

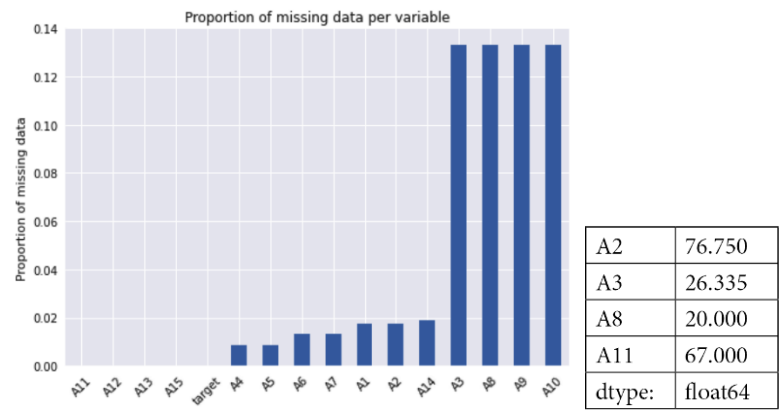
Chapter 1: Imputing Missing Data

← → ↻ Not secure archive.ics.uci.edu/ml/machine-learning-databases/credit-screening/

Index of /ml/machine-learning-databases/credit-screening

- Parent Directory
- Index
- credit.lisp
- credit.names
- crs.data
- crs.names

Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips SVN/1.7.14 Phusion_Passenger/4.0.53 mod_perl/2.0.10 Perl/v5.16.3 Server at archive.ics.uci.edu Port 80



	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	...	A13	A14	A15	A1_na	A3_na	A4_na	A5_na	A6_na	A7_na	A8_na
596	a	46.08	3.000	u	g	c	v	2.375	t	t	...	g	396.0	4159	0	0	0	0	0	0	0
303	a	15.92	2.875	u	g	q	v	0.085	f	f	...	g	120.0	0	0	0	0	0	0	0	0
204	b	36.33	2.125	y	p	w	v	0.085	t	t	...	g	50.0	1187	0	0	0	0	0	0	0
351	b	22.17	0.585	y	p	ff	ff	0.000	f	f	...	g	100.0	0	0	0	0	0	0	0	0
118	b	57.83	7.040	u	g	m	v	14.000	t	t	...	g	360.0	1332	0	0	0	0	0	0	0

Chapter 2: Encoding Categorical Variables

← → ⓘ Not secure | archive.ics.uci.edu/ml/machine-learning-databases/credit-screening/

Index of /ml/machine-learning-databases/credit-screening

- [Parent Directory](#)
- [index](#)
- [credit.data](#)
- [credit.names](#)
- [credit.names](#)

Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips SVN/1.7.14 Phusion_Passenger/4.0.53 mod_perl/2.0.10 Perl/v5.16.3 Server at archive.ics.uci.edu Port 80

Gender	Female	Male
Female	1	0
Male	0	1
Male	0	1
Female	1	0
Female	1	0

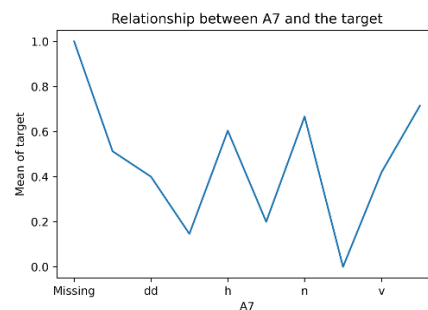
	A2	A3	A8	A11	A14	A15	A1_a	A1_b	A4_l	A4_u	...	A7_j	A7_n	A7_o	A7_v	A7_z	A9_t	A10_t	A12_t	A13_p	A13_s
596	46.08	3.000	2.375	8	396.0	4159	1	0	0	1	...	0	0	0	1	0	1	1	1	0	0
303	15.92	2.875	0.085	0	120.0	0	1	0	0	1	...	0	0	0	1	0	0	0	0	0	0
204	36.33	2.125	0.085	1	50.0	1187	0	1	0	0	...	0	0	0	1	0	1	1	0	0	0
351	22.17	0.585	0.000	0	100.0	0	0	1	0	0	...	0	0	0	0	0	0	0	0	0	0
118	57.83	7.040	14.000	6	360.0	1332	0	1	0	1	...	0	0	0	1	0	1	1	1	0	0

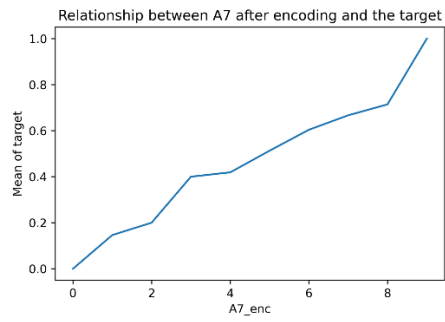
5 rows × 42 columns

```
[array(['Missing', 'a', 'b'], dtype=object),
array(['Missing', 'l', 'u', 'y'], dtype=object),
array(['Missing', 'g', 'gg', 'p'], dtype=object),
array(['Missing', 'aa', 'c', 'cc', 'd', 'e', 'ff', 'i', 'j', 'k', 'm',
      'q', 'r', 'w', 'x'], dtype=object),
array(['Missing', 'bb', 'dd', 'ff', 'h', 'j', 'n', 'o', 'v', 'z'],
      dtype=object),
array(['f', 't'], dtype=object),
array(['f', 't'], dtype=object),
array(['f', 't'], dtype=object),
array(['g', 'p', 's'], dtype=object)]
```

```
array(['A1_a', 'A1_b', 'A4_l', 'A4_u', 'A4_y', 'A5_g', 'A5_gg', 'A5_p',
      'A6_aa', 'A6_c', 'A6_cc', 'A6_d', 'A6_e', 'A6_ff', 'A6_i', 'A6_j',
      'A6_k', 'A6_m', 'A6_q', 'A6_r', 'A6_w', 'A6_x', 'A7_bb', 'A7_dd',
      'A7_ff', 'A7_h', 'A7_j', 'A7_n', 'A7_o', 'A7_v', 'A7_z', 'A9_t',
      'A10_t', 'A12_t', 'A13_p', 'A13_s'], dtype=object)
```

	A2	A3	A8	A11	A14	A15	A1_a	A1_b	A4_u	A4_y	...	A7_z	A7_bb	A7_j	A7_Missing	A7_n	A9_t	A10_t	A12_t	A13_g	A13_s
596	46.08	3.000	2.375	8	396.0	4159	1	0	1	0	...	0	0	0	0	0	1	1	1	1	0
303	15.92	2.875	0.085	0	120.0	0	1	0	1	0	...	0	0	0	0	0	0	0	0	1	0
204	36.33	2.125	0.085	1	50.0	1187	0	1	0	1	...	0	0	0	0	0	1	1	0	1	0
351	22.17	0.585	0.000	0	100.0	0	0	1	0	1	...	0	0	0	0	0	0	0	0	1	0
118	57.83	7.040	14.000	6	360.0	1332	0	1	1	0	...	0	0	0	0	0	1	1	1	1	0

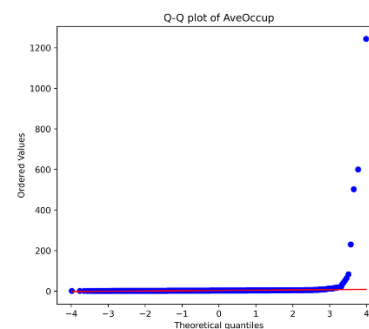
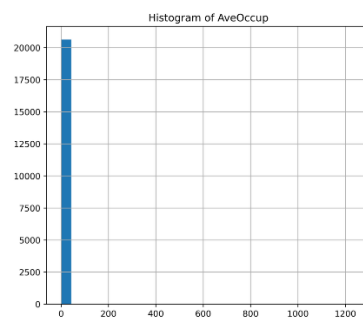
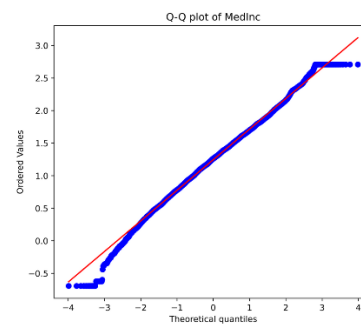
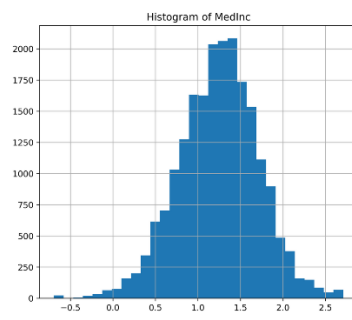
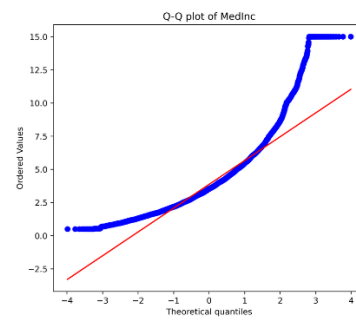
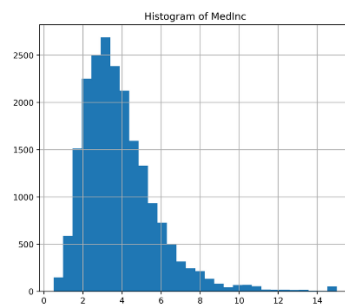
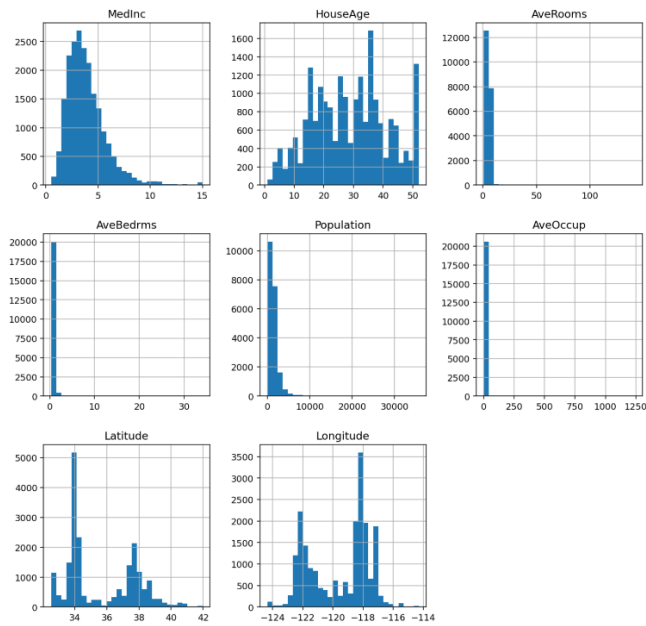


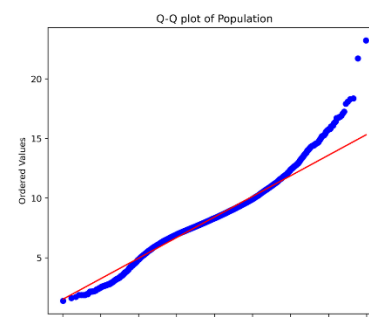
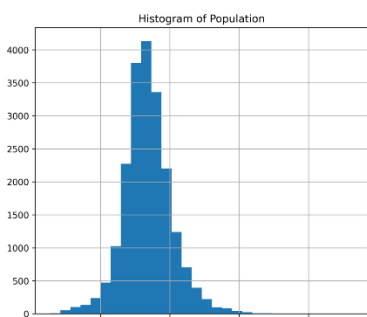
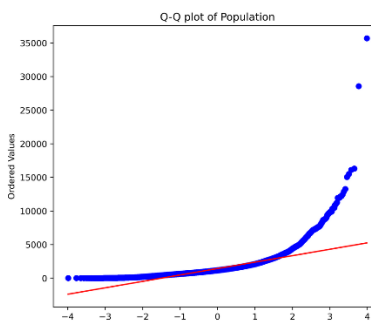
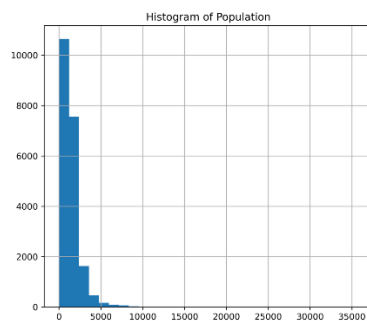
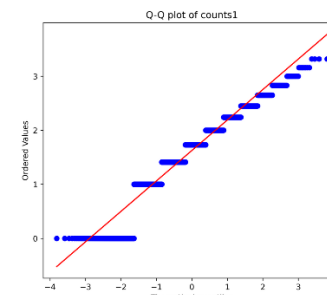
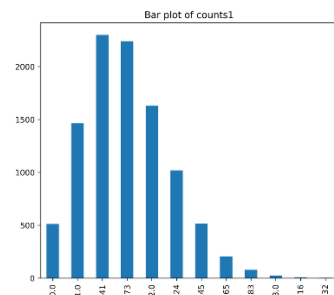
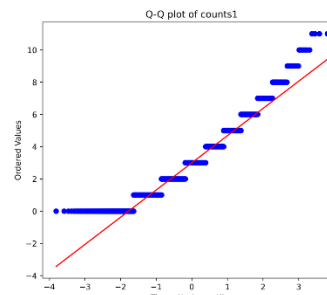
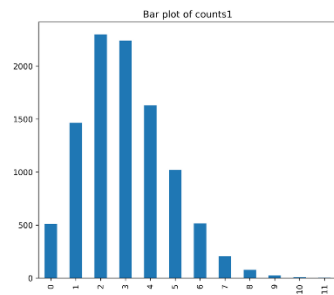
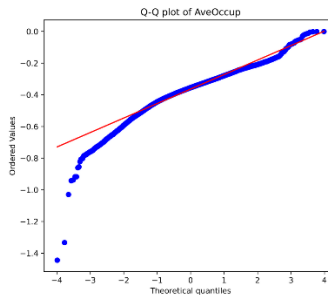
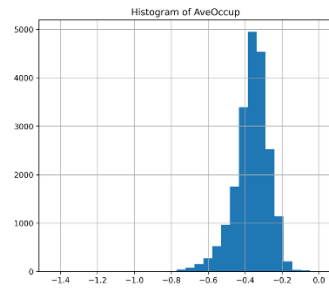


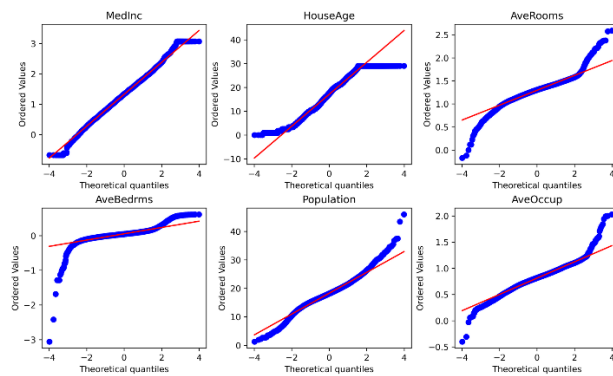
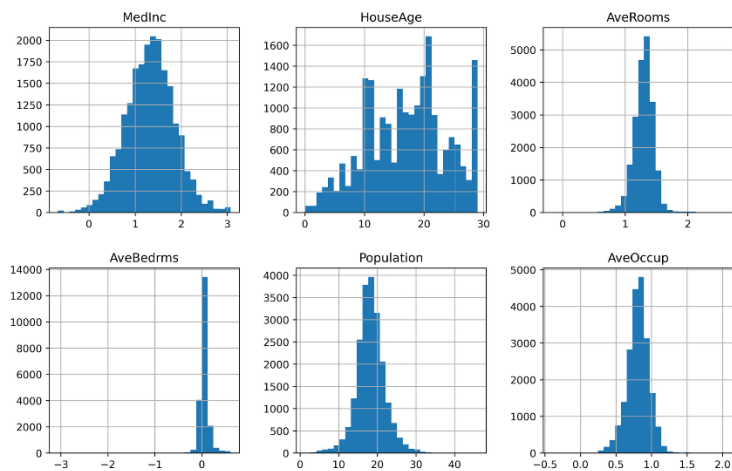
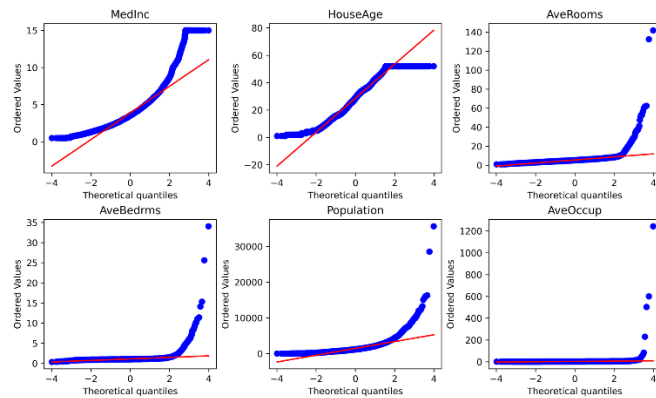
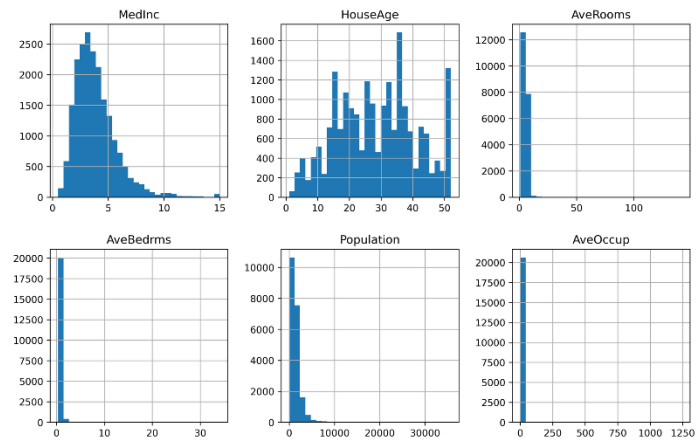
Color	Intermediate step	1st	2nd
Blue	1	1	0
Red	2	0	1
Green	3	1	1
Yellow	0	0	0

	A1	A2	A3	A4	A5	A6	A7_0	A7_1	A7_2	A7_3	A8	A9	A10	A11	A12	A13	A14	A15
596	a	46.08	3.000	u	g	c	0	0	0	1	2.375	t	t	8	t	g	396.0	4159
303	a	15.92	2.875	u	g	q	0	0	0	1	0.085	f	f	0	f	g	120.0	0
204	b	36.33	2.125	y	p	w	0	0	0	1	0.085	t	t	1	f	g	50.0	1187
351	b	22.17	0.585	y	p	ff	0	0	1	0	0.000	f	f	0	f	g	100.0	0
118	b	57.83	7.040	u	g	m	0	0	0	1	14.000	t	t	6	t	g	360.0	1332

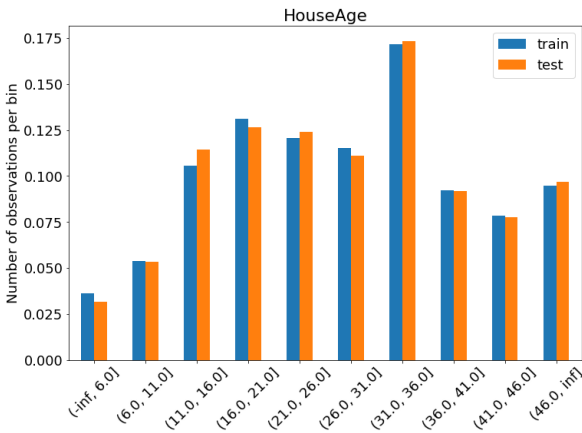
Chapter 3: Transforming Numerical Variables



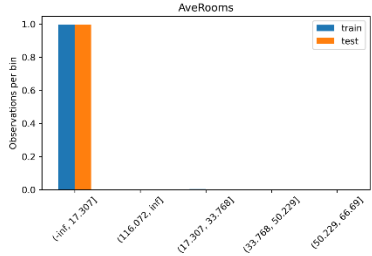
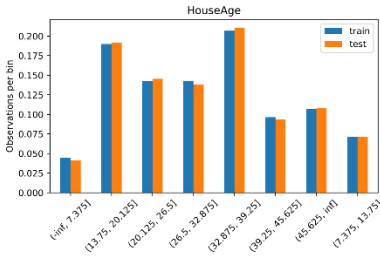
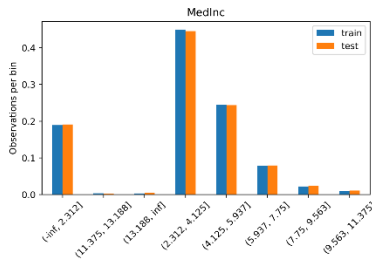


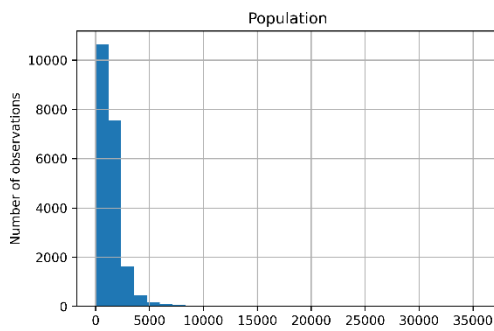
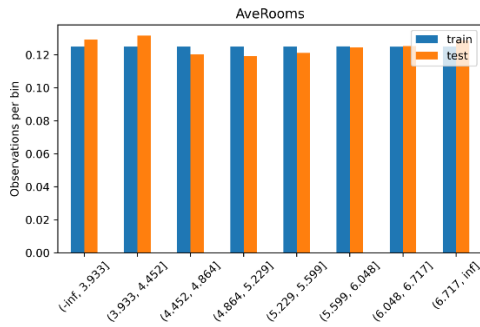
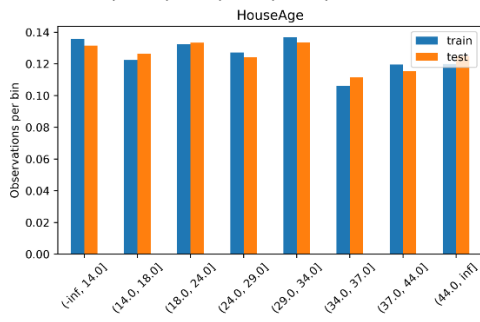
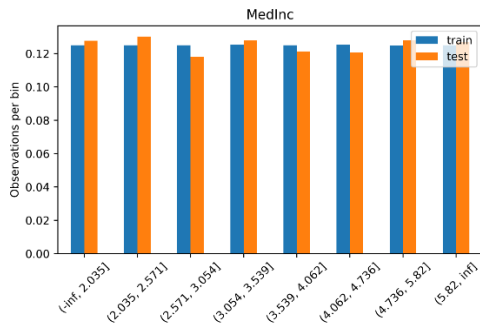
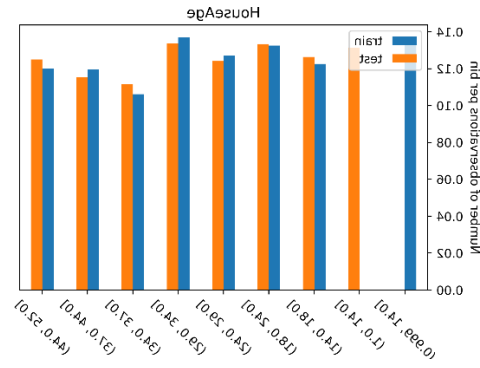


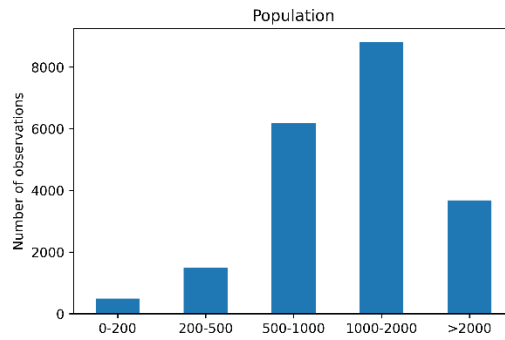
Chapter 4: Performing Variable Discretization



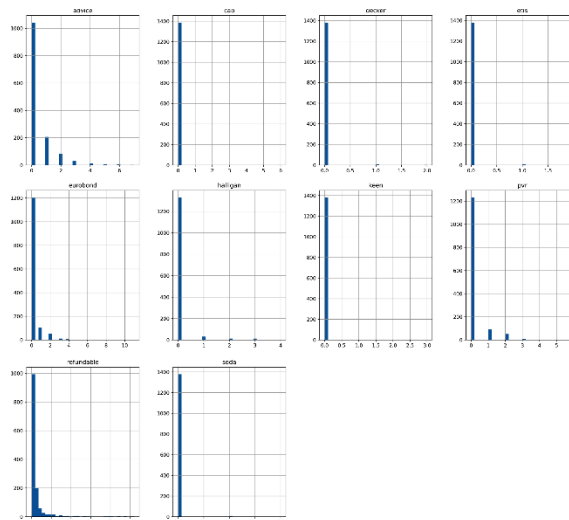
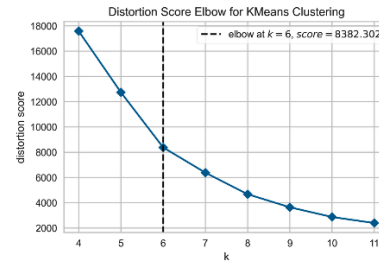
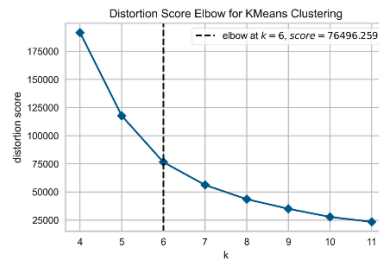
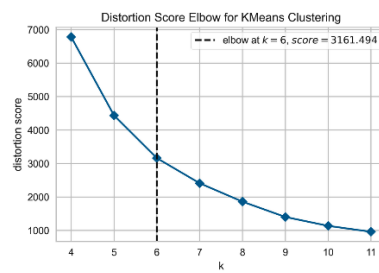
	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude
14740	(4.125, 5.937]	(20.125, 26.5]	(-inf, 17.307]	1.075472	1551.0	4.180593	32.58	-117.05
10101	(4.125, 5.937]	(26.5, 32.875]	(-inf, 17.307]	0.927739	1296.0	3.020979	33.92	-117.97
20566	(4.125, 5.937]	(26.5, 32.875]	(-inf, 17.307]	1.026217	1554.0	2.910112	38.65	-121.84
2670	(2.312, 4.125]	(32.875, 39.25]	(-inf, 17.307]	1.316901	390.0	2.746479	33.20	-115.60
15709	(4.125, 5.937]	(20.125, 26.5]	(-inf, 17.307]	1.039578	649.0	1.712401	37.79	-122.43

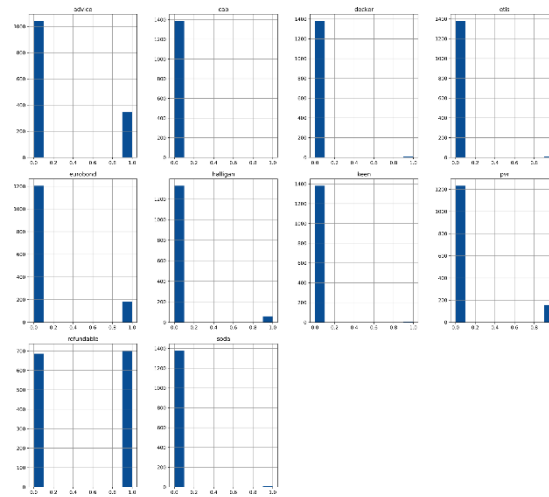




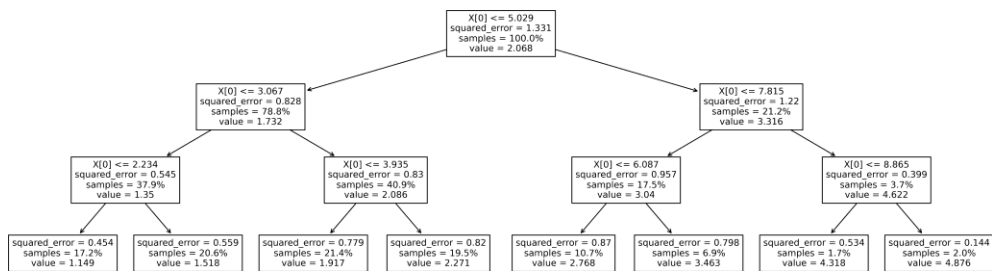
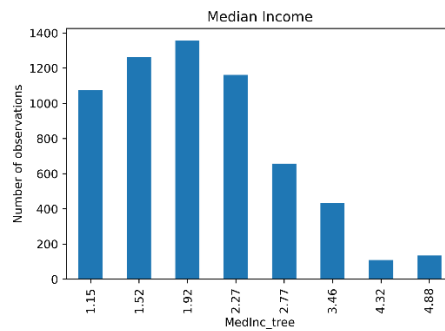
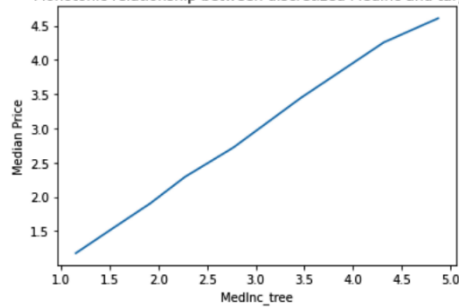


	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude
0	8.3252	41.0	6.984127	1.023810	(200.0, 500.0]	2.555556	37.88	-122.23
1	8.3014	21.0	6.238137	0.971880	(2000.0, inf]	2.109842	37.86	-122.22
2	7.2574	52.0	8.288136	1.073446	(200.0, 500.0]	2.802260	37.85	-122.24
3	5.6431	52.0	5.817352	1.073059	(500.0, 1000.0]	2.547945	37.85	-122.25
4	3.8462	52.0	6.281853	1.081081	(500.0, 1000.0]	2.181467	37.85	-122.25

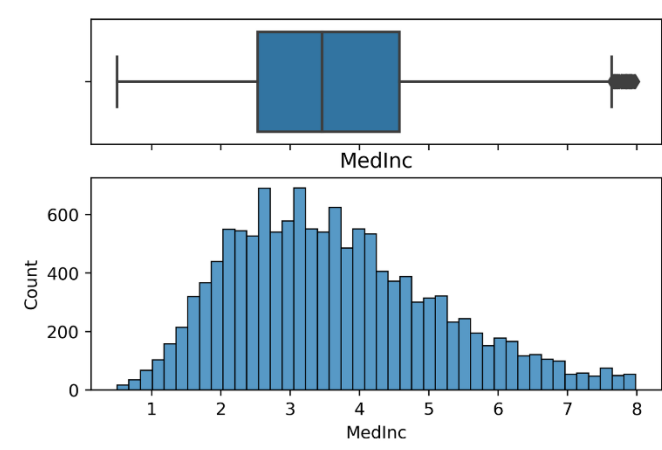
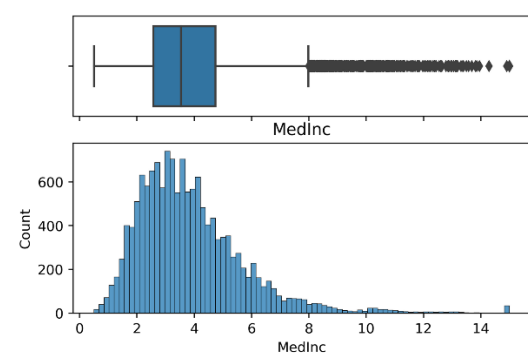
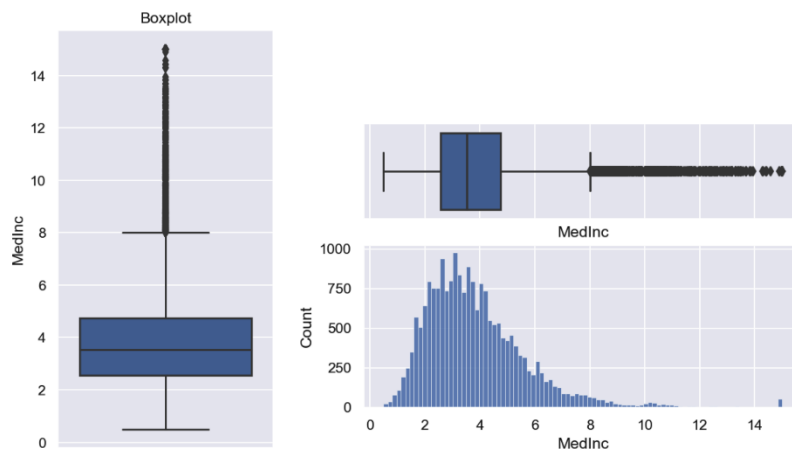




Monotonic relationship between discretized MedInc and target



Chapter 5: Working with Outliers



Chapter 6: Extracting Features from Date and Time

date		date		year	date		quarter	date		month
0	2019-03-05	0	2019-03-05	2019	0	2019-03-05	1	0	2019-03-05	3
1	2019-03-06	1	2019-03-06	2019	1	2019-03-06	1	1	2019-03-06	3
2	2019-03-07	2	2019-03-07	2019	2	2019-03-07	1	2	2019-03-07	3
3	2019-03-08	3	2019-03-08	2019	3	2019-03-08	1	3	2019-03-08	3
4	2019-03-09	4	2019-03-09	2019	4	2019-03-09	1	4	2019-03-09	3

date		week	date		day_mo	date		day_mo	day_week
0	2019-03-05	10	0	2019-03-05	5	0	2019-03-05	5	1
1	2019-03-06	10	1	2019-03-06	6	1	2019-03-06	6	2
2	2019-03-07	10	2	2019-03-07	7	2	2019-03-07	7	3
3	2019-03-08	10	3	2019-03-08	8	3	2019-03-08	8	4
4	2019-03-09	10	4	2019-03-09	9	4	2019-03-09	9	5

date	day_week	is_weekend	date		date	hour	min	sec
0	2019-03-05	1	0	2019-03-05	00:00:00	0	0	0
1	2019-03-06	2	1	2019-03-06	01:15:10	1	15	10
2	2019-03-07	3	2	2019-03-07	02:30:20	2	30	20
3	2019-03-08	4	3	2019-03-08	03:45:30	3	45	30
4	2019-03-09	5	4	2019-03-09	05:00:40	5	0	40

	date	hour	min	sec	h	m	s		date	hour	min	sec	h	m	s	is_morning	
0	2019-03-05	00:00:00	0	0	0	0	0	0	2019-03-05	00:00:00	0	0	0	0	0	0	
1	2019-03-05	01:15:10	1	15	10	1	15	10	2019-03-05	01:15:10	1	15	10	1	15	10	0
2	2019-03-05	02:30:20	2	30	20	2	30	20	2019-03-05	02:30:20	2	30	20	2	30	20	0
3	2019-03-05	03:45:30	3	45	30	3	45	30	2019-03-05	03:45:30	3	45	30	3	45	30	0
4	2019-03-05	05:00:40	5	0	40	5	0	40	2019-03-05	05:00:40	5	0	40	5	0	40	0

	date1	date2		date1	date2	elapsed_days		date1	date2	elapsed_days	months_passed	
0	2019-03-05 00:00:00	2019-03-31	0	2019-03-05 00:00:00	2019-03-31	26		0	2019-03-05 00:00:00	2019-03-31	26	1.0
1	2019-03-05 01:00:00	2019-04-30	1	2019-03-05 01:00:00	2019-04-30	55		1	2019-03-05 01:00:00	2019-04-30	55	2.0
2	2019-03-05 02:00:00	2019-05-31	2	2019-03-05 02:00:00	2019-05-31	86		2	2019-03-05 02:00:00	2019-05-31	86	3.0
3	2019-03-05 03:00:00	2019-06-30	3	2019-03-05 03:00:00	2019-06-30	116		3	2019-03-05 03:00:00	2019-06-30	116	4.0
4	2019-03-05 04:00:00	2019-07-31	4	2019-03-05 04:00:00	2019-07-31	147		4	2019-03-05 04:00:00	2019-07-31	147	5.0

	date1	date2	elapsed_days	months_passed	diff_seconds	diff_minutes
0	2019-03-05 00:00:00	2019-03-31	26	1.0	2246400.0	37440.0
1	2019-03-05 01:00:00	2019-04-30	55	2.0	4834800.0	80580.0
2	2019-03-05 02:00:00	2019-05-31	86	3.0	7509600.0	125160.0
3	2019-03-05 03:00:00	2019-06-30	116	4.0	10098000.0	168300.0
4	2019-03-05 04:00:00	2019-07-31	147	5.0	12772800.0	212880.0

date1	date2	elapsed_days	months_passed	diff_seconds	diff_minutes	to_today	time1	time2
0	2019-03-05 00:00:00	2019-03-31	26	1.0	2246400.0	37440.0	2015-06-10 09:00:00+02:00	2015-07-01 09:00:00+02:00
1	2019-03-05 01:00:00	2019-04-30	55	2.0	4834800.0	80580.0	2015-06-10 10:00:00+02:00	2015-07-01 10:00:00+02:00
2	2019-03-05 02:00:00	2019-05-31	86	3.0	7509600.0	125160.0	2015-06-10 11:00:00+02:00	2015-07-01 11:00:00+02:00
3	2019-03-05 03:00:00	2019-06-30	116	4.0	10098000.0	168300.0	2015-09-10 09:00:00-05:00	2015-08-01 09:00:00-05:00
4	2019-03-05 04:00:00	2019-07-31	147	5.0	12772800.0	212880.0	2015-09-10 10:00:00-05:00	2015-08-01 10:00:00-05:00

				time1_london	time2_berlin
time1	time2	time1_utc	time2_utc		
0	2015-06-10 09:00:00+02:00	2015-07-01 09:00:00+02:00	2015-06-10 07:00:00+00:00	2015-07-01 07:00:00+00:00	0 2015-06-10 08:00:00+01:00 2015-06-10 09:00:00+02:00
1	2015-06-10 10:00:00+02:00	2015-07-01 10:00:00+02:00	2015-06-10 08:00:00+00:00	2015-07-01 08:00:00+00:00	1 2015-06-10 09:00:00+01:00 2015-06-10 10:00:00+02:00
2	2015-06-10 11:00:00+02:00	2015-07-01 11:00:00+02:00	2015-06-10 09:00:00+00:00	2015-07-01 09:00:00+00:00	2 2015-06-10 10:00:00+01:00 2015-06-10 11:00:00+02:00
0	2015-09-10 09:00:00-05:00	2015-08-01 09:00:00-05:00	2015-09-10 14:00:00+00:00	2015-08-01 14:00:00+00:00	0 2015-09-10 15:00:00+01:00 2015-09-10 16:00:00+02:00
1	2015-09-10 10:00:00-05:00	2015-08-01 10:00:00-05:00	2015-09-10 15:00:00+00:00	2015-08-01 15:00:00+00:00	1 2015-09-10 16:00:00+01:00 2015-09-10 17:00:00+02:00
2	2015-09-10 11:00:00-05:00	2015-08-01 11:00:00-05:00	2015-09-10 16:00:00+00:00	2015-08-01 16:00:00+00:00	2 2015-09-10 17:00:00+01:00 2015-09-10 18:00:00+02:00

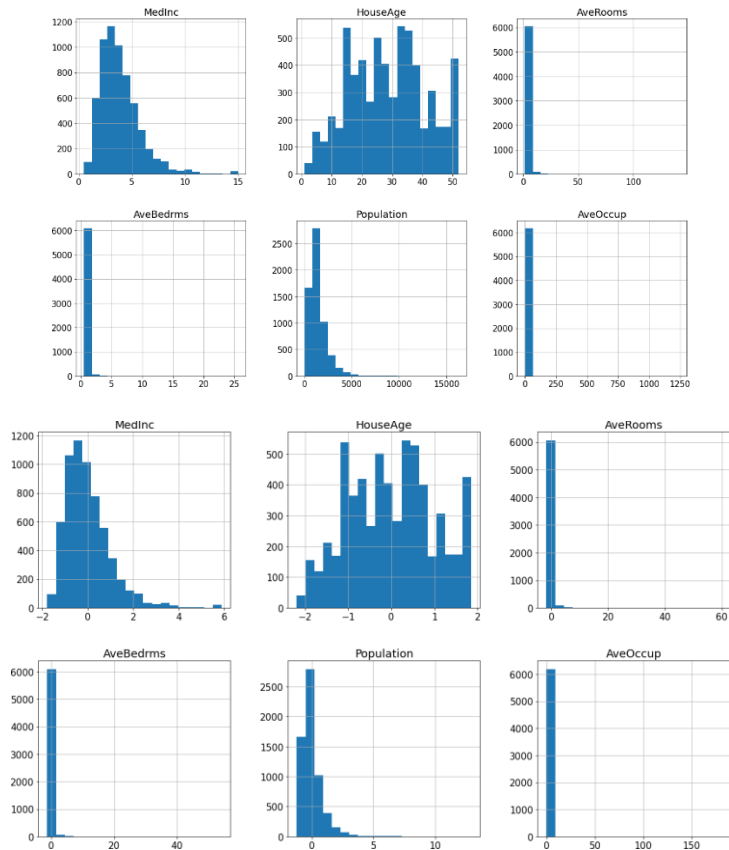
							time
date_month	date_year	date_day_of_week	date_day_of_month	date_hour	date_minute	date_second	
0	3	2019	1	5	0	0	0 2014-08-01 09:00:00+02:00
1	3	2019	2	6	0	0	1 2014-08-01 10:00:00+02:00
2	3	2019	3	7	0	0	2 2014-08-01 11:00:00+02:00
3	3	2019	4	8	0	0	0 2014-08-01 09:00:00-05:00
4	3	2019	5	9	0	0	1 2014-08-01 10:00:00-05:00
							2 2014-08-01 11:00:00-05:00

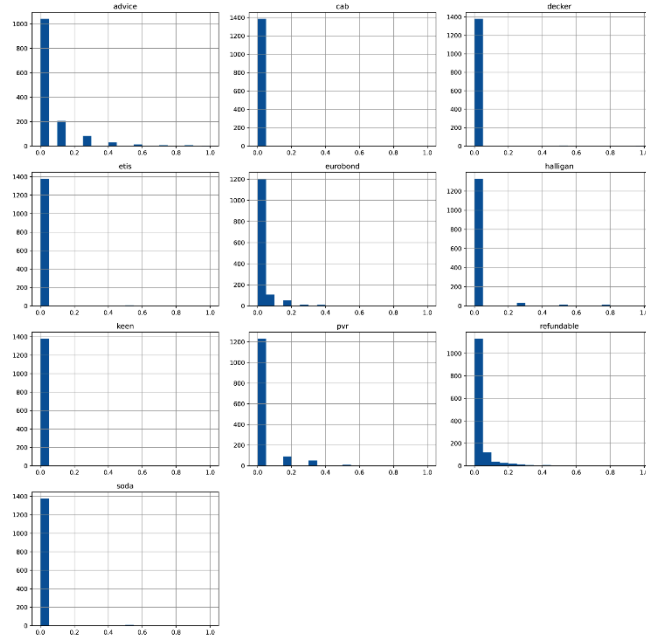
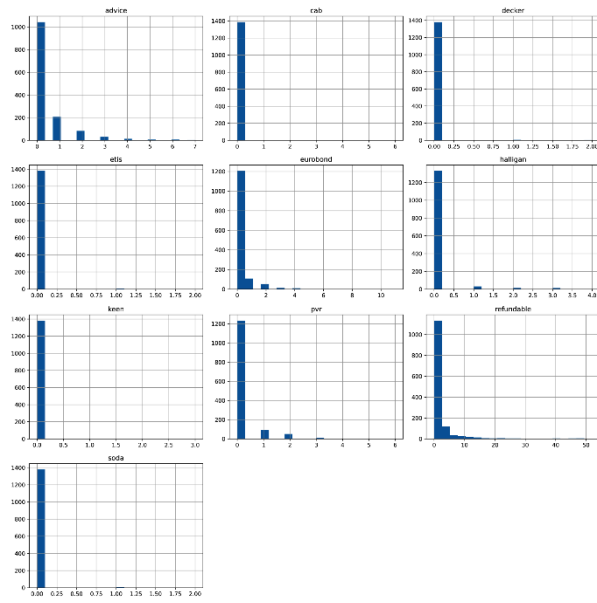
time	time_day_of_week	time_hour	time_minute
0	2014-08-01 09:00:00+02:00	4	7
1	2014-08-01 10:00:00+02:00	4	8
2	2014-08-01 11:00:00+02:00	4	9
0	2014-08-01 09:00:00-05:00	4	14
1	2014-08-01 10:00:00-05:00	4	15

Chapter 7: Performing Feature Scaling

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup
count	6192.000000	6192.000000	6192.000000	6192.000000	6192.000000	6192.000000
mean	3.880013	28.687984	5.442057	1.101109	1426.222061	3.140976
std	1.920007	12.560416	2.862733	0.519956	1091.567168	15.796292
min	0.499900	1.000000	1.465753	0.500000	8.000000	0.692308
25%	2.552150	18.000000	4.414452	1.006494	796.000000	2.436452
50%	3.529600	29.000000	5.227365	1.048741	1169.500000	2.825041
75%	4.768750	37.000000	6.064257	1.098434	1727.250000	3.285501
max	15.000100	52.000000	141.909091	25.636364	16305.000000	1243.333333

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup
count	6192.000000	6192.000000	6192.000000	6192.000000	6192.000000	6192.000000
mean	0.007057	0.005500	0.008154	0.013991	0.000926	0.014627
std	1.015290	0.997154	1.251446	1.148474	0.949568	2.300008
min	-1.780329	-2.192612	-1.730090	-1.313734	-1.232803	-0.341909
25%	-0.695110	-0.843004	-0.441064	-0.194995	-0.547312	-0.087955
50%	-0.178240	0.030271	-0.085698	-0.101679	-0.222400	-0.031374
75%	0.477017	0.665380	0.280150	0.008082	0.262794	0.035671
max	5.887302	1.856210	59.664826	54.207251	12.944167	180.591967



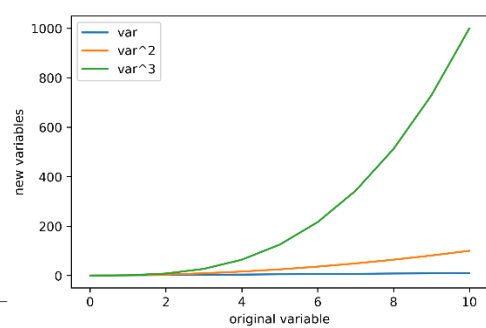
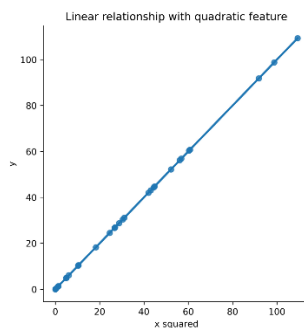
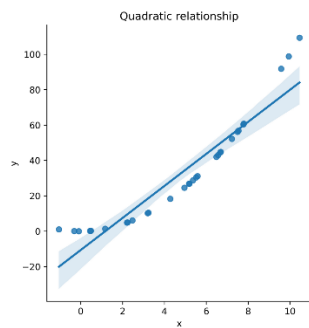


Chapter 8: Creating New Features

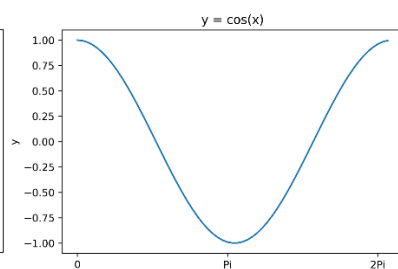
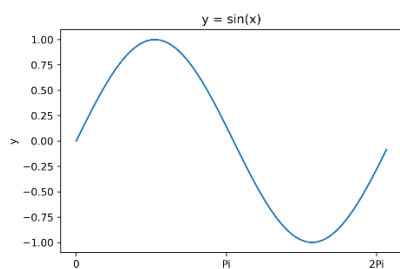
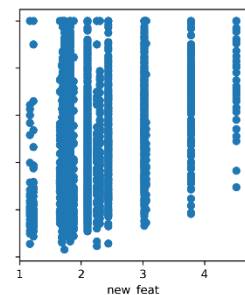
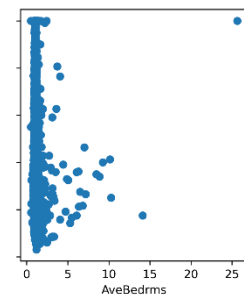
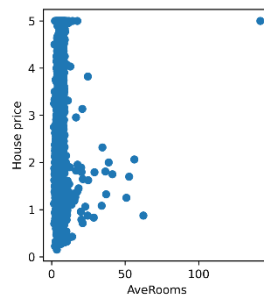
	sum	prod	mean	std	max	min
0	1.08510	0.000351	0.21702	0.080321	0.3001	0.11840
1	0.50165	0.000007	0.10033	0.045671	0.1812	0.07017
2	0.80170	0.000092	0.16034	0.042333	0.2069	0.10960
3	1.03270	0.000267	0.20654	0.078097	0.2839	0.10520
4	0.71630	0.000050	0.14326	0.044402	0.1980	0.10030

	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	sum_f	prod_f	mean_f	std_f	max_f	min_f
0	0.11840	0.27760	0.3001	0.14710	0.2419	1.08510	0.000351	0.21702	0.080321	0.3001	0.11840
1	0.08474	0.07864	0.0869	0.07017	0.1812	0.50165	0.000007	0.10033	0.045671	0.1812	0.07017
2	0.10960	0.15990	0.1974	0.12790	0.2069	0.80170	0.000092	0.16034	0.042333	0.2069	0.10960
3	0.14250	0.28390	0.2414	0.10520	0.2597	1.03270	0.000267	0.20654	0.078097	0.2839	0.10520
4	0.10030	0.13280	0.1980	0.10430	0.1809	0.71630	0.000050	0.14326	0.044402	0.1980	0.10030

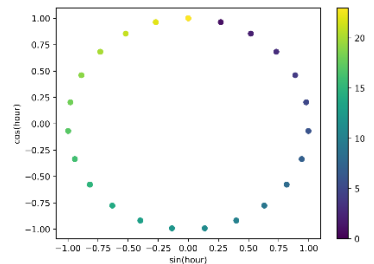
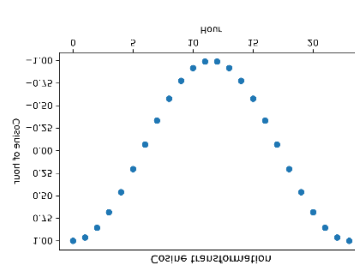
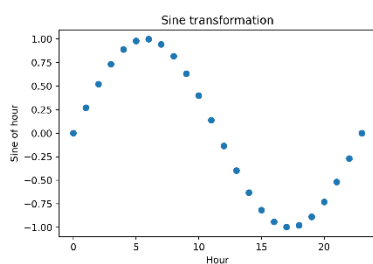
	mean smoothness_sub_mean radius	mean compactness_sub_mean radius	mean concavity_sub_mean radius	mean symmetry_sub_mean radius	mean smoothness_sub_mean area	mean compactness_sub_mean area
0	-17.87160	-17.71240	-17.6899	-17.7481	-1000.88160	-1000.72240
1	-20.48526	-20.49136	-20.4831	-20.3888	-1325.91526	-1325.92136
2	-19.58040	-19.53010	-19.4926	-19.4831	-1202.89040	-1202.84010
3	-11.27750	-11.13610	-11.1786	-11.1603	-385.95750	-385.81610
4	-20.18970	-20.15720	-20.0920	-20.1091	-1296.89970	-1296.86720



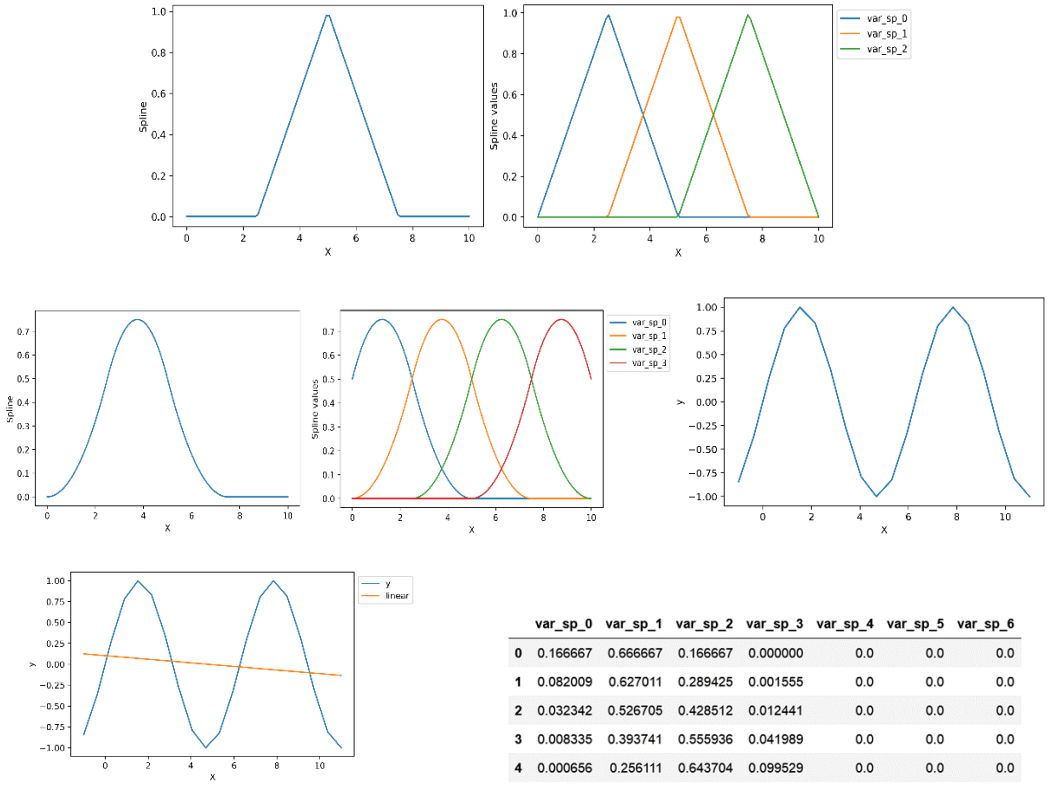
	var	col	feat	var col	var feat	col feat
0	0.0	0.0	0.0	0.0	0.0	0.00
1	1.0	0.5	0.5	0.5	0.5	0.25
2	2.0	1.0	1.0	2.0	2.0	1.00
3	3.0	1.5	1.5	4.5	4.5	2.25
4	4.0	2.0	2.0	8.0	8.0	4.00
5	5.0	2.5	2.5	12.5	12.5	6.25
6	6.0	3.0	3.0	18.0	18.0	9.00
7	7.0	3.5	3.5	24.5	24.5	12.25
8	8.0	4.0	4.0	32.0	32.0	16.00
9	9.0	4.5	4.5	40.5	40.5	20.25
10	10.0	5.0	5.0	50.0	50.0	25.00



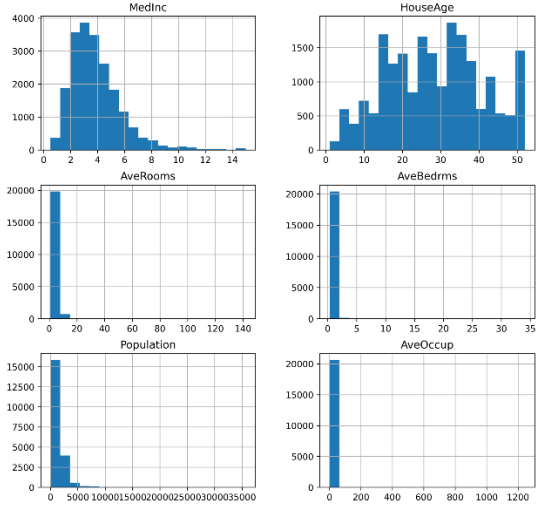
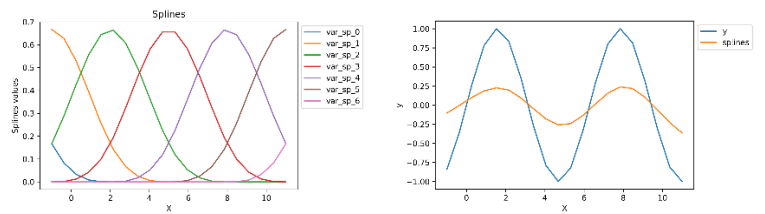
	hour	hour_sin	hour_cos
0	0	0.000000	1.000000
1	1	0.269797	0.962917
2	2	0.519584	0.854419
3	3	0.730836	0.682553
4	4	0.887885	0.460065



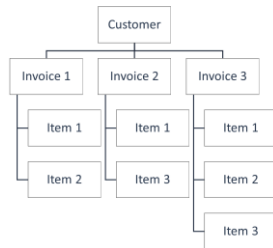
hour	month	week	hour	month	week	hour_sin	hour_cos	month_sin	month_cos	week_sin	week_cos
0	0	1	0	0	1	0.000000	1.000000	0.500000	8.660254e-01	0.000000e+00	1.0
1	1	2	1	1	2	0.269797	0.962917	0.866025	5.000000e-01	8.660254e-01	0.5
2	2	3	2	2	3	0.519584	0.854419	1.000000	6.123234e-17	8.660254e-01	-0.5
3	3	4	3	3	4	0.730836	0.682553	0.866025	-5.000000e-01	1.224647e-16	-1.0
4	4	5	4	4	5	0.887885	0.460065	0.500000	-8.660254e-01	-8.660254e-01	-0.5



	var_sp_0	var_sp_1	var_sp_2	var_sp_3	var_sp_4	var_sp_5	var_sp_6
0	0.166667	0.666667	0.166667	0.000000	0.0	0.0	0.0
1	0.082009	0.627011	0.289425	0.001555	0.0	0.0	0.0
2	0.032342	0.526705	0.428512	0.012441	0.0	0.0	0.0
3	0.008335	0.393741	0.555936	0.041989	0.0	0.0	0.0
4	0.000656	0.256111	0.643704	0.099529	0.0	0.0	0.0

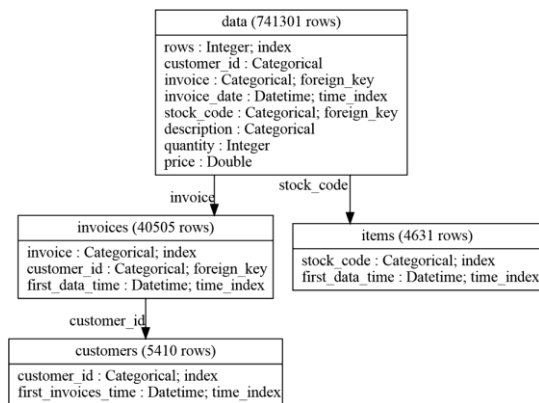


Chapter 9: Extracting Features from Relational Data with Featuretools



	customer_id	invoice	invoice_date	stock_code	description	quantity	price
0	13085.0	489434	2009-12-01 07:45:00	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	6.95
1	13085.0	489434	2009-12-01 07:45:00	79323P	PINK CHERRY LIGHTS	12	6.75
2	13085.0	489434	2009-12-01 07:45:00	79323W	WHITE CHERRY LIGHTS	12	6.75
3	13085.0	489434	2009-12-01 07:45:00	22041	RECORD FRAME 7" SINGLE SIZE	48	2.10
4	13085.0	489434	2009-12-01 07:45:00	21232	STRAWBERRY CERAMIC TRINKET BOX	24	1.25

invoice	customer_id	first_data_time	customer_id	first_data_time
489434	489434	13085.0 2009-12-01 07:45:00	13085.0	13085.0 2009-12-01 07:45:00
489435	489435	13085.0 2009-12-01 07:46:00	13078.0	13078.0 2009-12-01 09:06:00
489436	489436	13078.0 2009-12-01 09:06:00	15362.0	15362.0 2009-12-01 09:08:00
489437	489437	15362.0 2009-12-01 09:08:00	18102.0	18102.0 2009-12-01 09:24:00
489438	489438	18102.0 2009-12-01 09:24:00	18087.0	18087.0 2009-12-01 09:43:00



customer_id	MIN(data.price)	MIN(data.quantity)	MODE(data.description)	MODE(data.stock_code)	NUM_UNIQUE(data.description)
13085.0	0.55	-48.0	RECORD FRAME 7" SINGLE SIZE	22041	52
13078.0	0.19	-14.0	AREA PATROLLED METAL SIGN	82582	165
15362.0	0.21	1.0	BLUE PADDED SOFT MOBILE	20703	38
18102.0	0.27	-324.0	CREAM HEART CARD HOLDER	22189	415
18087.0	0.36	-96.0	WHITE HANGING HEART T-LIGHT HOLDER	85123A	48

	customer_id	invoice	stock_code	description	quantity	price	SINE(price)	SINE(quantity)	CUM_MAX(price) by invoice	CUM_MAX(quantity) by invoice
rows										
0	13085.0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	6.95	0.618486	-0.536573	6.95	12.0
1	13085.0	489434	79323P	PINK CHERRY LIGHTS	12	6.75	0.450044	-0.536573	6.95	12.0
2	13085.0	489434	79323W	WHITE CHERRY LIGHTS	12	6.75	0.450044	-0.536573	6.95	12.0
3	13085.0	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2.10	0.863209	-0.768255	6.95	48.0
4	13085.0	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	1.25	0.948985	-0.905578	6.95	48.0
				CUM_SUM(price) by invoice	CUM_SUM(quantity) by invoice	DIFF(price) by invoice	DIFF(quantity) by invoice	TIME_SINCE_PREVIOUS(invoice_date) by invoice		
				6.95	12.0	NaN	NaN			NaN
				13.70	24.0	-0.20	0.0			0.0
				20.45	36.0	0.00	0.0			0.0
				22.55	84.0	-4.65	36.0			0.0
				23.80	108.0	-0.85	-24.0			0.0

	customer_id	invoice	stock_code		description	quantity	price	price * quantity
rows								
0	13085.0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	6.95	83.4	
1	13085.0	489434	79323P	PINK CHERRY LIGHTS	12	6.75	81.0	
2	13085.0	489434	79323W	WHITE CHERRY LIGHTS	12	6.75	81.0	
3	13085.0	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2.10	100.8	
4	13085.0	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	1.25	30.0	

	customer_id	DAY(first_data_time)	DAYS_IN_MONTH(first_data_time)	HOUR(first_data_time)	IS_FEDERAL_HOLIDAY(first_data_time)	
invoice						
489434	13085.0		1	31	7	False
489435	13085.0		1	31	7	False
489436	13078.0		1	31	9	False
489437	15362.0		1	31	9	False
489438	18102.0		1	31	9	False
	MINUTE(first_data_time)	MONTH(first_data_time)	PART_OF_DAY(first_data_time)	WEEKDAY(first_data_time)	YEAR(first_data_time)	
	45	12	early morning	1	2009	
	46	12	early morning	1	2009	
	6	12	late morning	1	2009	
	8	12	late morning	1	2009	
	24	12	late morning	1	2009	

customer_id DISTANCE_TO_HOLIDAY(first_data_time, holiday=Boxing Day, country=UK)

invoice

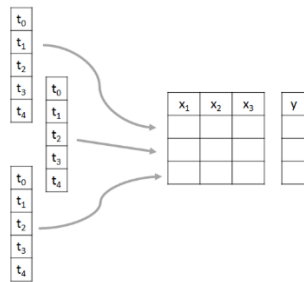
489434	13085.0	25.0
489435	13085.0	25.0
489436	13078.0	25.0
489437	15362.0	25.0
489438	18102.0	25.0

	customer_id	invoice	stock_code	quantity	price	NUM_CHARACTERS(description)	NUM_WORDS(description)
rows							
0	13085.0	489434	85048	12	6.95	35.0	6.0
1	13085.0	489434	79323P	12	6.75	18.0	3.0
2	13085.0	489434	79323W	12	6.75	20.0	4.0
3	13085.0	489434	22041	48	2.10	28.0	6.0
4	13085.0	489434	21232	24	1.25	30.0	4.0

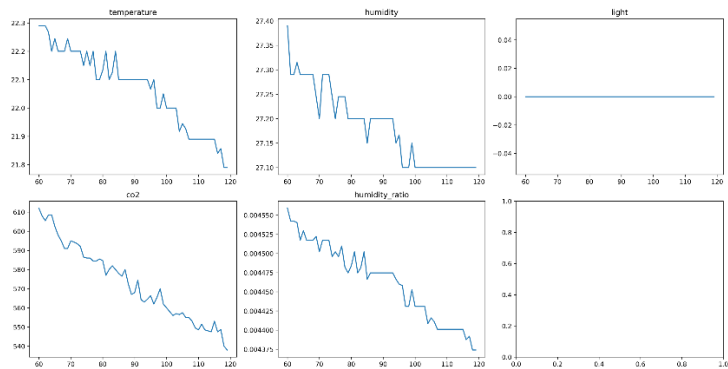
	DIVERSITY_SCORE(description)	MEAN_CHARACTERS_PER_WORD(description)	
rows			
0	0.833333	5.000000	
1	1.000000	5.333333	
2	1.000000	5.666667	
3	1.000000	4.600000	
4	1.000000	6.750000	
NUM_CHARACTERS(description)	NUM_WORDS(description)	PUNCTUATION_COUNT(description)	
	35.0	6.0	0
	18.0	3.0	0
	20.0	4.0	0
	28.0	6.0	1
	30.0	4.0	0

	MAX(data.price)	MAX(data.quantity)	MEAN(data.price)	MEAN(data.quantity)	MIN(data.price)	MIN(data.quantity)	SUM(data.price)
customer_id							
13085.0	830.12	48.0	12.413587	9.076087	0.55	-48.0	1142.05
13078.0	12.75	300.0	3.961193	14.061988	0.19	-14.0	3386.82
15362.0	9.95	48.0	3.612000	9.200000	0.21	1.0	144.48
18102.0	3580.80	1008.0	10.831367	175.196629	0.27	-324.0	11567.90
18087.0	852.80	3906.0	11.971368	78.189474	0.36	-96.0	1137.28
SUM(data.quantity)	MAX(invoices.MEAN(data.price))	MAX(invoices.MEAN(data.quantity))	...	MIN(invoices.MEAN(data.price))	MIN(invoices.MEAN(data.quantity))		
835.0	830.120000	20.750000	...	1.828571	-15.428571		
12023.0	12.750000	61.333333	...	0.190000	-14.000000		
368.0	3.628261	13.117647	...	3.590000	6.304348		
187110.0	3580.800000	624.000000	...	0.480000	-324.000000		
7428.0	852.800000	3906.000000	...	0.820000	-96.000000		

Chapter 10: Creating Features from Time Series with tsfresh



	id	date	temperature	humidity	light	co2	humidity_ratio
0	1	2015-02-04 18:00:00	23.075	27.175000	419.0	688.00	0.004745
1	1	2015-02-04 18:01:00	23.075	27.150000	419.0	690.25	0.004741
2	1	2015-02-04 18:02:00	23.100	27.100000	419.0	691.00	0.004739
3	1	2015-02-04 18:03:00	23.100	27.166667	419.0	683.50	0.004751
4	1	2015-02-04 18:04:00	23.050	27.150000	419.0	687.50	0.004734



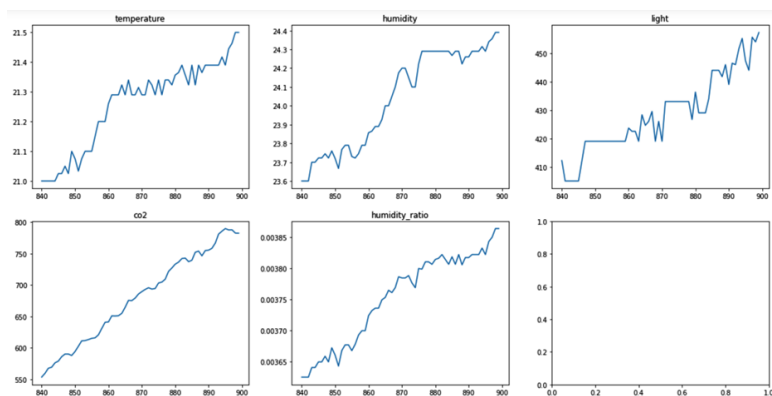
	light_mean	light_length	light_standard_deviation	light_variation_coefficient	light_variance
1	48.875	60.0	134.485582	2.751623	18086.371875
2	0.000	60.0	0.000000	NaN	0.000000
3	0.000	60.0	0.000000	NaN	0.000000
4	0.000	60.0	0.000000	NaN	0.000000
5	0.000	60.0	0.000000	NaN	0.000000

	light_minimum	light_agg_linear_trend_attr_"intercept"__chunk_len_50_f_agg_"min"
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0
5	0.0	0.0

	light_quantile_q_0.1	light_quantile_q_0.3	light_quantile_q_0.4
1	0.0	0.0	0.0
2	0.0	0.0	0.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0
5	0.0	0.0	0.0

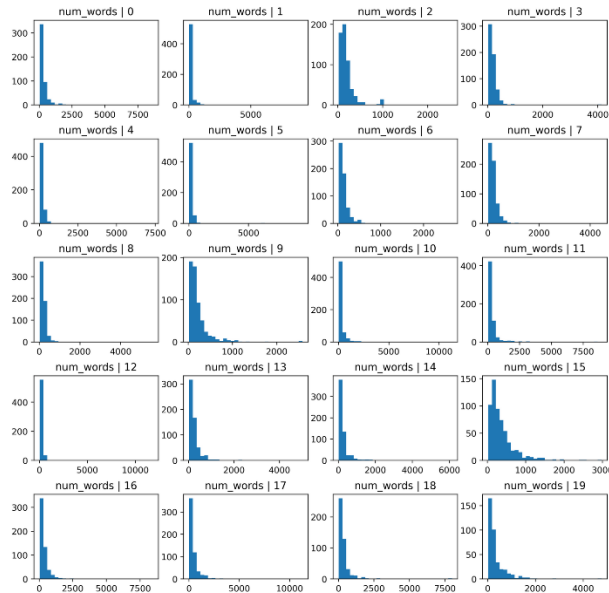
	light_sum_values	light_median	light_mean	light_length	light_standard_deviation	light_variance
1	2932.5	0.0	48.875	60.0	134.485582	18086.371875
2	0.0	0.0	0.000	60.0	0.000000	0.000000
3	0.0	0.0	0.000	60.0	0.000000	0.000000
4	0.0	0.0	0.000	60.0	0.000000	0.000000
5	0.0	0.0	0.000	60.0	0.000000	0.000000

	light_variance	light_root_mean_square	light_maximum	light_absolute_maximum	light_minimum
1	18086.371875	143.091361	419.0	419.0	0.0
2	0.000000	0.000000	0.0	0.0	0.0
3	0.000000	0.000000	0.0	0.0	0.0
4	0.000000	0.000000	0.0	0.0	0.0
5	0.000000	0.000000	0.0	0.0	0.0



Chapter 11: Extracting Features from Text Variables

	text	num_char	num_words	num_vocab	lexical_div	ave_word_length
0	From: lxxstl@wam.umd.edu (where's my thing)\nS...	716	123	93	1.322581	5.821138
1	From: guykuo@carson.u.washington.edu (Guy Kuo)...	857	123	99	1.242424	6.967480
2	From: twillis@ec.ecn.purdue.edu (Thomas E Will...	1980	339	219	1.547945	5.840708
3	From: jgreen@amber (Joe Green)\nSubject: Re: W...	814	113	96	1.177083	7.203540
4	From: jcm@head-cfa.harvard.edu (Jonathan McDow...	1117	171	139	1.230216	6.532164



	text	num_sent
1	11\nNntp-Posting-Host: carson.u.washington.ed...	6
2	36\n\nI well folks, my mac plus finally gave up...	9
3	14\nDistribution: world\nNntp-Posting-Host: a...	7
4	23\n\nFrom article <C5owCB.n3p@world.std.com>...	10
5	58\n\n\narticle <1r1eu1\$4t@transfer.stratus....	21
6	12\n\n\nThere were a few people who responded t...	8
7	44\nDistribution: world\nNntp-Posting-Host: d...	15
8	10\n\nI have win 3.0 and downloaded several i...	3
9	29\n\njap10@po.CWRU.Edu (Joseph A. Pelletier...	12
10	13\n\nI have a line on a Ducati 900GTS 1978 m...	11

	able	access	actually	ago	apr	article	articleid	ask	available	away	...	works	world	writes	wrong	wrote	xnewsreader	year	years	yes	your
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	...	0	1	1	0	1	0	0	0	0	0
4	0	0	0	0	0	2	0	0	0	0	...	0	0	1	0	0	0	0	0	1	0

5 rows × 191 columns

	able	access	actually	ago	apr	article	articleid	ask	available	away	...
0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0	...
1	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.356469	0.0	0.0	0.0	...
2	0.0	0.135765	0.123914	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0	...
3	0.0	0.000000	0.000000	0.0	0.0	0.110035	0.000000	0.0	0.0	0.0	...
4	0.0	0.000000	0.000000	0.0	0.0	0.262692	0.000000	0.0	0.0	0.0	...

	works	world	writes	wrong	wrote	xnewsreader	year	years	yes	your
0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.27302	0.000000	0.0	
0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.00000	0.000000	0.0	
0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.00000	0.000000	0.0	
0.0	0.169635	0.100554	0.0	0.218197	0.233578	0.0	0.00000	0.000000	0.0	
0.0	0.000000	0.120029	0.0	0.000000	0.000000	0.0	0.00000	0.264836	0.0	