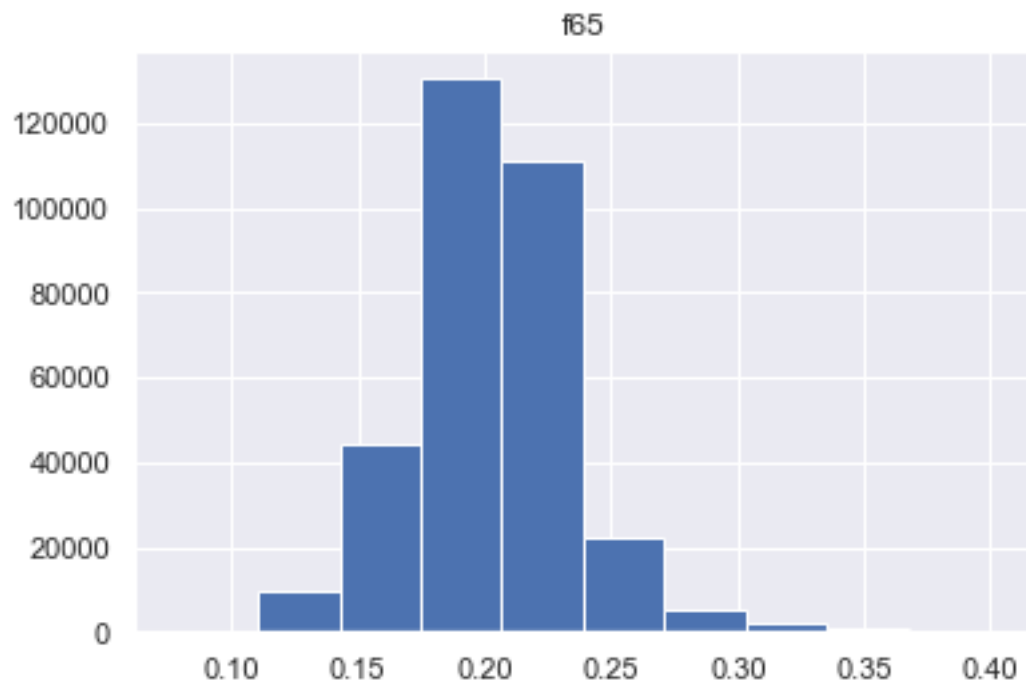
<Figure size 432x288 with 0 Axes>



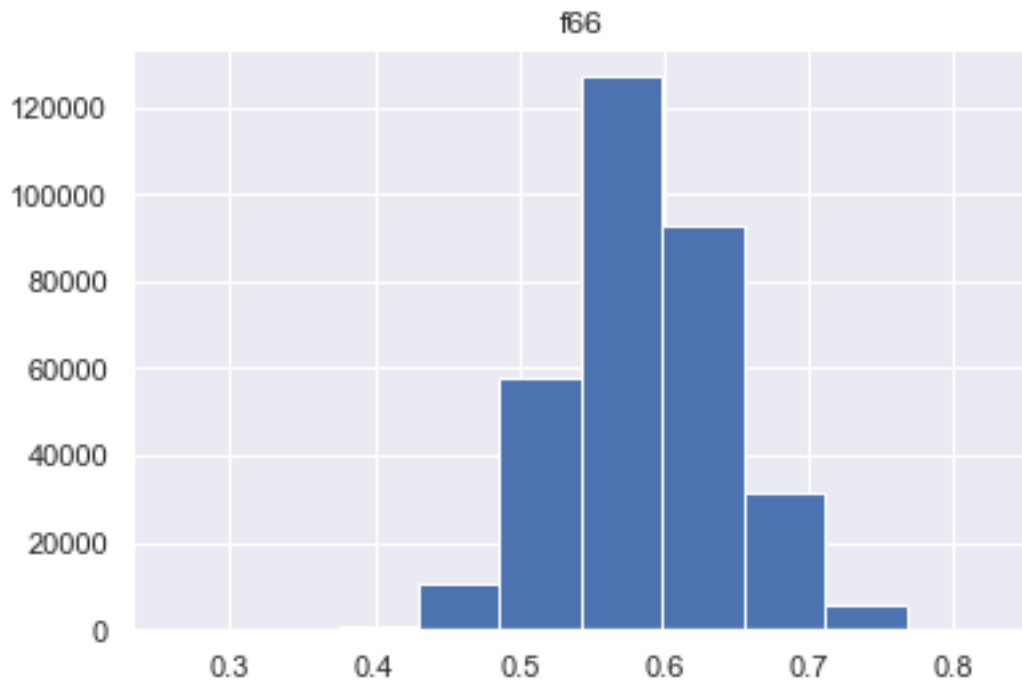
<Figure size 432x288 with 0 Axes>



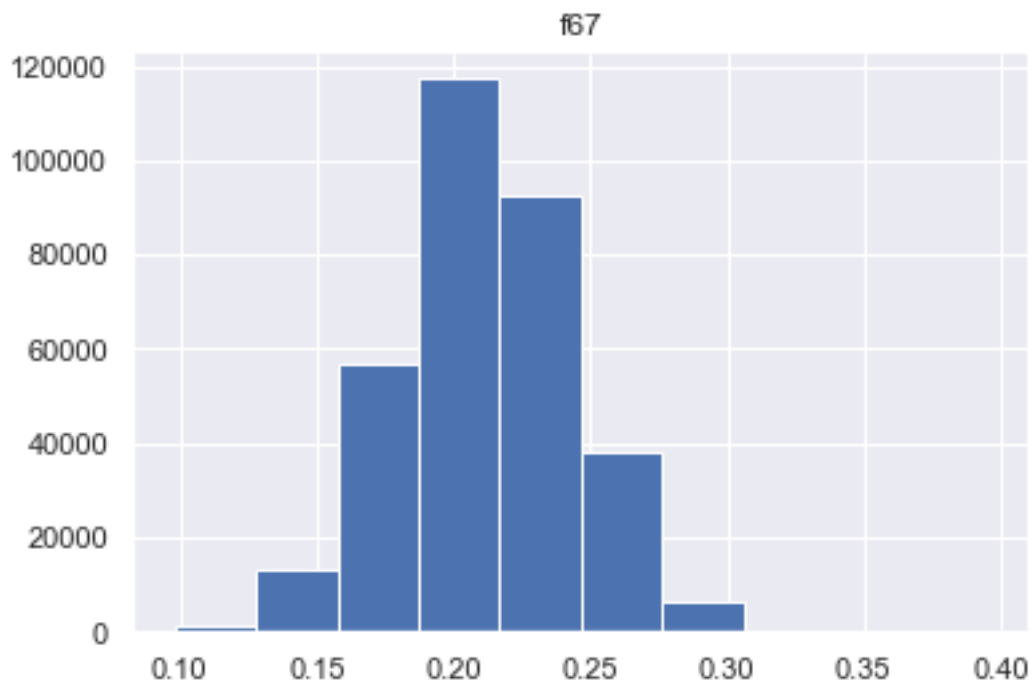
<Figure size 432x288 with 0 Axes>



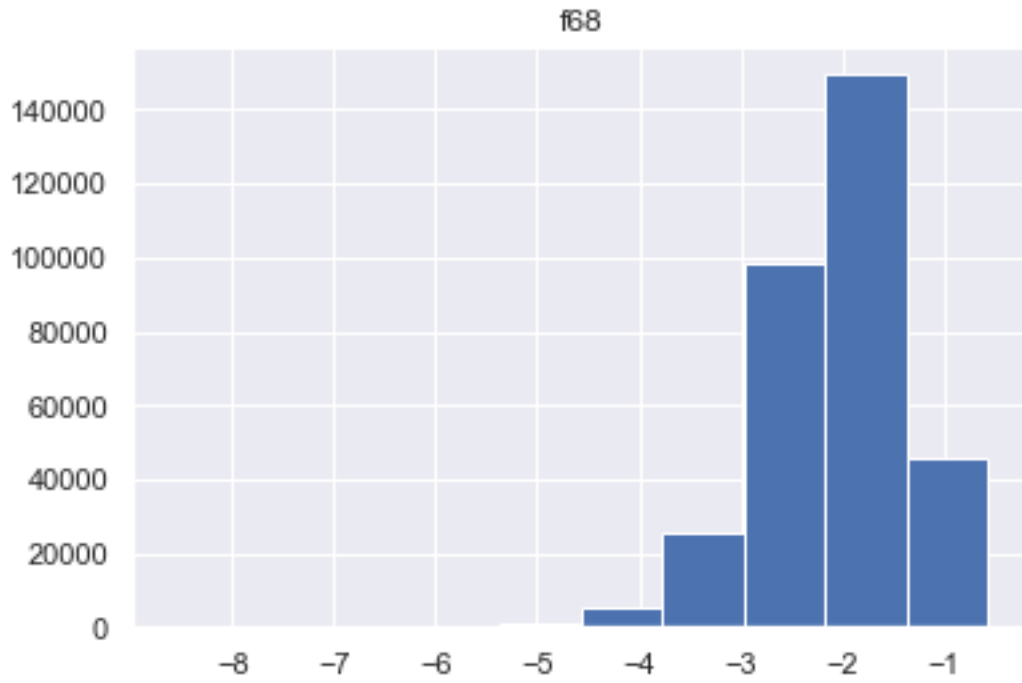
<Figure size 432x288 with 0 Axes>



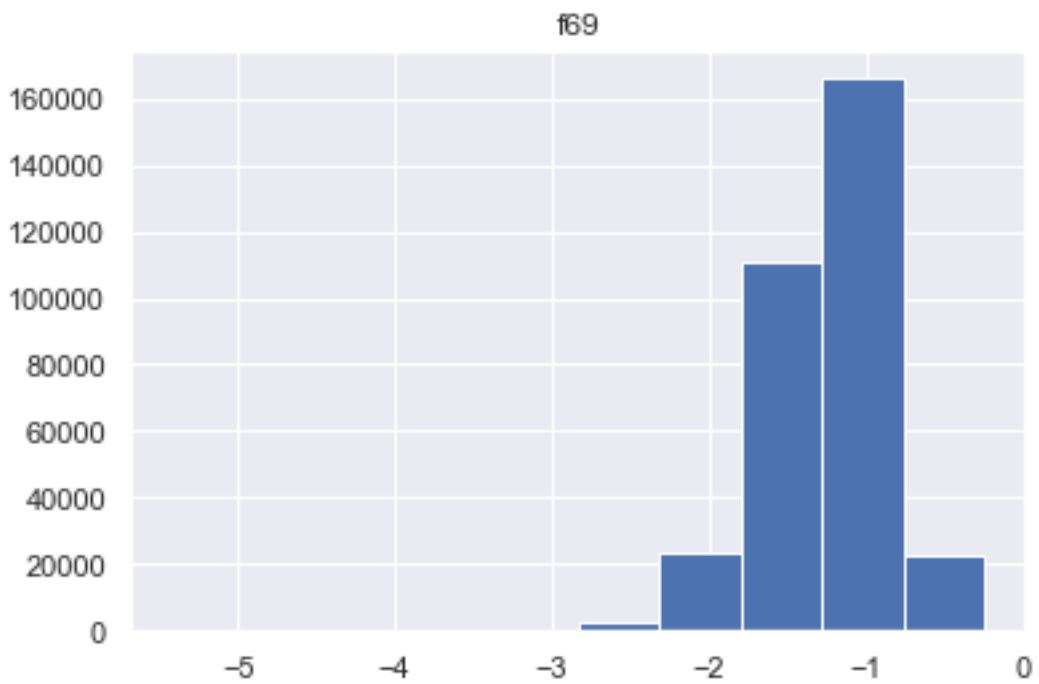
<Figure size 432x288 with 0 Axes>



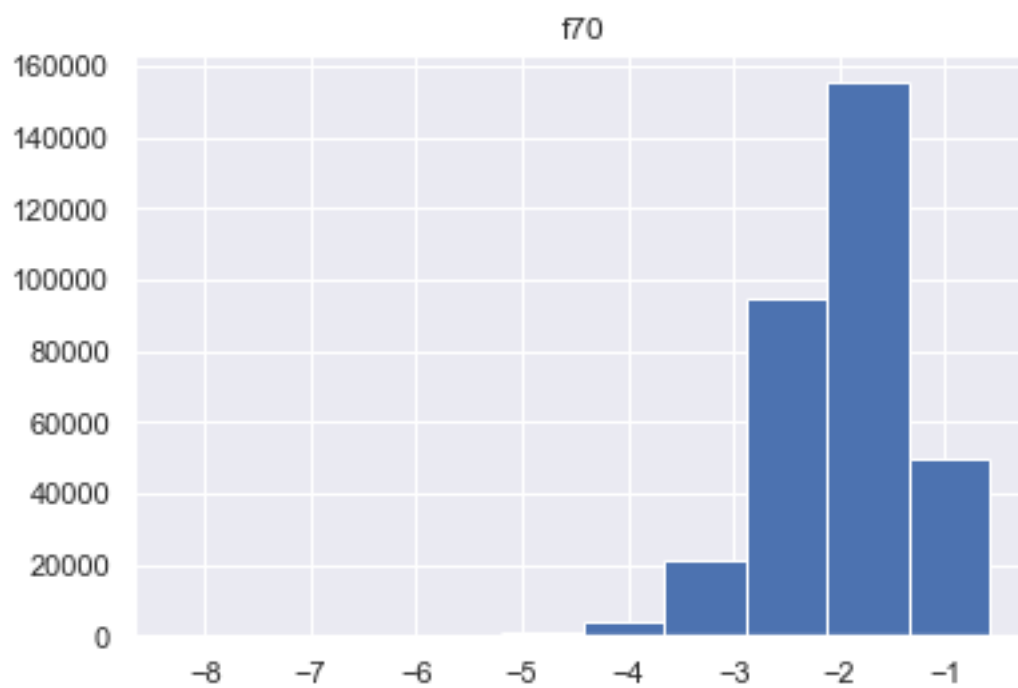
<Figure size 432x288 with 0 Axes>



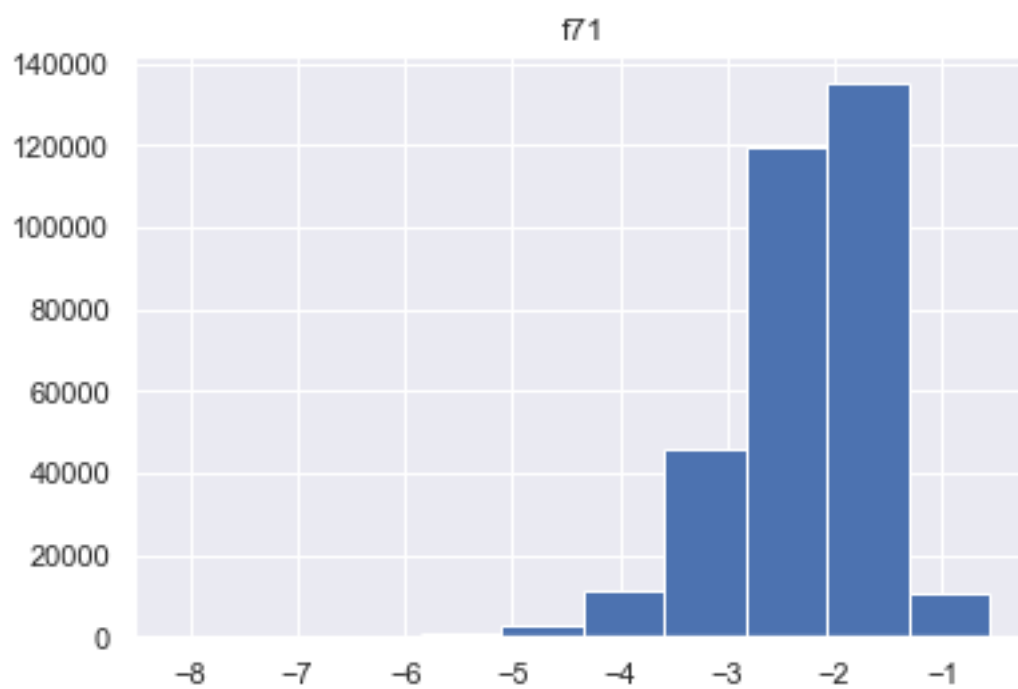
<Figure size 432x288 with 0 Axes>



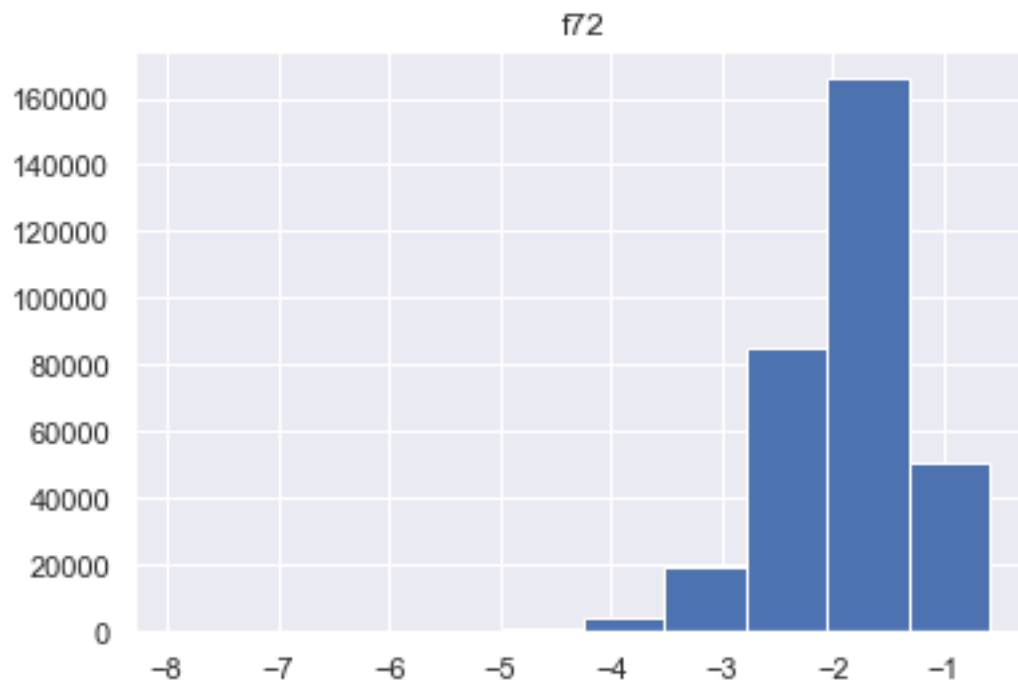
<Figure size 432x288 with 0 Axes>



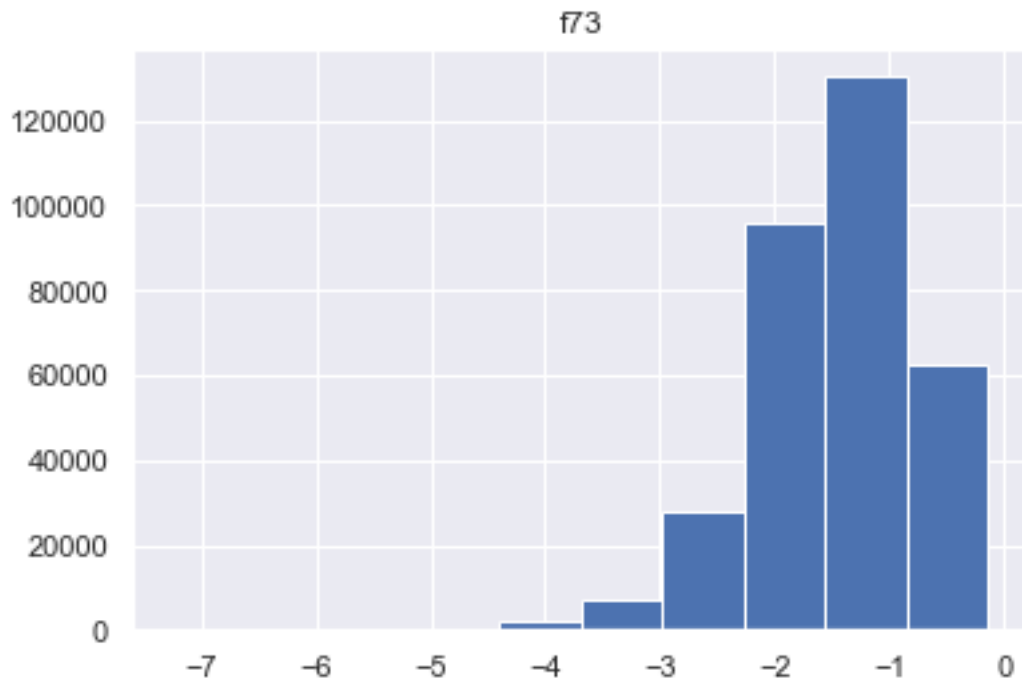
<Figure size 432x288 with 0 Axes>



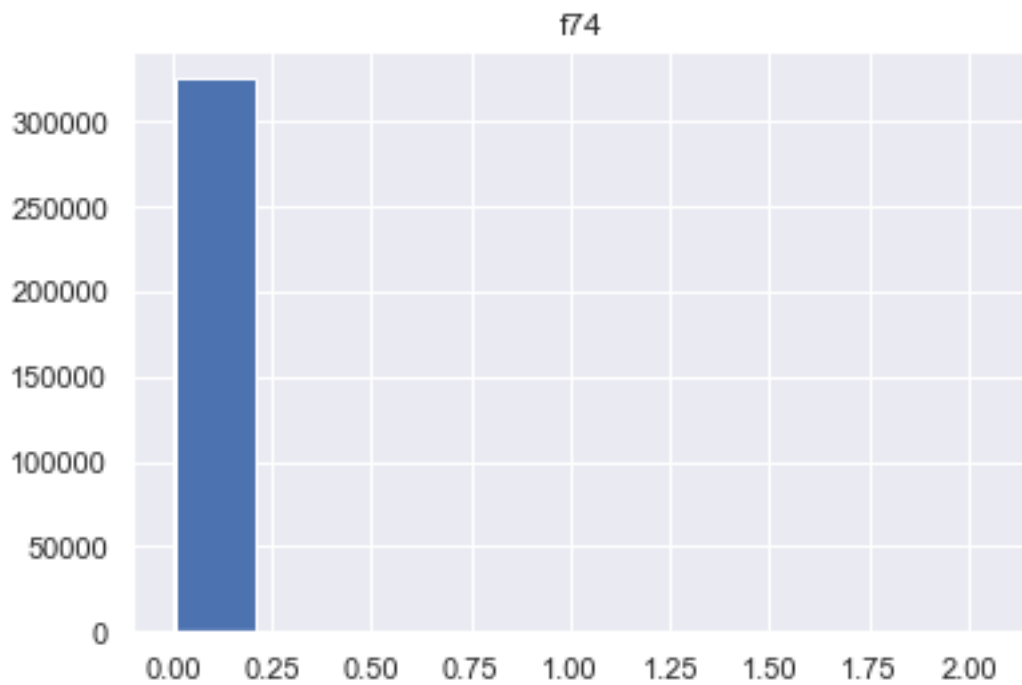
<Figure size 432x288 with 0 Axes>



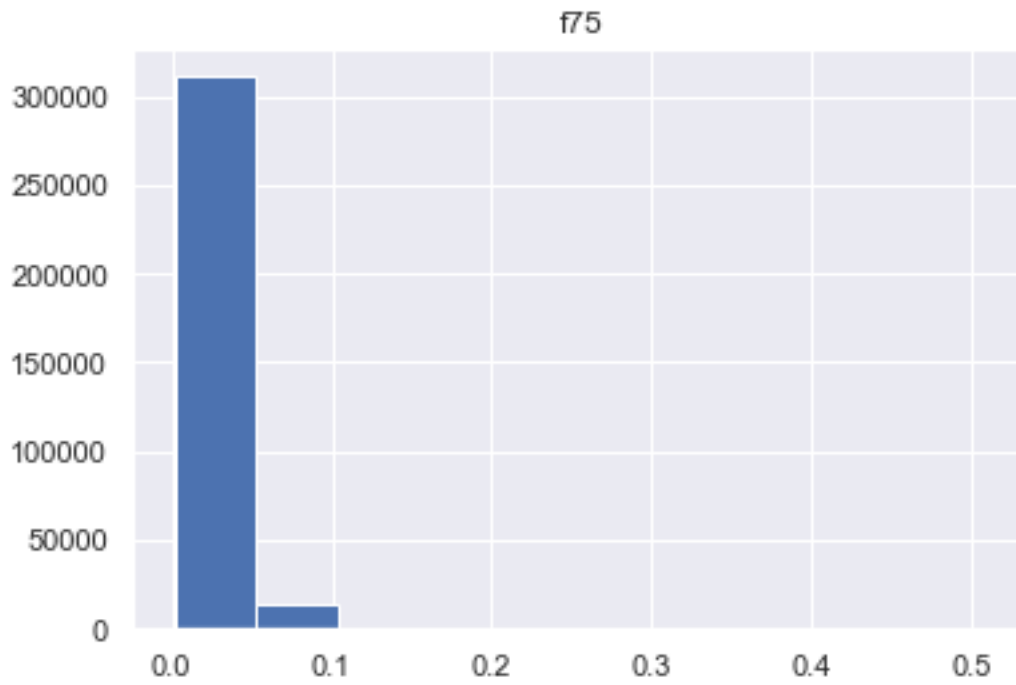
<Figure size 432x288 with 0 Axes>



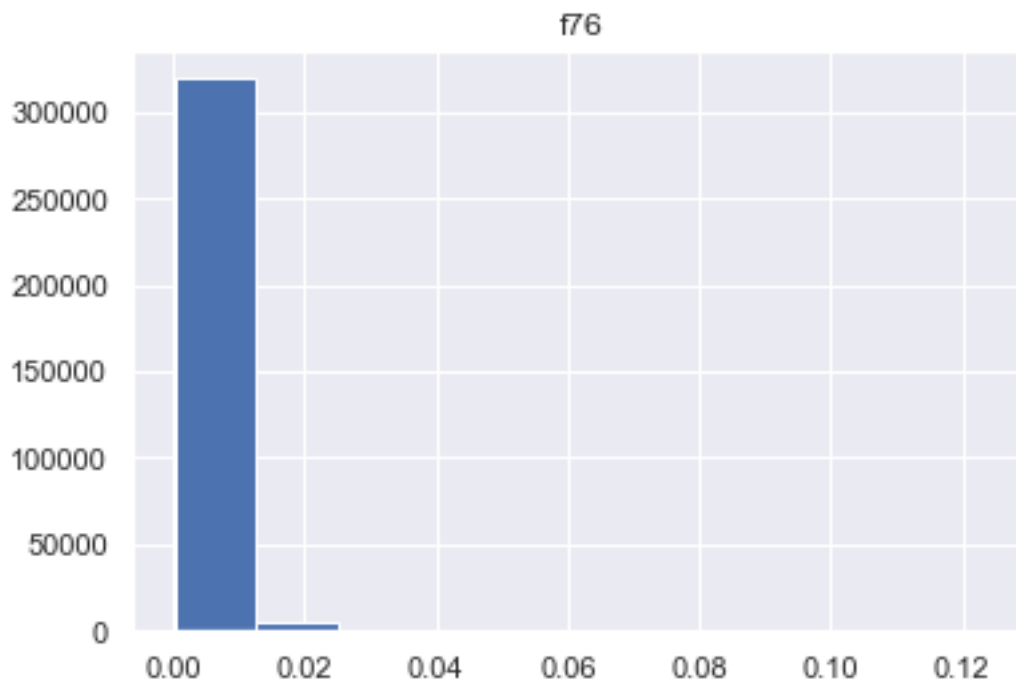
<Figure size 432x288 with 0 Axes>



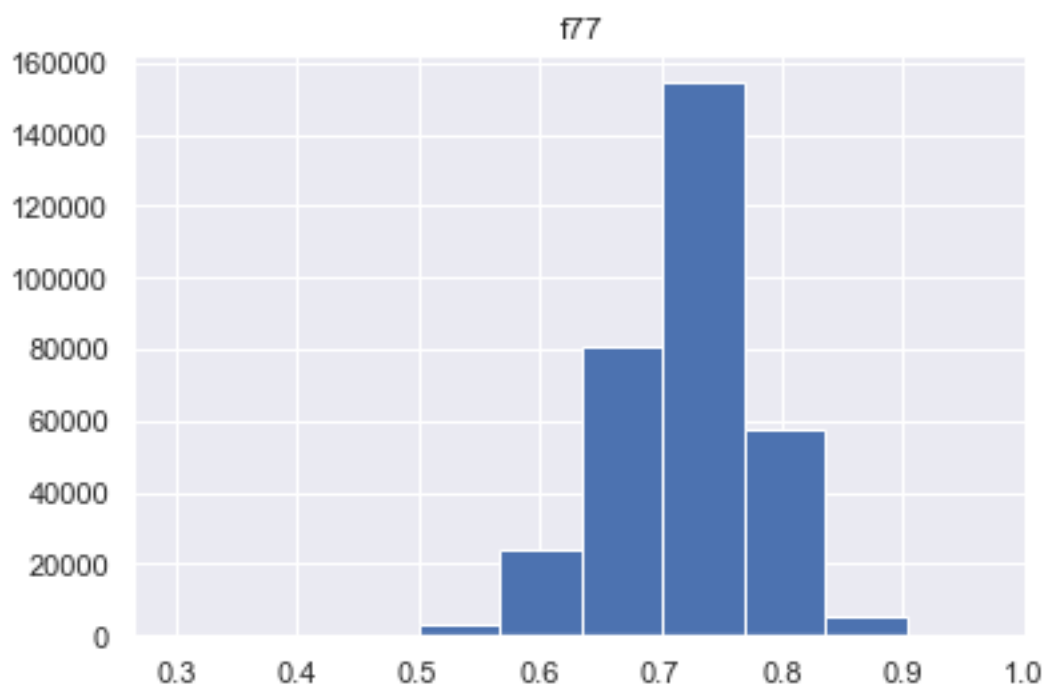
<Figure size 432x288 with 0 Axes>



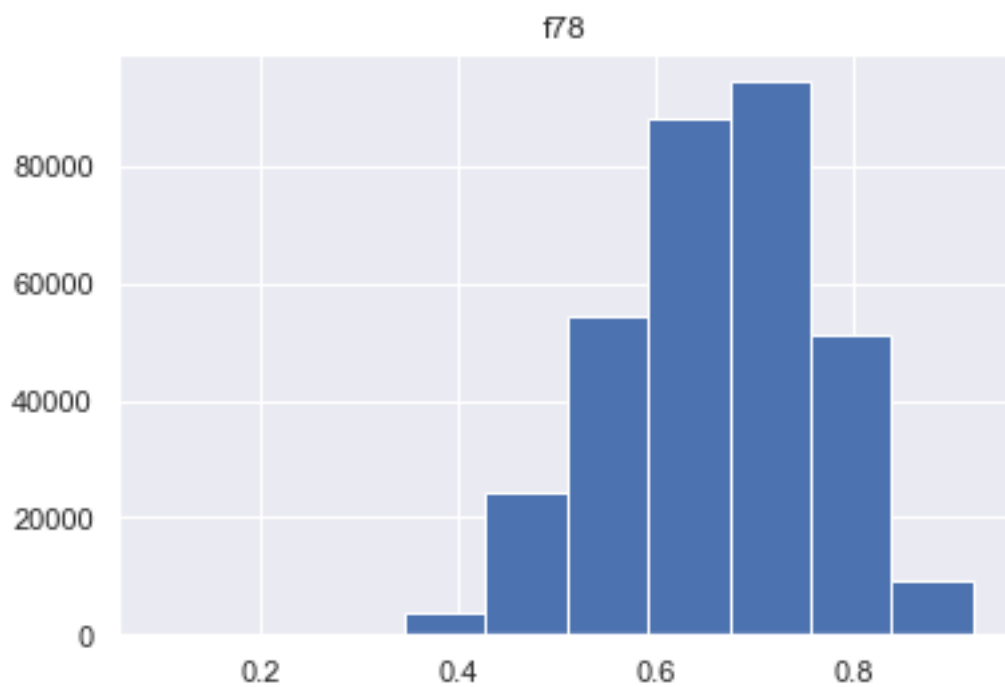
<Figure size 432x288 with 0 Axes>



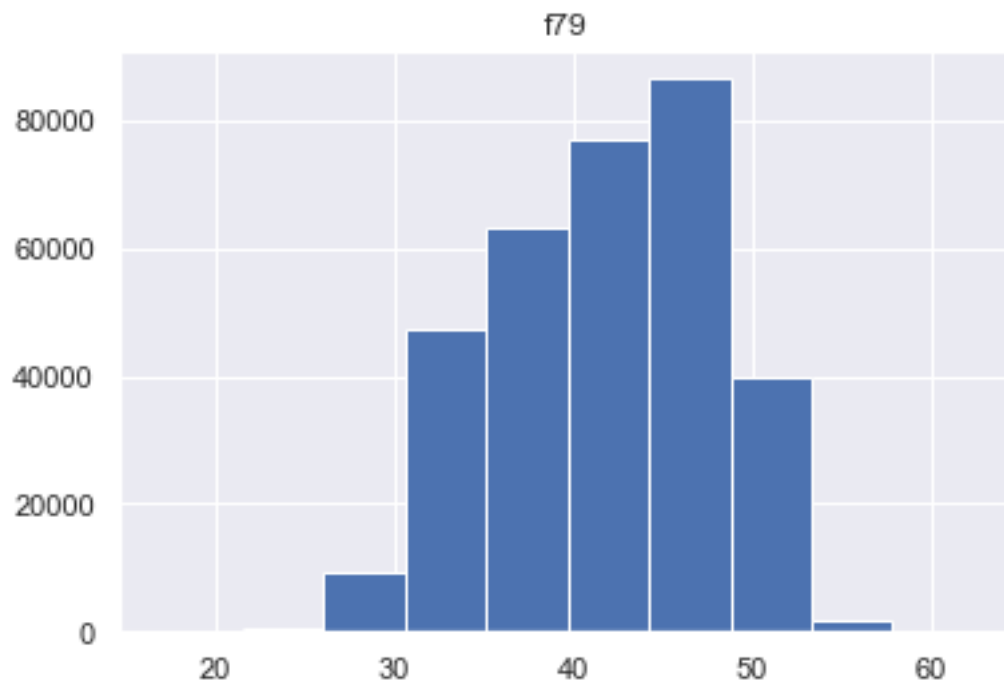
<Figure size 432x288 with 0 Axes>



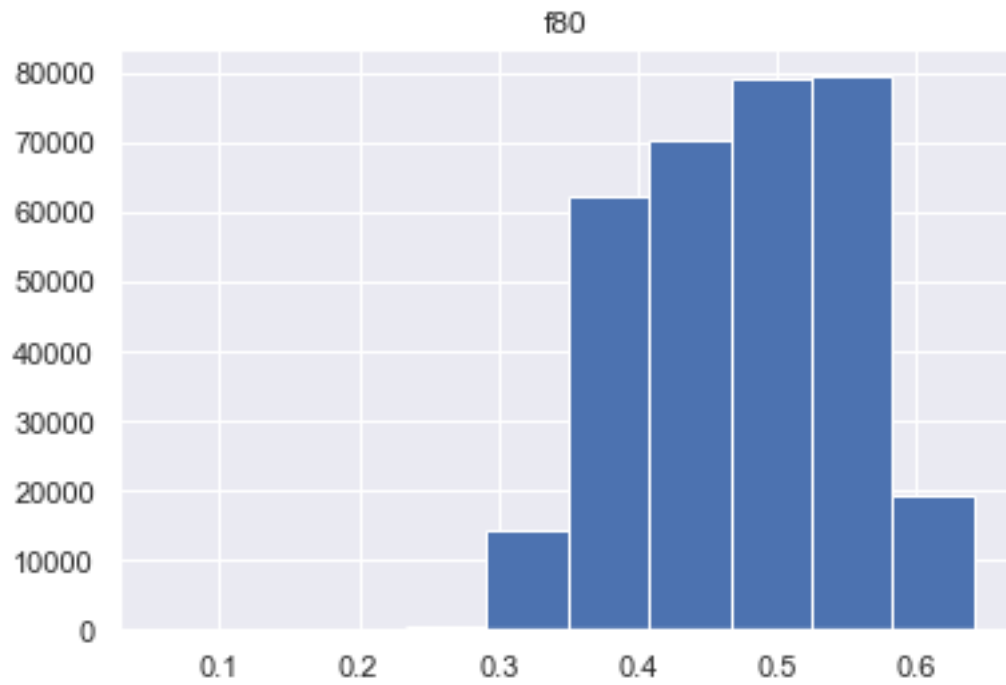
<Figure size 432x288 with 0 Axes>



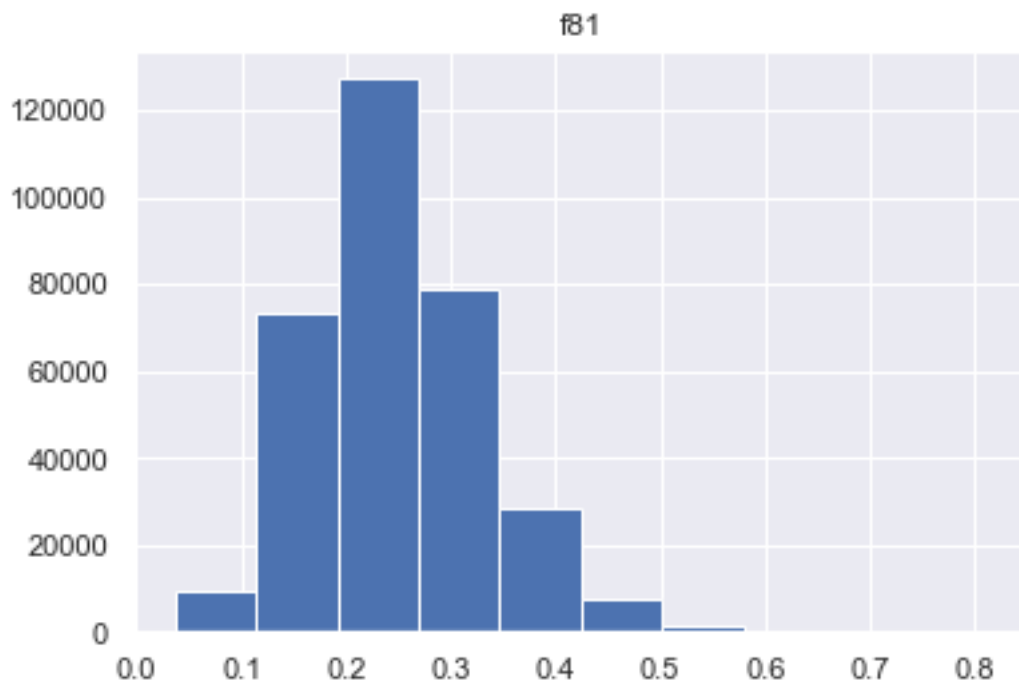
<Figure size 432x288 with 0 Axes>



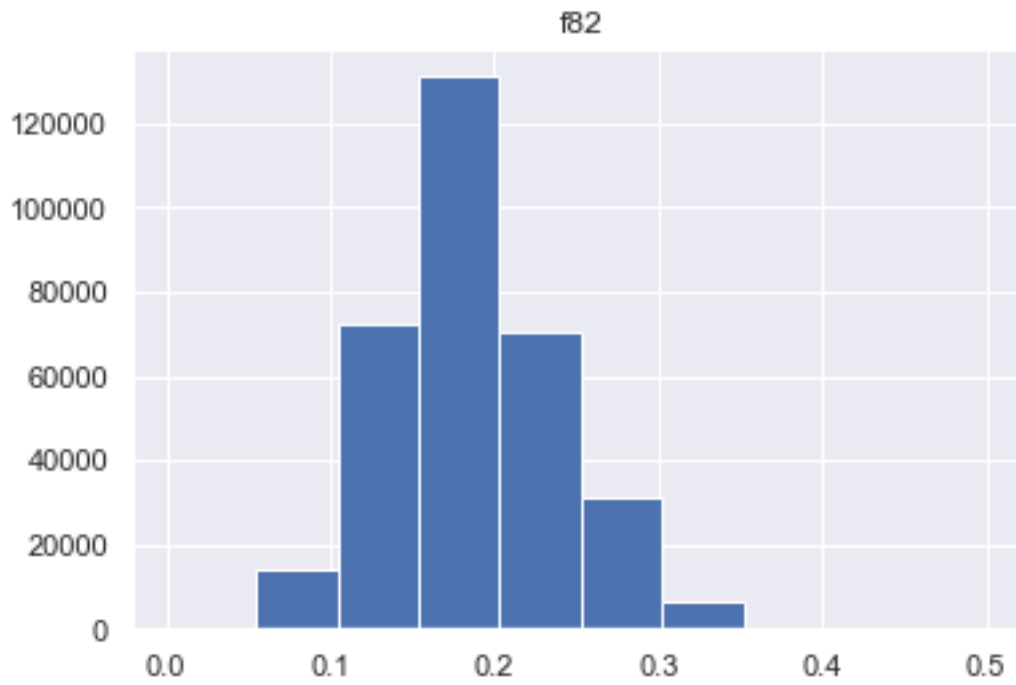
<Figure size 432x288 with 0 Axes>



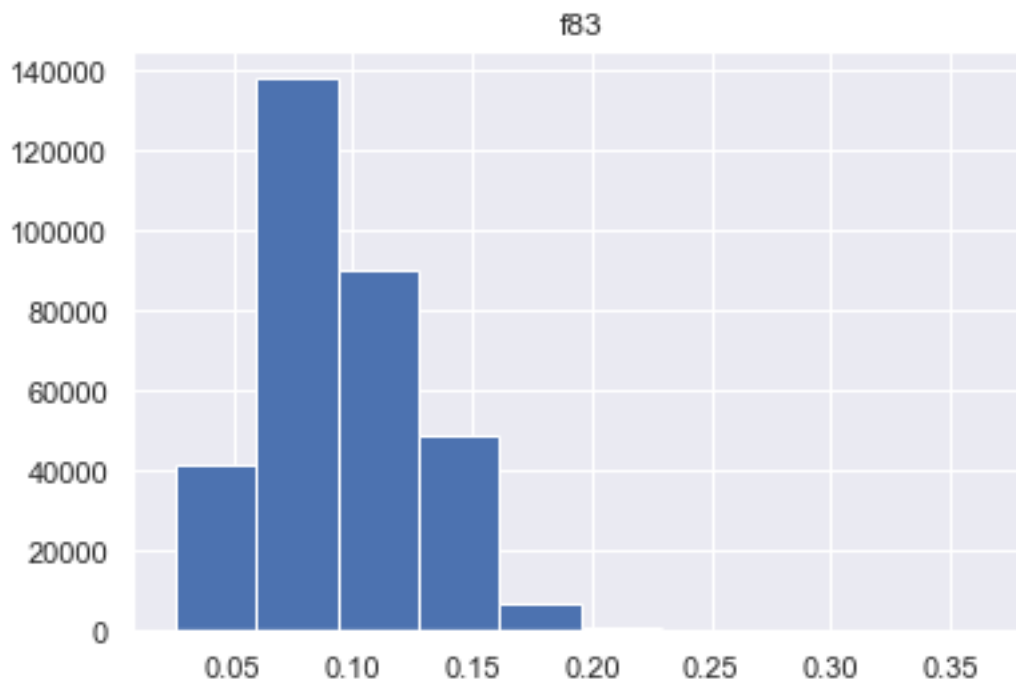
<Figure size 432x288 with 0 Axes>



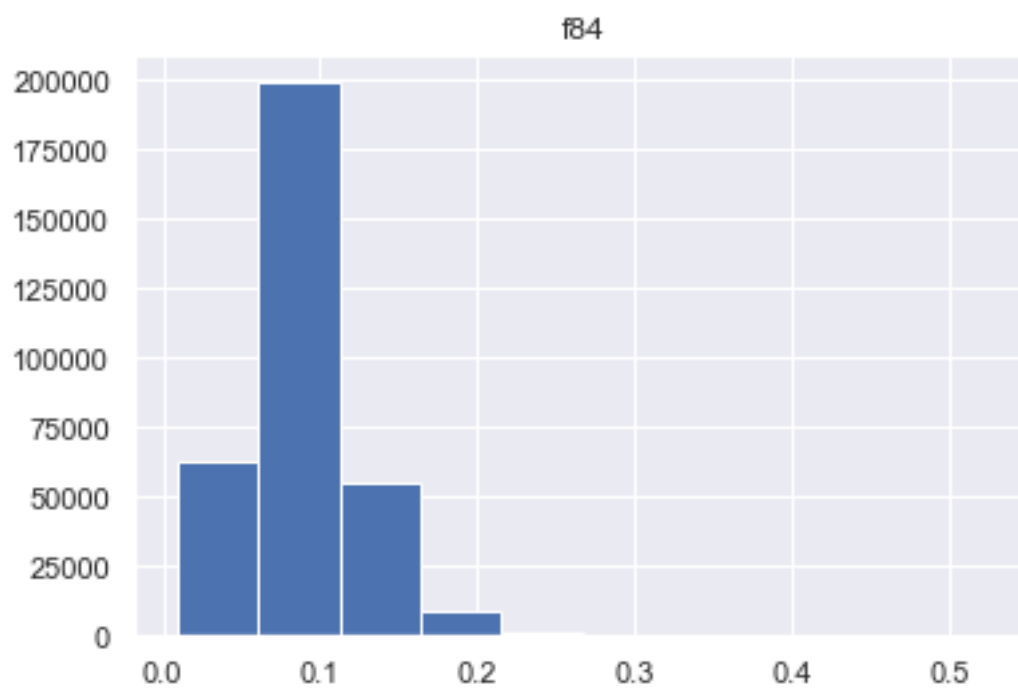
<Figure size 432x288 with 0 Axes>



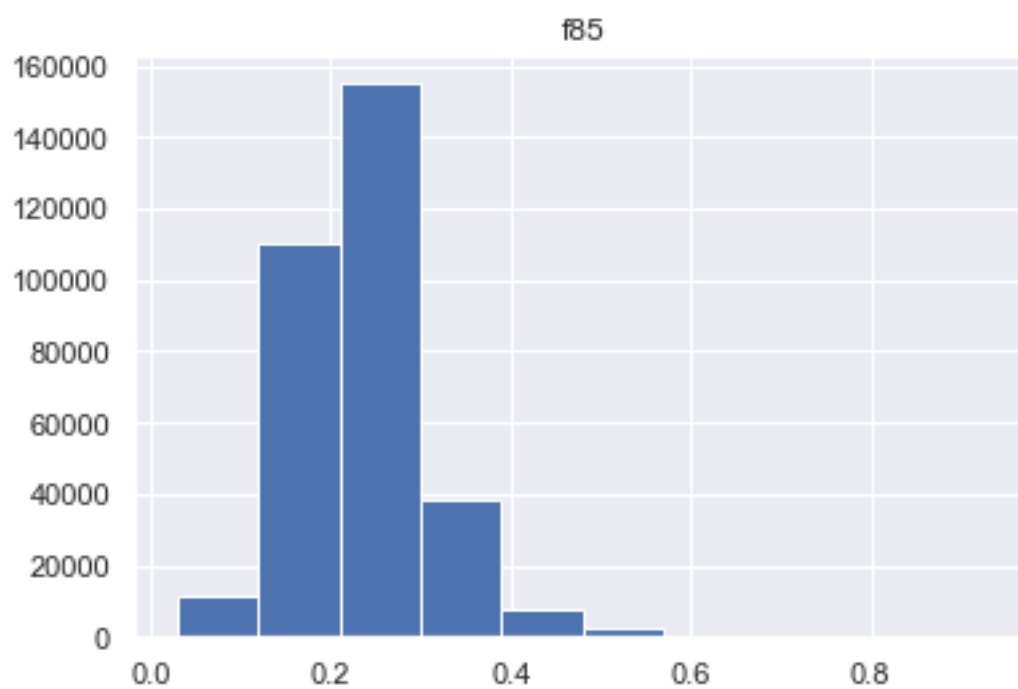
<Figure size 432x288 with 0 Axes>



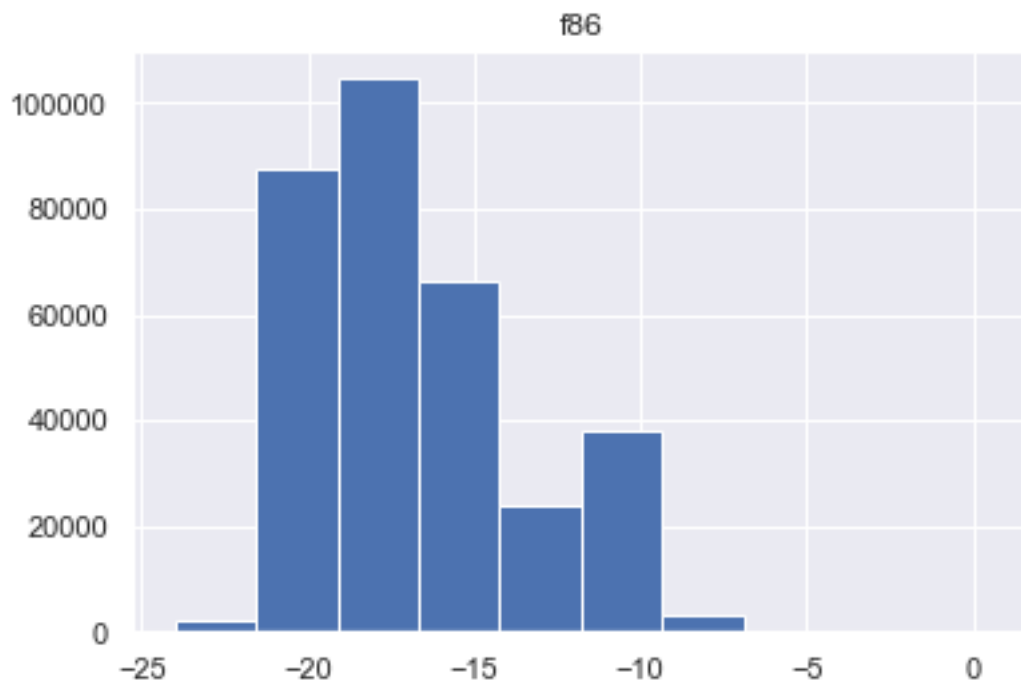
<Figure size 432x288 with 0 Axes>



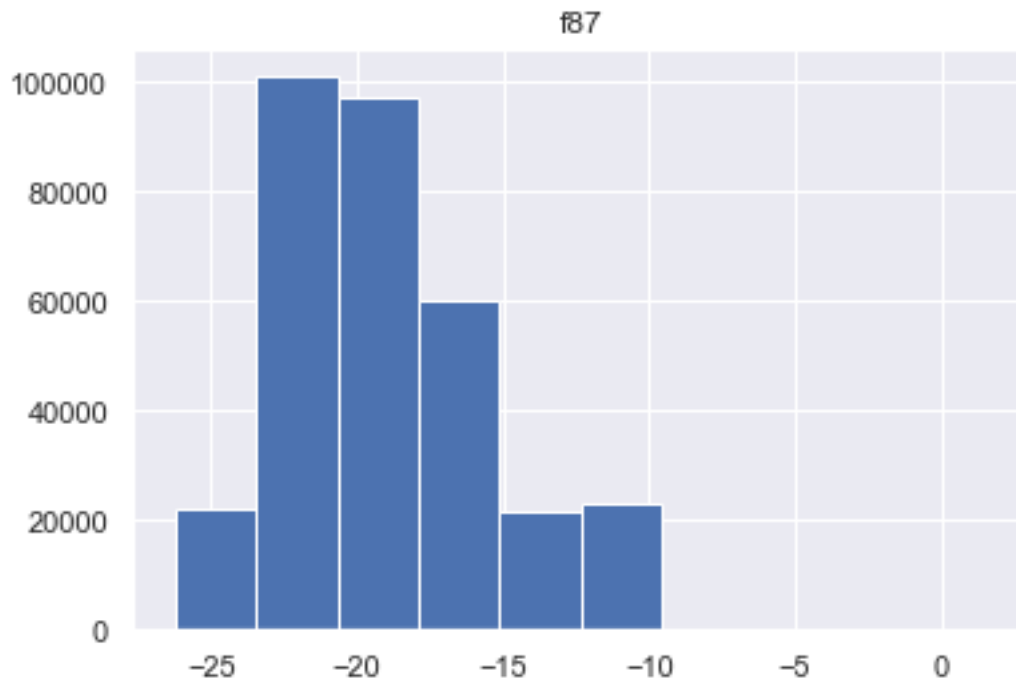
<Figure size 432x288 with 0 Axes>



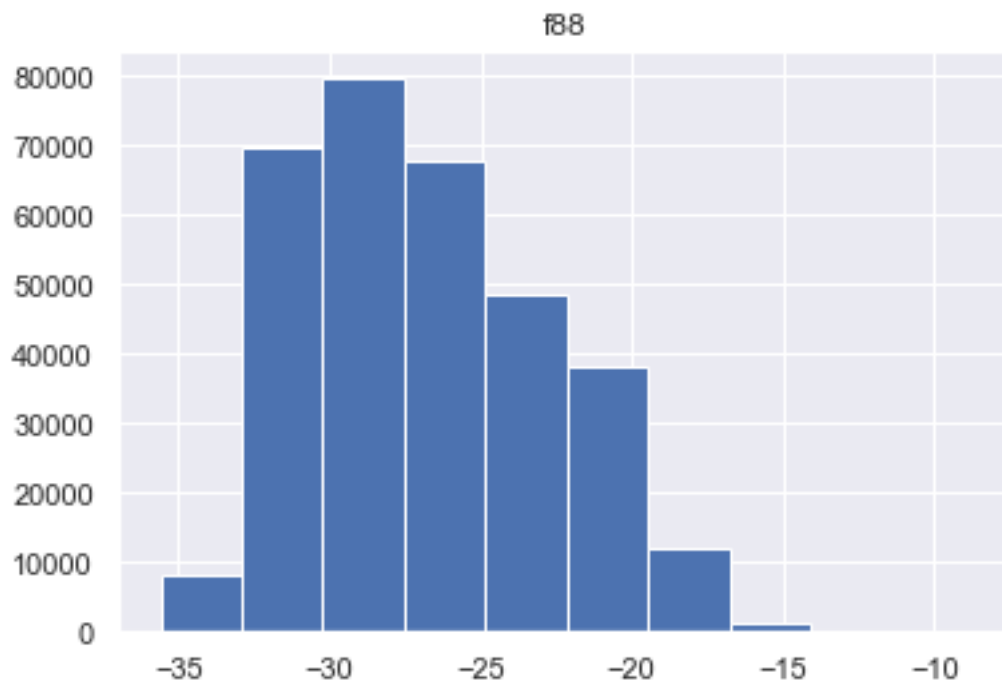
<Figure size 432x288 with 0 Axes>



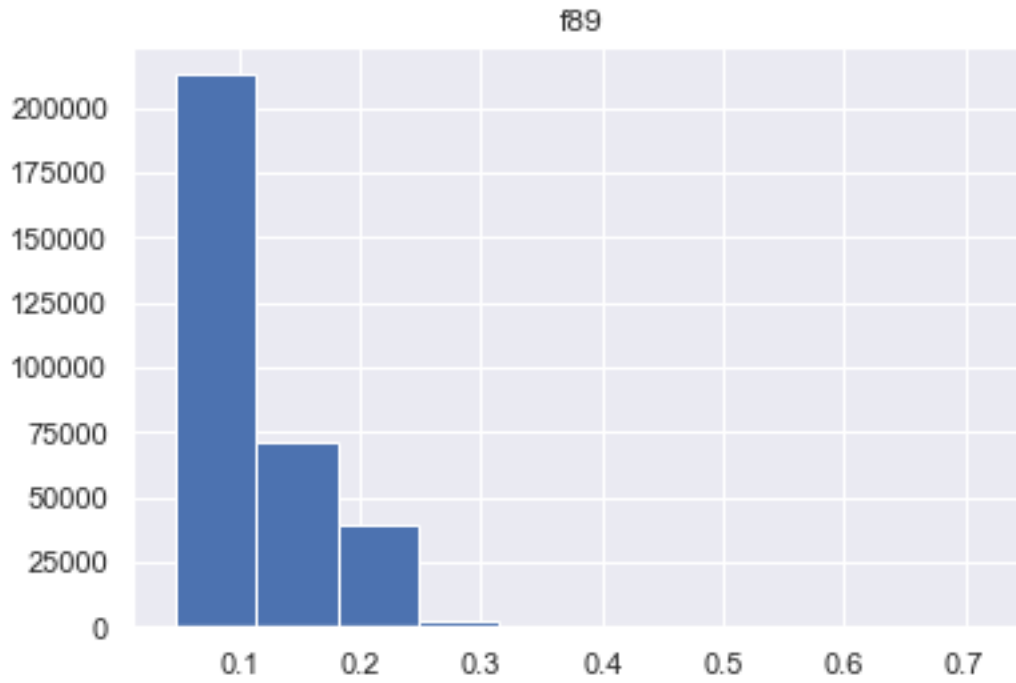
<Figure size 432x288 with 0 Axes>



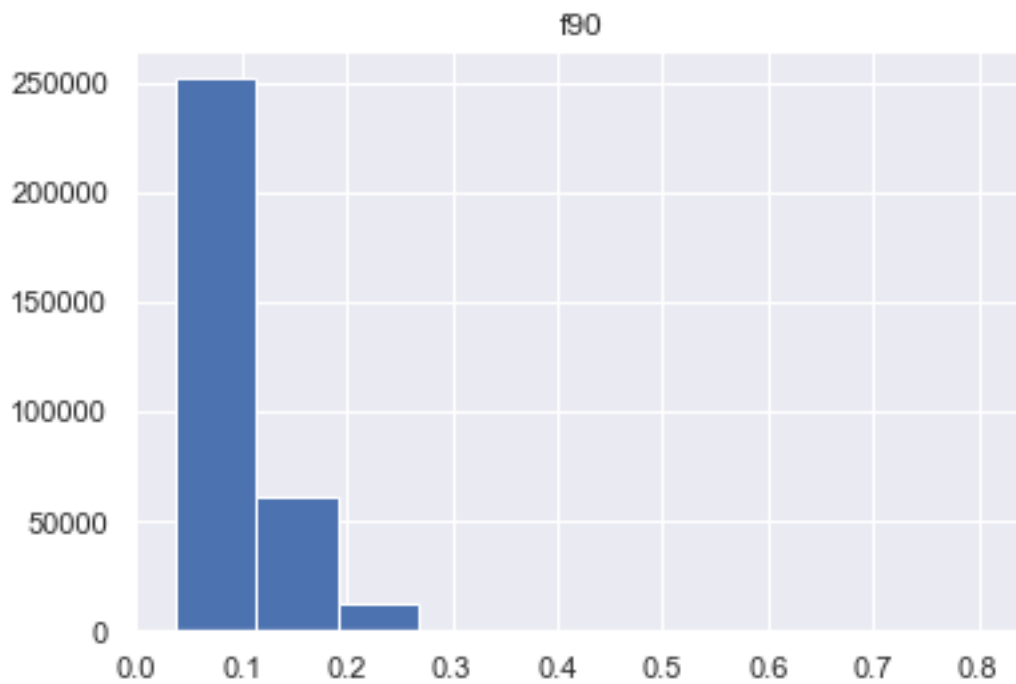
<Figure size 432x288 with 0 Axes>



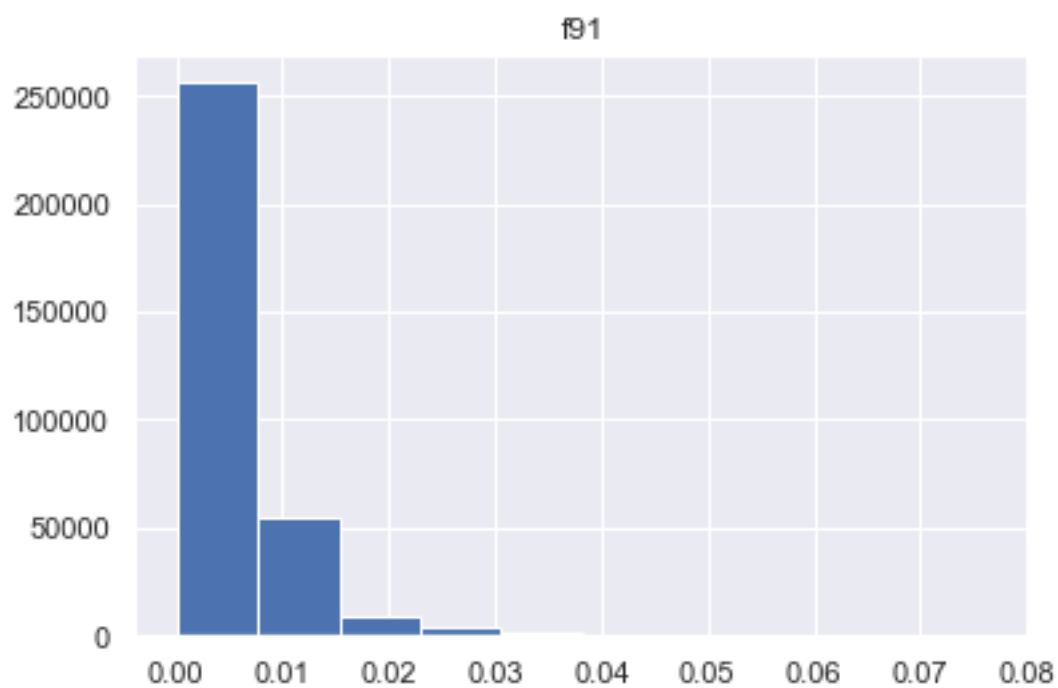
<Figure size 432x288 with 0 Axes>



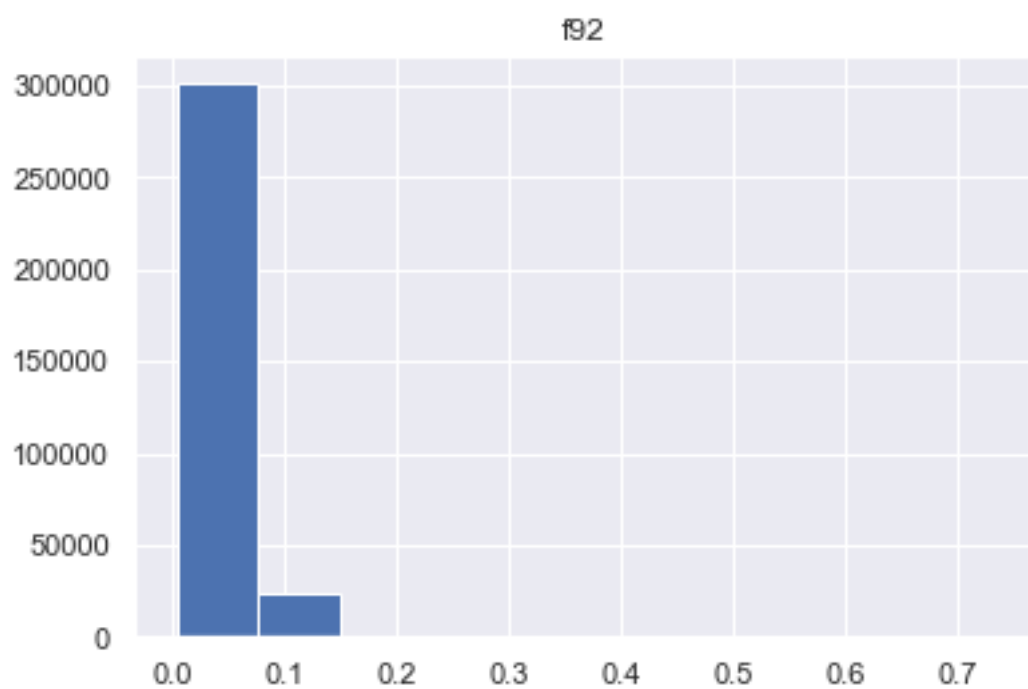
<Figure size 432x288 with 0 Axes>



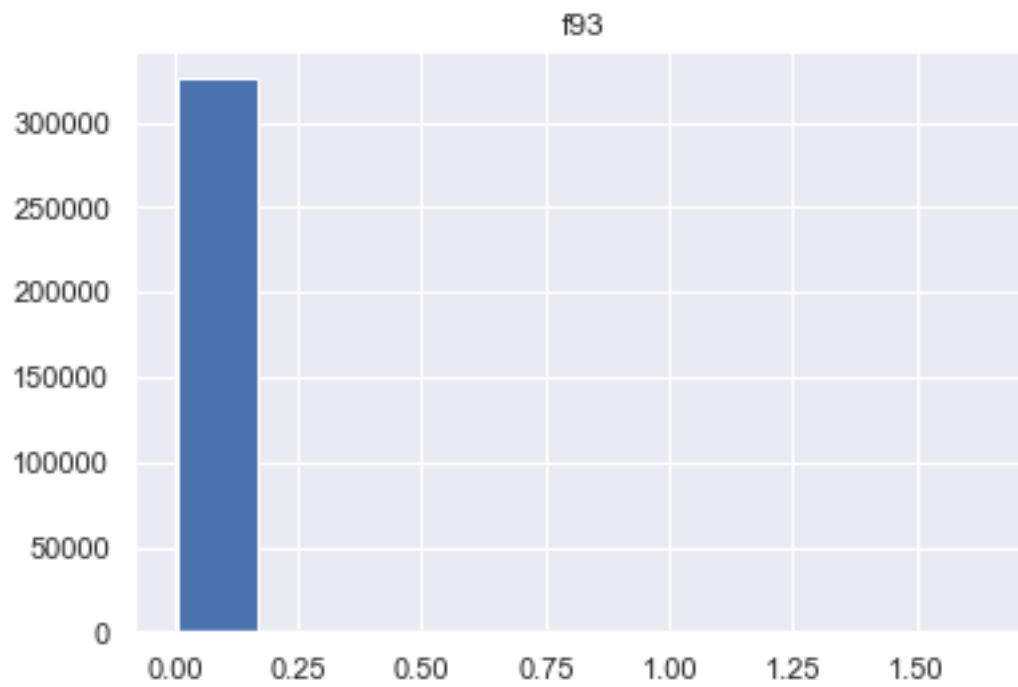
<Figure size 432x288 with 0 Axes>



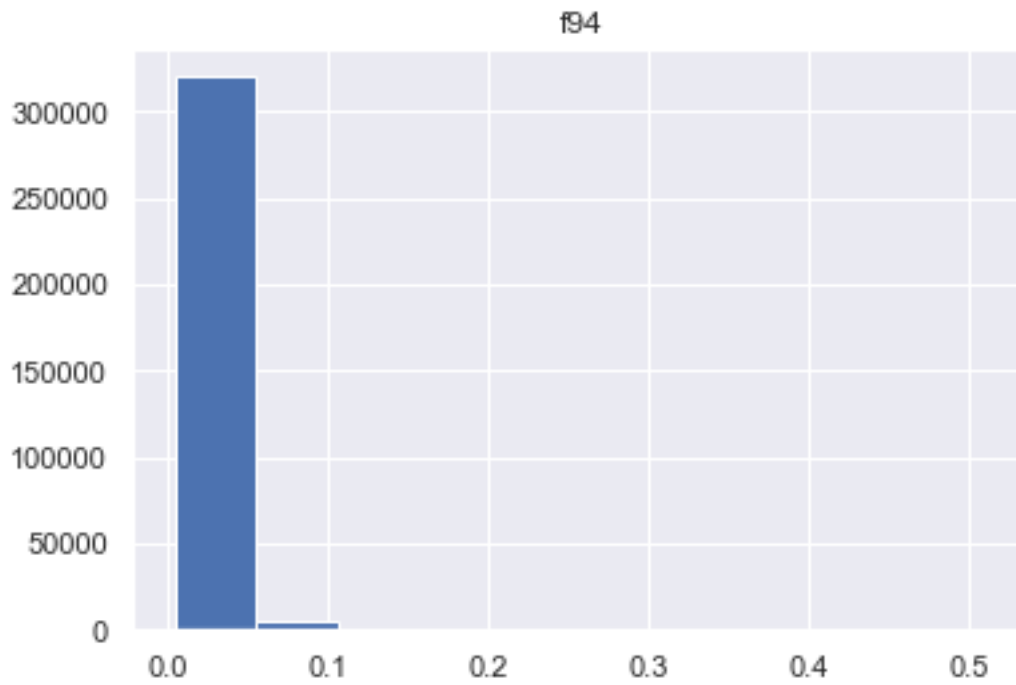
<Figure size 432x288 with 0 Axes>



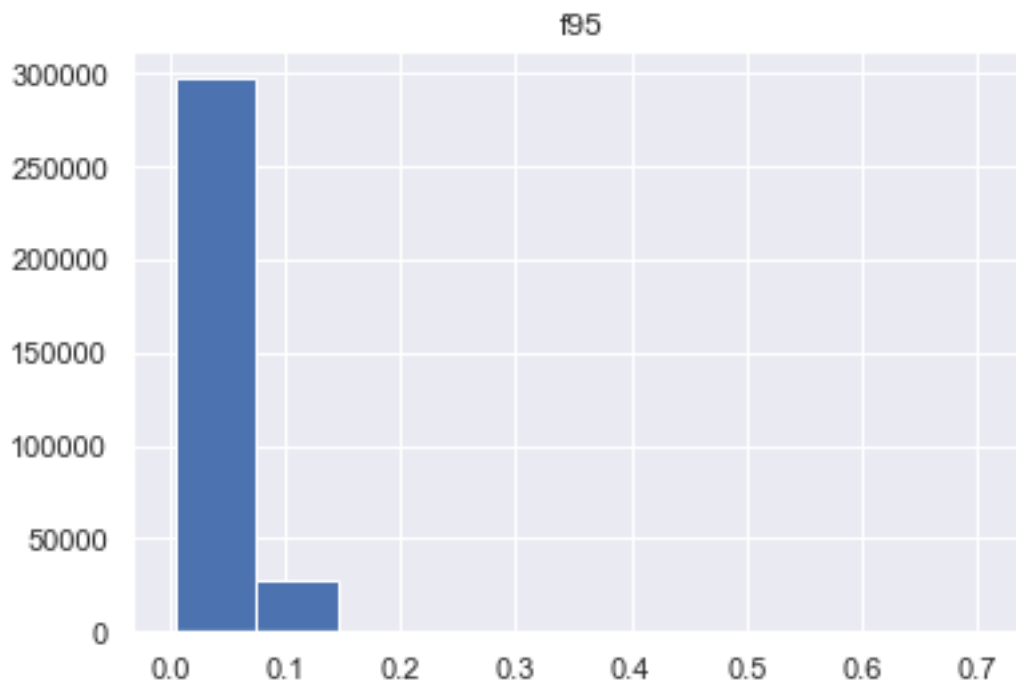
<Figure size 432x288 with 0 Axes>



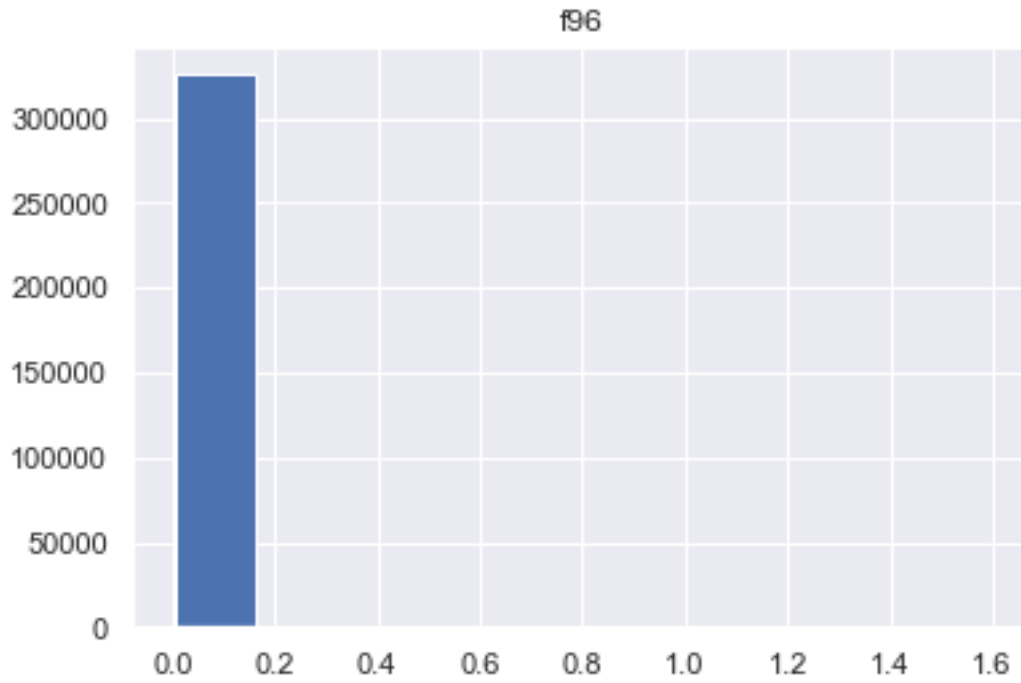
<Figure size 432x288 with 0 Axes>



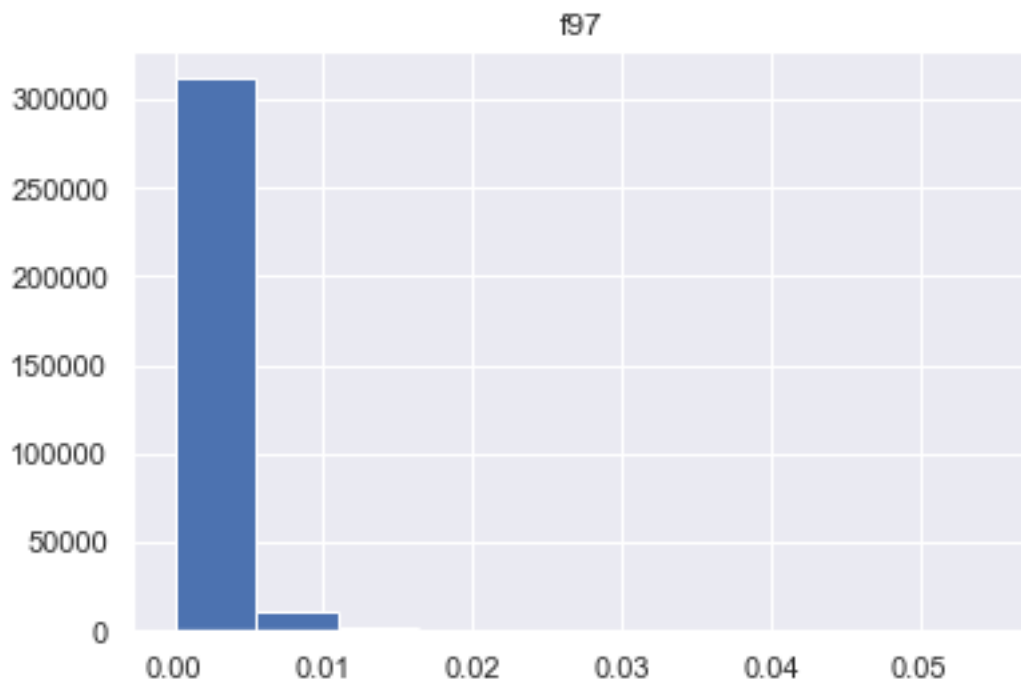
<Figure size 432x288 with 0 Axes>



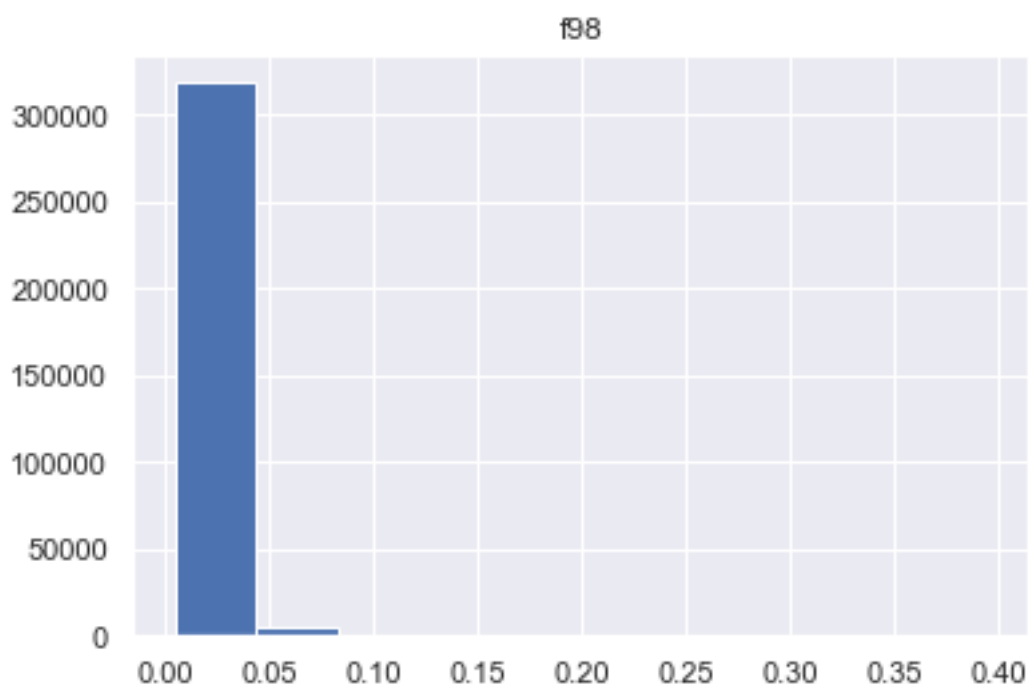
<Figure size 432x288 with 0 Axes>



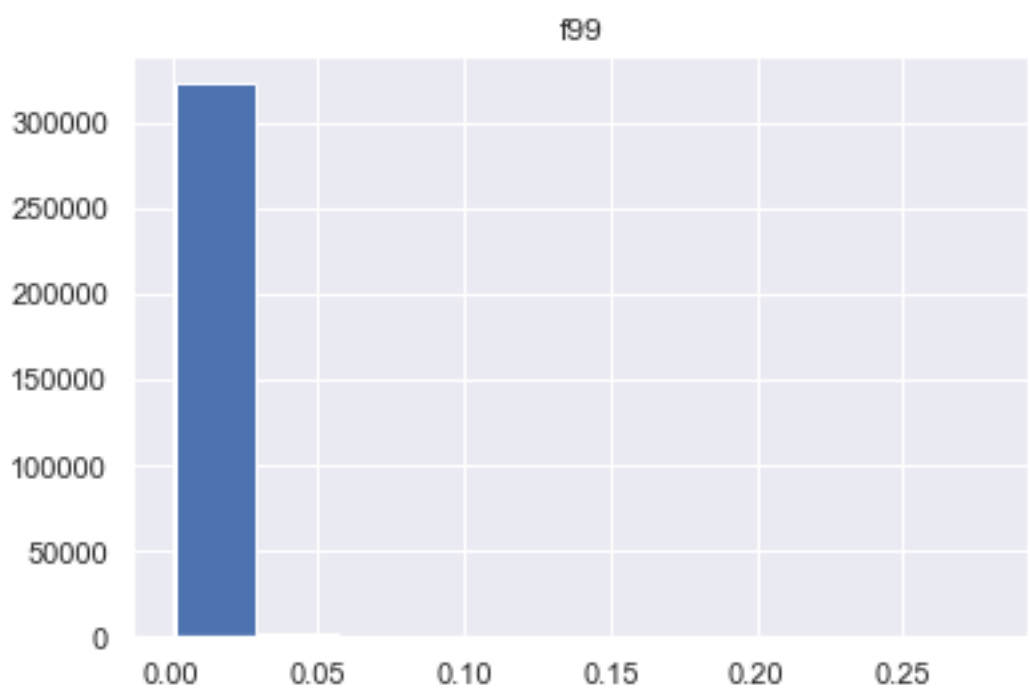
<Figure size 432x288 with 0 Axes>



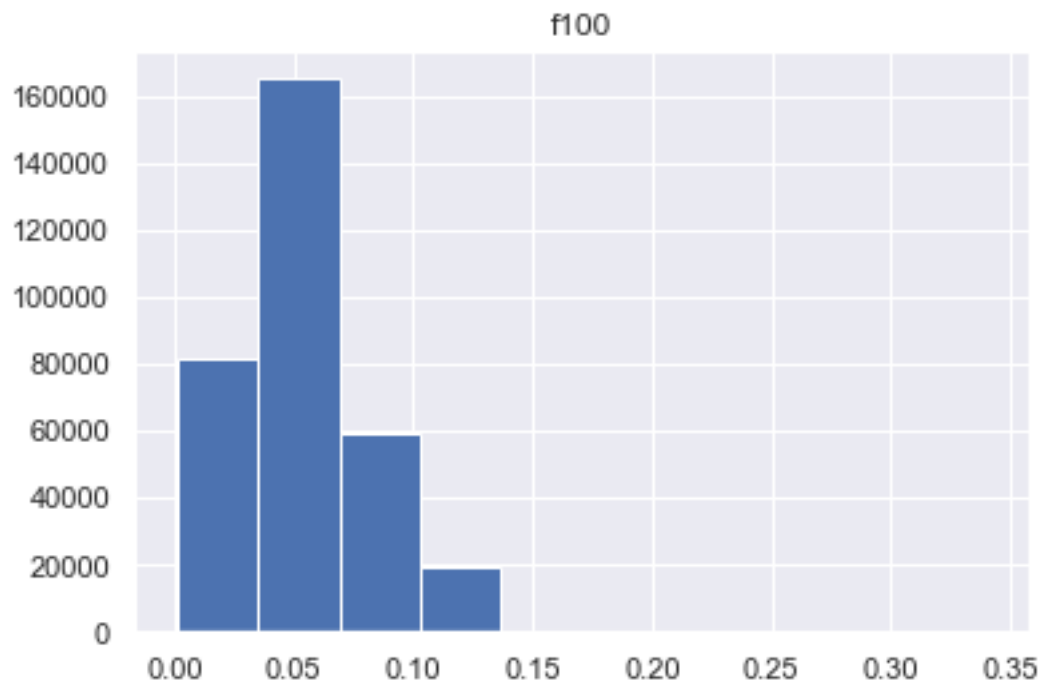
<Figure size 432x288 with 0 Axes>



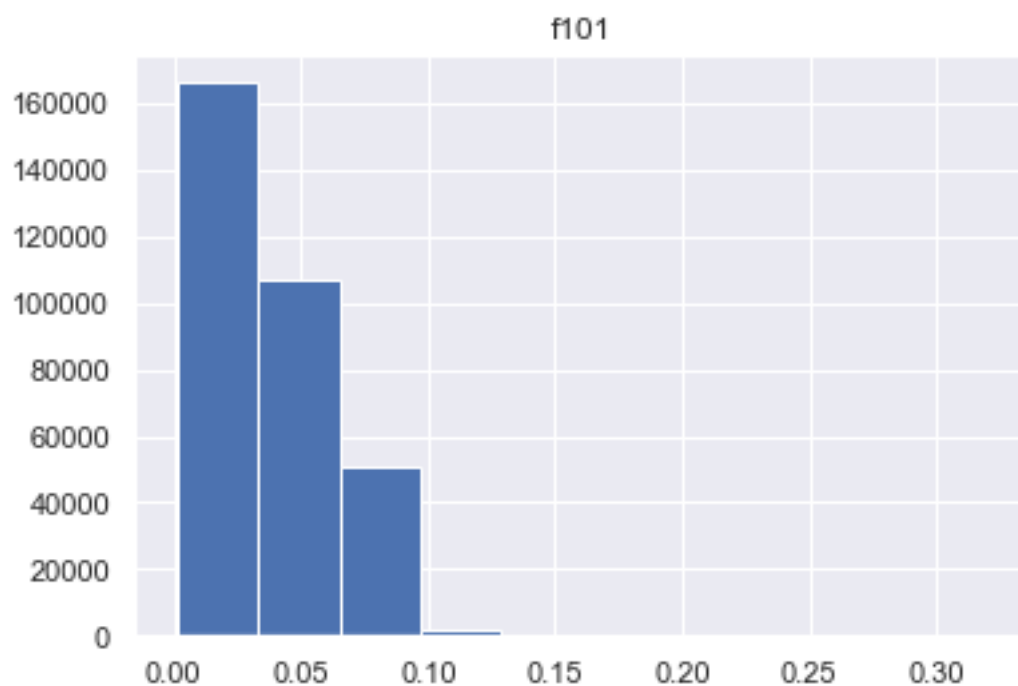
<Figure size 432x288 with 0 Axes>



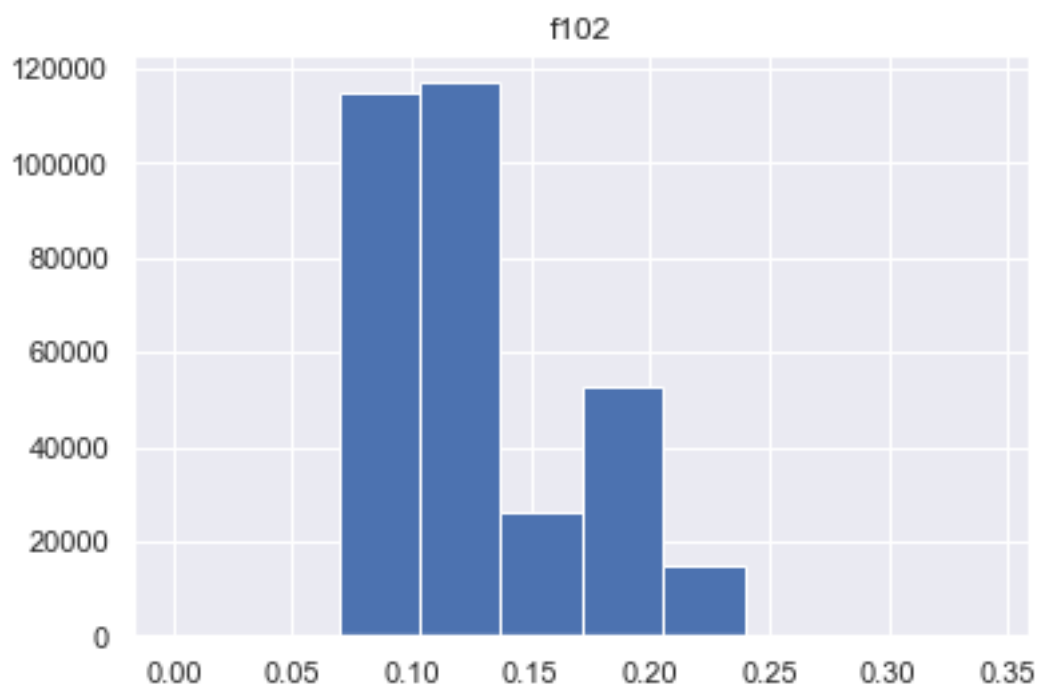
<Figure size 432x288 with 0 Axes>



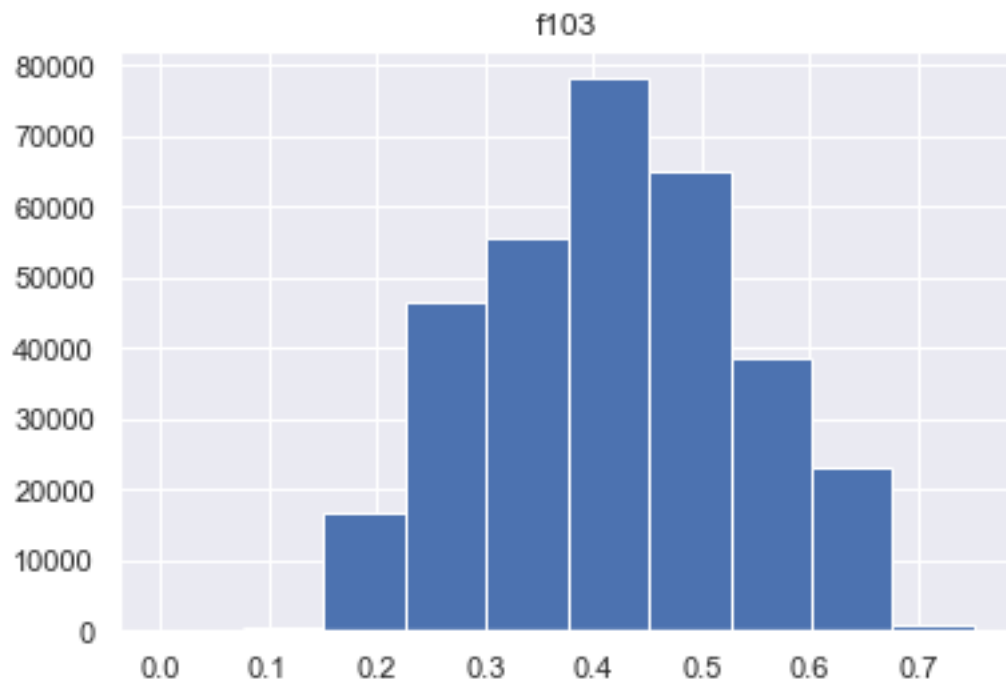
<Figure size 432x288 with 0 Axes>



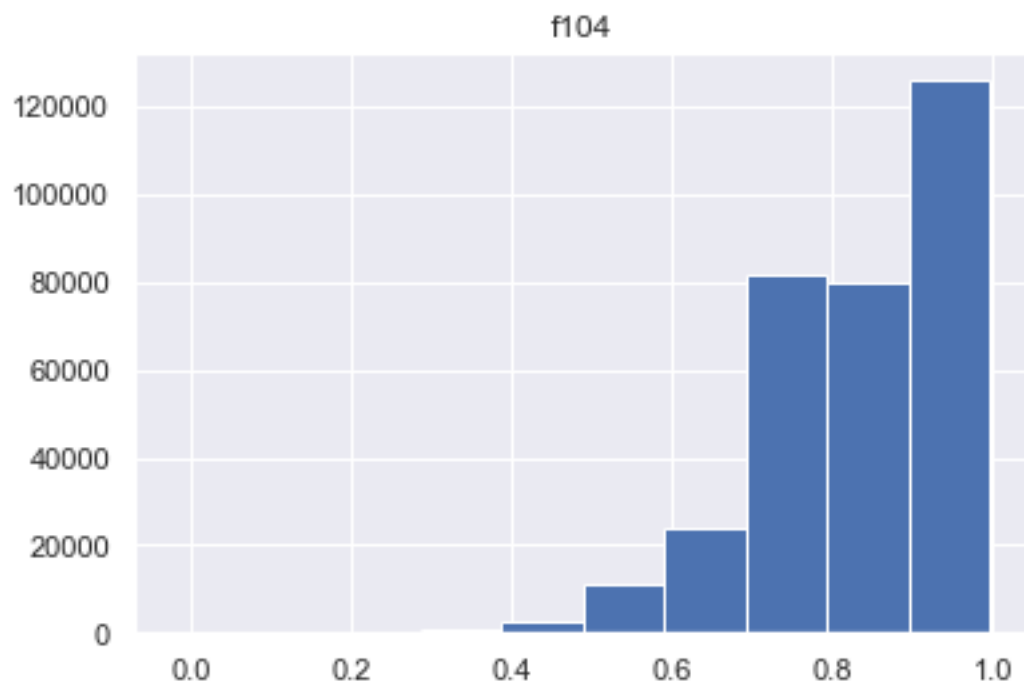
<Figure size 432x288 with 0 Axes>



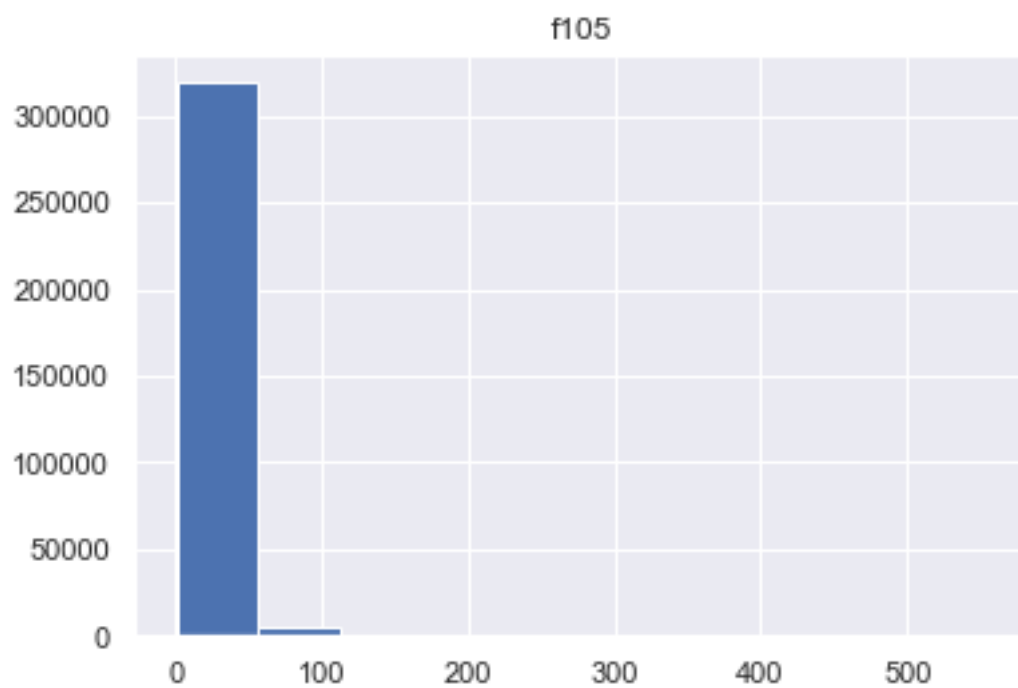
<Figure size 432x288 with 0 Axes>



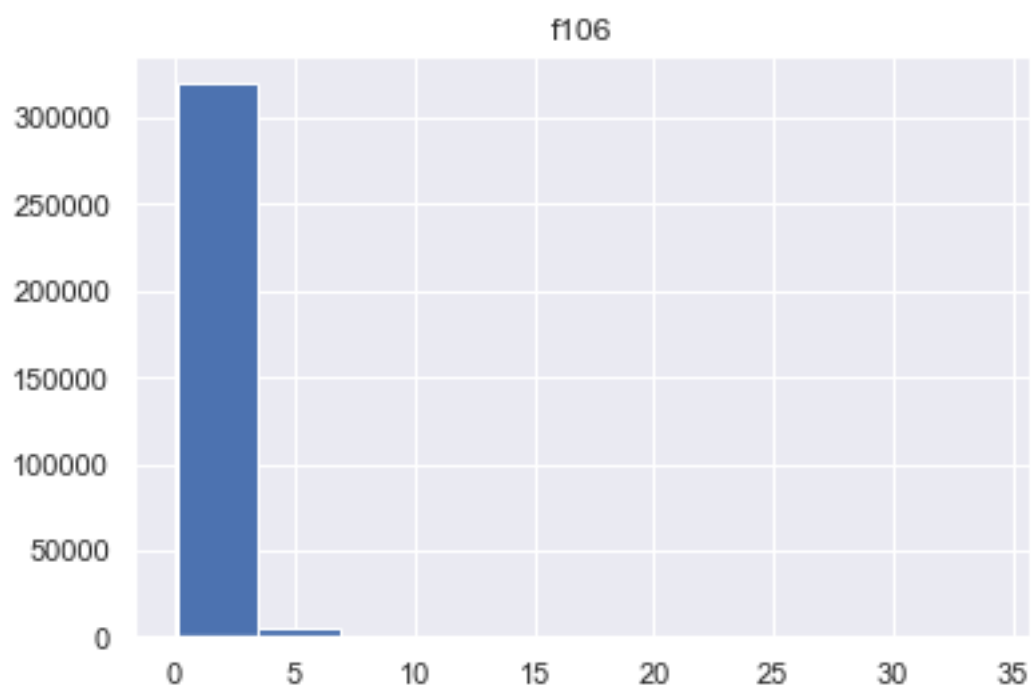
<Figure size 432x288 with 0 Axes>



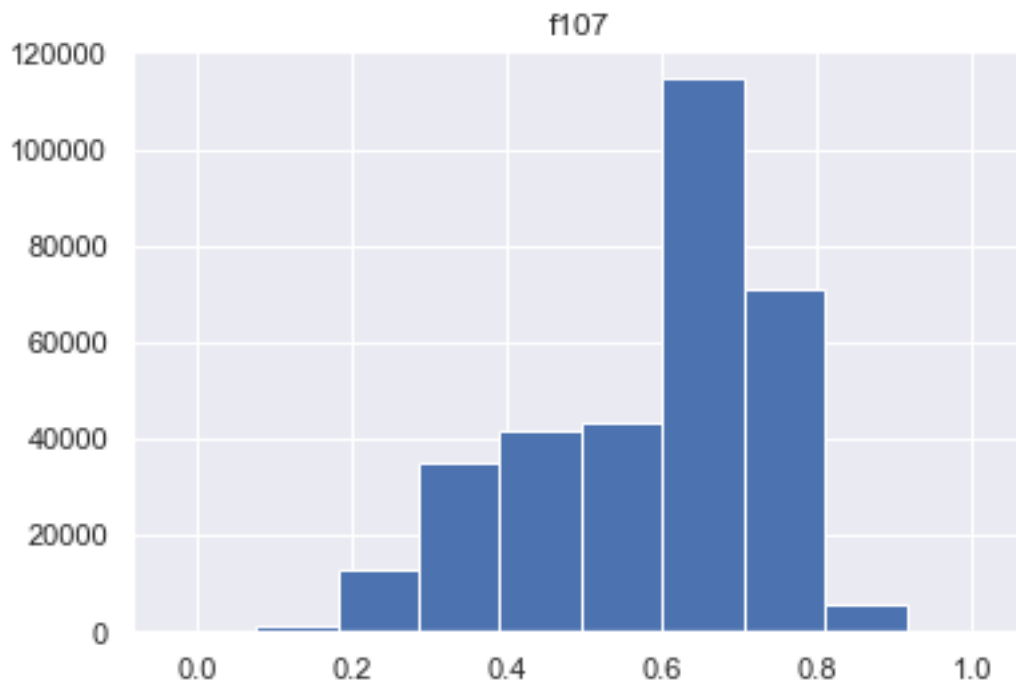
<Figure size 432x288 with 0 Axes>



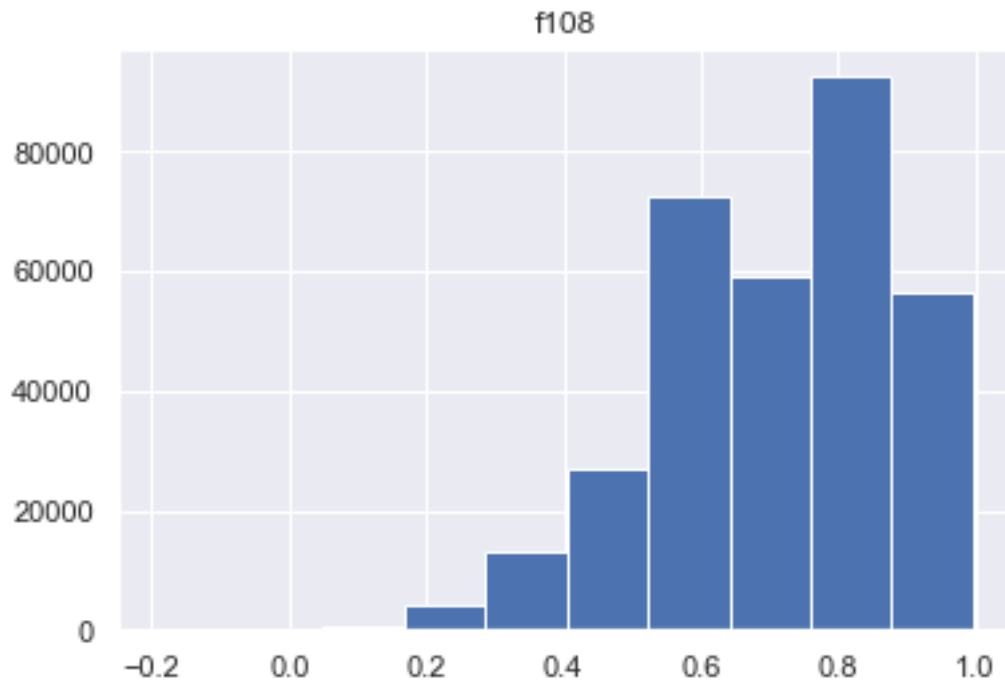
<Figure size 432x288 with 0 Axes>



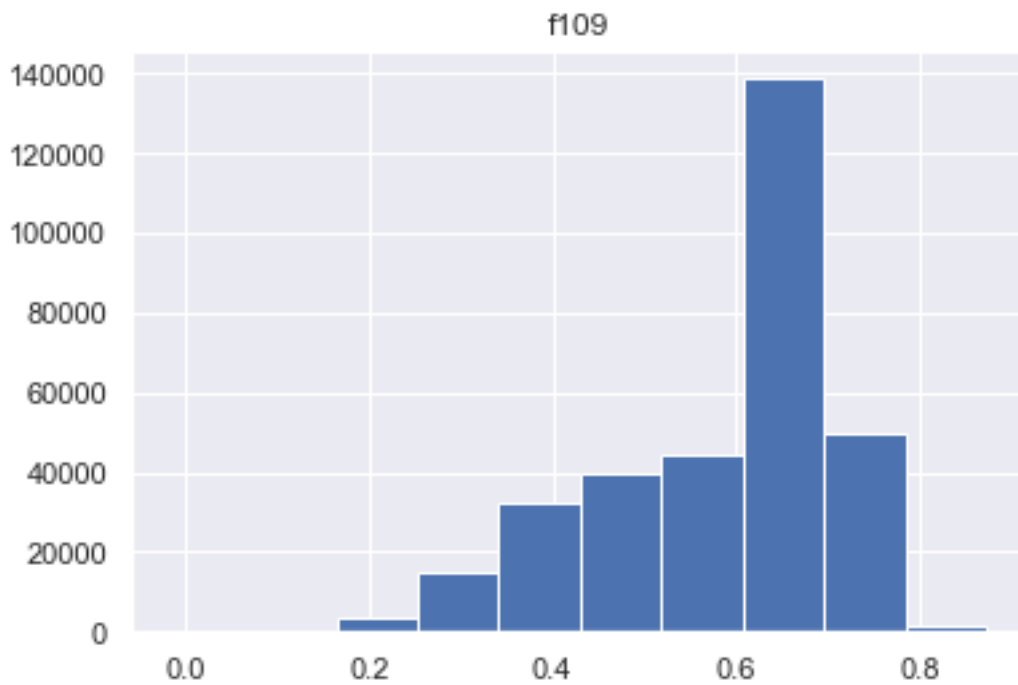
<Figure size 432x288 with 0 Axes>



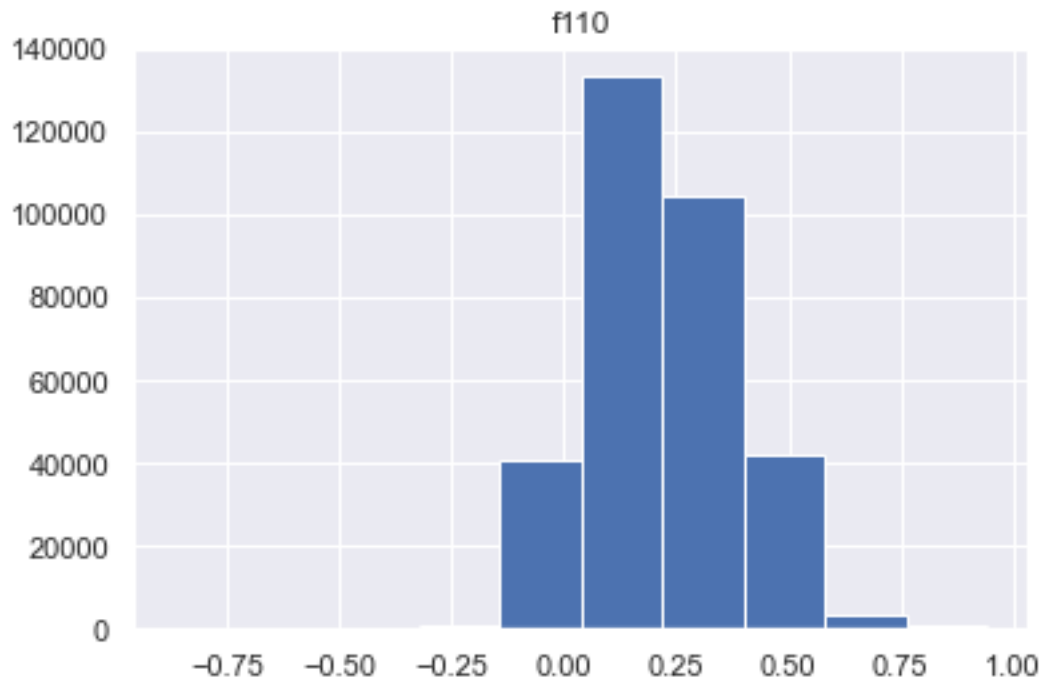
<Figure size 432x288 with 0 Axes>



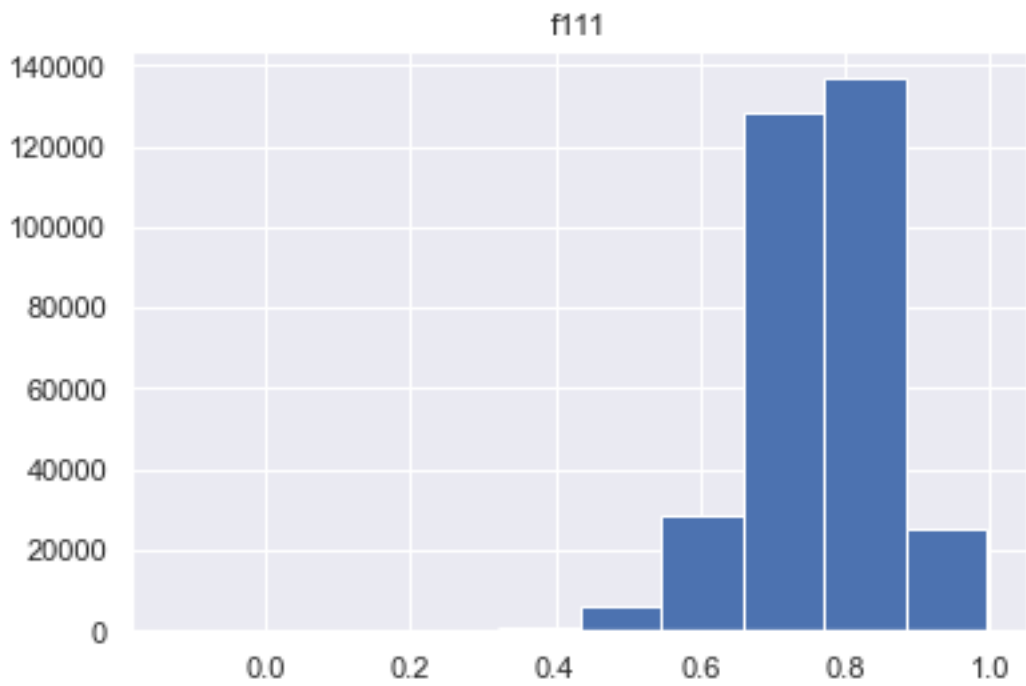
<Figure size 432x288 with 0 Axes>



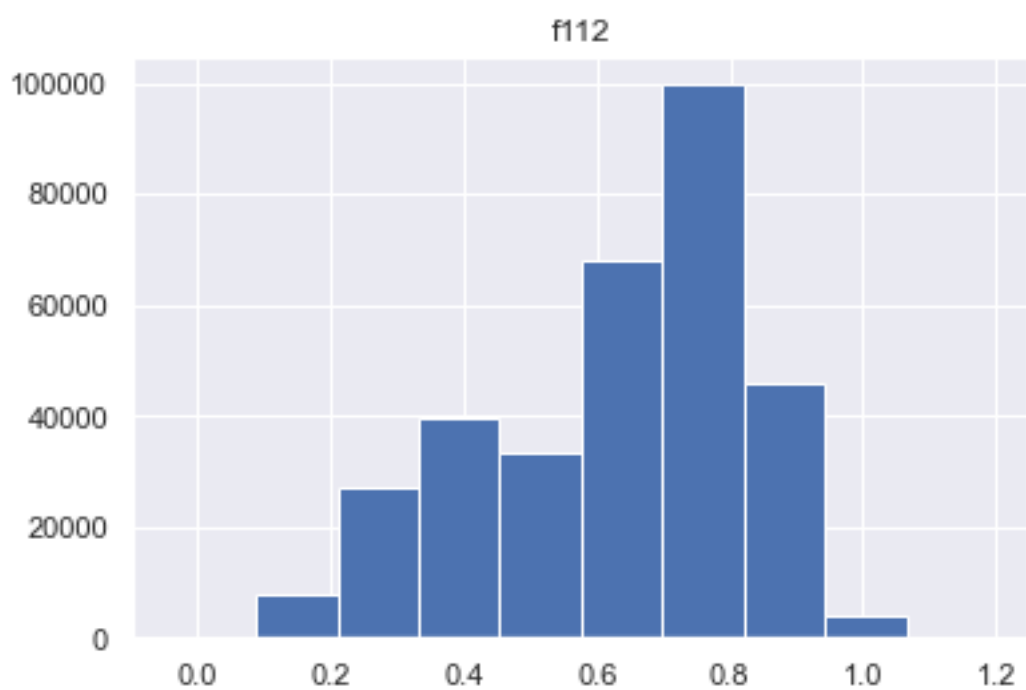
<Figure size 432x288 with 0 Axes>



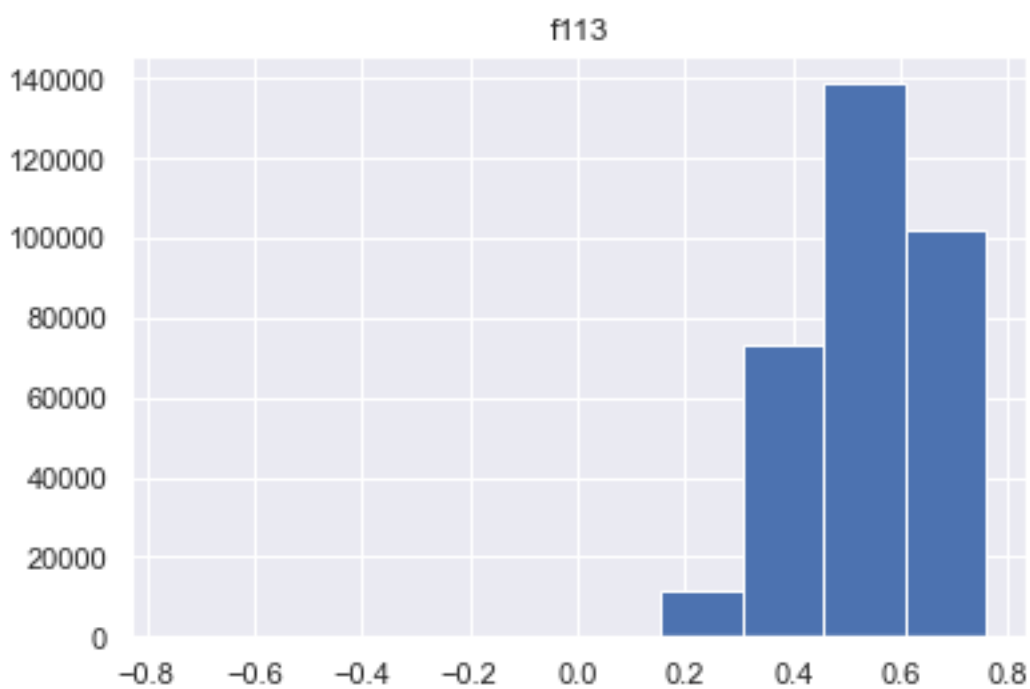
<Figure size 432x288 with 0 Axes>



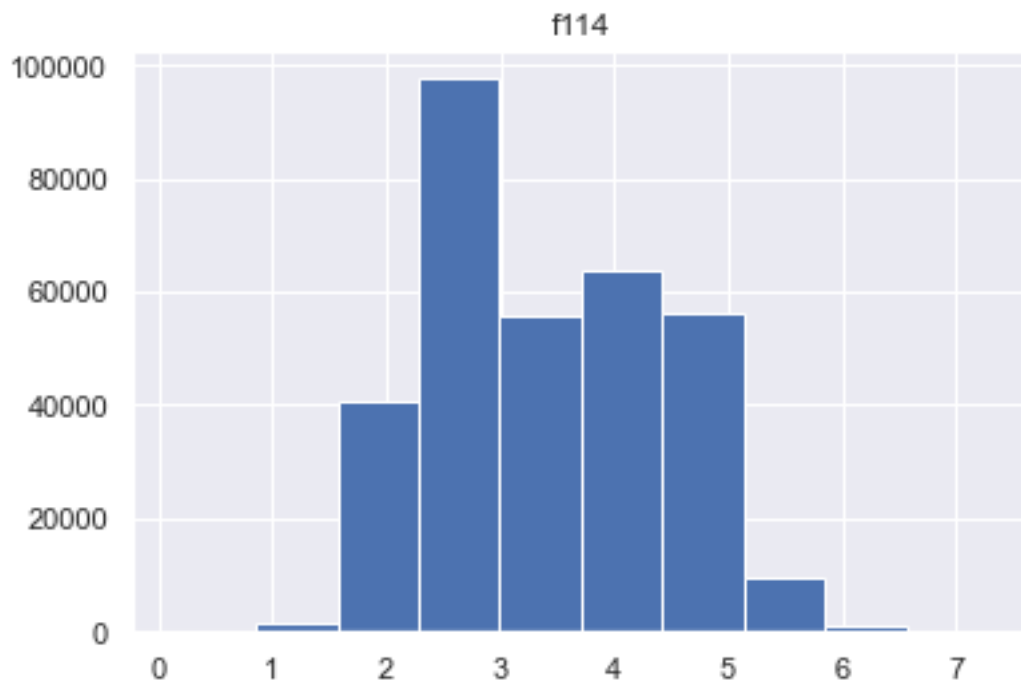
<Figure size 432x288 with 0 Axes>



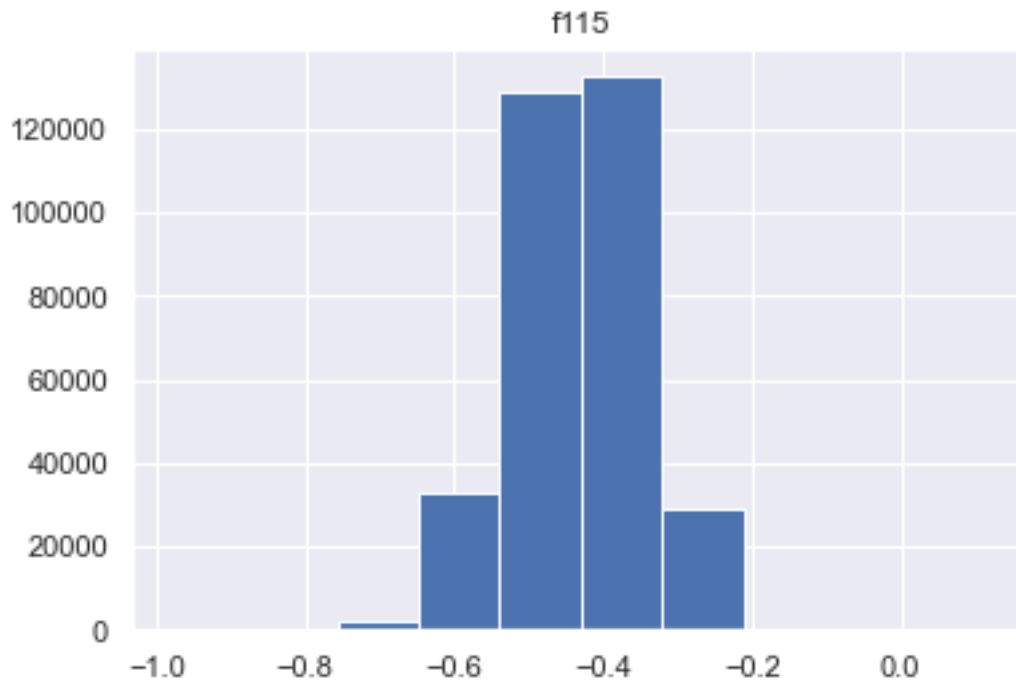
<Figure size 432x288 with 0 Axes>



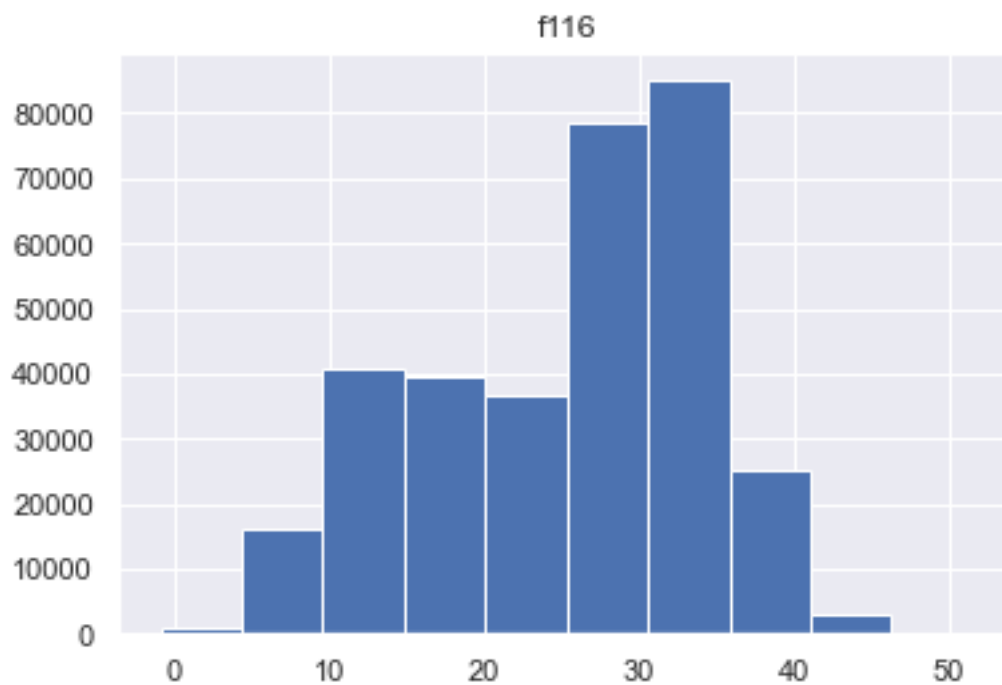
<Figure size 432x288 with 0 Axes>



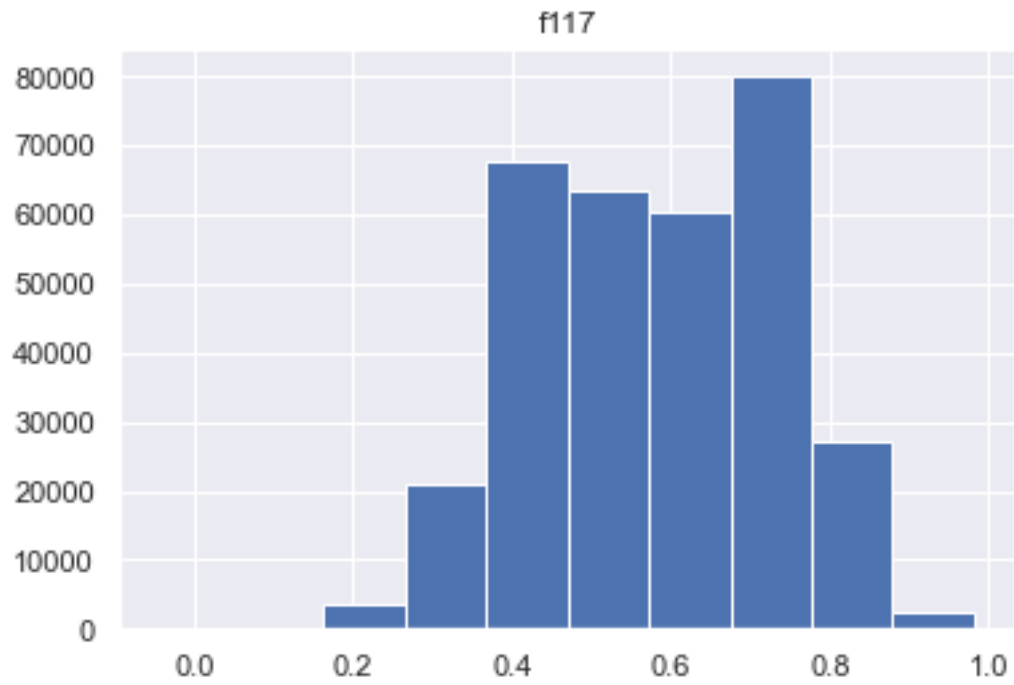
<Figure size 432x288 with 0 Axes>



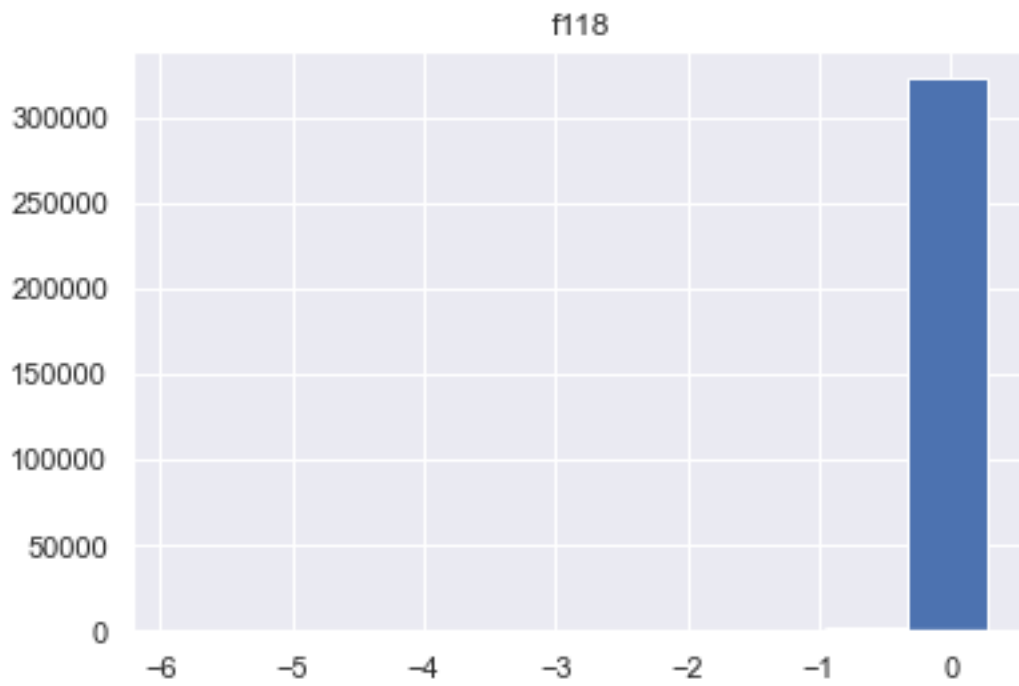
<Figure size 432x288 with 0 Axes>



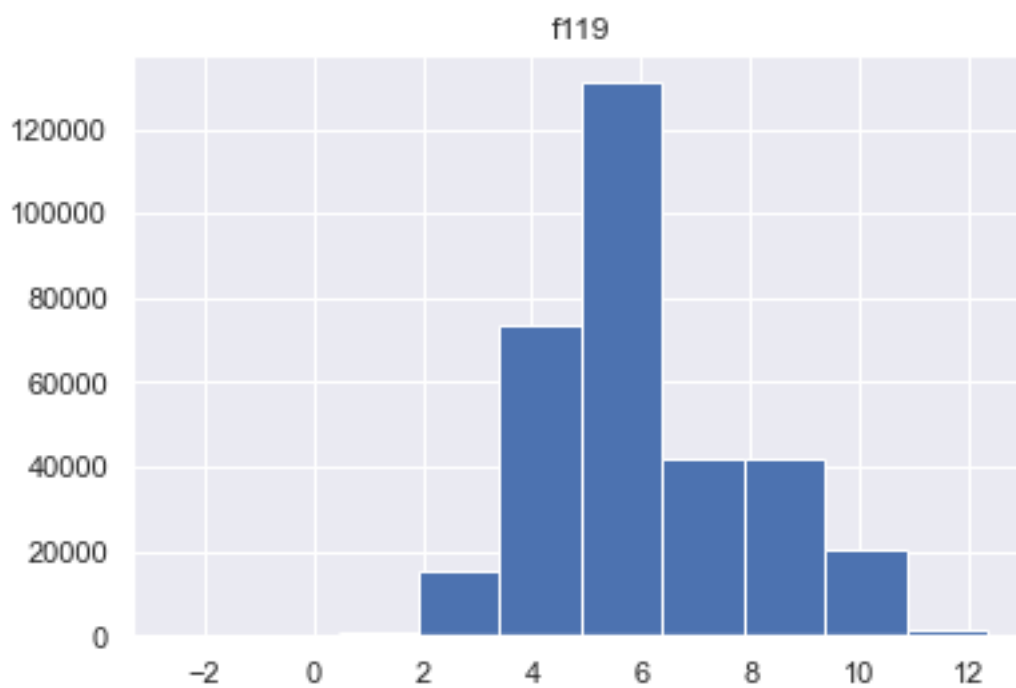
<Figure size 432x288 with 0 Axes>



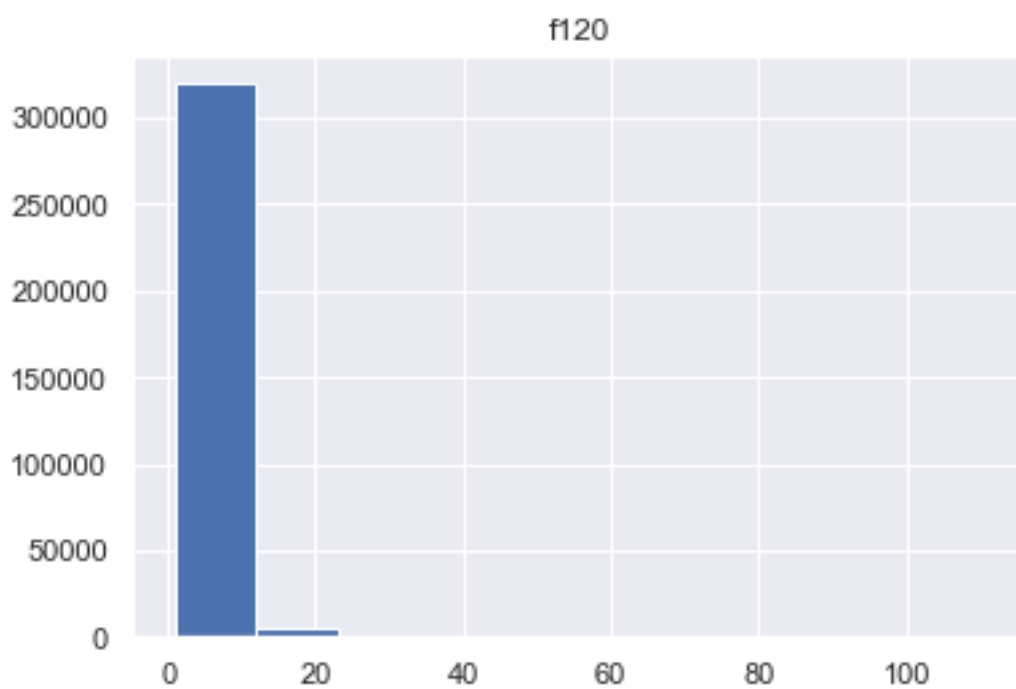
<Figure size 432x288 with 0 Axes>



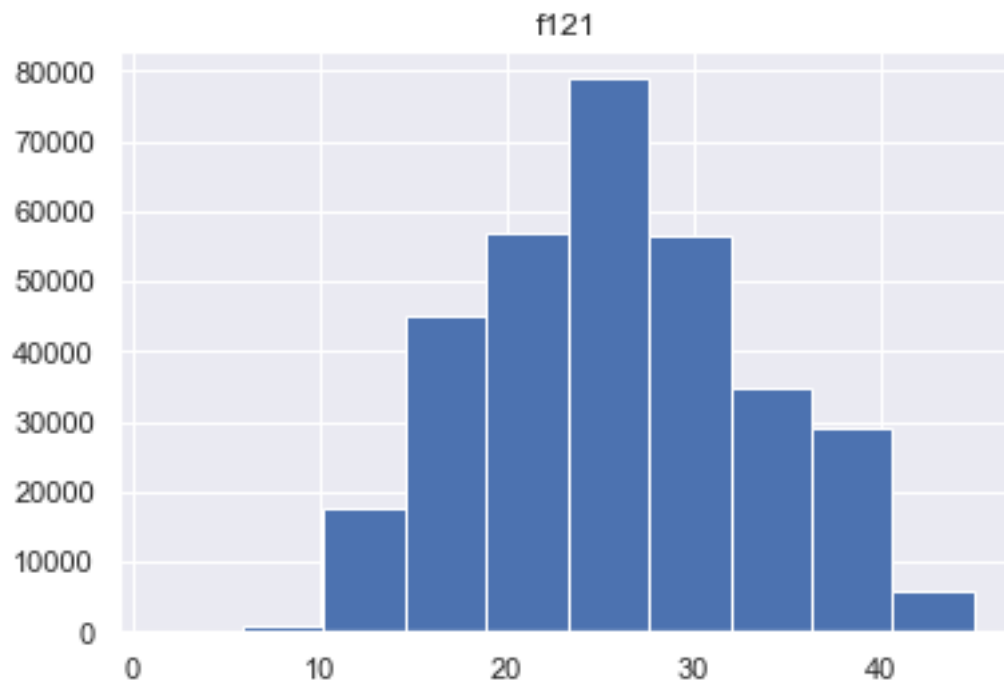
<Figure size 432x288 with 0 Axes>



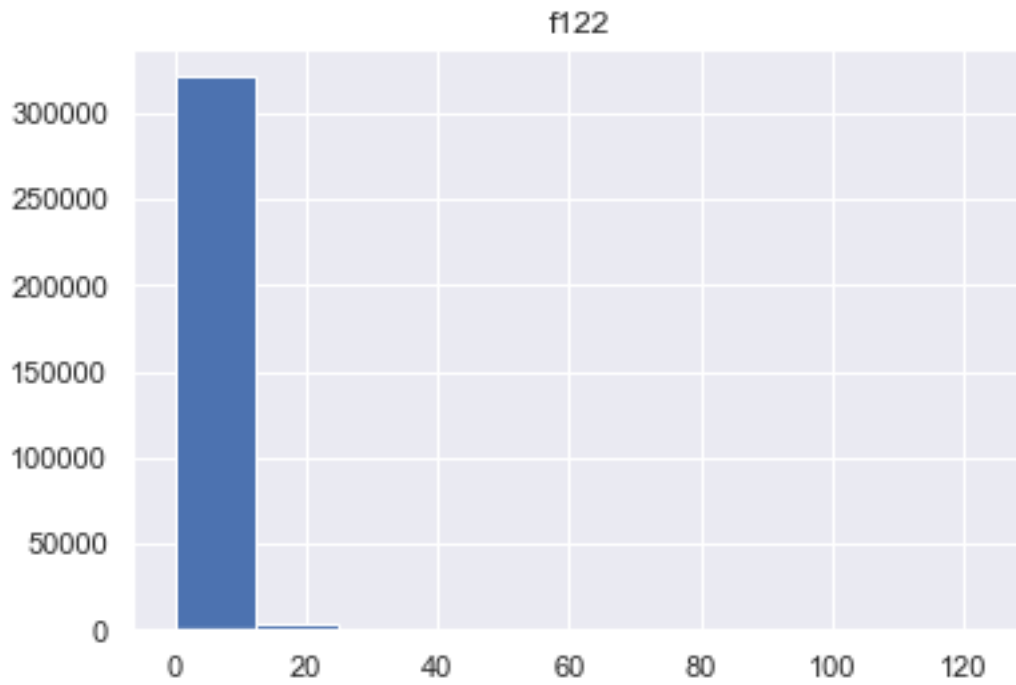
<Figure size 432x288 with 0 Axes>



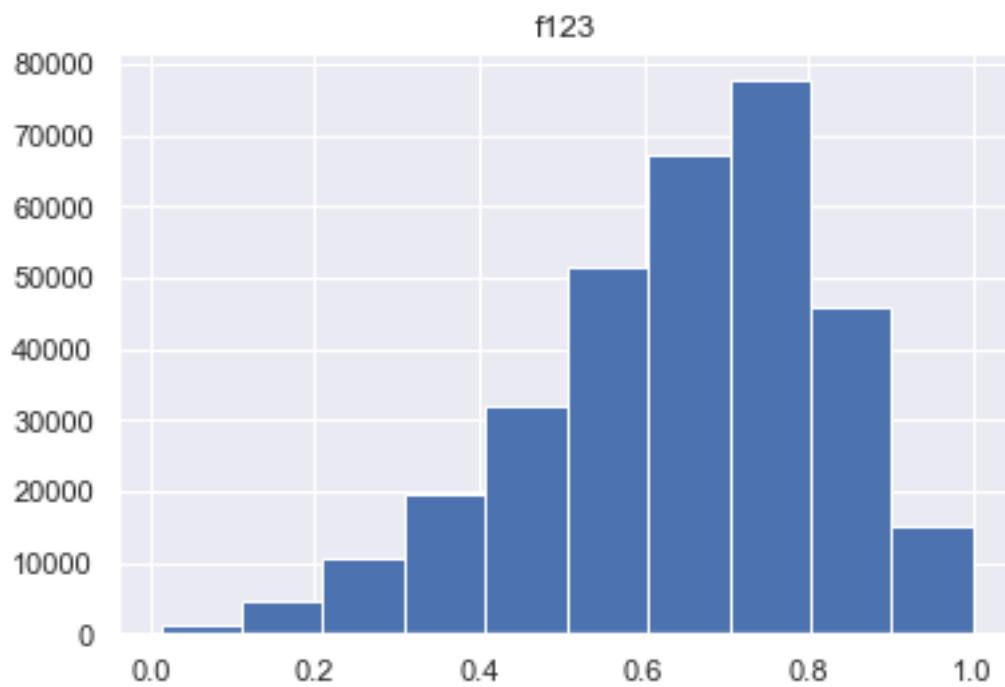
<Figure size 432x288 with 0 Axes>



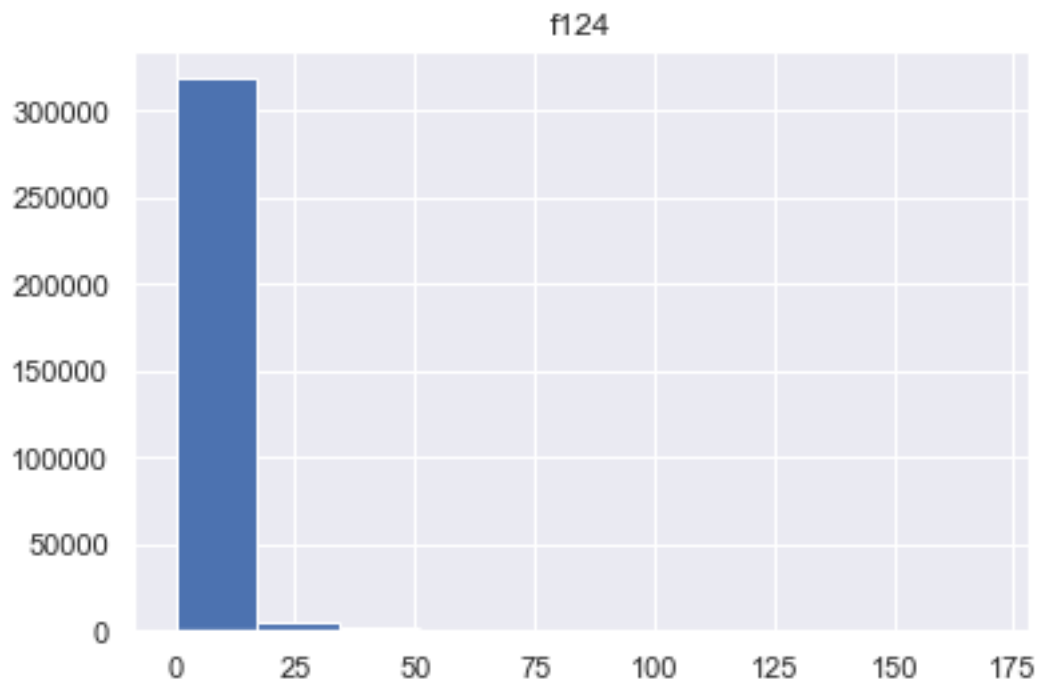
<Figure size 432x288 with 0 Axes>



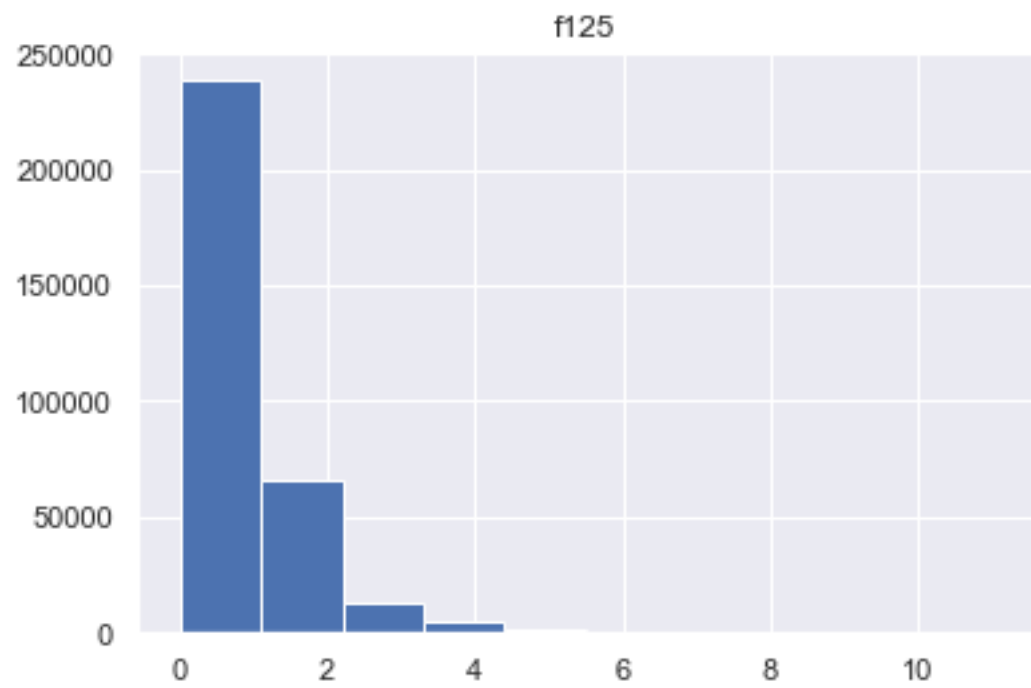
<Figure size 432x288 with 0 Axes>



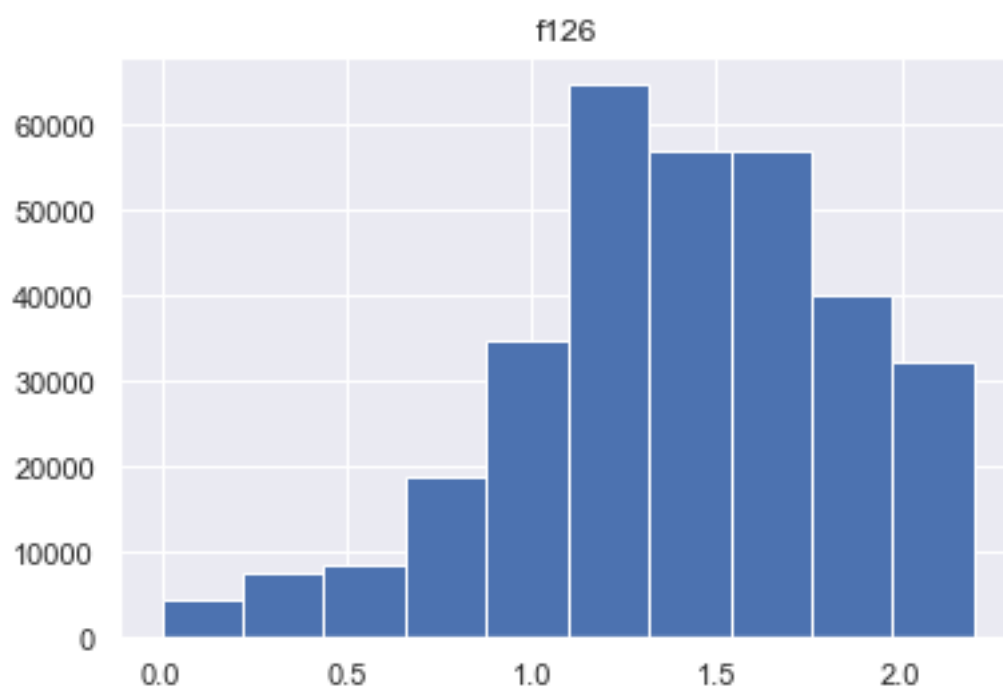
<Figure size 432x288 with 0 Axes>



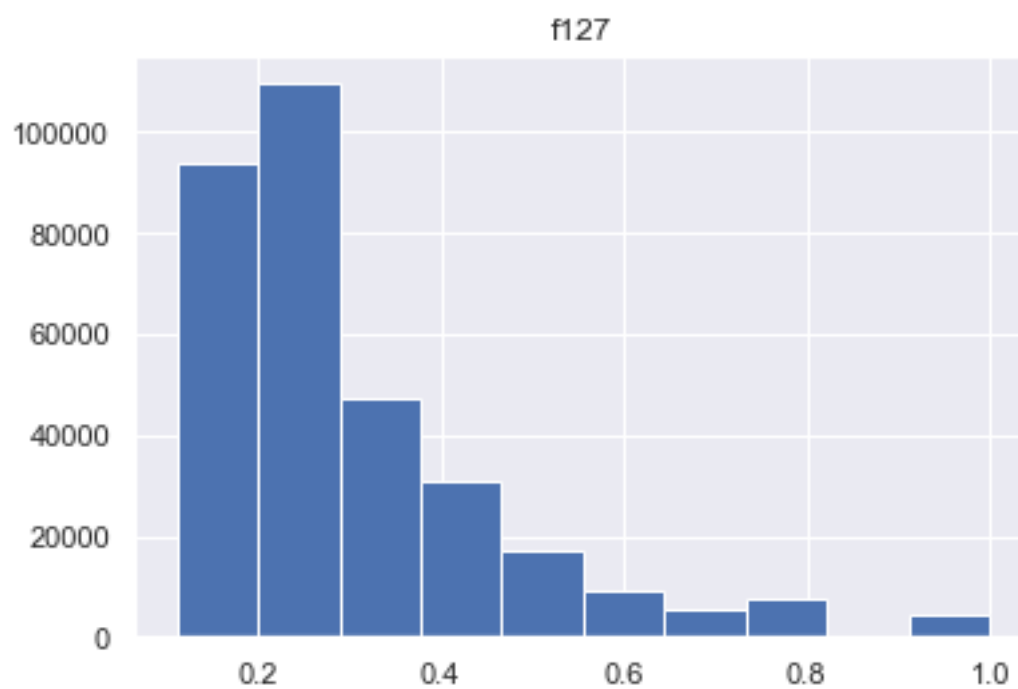
<Figure size 432x288 with 0 Axes>



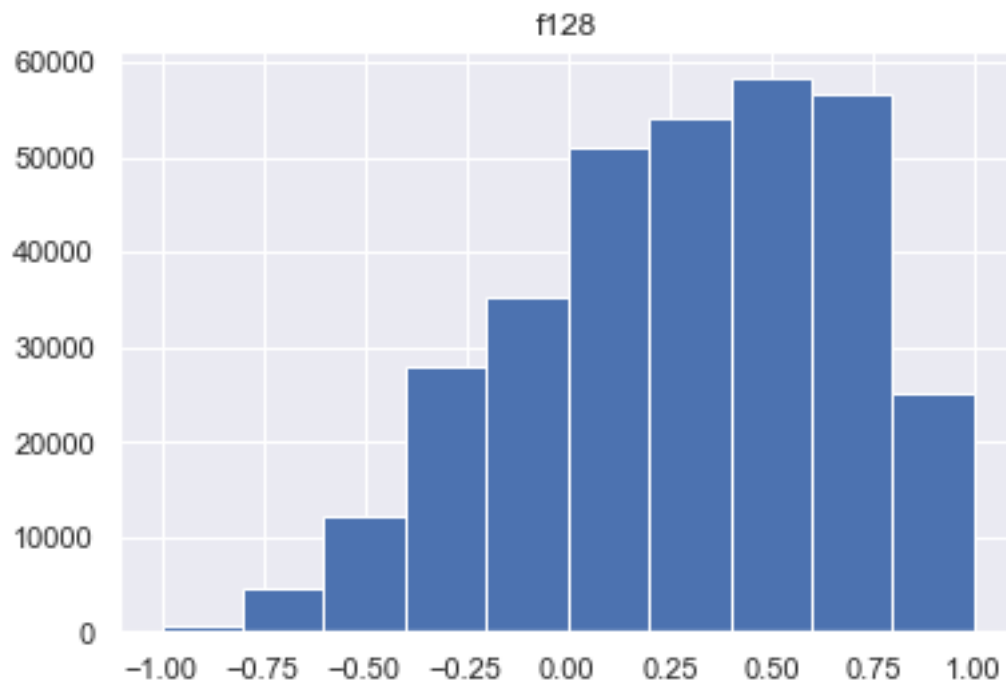
<Figure size 432x288 with 0 Axes>



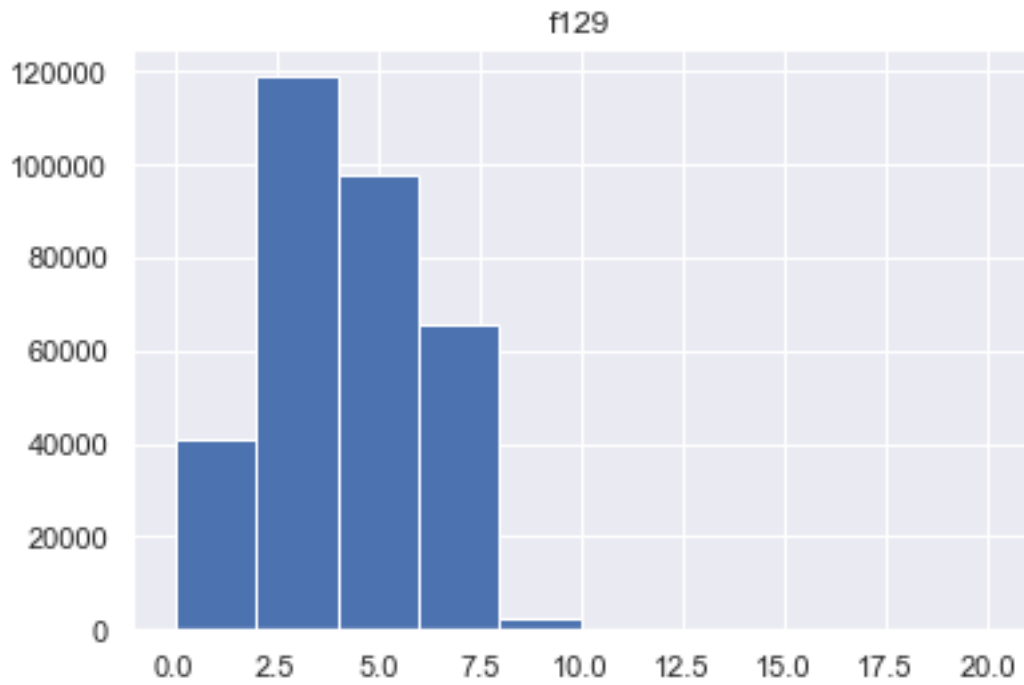
<Figure size 432x288 with 0 Axes>



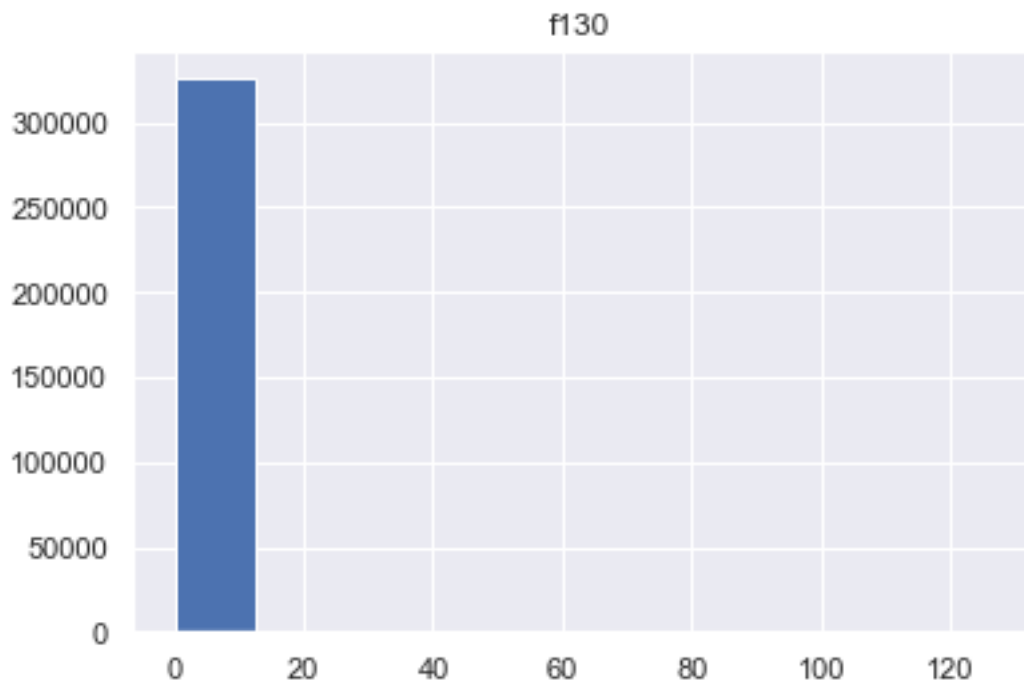
<Figure size 432x288 with 0 Axes>



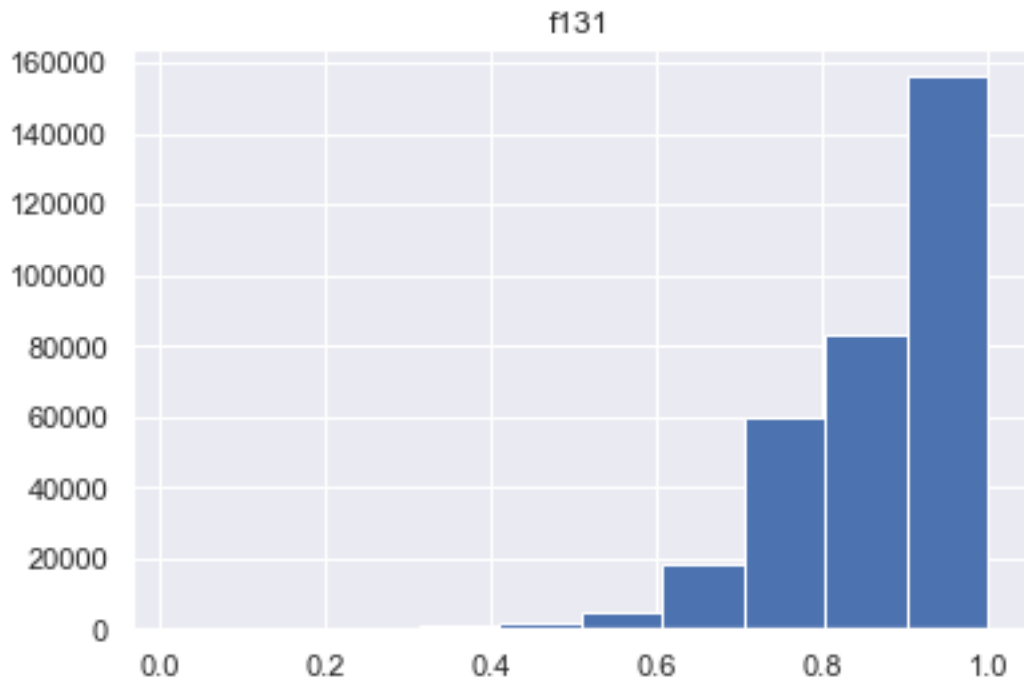
<Figure size 432x288 with 0 Axes>



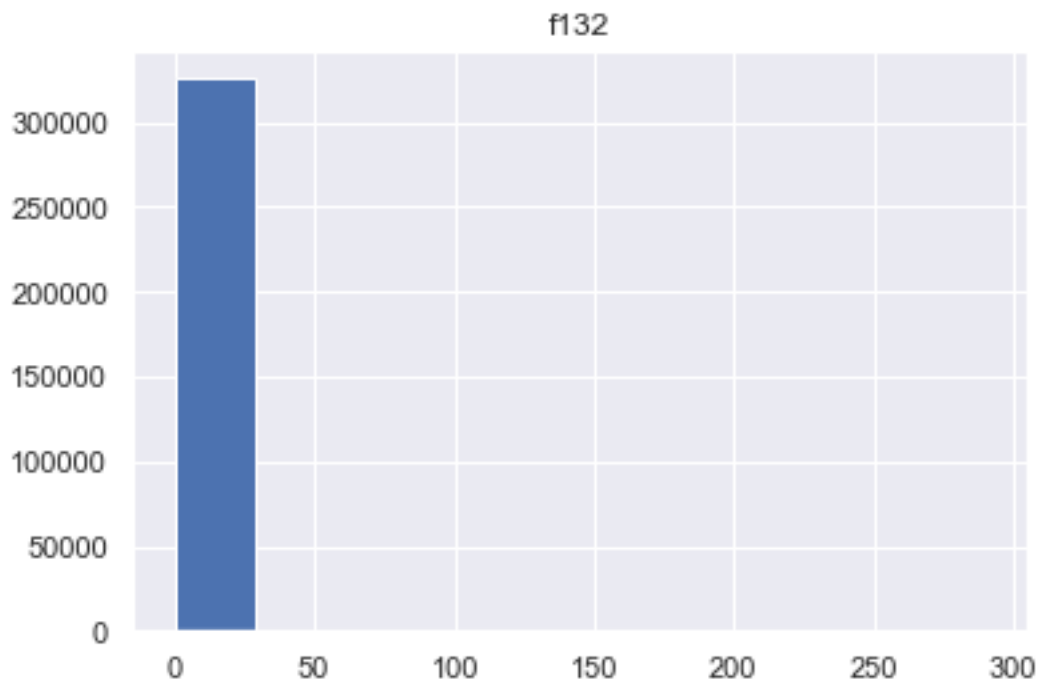
<Figure size 432x288 with 0 Axes>



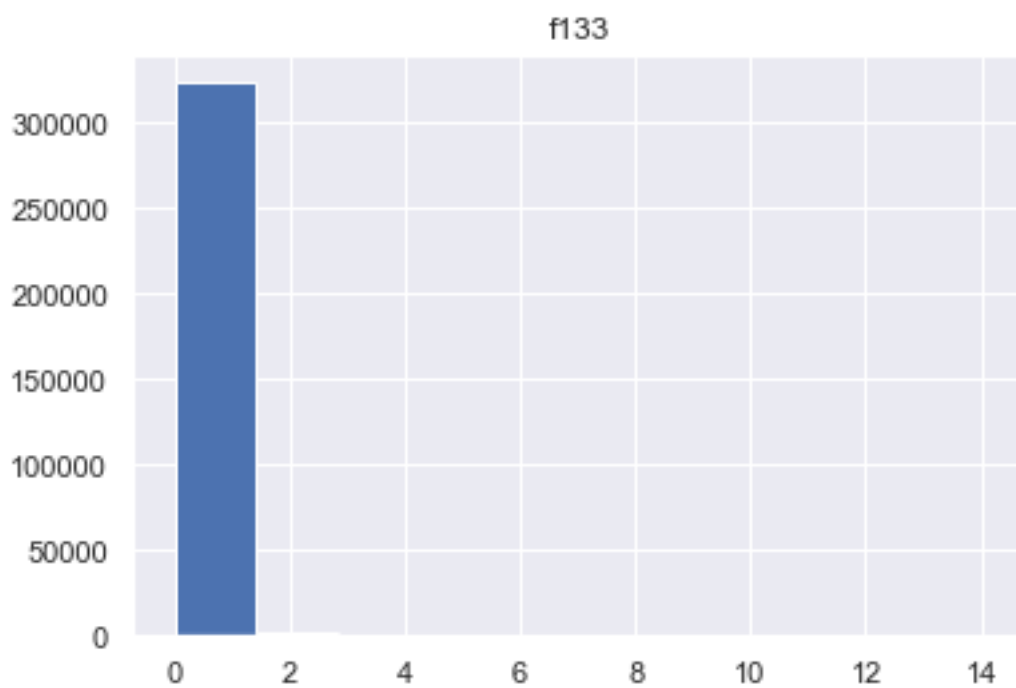
<Figure size 432x288 with 0 Axes>



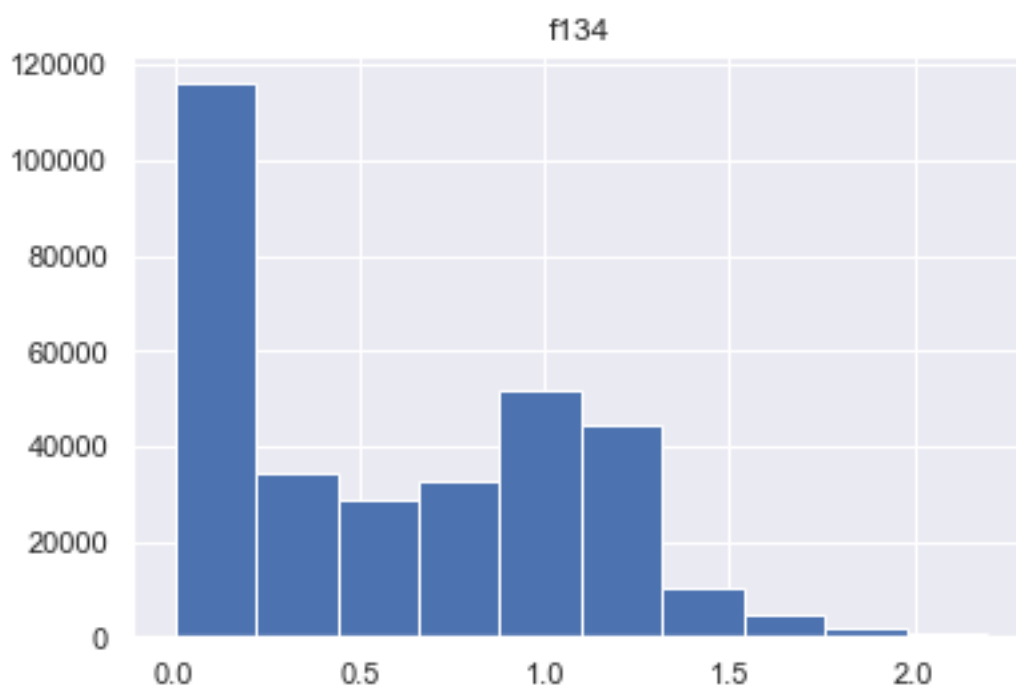
<Figure size 432x288 with 0 Axes>



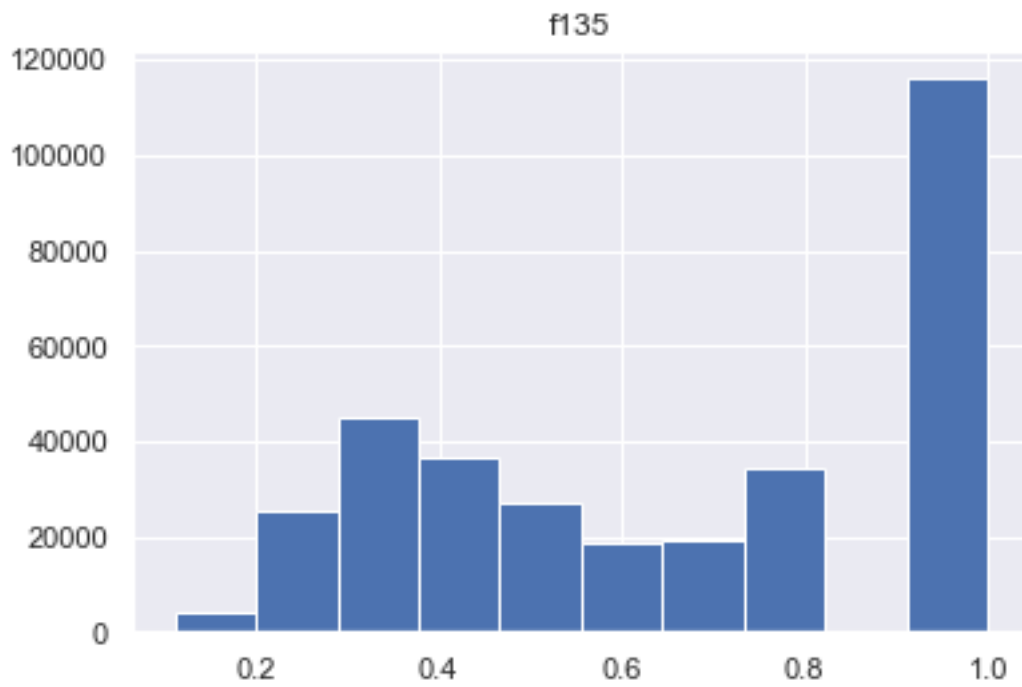
<Figure size 432x288 with 0 Axes>



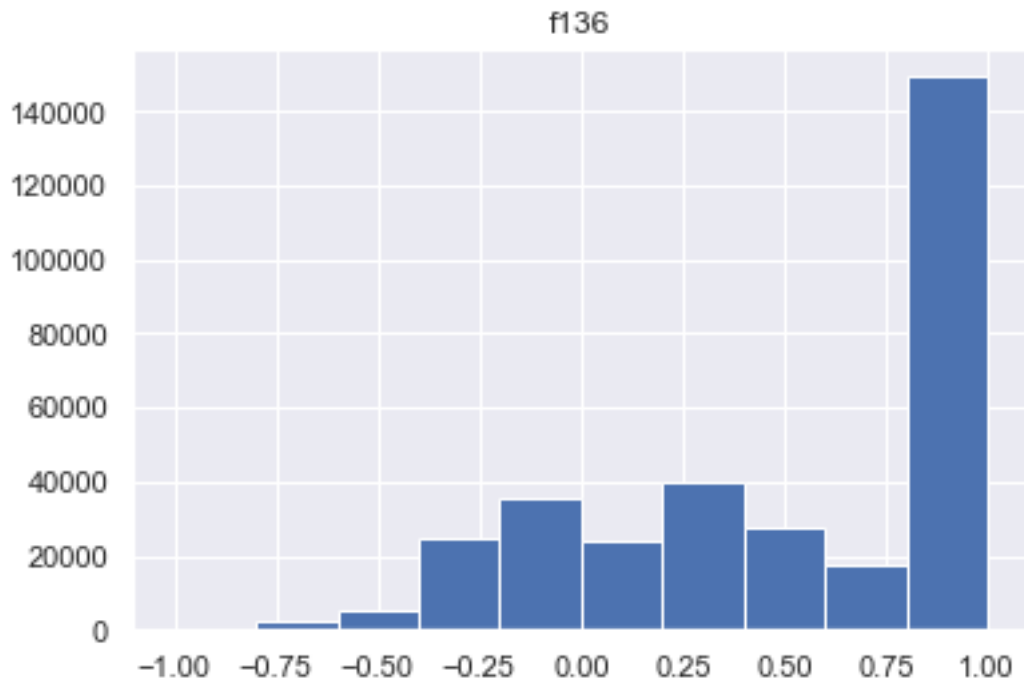
<Figure size 432x288 with 0 Axes>



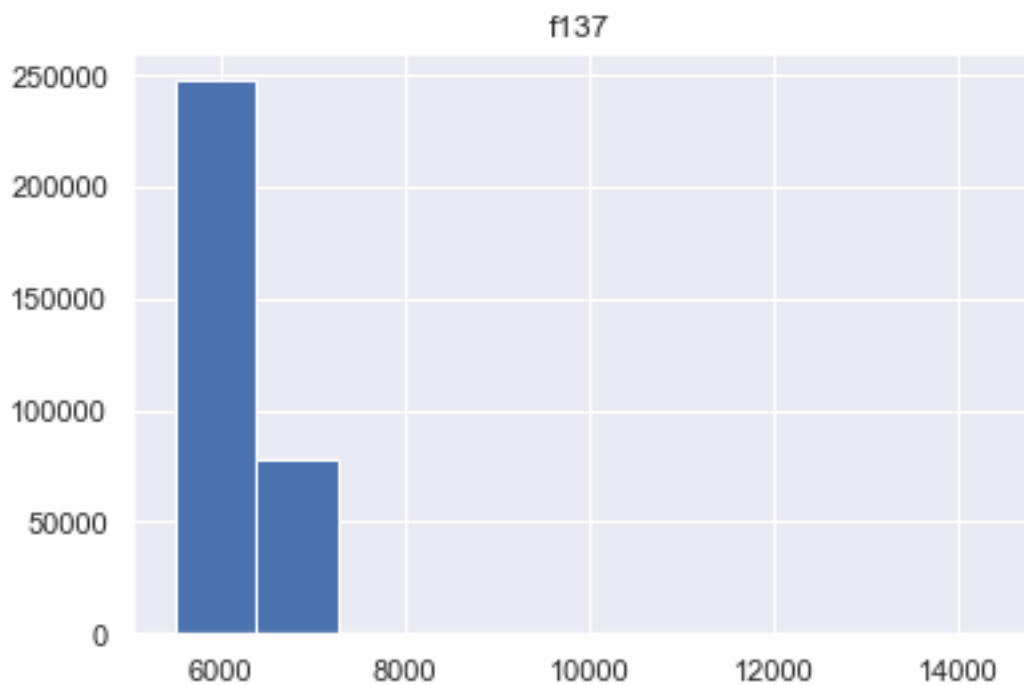
<Figure size 432x288 with 0 Axes>



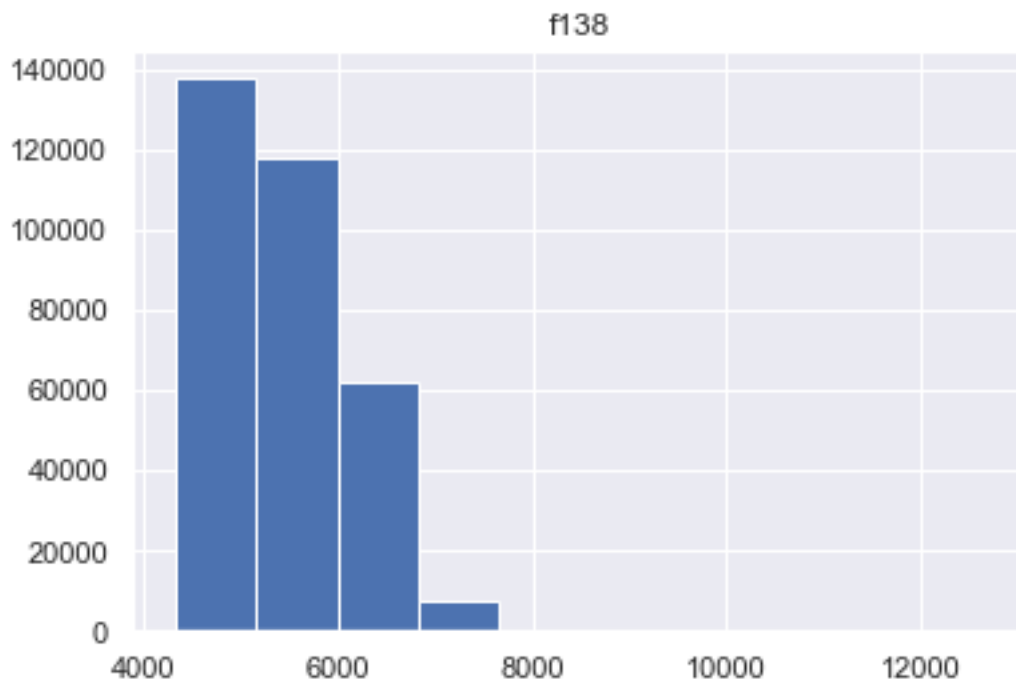
<Figure size 432x288 with 0 Axes>



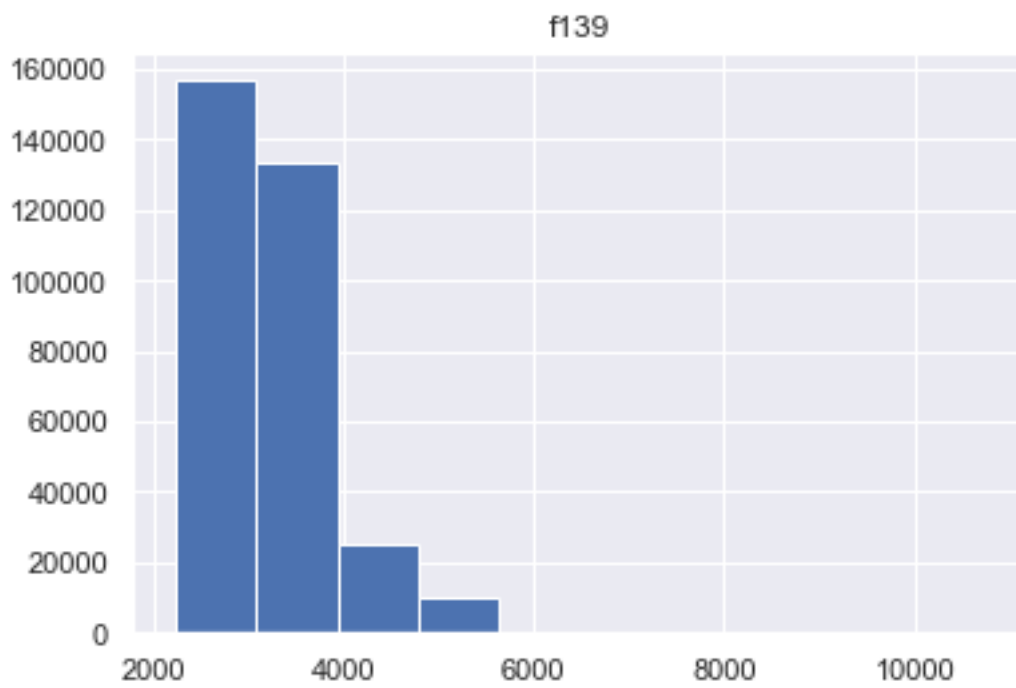
<Figure size 432x288 with 0 Axes>



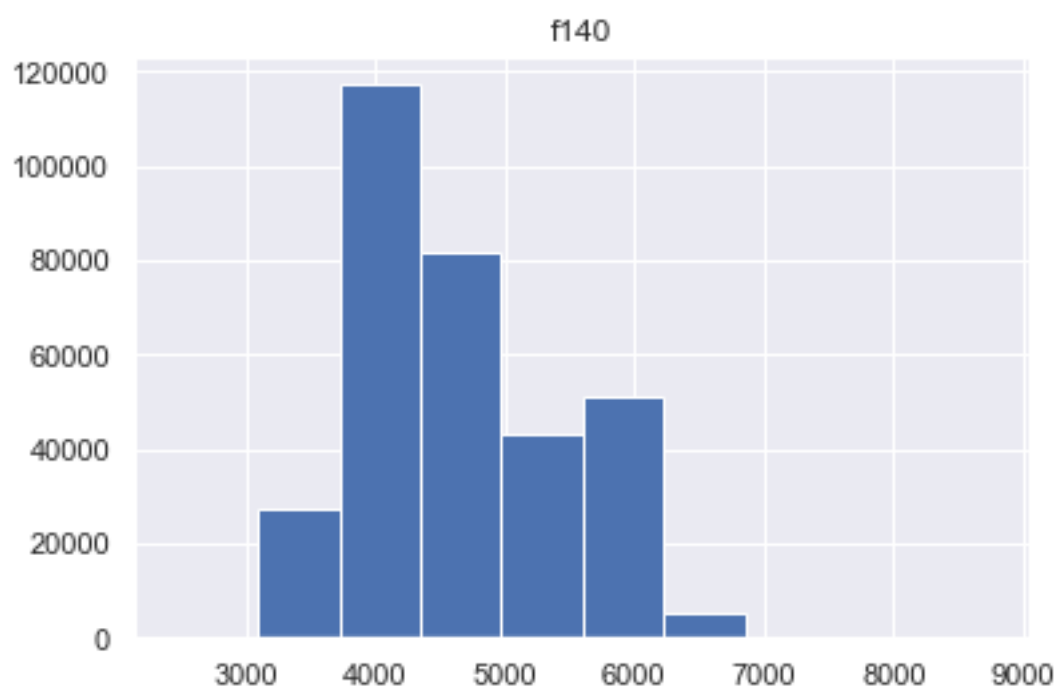
<Figure size 432x288 with 0 Axes>



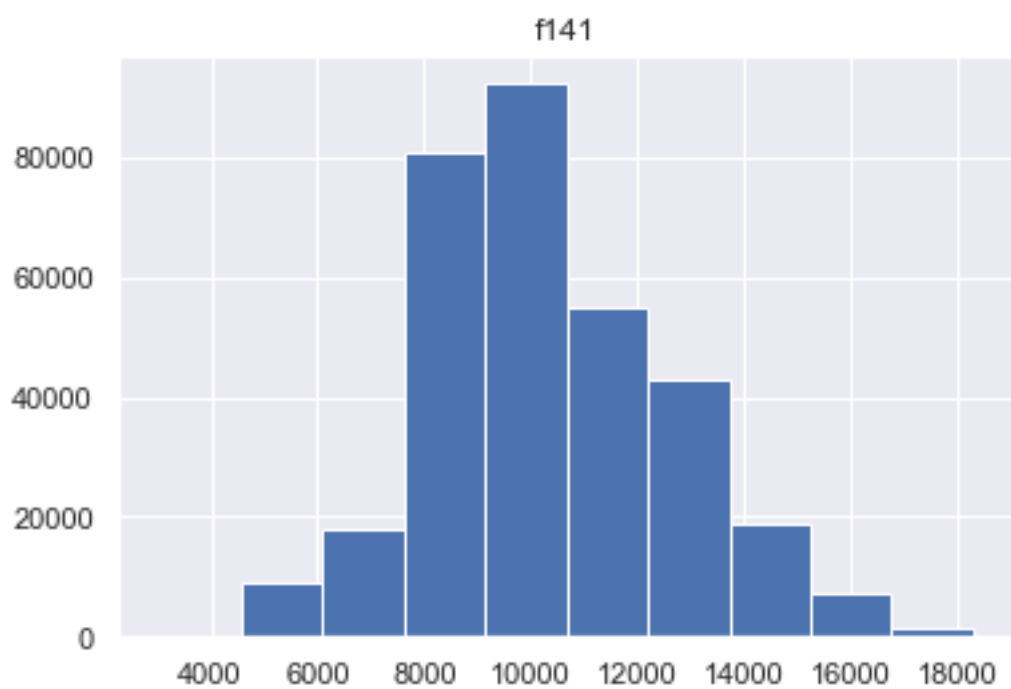
<Figure size 432x288 with 0 Axes>



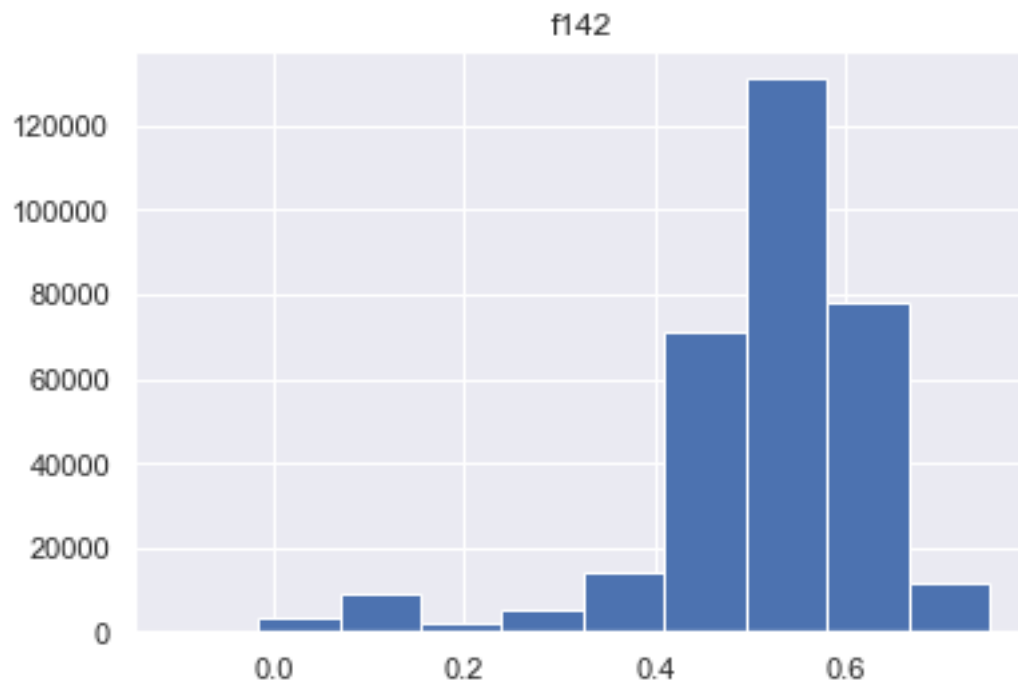
<Figure size 432x288 with 0 Axes>



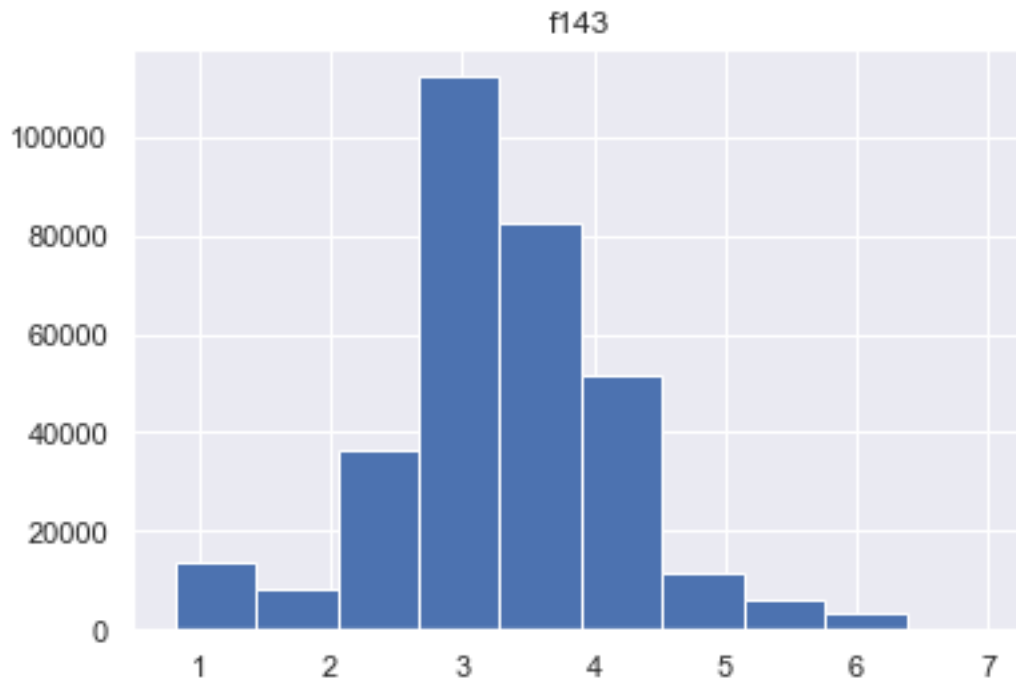
<Figure size 432x288 with 0 Axes>



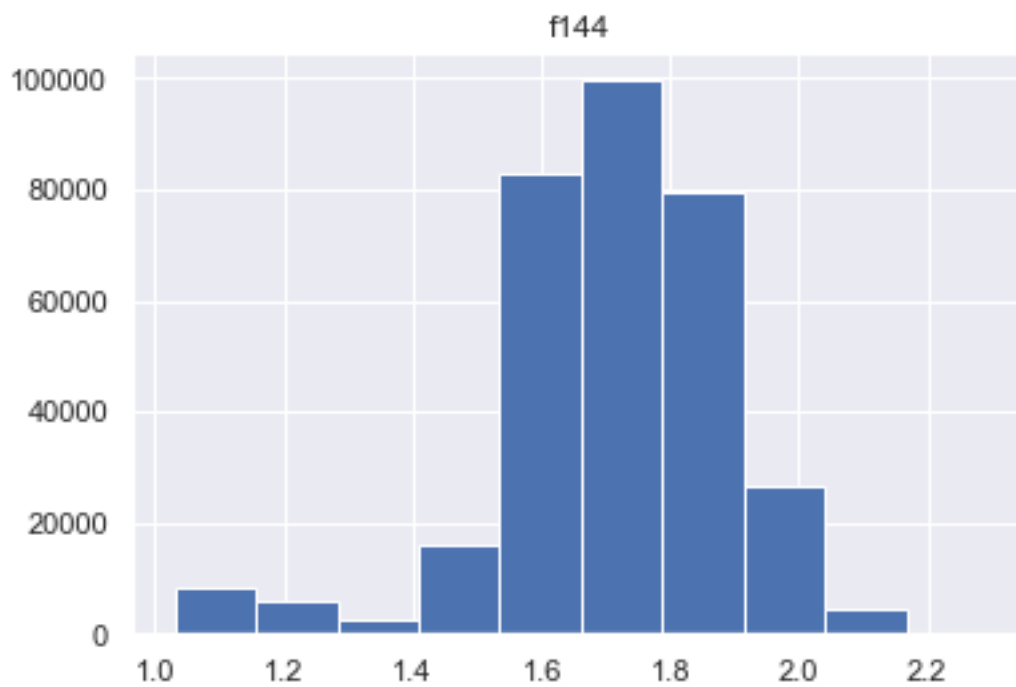
<Figure size 432x288 with 0 Axes>



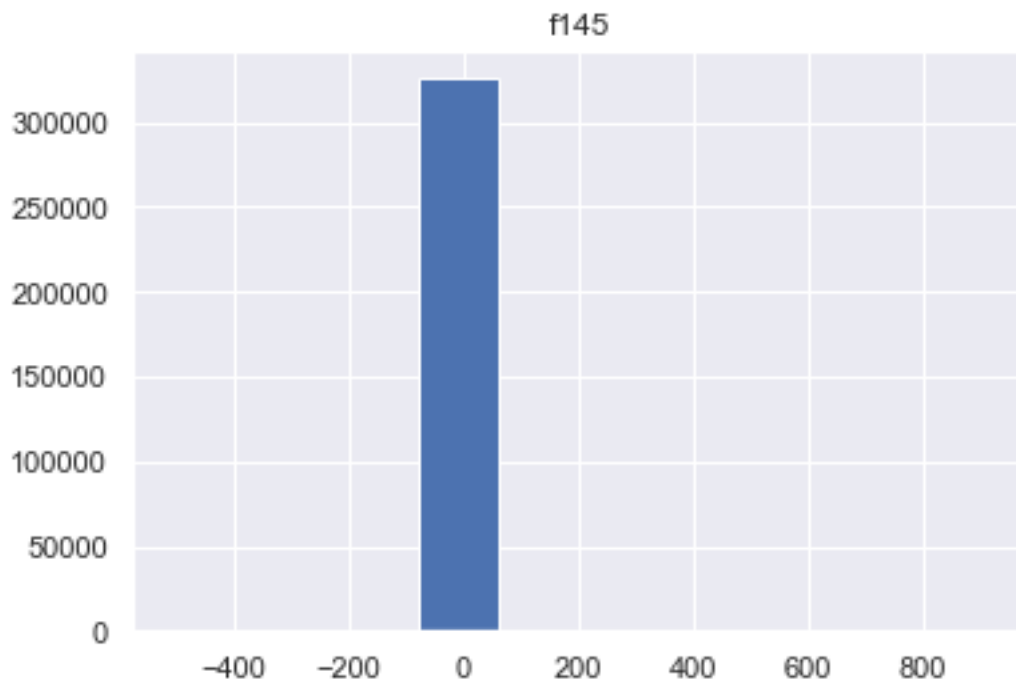
<Figure size 432x288 with 0 Axes>



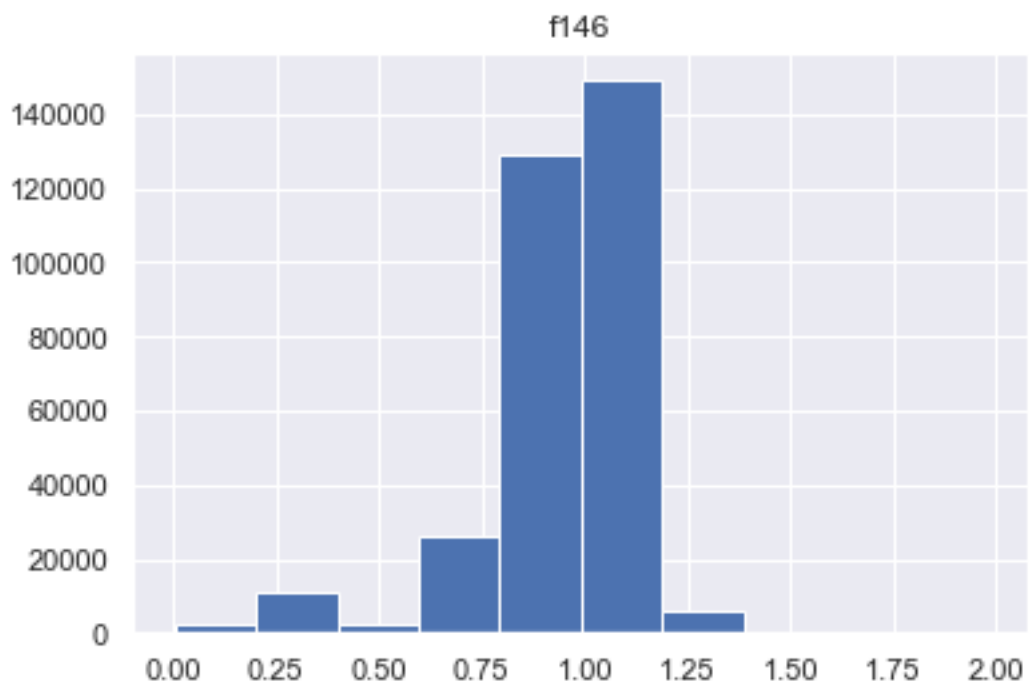
<Figure size 432x288 with 0 Axes>



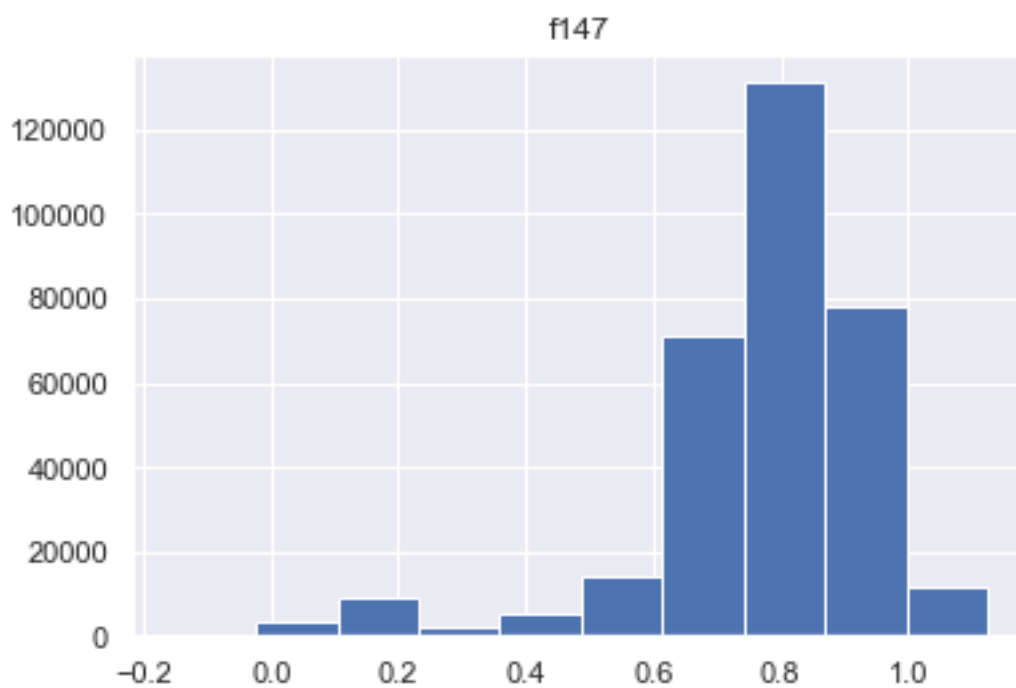
<Figure size 432x288 with 0 Axes>



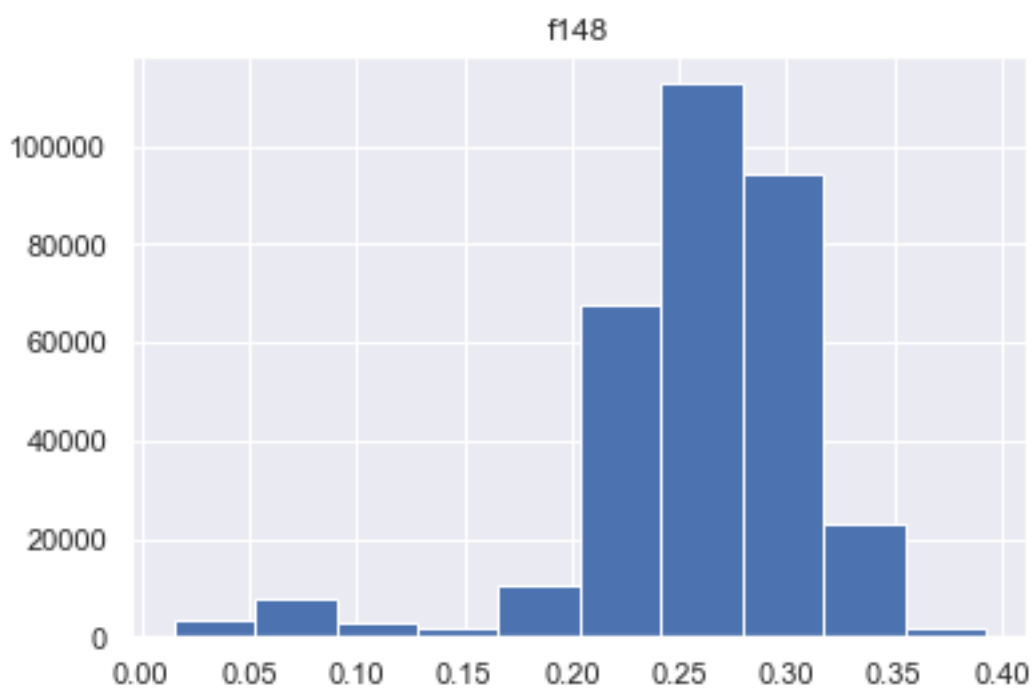
<Figure size 432x288 with 0 Axes>



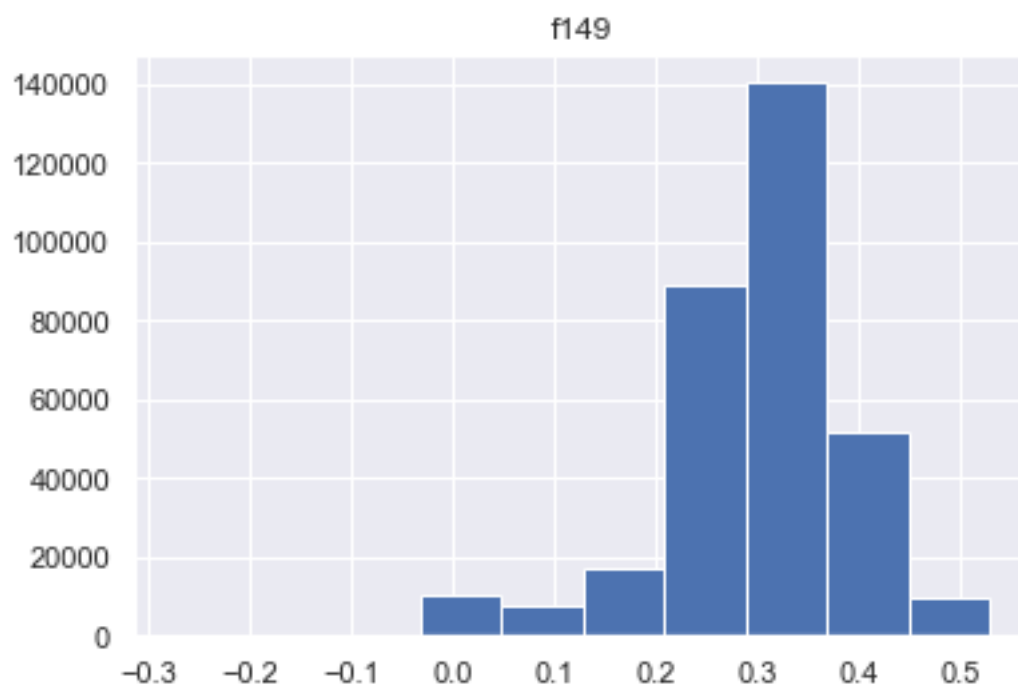
<Figure size 432x288 with 0 Axes>



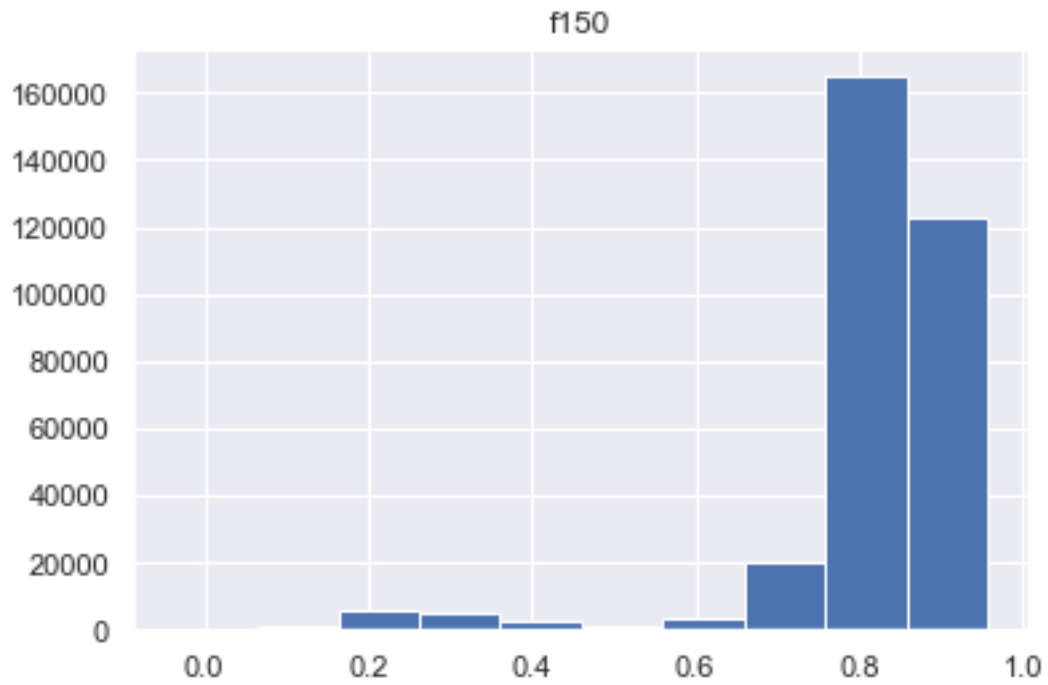
<Figure size 432x288 with 0 Axes>



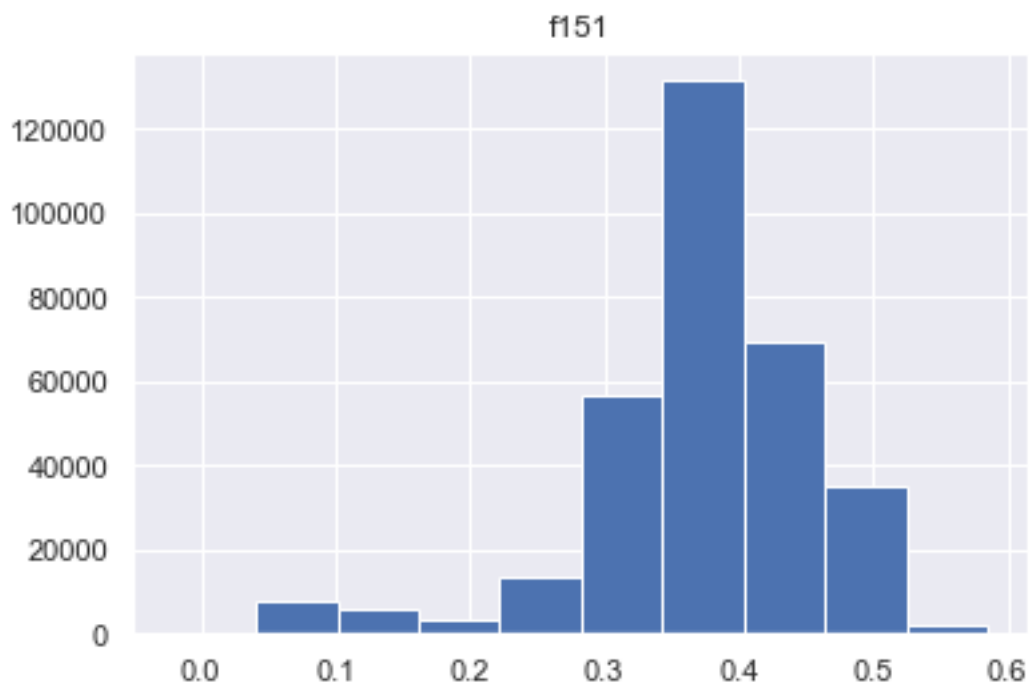
<Figure size 432x288 with 0 Axes>



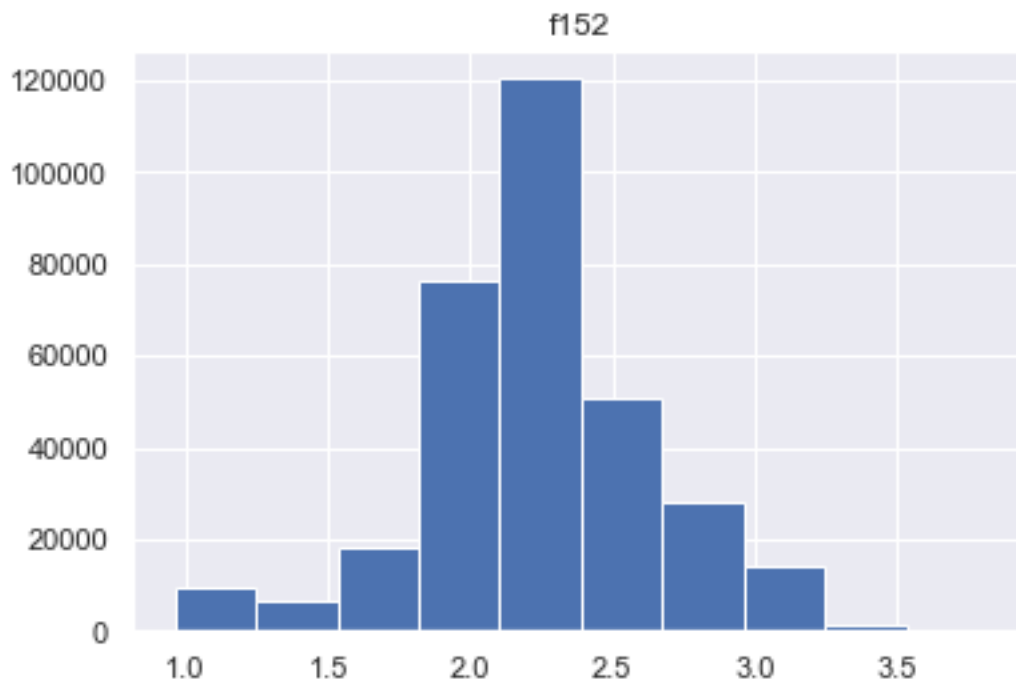
<Figure size 432x288 with 0 Axes>



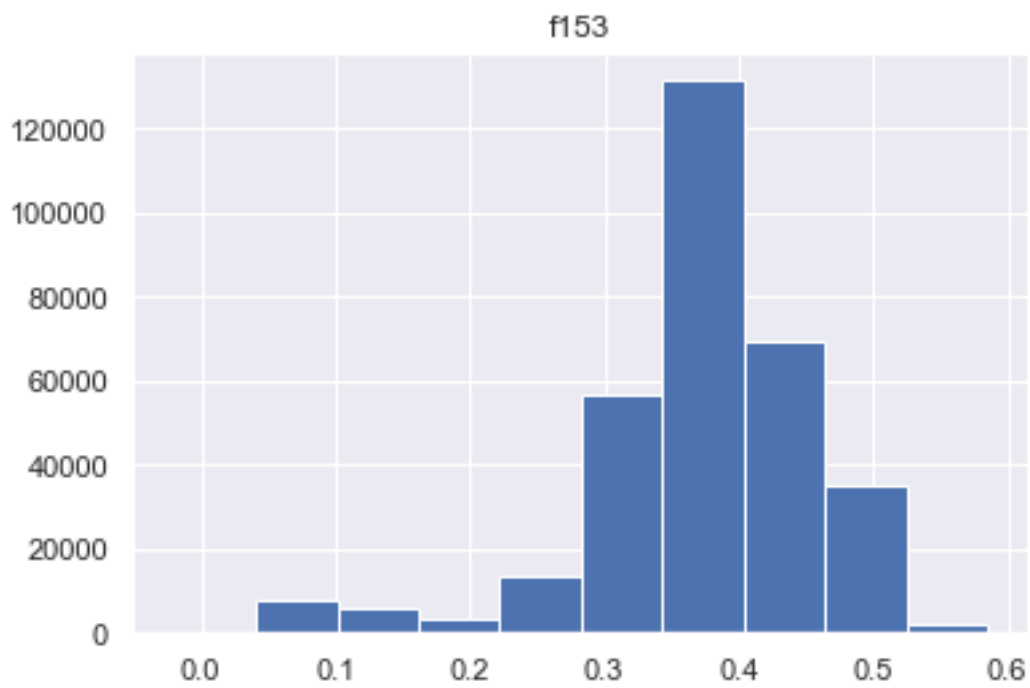
<Figure size 432x288 with 0 Axes>



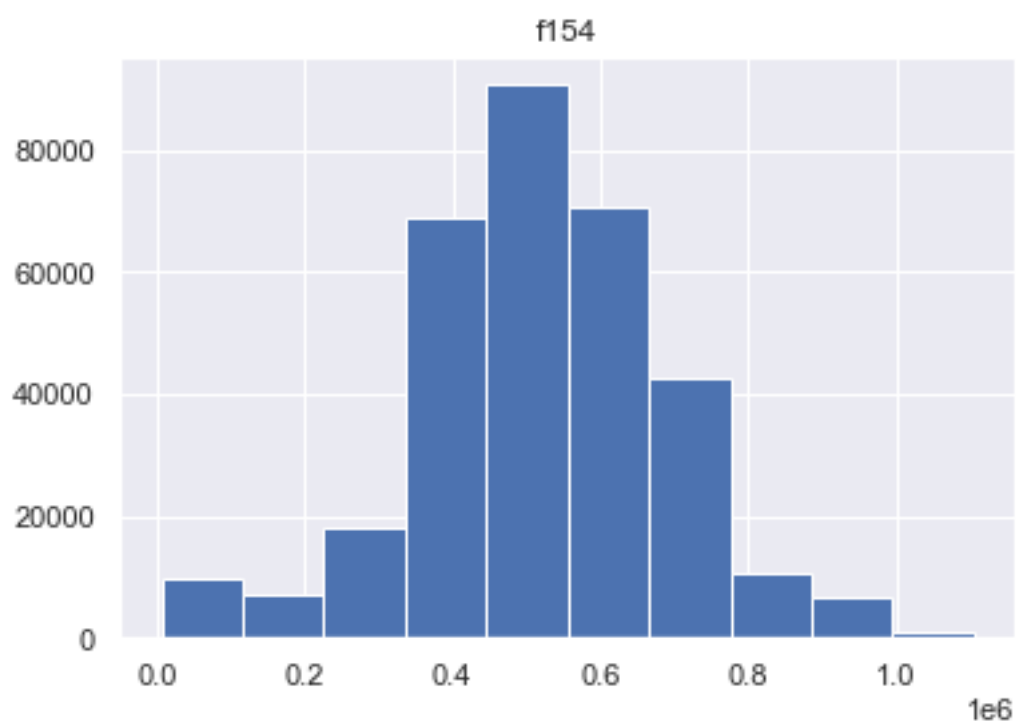
<Figure size 432x288 with 0 Axes>



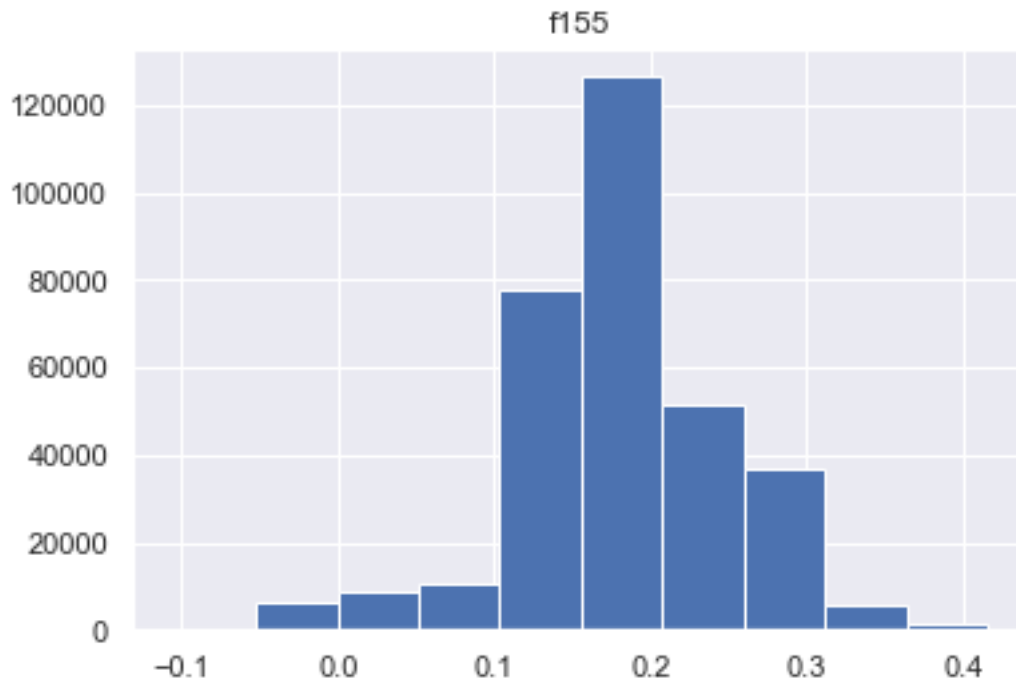
<Figure size 432x288 with 0 Axes>



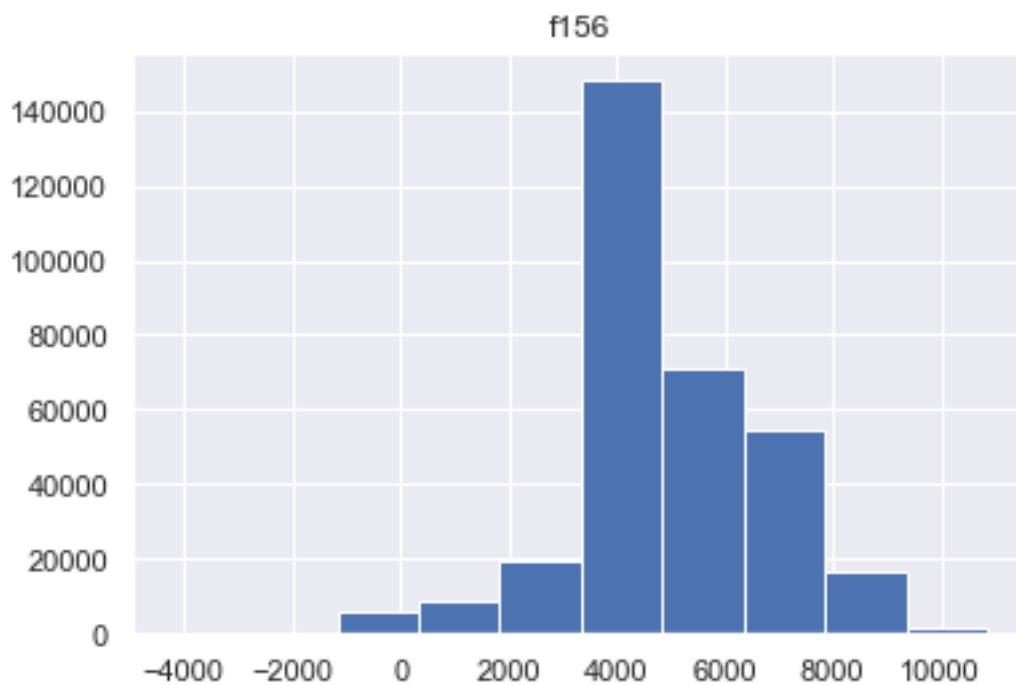
<Figure size 432x288 with 0 Axes>



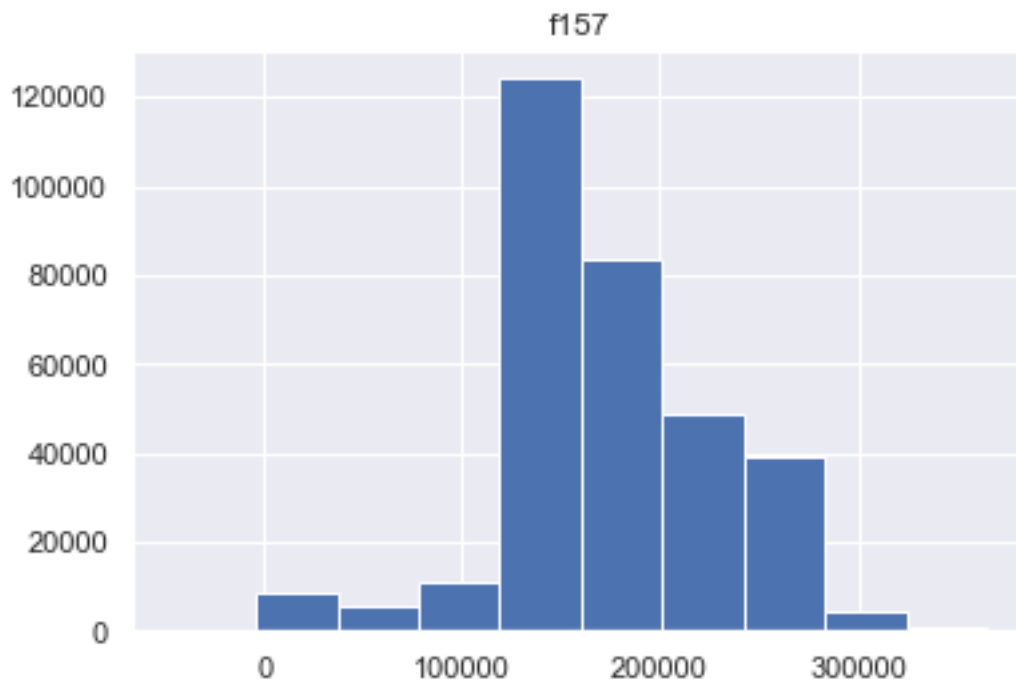
<Figure size 432x288 with 0 Axes>



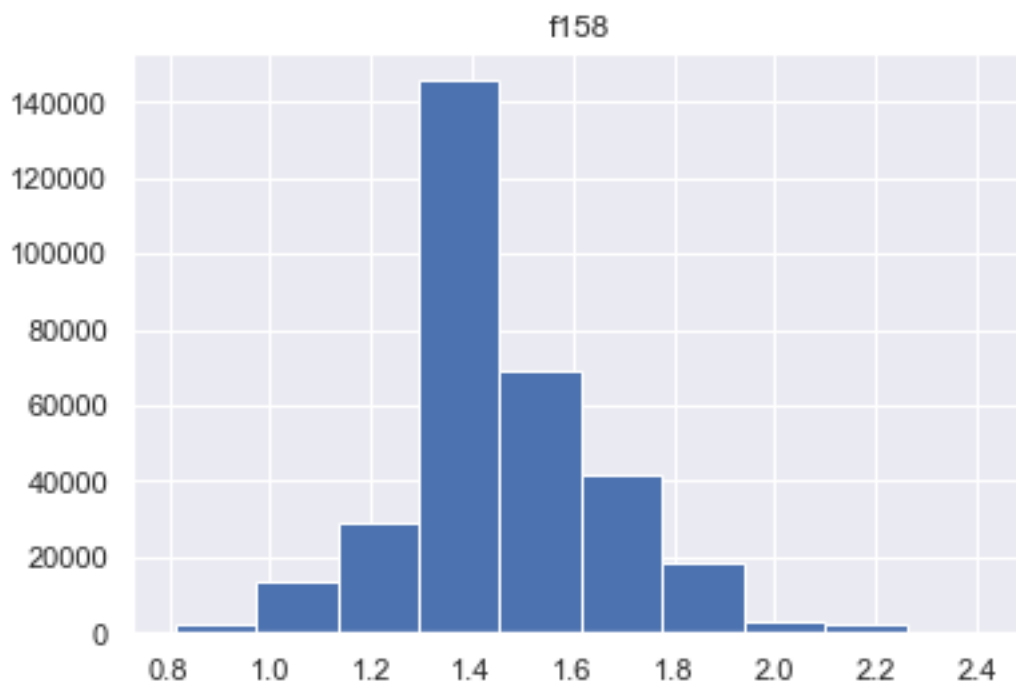
<Figure size 432x288 with 0 Axes>



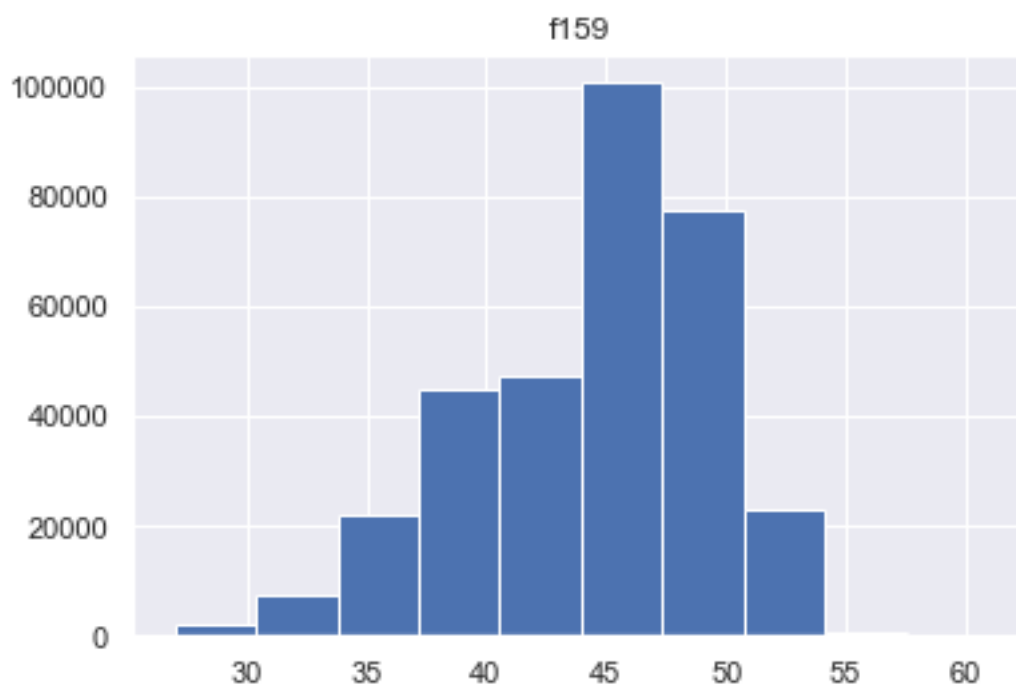
<Figure size 432x288 with 0 Axes>



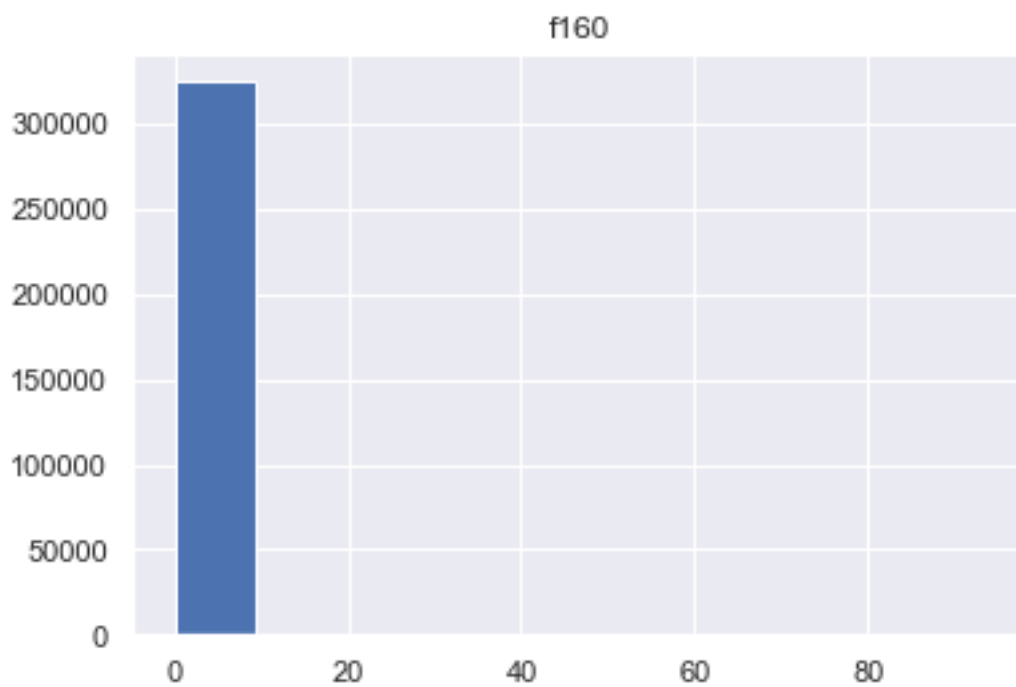
<Figure size 432x288 with 0 Axes>



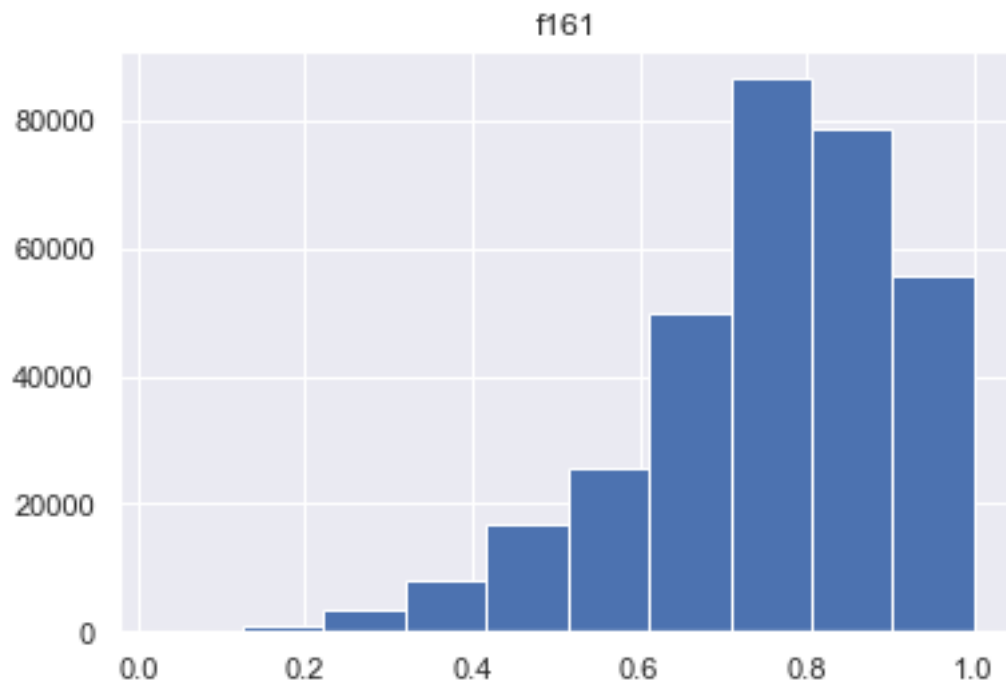
<Figure size 432x288 with 0 Axes>



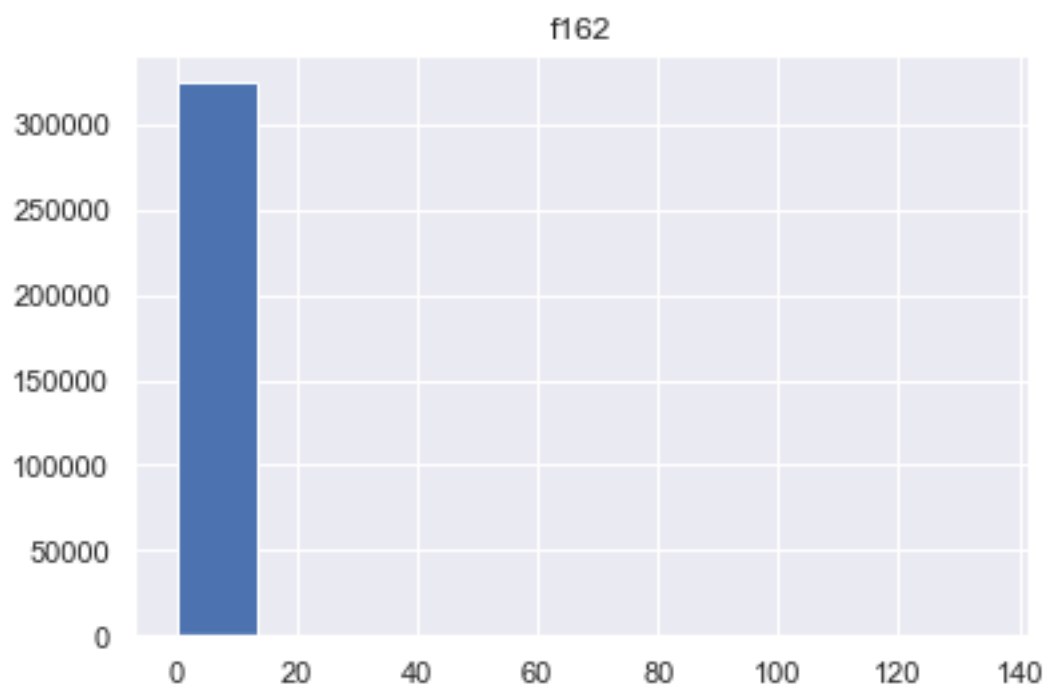
<Figure size 432x288 with 0 Axes>



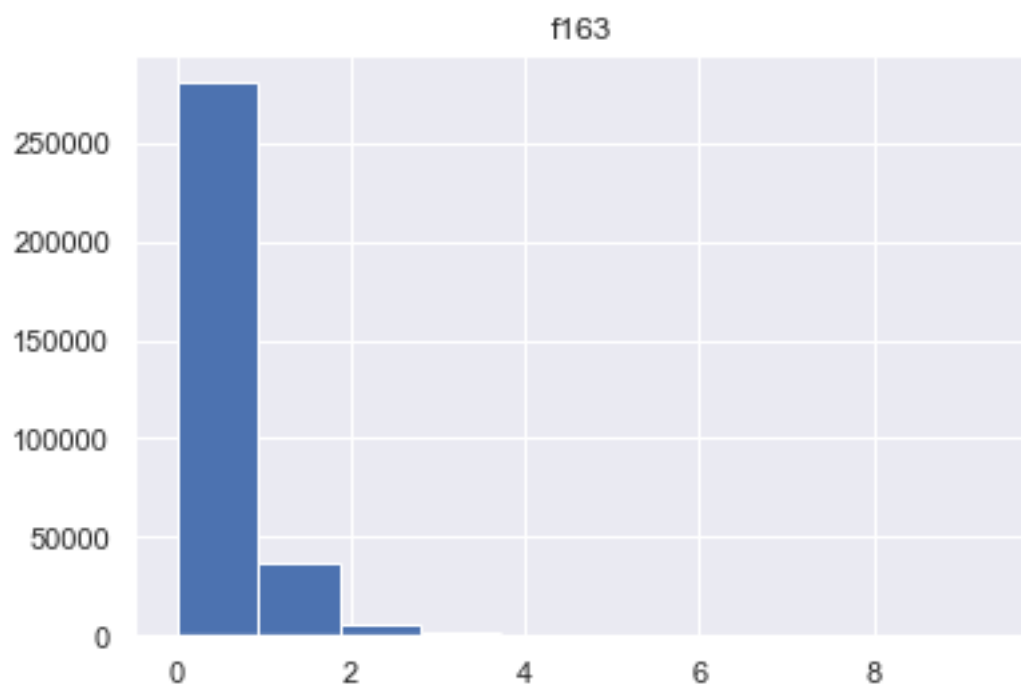
<Figure size 432x288 with 0 Axes>



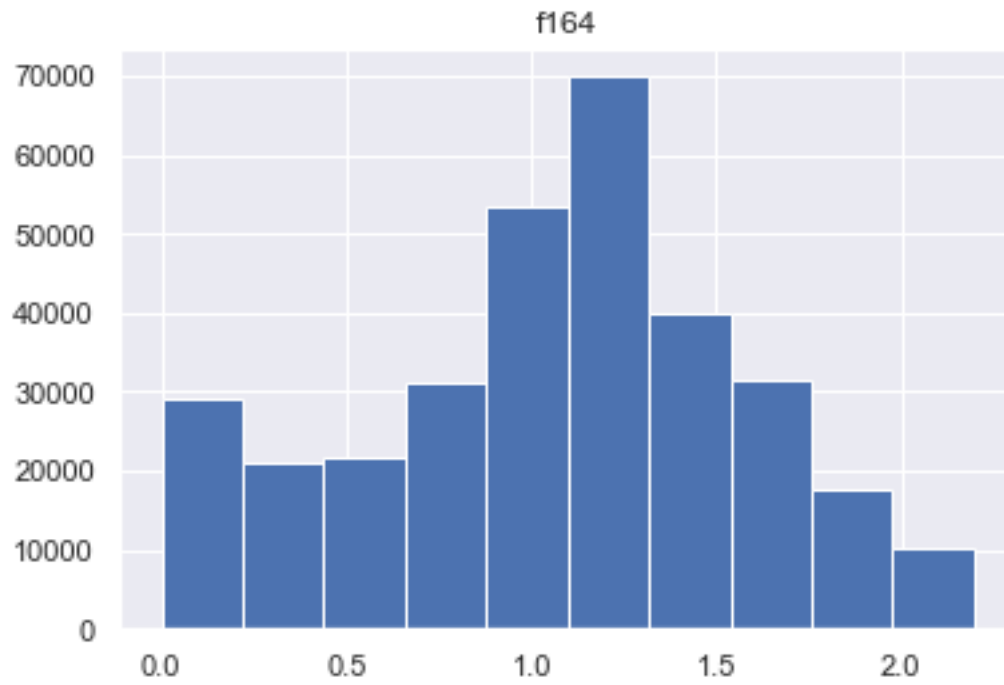
<Figure size 432x288 with 0 Axes>



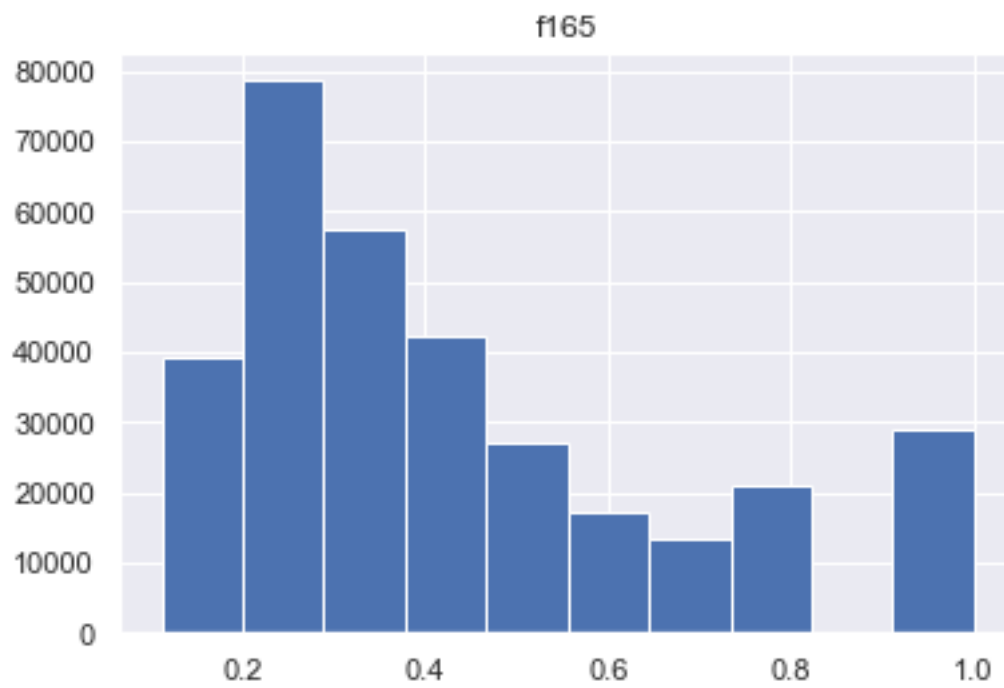
<Figure size 432x288 with 0 Axes>



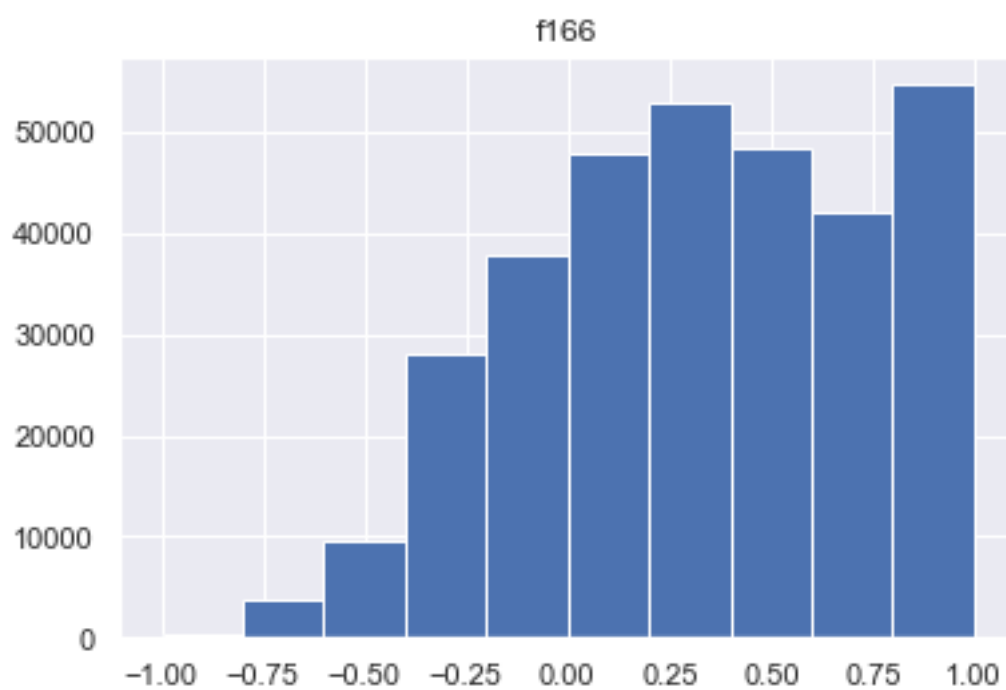
<Figure size 432x288 with 0 Axes>



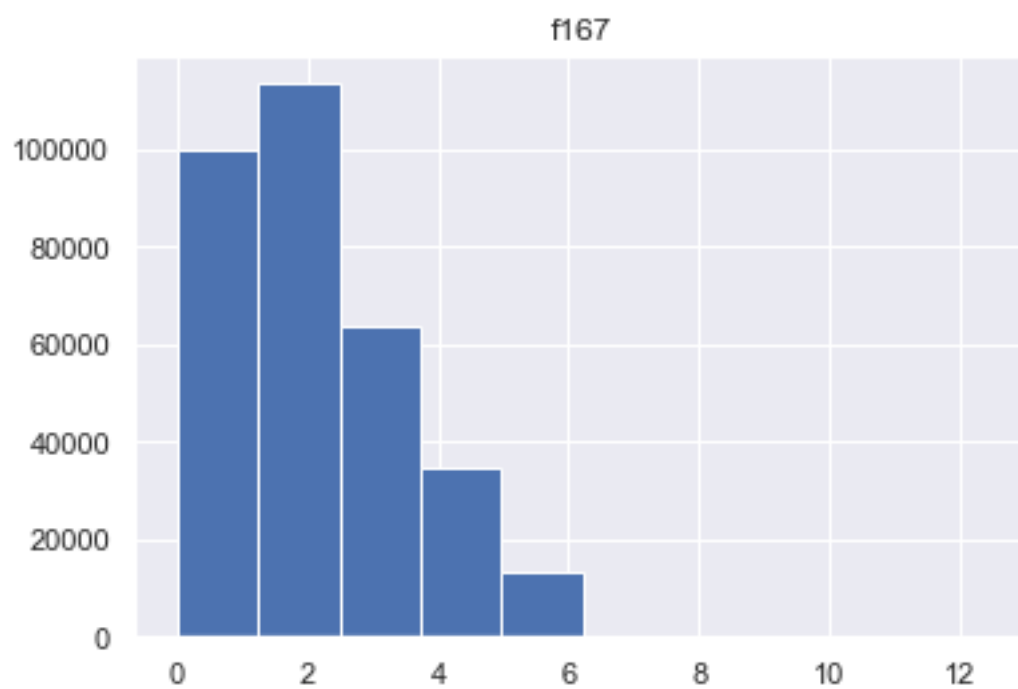
<Figure size 432x288 with 0 Axes>



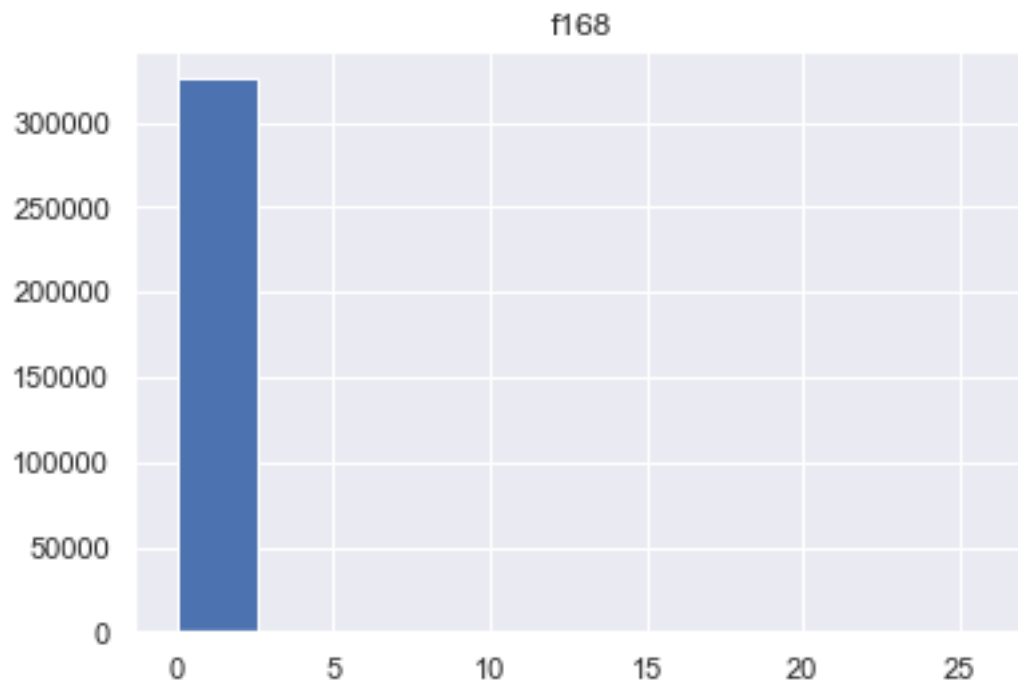
<Figure size 432x288 with 0 Axes>



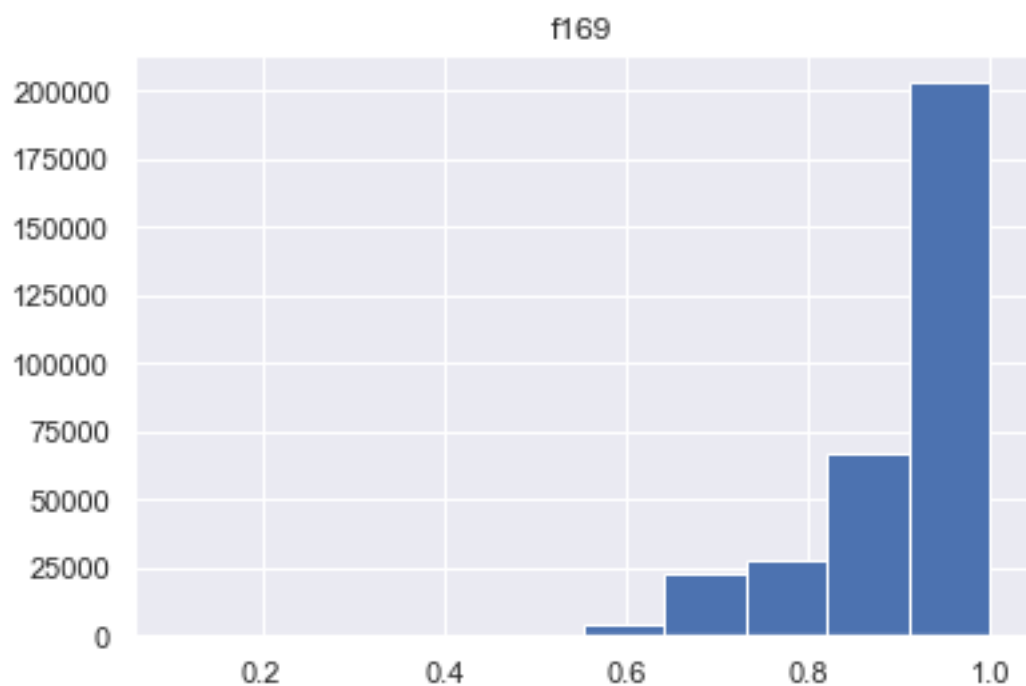
<Figure size 432x288 with 0 Axes>



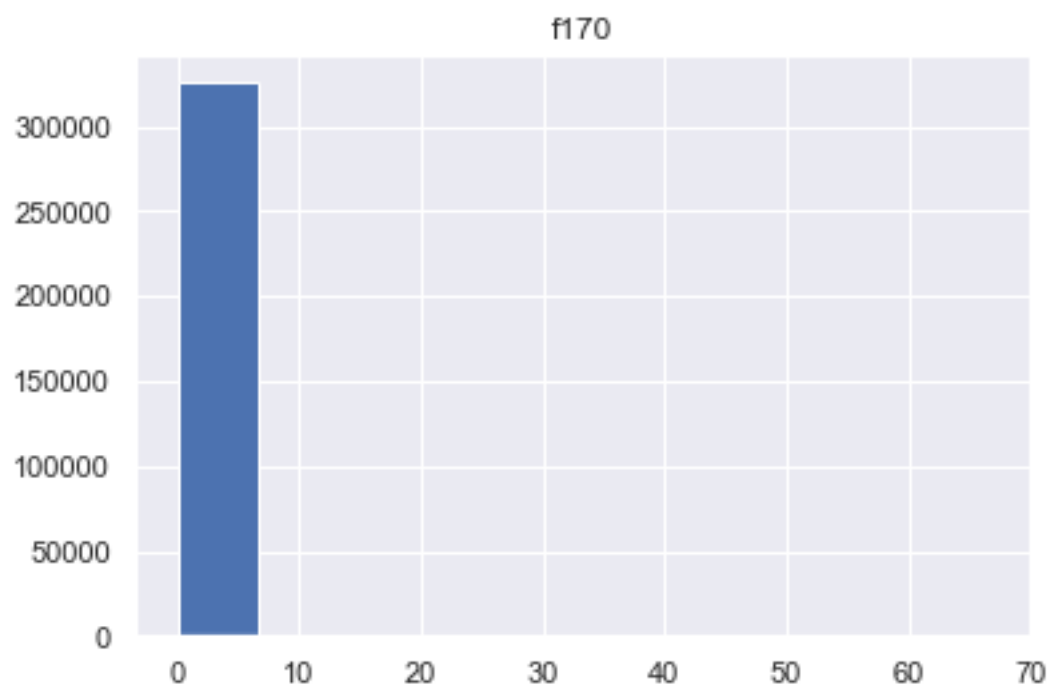
<Figure size 432x288 with 0 Axes>



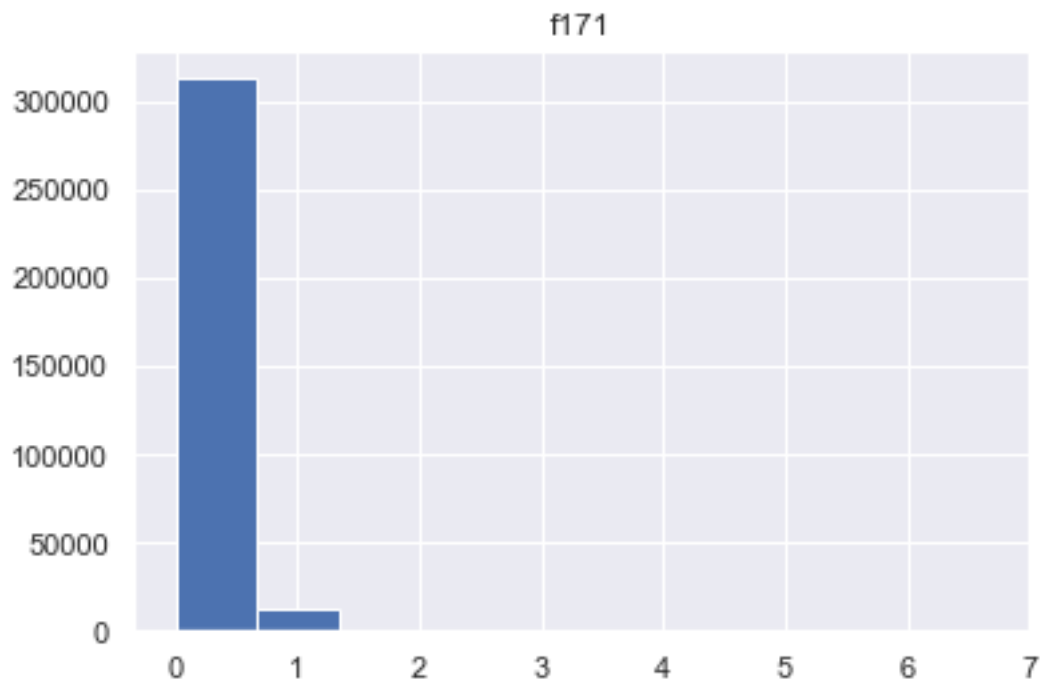
<Figure size 432x288 with 0 Axes>



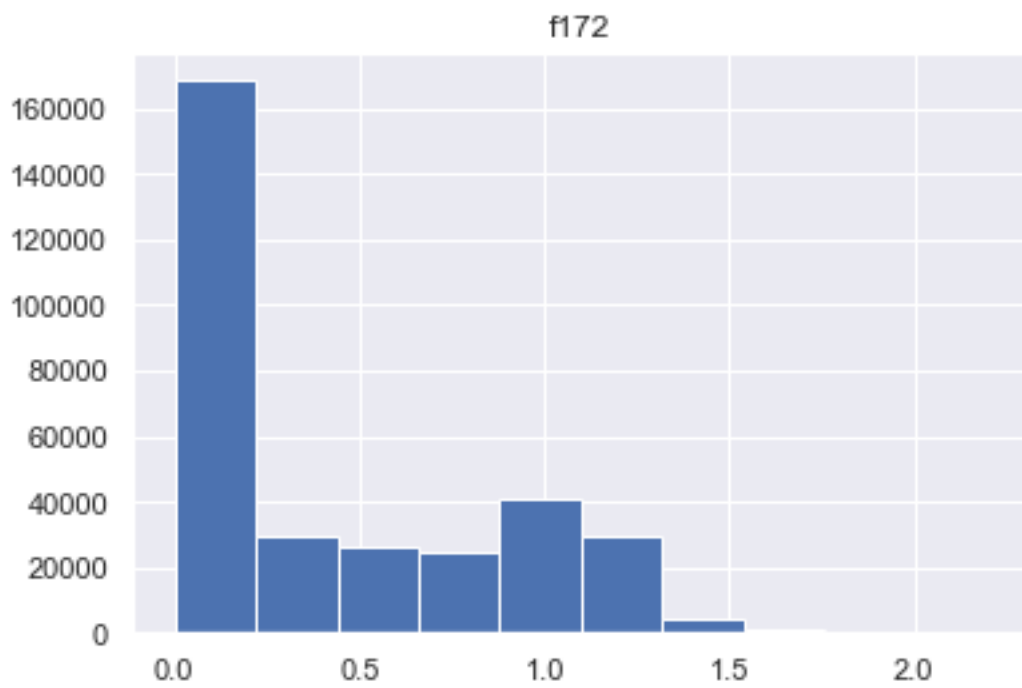
<Figure size 432x288 with 0 Axes>



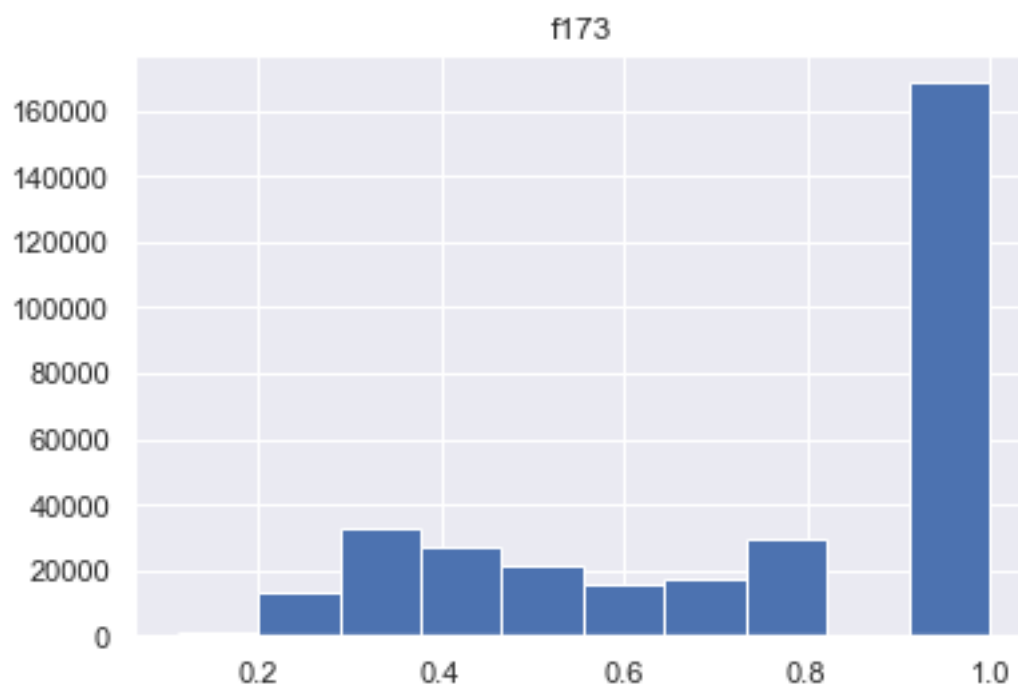
<Figure size 432x288 with 0 Axes>



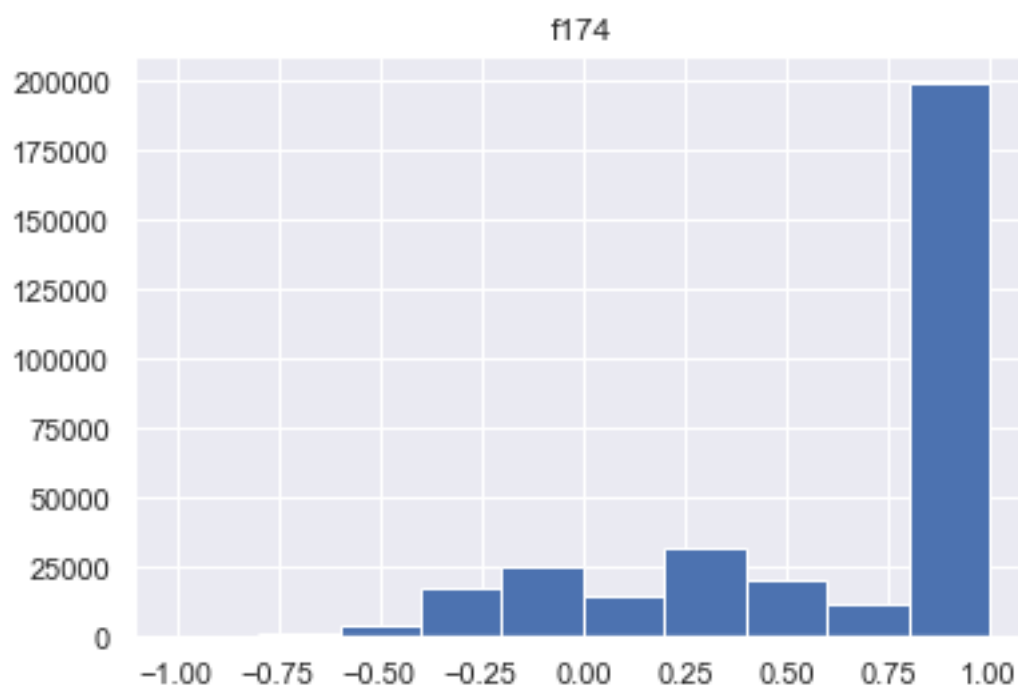
<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>



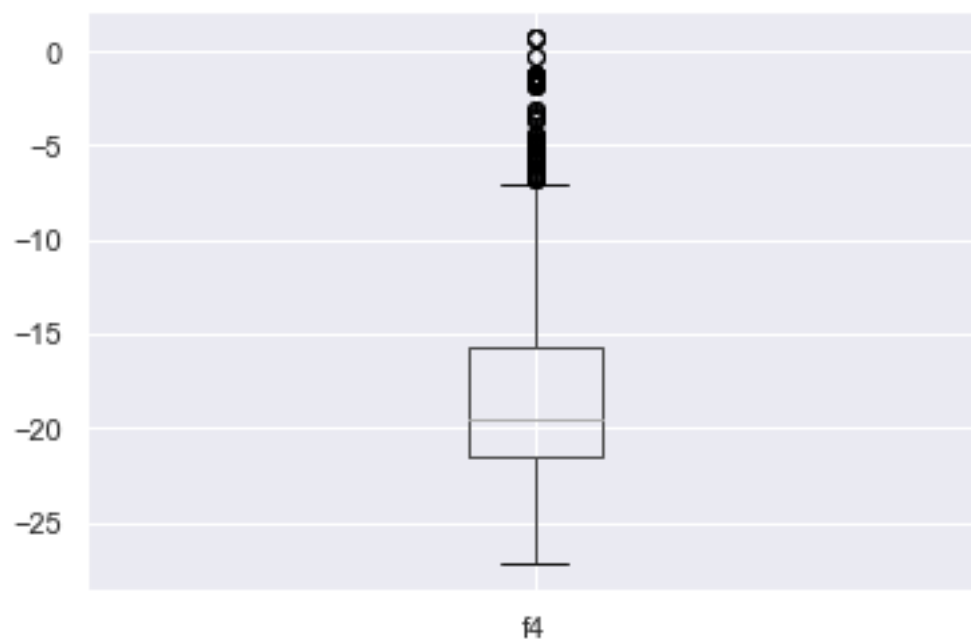
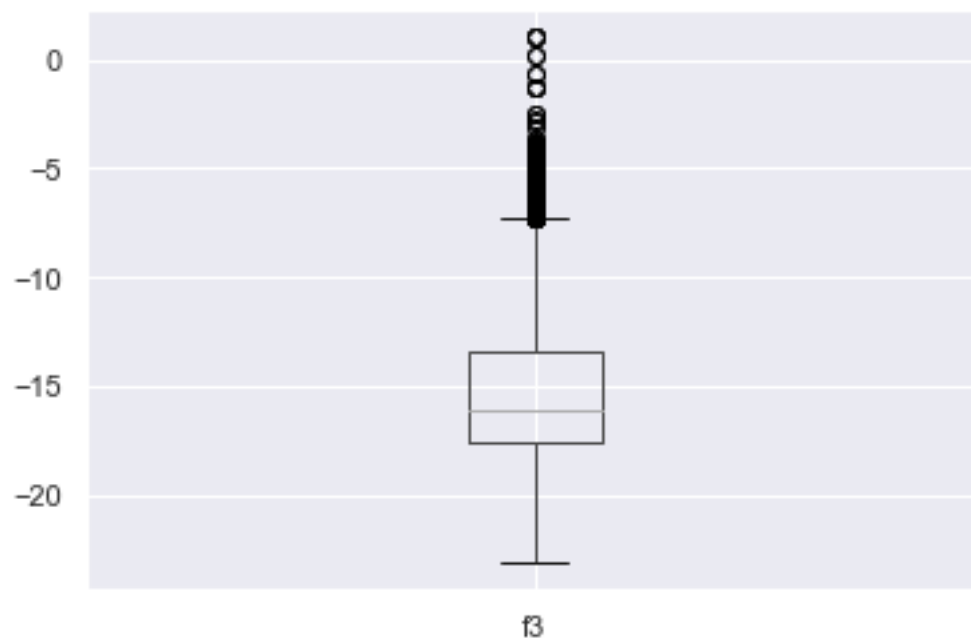
```
[7]: # boxplots of dataset attributes
for column in df:
    plt.figure()
    df.boxplot([column])
```

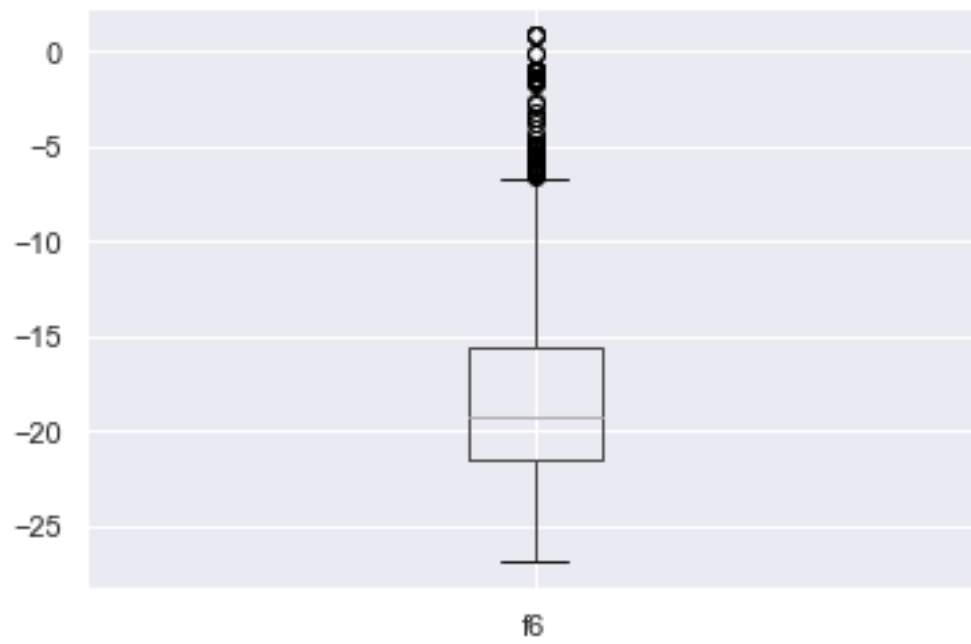
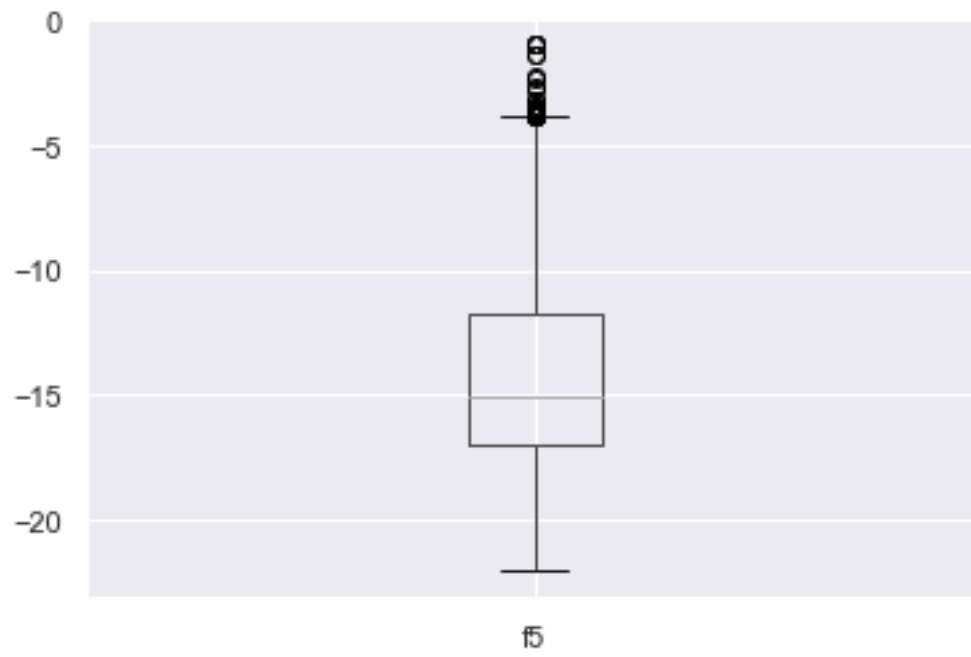
<ipython-input-7-3ce495de62b0>:3: RuntimeWarning: More than 20 figures have been opened. Figures created through the pyplot interface (`matplotlib.pyplot.figure`) are retained until explicitly closed and may consume too much memory. (To control this warning, see the rcParam `figure.max_open_warning`).

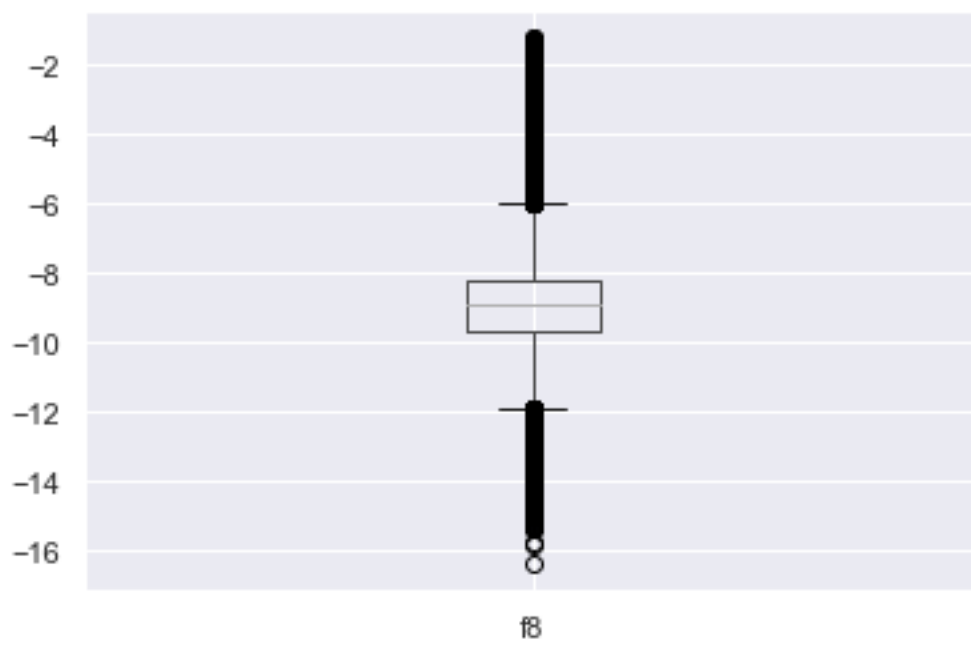
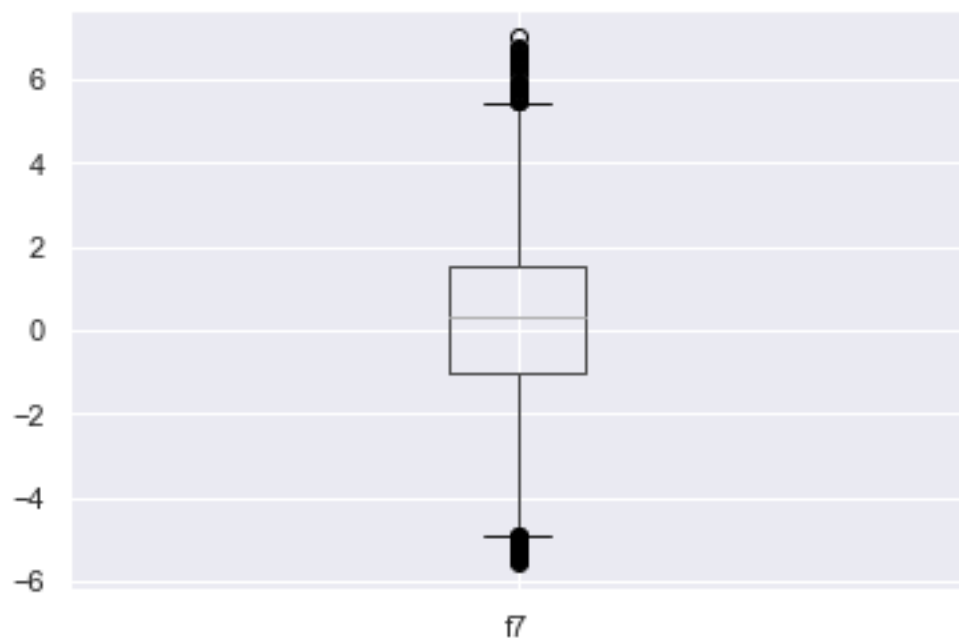
```
plt.figure()
```

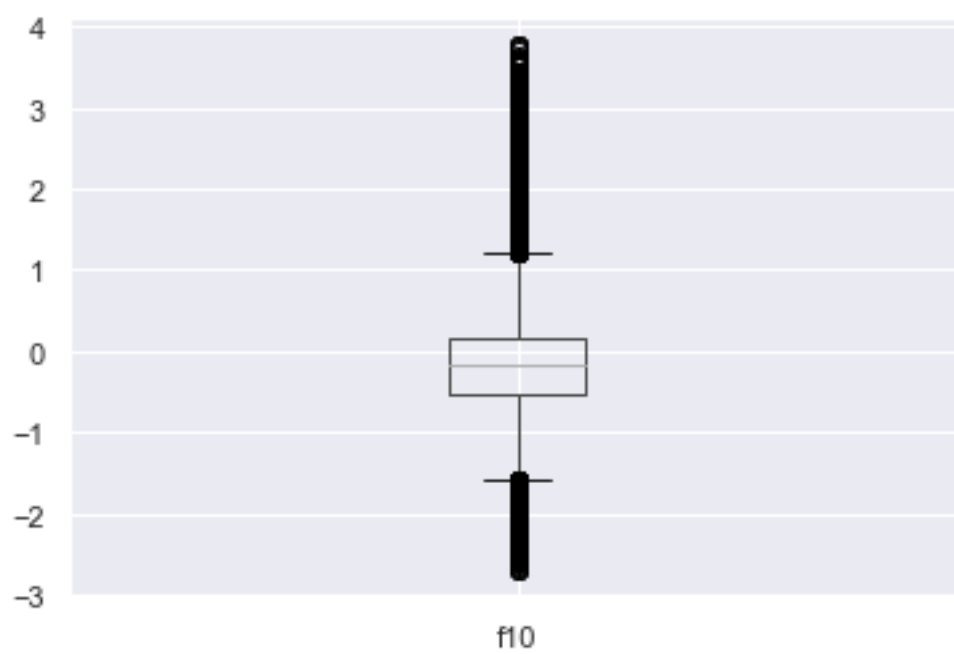
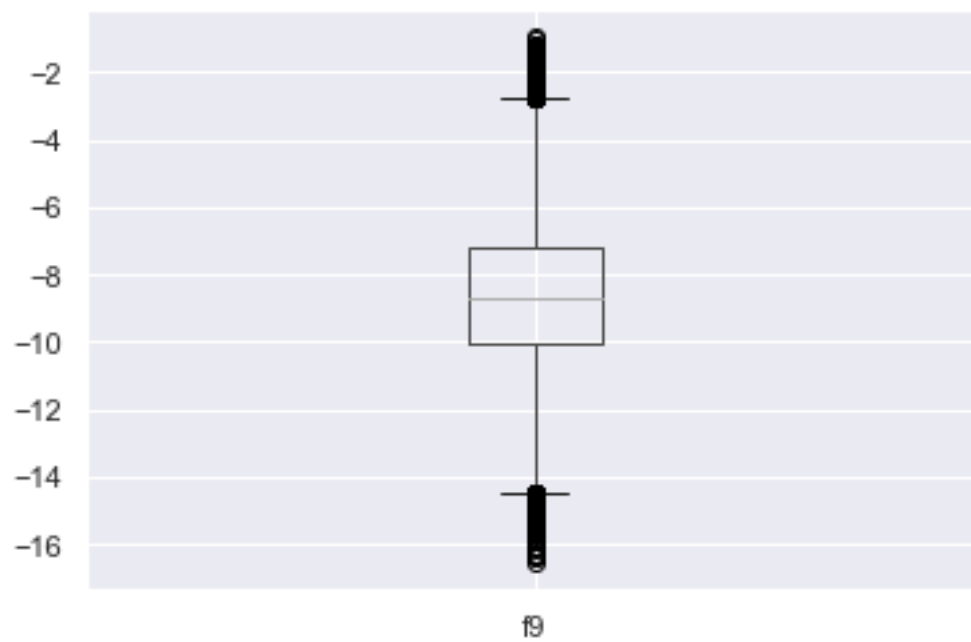


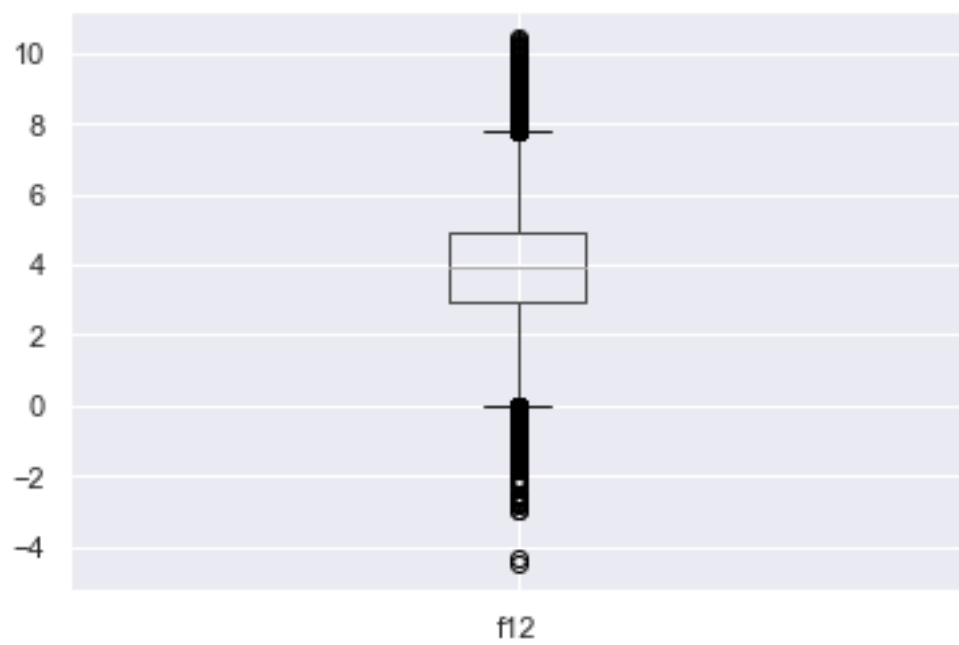
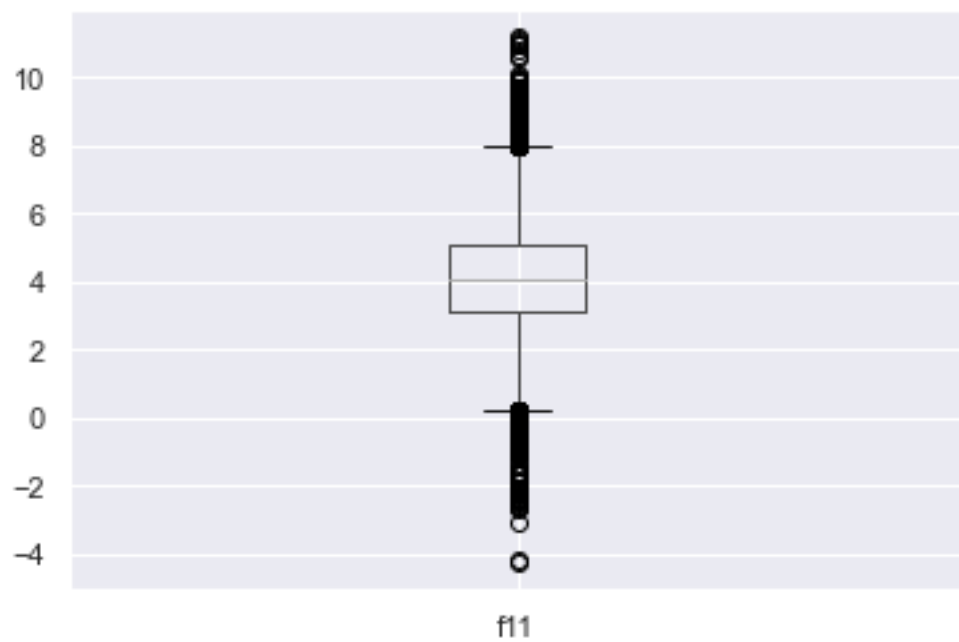


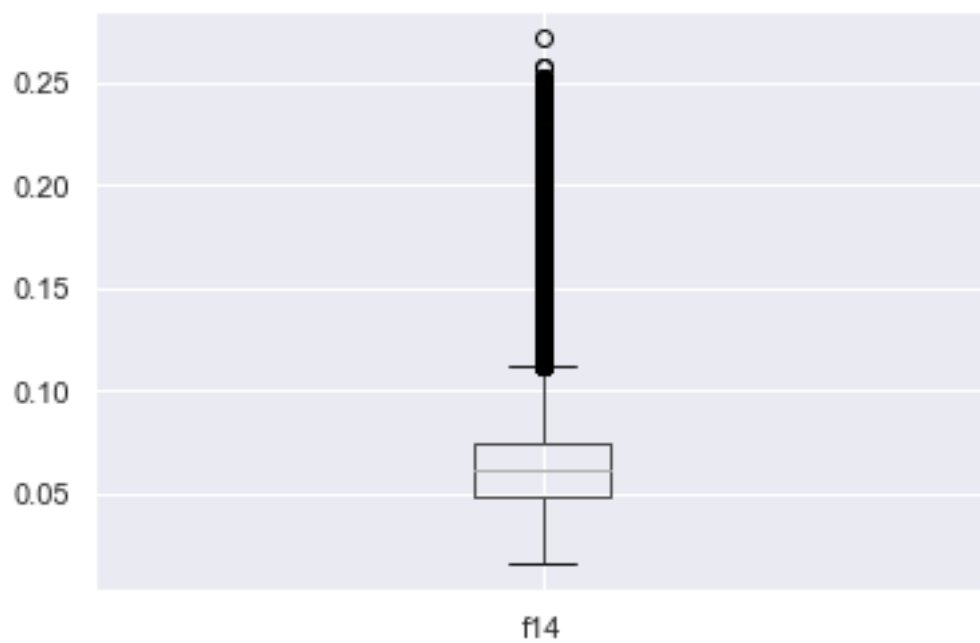
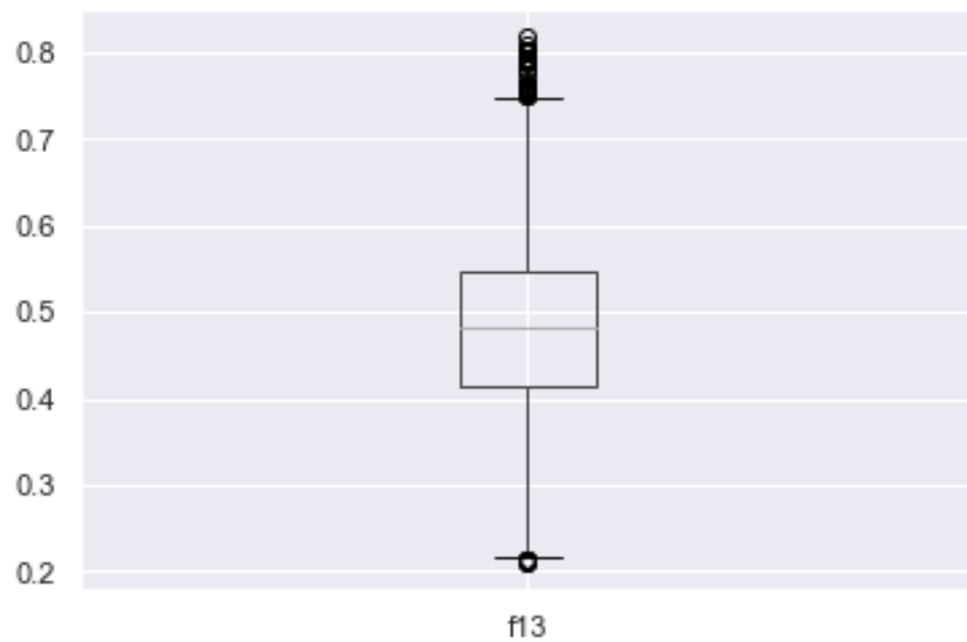


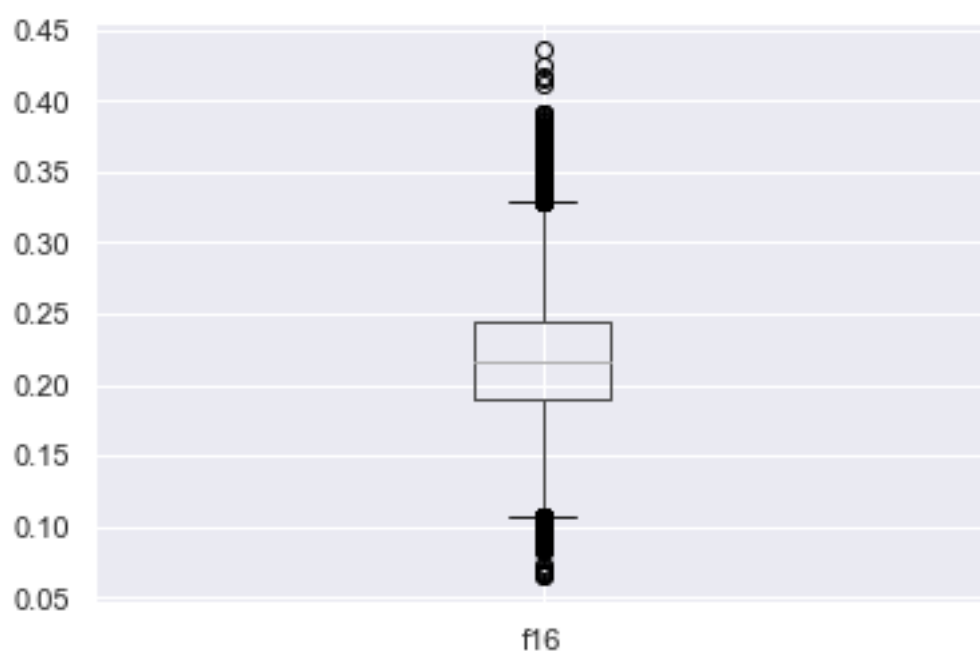
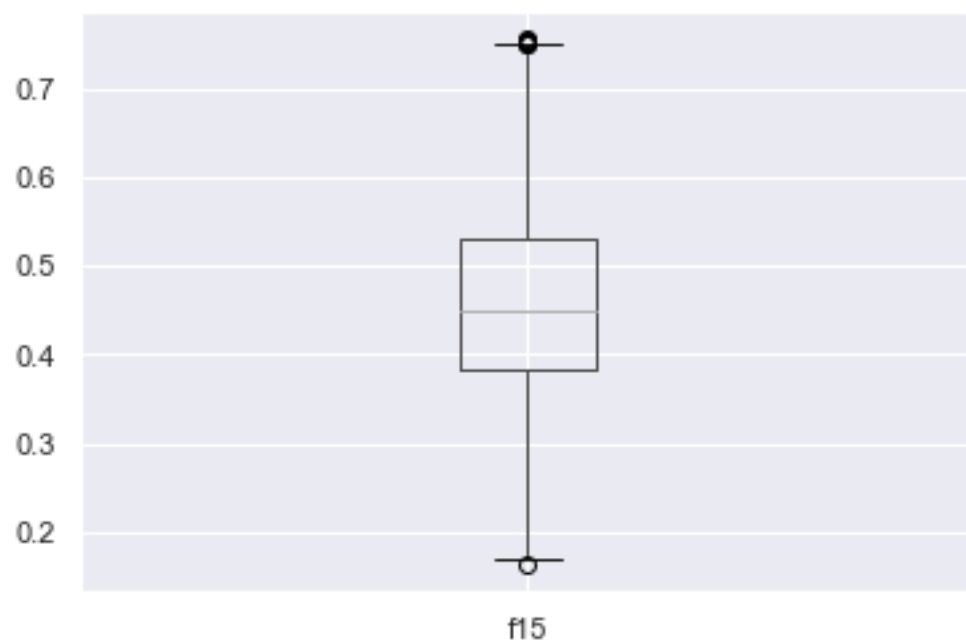


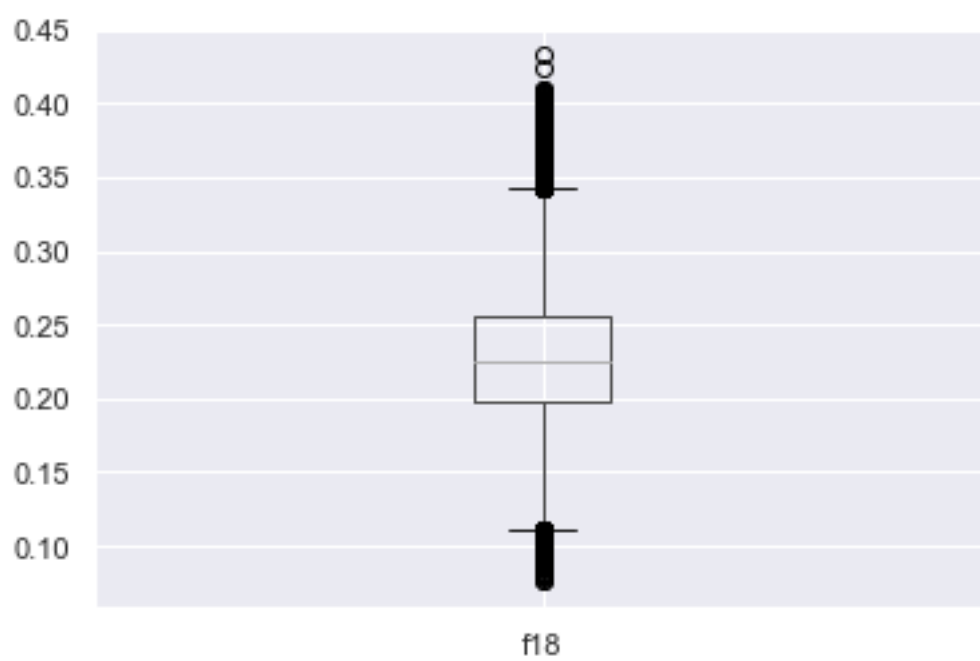
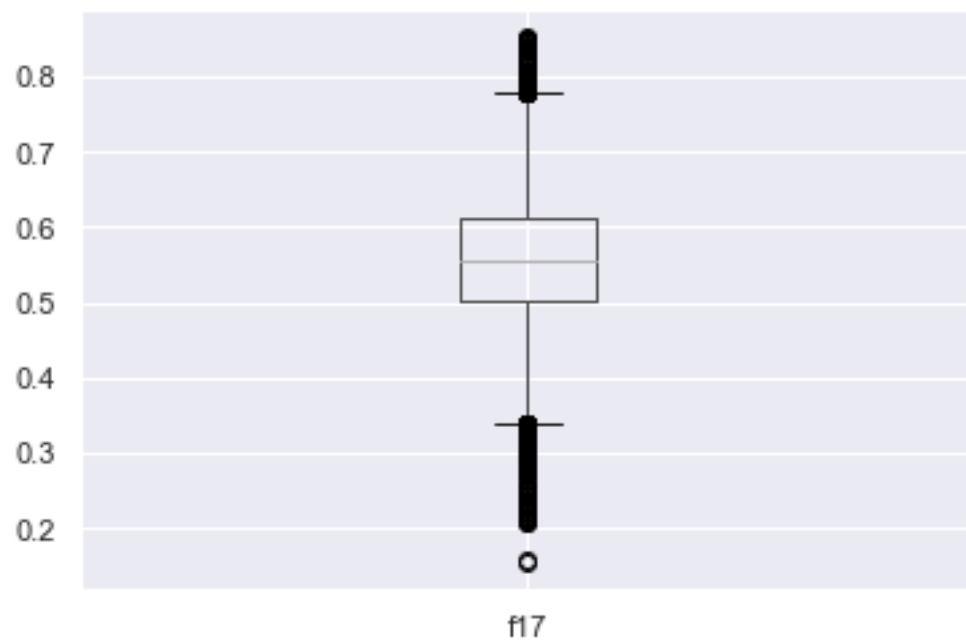


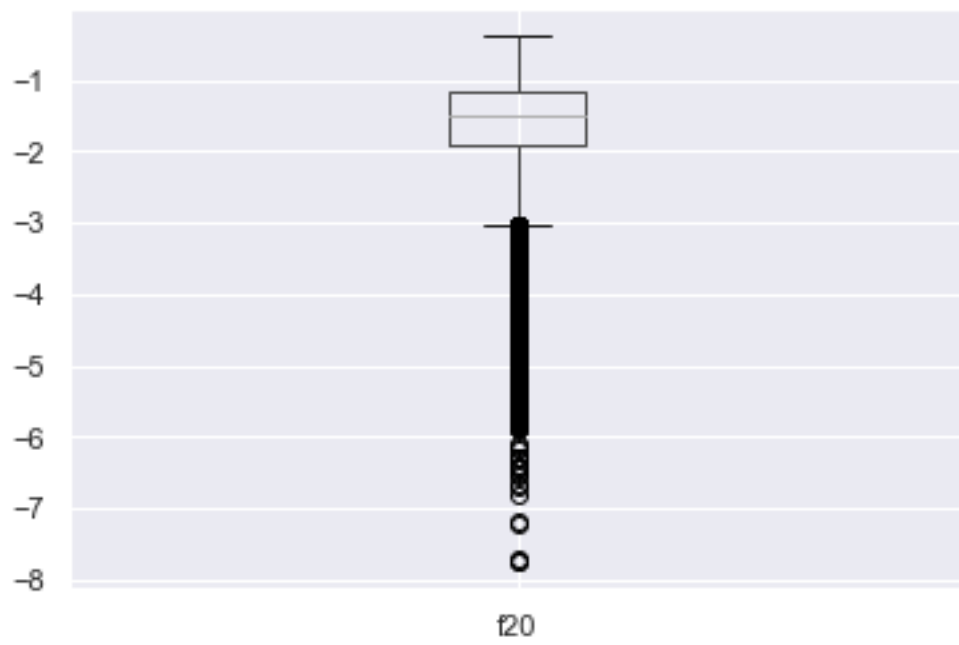
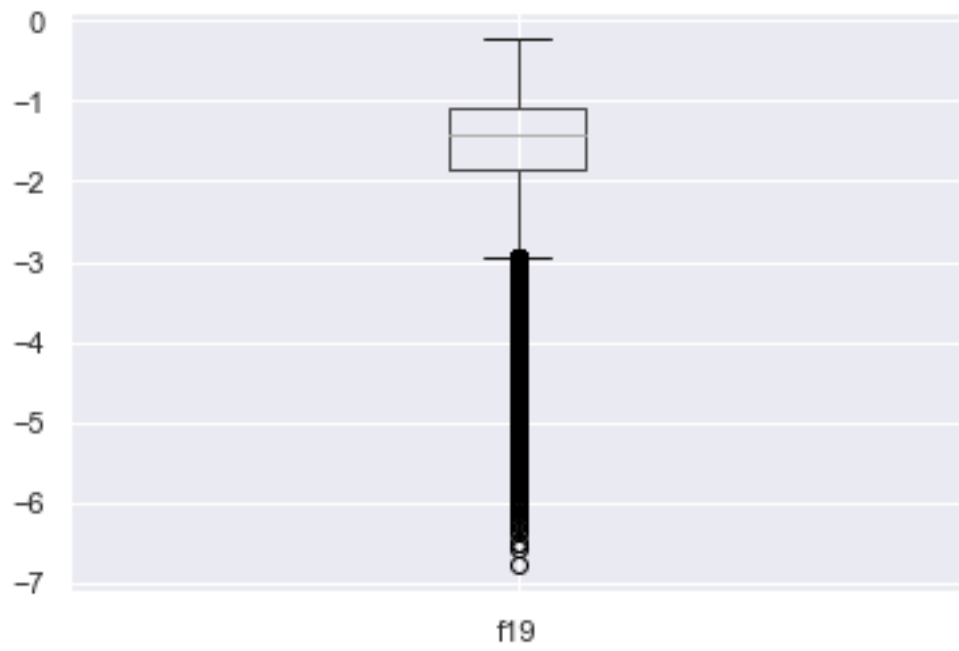


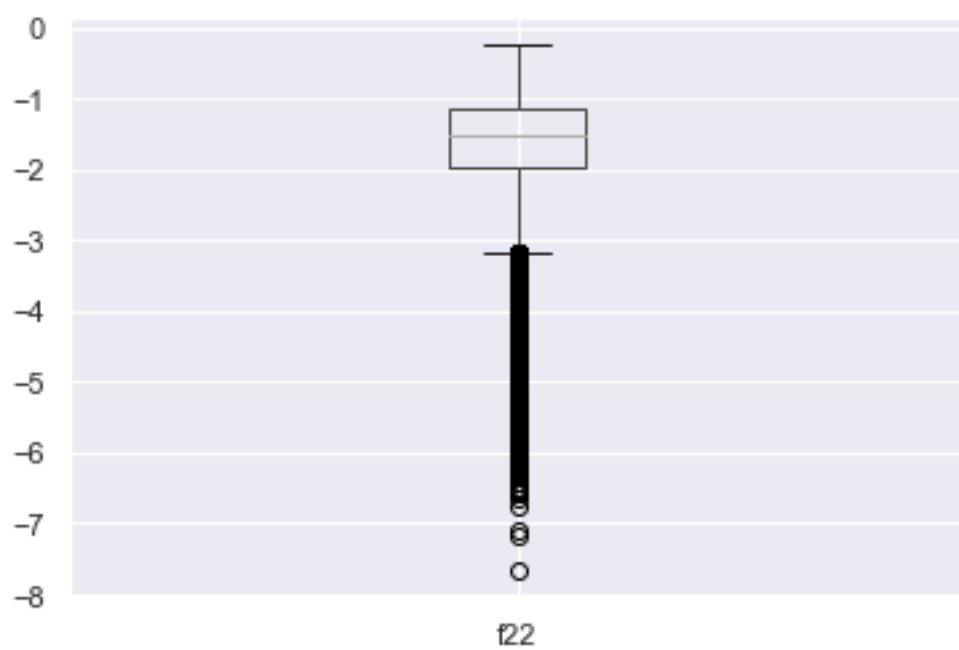
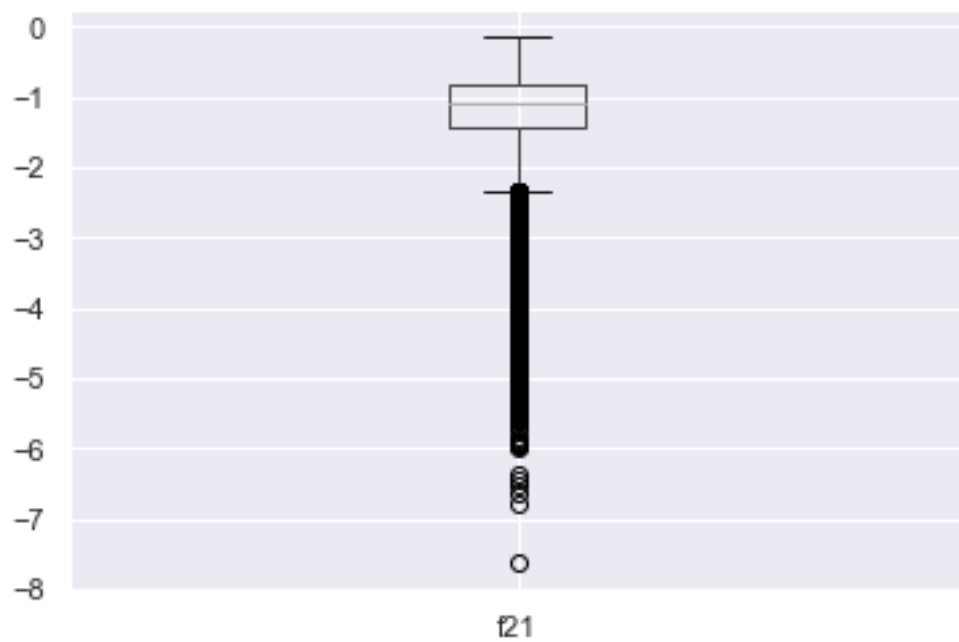


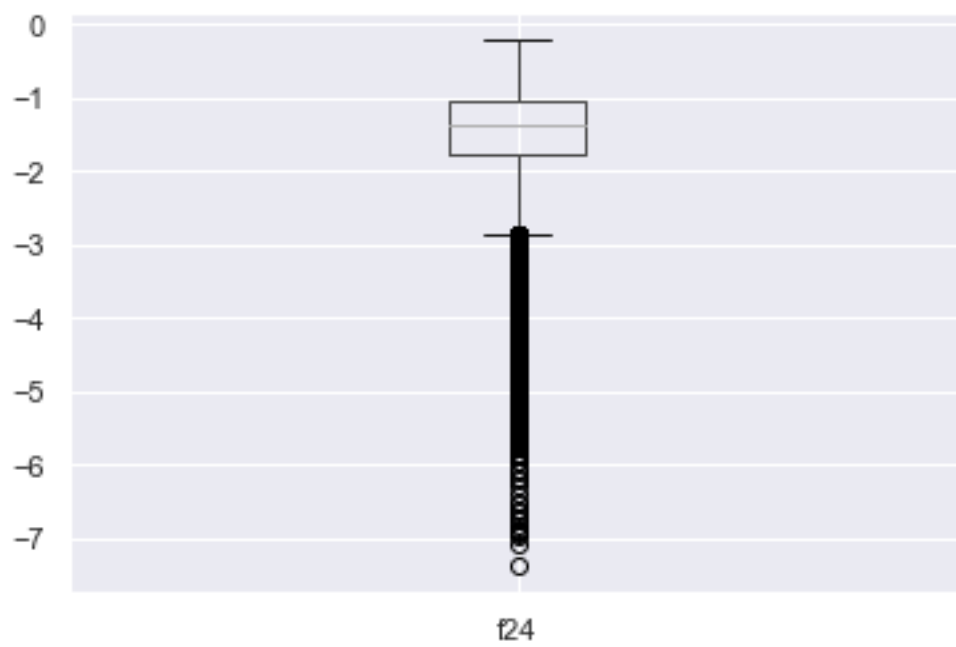
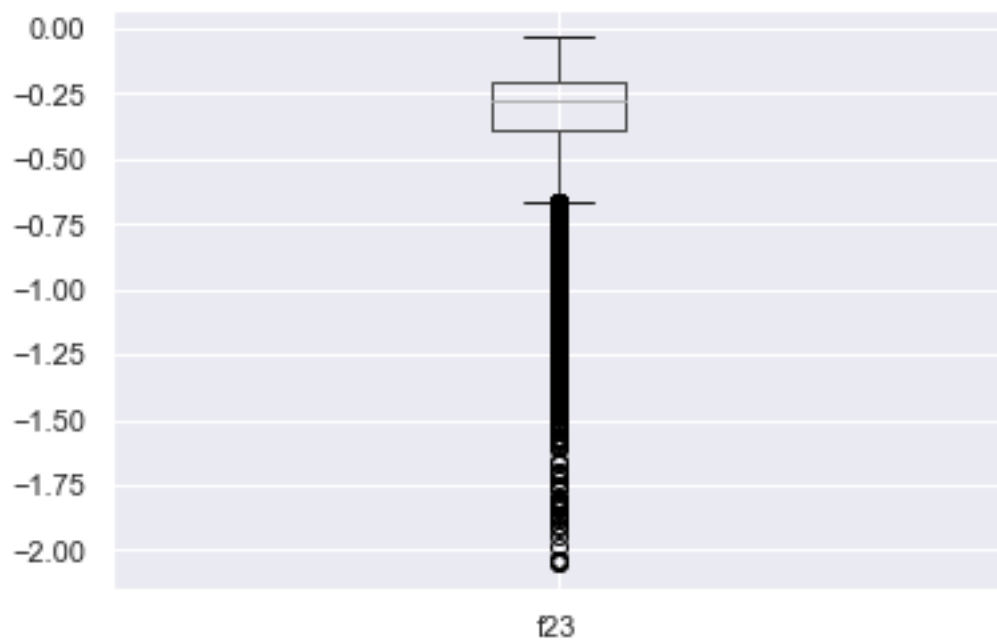


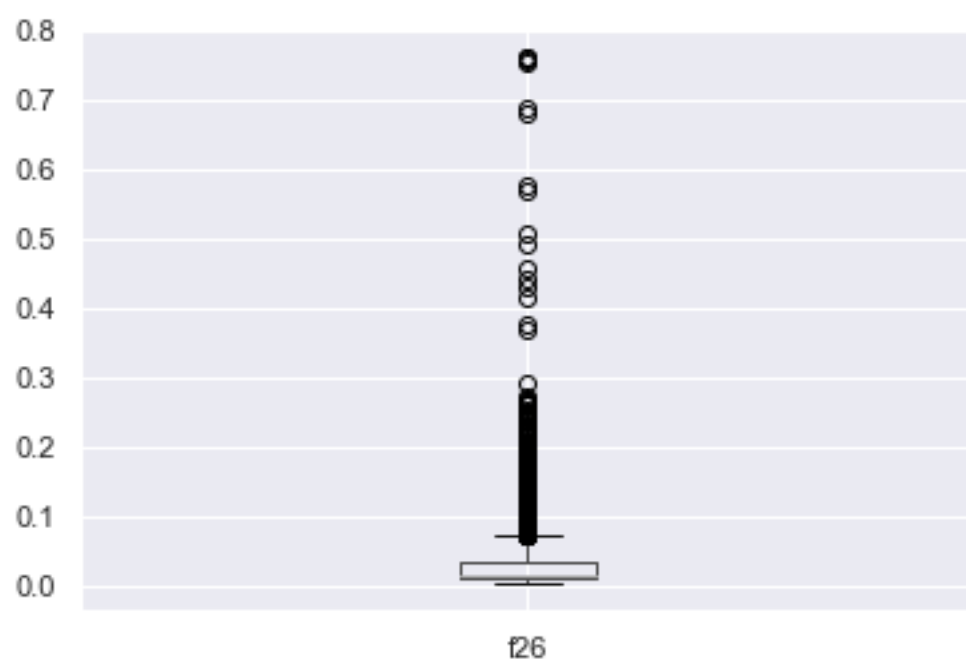
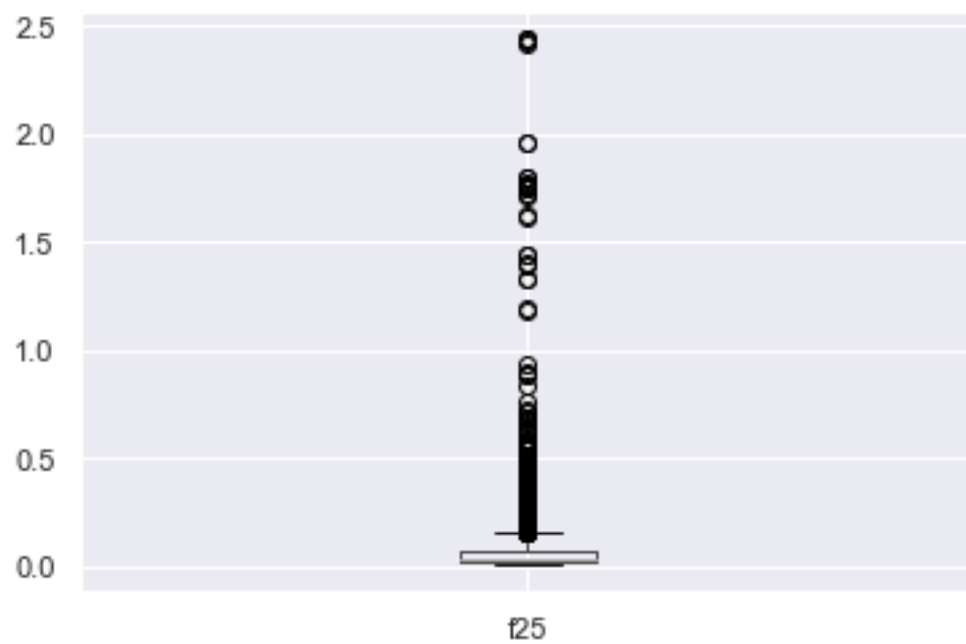


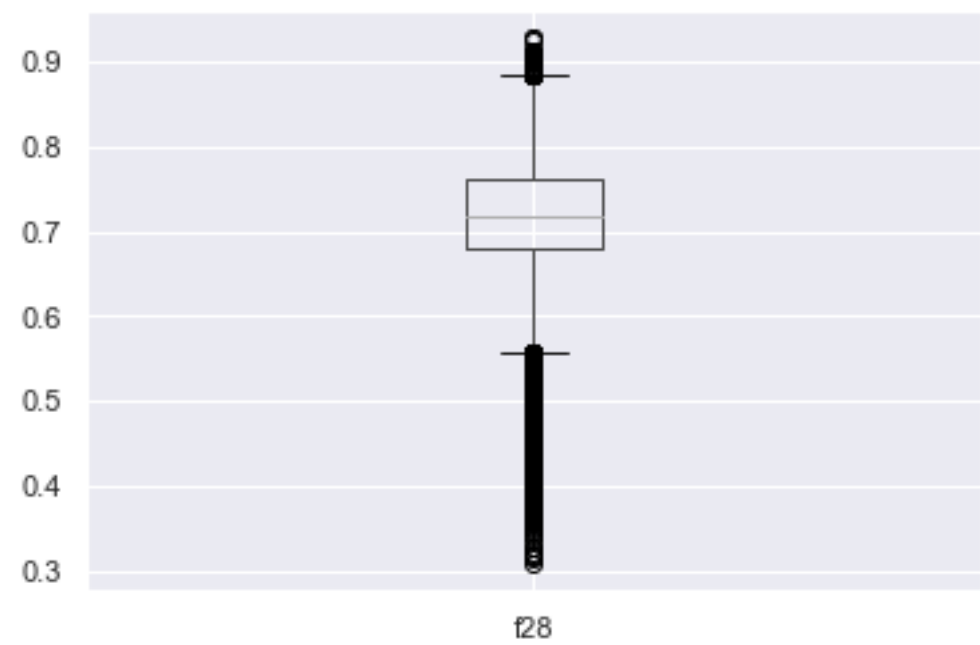
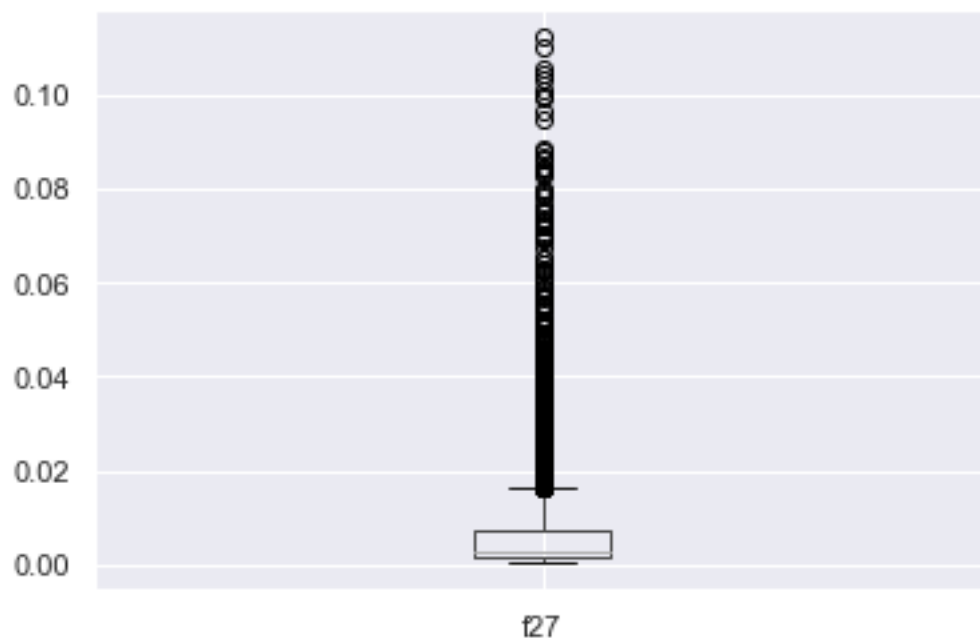


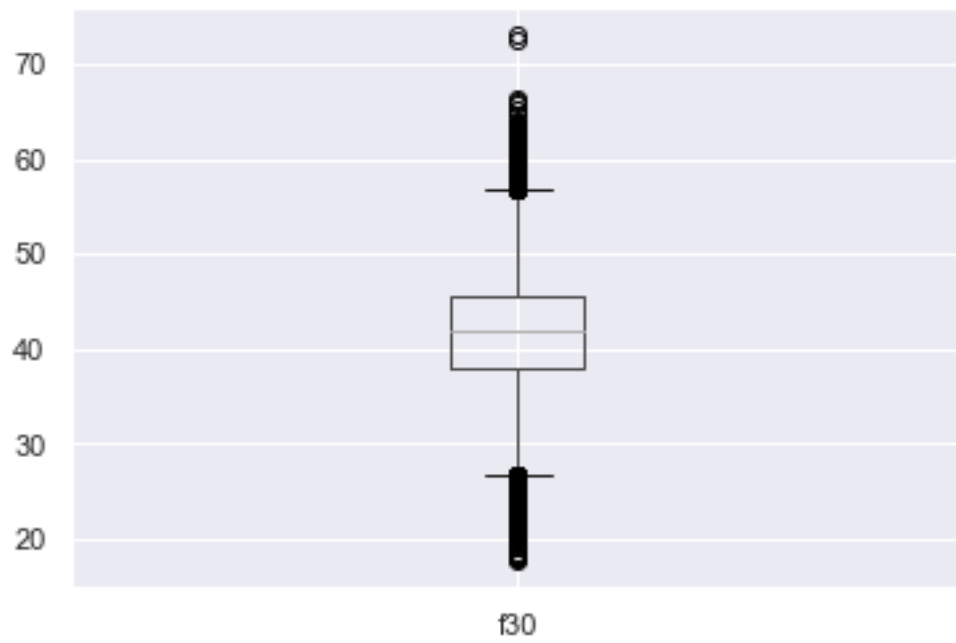
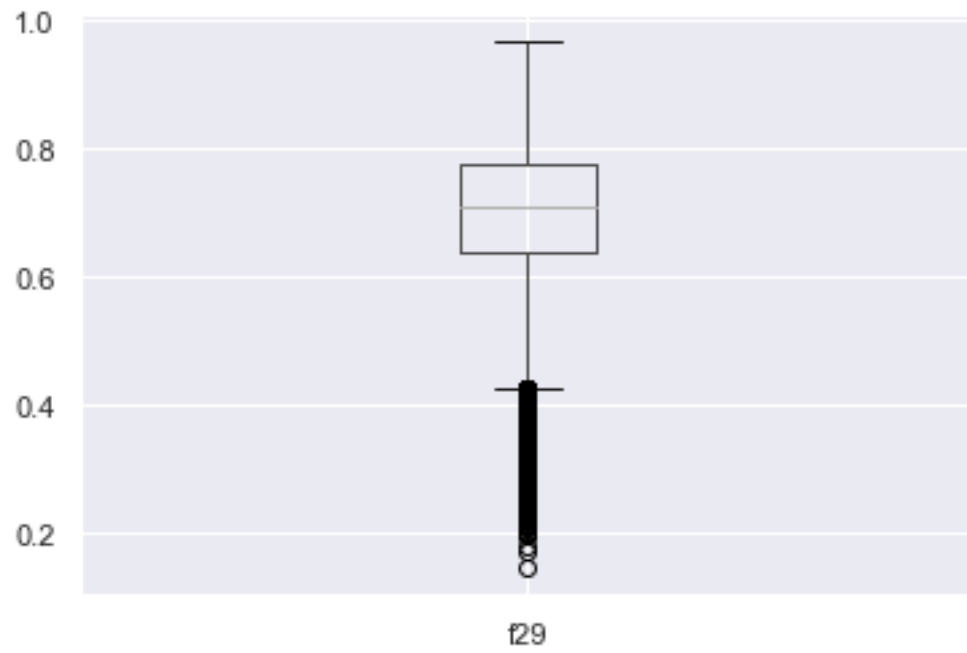


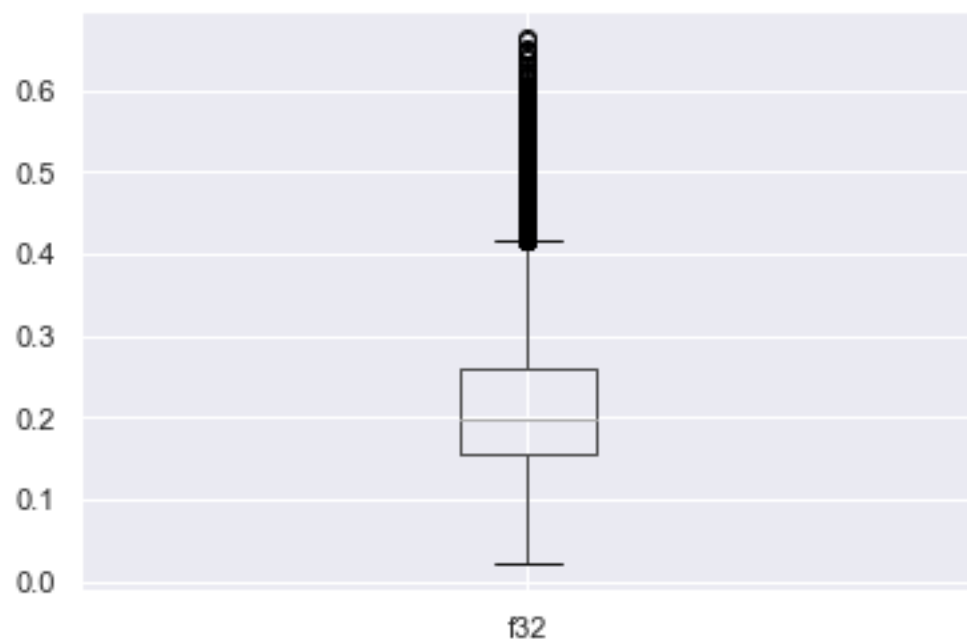
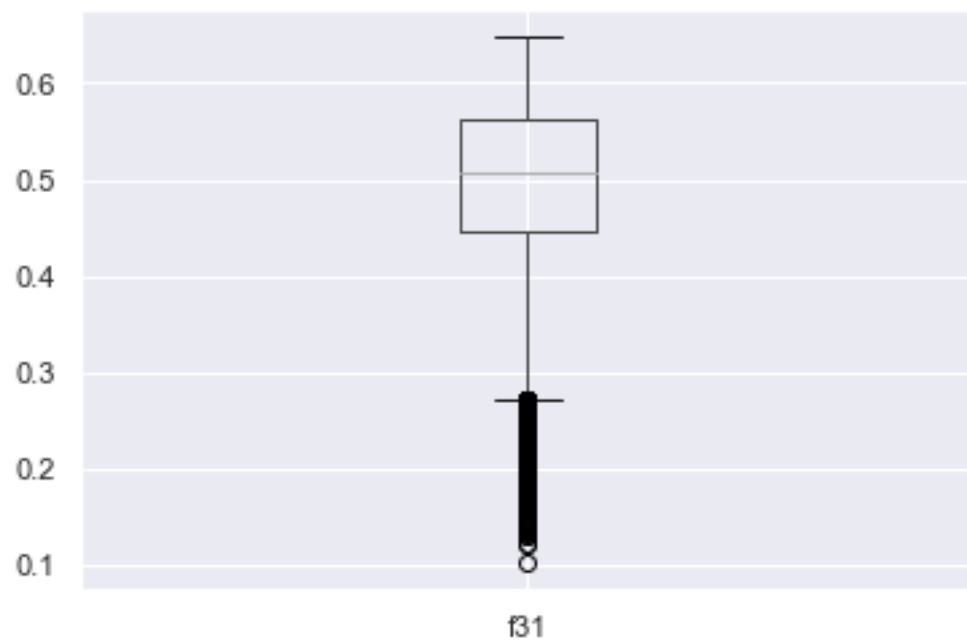


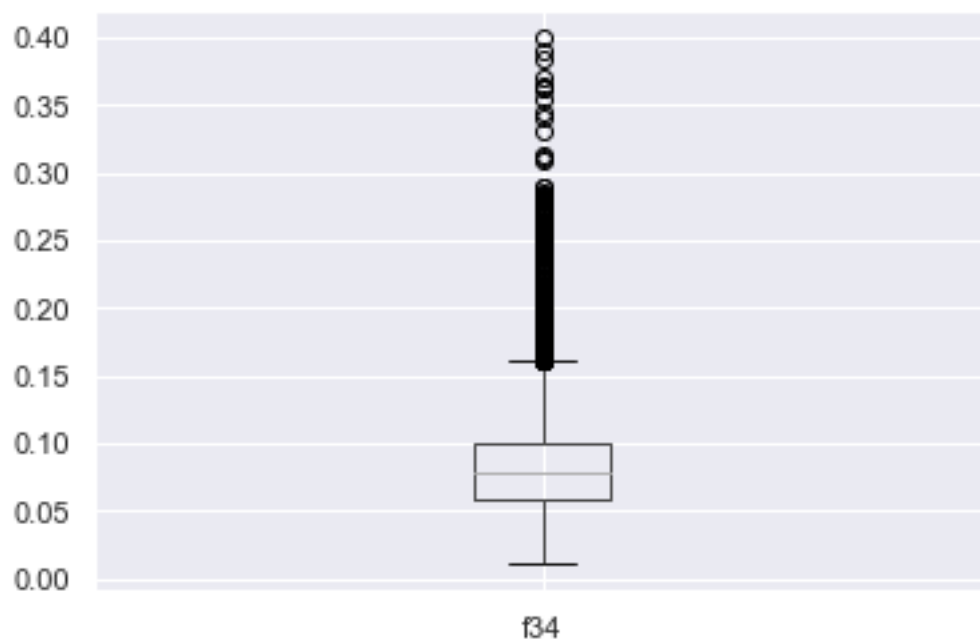
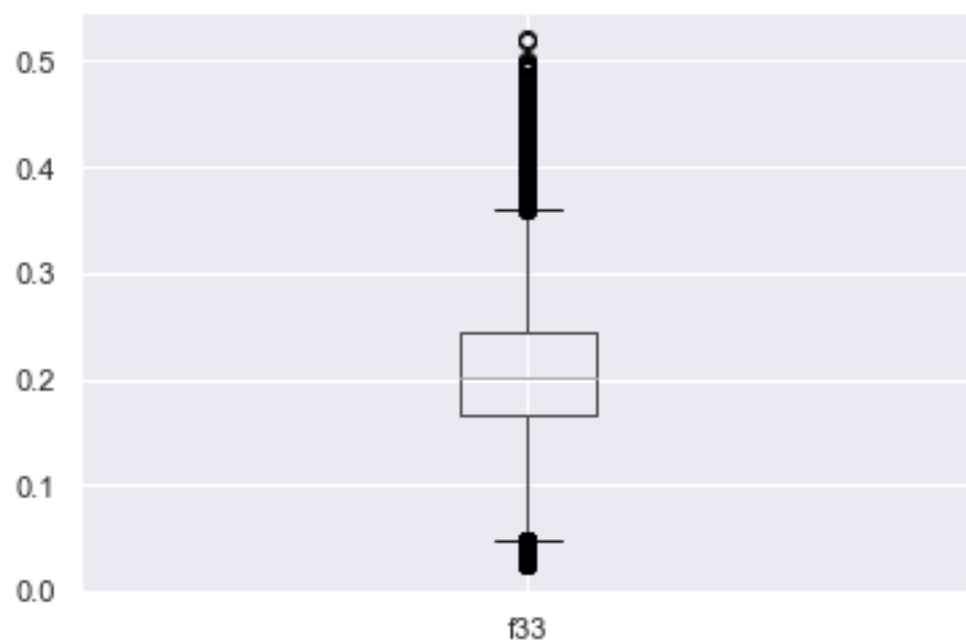


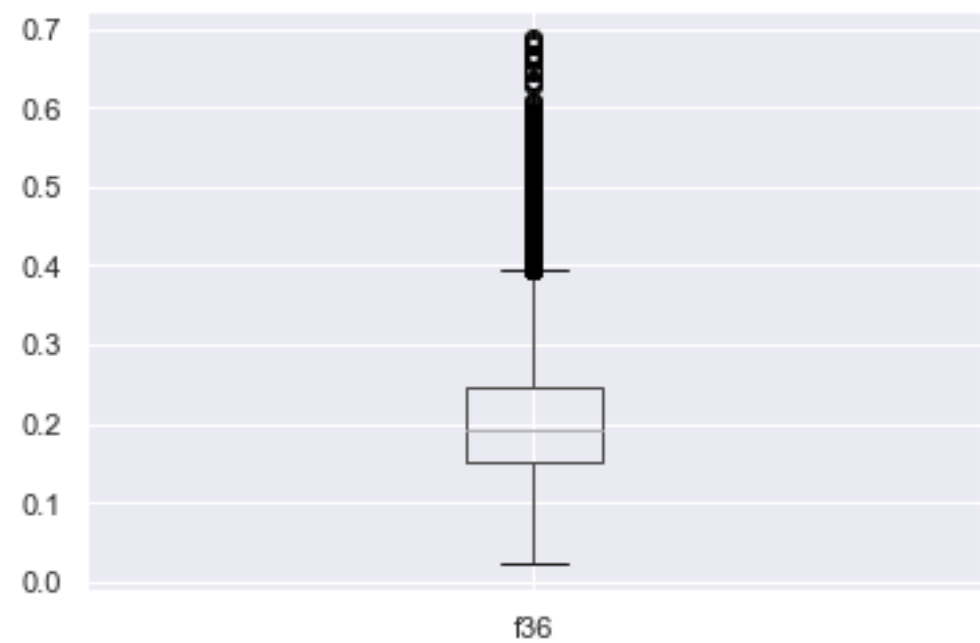
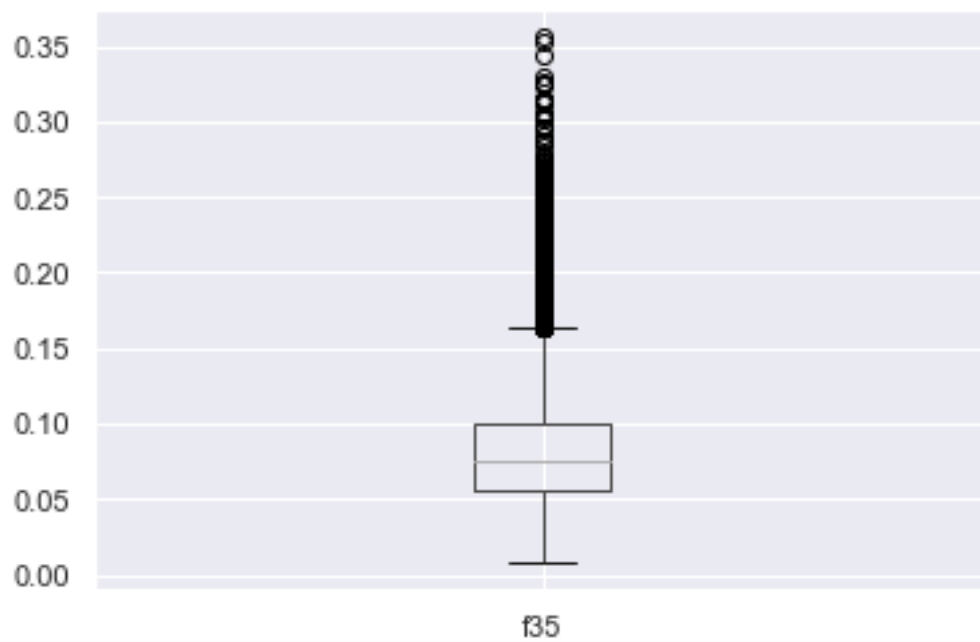


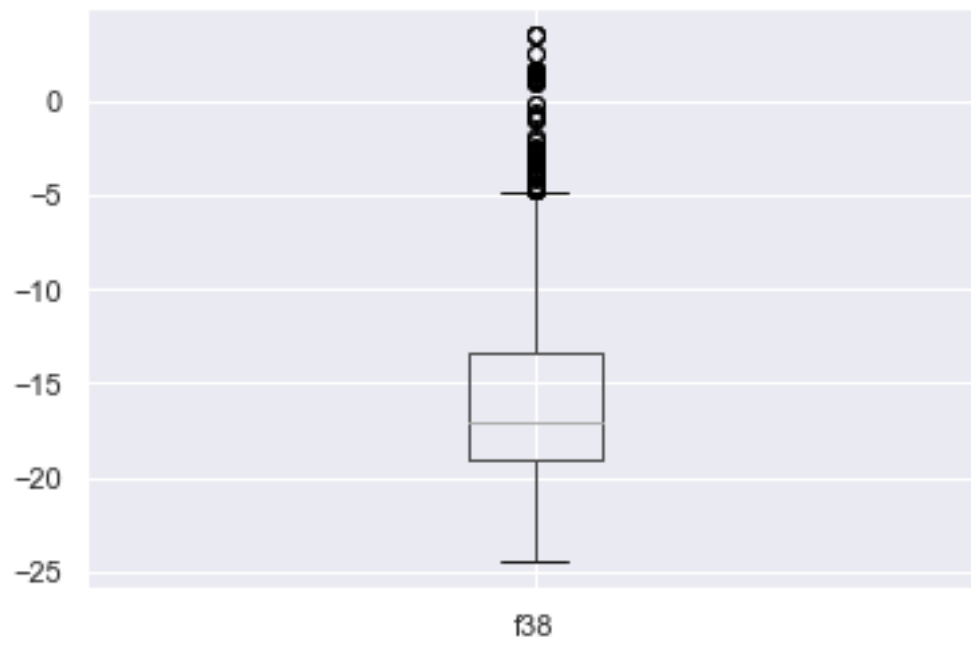
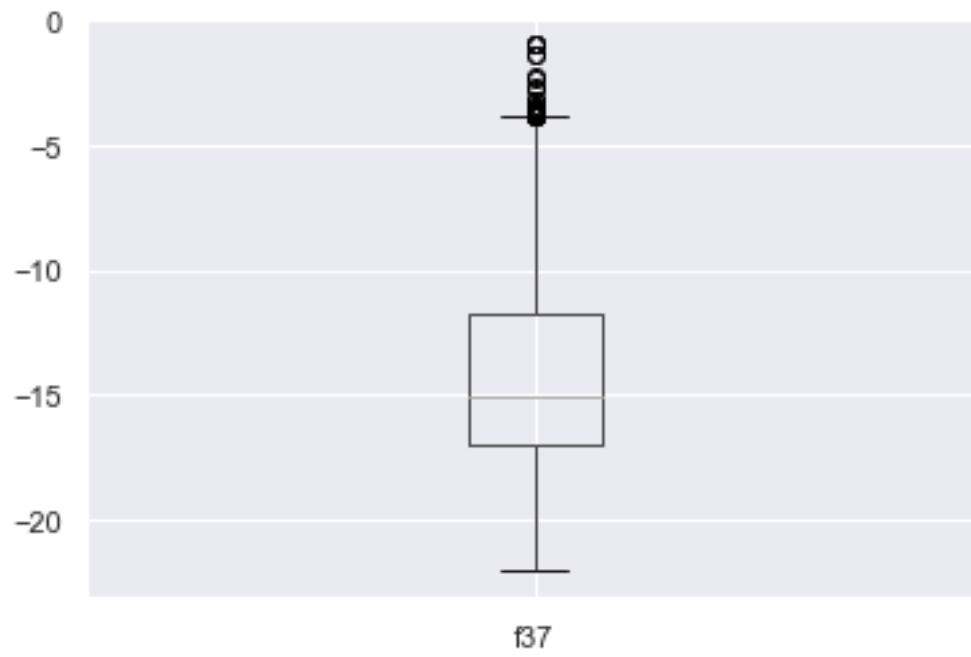


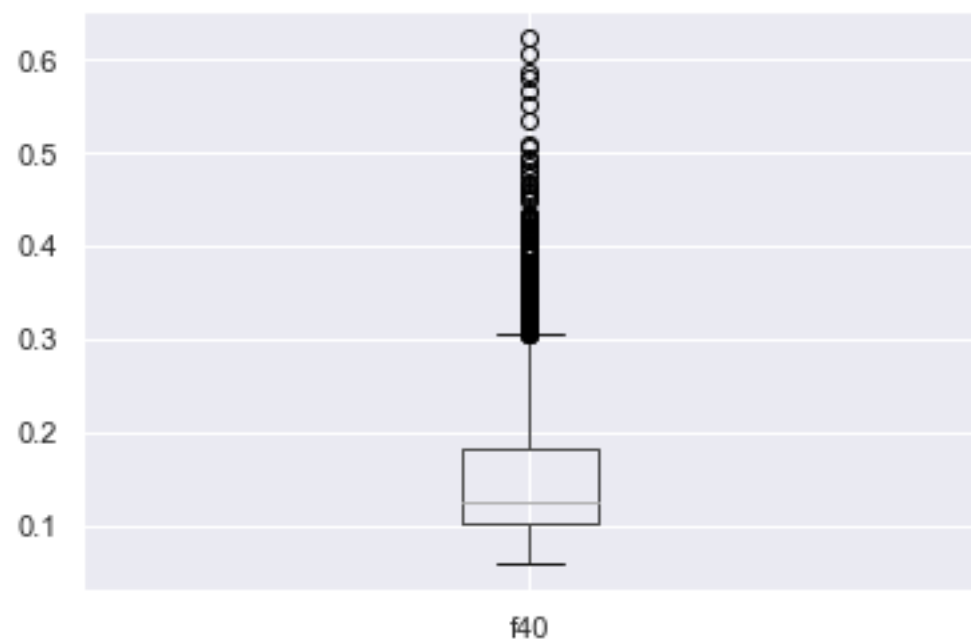
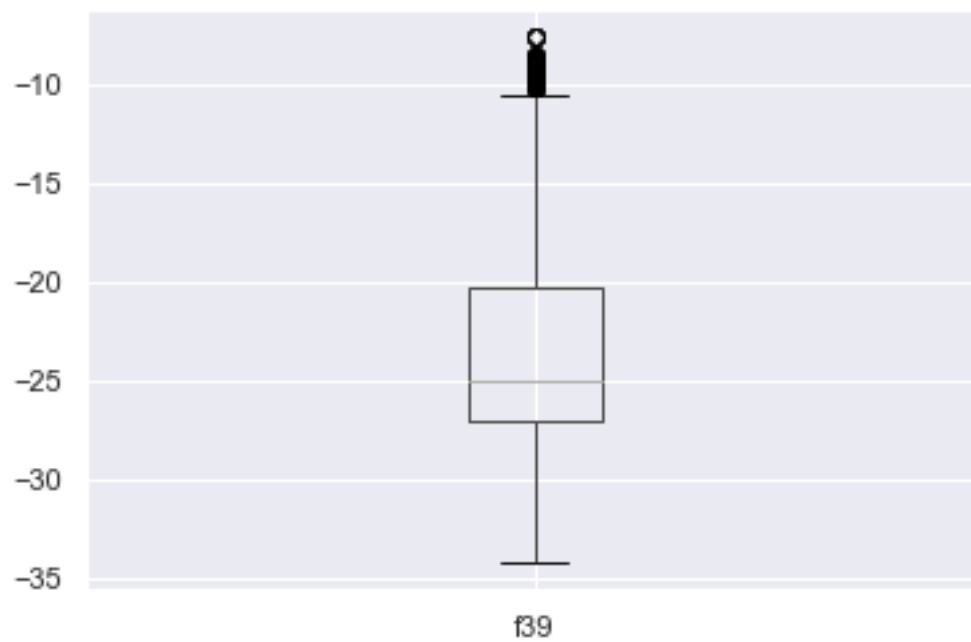


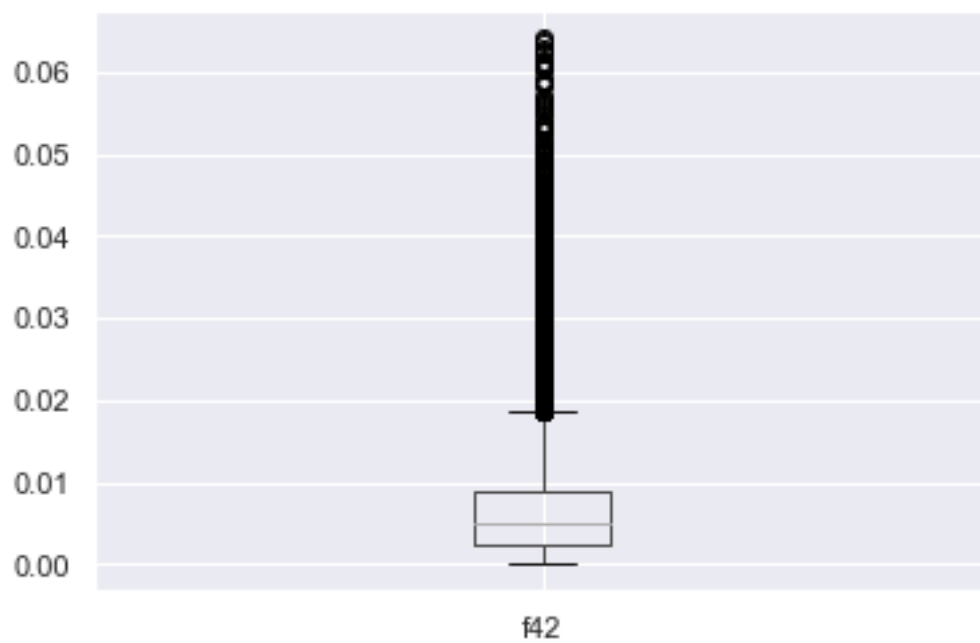
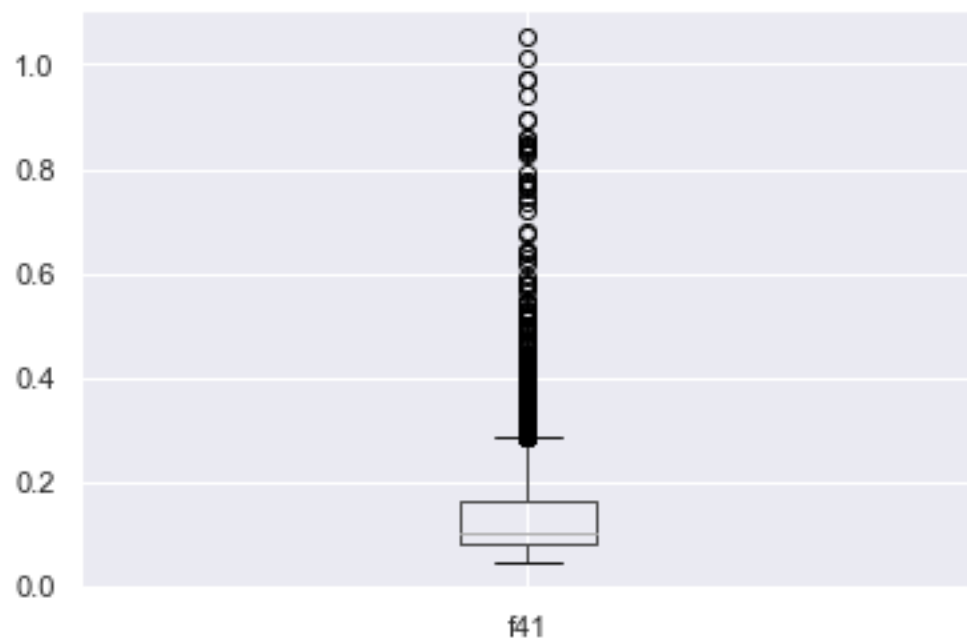


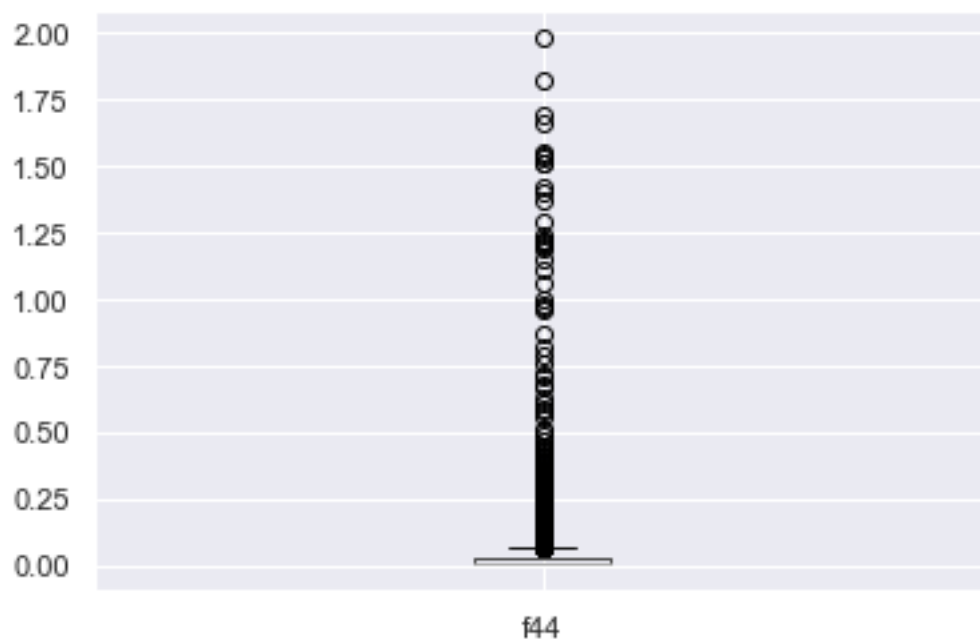
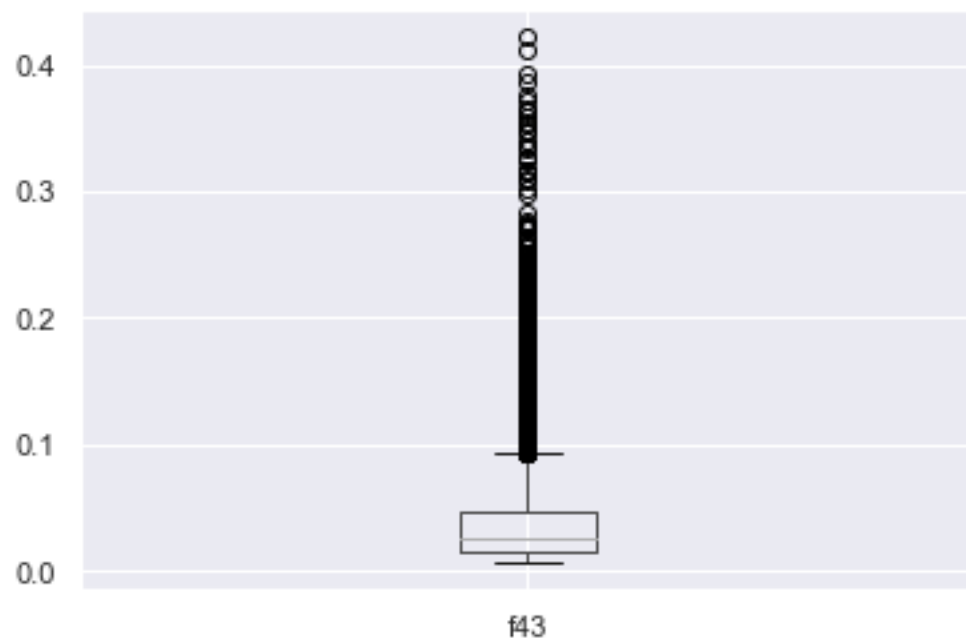


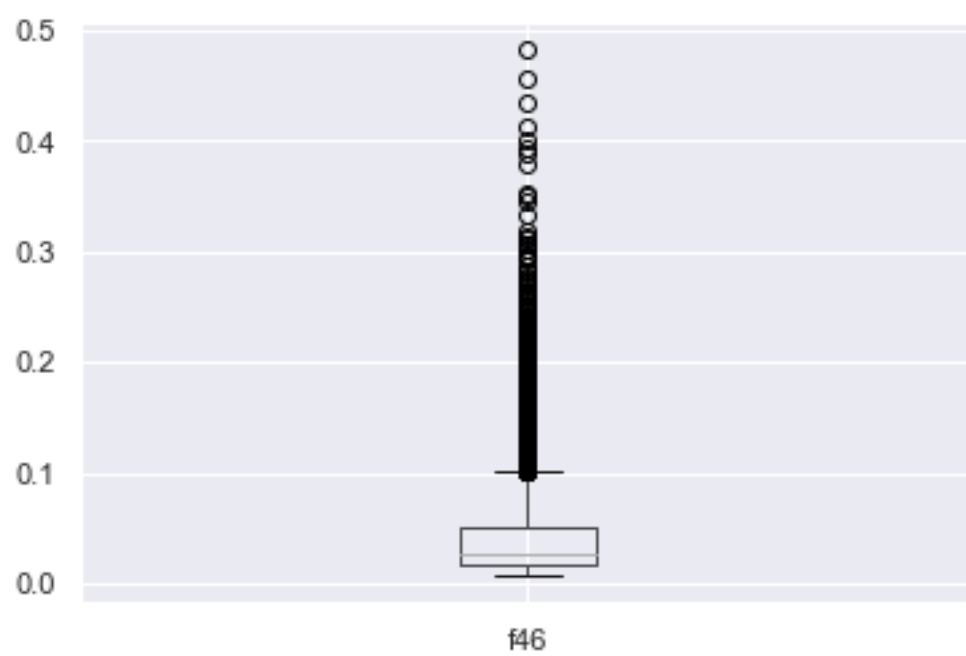
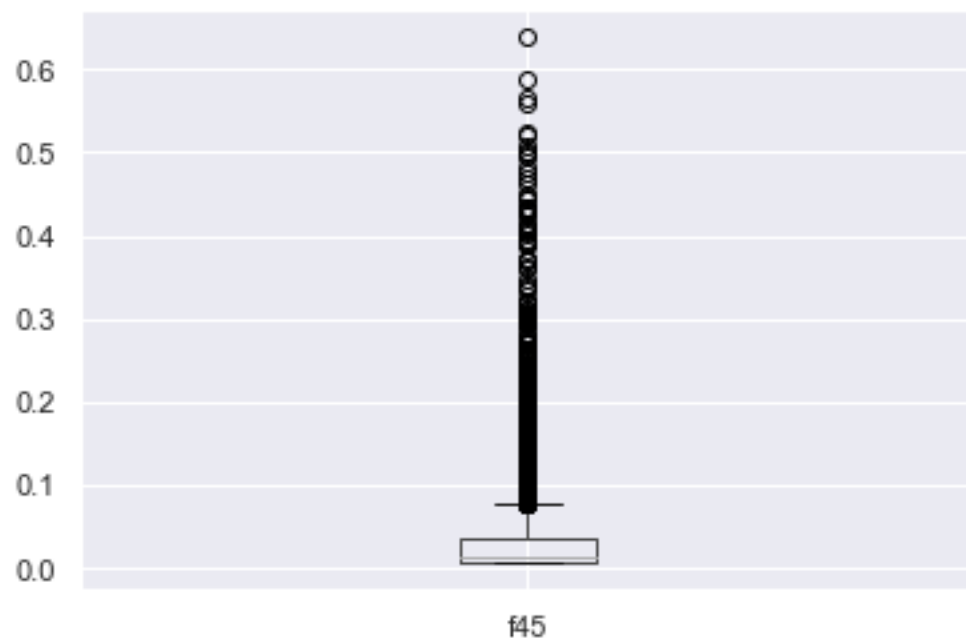


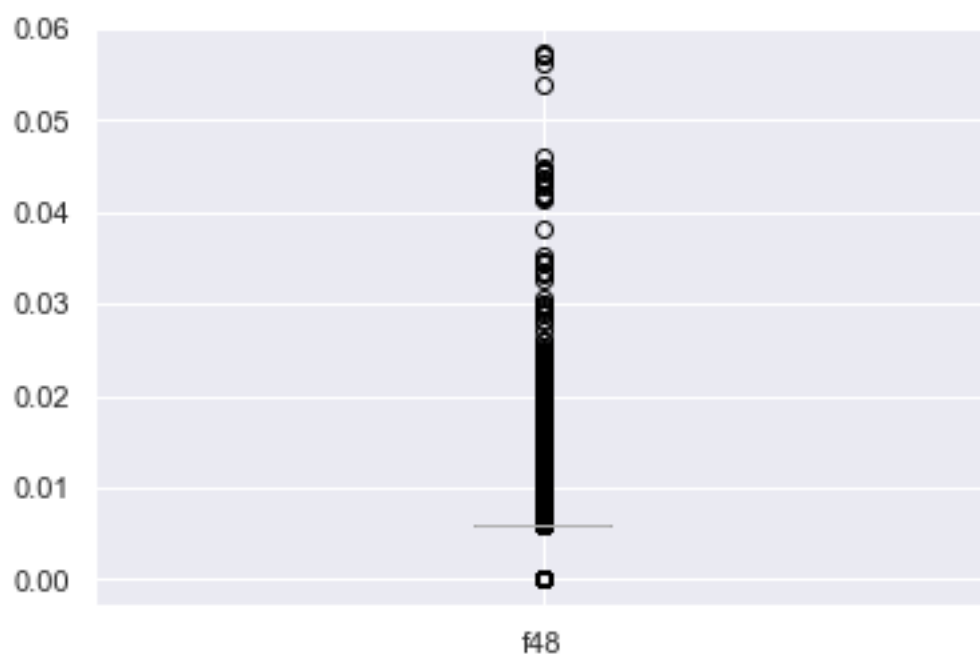
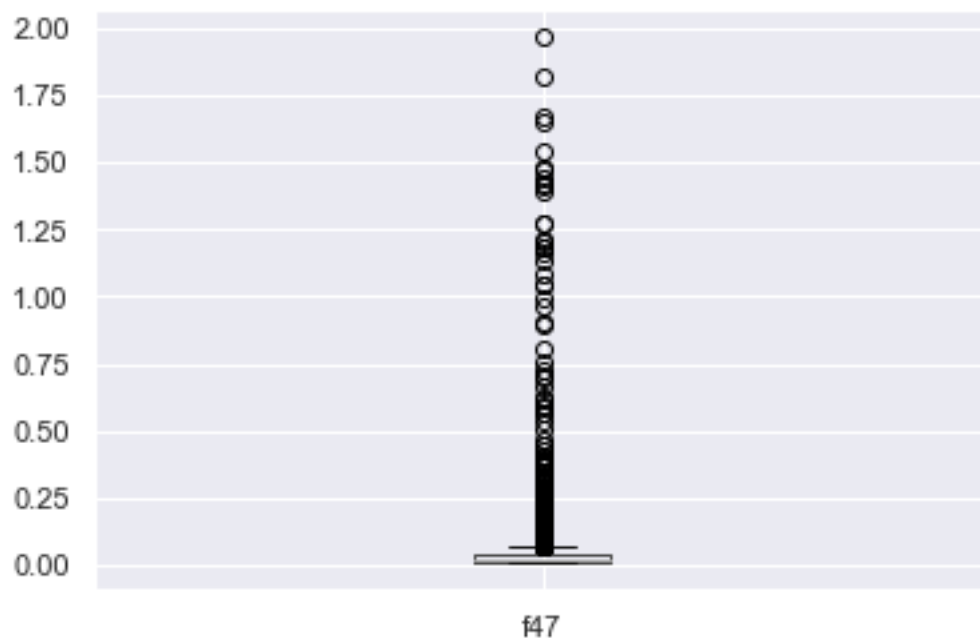


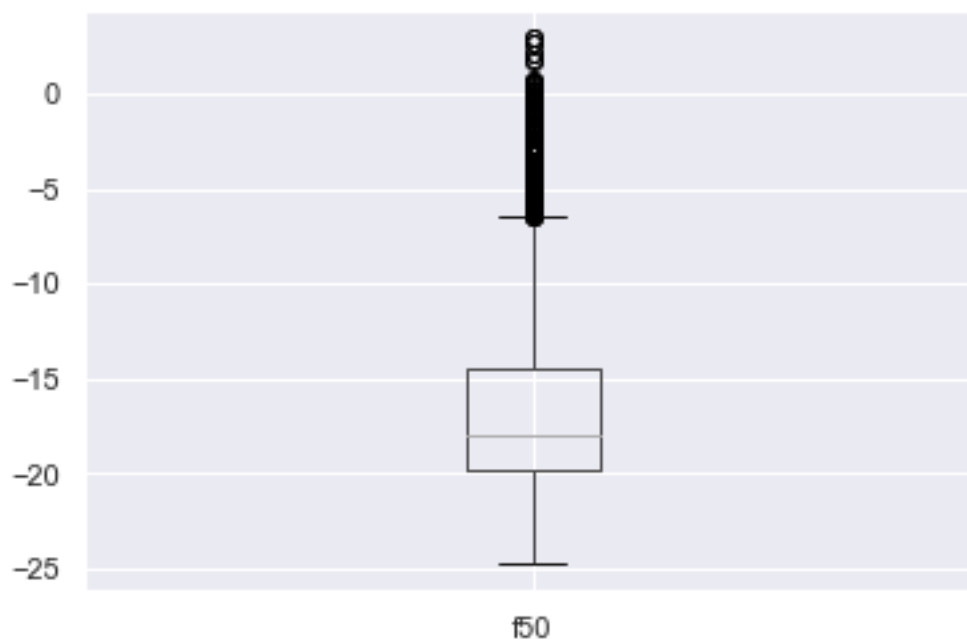
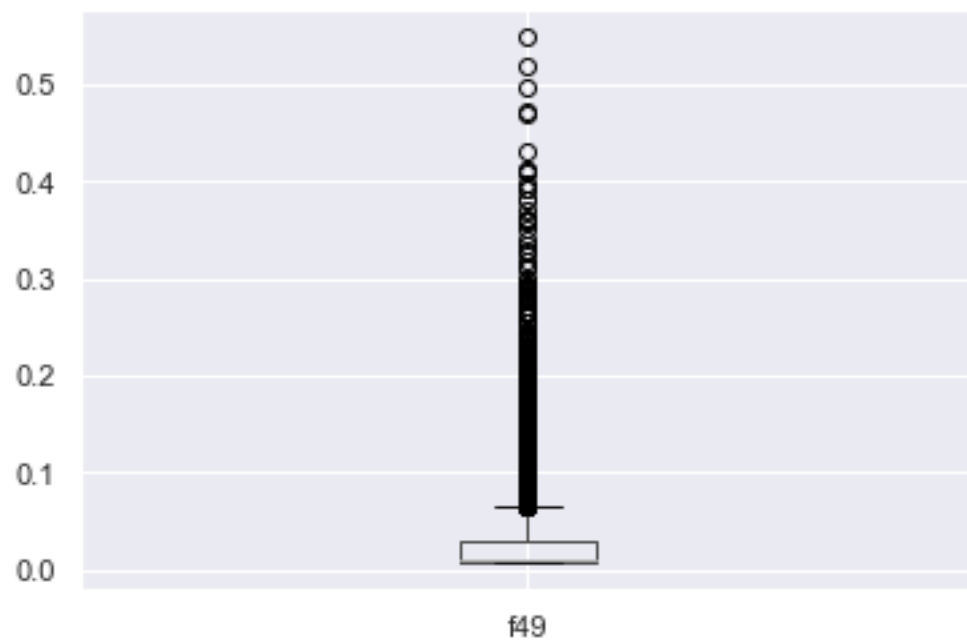


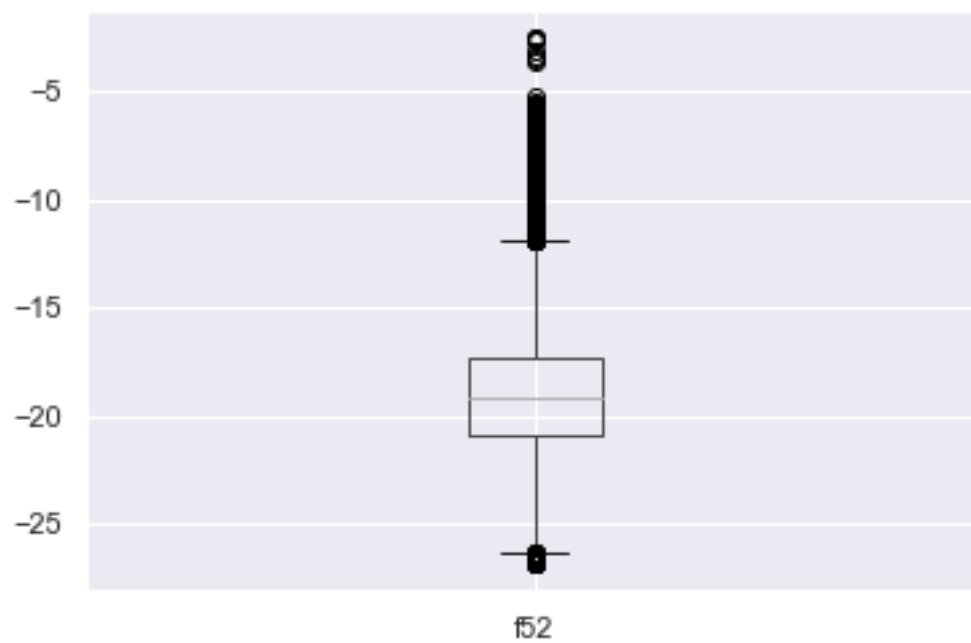
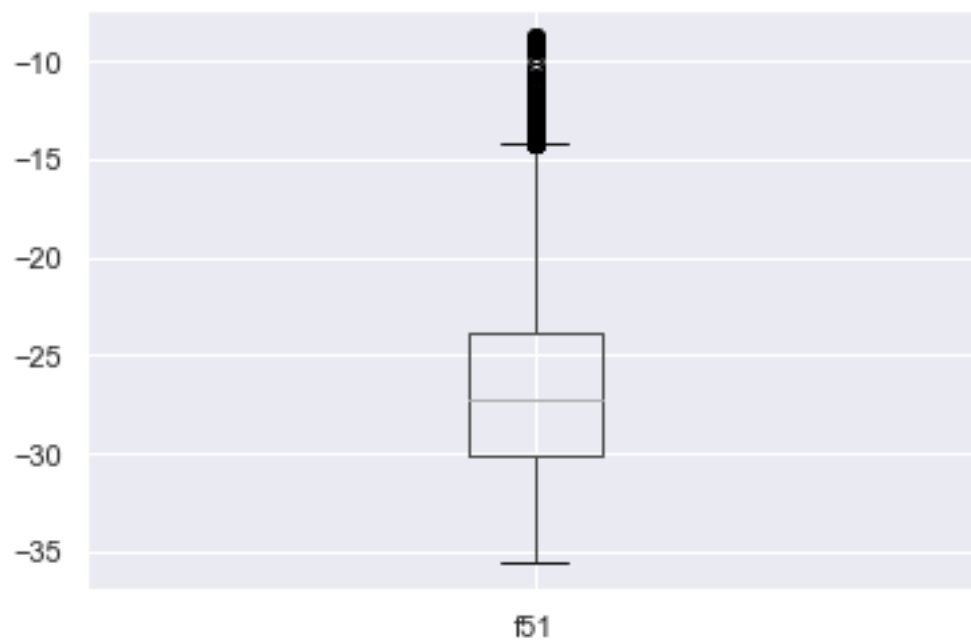


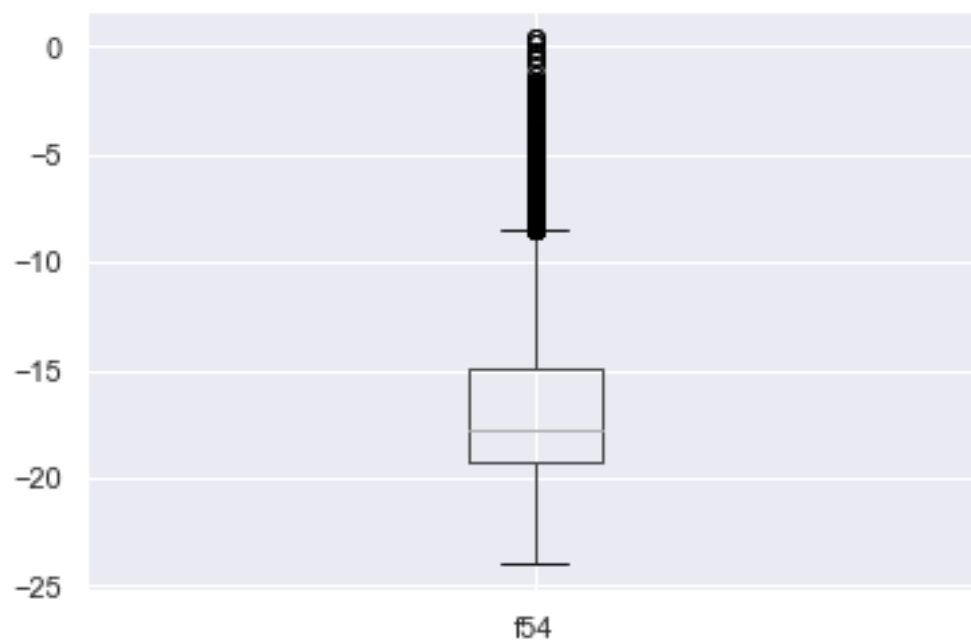
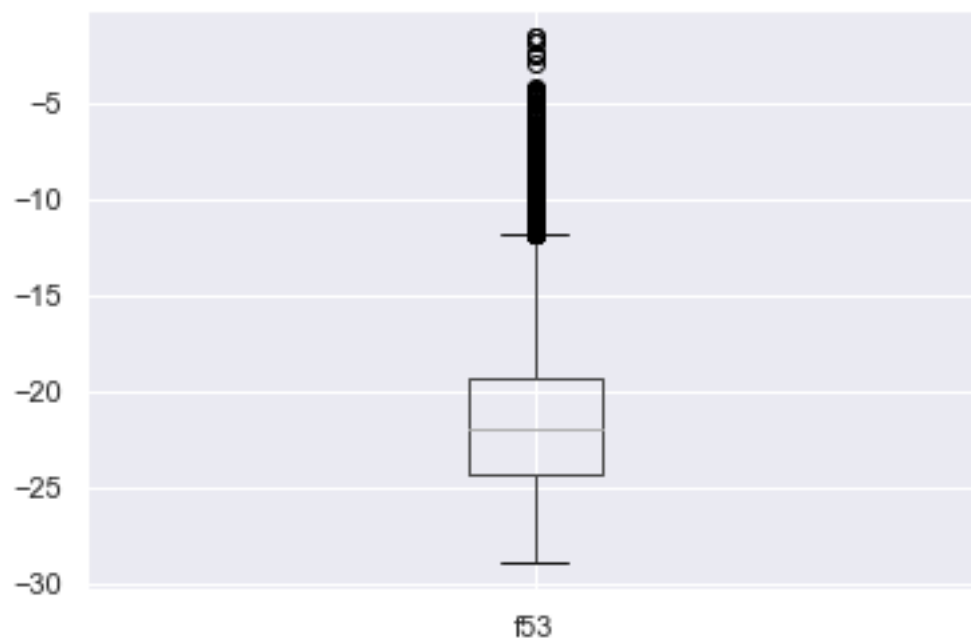


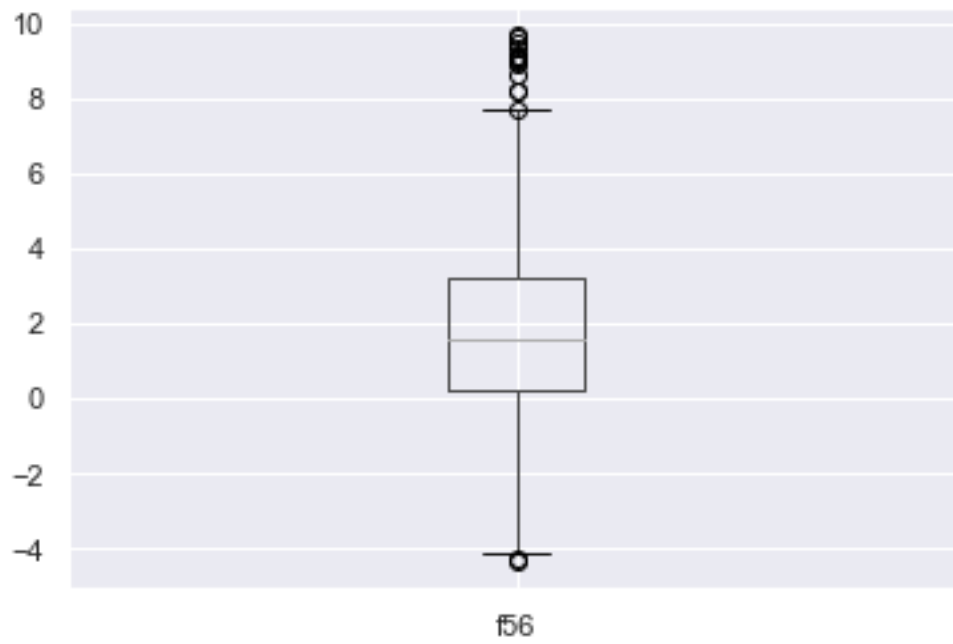


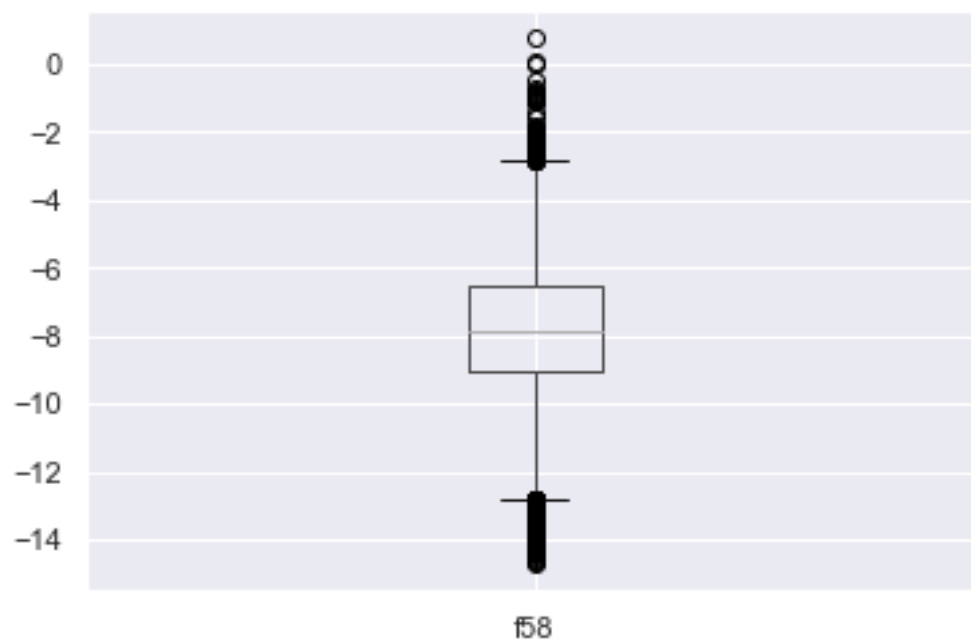
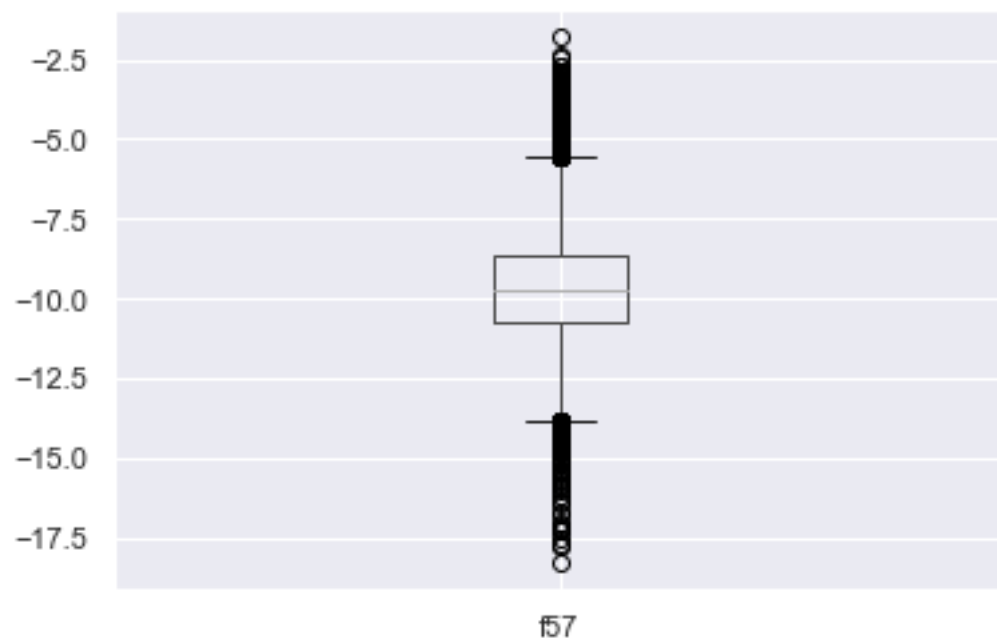


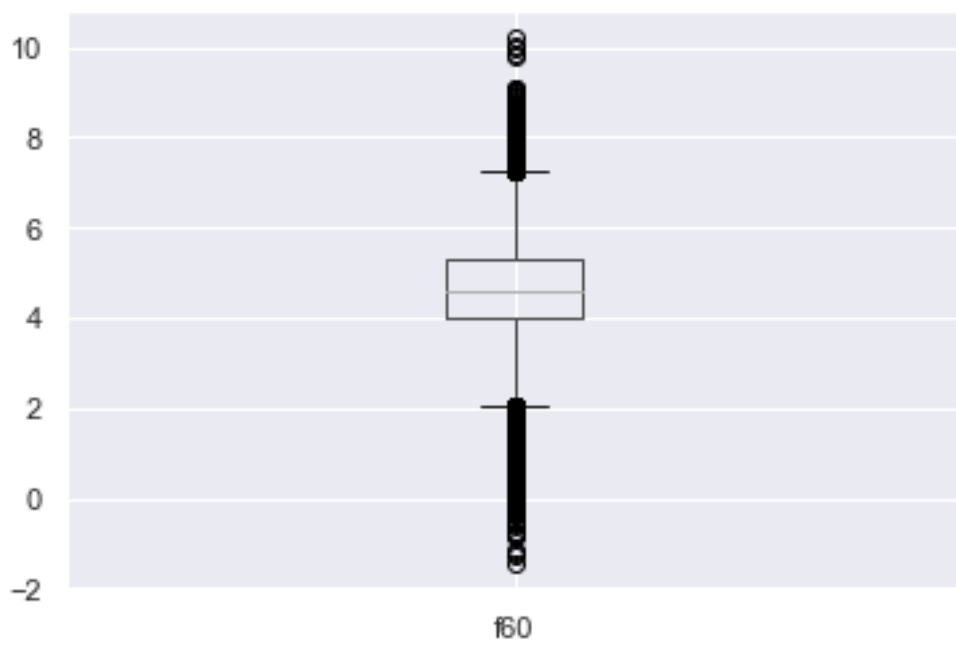
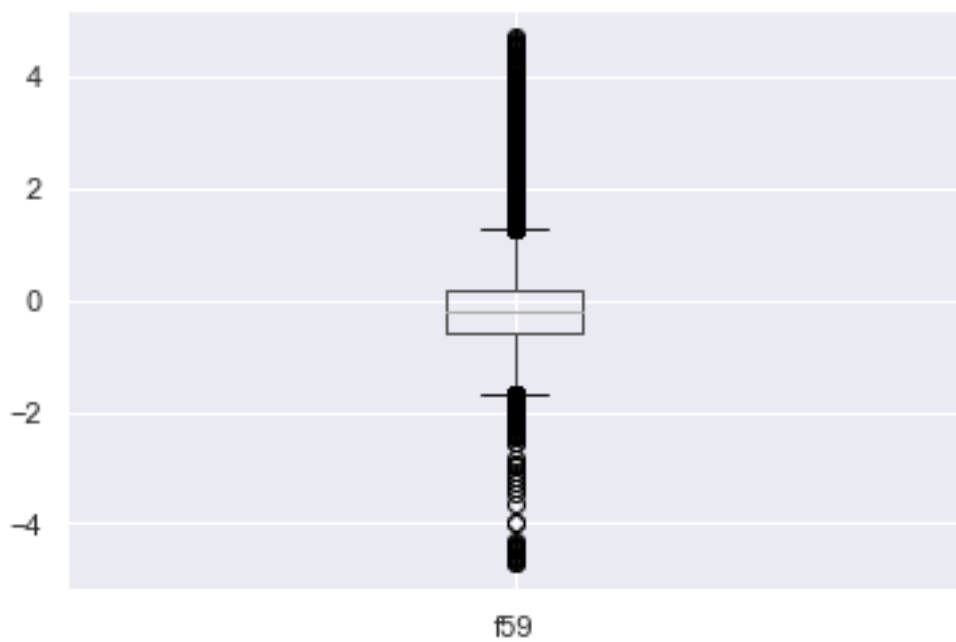


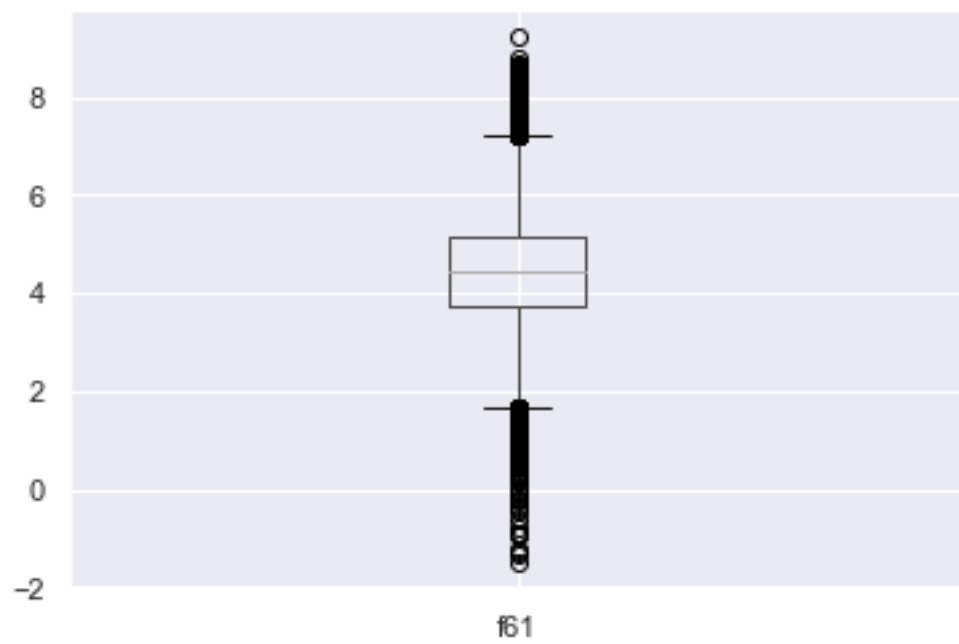


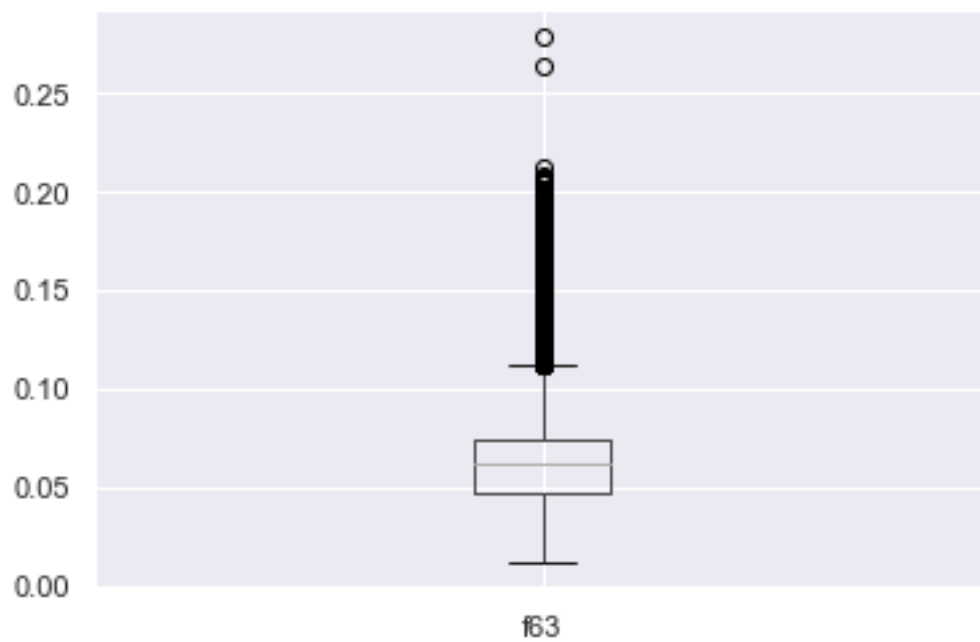


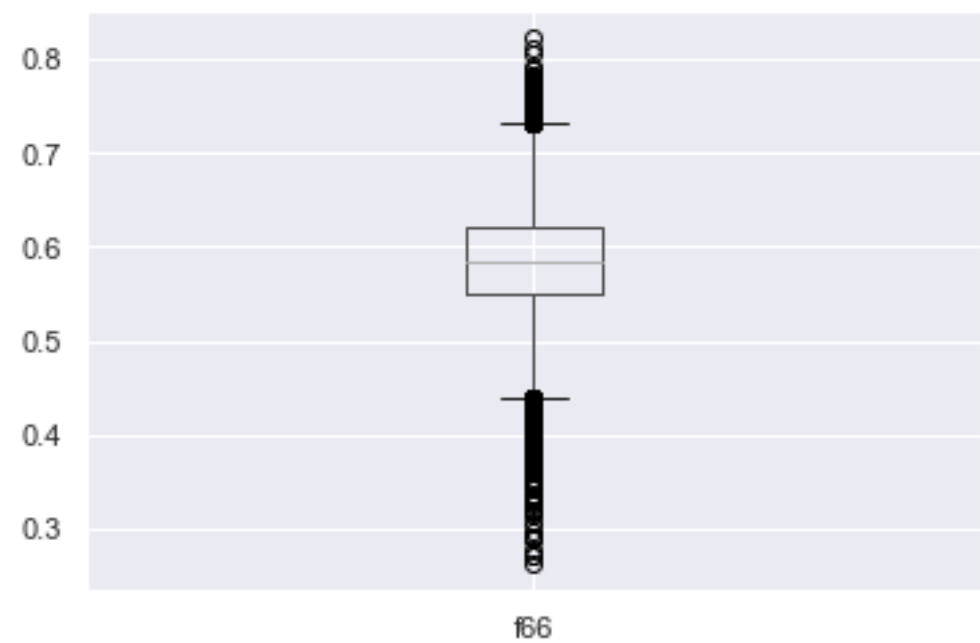
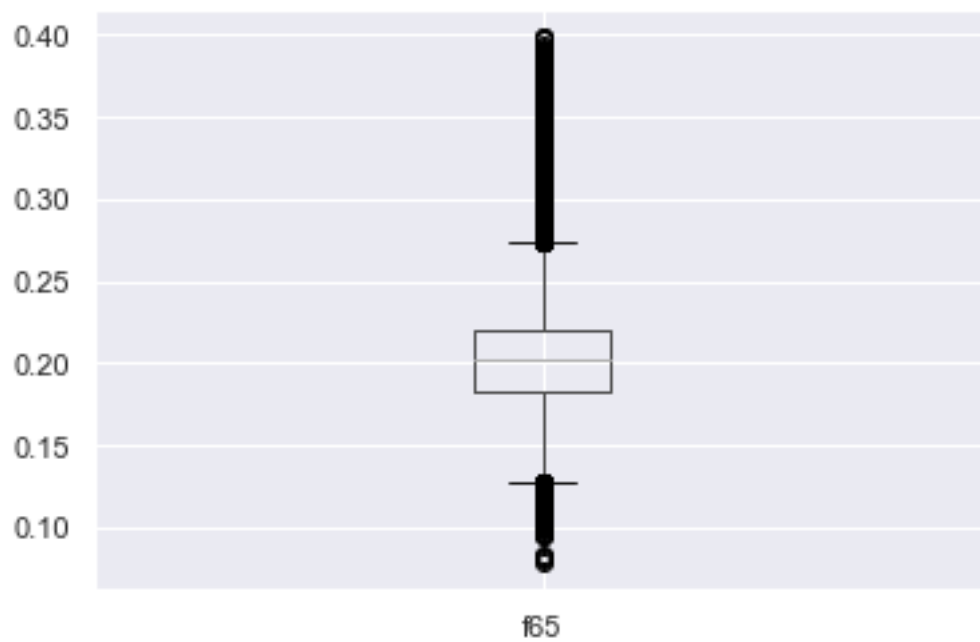


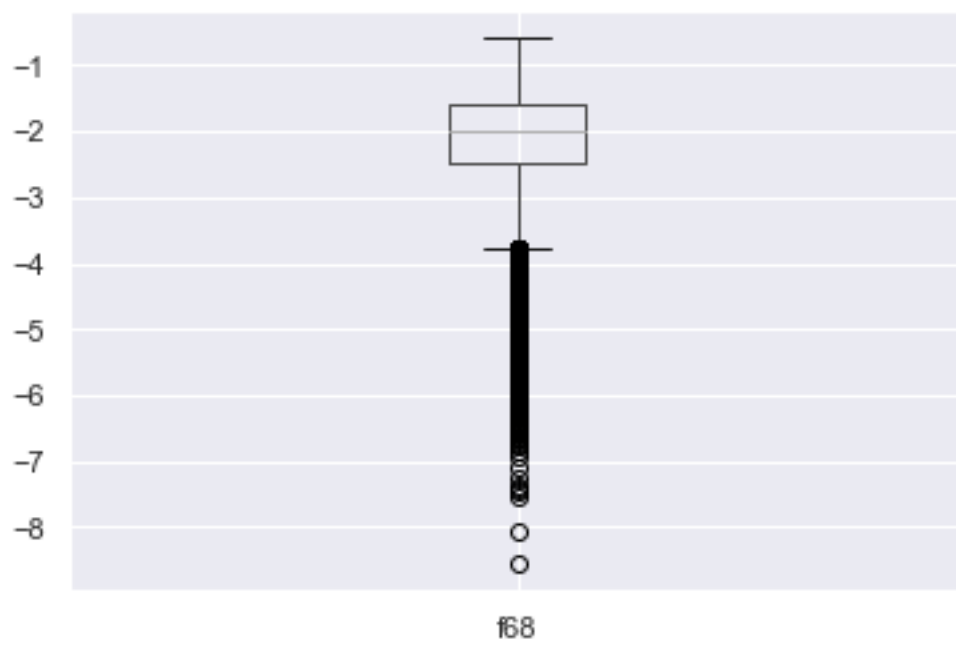
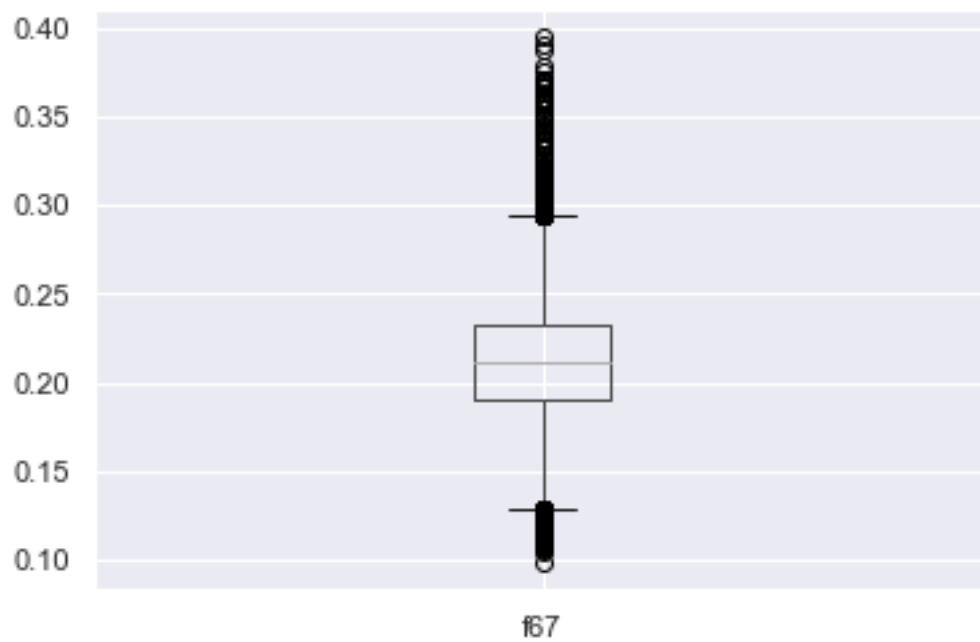


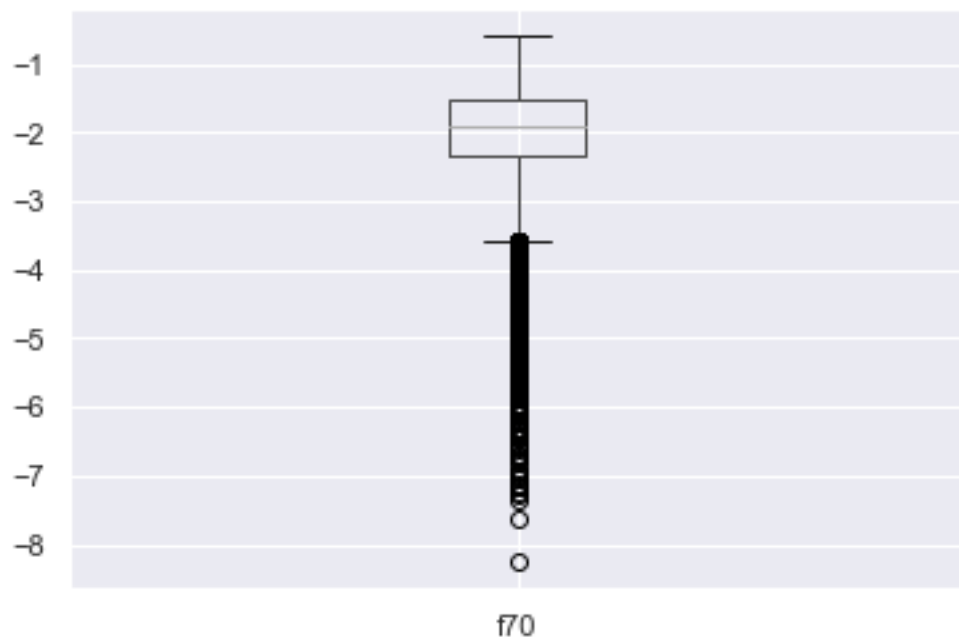
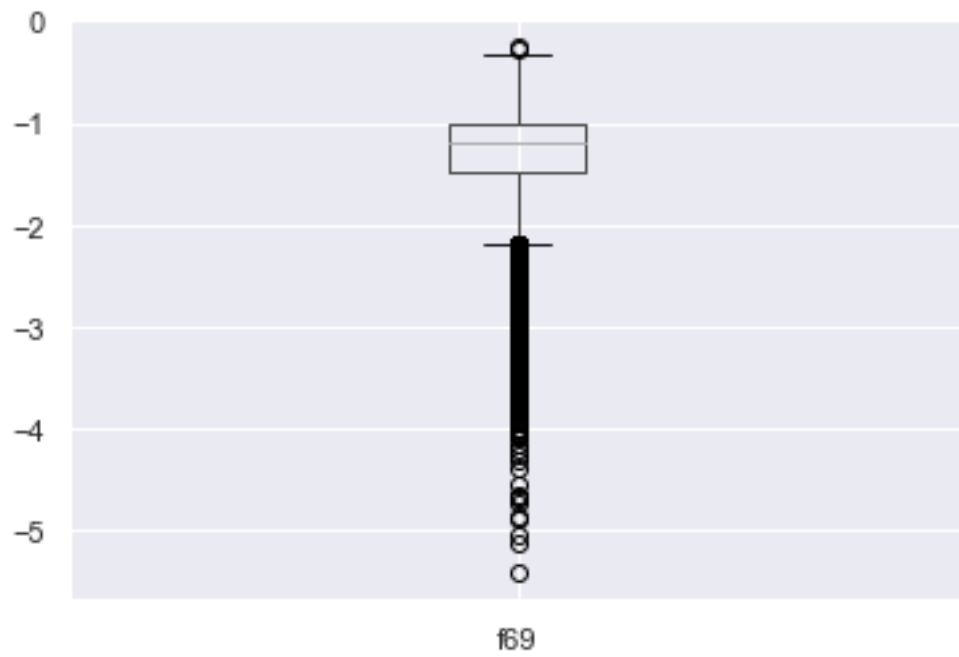


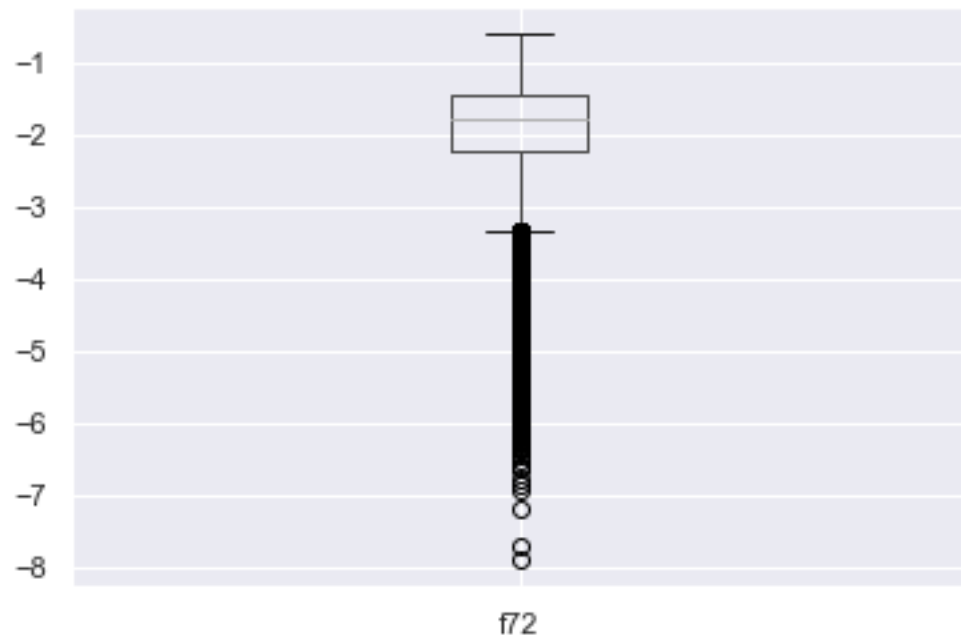
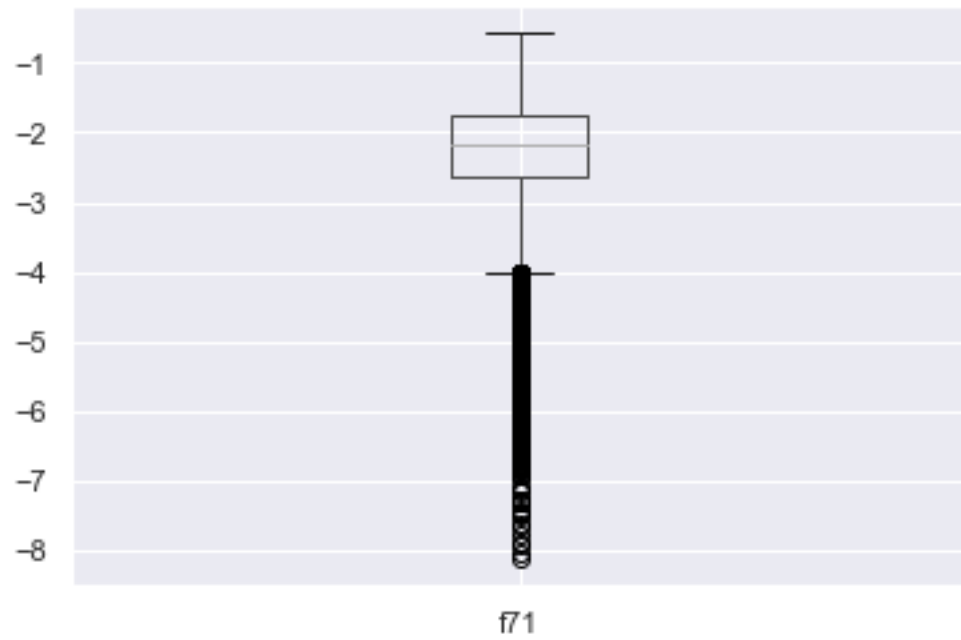


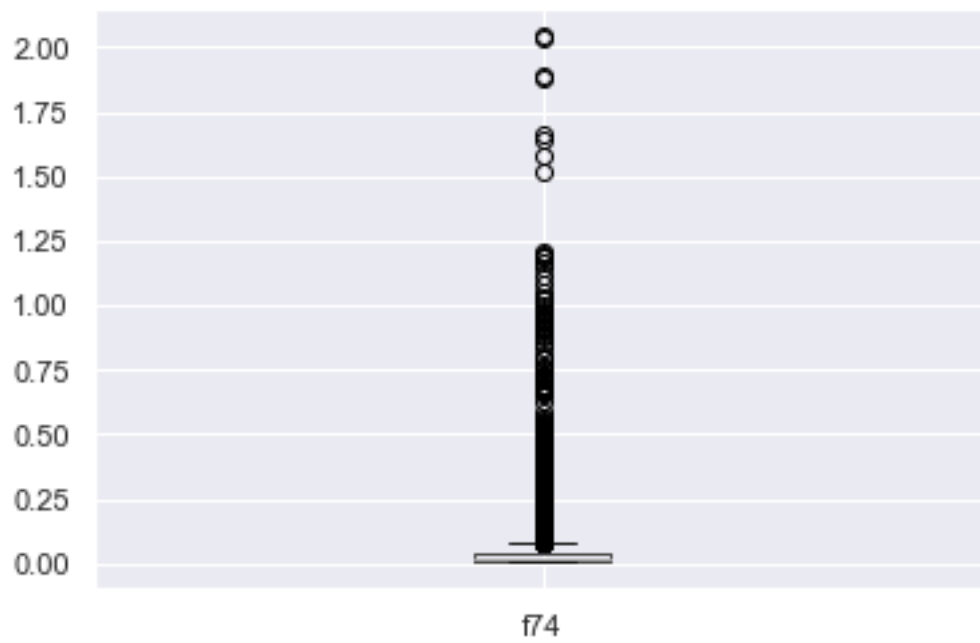
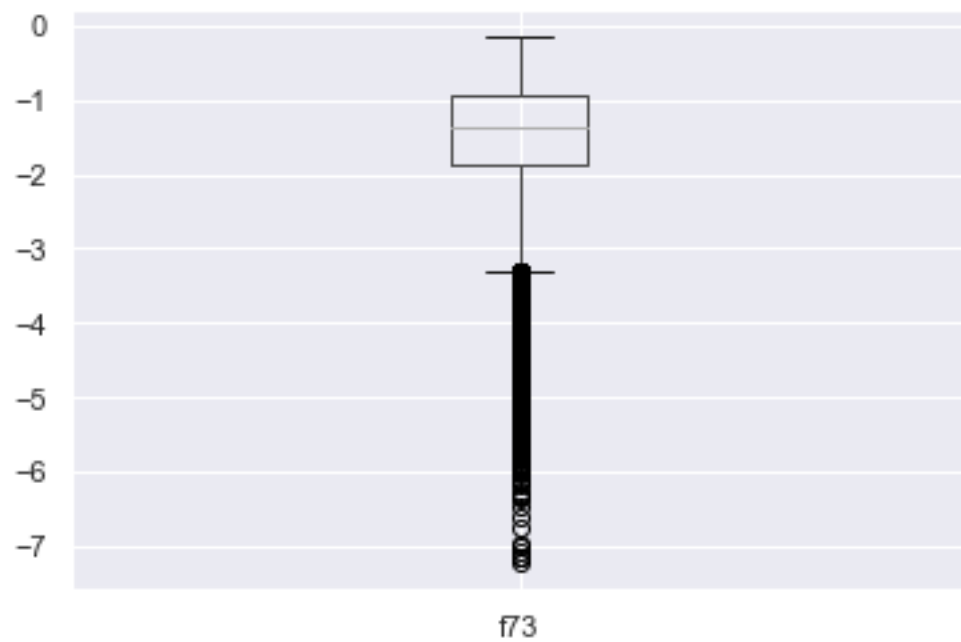


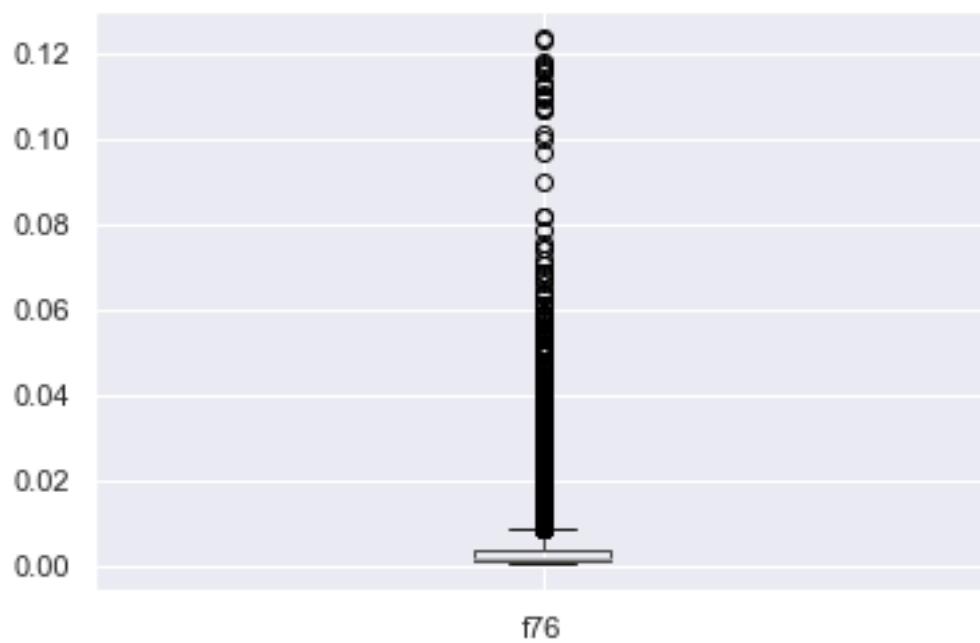
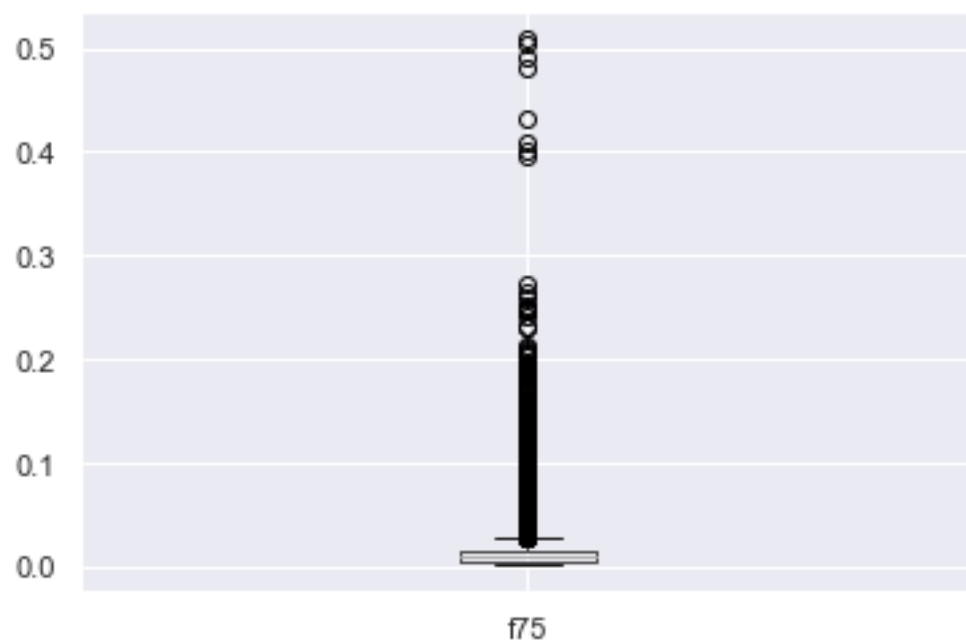


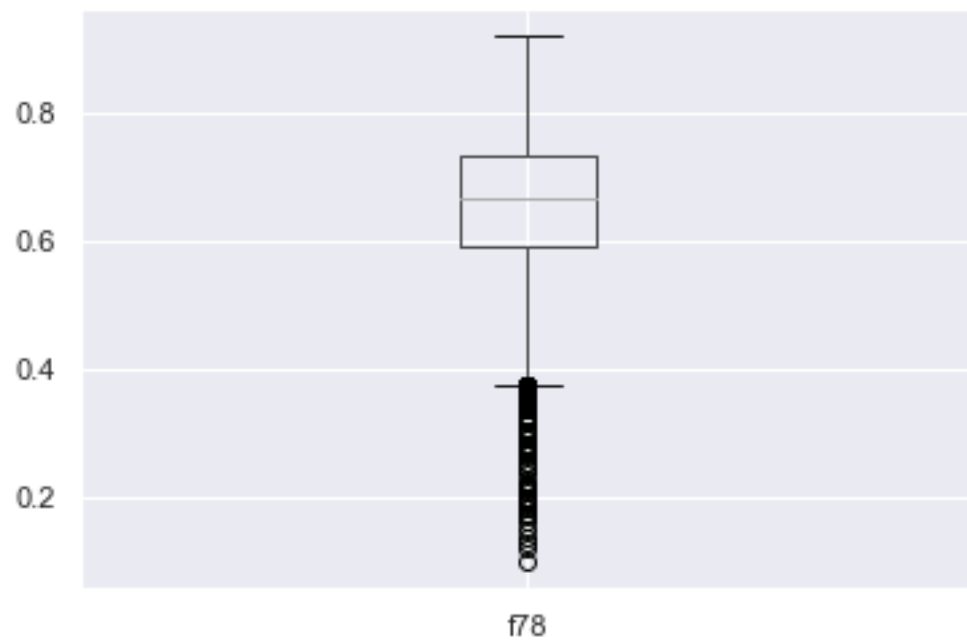
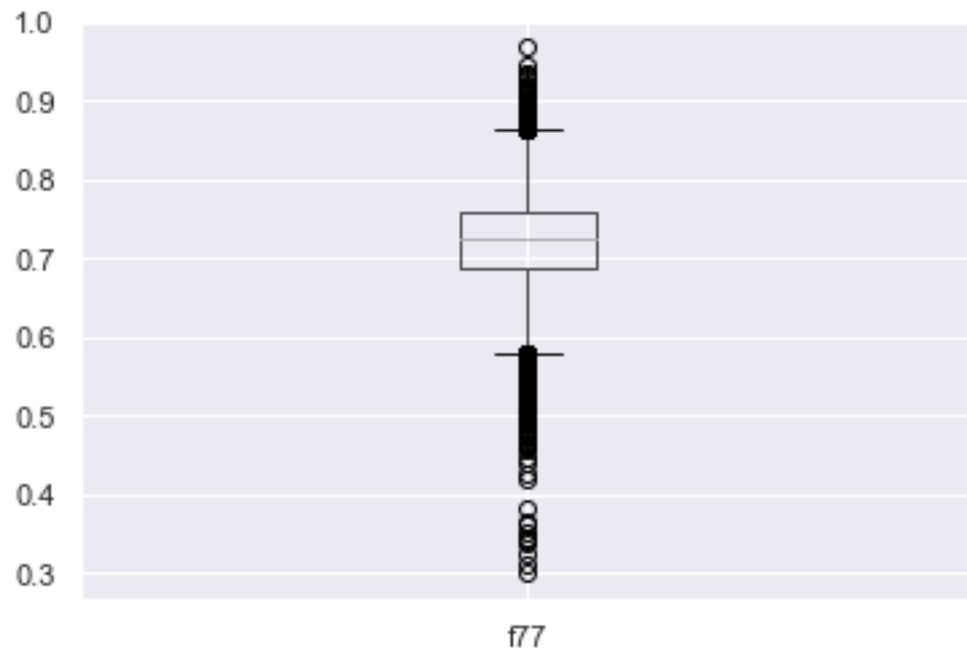


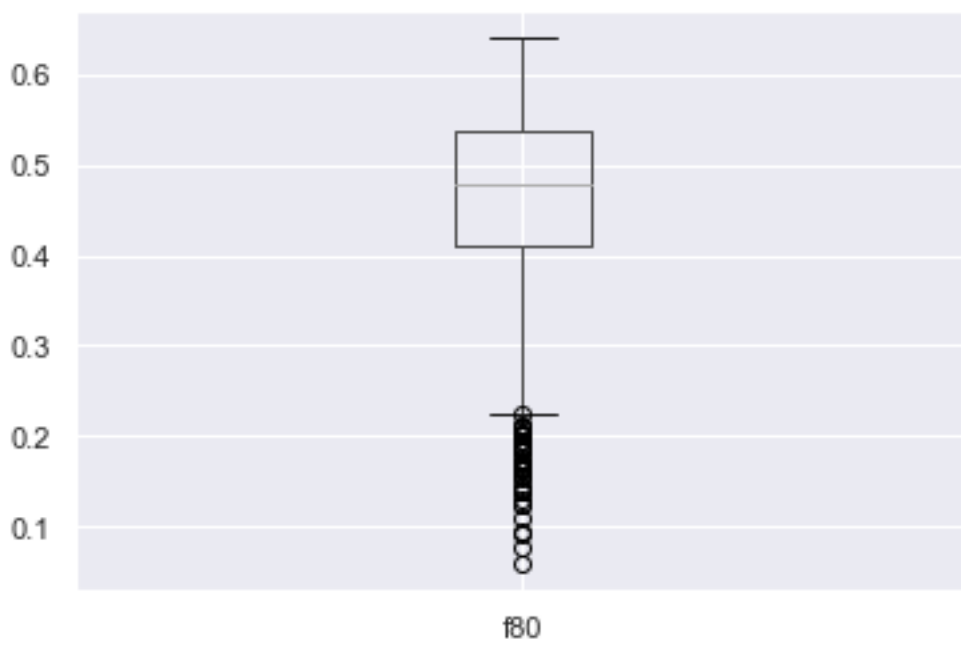
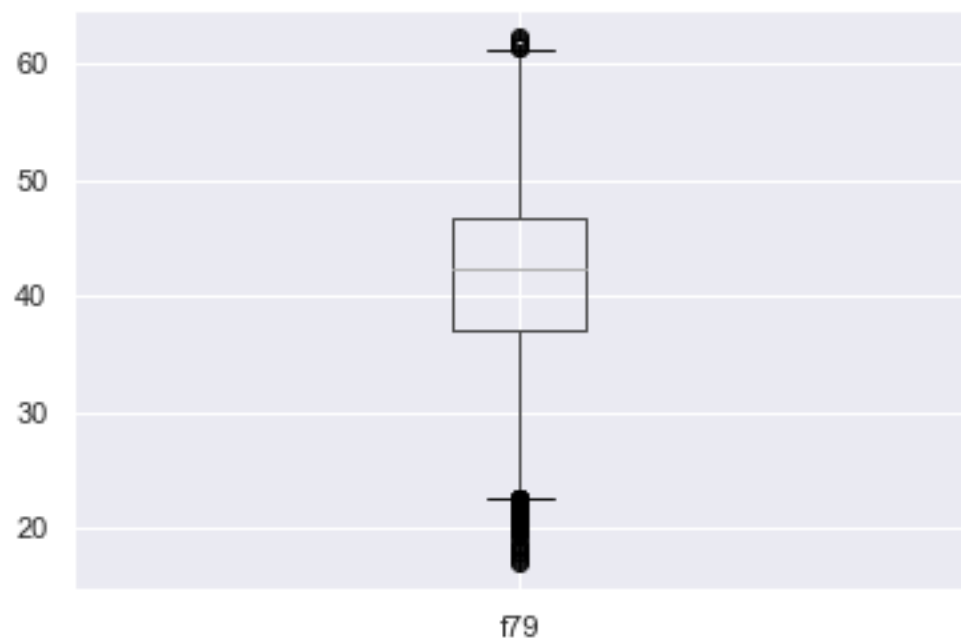


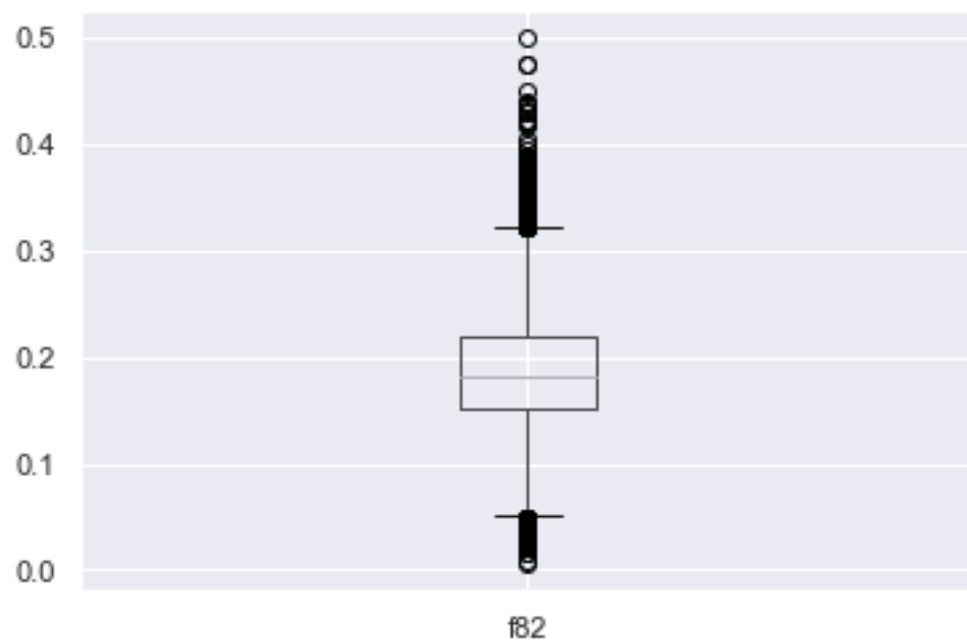
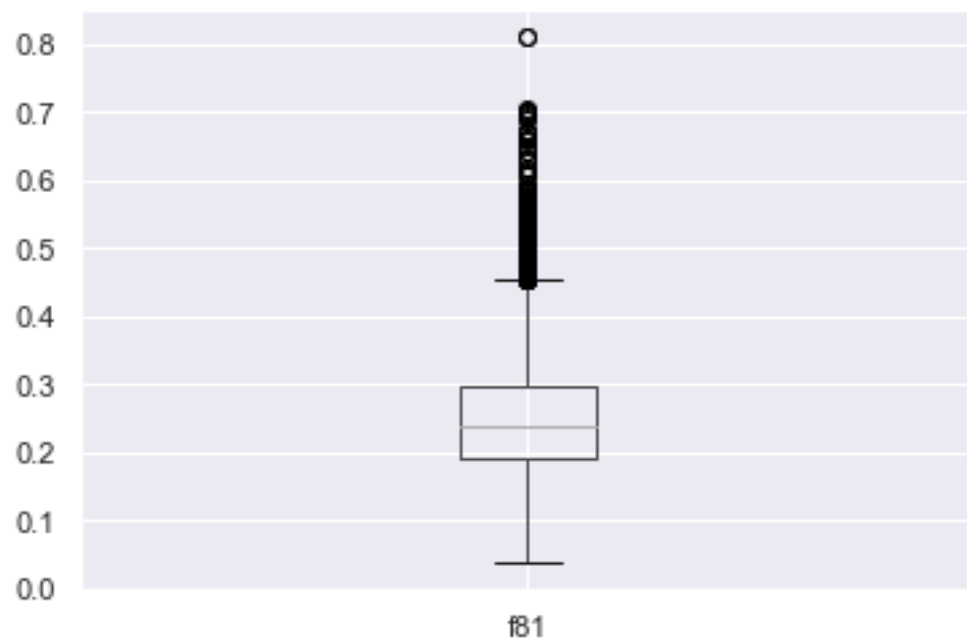


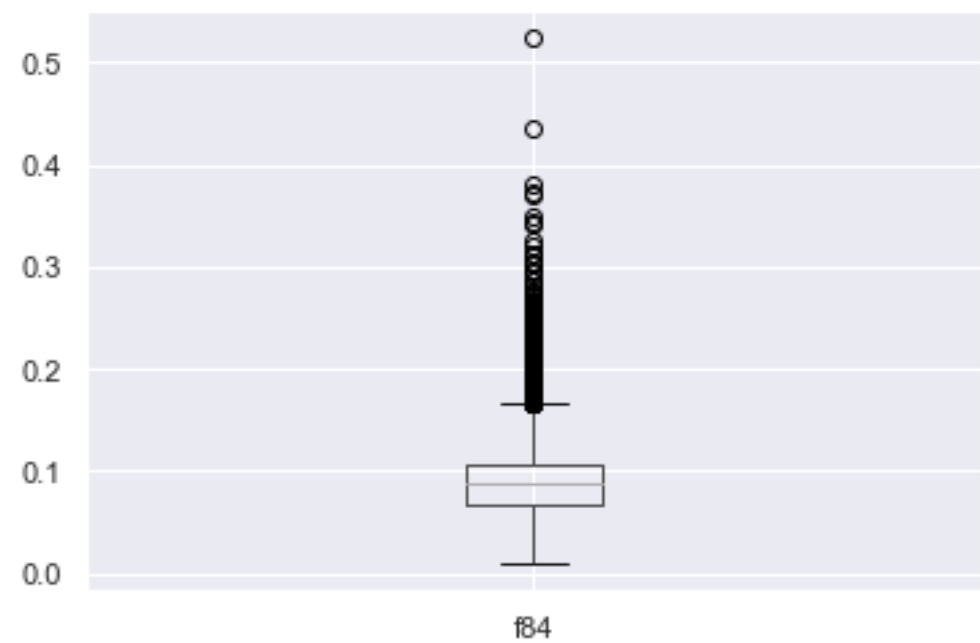
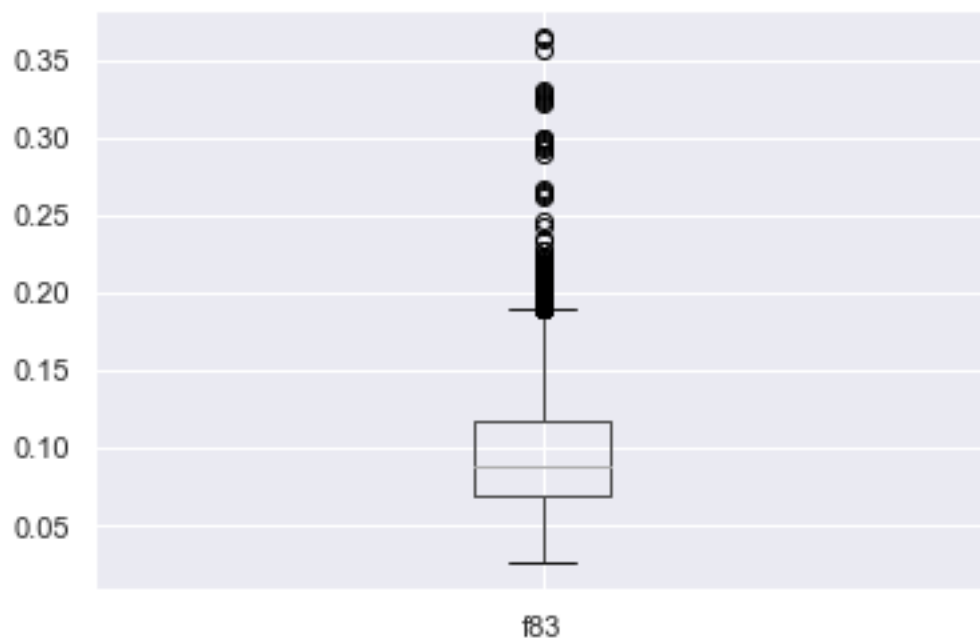


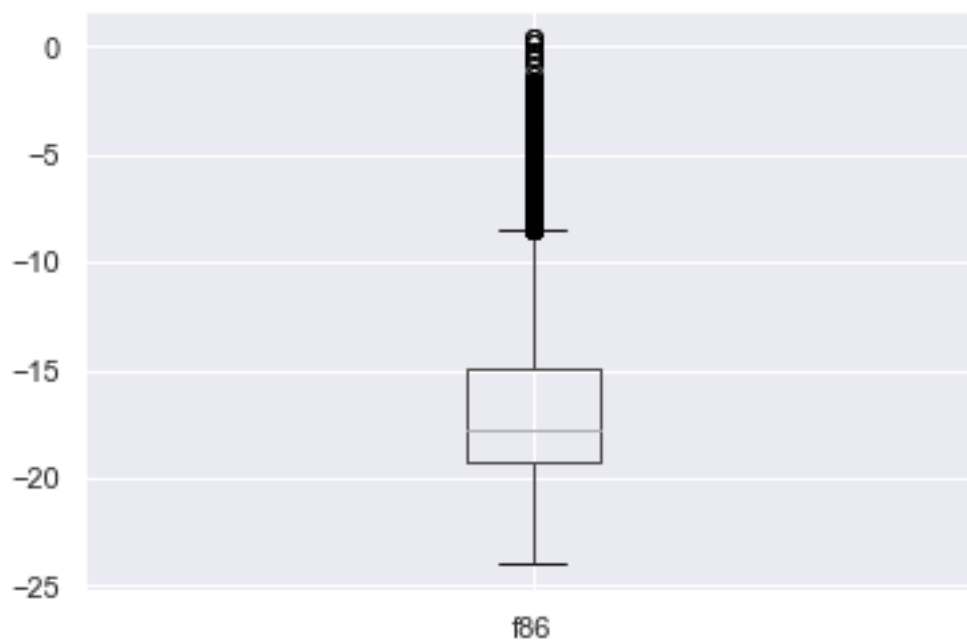
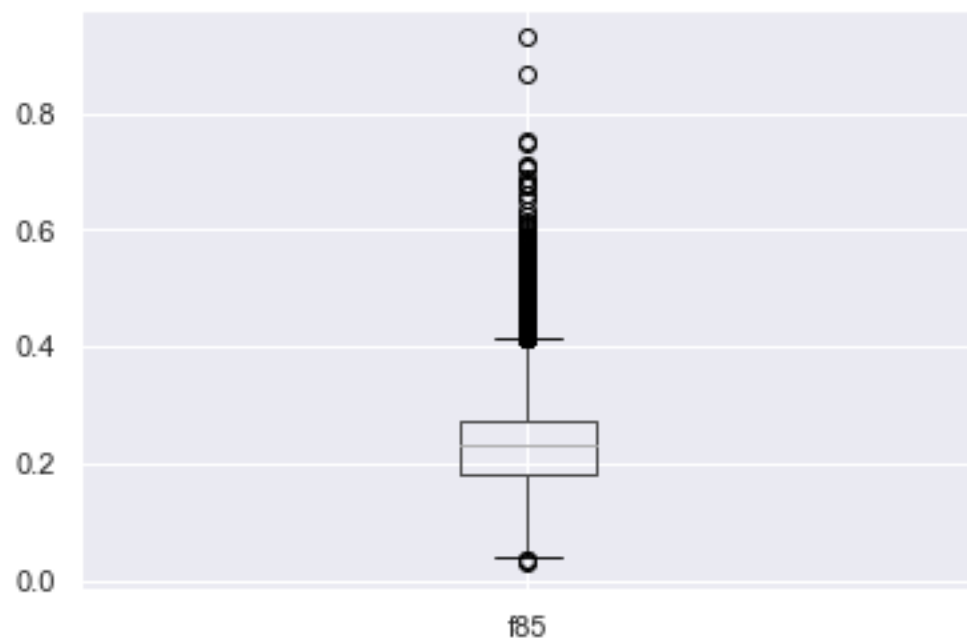


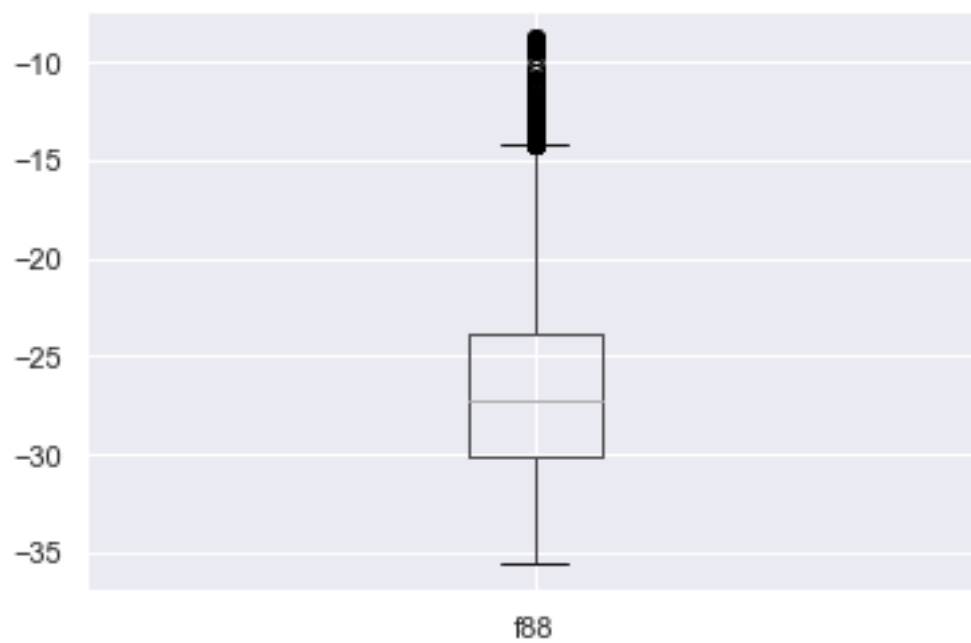
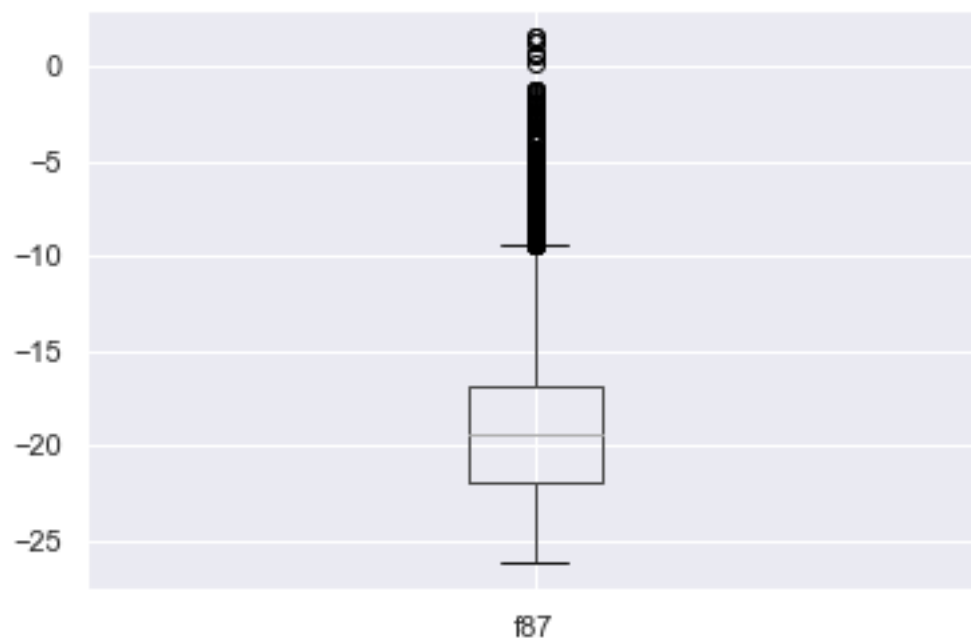


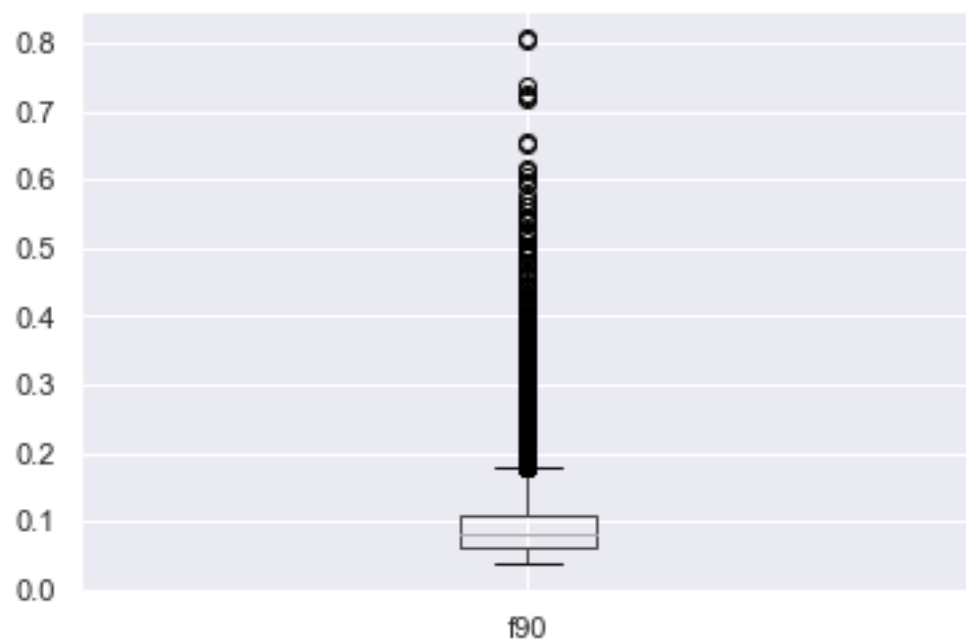
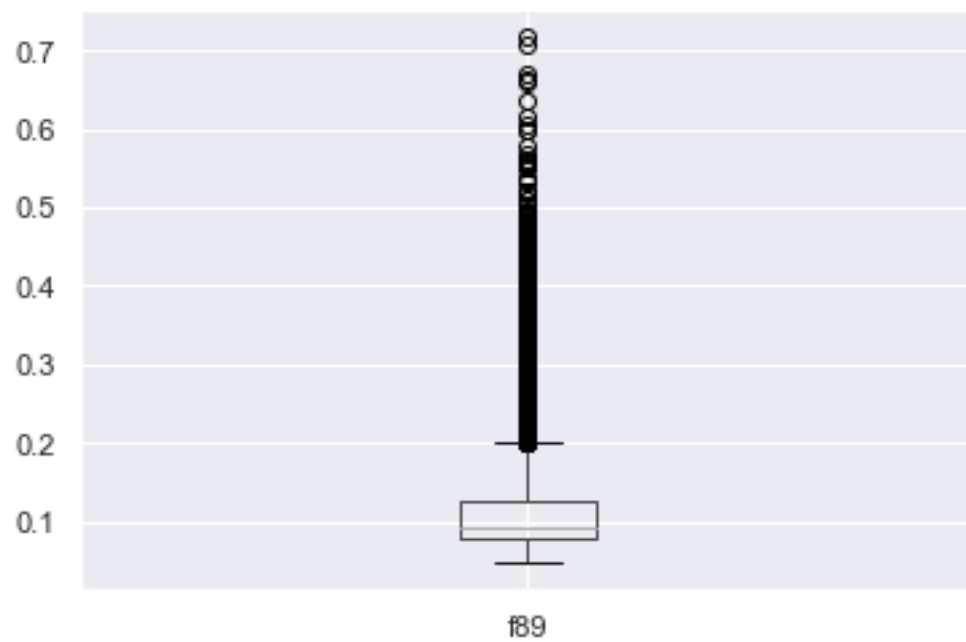


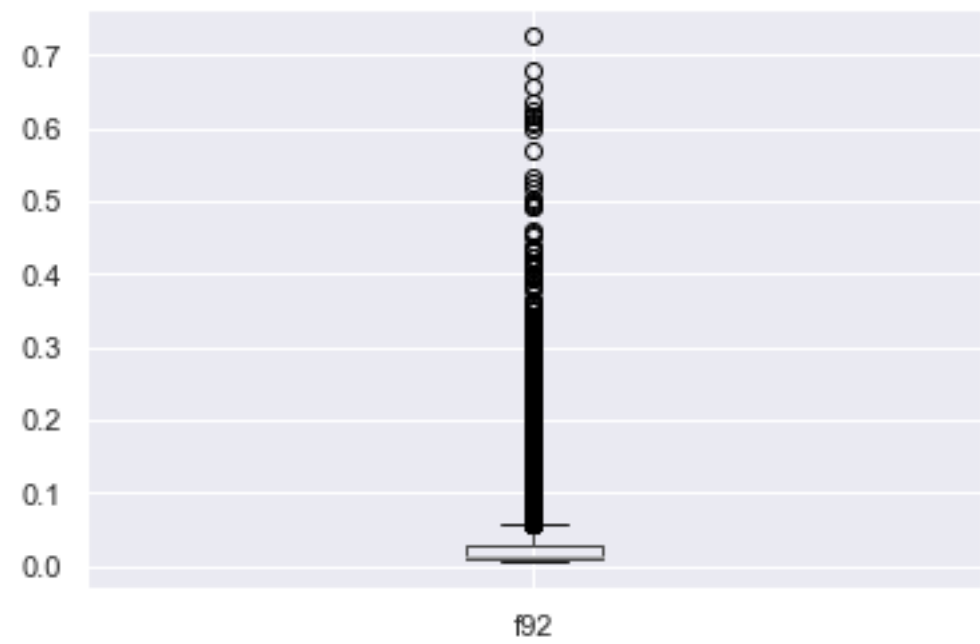
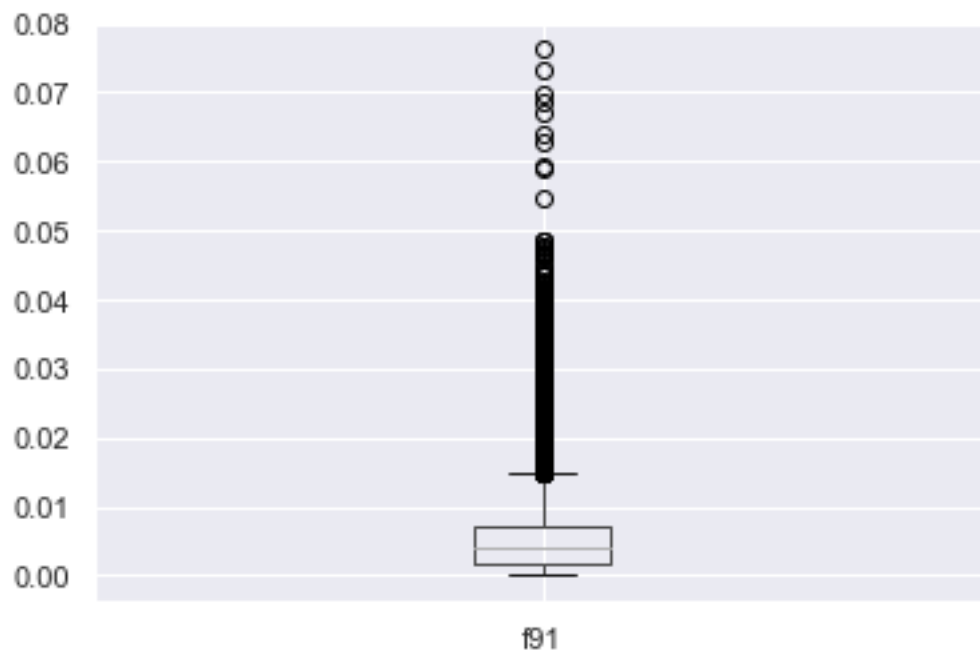


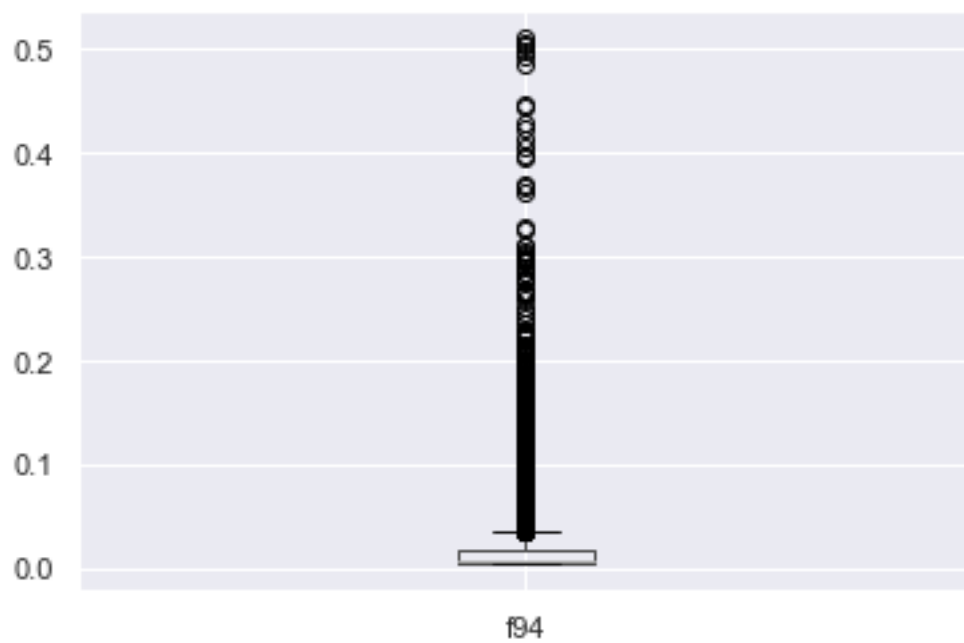
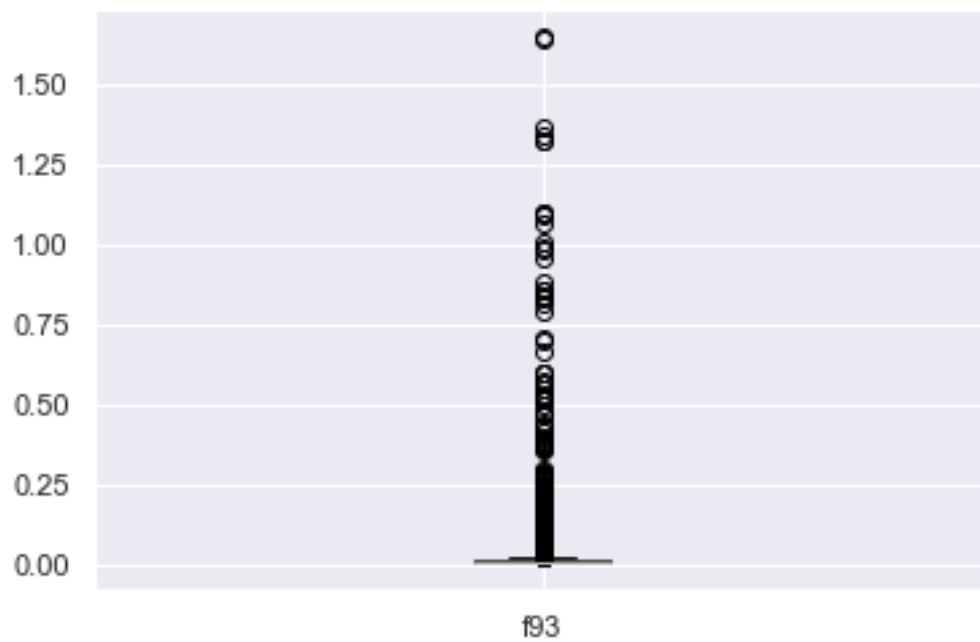


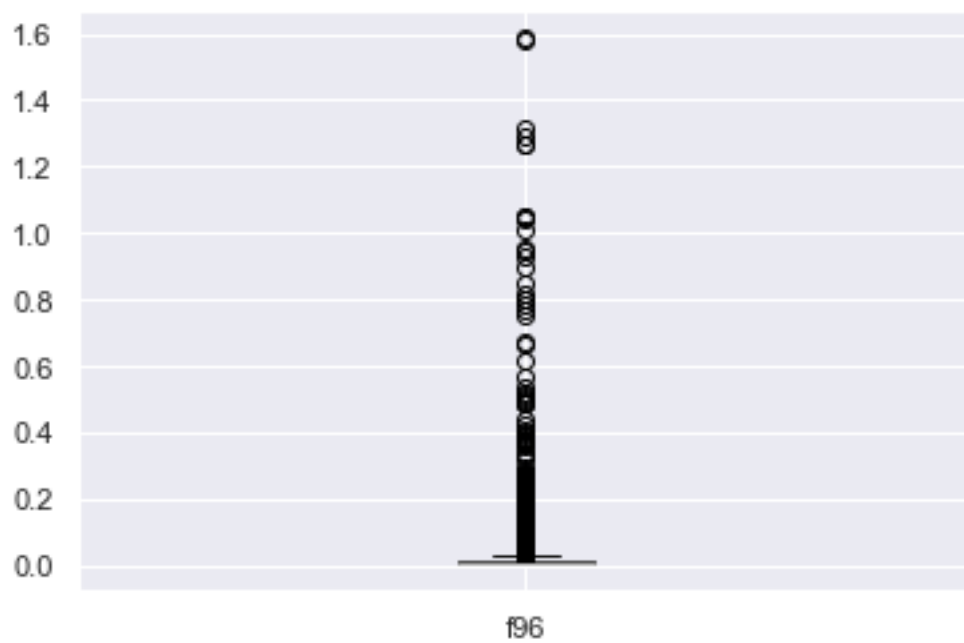
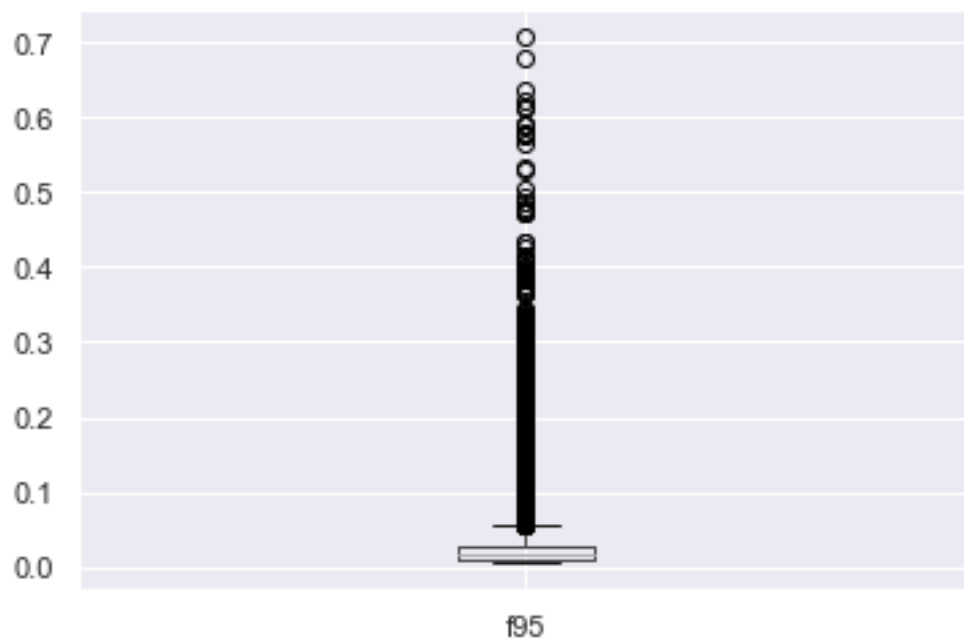


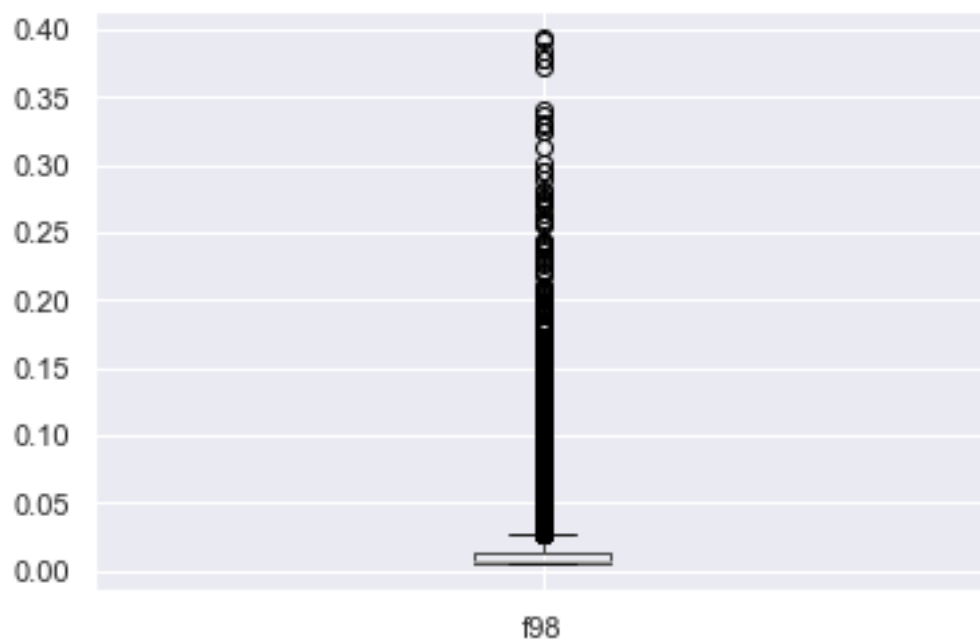
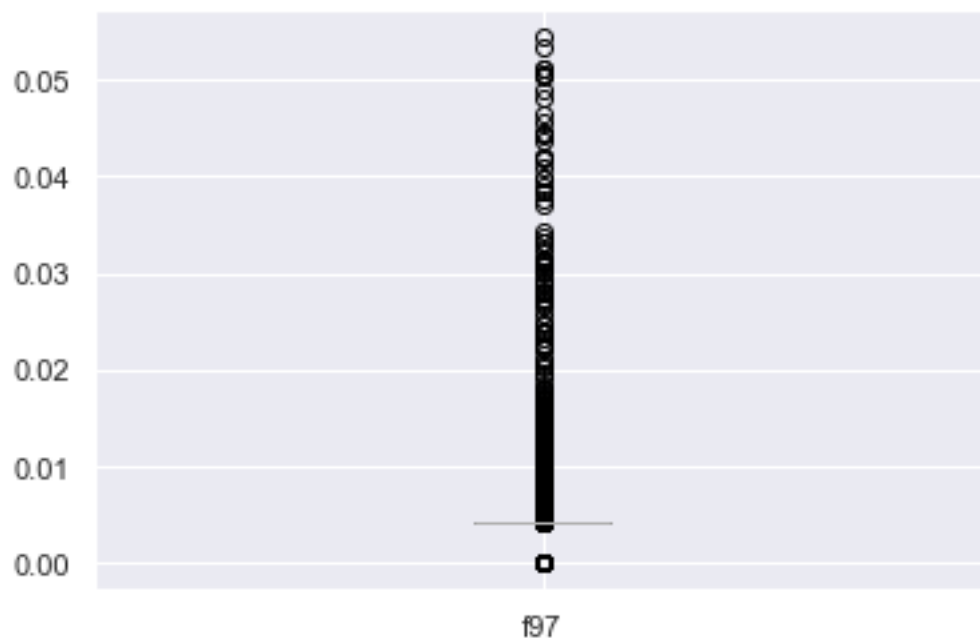


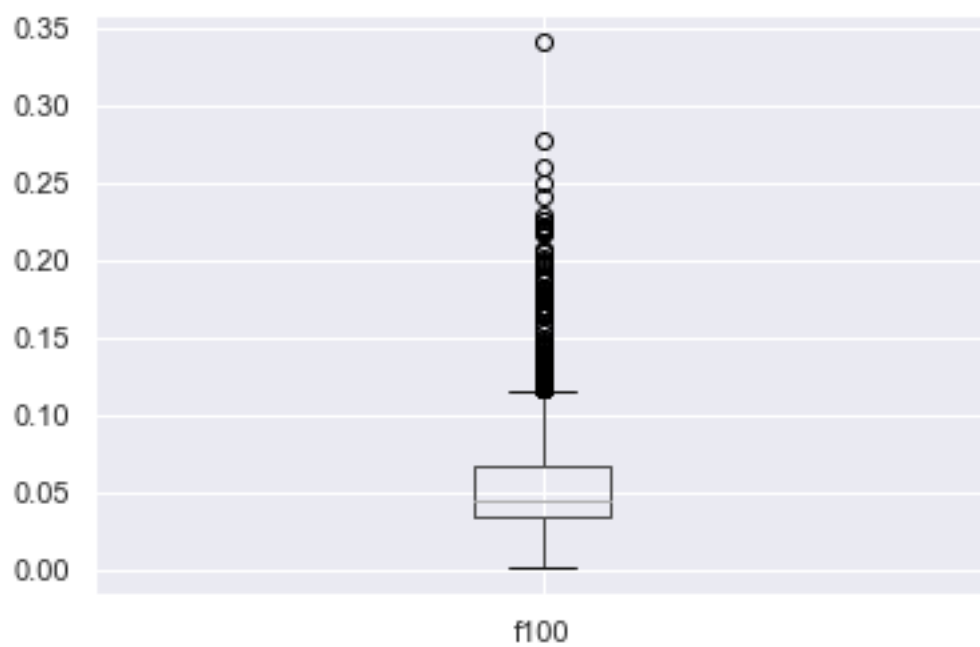
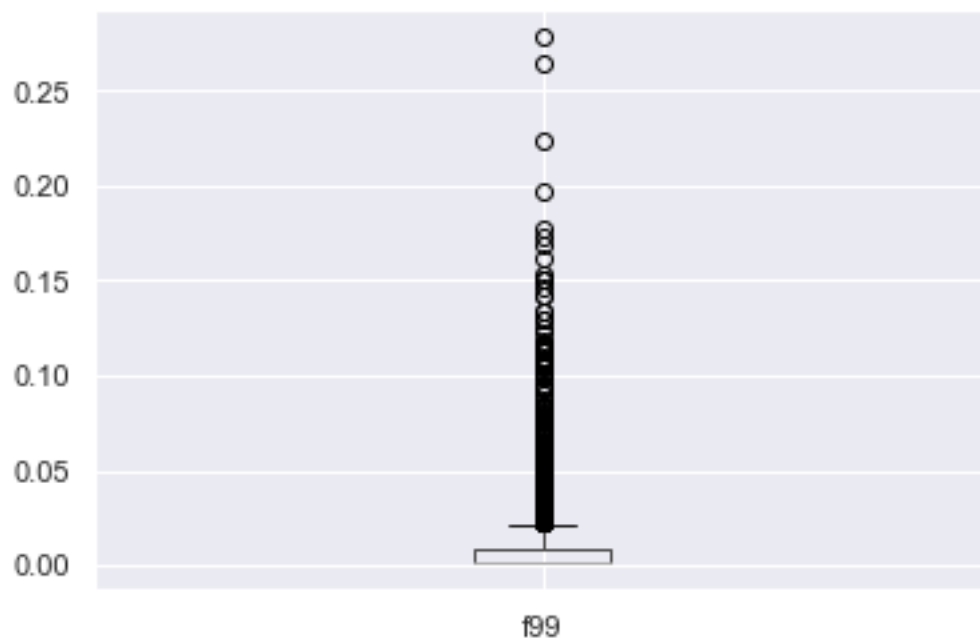


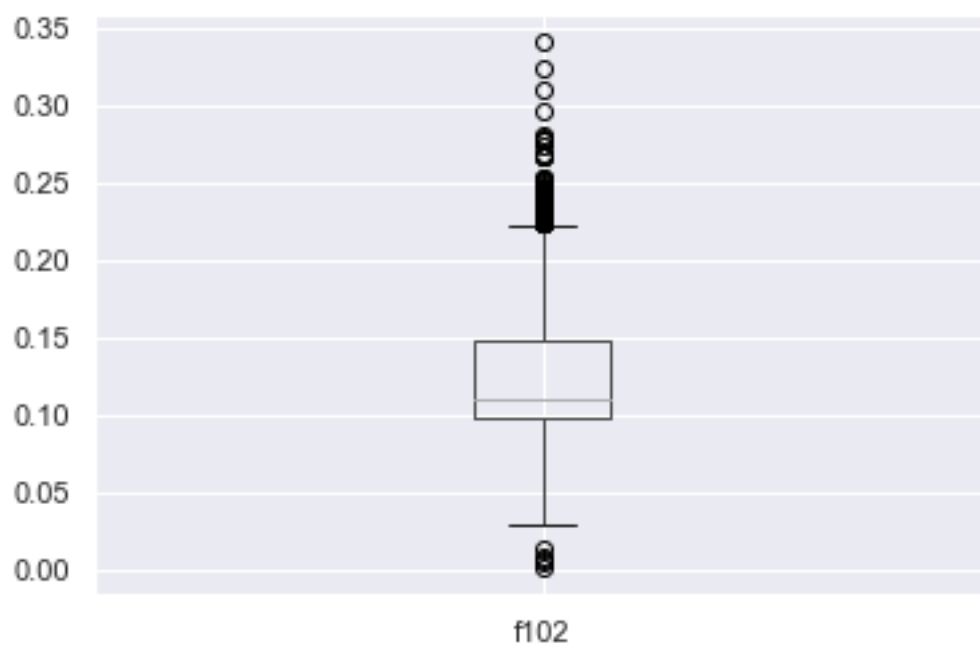
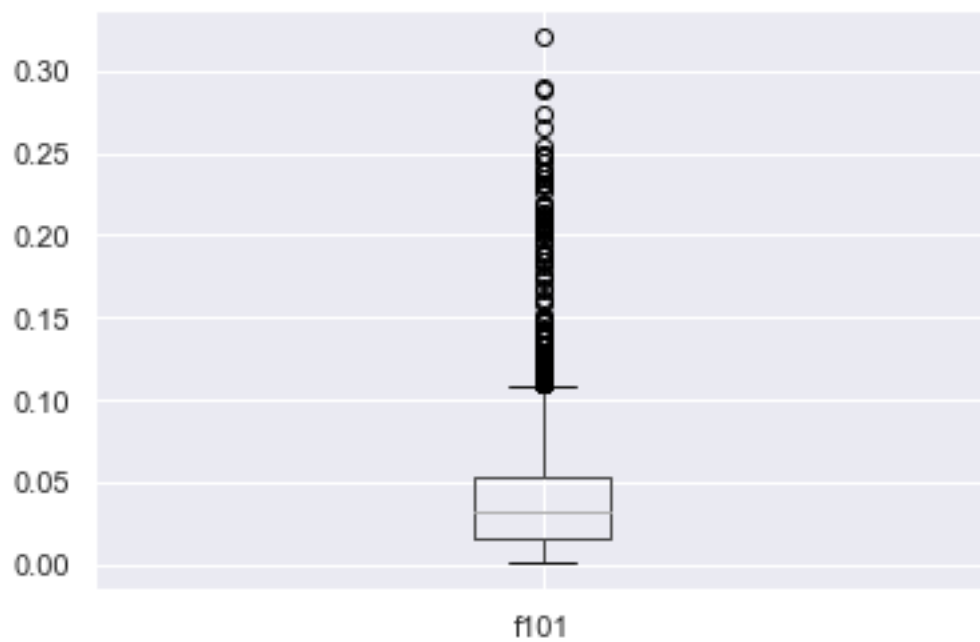


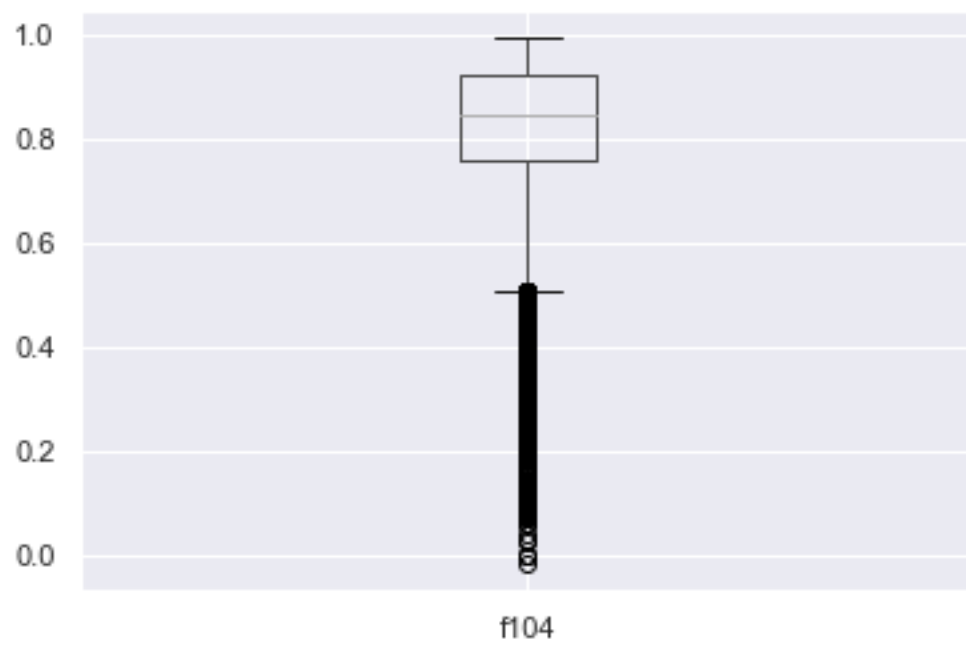
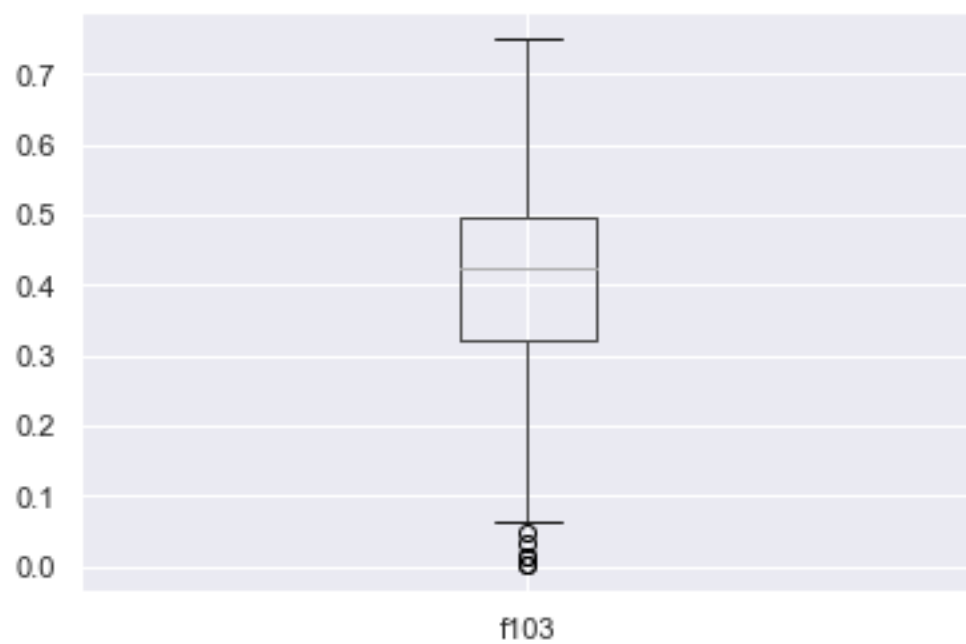


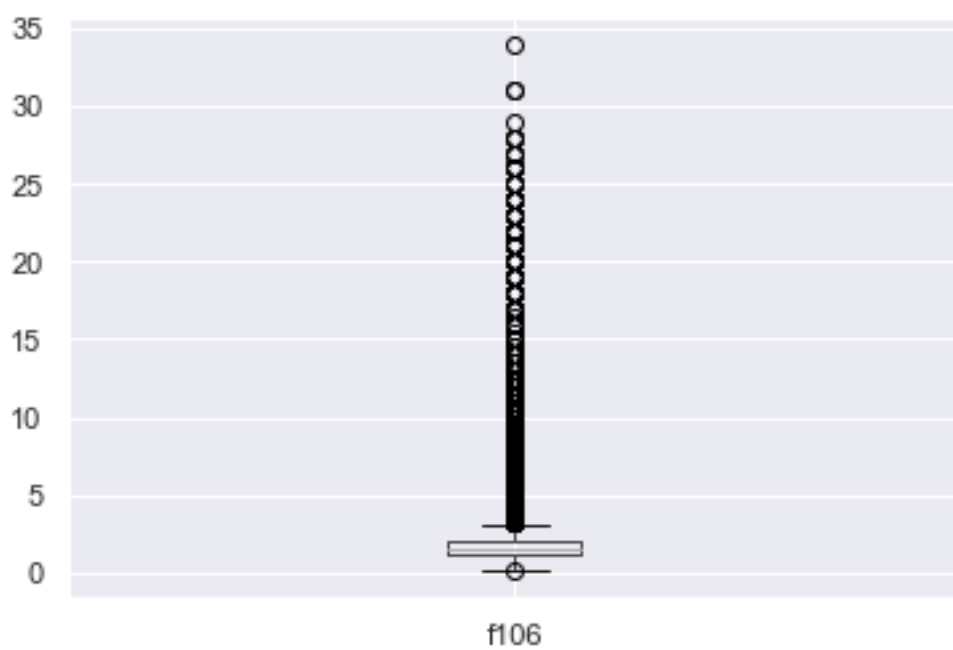
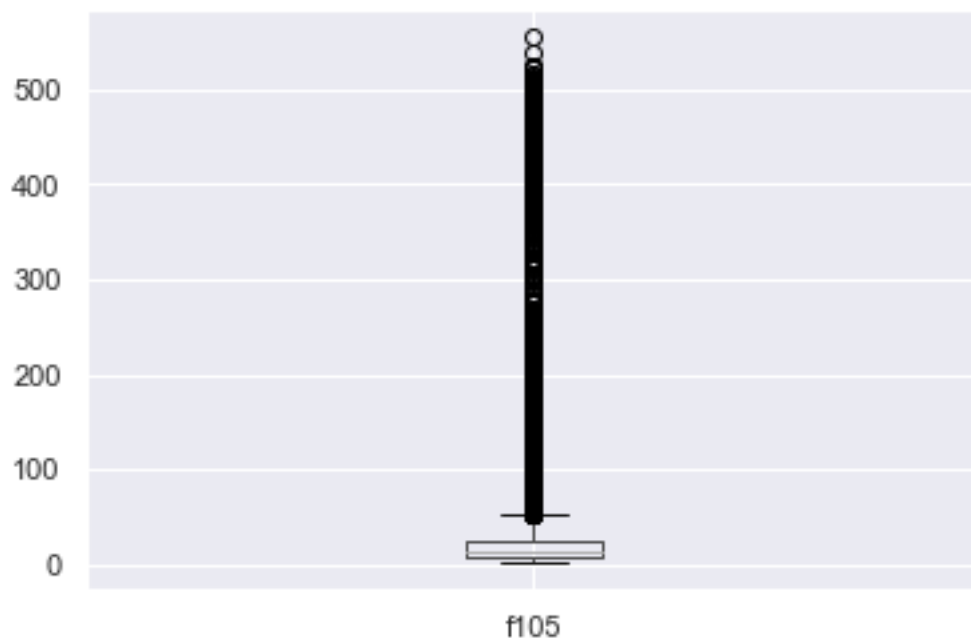


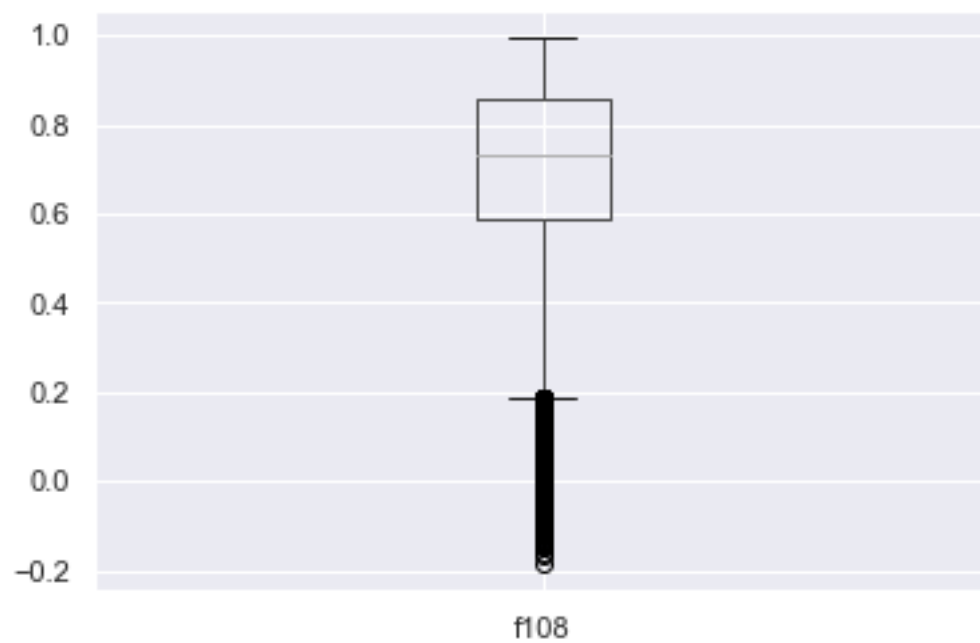
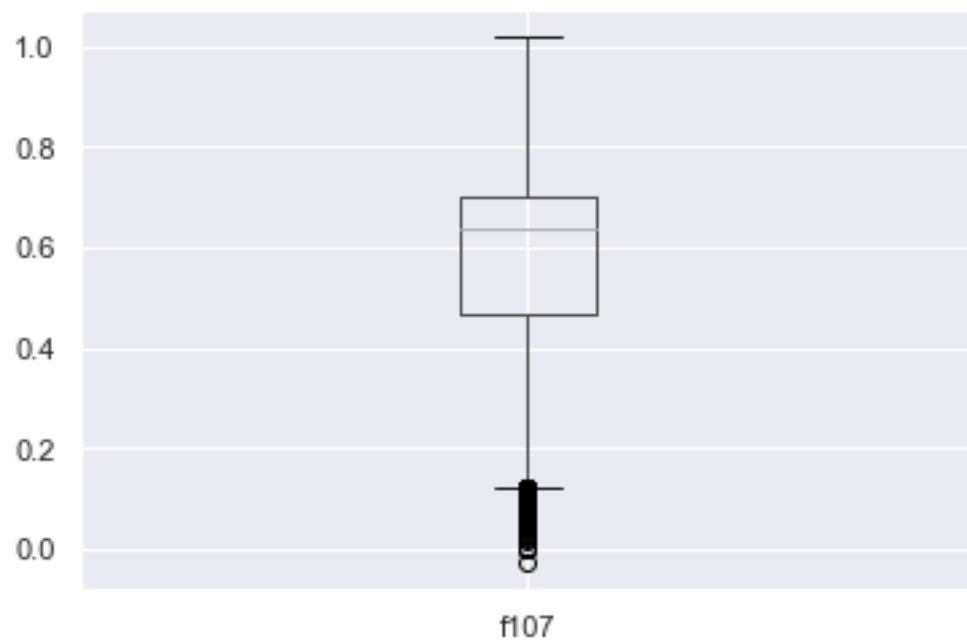


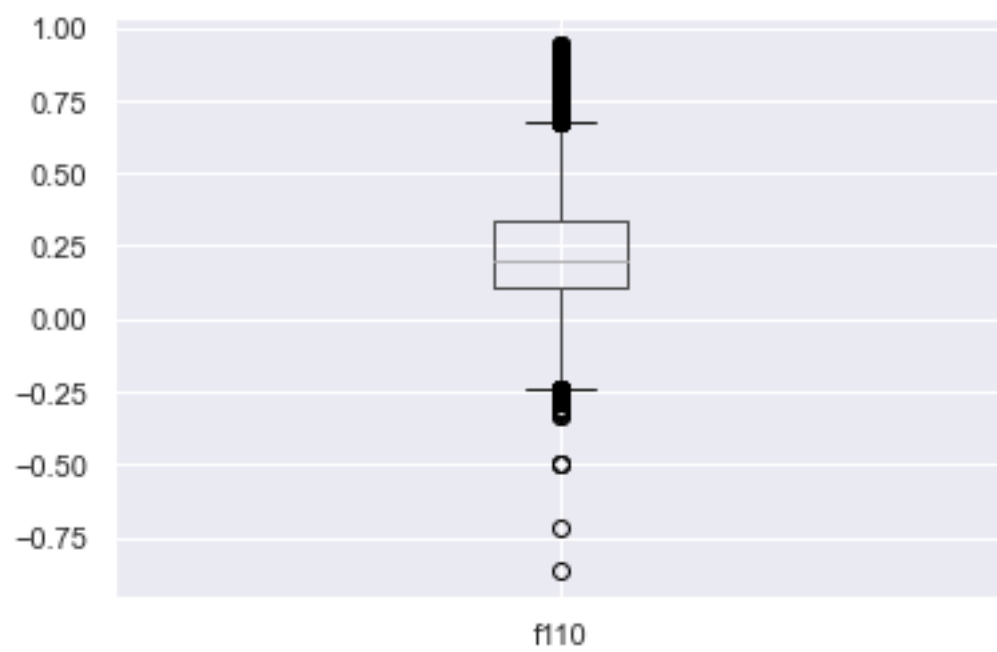
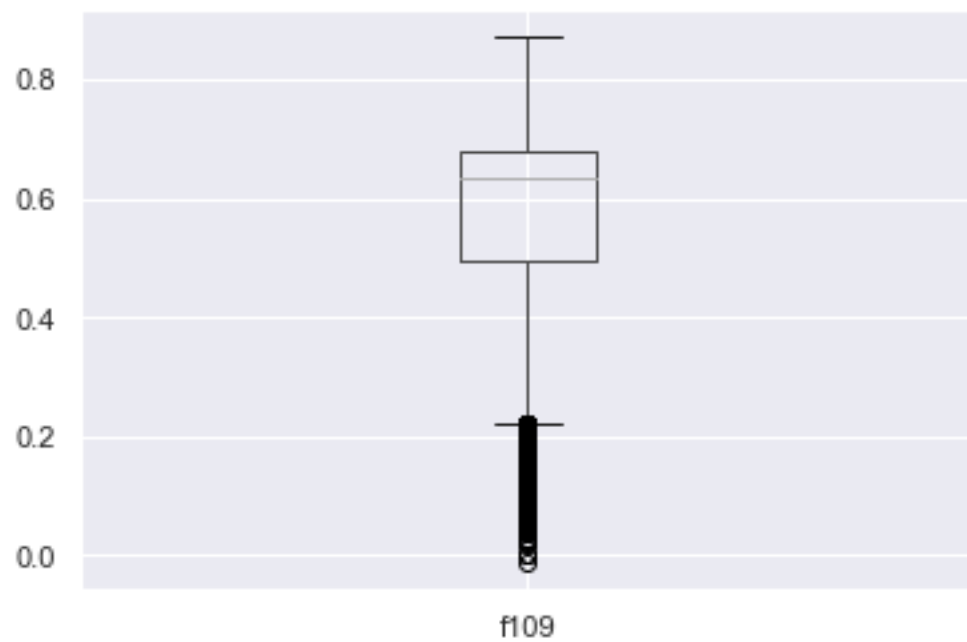


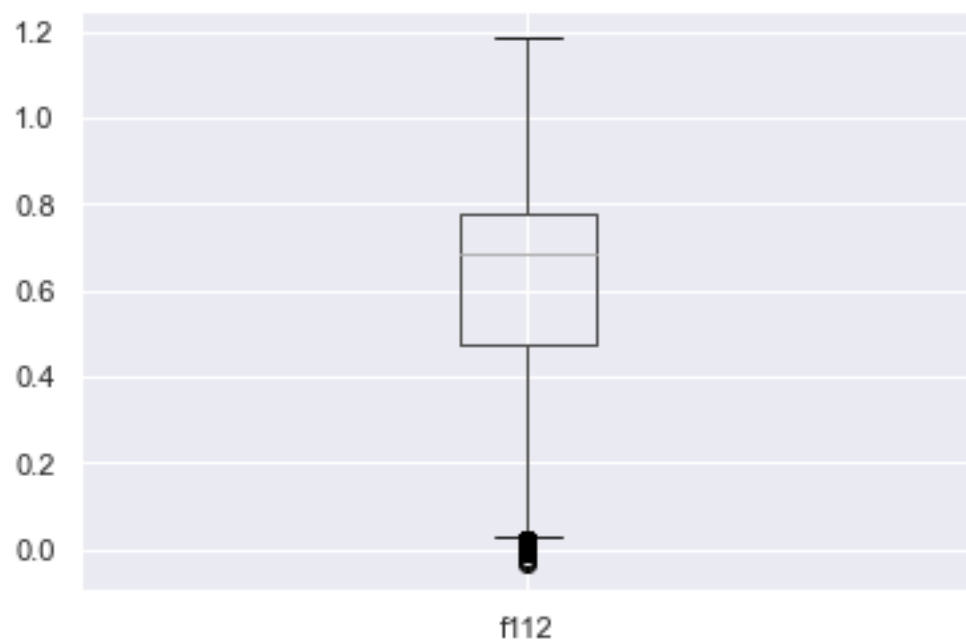
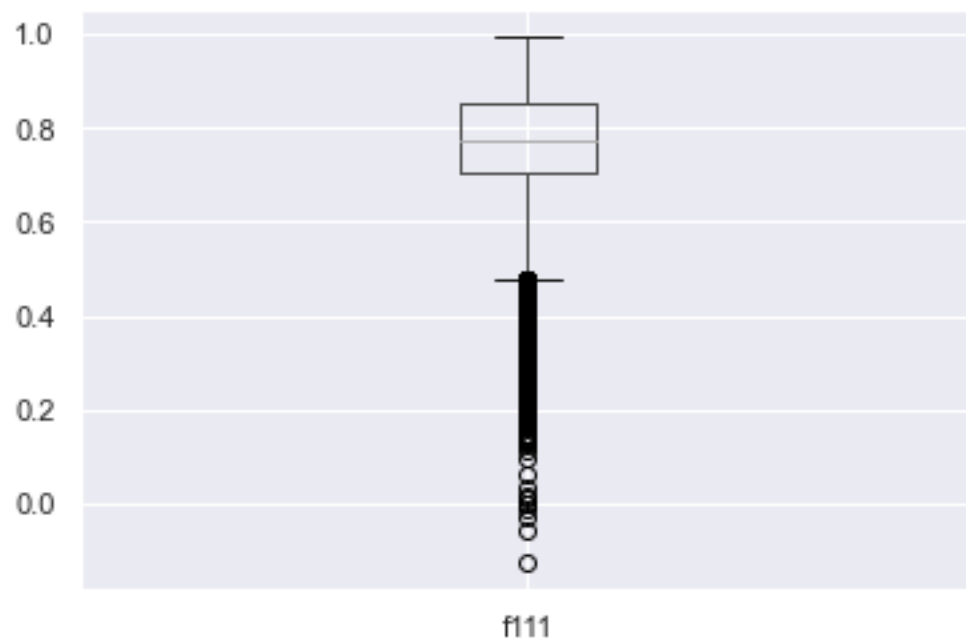


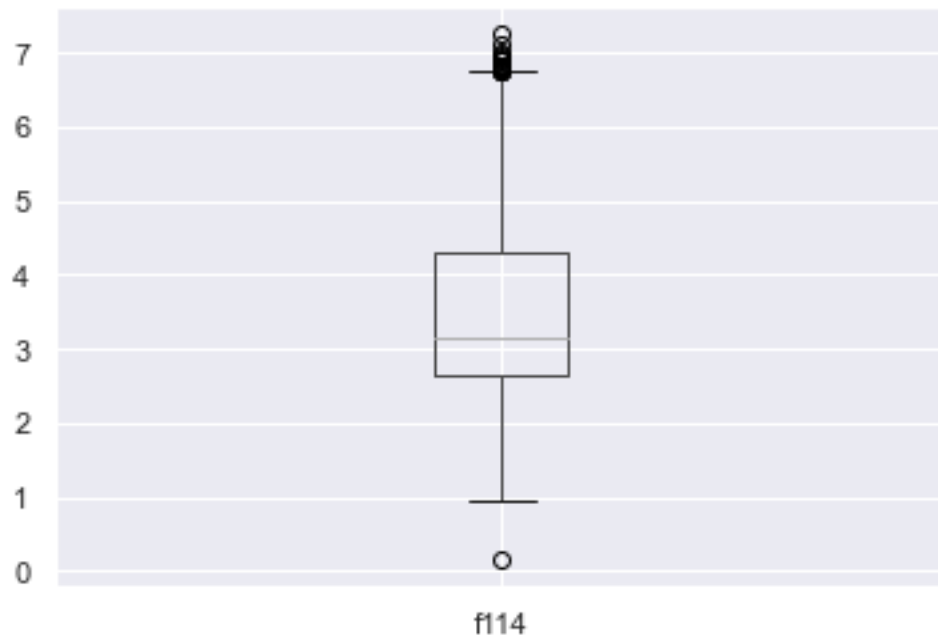
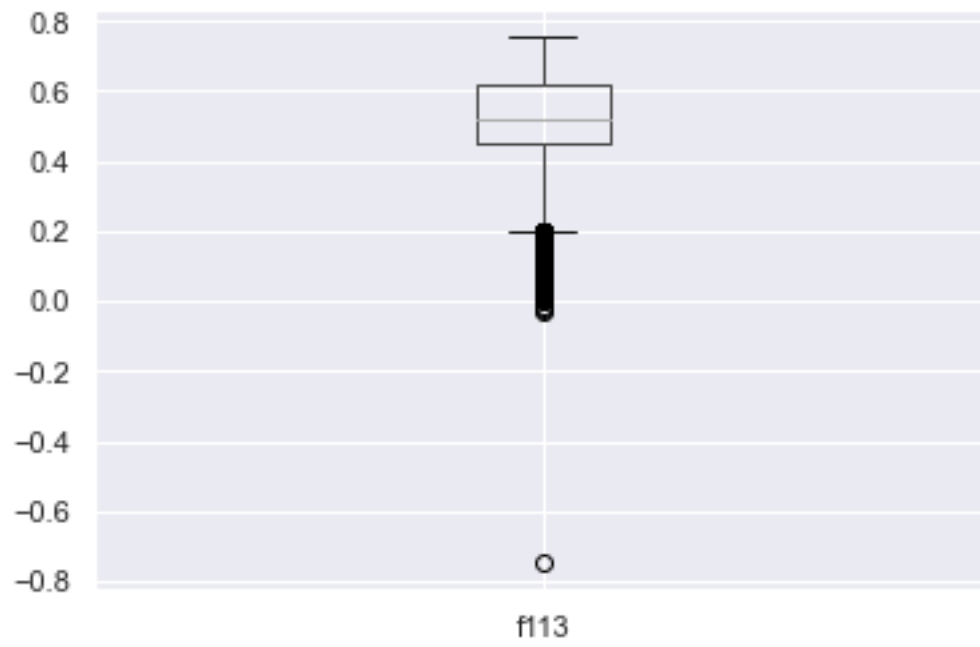


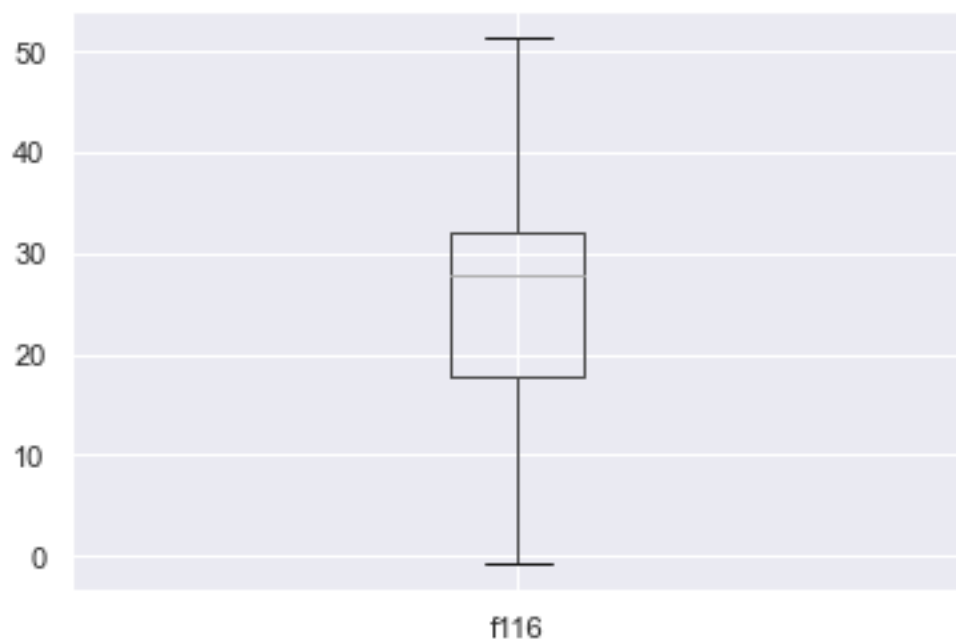
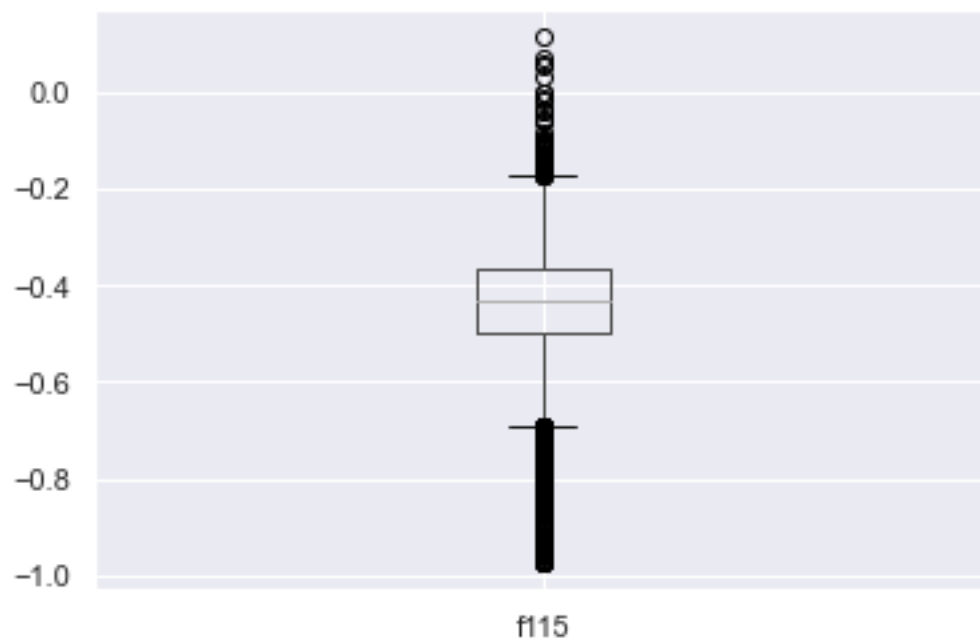


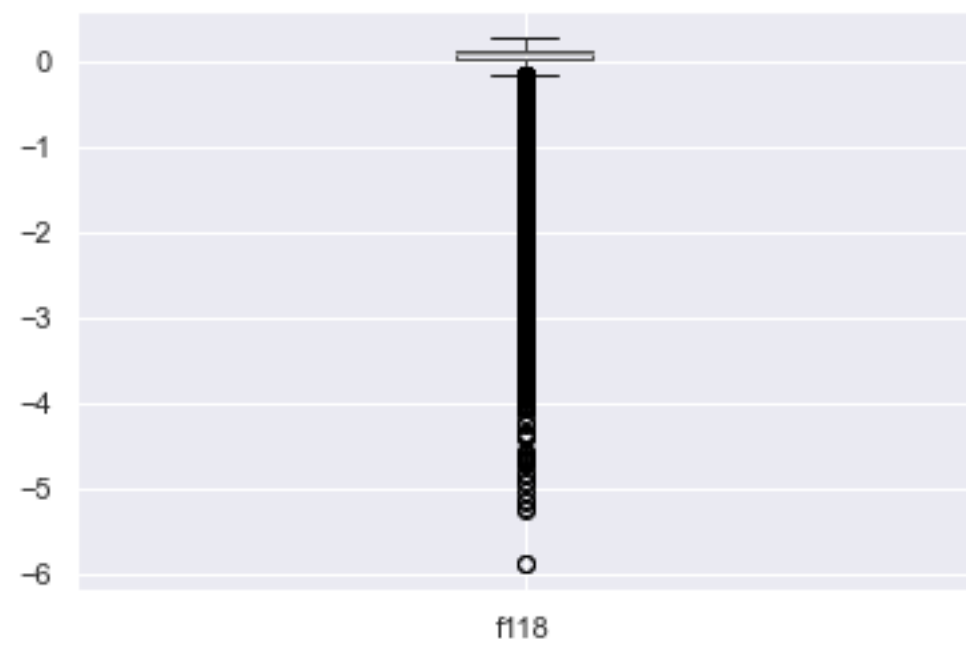
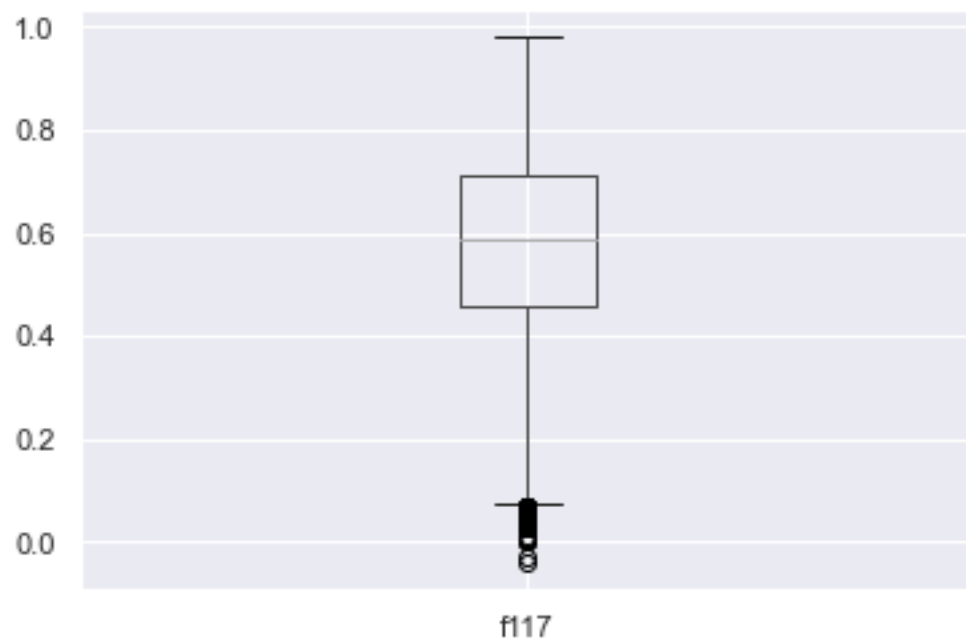


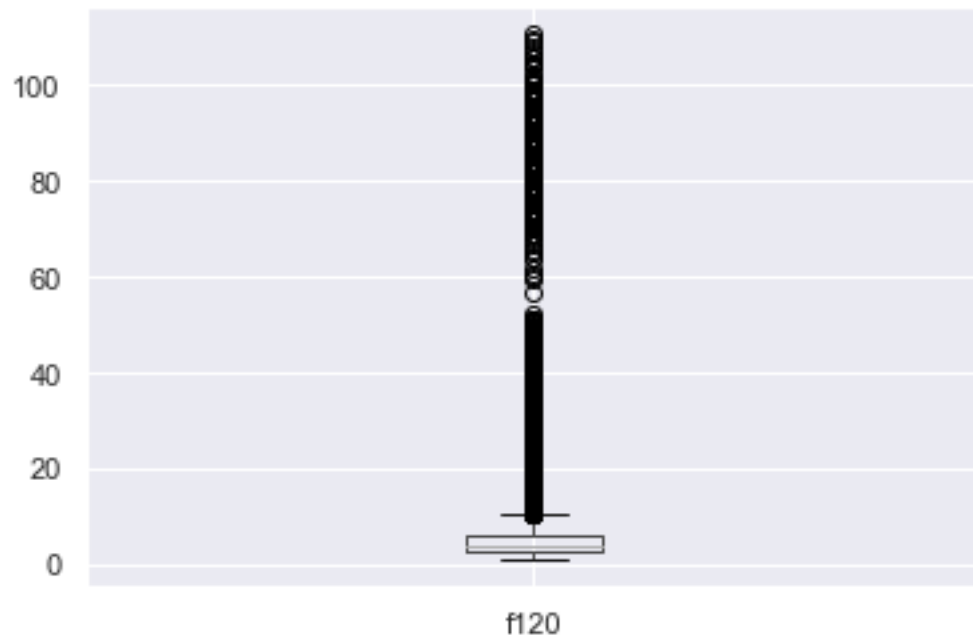
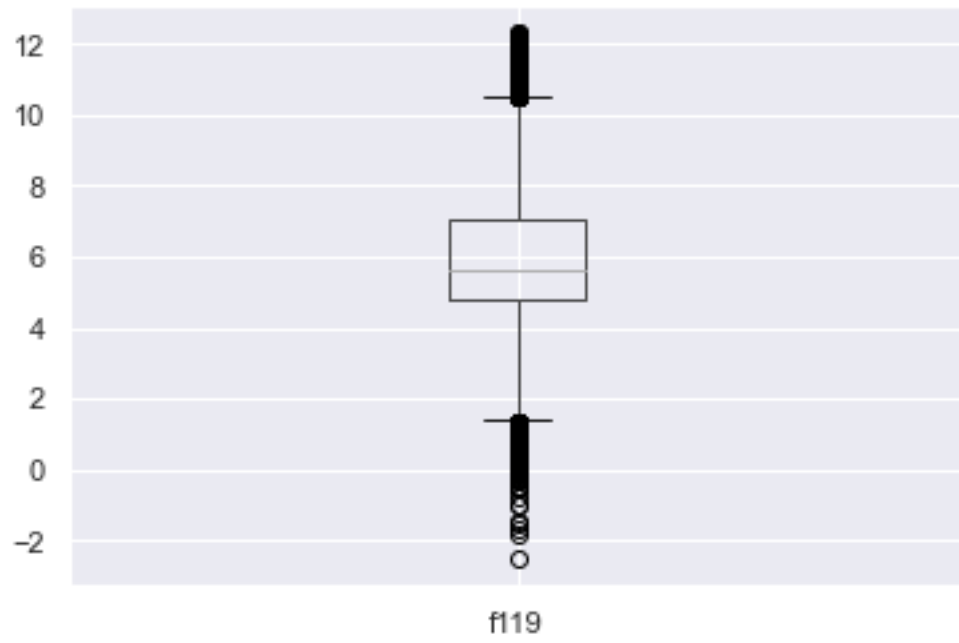


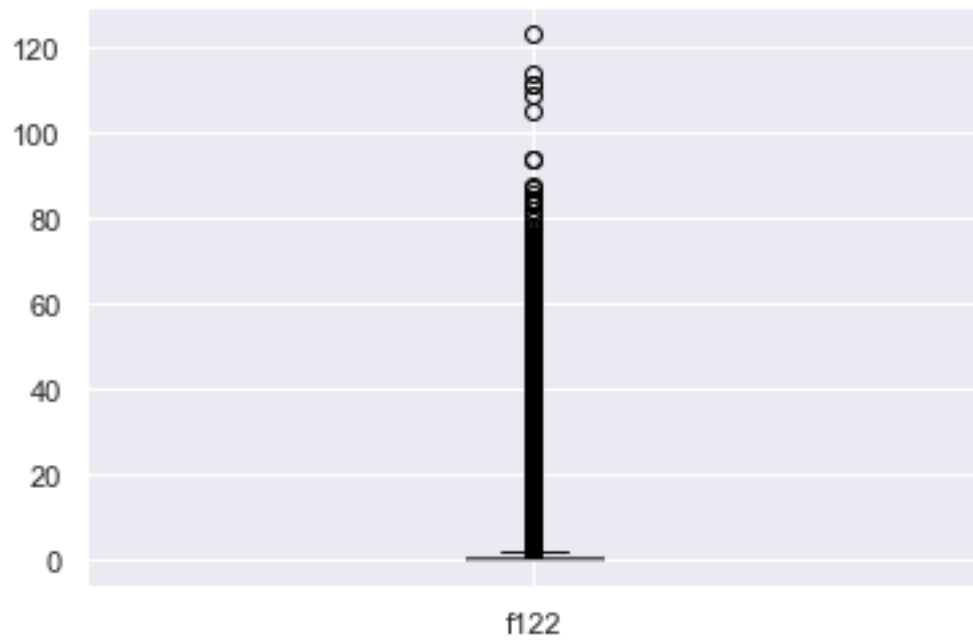
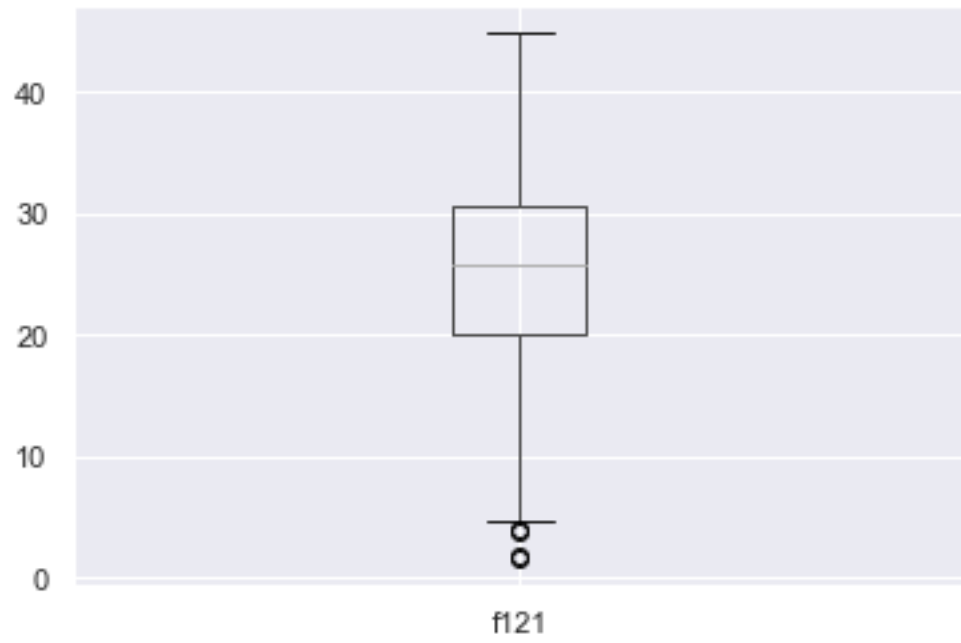


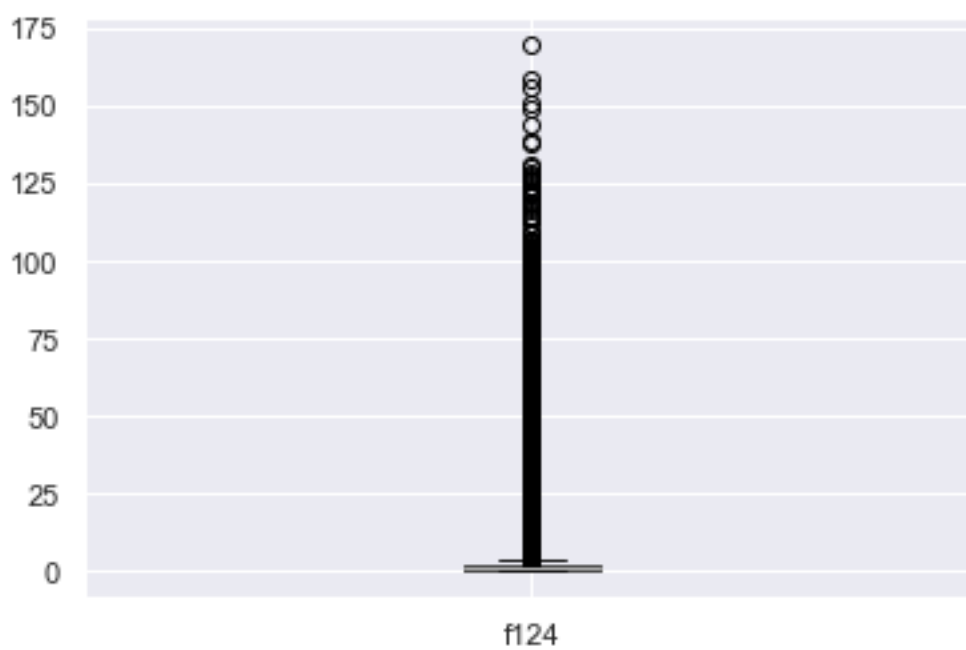
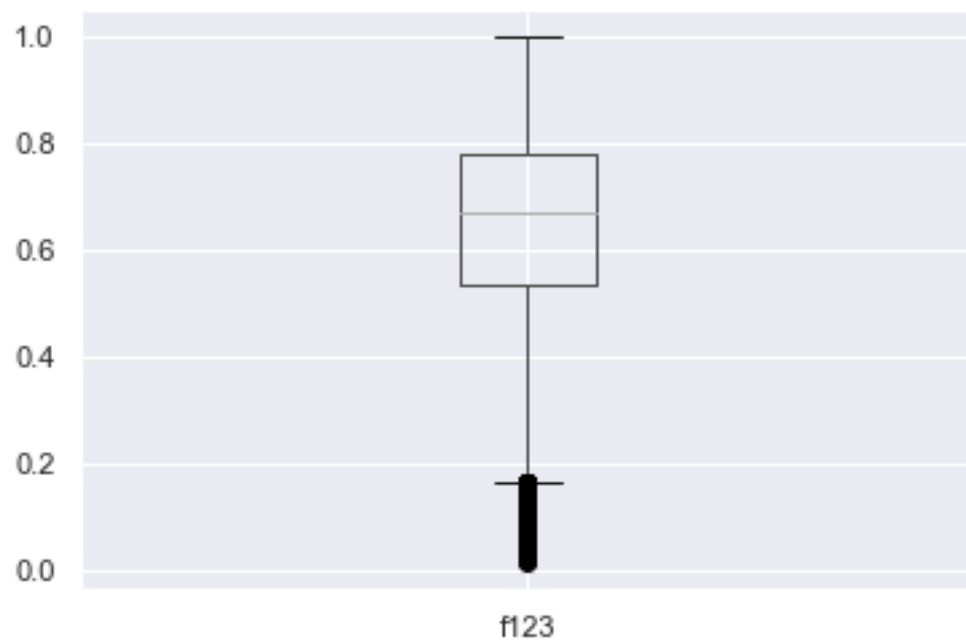


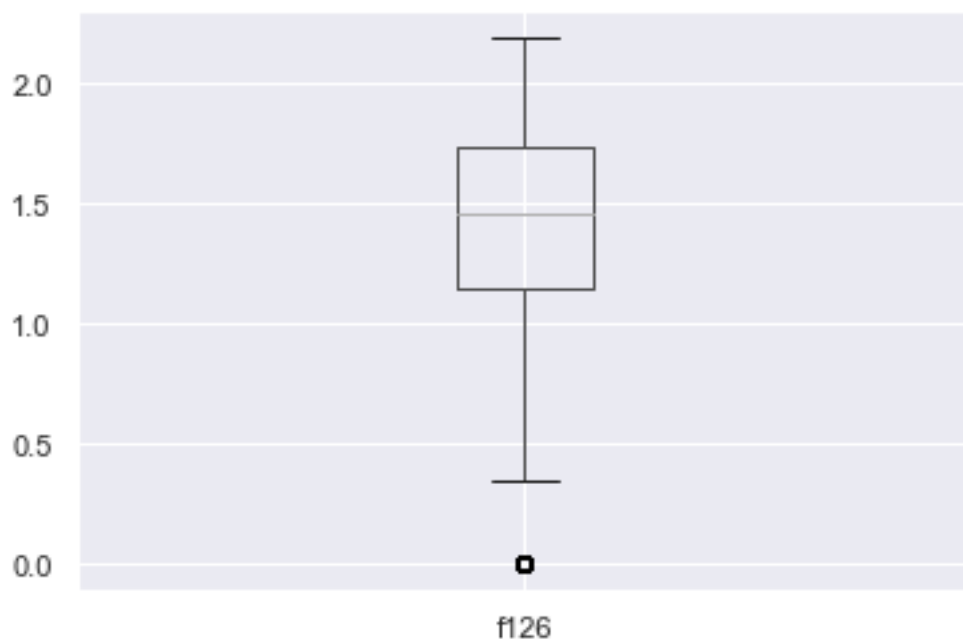
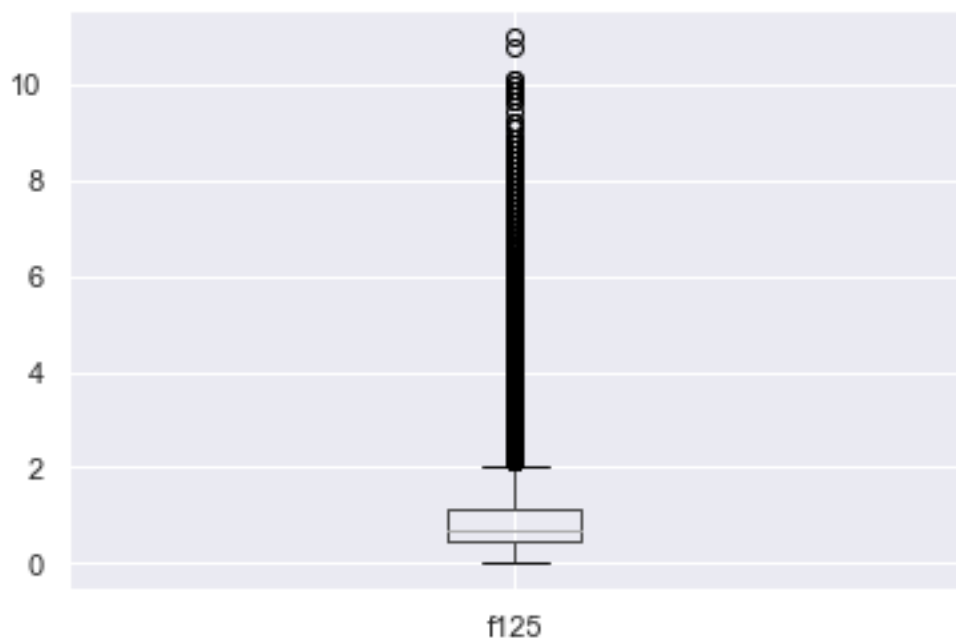


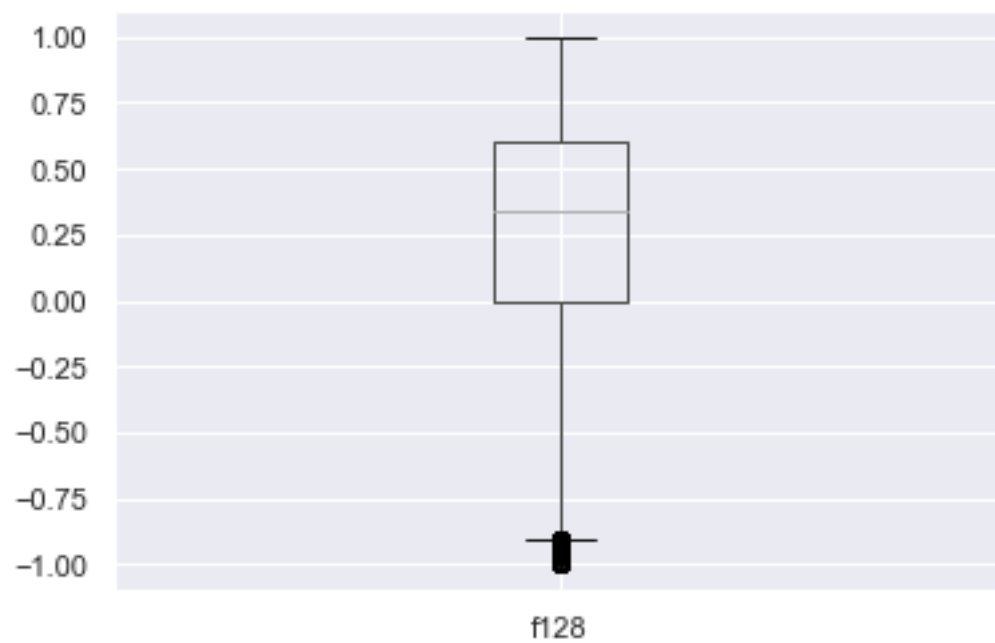
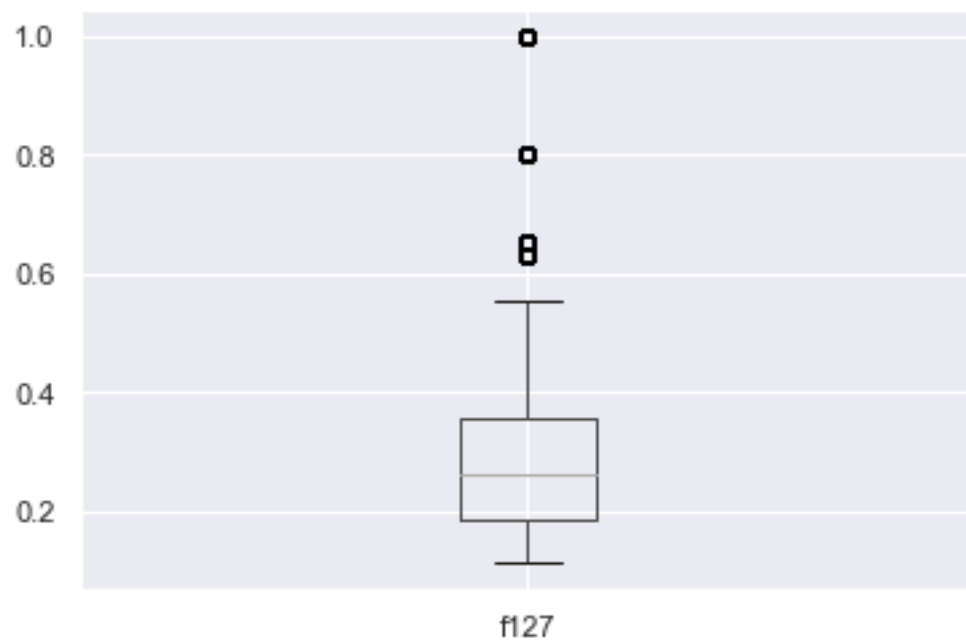


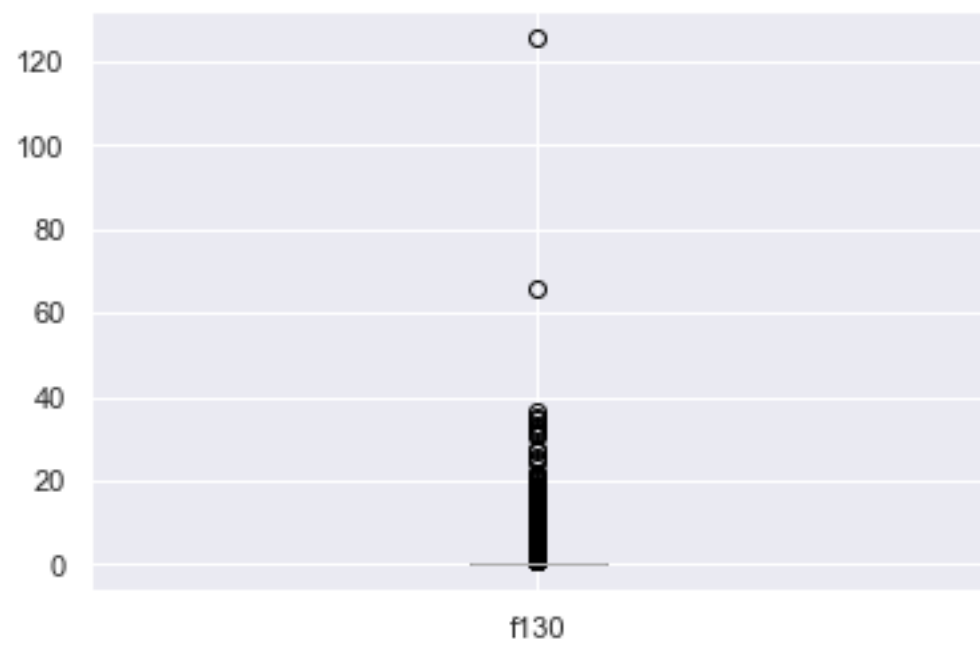
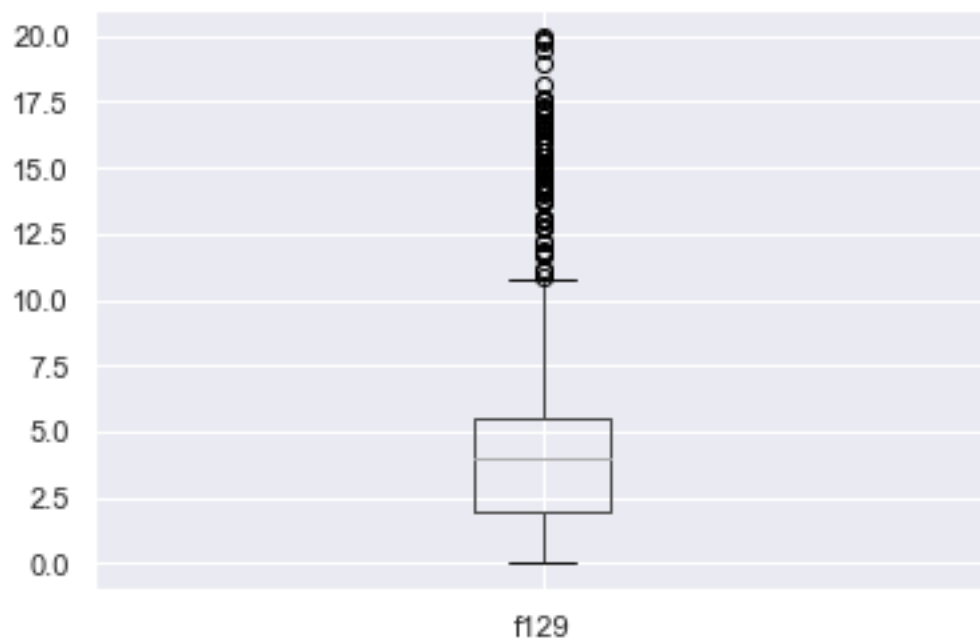


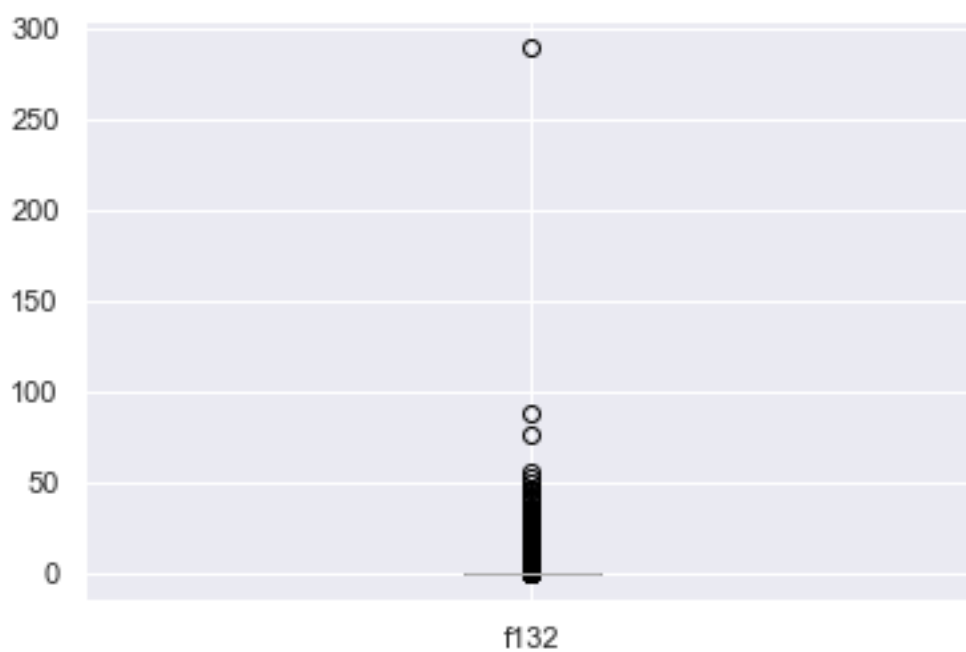
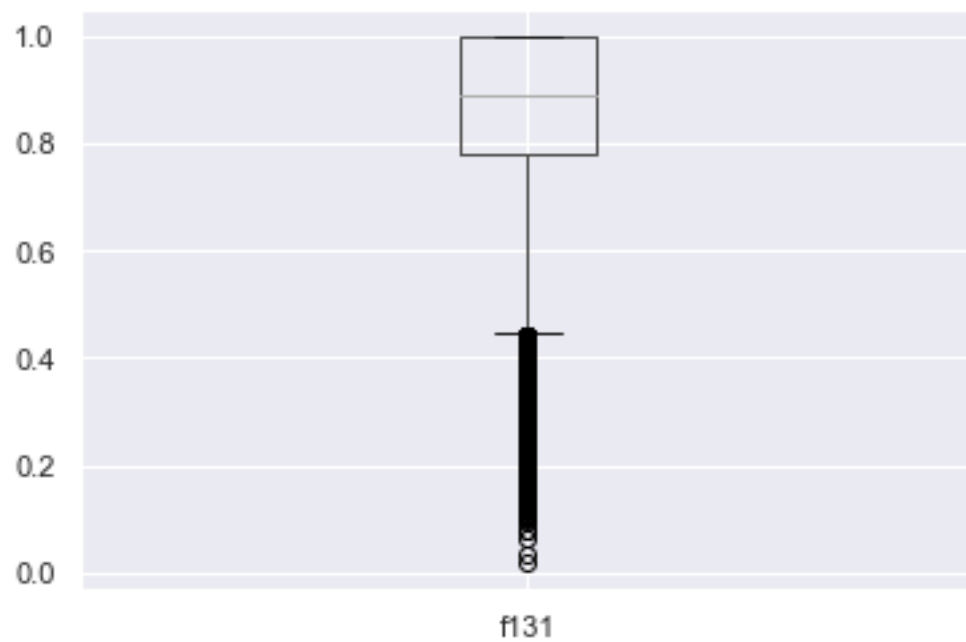


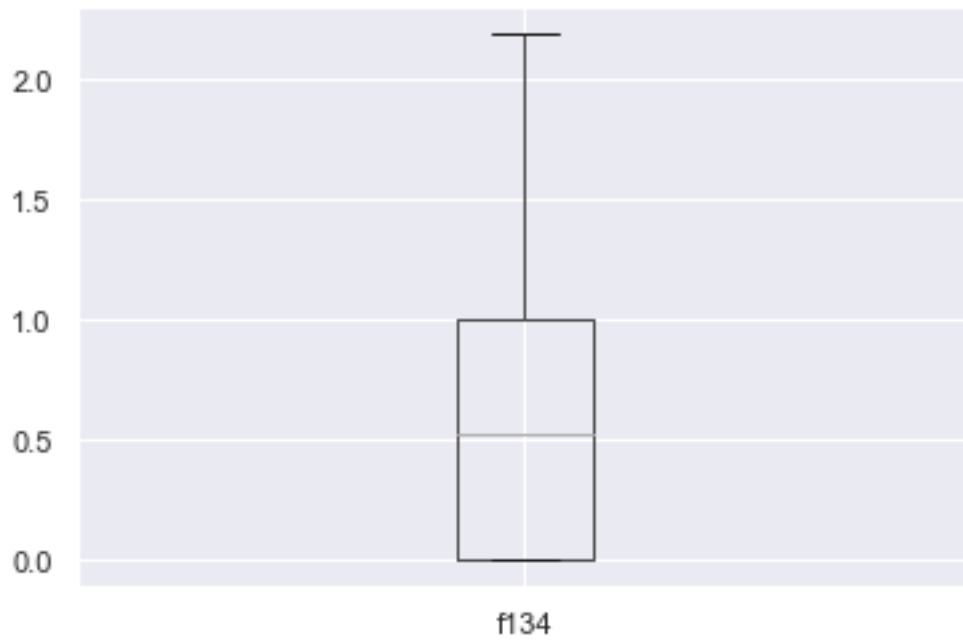
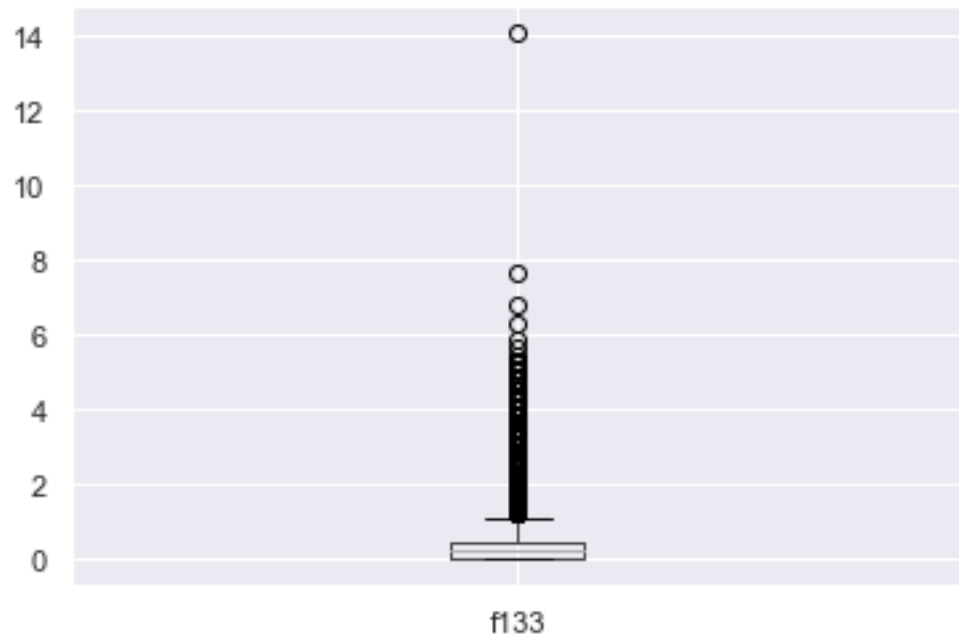


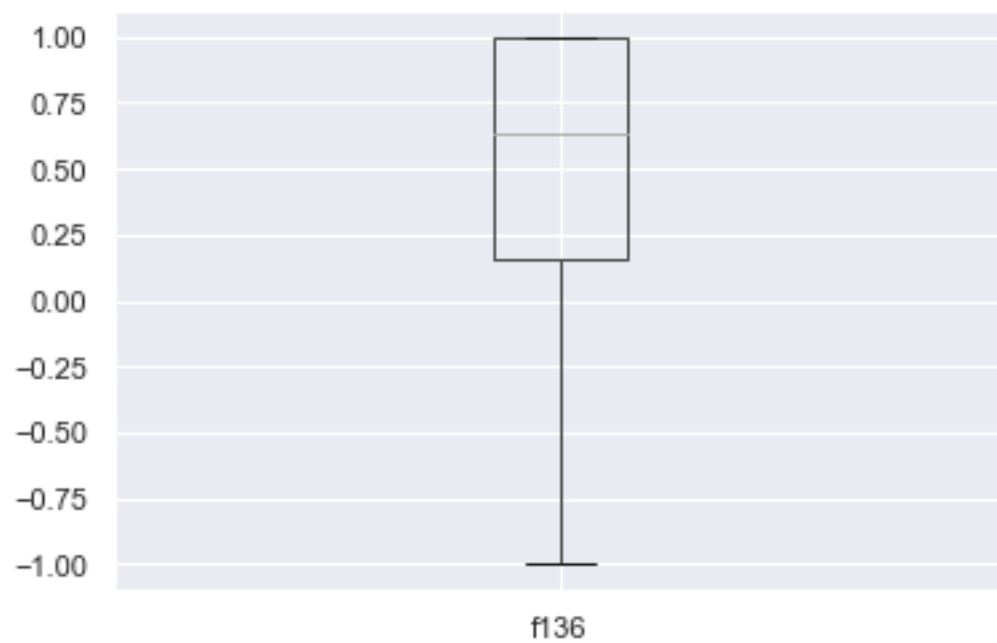


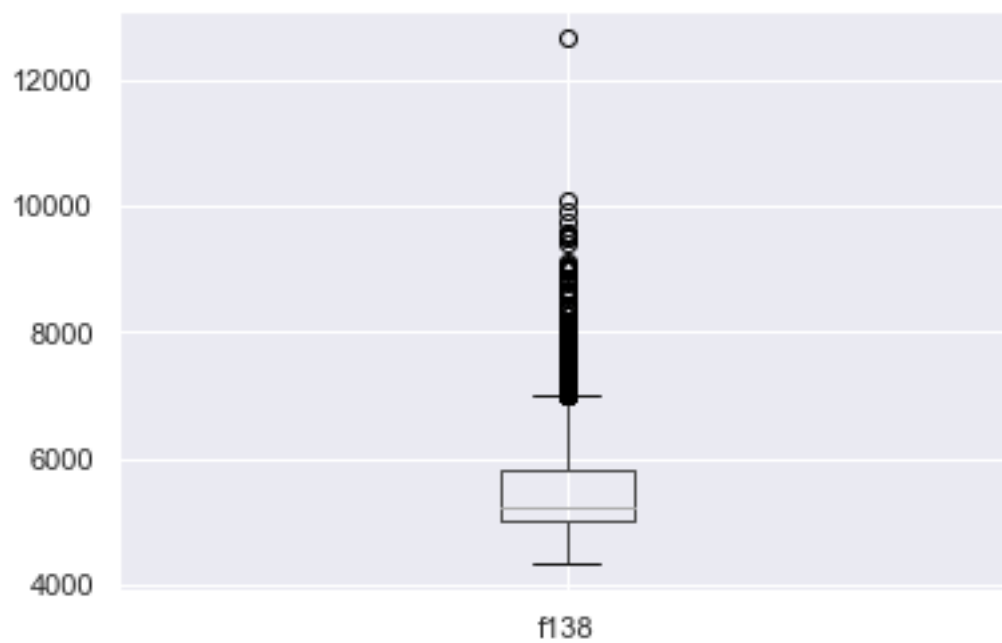
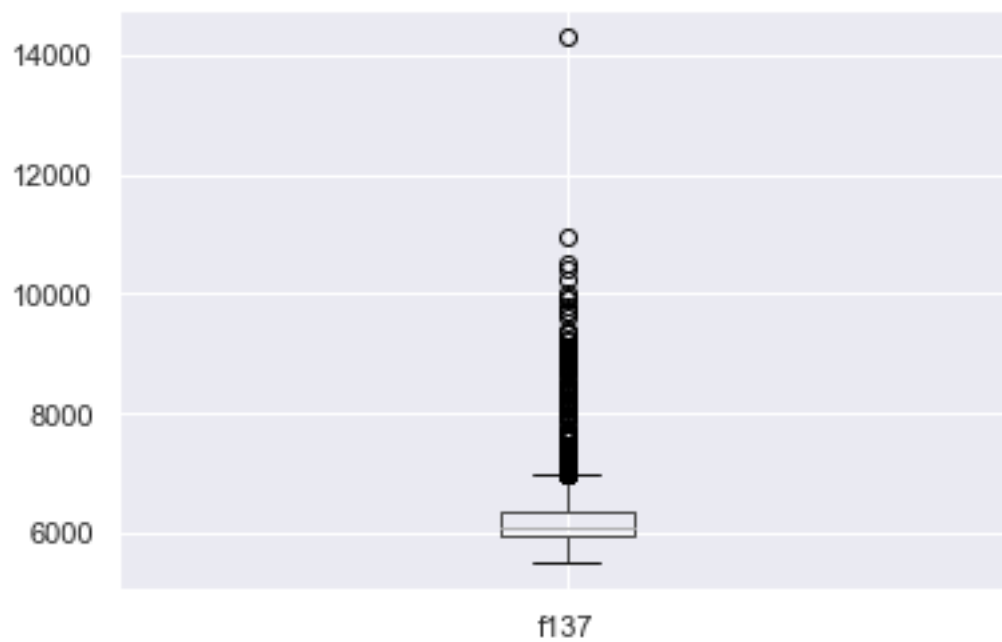


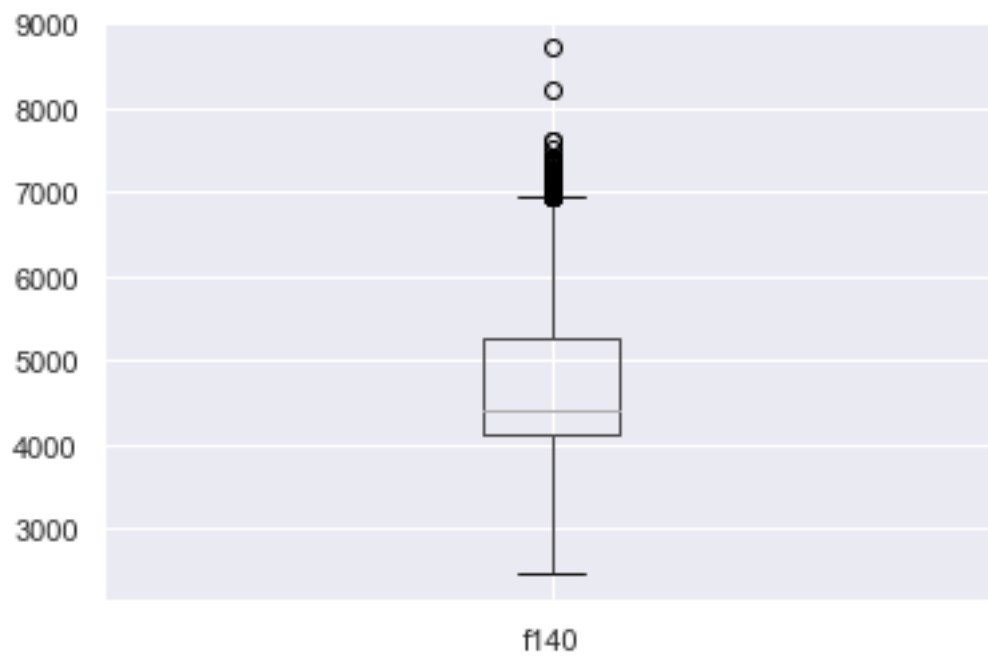
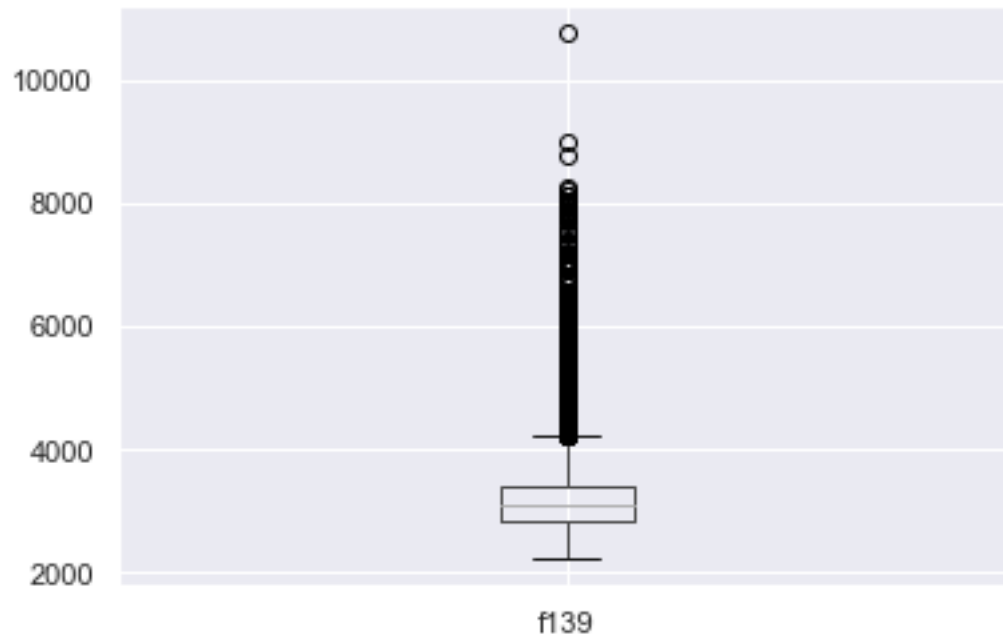


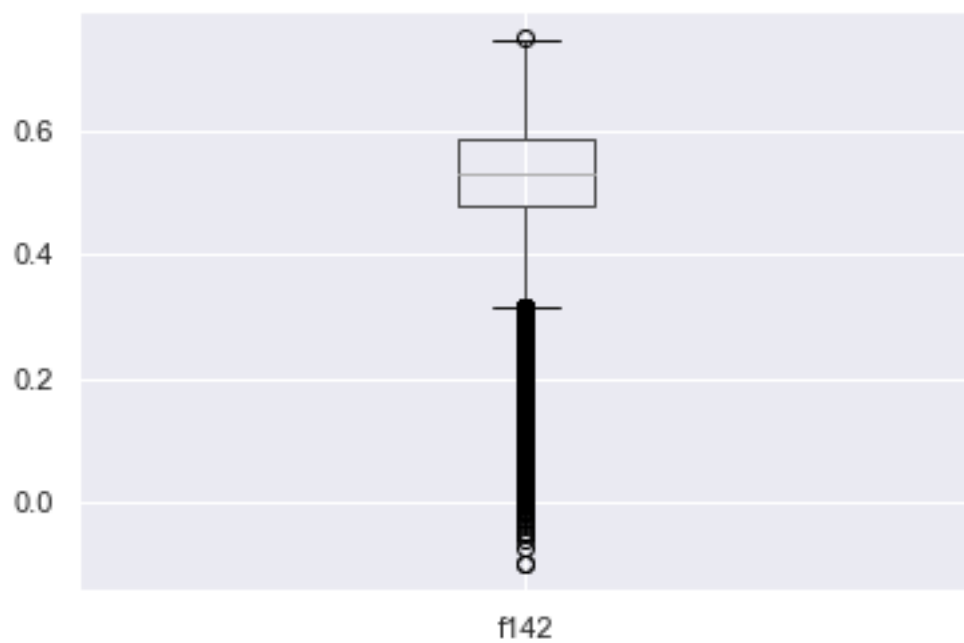
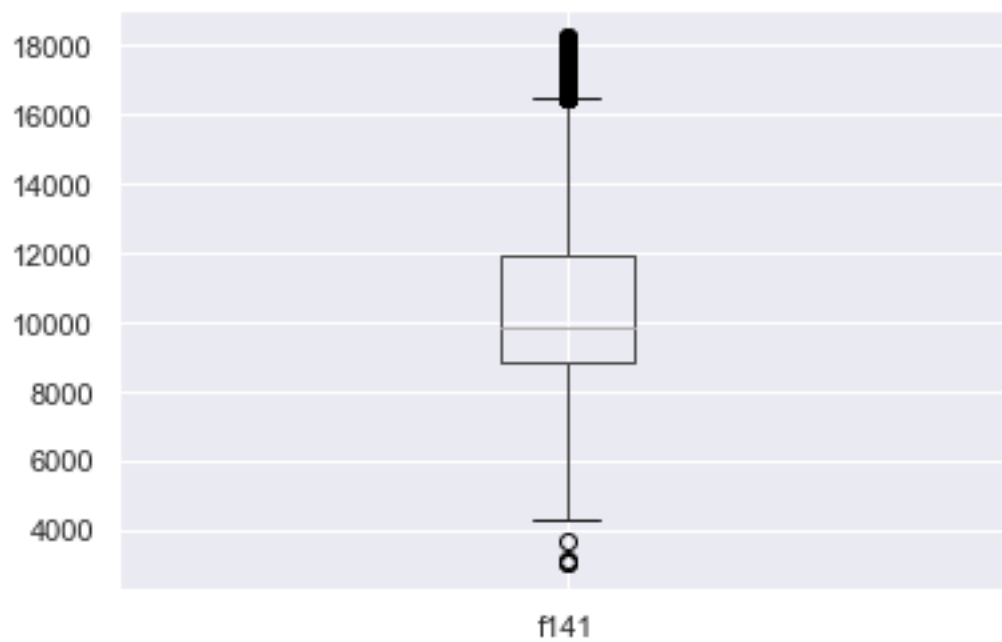


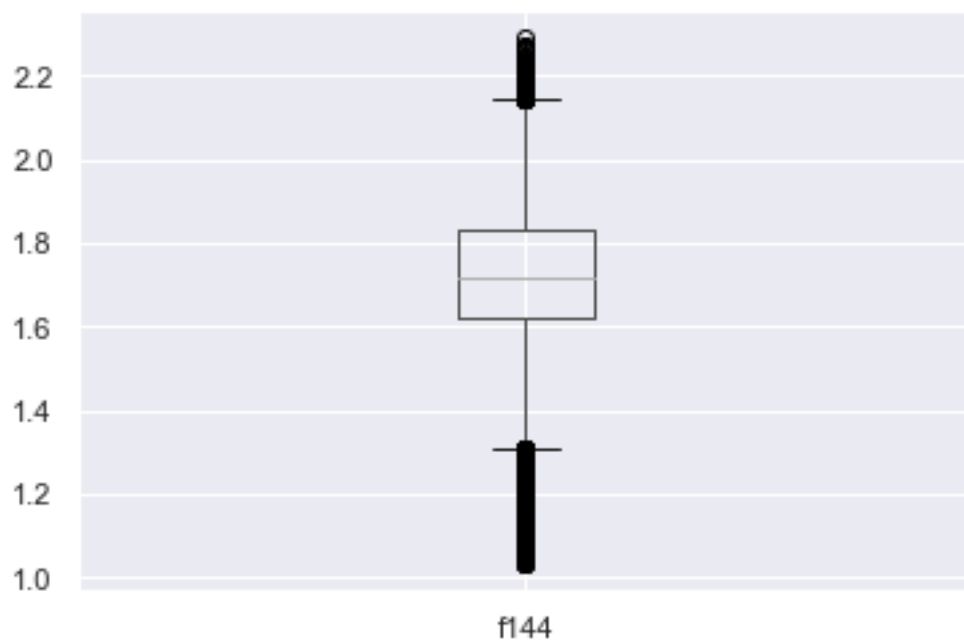
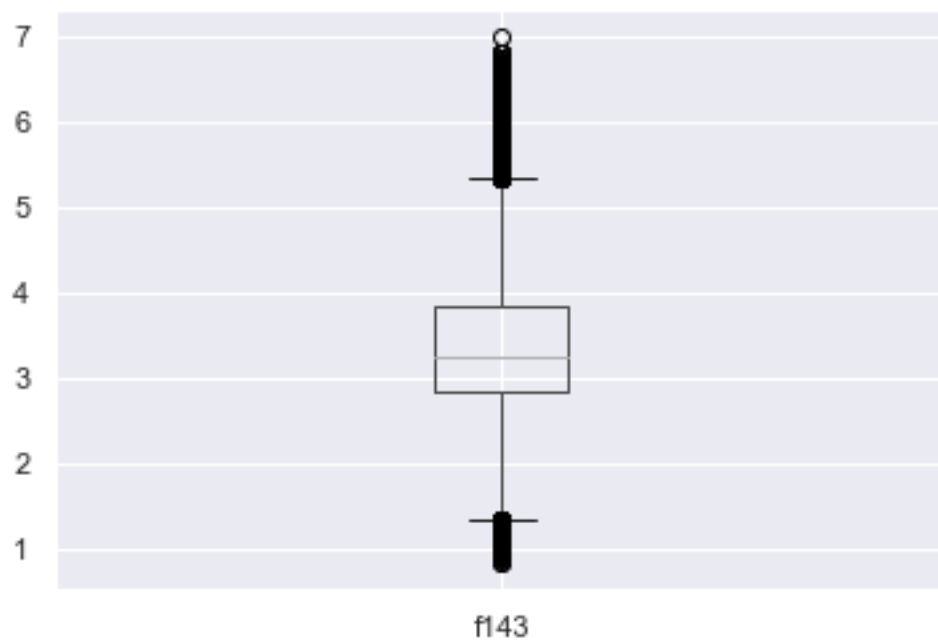


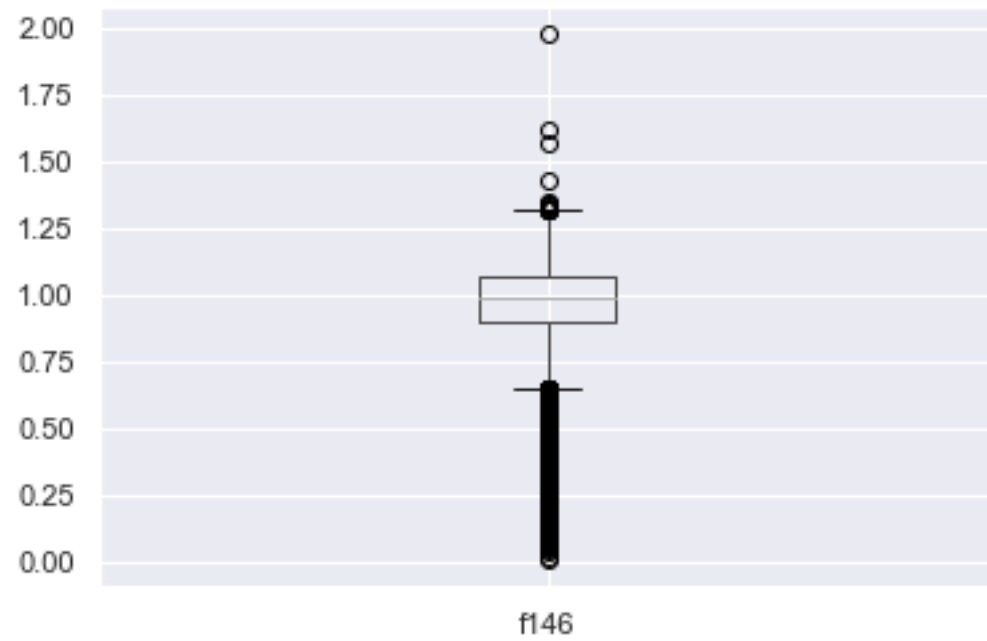
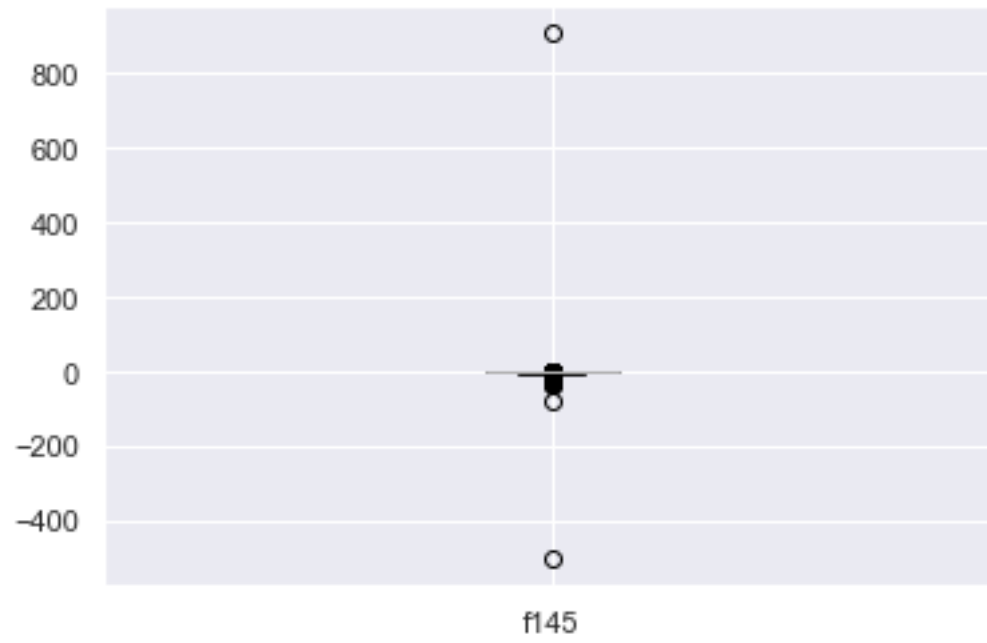


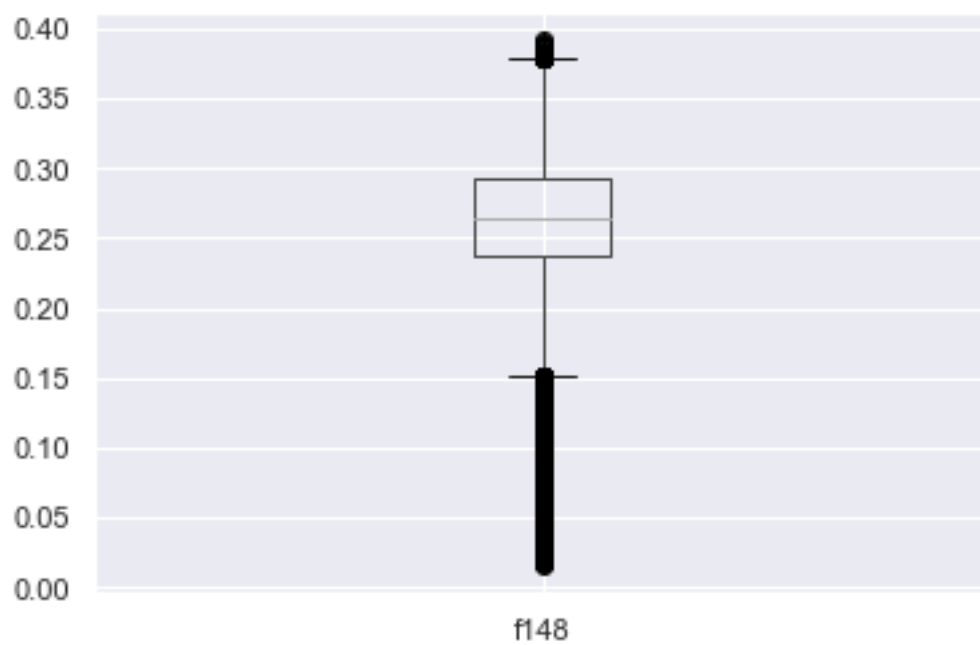
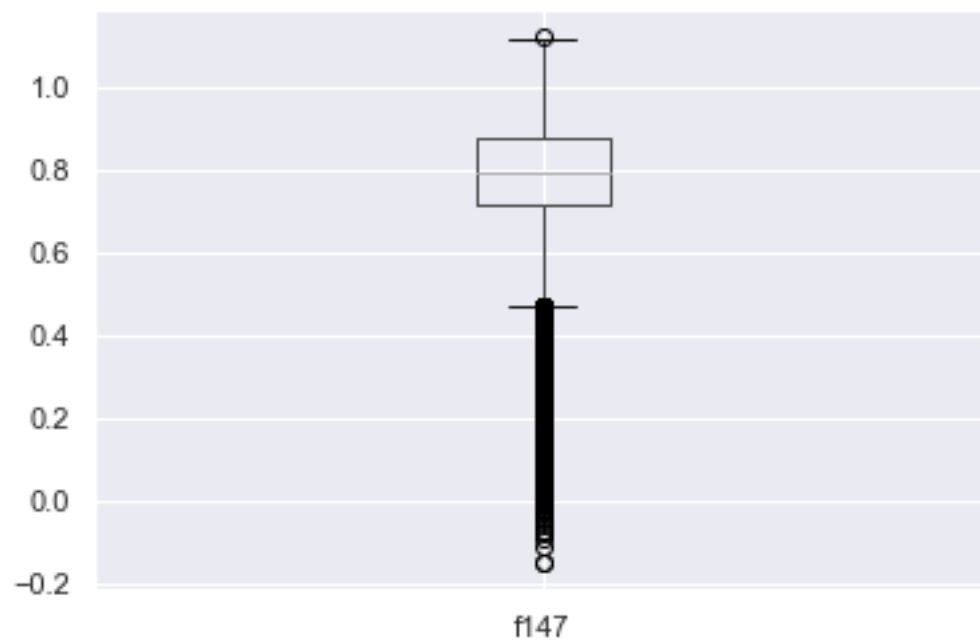


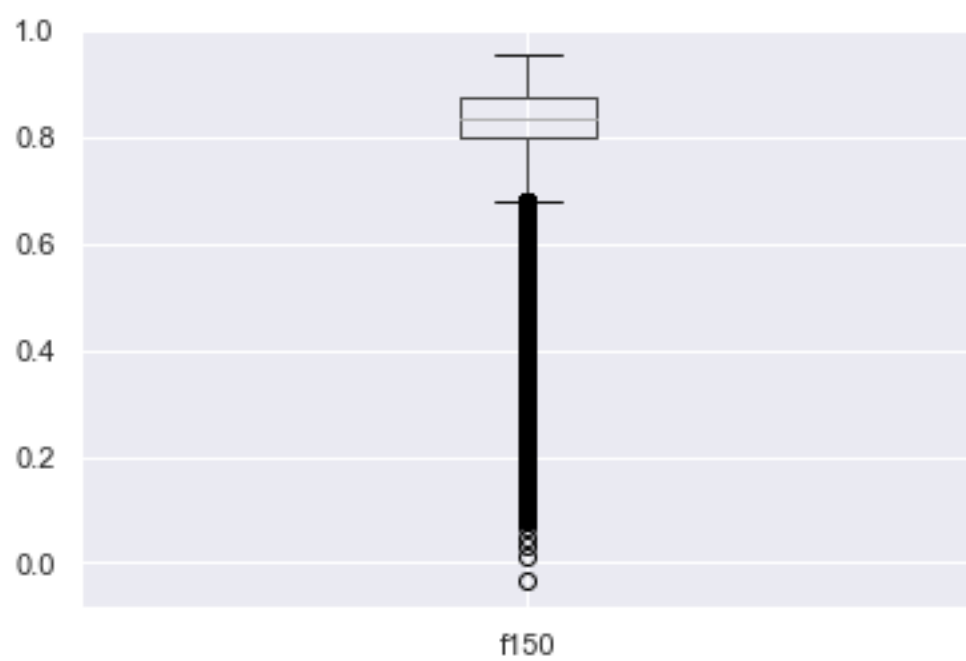
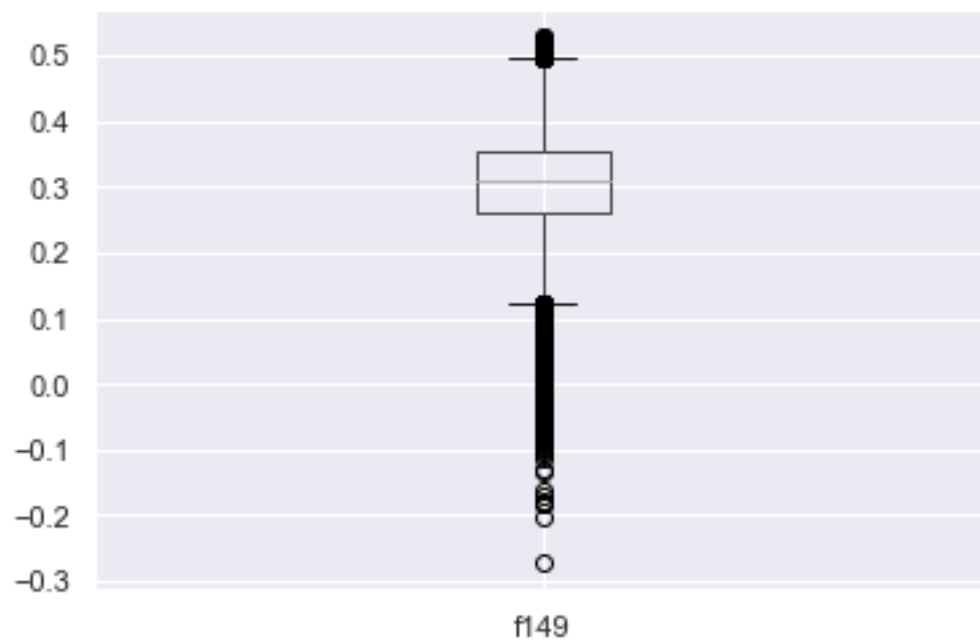


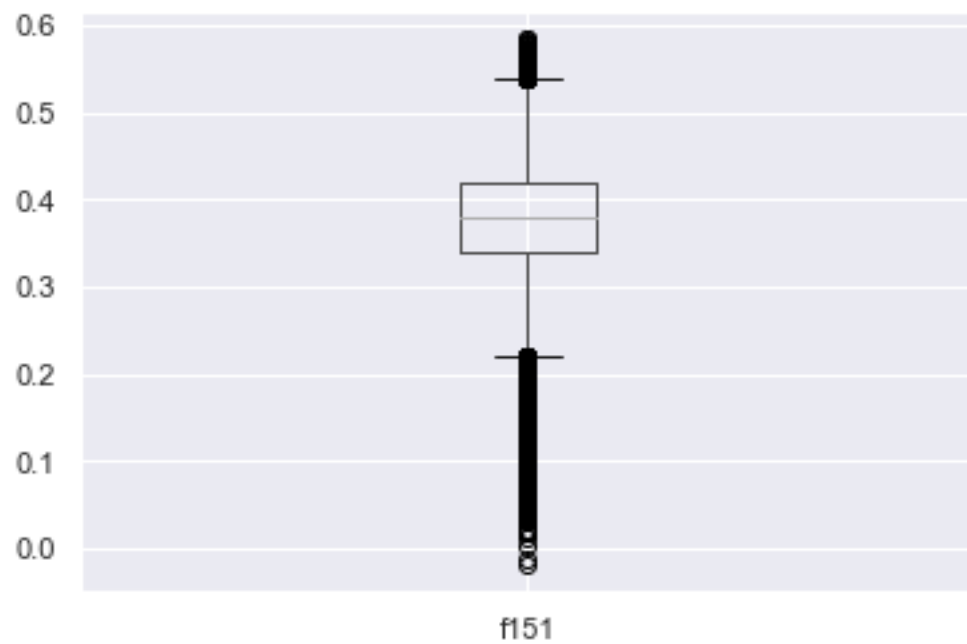


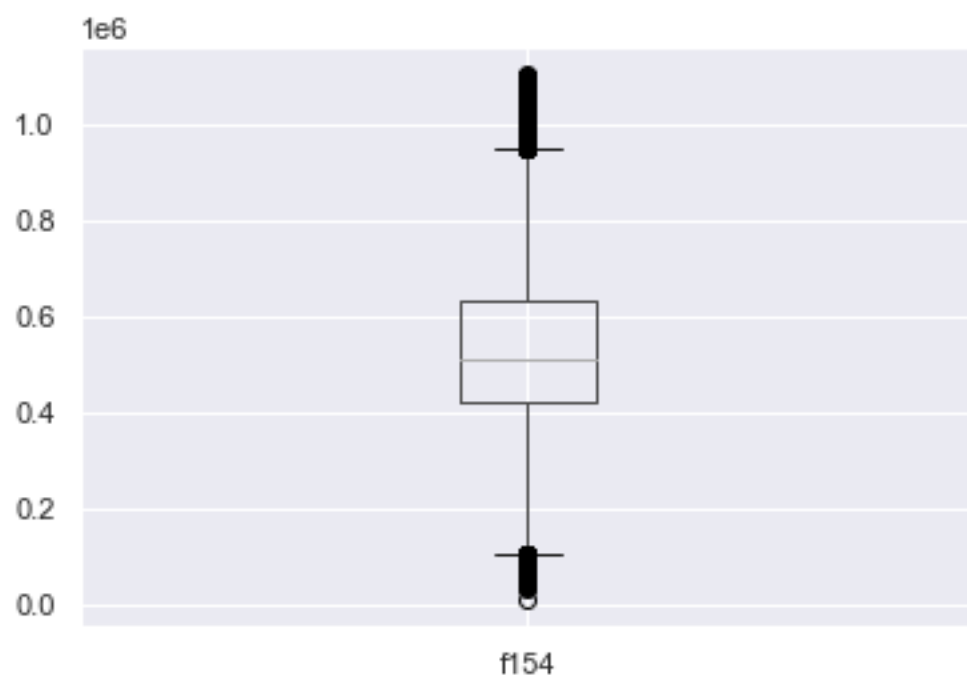
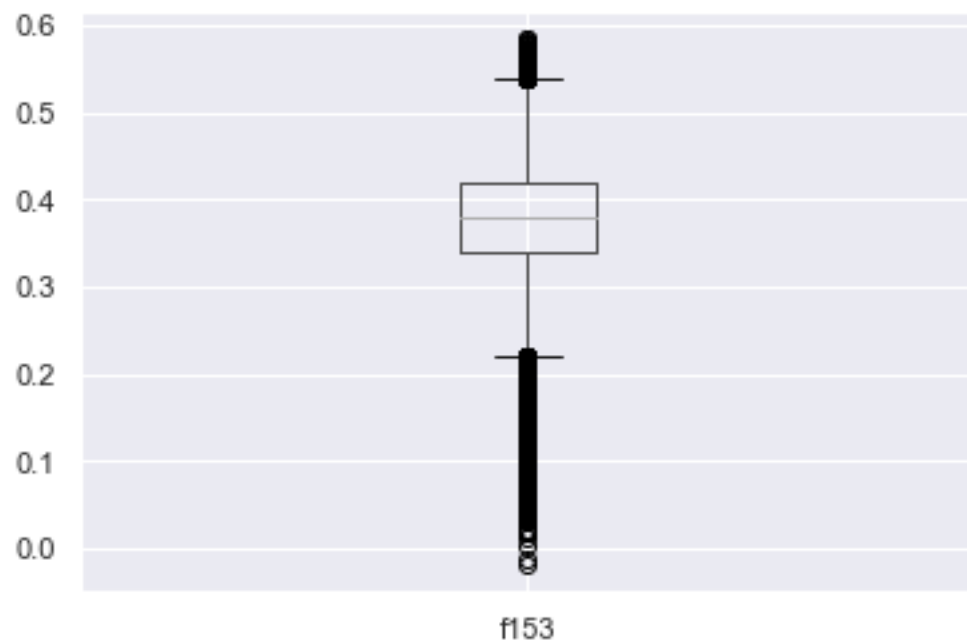


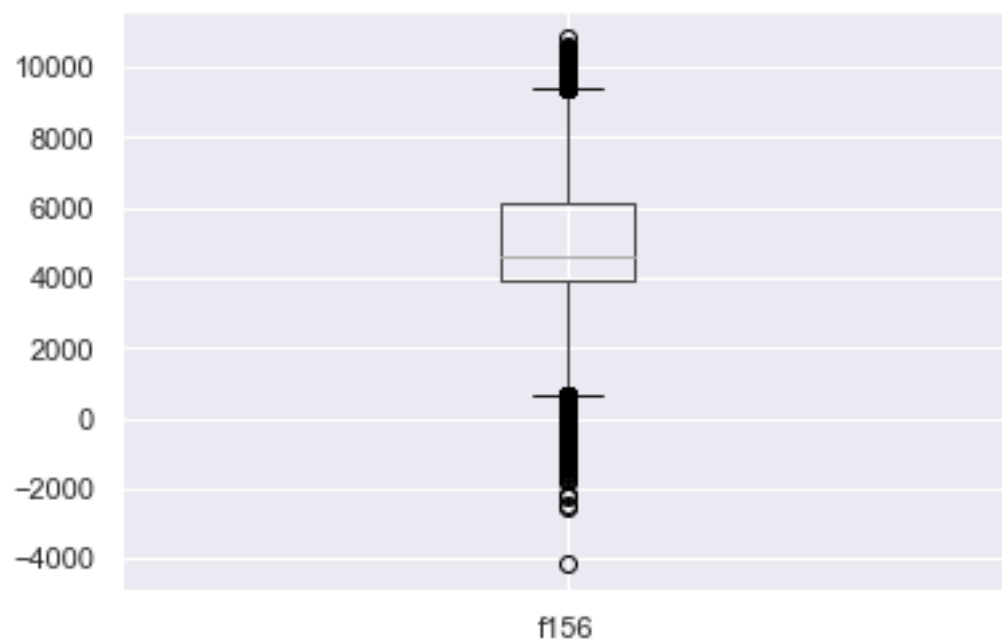
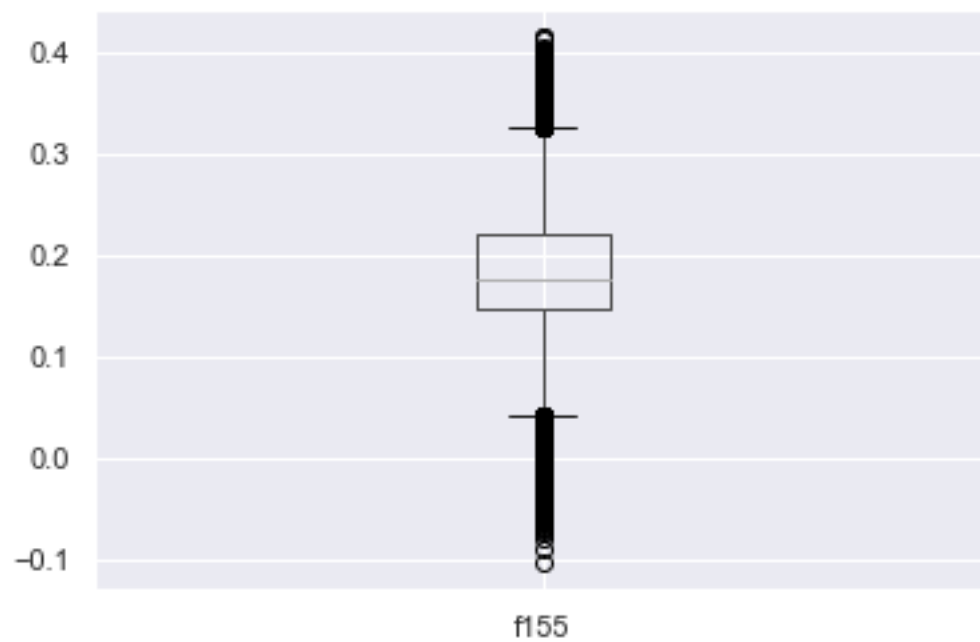


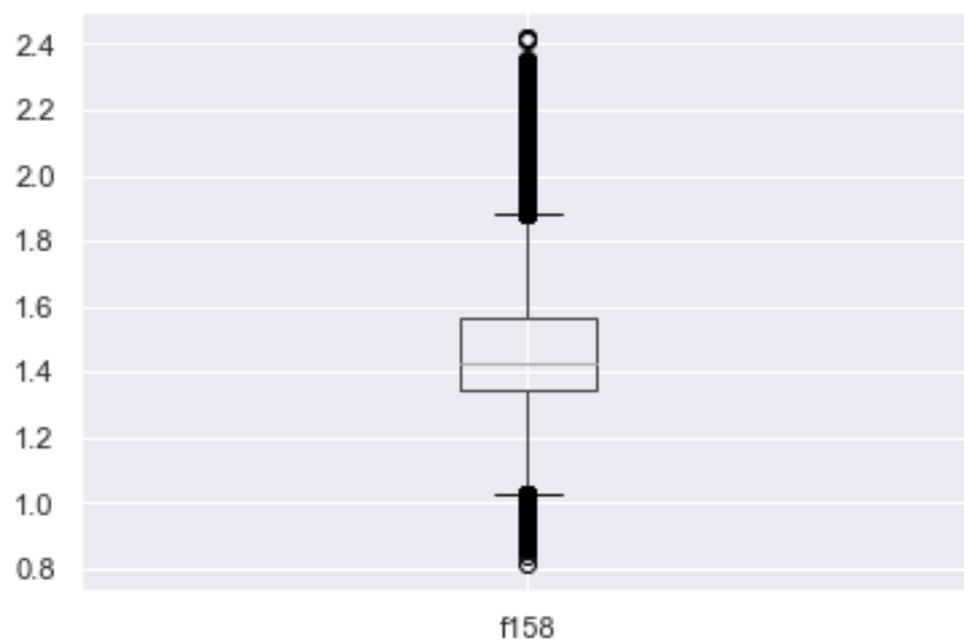
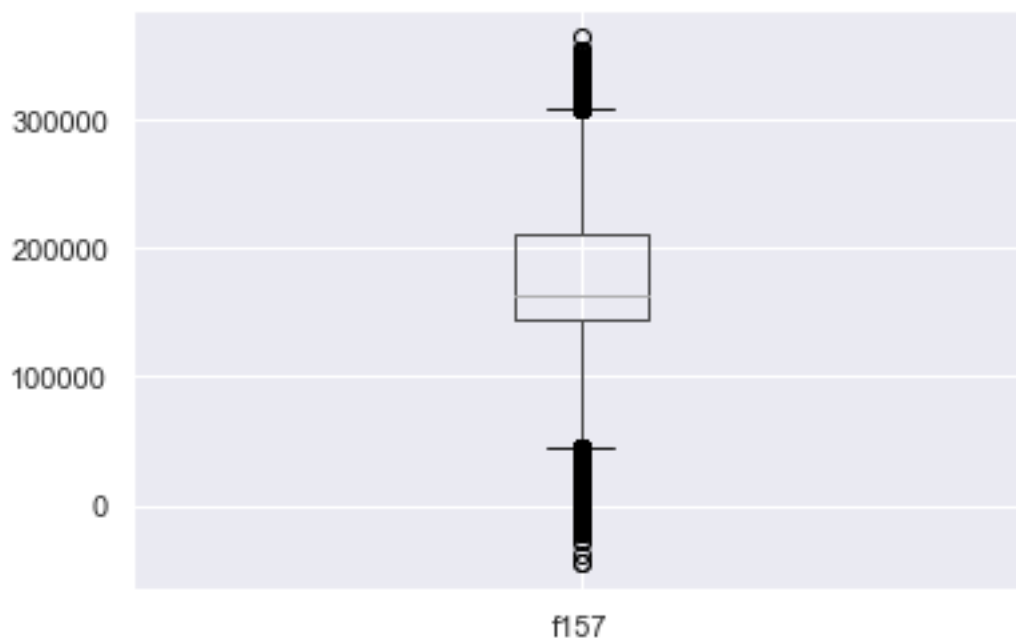


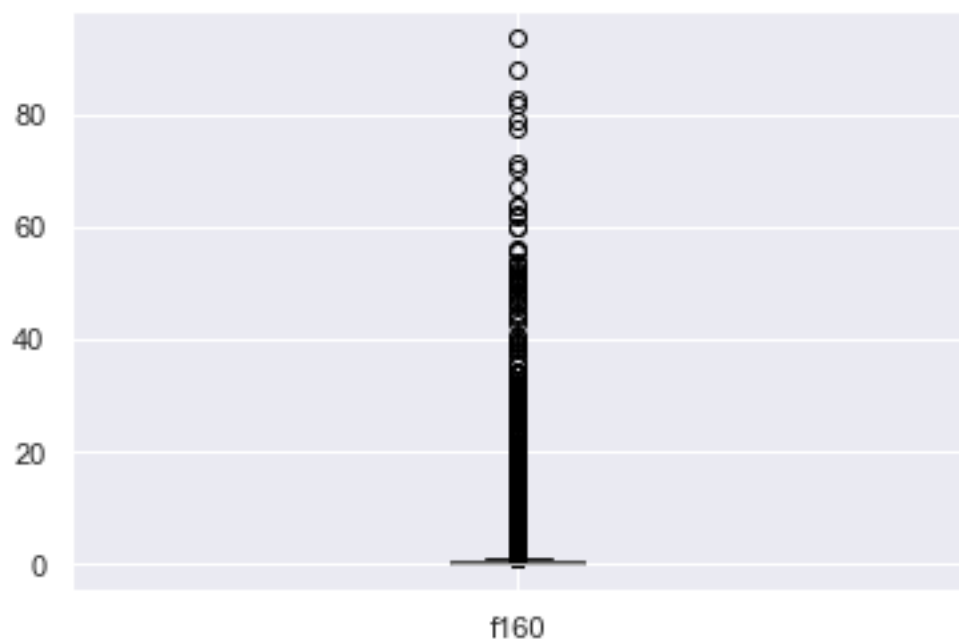
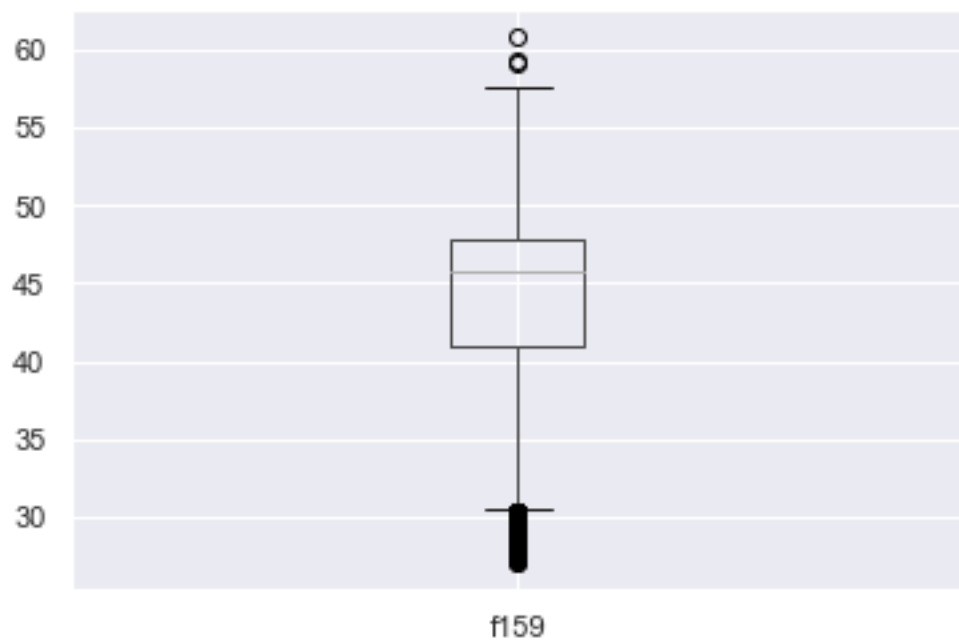


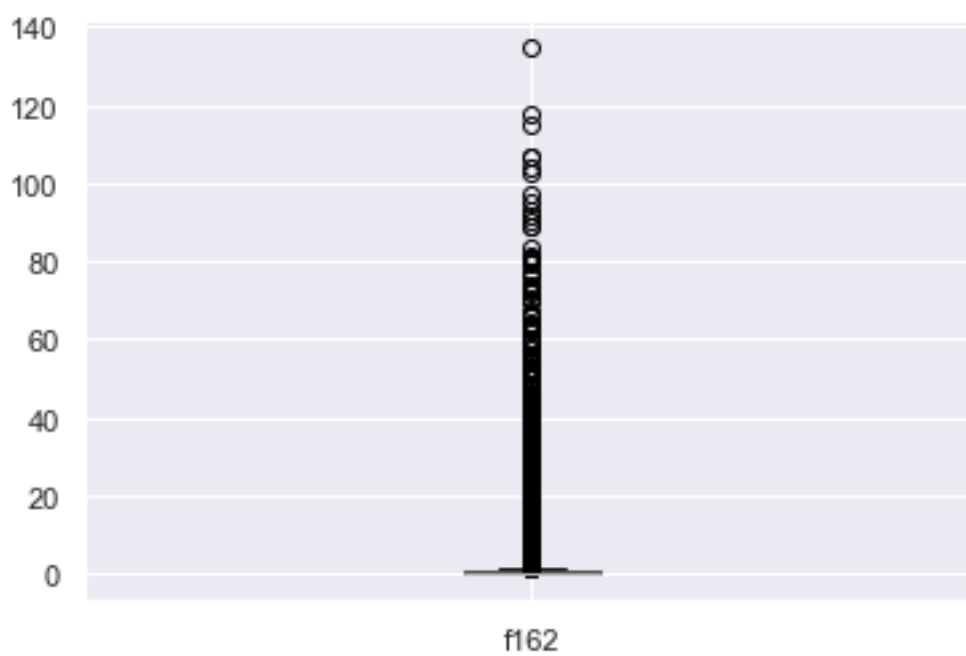
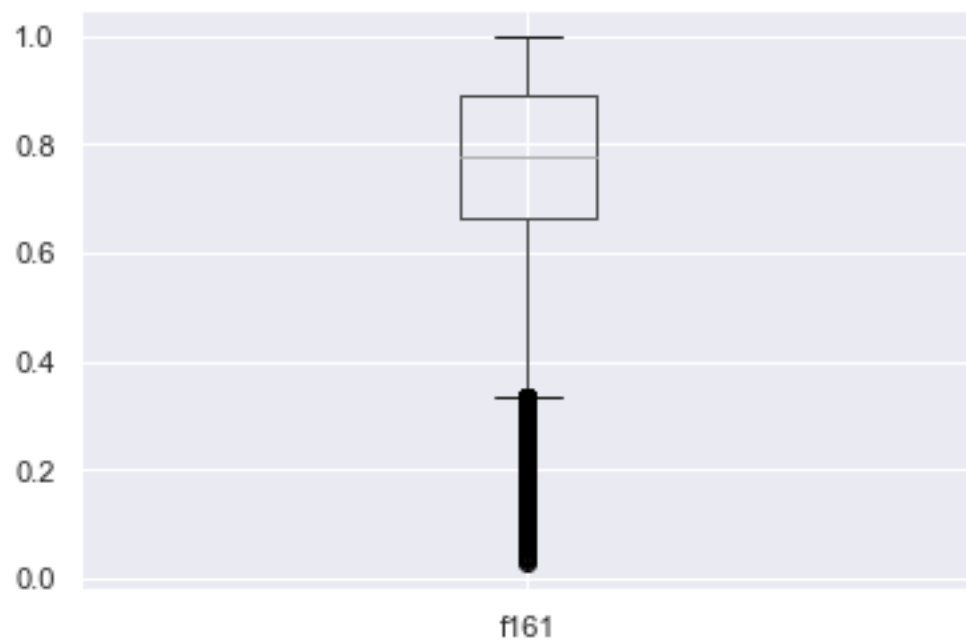


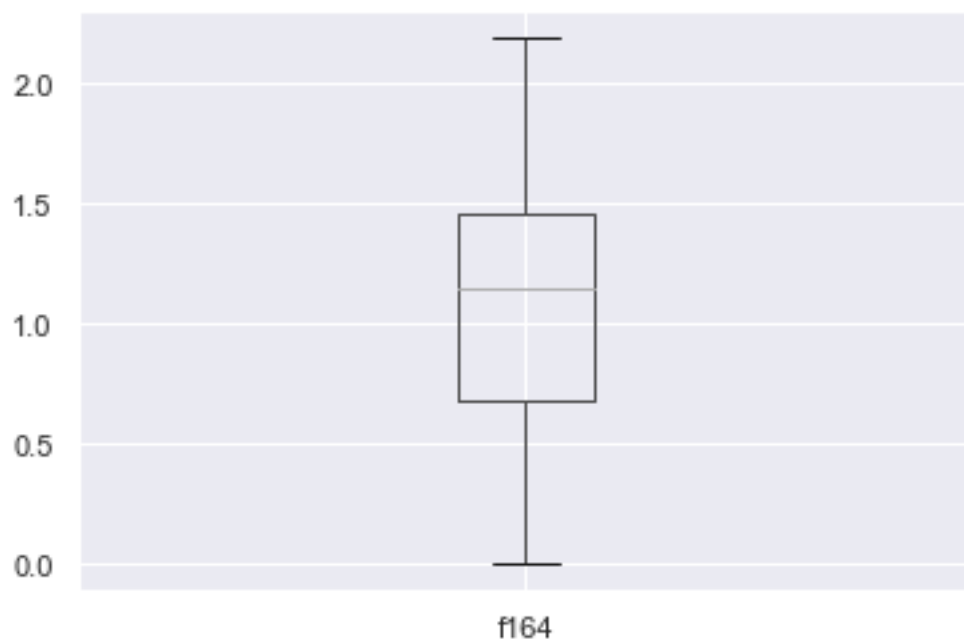
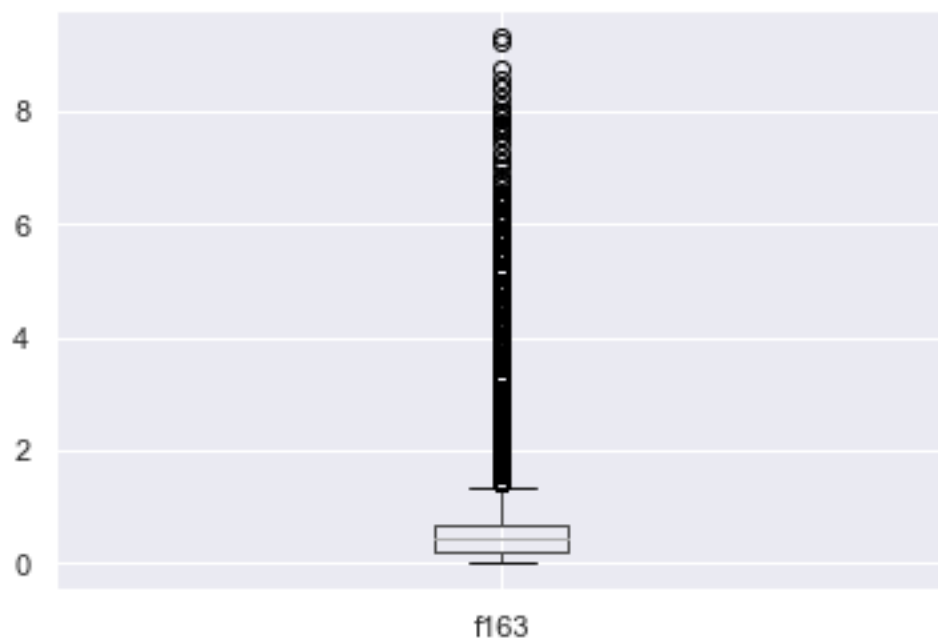


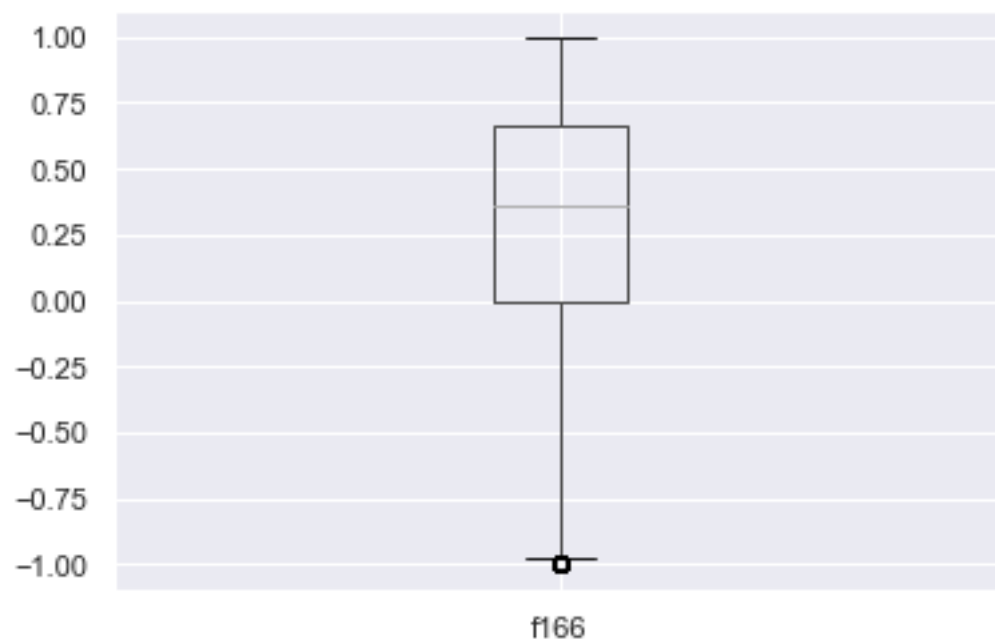
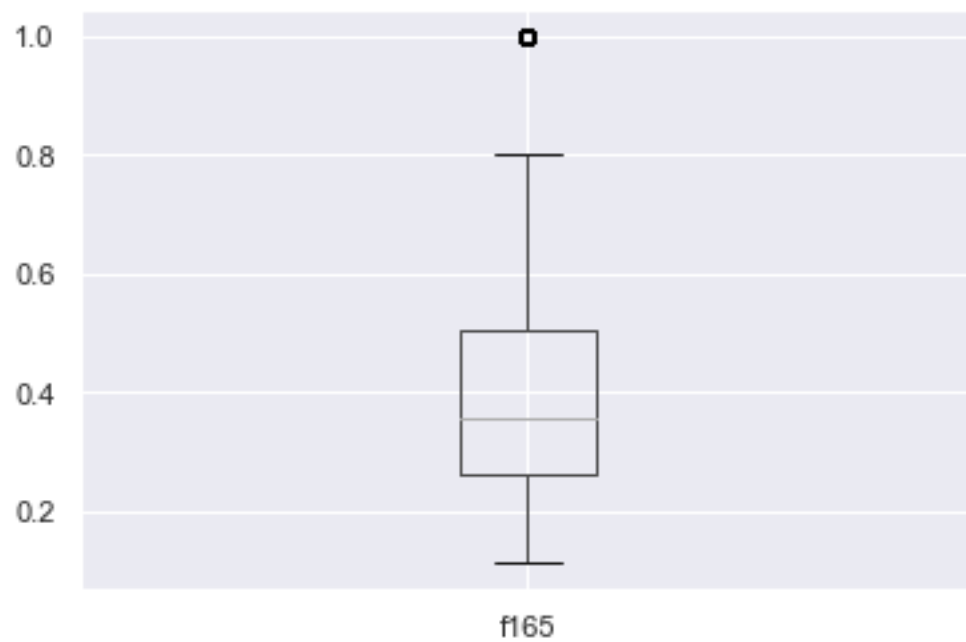


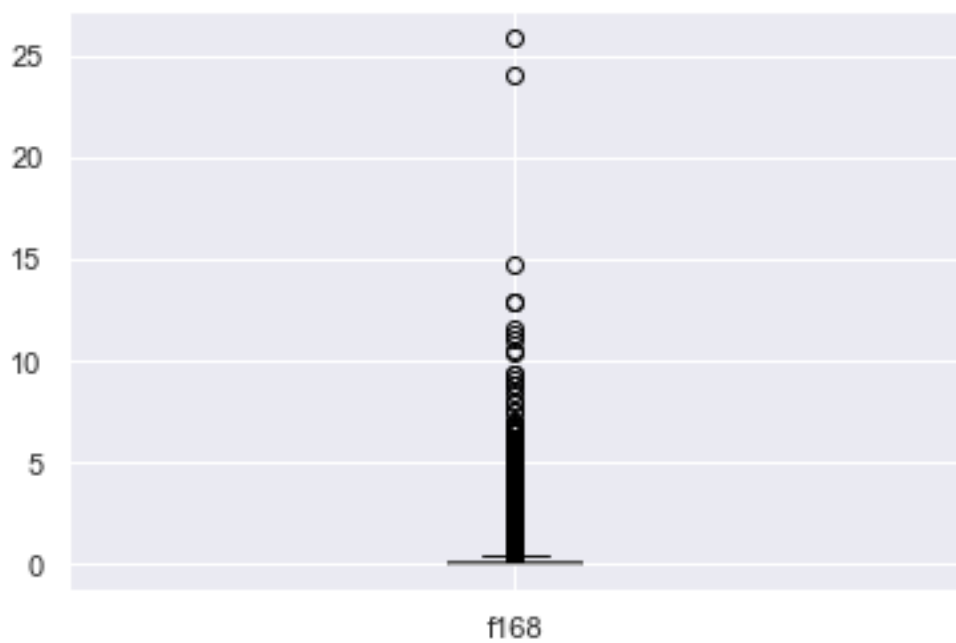
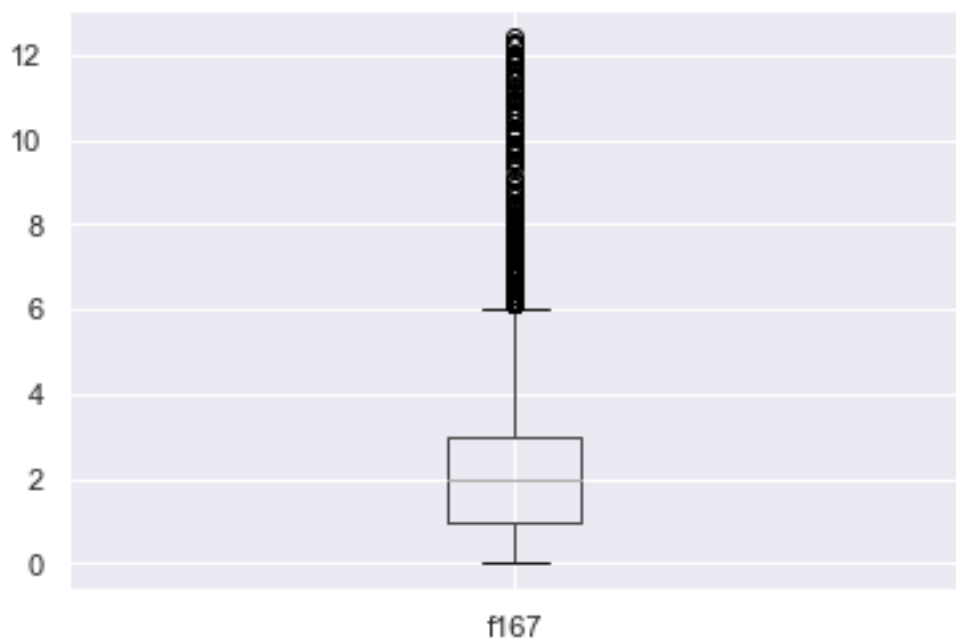


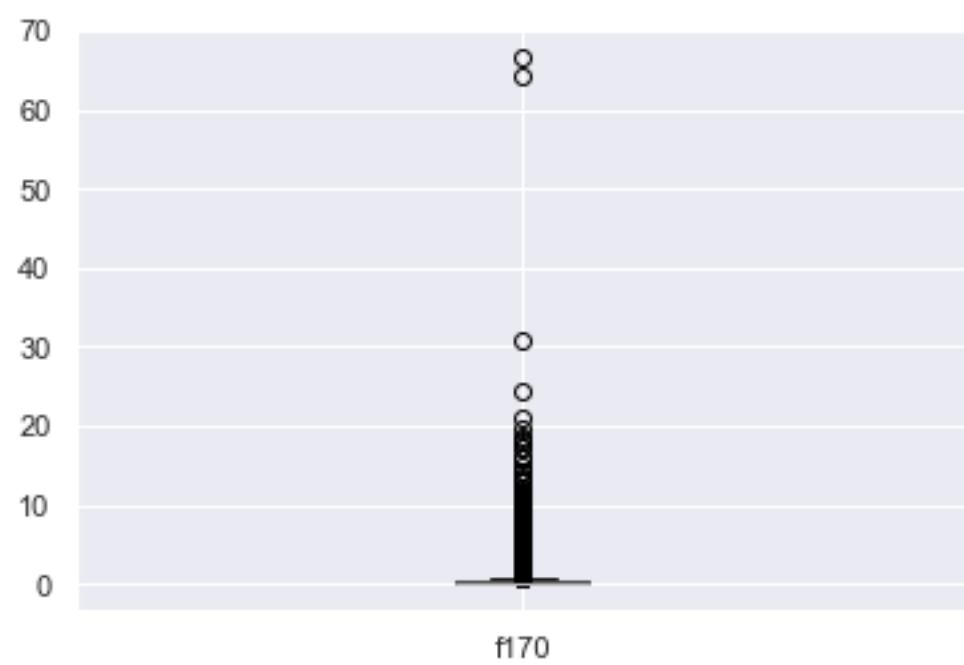
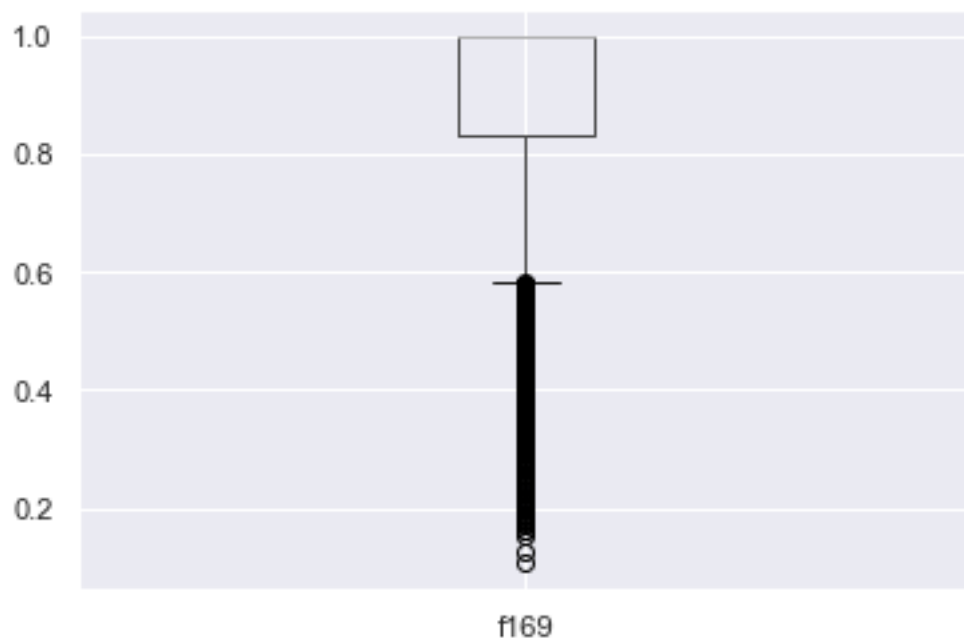


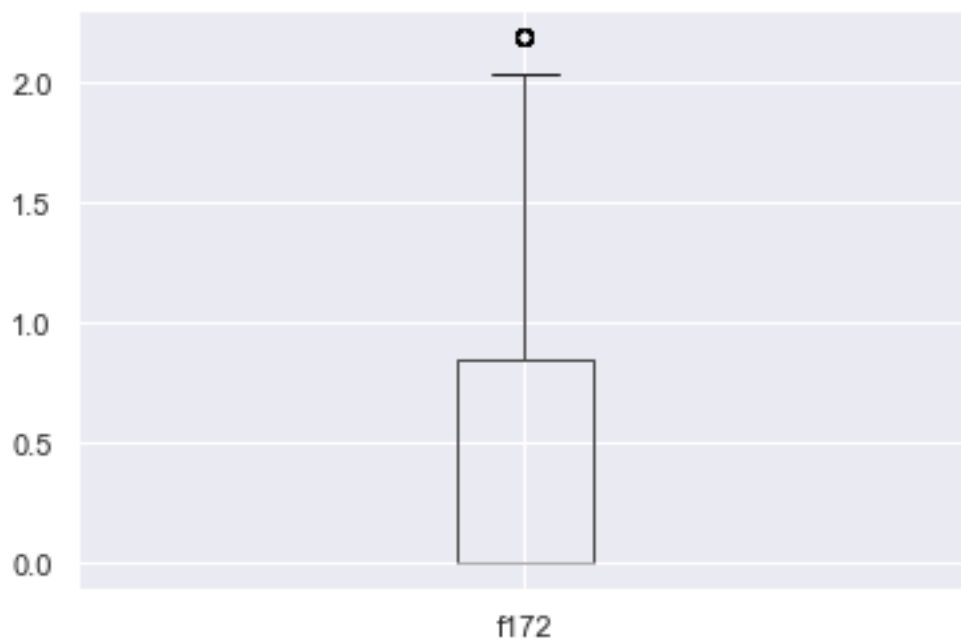
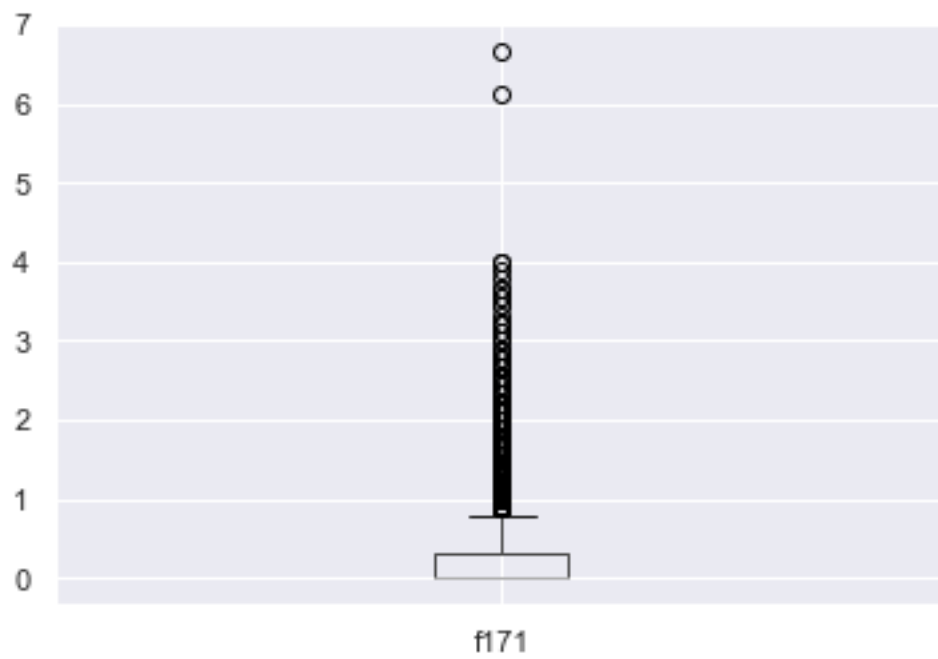


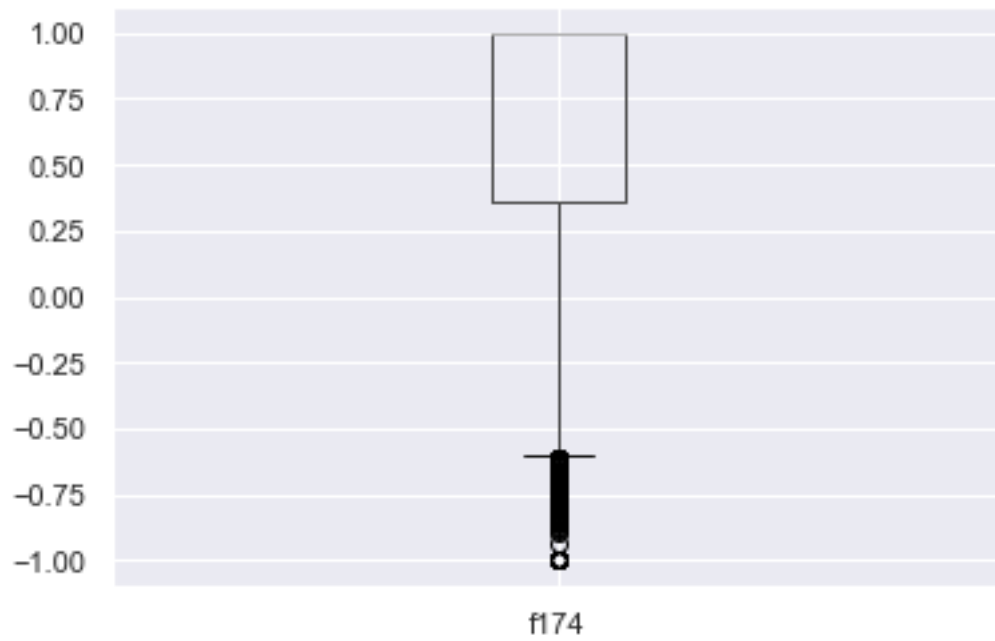












```
[8]: # descriptive statistics of attributes
print(df.describe(include='all'))
```

label	f1	f2	f3	f4	f5	\
-------	----	----	----	----	----	---

count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	4.062	-15.144	-24.026	-15.400	-18.594	-14.493
std	1.605	3.504	4.054	3.268	3.712	3.189
min	1.000	-23.971	-34.308	-23.161	-27.245	-22.103
25%	3.000	-17.848	-27.119	-17.563	-21.449	-16.989
50%	4.000	-15.992	-25.064	-16.164	-19.588	-15.099
75%	6.000	-11.786	-20.387	-13.427	-15.650	-11.735
max	7.000	2.536	-7.589	1.104	0.775	-0.887

	f6	f7	f8	f9	f10	f11 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	-18.427	0.256	-8.882	-8.626	-0.166	4.100
std	3.676	1.682	1.204	2.003	0.596	1.500
min	-26.897	-5.583	-16.389	-16.558	-2.717	-4.282
25%	-21.450	-1.038	-9.659	-10.070	-0.526	3.141
50%	-19.282	0.309	-8.916	-8.733	-0.175	4.079
75%	-15.514	1.558	-8.190	-7.158	0.172	5.090
max	0.952	7.030	-1.175	-0.925	3.785	11.212

	f12	f13	f14	f15	f16	f17 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	3.934	0.480	0.064	0.456	0.218	0.556
std	1.539	0.084	0.020	0.093	0.042	0.082
min	-4.498	0.209	0.016	0.162	0.065	0.154
25%	2.933	0.415	0.049	0.383	0.190	0.503
50%	3.958	0.483	0.061	0.450	0.217	0.556
75%	4.880	0.548	0.074	0.529	0.245	0.612
max	10.452	0.819	0.272	0.756	0.437	0.854

	f18	f19	f20	f21	f22	f23 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.227	-1.517	-1.609	-1.181	-1.593	-0.309
std	0.045	0.644	0.622	0.520	0.668	0.155
min	0.076	-6.793	-7.757	-7.643	-7.677	-2.051
25%	0.198	-1.836	-1.917	-1.436	-1.957	-0.386
50%	0.224	-1.424	-1.496	-1.083	-1.510	-0.277
75%	0.256	-1.084	-1.178	-0.837	-1.143	-0.203
max	0.434	-0.210	-0.374	-0.124	-0.231	-0.028

	f24	f25	f26	f27	f28	f29 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	-1.487	0.058	0.025	0.005	0.715	0.703
std	0.635	0.054	0.020	0.005	0.067	0.105
min	-7.384	0.006	0.003	0.000	0.307	0.145
25%	-1.786	0.023	0.011	0.002	0.679	0.637
50%	-1.377	0.035	0.016	0.003	0.718	0.710
75%	-1.059	0.078	0.036	0.007	0.762	0.778
max	-0.200	2.445	0.763	0.113	0.929	0.969

	f30	f31	f32	f33	f34	f35	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	41.669	0.500	0.215	0.202	0.082	0.083	
std	5.776	0.076	0.088	0.060	0.030	0.038	
min	17.538	0.102	0.020	0.024	0.011	0.007	
25%	37.920	0.447	0.157	0.165	0.059	0.056	
50%	41.985	0.506	0.199	0.202	0.078	0.076	
75%	45.504	0.563	0.260	0.243	0.100	0.099	
max	73.211	0.649	0.665	0.522	0.400	0.357	

	f36	f37	f38	f39	f40	f41	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	0.210	-14.493	-16.191	-24.026	0.143	0.128	
std	0.084	3.189	3.664	4.054	0.054	0.060	
min	0.022	-22.103	-24.520	-34.308	0.057	0.045	
25%	0.151	-16.989	-19.112	-27.119	0.101	0.082	
50%	0.194	-15.099	-17.046	-25.064	0.125	0.104	
75%	0.248	-11.735	-13.388	-20.387	0.183	0.164	
max	0.690	-0.887	3.555	-7.589	0.625	1.055	

	f42	f43	f44	f45	f46	f47	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	0.006	0.035	0.029	0.025	0.037	0.029	
std	0.005	0.028	0.038	0.024	0.027	0.035	
min	0.000	0.006	0.006	0.006	0.006	0.006	
25%	0.002	0.016	0.009	0.008	0.017	0.010	
50%	0.005	0.025	0.015	0.013	0.027	0.016	
75%	0.009	0.047	0.032	0.036	0.051	0.036	
max	0.064	0.423	1.987	0.640	0.484	1.971	

	f48	f49	f50	f51	f52	f53	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	0.006	0.021	-17.033	-26.718	-18.747	-21.423	
std	0.001	0.022	3.568	4.046	3.181	3.505	
min	0.000	0.006	-24.812	-35.604	-26.894	-28.954	
25%	0.006	0.006	-19.843	-30.148	-20.944	-24.245	
50%	0.006	0.010	-18.036	-27.312	-19.225	-21.941	
75%	0.006	0.030	-14.484	-23.788	-17.326	-19.261	
max	0.057	0.551	2.982	-8.711	-2.500	-1.481	

	f54	f55	f56	f57	f58	f59	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	-16.771	-21.236	1.714	-9.685	-7.970	-0.187	
std	3.250	3.613	1.744	1.599	1.885	0.575	
min	-24.015	-28.315	-4.367	-18.304	-14.727	-4.721	
25%	-19.258	-24.232	0.240	-10.706	-9.067	-0.568	
50%	-17.750	-21.583	1.549	-9.749	-7.860	-0.217	

75%	-14.935	-19.010	3.228	-8.629	-6.575	0.170
max	0.444	-1.133	9.722	-1.760	0.779	4.729

	f60	f61	f62	f63	f64	f65 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	4.652	4.464	0.557	0.062	0.381	0.202
std	1.070	1.053	0.089	0.020	0.089	0.032
min	-1.471	-1.516	0.243	0.011	0.094	0.078
25%	3.993	3.738	0.481	0.048	0.300	0.183
50%	4.601	4.418	0.555	0.062	0.387	0.202
75%	5.293	5.126	0.630	0.074	0.454	0.220
max	10.230	9.245	0.885	0.279	0.681	0.400

	f66	f67	f68	f69	f70	f71 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.586	0.211	-2.084	-1.259	-1.976	-2.266
std	0.056	0.031	0.681	0.370	0.627	0.690
min	0.262	0.098	-8.565	-5.417	-8.266	-8.127
25%	0.549	0.190	-2.467	-1.474	-2.340	-2.651
50%	0.584	0.211	-2.011	-1.208	-1.905	-2.163
75%	0.622	0.232	-1.593	-1.004	-1.510	-1.744
max	0.824	0.395	-0.573	-0.248	-0.581	-0.558

	f72	f73	f74	f75	f76	f77 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	-1.884	-1.478	0.033	0.015	0.003	0.719
std	0.599	0.668	0.035	0.015	0.004	0.057
min	-7.913	-7.232	0.005	0.002	0.000	0.300
25%	-2.218	-1.877	0.013	0.006	0.001	0.686
50%	-1.788	-1.385	0.019	0.009	0.002	0.723
75%	-1.461	-0.928	0.038	0.015	0.004	0.757
max	-0.591	-0.144	2.047	0.510	0.124	0.969

	f78	f79	f80	f81	f82	f83 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.660	41.743	0.474	0.246	0.187	0.094
std	0.102	6.021	0.075	0.081	0.052	0.031
min	0.098	16.952	0.058	0.036	0.005	0.026
25%	0.590	37.074	0.412	0.190	0.152	0.069
50%	0.667	42.301	0.477	0.237	0.181	0.089
75%	0.734	46.773	0.536	0.296	0.220	0.117
max	0.921	62.363	0.641	0.812	0.501	0.365

	f84	f85	f86	f87	f88	f89 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.090	0.232	-16.771	-19.029	-26.718	0.111
std	0.032	0.074	3.250	3.513	4.046	0.048
min	0.009	0.029	-24.015	-26.255	-35.604	0.047

25%	0.067	0.179	-19.258	-21.923	-30.148	0.077
50%	0.088	0.230	-17.750	-19.450	-27.312	0.092
75%	0.107	0.273	-14.935	-16.912	-23.788	0.127
max	0.525	0.932	0.444	1.589	-8.711	0.719

	f90	f91	f92	f93	f94	f95	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	0.091	0.005	0.025	0.013	0.014	0.025	
std	0.042	0.005	0.025	0.017	0.015	0.025	
min	0.036	0.000	0.004	0.004	0.004	0.004	
25%	0.060	0.002	0.010	0.005	0.004	0.011	
50%	0.079	0.004	0.014	0.008	0.008	0.015	
75%	0.108	0.007	0.029	0.013	0.017	0.029	
max	0.809	0.076	0.727	1.648	0.512	0.709	

	f96	f97	f98	f99	f100	f101	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	0.014	0.004	0.011	0.005	0.053	0.037	
std	0.017	0.001	0.013	0.007	0.025	0.023	
min	0.004	0.000	0.004	0.001	0.001	0.001	
25%	0.005	0.004	0.004	0.001	0.034	0.017	
50%	0.008	0.004	0.006	0.001	0.044	0.032	
75%	0.015	0.004	0.014	0.009	0.067	0.054	
max	1.590	0.055	0.394	0.279	0.341	0.321	

	f102	f103	f104	f105	f106	f107	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	0.127	0.417	0.826	18.178	1.702	0.586	
std	0.039	0.119	0.115	20.225	0.843	0.154	
min	0.001	0.001	-0.018	0.965	0.071	-0.029	
25%	0.098	0.322	0.757	7.233	1.234	0.470	
50%	0.111	0.423	0.846	12.000	1.500	0.634	
75%	0.148	0.497	0.924	25.381	2.000	0.702	
max	0.342	0.752	0.996	556.000	34.000	1.020	

	f108	f109	f110	f111	f112	f113	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	0.708	0.584	0.219	0.769	0.630	0.523	
std	0.170	0.128	0.160	0.095	0.199	0.110	
min	-0.188	-0.013	-0.867	-0.127	-0.034	-0.750	
25%	0.590	0.495	0.105	0.702	0.478	0.452	
50%	0.729	0.632	0.200	0.771	0.682	0.520	
75%	0.858	0.678	0.333	0.852	0.778	0.622	
max	0.996	0.873	0.943	0.996	1.187	0.758	

	f114	f115	f116	f117	f118	f119	\
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000	
mean	3.411	-0.433	25.317	0.581	0.065	6.025	

std	0.991	0.085	8.941	0.152	0.157	1.866
min	0.143	-0.976	-0.860	-0.042	-5.889	-2.540
25%	2.648	-0.496	17.760	0.457	0.015	4.800
50%	3.168	-0.431	27.730	0.586	0.087	5.620
75%	4.293	-0.366	32.200	0.712	0.130	7.080
max	7.273	0.116	51.500	0.982	0.279	12.340

	f120	f121	f122	f123	f124	f125 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	4.694	25.880	1.263	0.646	2.399	0.922
std	3.730	7.339	3.450	0.182	5.784	0.813
min	0.920	1.556	0.000	0.012	0.000	0.000
25%	2.680	20.111	0.222	0.533	0.444	0.444
50%	3.833	25.889	0.444	0.667	0.889	0.667
75%	5.941	30.556	0.988	0.778	1.889	1.111
max	111.000	45.000	123.560	1.000	169.780	11.000

	f126	f127	f128	f129	f130	f131 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	1.414	0.301	0.279	3.904	0.139	0.881
std	0.454	0.167	0.404	1.932	0.487	0.124
min	-0.000	0.111	-1.000	0.000	0.000	0.018
25%	1.149	0.185	0.000	2.000	0.000	0.778
50%	1.465	0.259	0.335	4.000	0.099	0.889
75%	1.735	0.358	0.601	5.556	0.222	1.000
max	2.197	1.000	1.000	20.000	125.730	1.000

	f132	f133	f134	f135	f136	f137 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.296	0.246	0.579	0.670	0.536	6155.009
std	0.905	0.286	0.524	0.287	0.501	245.682
min	0.000	0.000	-0.000	0.111	-1.000	5495.000
25%	0.000	0.000	-0.000	0.407	0.158	5959.000
50%	0.222	0.222	0.530	0.654	0.632	6098.000
75%	0.444	0.444	1.003	1.000	1.000	6368.000
max	290.330	14.111	2.197	1.000	1.000	14311.000

	f138	f139	f140	f141	f142	f143 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	5429.190	3224.351	4651.873	10335.964	0.513	3.302
std	582.647	572.939	746.884	2262.584	0.117	0.869
min	4310.000	2219.000	2462.000	3050.000	-0.101	0.817
25%	5011.000	2847.000	4129.000	8827.000	0.477	2.826
50%	5214.000	3091.000	4415.000	9872.000	0.529	3.247
75%	5800.000	3404.000	5254.000	11906.000	0.586	3.827
max	12687.000	10768.000	8734.000	18281.000	0.750	6.994

	f144	f145	f146	f147	f148	f149 \
--	------	------	------	------	------	--------

count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	1.709	-1.166	0.949	0.769	0.258	0.300
std	0.180	1.970	0.187	0.176	0.054	0.090
min	1.030	-502.770	0.003	-0.151	0.015	-0.273
25%	1.621	-1.343	0.895	0.716	0.237	0.263
50%	1.715	-0.977	0.985	0.794	0.263	0.309
75%	1.831	-0.799	1.063	0.879	0.294	0.356
max	2.293	912.600	1.980	1.125	0.393	0.530

	f150	f151	f152	f153	f154	f155 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.811	0.372	2.235	0.372	519341.312	0.181
std	0.127	0.085	0.400	0.085	169918.820	0.067
min	-0.035	-0.020	0.961	-0.020	6390.000	-0.104
25%	0.798	0.339	2.026	0.339	419620.000	0.148
50%	0.837	0.379	2.220	0.379	511300.000	0.175
75%	0.877	0.419	2.441	0.419	631180.000	0.219
max	0.956	0.585	3.819	0.585	1105900.000	0.416

	f156	f157	f158	f159	f160	f161 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	4939.284	173844.894	1.459	44.453	0.480	0.756
std	1758.171	54569.523	0.203	5.066	1.264	0.165
min	-4178.200	-45280.000	0.811	27.000	0.000	0.028
25%	3933.000	143760.000	1.347	40.889	0.099	0.667
50%	4624.100	163940.000	1.424	45.778	0.222	0.778
75%	6116.150	210115.000	1.562	47.889	0.444	0.889
max	10862.000	365400.000	2.422	60.889	93.802	1.000

	f162	f163	f164	f165	f166	f167 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.930	0.554	1.075	0.434	0.329	2.171
std	2.121	0.491	0.542	0.246	0.442	1.285
min	0.000	0.000	-0.000	0.111	-1.000	0.000
25%	0.222	0.222	0.687	0.259	0.000	1.000
50%	0.444	0.444	1.149	0.358	0.357	2.000
75%	0.889	0.667	1.465	0.506	0.661	3.000
max	134.890	9.333	2.197	1.000	1.000	12.444

	f168	f169	f170	f171	f172	f173 \
count	325834.000	325834.000	325834.000	325834.000	325834.000	325834.000
mean	0.080	0.920	0.174	0.163	0.403	0.765
std	0.183	0.106	0.374	0.222	0.480	0.274
min	0.000	0.106	0.000	0.000	-0.000	0.111
25%	0.000	0.833	0.000	0.000	-0.000	0.506
50%	0.000	1.000	0.000	0.000	-0.000	1.000
75%	0.173	1.000	0.333	0.333	0.849	1.000
max	25.951	1.000	66.667	6.667	2.197	1.000

```

                f174
count  325834.000
mean    0.668
std     0.471
min     -1.000
25%     0.357
50%     1.000
75%     1.000
max      1.000

```

It was decided not to drop any features or remove any outliers. Outliers were not removed because in remote sensing each value of attribute corresponds to a pixel in the output remote sensing image. In essence by removing that value you would not be able to classify that pixel.

0.2 Split Dataset

```

[10]: # prepare data for train/test split
X = df.drop("label", 1)
y = df['label']

```

```

[11]: # Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
→random_state=0)

```

```

[12]: # standardize the X dataset
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

```

0.3 Reduce Dimensions Using Principal Component Analysis

```

[13]: # run base PCA using all components
from sklearn.decomposition import PCA

pca = PCA()
X_train_pca = pca.fit_transform(X_train)
X_test_pca = pca.transform(X_test)

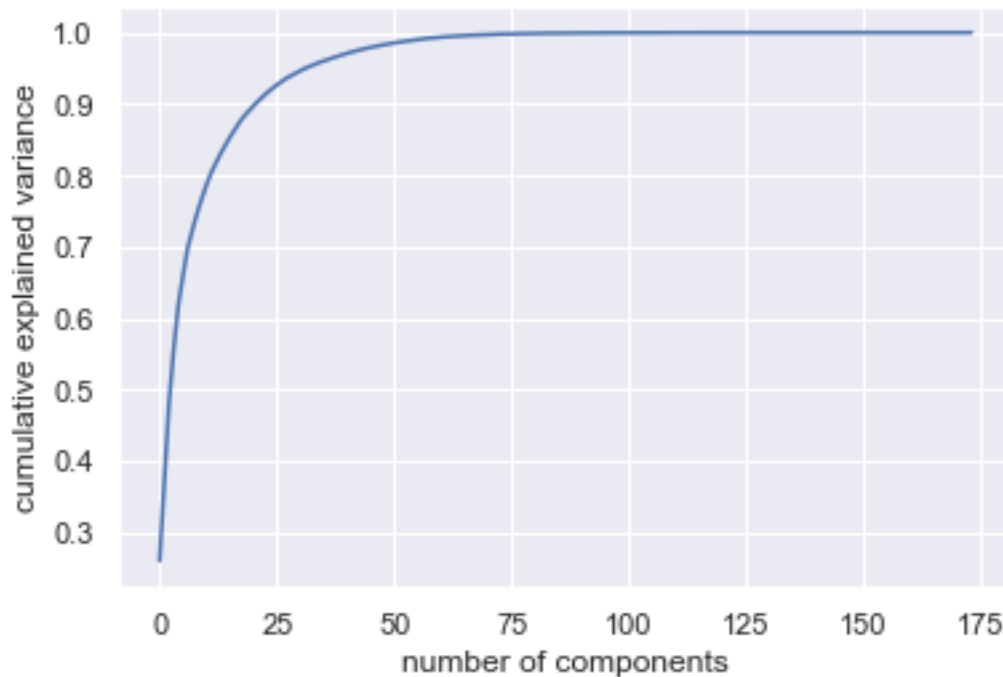
```

```

[14]: # plot variance explained by number of components in PCA
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')

```

```
[14]: Text(0, 0.5, 'cumulative explained variance')
```



```
[15]: # 11 components were determined to give the highest accuracy when running base_
      ↪ RandomForestClassifier
pca = PCA(n_components = 11)
X_train_pca11 = pca.fit_transform(X_train)
X_test_pca11 = pca.transform(X_test)
```

0.4 Fine Tune Hyperparameters of Random Forest Classifier

```
[16]: # next sections uses random grid and search to fine tune hyperparameters of_
      ↪ RandomForest classifier

n_estimators = [5,20,50,100] # number of trees in the random forest
max_features = ['auto', 'sqrt'] # number of features in consideration at every_
      ↪ split
max_depth = [int(x) for x in np.linspace(10, 120, num = 12)] # maximum number_
      ↪ of levels allowed in each decision tree
min_samples_split = [2, 6, 10] # minimum sample number to split a node
min_samples_leaf = [1, 3, 4] # minimum sample number that can be stored in a_
      ↪ leaf node
bootstrap = [True, False] # method used to sample data points

random_grid = {'n_estimators': n_estimators,
```

```

'max_features': max_features,

'max_depth': max_depth,

'min_samples_split': min_samples_split,

'min_samples_leaf': min_samples_leaf,

'bootstrap': bootstrap}

```

```

[17]: # create RandomForest classifier
from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier()

```

```

[18]: # create RandomizedSearch object with parameters
from sklearn.model_selection import RandomizedSearchCV
classifier_random = RandomizedSearchCV(estimator =
    ↳ classifier, param_distributions = random_grid,
        n_iter = 100, cv = 5, verbose=2, random_state=35, n_jobs = -1)

```

```

[ ]: # run the RandomizedSearch classifier on training data to determine best
    ↳ parameters for RandomForest classifier run
classifier_random.fit(X_train_pca11, y_train)

```

```

[ ]: # print random grid parameters
print ('Random grid: ', random_grid, '\n')
# print the best parameters
print ('Best Parameters: ', classifier_random.best_params_, ' \n')

```

0.5 Random Forest Classifier Model

```

[19]: # run the RandomForest classifier using fine tuned settings
from sklearn.ensemble import RandomForestClassifier

classifier_finetime = RandomForestClassifier(n_estimators=50,
    ↳ min_samples_split=6, min_samples_leaf=1, max_features='sqrt', max_depth=100,
    ↳ bootstrap=False, random_state=0)
classifier_finetime.fit(X_train_pca11, y_train)

# Predicting the Test set results
y_pred = classifier_finetime.predict(X_test_pca11)

```

```

[20]: # create confusion matrix and calculate accuracy and cohen's kappa
from sklearn.metrics import confusion_matrix

```

```

from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
from sklearn.metrics import cohen_kappa_score

cm = confusion_matrix(y_test, y_pred)
print(cm, '\n')
print(classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Cohen's Kappa:", cohen_kappa_score(y_test, y_pred))

```

```

[[11806      3      5     27      3      5      7]
 [      2 1065      1      2      0      0      0]
 [      2      2 22794     22      3     11      0]
 [     27      0     28 22219     32     30      0]
 [     25      0     12     55 13510     365      6]
 [     19      0     14     34     178 25089      3]
 [     15      3      0      0      1      3    323]]

```

	precision	recall	f1-score	support
1	0.99	1.00	0.99	11856
2	0.99	1.00	0.99	1070
3	1.00	1.00	1.00	22834
4	0.99	0.99	0.99	22336
5	0.98	0.97	0.98	13973
6	0.98	0.99	0.99	25337
7	0.95	0.94	0.94	345
accuracy			0.99	97751
macro avg	0.99	0.98	0.98	97751
weighted avg	0.99	0.99	0.99	97751

Accuracy: 0.9903325797178546

Cohen's Kappa: 0.9877709636057277

[]: