数据探索性分析与数据预处理

学院： 计算机学院
专业： 生物医学工程
姓名： 陈晓珍
学号：2120170997

1. 问题描述

对数据集 San Francisco Building Permits 和数据集 NFL Play-by-Play 2009-2017 进行探索性分析与预处理。

2. 数据说明

数据集：San Francisco Building Permits，记做 S_data 数据集
数据集：NFL Play-by-Play 2009-2017，记做 N_data 数据集

3. 数据分析

3.1 数据可视化和摘要

3.1.1 数据摘要

（1）标称属性，给出其每个可能取值的频数

数据集	结果
S_data	/Building_Permits/result_Build_nominal.txt
N_data	/NFL/result_NFL_nominal.txt

（2）S_data 数据集数值属性，给出最大、最小、均值、中位数、四分位数及缺失值的个数。

	Permit Type	Street Number	Unit	Number of Existing	Number of Proposed Stories	Estimated Cost	Revised Cost	Existing Units	Proposed Units	Plansets	Existing Construction Type	Proposed Construction Type	Supervisor District	Zipcode	Record ID
count	198900.000000	198900.000000	29479.000000	156116.000000	156032.000000	1.608340e+05	1.928340e+05	147362.00000	147989.00	161591.000000	155534.000000	155738.000000	197183.000000	197184.000000	1.98000e+05
mean	7.522323	1121.728944	78.517182	5.705773	5.745043	1.689554e+05	1.328562e+05	15.666164	16.510950	1.274650	4.072878	4.089529	5.538403	94115.500558	1.162048e+12
Std	1.457451	1135.768948	326.981324	8.613455	8.613284	3.630386e+06	3.584903e+06	74.47632	75.220444	22.407345	1.585756	1.578766	2.887041	9.270131	4.918215e+11
min	1.000000	0.000000	0.000000	0.000000	0.000000	1.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000	94102.000000	1.293532e+10
25%	8.000000	235.000000	0.000000	2.000000	2.000000	3.300000e+03	1.000000e+00	1.000000	1.000000	0.000000	3.000000	3.000000	3.000000	94109.000000	1.308567e+12
50%	8.000000	710.000000	0.000000	3.000000	3.000000	1.100000e+04	7.000000e+03	1.000000	2.000000	2.000000	5.000000	5.000000	6.000000	94114.000000	1.371840e+12
75%	8.000000	1700.000000	1.000000	4.000000	4.000000	3.500000e+04	2.870750e+04	4.000000	4.000000	2.000000	5.000000	5.000000	8.000000	94122.000000	1.435000e+12
Max	8.000000	8400.000000	6004.00000	78.000000	78.000000	5.379586e+08	7.805000e+08	1907.000000	1911.0	9000.000000	5.000000	5.000000	11.000000	94158.000000	1.498342e+12

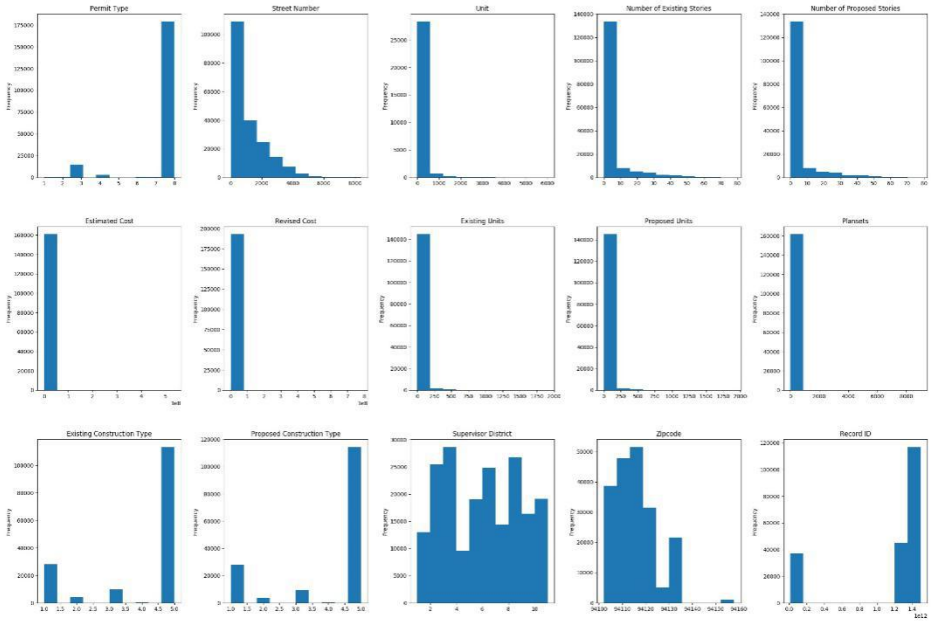
（3）N_data 数据集数值属性，给出最大、最小、均值、中位数、四分位数及缺失值的个数。

	GameID	Drive	qtr	down	TimeUnder	TimeSecs	PlayTime Diff	yrdln	yrdline100	ydstogo	yacEPA
count	407688.000000	407464.000000	407244.000000	406848.000000	4.076880e+05	407688.000000	407688.000000	346534.000000	406848.000000	407688.000000	159190.000000
mean	7.374200	1695.268944	20.576762	28.488327	2.013158e+09	12.316158	2.577412	2.002476	48.644081	7.309403	-0.386086
std	4.642388	1062.801012	17.969326	12.946471	2.572839e+06	7.149527	1.129750	1.006353	25.070416	4.869987	1.972715
min	0.000000	-900.000000	0.000000	1.000000	2.009091e+09	1.000000	1.000000	1.000000	1.000000	0.000000	-14.000000
25%	3.000000	778.000000	5.000000	20.000000	2.011101e+09	6.000000	2.000000	1.000000	30.000000	3.000000	-0.961115
50%	7.000000	1800.000000	17.000000	30.000000	2.013111e+09	12.000000	3.000000	2.000000	49.000000	9.000000	0.000000
75%	11.000000	2585.000000	37.000000	39.000000	2.015121e+09	18.000000	4.000000	3.000000	70.000000	10.000000	0.485508
max	15.000000	3600.000000	943.000000	50.000000	2.017123e+09	35.000000	5.000000	4.000000	99.000000	50.000000	9.559834

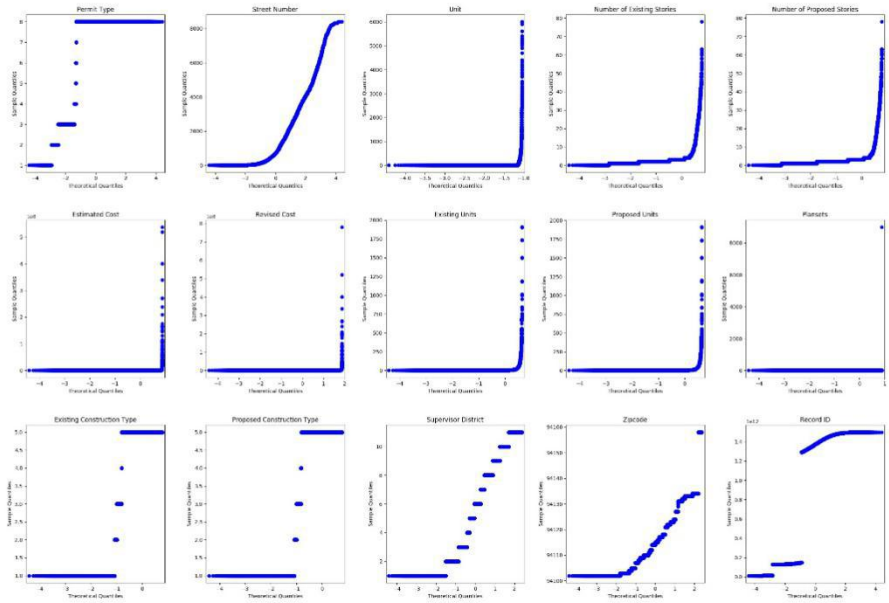
	Home_WP_pre	Away_WP_pre	Home_WP_post	Away_WP_post	Win_Prob	WPA	airWPA	yacWPA	Season
count	382734.000000	382734.000000	381101.000000	381101.000000	382679.000000	402147.000000	159187.000000	158926.000000	407688.000000
mean	0.534488	0.465965	0.534791	0.465613	0.501320	0.002099	0.015135	-0.010480	2013.018985
std	0.285574	0.285629	0.287818	0.287867	0.287445	0.045363	0.056490	0.068139	2.576962
min	0.000000	0.000000	0.000000	0.000000	0.000000	-0.997214	-0.999881	-0.986673	2009.000000
25%	0.325123	0.231411	0.321701	0.227694	0.276472	-0.014728	-0.011518	-0.018683	2011.000000
50%	0.531274	0.469052	0.533609	0.466670	0.504470	0.000000	0.003441	0.000000	2013.000000
75%	0.769232	0.675530	0.772882	0.678833	0.725477	0.014684	0.035792	0.011431	2015.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	0.994848	0.994848	1.000000	2017.000000

3.1.2 S_data 数据可视化

（1）针对数值属性，绘制直方图，用 qq 图检验其分布是否为正态分布。直方图如下所示:

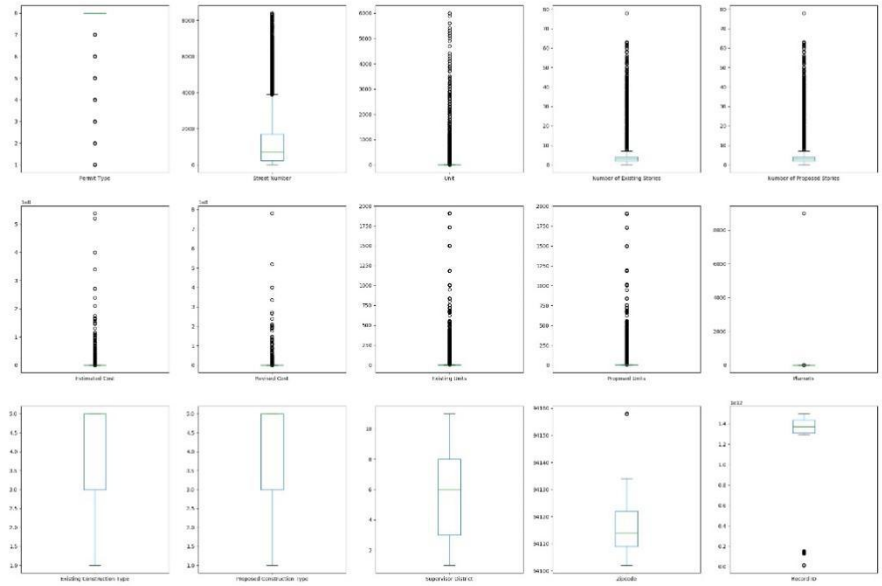


（2）qq 图如下所示:



由各个属性的 qq 图可以看出,无属性满足正态分布

（3）绘制盒图，对离群值进行识别。盒图如下所示:



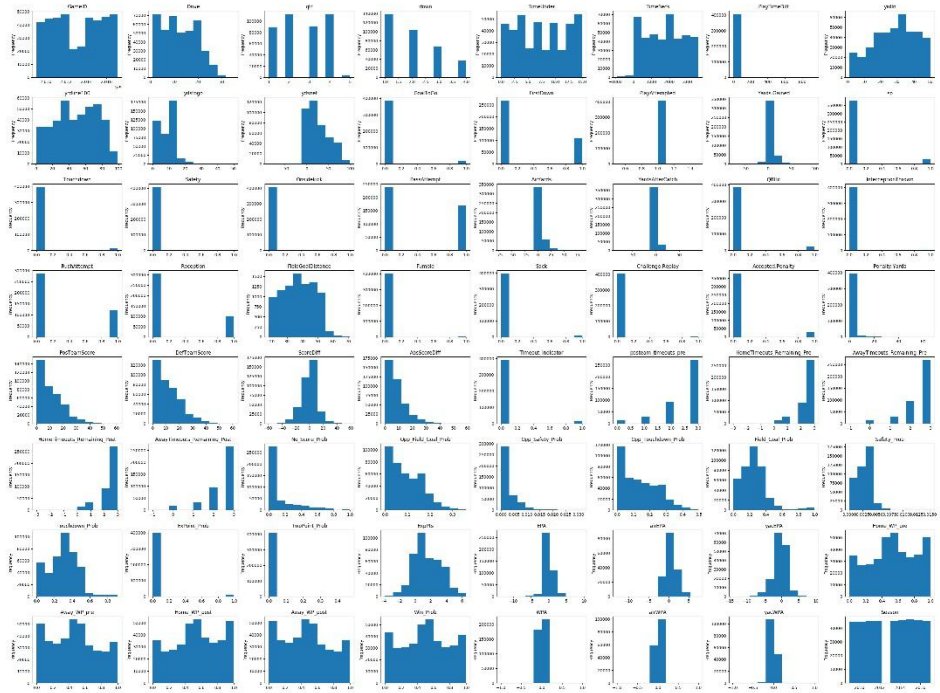
从各个属性的盒图观察可得,属性 Permit Type、Street Number、Unit、Number of Existing Stories、Number of Proposed Stories、Estimated Cost、Revised Cost、Existing Units、Proposed Units、Plansets、Zipcode、Record ID 存在离群值

3.1.3 N_data 数据可视化

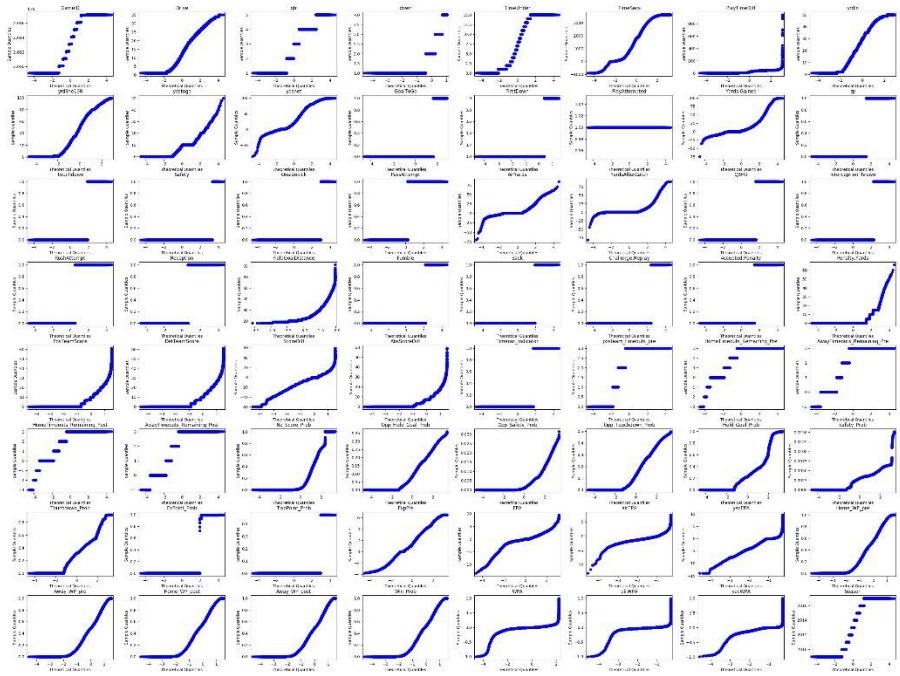
针对数值属性，

绘制直方图，用 qq 图检验其分布是否为正态分布。

直方图如下所示:



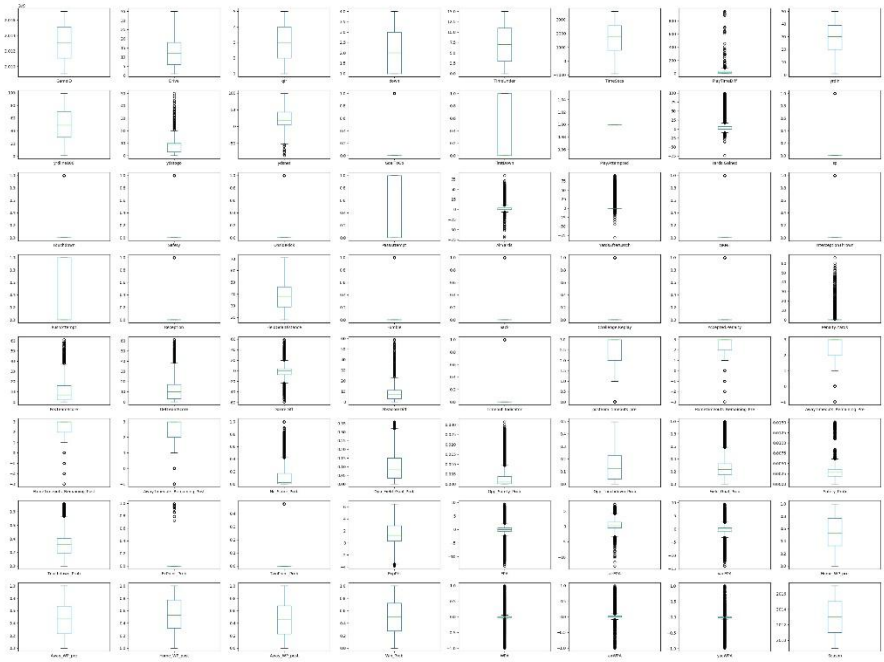
qq 图如下所示:



由各个属性的 qq 图可以看出,属性 ExpPts 和 EPA 满足正态分布

绘制盒图，对离群值进行识别

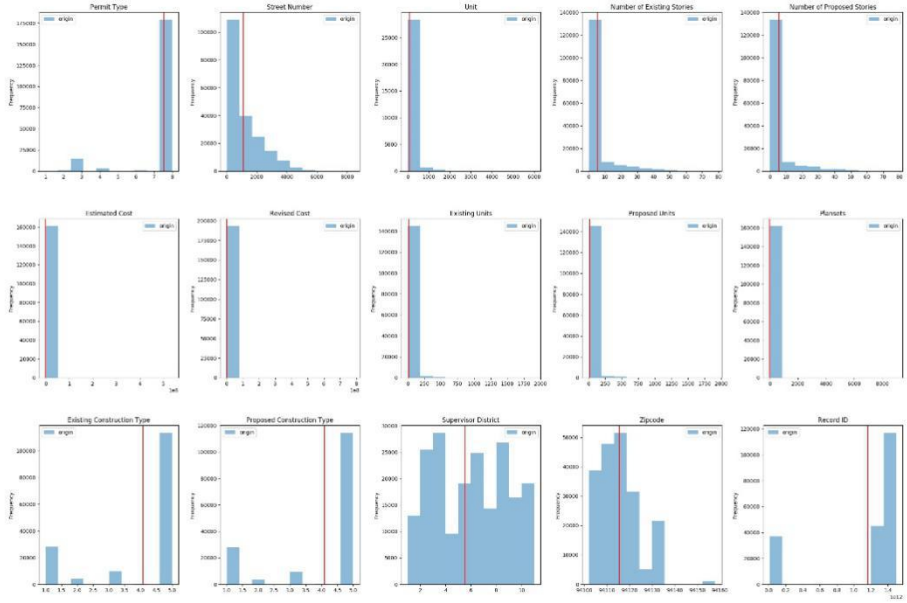
盒图如下所示:



从各个属性的盒图观察可得,属性 PlayTimeDiff、ydstogo、ydsnet、GoalToGo、Yards.Gained、sp、Touchdown、Safety、Onsidekick、AirYards、YardsAfterCatch、QBHit、Interception Thrown、Reception、Fumble、Sack、Challenge.Replay、Accepted.Penalty、Penalty.Yards、PosTeamScore、DefTeamScore、ScoreDiff、AbsScoreDiff、Timeout_Indicator、posteam_timeouts_pre、HomeTimeouts_Remaining_Pre、AwayTimeouts_Remaining_Pre、HomeTimeouts_Remaining_Post、AwayTimeouts_Remaining_Post、No_Score_Prob、Opp_Field_Goal_Prob、Opp_Safety_Prob、Field_Goal_Prob、Safety_Prob、Touchdown_Prob、ExPoint_Prob、TwoPoint_Prob、ExpPts、EPA、airEPA、yacEPA、WPA、airWPA、yacWPA 存在离群值

3.2 S_data 数据缺失处理

（1）将缺失部分剔除



（2）用最高频率值来填补缺失值

填补前,csv 数据文件部分情况如下所示:

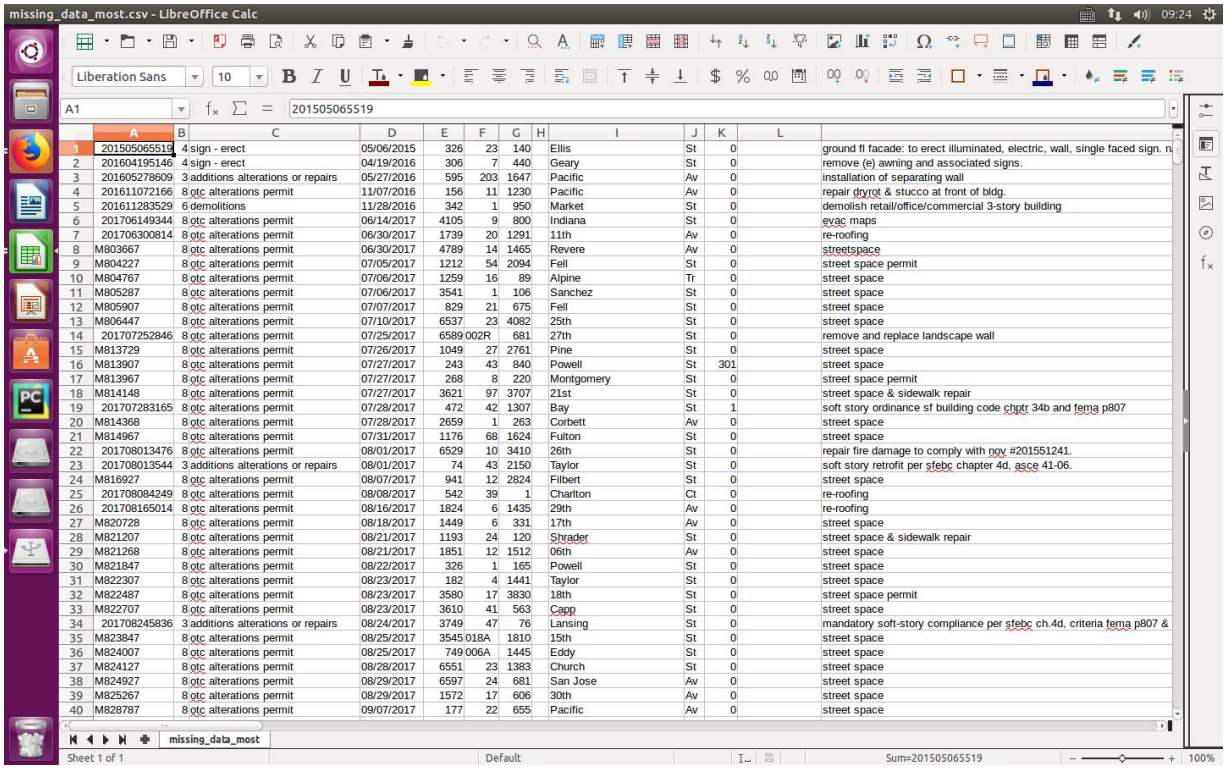
Building_Permits.csv - LibreOffice Calc

	D	E	F	G	H	I	J	K	L	
1	Permit Creation Date	Block	Lot	Street Number	Street Number Suffix	Street Name	Street Suffix	Unit	Unit Suffix	Description
2	05/06/2015	326	23	140		Ellis	St			ground fl facade: to erect illuminated, electric, wall, single faced sign. n/a for maher
3	04/19/2016	306	7	440		Geary	St	0		remove (e) awning and associated signs.
4	05/27/2016	595	203	1647		Pacific	Av			installation of separating wall
5	11/07/2016	156	11	1230		Pacific	Av	0		repair dryrot & stucco at front of bldg.
6	11/28/2016	342	1	950		Market	St			demolish retail/office/commercial 3-story building
7	06/14/2017	4105	9	800		Indiana	St			evac maps
8	06/30/2017	1739	20	1291		11th	Av	0		re-roofing
9	06/30/2017	4789	14	1465		Revere	Av			streetspace
10	07/05/2017	1212	54	2094		Fell	St			street space permit
11	07/06/2017	1259	16	89		Alpine	Tr			street space
12	07/06/2017	3541	1	106		Sanchez	St			street space
13	07/07/2017	829	21	675		Fell	St			street space
14	07/10/2017	6537	23	4082		25th	St			street space
15	07/25/2017	6589	002R	681		27th	St			remove and replace landscape wall
16	07/26/2017	1049	27	2761		Pine	St			street space
17	07/27/2017	243	43	840		Powell	St	301		street space
18	07/27/2017	268	8	220		Montgomery	St			street space permit
19	07/27/2017	3621	97	3707		21st	St	0		street space & sidewalk repair
20	07/28/2017	472	42	1307		Bay	St	1		soft story ordinance sf building code chptr 34b and fema p807
21	07/28/2017	2659	1	263		Corbett	Av	0		street space
22	07/31/2017	1176	68	1624		Fulton	St			street space
23	08/01/2017	6529	10	3410		26th	St			repair fire damage to comply with nov #201551241.
24	08/01/2017	74	43	2150		Taylor	St			soft story retrofit per sfebc chapter 4d, asce 41-06.
25	08/07/2017	941	12	2824		Filbert	St			street space
26	08/08/2017	542	39	1		Charlton	Ct			re-roofing
27	08/16/2017	1824	6	1435		29th	Av			re-roofing
28	08/18/2017	1449	6	331		17th	Av			street space
29	08/21/2017	1193	24	120		Shrader	St			street space & sidewalk repair
30	08/21/2017	1851	12	1512		06th	Av			street space
31	08/22/2017	326	1	165		Powell	St			street space
32	08/23/2017	182	4	1441		Taylor	St	0		street space
33	08/23/2017	3580	17	3830		18th	St	0		street space permit
34	08/23/2017	3610	41	563		Capp	St			street space
35	08/24/2017	3749	47	76		Lansing	St			mandatory soft-story compliance per sfebc ch.4d, criteria fema p807 & asce 41-13.
36	08/25/2017	3545	018A	1810		15th	St			street space
37	08/25/2017	749	006A	1445		Eddy	St			street space
38	08/28/2017	6551	23	1383		Church	St			street space
39	08/29/2017	6597	24	681		San Jose	Av			street space
40	08/29/2017	1572	17	606		30th	Av			street space

Building_Permits

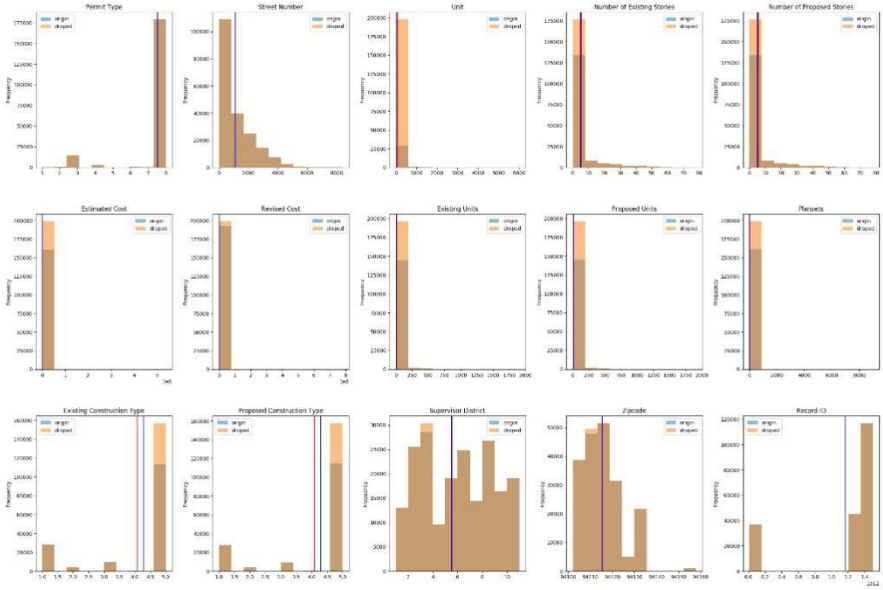
Sheet 1 of 1 | Default | Sum=0 | 100%

填补后,csv 数据文件部分情况如下所示:



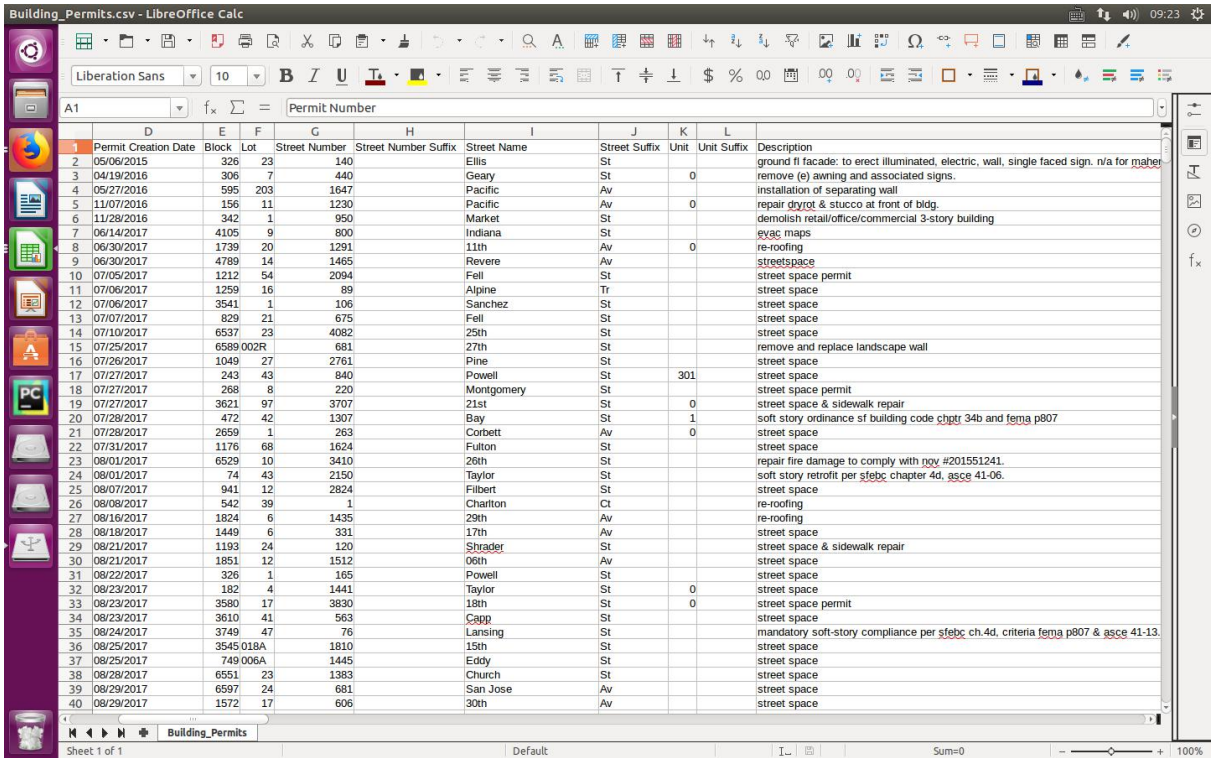
	A	B	C	D	E	F	G	H	I	J	K	L	
1	201505065519	4	sign - erect	05/06/2015	326	23	140		Ellis	St	0		ground fl facade: to erect illuminated, electric, wall, single faced sign. n
2	201604195146	4	sign - erect	04/19/2016	306	7	440		Geary	St	0		remove (e) awning and associated signs.
3	201605278609	3	additions alterations or repairs	05/27/2016	595	203	1647		Pacific	Av	0		installation of separating wall
4	201611072166	8	otc alterations permit	11/07/2016	156	11	1230		Pacific	Av	0		repair dryrot & stucco at front of bldg.
5	201611283529	6	demolitions	11/28/2016	342	1	950		Market	St	0		demolish retail/office/commercial 3-story building
6	201706149344	8	otc alterations permit	06/14/2017	4105	9	800		Indiana	St	0		evac maps
7	201706300814	8	otc alterations permit	06/30/2017	1739	20	1291		11th	Av	0		re-roofing
8	M803667	8	otc alterations permit	06/30/2017	4789	14	1465		Revere	Av	0		streetspace
9	M804227	8	otc alterations permit	07/05/2017	1212	54	2094		Fell	St	0		street space permit
10	M804767	8	otc alterations permit	07/06/2017	1259	16	89		Alpine	Tr	0		street space
11	M805287	8	otc alterations permit	07/06/2017	3541	1	106		Sanchez	St	0		street space
12	M805907	8	otc alterations permit	07/07/2017	829	21	675		Fell	St	0		street space
13	M806447	8	otc alterations permit	07/10/2017	6537	23	4082		25th	St	0		street space
14	201707252846	8	otc alterations permit	07/25/2017	6589	002R	681		27th	St	0		remove and replace landscape wall
15	M813729	8	otc alterations permit	07/26/2017	1049	27	2761		Pine	St	0		street space
16	M813907	8	otc alterations permit	07/27/2017	243	43	840		Powell	St	301		street space
17	M813967	8	otc alterations permit	07/27/2017	268	8	220		Montgomery	St	0		street space permit
18	M814148	8	otc alterations permit	07/27/2017	3621	97	3707		21st	St	0		street space & sidewalk repair
19	201707283165	8	otc alterations permit	07/28/2017	472	42	1307		Bay	St	1		soft story ordinance sf building code chptr 34b and fema p807
20	M814368	8	otc alterations permit	07/28/2017	2659	1	263		Corbett	Av	0		street space
21	M814967	8	otc alterations permit	07/31/2017	1176	68	1624		Fulton	St	0		street space
22	201708013476	8	otc alterations permit	08/01/2017	6529	10	3410		26th	St	0		repair fire damage to comply with nov #201551241.
23	201708013544	3	additions alterations or repairs	08/01/2017	74	43	2150		Taylor	St	0		soft story retrofit per sfebc chapter 4d, asce 41-06.
24	M816927	8	otc alterations permit	08/07/2017	941	12	2824		Filbert	St	0		street space
25	201708084249	8	otc alterations permit	08/08/2017	542	39	1		Charlton	Ct	0		re-roofing
26	201708165014	8	otc alterations permit	08/16/2017	1824	6	1435		29th	Av	0		re-roofing
27	M820728	8	otc alterations permit	08/18/2017	1449	6	331		17th	Av	0		street space
28	M821207	8	otc alterations permit	08/21/2017	1193	24	120		Shrader	St	0		street space & sidewalk repair
29	M821268	8	otc alterations permit	08/21/2017	1851	12	1512		06th	Av	0		street space
30	M821847	8	otc alterations permit	08/22/2017	326	1	165		Powell	St	0		street space
31	M822307	8	otc alterations permit	08/23/2017	182	4	1441		Taylor	St	0		street space
32	M822487	8	otc alterations permit	08/23/2017	3580	17	3830		18th	St	0		street space permit
33	M822707	8	otc alterations permit	08/23/2017	3610	41	563		Capp	St	0		street space
34	201708245836	3	additions alterations or repairs	08/24/2017	3749	47	76		Lansing	St	0		mandatory soft-story compliance per sfebc ch.4d, criteria fema p807 &
35	M823847	8	otc alterations permit	08/25/2017	3545	018A	1810		15th	St	0		street space
36	M824007	8	otc alterations permit	08/25/2017	749	006A	1445		Eddy	St	0		street space
37	M824127	8	otc alterations permit	08/28/2017	6551	23	1383		Church	St	0		street space
38	M824927	8	otc alterations permit	08/29/2017	6597	24	681		San Jose	Av	0		street space
39	M825267	8	otc alterations permit	08/29/2017	1572	17	606		30th	Av	0		street space
40	M828787	8	otc alterations permit	09/07/2017	177	22	655		Pacific	Av	0		street space

可视化对比填补前后数据,直方图如下所示:

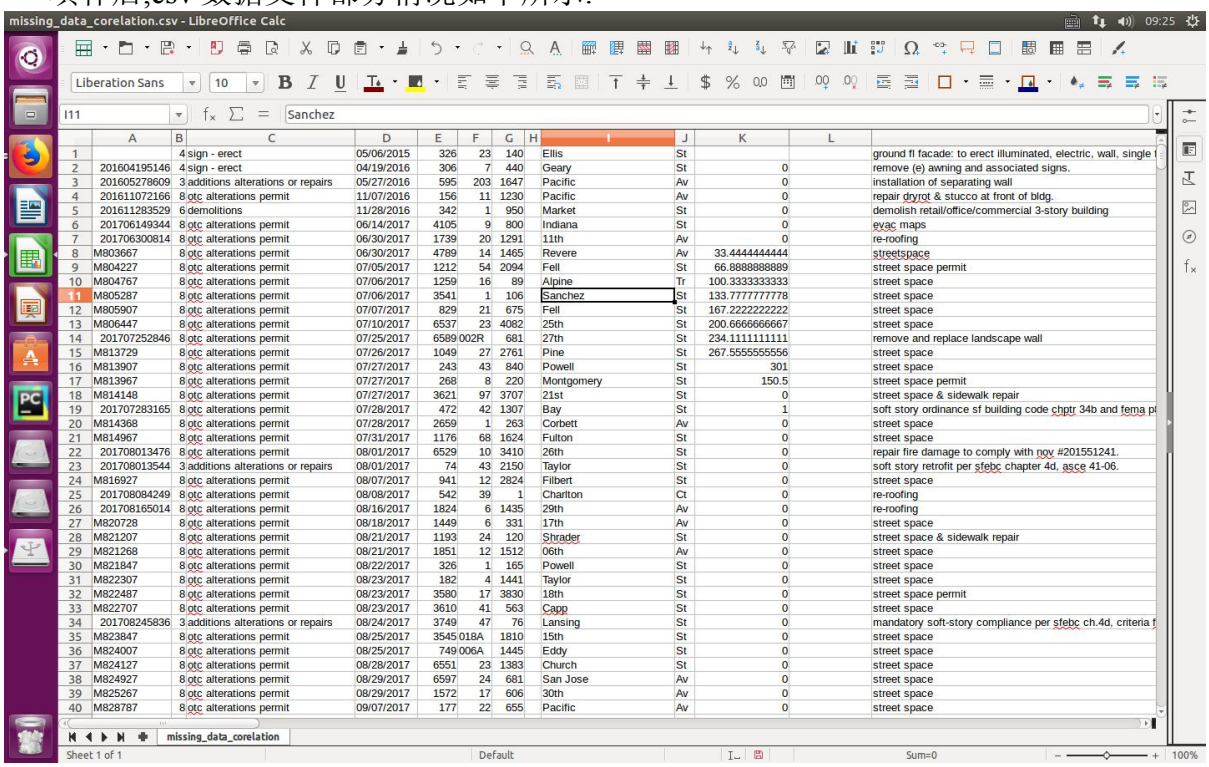


- 通过属性的相关关系来填补缺失值

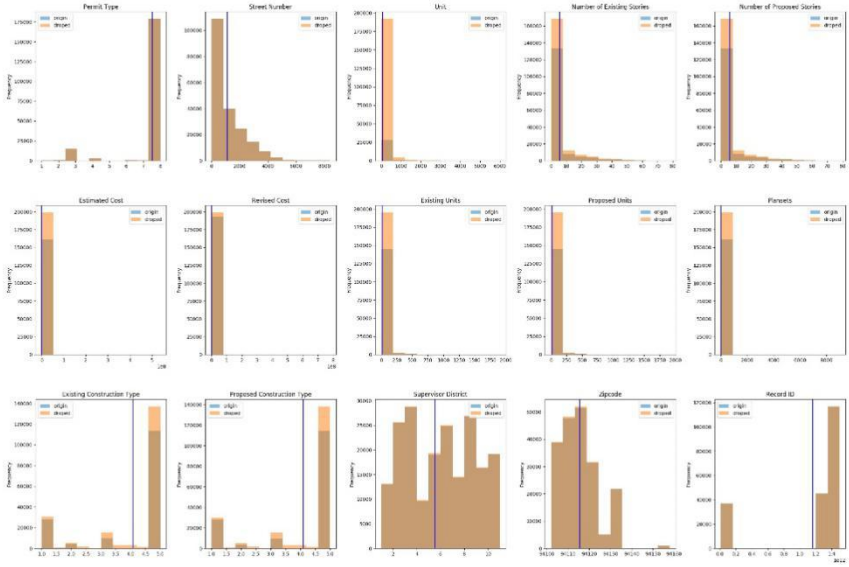
填补前,csv 数据文件部分情况如下所示:



填补后,csv 数据文件部分情况如下所示:



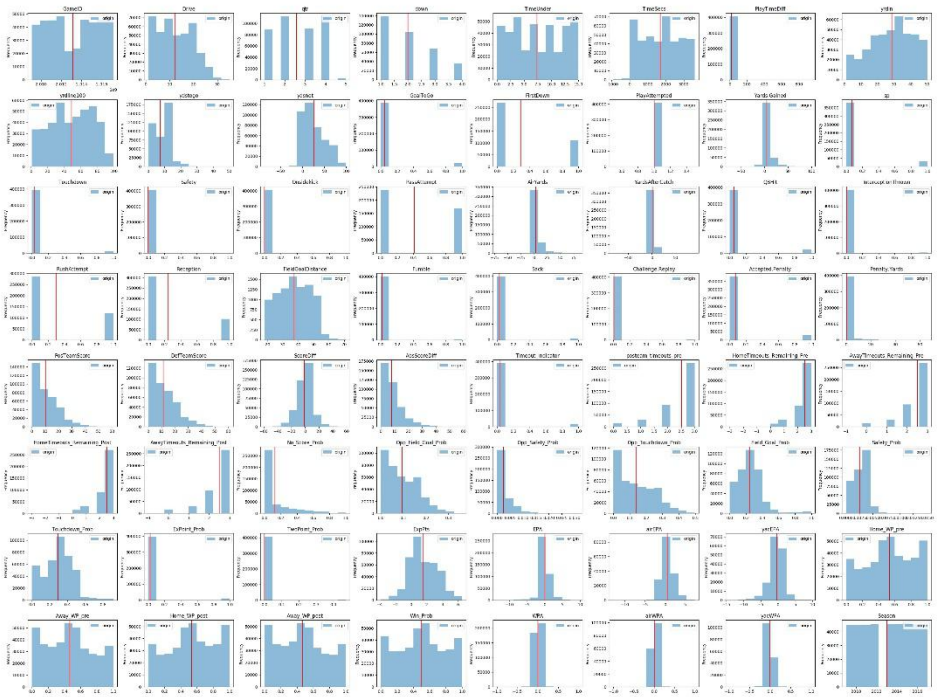
可视化对比填补前后数据,直方图如下所示:



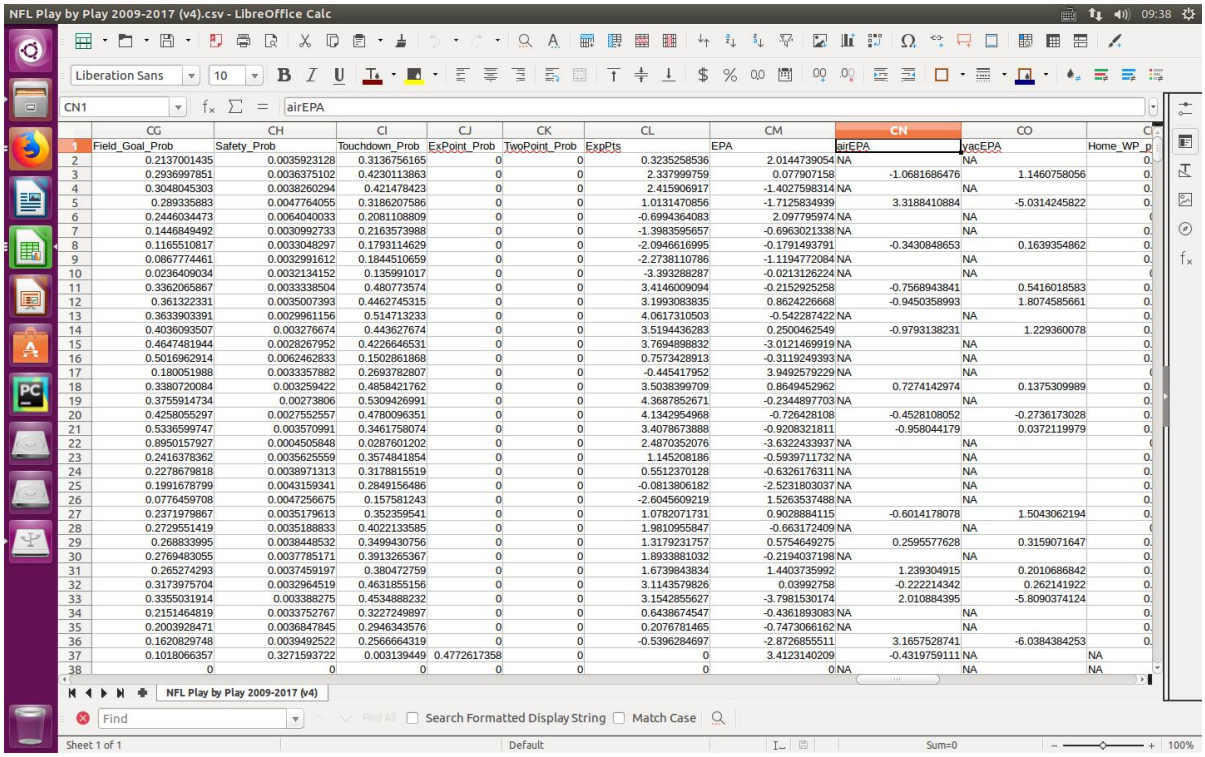
- 通过数据对象之间的相似性来填补缺失值

3.3 N_data 数据缺失处理

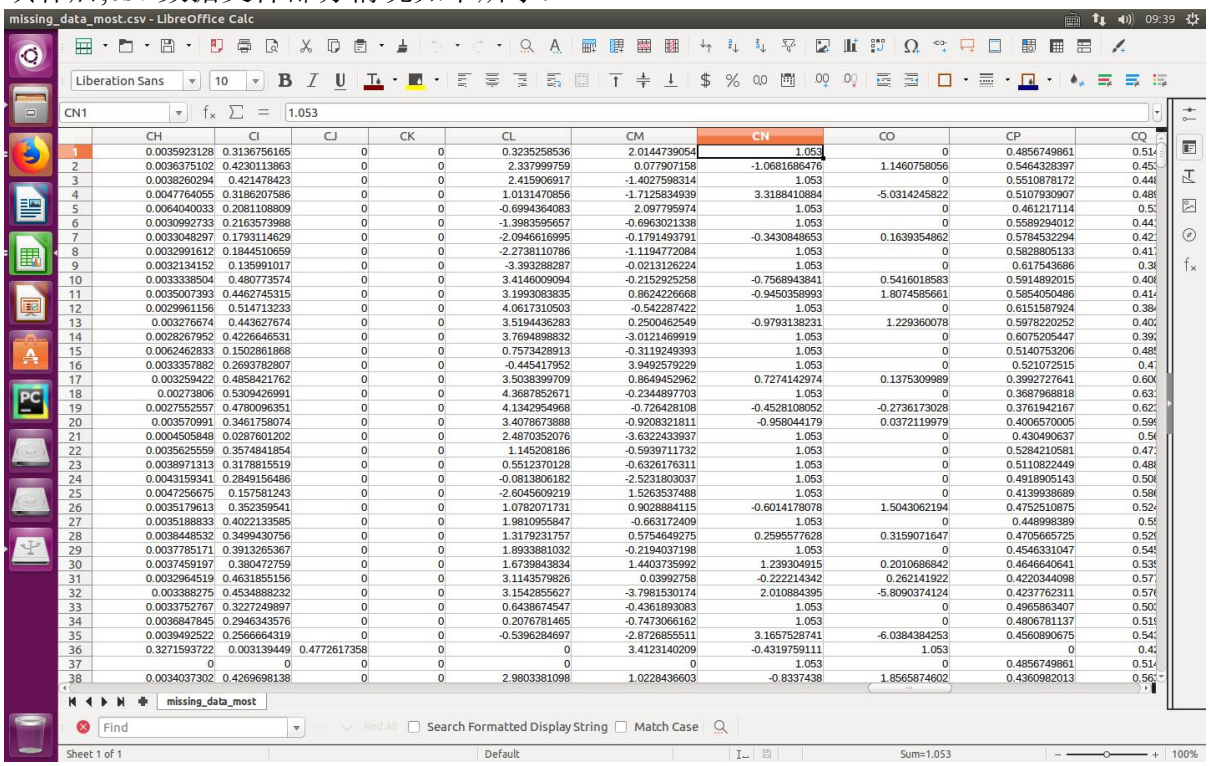
将缺失部分剔除



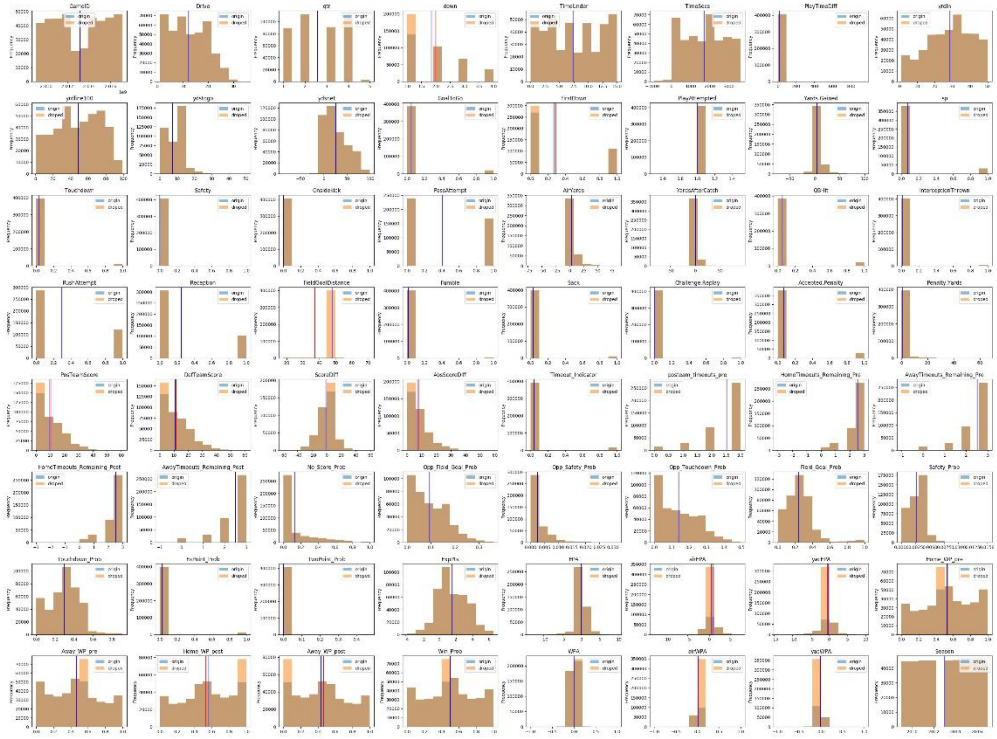
用最高频率值来填补缺失值
填补前,csv 数据文件部分情况如下所示:



填补后,csv 数据文件部分情况如下所示:

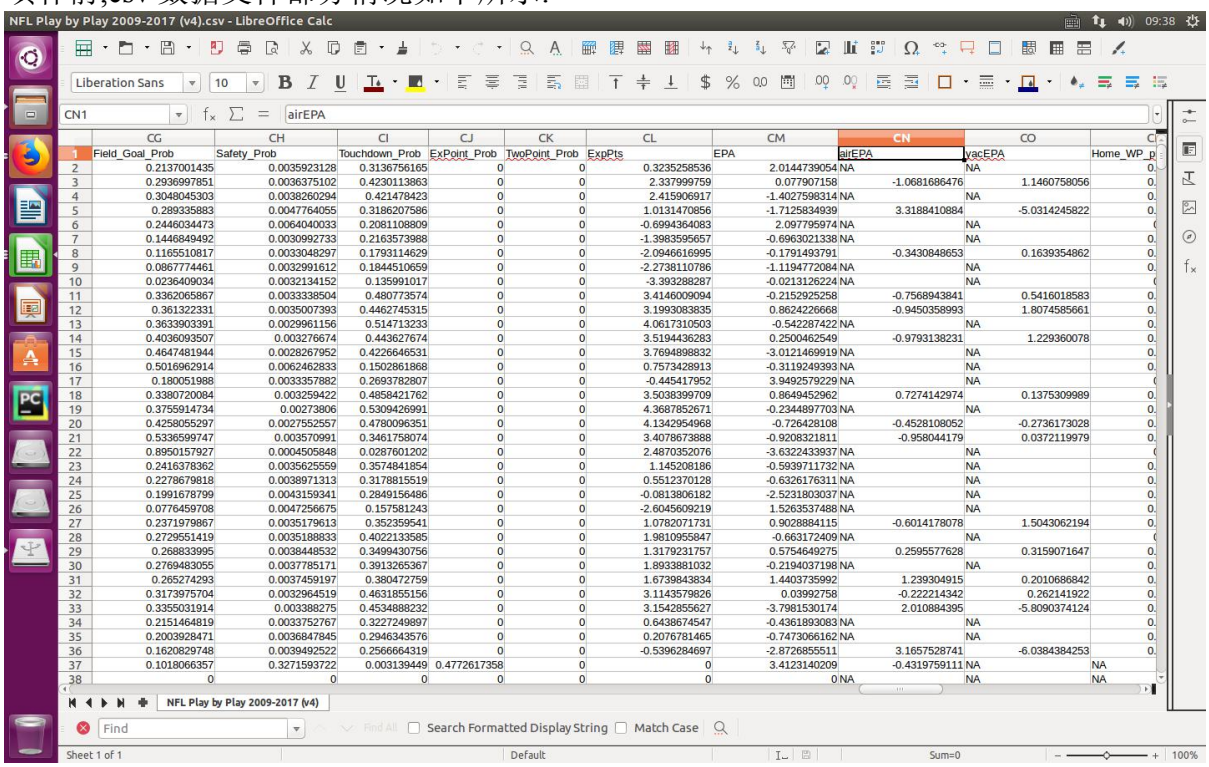


可视化对比填补前后数据,直方图如下所示:



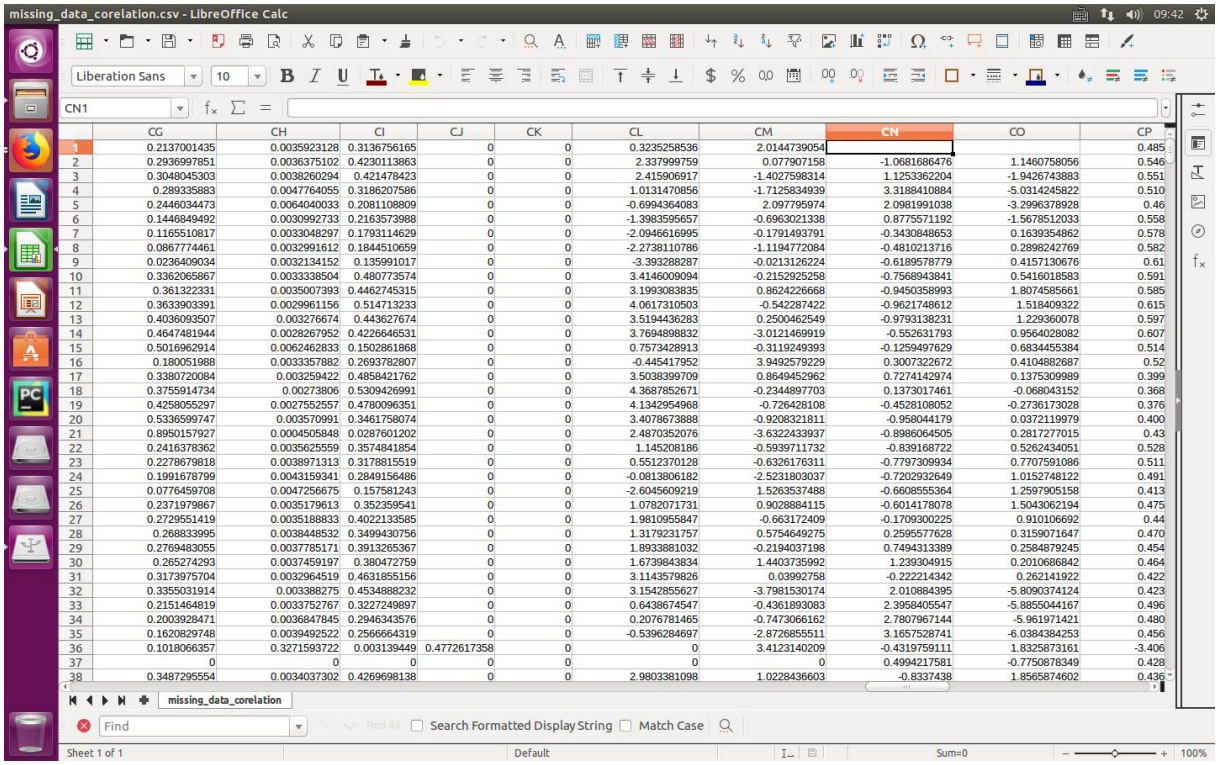
通过属性的相关关系来填补缺失值

填补前,csv 数据文件部分情况如下所示:



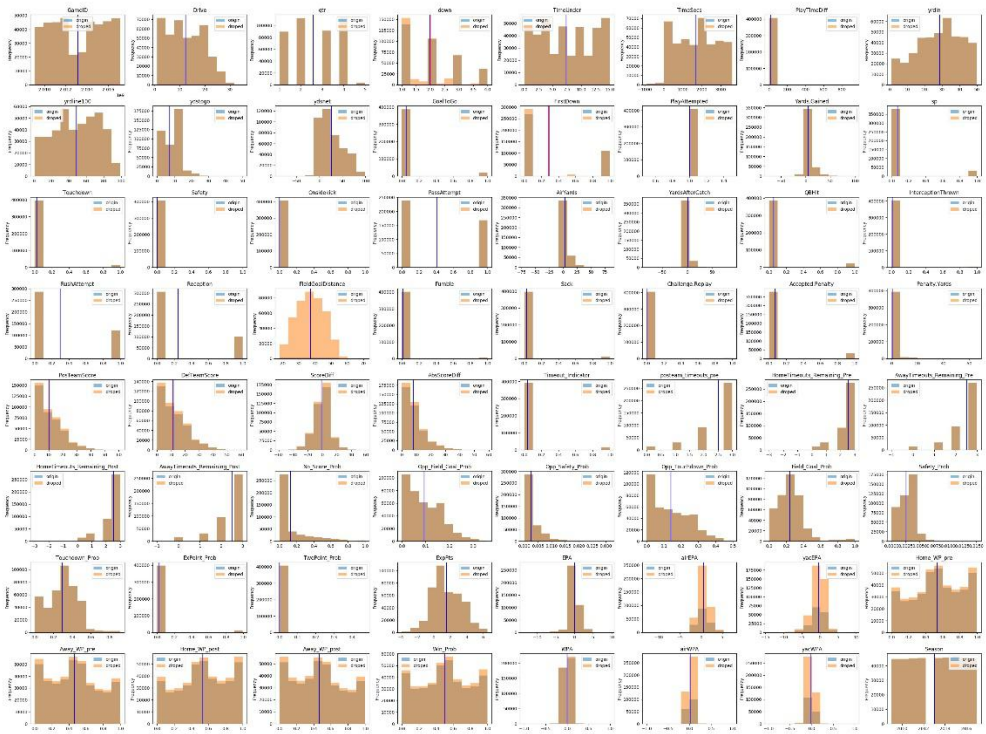
	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP
1	Field_Goal_Prob	Safety_Prob	Touchdown_Prob	ExPoint_Prob	TwoPoint_Prob	ExpPts	EPA	airEPA	yacEPA	Home_WP_p
2	0.2137001435	0.0035923128	0.3136756165	0	0	0.3235258536	2.0144739054	NA		0.
3	0.2936997851	0.0036375102	0.4230113863	0	0	2.337999759	0.077907158	-1.0681686476	1.1460758056	0.
4	0.3048045303	0.0038260294	0.421478423	0	0	2.415906917	-1.4027598314	NA	NA	0.
5	0.289335883	0.0047764055	0.3186207586	0	0	1.0131470856	-1.7125834939	3.3188410884	-5.0314245822	0.
6	0.2446034473	0.0064040033	0.2081108809	0	0	-0.6994364083	2.097795974	NA	NA	0
7	0.1446849492	0.0030992733	0.2163573988	0	0	-1.3983595657	-0.6963021338	NA	NA	0.
8	0.1165510817	0.0033048297	0.1793114629	0	0	-2.0946616995	-0.1791493791	-0.3430848653	0.1639354862	0.
9	0.0867774461	0.0032991612	0.1844510659	0	0	-2.2738110786	-1.1194772084	NA	NA	0.
10	0.0236409034	0.0032134152	0.135991017	0	0	-3.393288287	-0.0213126224	NA	NA	0
11	0.3362065867	0.0033338504	0.480773574	0	0	3.4146009094	-0.2152925258	-0.7568943841	0.5416018583	0.
12	0.361322331	0.0035007393	0.4462745315	0	0	3.1993083835	0.8624226668	-0.9450358993	1.8074585661	0.
13	0.3633903391	0.0029961156	0.514713233	0	0	4.0617310503	-0.542287422	NA	NA	0.
14	0.4036093507	0.003276674	0.443627674	0	0	3.5194436283	0.2500462549	-0.9793138231	1.229360078	0.
15	0.4647481944	0.0028267952	0.4226646531	0	0	3.7694898832	-3.0121469919	NA	NA	0.
16	0.5016962914	0.0062462833	0.1502861868	0	0	0.7573428913	-0.3119249393	NA	NA	0.
17	0.180051988	0.0033357882	0.2693782807	0	0	-0.445417952	3.9492579229	NA	NA	0
18	0.3380720084	0.003259422	0.4858421762	0	0	3.5038399709	0.8649452962	0.7274142974	0.1375309989	0.
19	0.3755914734	0.00273806	0.5309426991	0	0	4.3687852671	-0.2344897703	NA	NA	0.
20	0.4258055297	0.0027552557	0.4780096351	0	0	4.1342954968	-0.726428108	-0.4528108052	-0.2736173028	0.
21	0.5336599747	0.003570991	0.3461758074	0	0	3.4078673888	-0.9208321811	-0.958044179	0.0372119979	0.
22	0.8950157927	0.0004505848	0.0287601202	0	0	2.4870352076	-3.6322433937	NA	NA	0
23	0.2416378362	0.0035625559	0.3574841854	0	0	1.145208186	-0.5939711732	NA	NA	0.
24	0.2278679818	0.0038971313	0.3178815519	0	0	0.5512370128	-0.6326176311	NA	NA	0.
25	0.1991678799	0.0043159341	0.2849156486	0	0	-0.0813806182	-2.5231803037	NA	NA	0.
26	0.0776459708	0.0047256675	0.157581243	0	0	-2.6045609219	1.5263537488	NA	NA	0.
27	0.2371979867	0.0035179613	0.352359541	0	0	1.0782071731	0.9028884115	-0.6014178078	1.5043062194	0.
28	0.2729551419	0.0035188833	0.4022133585	0	0	1.9810955847	-0.663172409	NA	NA	0
29	0.268833995	0.0038448532	0.3499430756	0	0	1.3179231757	0.5754649275	0.2595577628	0.3159071647	0.
30	0.2769483055	0.0037785171	0.3913265367	0	0	1.8933881032	-0.2194037198	NA	NA	0.
31	0.265274293	0.0037459197	0.380472759	0	0	1.6739843834	1.4403735992	1.239304915	0.2010686842	0.
32	0.3173975704	0.0032964519	0.4631855156	0	0	3.1143579826	0.03992758	-0.222214342	0.262141922	0.
33	0.3355031914	0.003388275	0.4534888232	0	0	3.1542855627	-3.7981530174	2.010884395	-5.8090374124	0.
34	0.2151464819	0.0033752767	0.3227249897	0	0	0.6438674547	-0.4361893083	NA	NA	0.
35	0.2003928471	0.0036847845	0.2946343576	0	0	0.2076781465	-0.7473066162	NA	NA	0.
36	0.1620829748	0.0039492522	0.2566664319	0	0	-0.5396284697	-2.8726855511	3.1657528741	-6.0384384253	0.
37	0.1018066357	0.3271593722	0.003139449	0.4772617358	0	0	3.4123140209	-0.4319759111	NA	NA
38	0	0	0	0	0	0	0 NA		NA	NA

填补后,csv 数据文件部分情况如下所示:



	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP
1	0.2137001435	0.0035923128	0.3136756165	0	0	0.3235258536	2.0144739054			0.485
2	0.2936997851	0.0036375102	0.4230113863	0	0	2.337999759	0.077907158	-1.0681686476	1.1460758056	0.546
3	0.3048045303	0.0038260294	0.421478423	0	0	2.415906917	-1.4027598314	1.1253362204	-1.9426743883	0.551
4	0.289335883	0.0047764055	0.3186207586	0	0	1.0131470856	-1.7125834939	3.3188410884	-5.0314245822	0.510
5	0.2446034473	0.0064040033	0.2081108809	0	0	-0.6994364083	2.097795974	2.0981991038	-3.2996378928	0.46
6	0.1446849492	0.0030992733	0.2163573988	0	0	-1.3983595657	-0.6963021338	0.8775571192	-1.5678512033	0.558
7	0.1165510817	0.0033048297	0.1793114629	0	0	-2.0946616995	-0.1791493791	-0.3430848653	0.1639354862	0.578
8	0.0867774461	0.0032991612	0.1844510659	0	0	-2.2738110786	-1.1194772084	-0.4810213716	0.2898242769	0.582
9	0.0236409034	0.0032134152	0.135991017	0	0	-3.393288287	-0.0213126224	-0.6189578779	0.4157130676	0.61
10	0.3362065867	0.0033338504	0.480773574	0	0	3.4146009094	-0.2152925258	-0.7568943841	0.5416018583	0.591
11	0.361322331	0.0035007393	0.4462745315	0	0	3.1993083835	0.8624226668	-0.9450358993	1.8074585661	0.585
12	0.3633903391	0.0029961156	0.514713233	0	0	4.0617310503	-0.542287422	-0.9621748612	1.518409322	0.615
13	0.4036093507	0.003276674	0.443627674	0	0	3.5194436283	0.2500462549	-0.9793138231	1.1293600078	0.597
14	0.4647481944	0.0028267952	0.4226646531	0	0	3.7694898832	-3.0121469919	-0.552631793	0.9564028082	0.607
15	0.5016962914	0.0062462833	0.1502861868	0	0	0.7573428913	-0.3119249393	-0.1259497629	0.6834455384	0.514
16	0.180051988	0.0033357882	0.2693782807	0	0	-0.445417952	3.9492579229	0.3007322672	0.4104882687	0.52
17	0.3380720084	0.003259422	0.4858421762	0	0	3.5038399709	0.8649452962	0.7274142974	0.1375309989	0.399
18	0.3755914734	0.00273806	0.5309426991	0	0	4.3687852671	-0.2344897703	0.1373017461	-0.068043152	0.368
19	0.4258055297	0.0027552557	0.4780096351	0	0	4.1342954968	-0.726428108	-0.4528108052	-0.2736173028	0.376
20	0.5336599747	0.003570991	0.3461758074	0	0	3.4078673888	-0.9208321811	-0.958044179	0.0372119979	0.400
21	0.8950157927	0.0004505848	0.0287601202	0	0	2.4870352076	-3.6322433937	-0.8986064505	0.2817277015	0.43
22	0.2416378362	0.0035625559	0.3574841854	0	0	1.145208186	-0.5939711732	-0.839168722	0.5262434051	0.528
23	0.2278679818	0.0038971313	0.3178815519	0	0	0.5512370128	-0.6326176311	-0.7797309934	0.7707591086	0.511
24	0.1991678799	0.0043159341	0.2849156486	0	0	-0.0813806182	-2.5231803037	-0.7202932649	1.0152748122	0.491
25	0.0776459708	0.0047256675	0.157581243	0	0	-2.6045609219	1.5263537488	-0.6608555364	1.2597905158	0.413
26	0.2371979867	0.0035179613	0.352359541	0	0	1.0782071731	0.9028884115	-0.6014178078	1.5043062194	0.475
27	0.2729551419	0.0035188833	0.4022133585	0	0	1.9810955847	-0.663172409	-0.1709300225	0.910106692	0.44
28	0.268833995	0.0038448532	0.3499430756	0	0	1.3179231757	0.5754649275	0.2595577628	0.3159071647	0.470
29	0.2769483055	0.0037785171	0.3913265367	0	0	1.8933881032	-0.2194037198	0.7494313389	0.2584879245	0.454
30	0.265274293	0.0037459197	0.380472759	0	0	1.6739843834	1.4403735992	1.239304915	0.2010686842	0.464
31	0.3173975704	0.0032964519	0.4631855156	0	0	3.1143579826	0.03992758	-0.222214342	0.262141922	0.422
32	0.3355031914	0.003388275	0.4534888232	0	0	3.1542855627	-3.7981530174	2.010884395	-5.8090374124	0.423
33	0.2151464819	0.0033752767	0.3227249897	0	0	0.6438674547	-0.4361893083	2.3958405547	-5.8855044167	0.496
34	0.2003928471	0.0036847845	0.2946343576	0	0	0.2076781465	-0.7473066162	2.7807967144	-5.961971421	0.480
35	0.1620829748	0.0039492522	0.2566664319	0	0	-0.5396284697	-2.8726855511	3.1657528741	-6.0384384253	0.456
36	0.1018066357	0.3271593722	0.003139449	0.4772617358		0	3.4123140209	-0.4319759111	1.8325873161	-3.406
37	0	0	0	0	0	0	0	0.4994217581	-0.7750878349	0.428
38	0.3487295554	0.0034037302	0.4269698138	0	0	2.9803381098	1.0228436603	-0.8337438	1.8565874602	0.436

可视化对比填补前后数据,直方图如下所示:



通过数据对象之间的相似性来填补缺失值