Data Quality Report

Data Information

- This dataset records transactions in government spending in Tennessee, Nashville.
- Dimension: 96753 rows, 10 columns
- Time frame: 01/01/2010 – 12/31/2010

Summary Table of Numerical Variables

	Recnum	Cardnum	Merch zip	Amount
count	96753.000000	9.675300e+04	92097.000000	9.675300e+04
mean	48377.000000	5.142202e+09	44706.596740	4.278857e+02
std	27930.329635	5.567084e+04	28369.537945	1.000614e+04
min	1.000000	5.142110e+09	1.000000	1.000000e-02
25%	24189.000000	5.142152e+09	20855.000000	3.348000e+01
50%	48377.000000	5.142196e+09	38118.000000	1.379800e+02
75%	72565.000000	5.142246e+09	63103.000000	4.282000e+02
max	96753.000000	5.142847e+09	99999.000000	3.102046e+06

	Records populated	% populated
Recnum	96753	100%
Cardnum	96753	100%
Date	96753	100%
Merchnum	93378	97%
Merch description	96753	100%
Merch state	95558	99%
Merch zip	92097	95%
Transtype	96753	100%
Amount	96753	100%
Fraud	96753	100%

Recnum (categorical)
Description: Uniquely identifies transactions in the dataset.

Cardnum (categorical)

Description:
- Credit card number if a transaction is paid for by credit card
- 1645 unique values

Cardnum	Count
5142148452	1192
5142184598	921
5142189108	663
5142297710	583
5142223373	579
5142187452	526
5142299634	515
5142189945	512

... ...

Date (categorical)

Description:
- Date when transactions occurred
- 365 unique values

Date
2010

Day of D..	January	February	March	April	May	June	July	August	Septemb..	October	November	December
1	51	368	431	92	315	379	96	521	418	308	191	316
2	29	387	321	35	442	367	49	436	91	46	237	331
3	159	393	352	312	406	186	209	454	58	10	324	276
4	229	325	137	343	362	24	120	402	80	204	285	44
5	309	63	22	374	329	353	325	177	404	249	310	43
6	330	33	358	387	103	492	311	38	440	231	54	259
7	307	372	418	267	30	383	345	386	536	354	11	290
8	104	295	372	117	308	439	77	446	411	214	218	367
9	19	496	359	21	374	331	27	536	97	35	298	323
10	323	444	320	262	386	90	328	610	19	16	330	311
11	318	372	113	335	453	37	415	427	390	100	207	75
12	442	95	26	450	316	287	410	73	373	187	200	33
13	374	18	313	386	127	415	469	177	564	218	29	308
14	350	259	339	313	25	356	348	403	533	228	18	300
15	139	396	594	98	363	440	150	386	459	232	228	449
16	9	501	451	29	470	328	24	390	171	32	275	352
17	177	379	337	285	426	198	313	392	30	6	354	322
18	235	317	121	336	346	26	426	392	436	206	300	68
19	302	54	18	346	367	358	460	139	497	236	264	18
20	359	11	400	362	203	382	400	202	491	186	67	328
21	283	184	401	256	38	408	393	305	522	240	17	264
22	46	331	465	65	326	428	135	454	315	213	273	365
23	59	327	385	48	394	337	26	470	171	31	304	217
24	325	286	320	282	401	125	300	495	33	16	269	43
25	329	274	109	407	405	33	429	384	374	197	59	14
26	237	79	14	363	310	340	384	96	374	236	116	6
27	304	13	291	379	149	437	409	41	429	280	18	199
28	297	684	423	411	33	413	351	335	503	270	38	216
29	89		416	342	66	472	145	503	496	263	218	263
30	13		381	28	345	385	33	462	180	46	289	200
31	263		414		364		389	518		19		53

Merchnum (categorical)

Description:
- Digits that uniquely identifies merchants
- 13092 unique values

Merchnum	Count

930090121224	9310
5509006296254	2131
9900020006406	1714
602608969534	1092
4353000719908	1020
410000971343	982
…	…

Merch description (categorical)

Description:
- Text description of merchants
- 13126 unique values

Merch description	Count
GSA-FSS-ADV	1688
SIGMA-ALDRICH	1635
STAPLES #941	1174
FISHER SCI ATL	1093
MWI*MICRO WAREHOUSE	958
CDW*GOVERNMENT INC	872
DELL MARKETING L.P.	816
FISHER SCI CHI	783
AMAZON.COM *SUPERSTOR	750
OFFICE DEPOT #1082	748
…	…

Merch state (categorical)

Description:
- State of merchant location
- 228 unique values

Merch state	Count
TN	12035
VA	7872
CA	6817
IL	6508
MD	5398
GA	5025
PA	4899
…	…
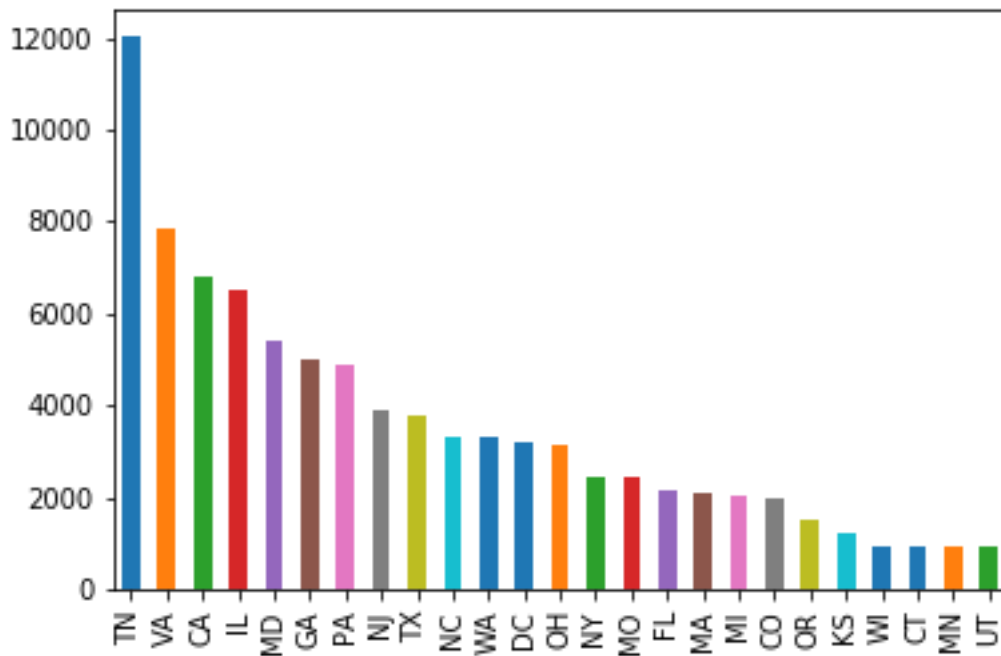
Merch zip (categorical)

Description:
- Zip code of merchant location
- 4568 unique values

Merch zip	Count
38118	11868
63103	1650
08701	1267
22202	1250
60061	1221
98101	1197
… …	

Transtype (categorical)

Description:
- Transaction type
- 4 unique values
- "p" represents credit card transactions

Transtype	Count
P	96398
A	181
D	173
Y	1

Amount (numerical)

Description:
Dollar amount paid in transactions.
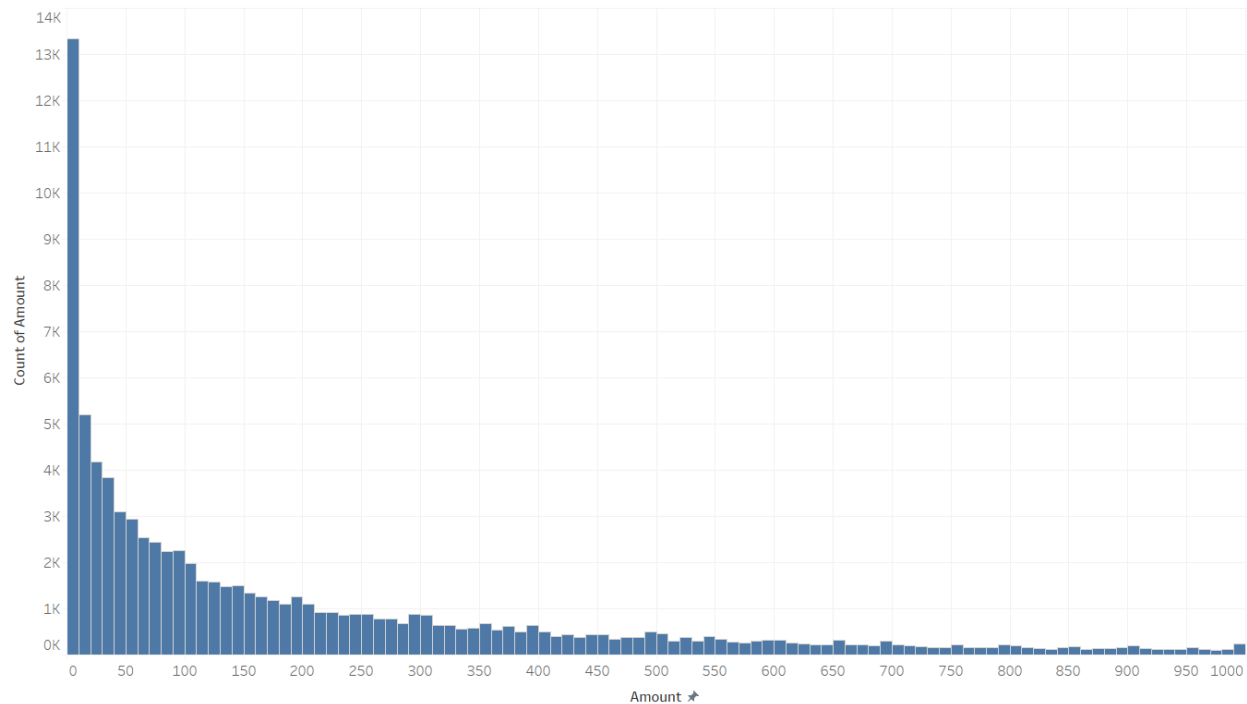
Fraud (numeric)

Description:
- Labels: 1-fraud, 0-normal
- 2 unique values

Fraud	
0	95,694
1	1,059