```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df.head()
```

	trans_date_trans_time	cc_num	merchant	category	amt	first	last	gender	street
0	2019-01-01 00:00:18	2703186189652095	fraud_Rippin, Kub and Mann	misc_net	4.97	Jennifer	Banks	F	561 Perry Cove
1	2019-01-01 00:00:44	630423337322345	fraud_Heidenreich, Lesch and Ullrich	grocery_pos	107.23	Stephanie	Gill	F	43039 Riley Greens Suite 393
2	2019-01-01 00:00:51	38859492057661	fraud_Lind-Buckridge	entertainment	220.11	Edward	Sanchez	M	594 White Dale Suite 530
3	2019-01-01 00:01:16	3018685985130206	fraud_Kutch, Hermiston and Farrell	gas_transport	45.00	Jeremy	White	M	9443 Cynthia Court Apt. 038
4	2019-01-01 00:03:06	3575051765865946	fraud_Keeling-Crist	misc_pos	41.60	Tyler	Garcia	M	408 Bradley Rest

5 rows × 23 columns

```
df.info()
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex:1296675 entries,0 to 1296674
Data columns(total 23 columns):
 0   trans_date_trans_time  1296675 non-null  object
 1   cc_num                 1296675 non-null  int64
 2   merchant               1296675 non-null  object
 3   category               1296675 non-null  object
 4   amt                    1296675 non-null  float64
 5   first                  1296675 non-null  object
 6   last                   1296675 non-null  object
 7   gender                 1296675 non-null  object
 8   street                 1296675 non-null  object
 9   city                   1296675 non-null  object
 10  state                  1296675 non-null  object
 11  zip                    1296675 non-null  int64
 12  lat                    1296675 non-null  float64
 13  long                   1296675 non-null  float64
 14  city_pop               1296675 non-null  int64
 15  job                    1296675 non-null  object
 16  dob                    1296675 non-null  object
 17  trans_num              1296675 non-null  object
 18  unix_time              1296675 non-null  int64
 19  merch_lat              1296675 non-null  float64
 20  merch_long             1296675 non-null  float64
 21  is_fraud               1296675 non-null  int64
 22  merch_zipcode          1100702 non-null  float64
dtype:float64,int64,object
memory usage:227.5+MB
```

list_long_merch_list_merch_long

convert lat_long_merch_list_merch_long to a distance

new column: distance

drop column: lat_long_merch_list_merch_long

drop: transaction 0

```
df[df.lat_long_merch_list_merch_long.isnull()]
```

	trans_date_trans_time	cc_num	merchant	category	amt	first	last	gender	street	city	state	zip	lat	long	city_pop	job	dob	trans_num	unix_time	merch_lat	merch_long	is_fraud	merch_zipcode
0	2019-01-01 00:00:18	2703186189652095	fraud_Rippin, Kub and Mann	misc_net	4.97	Jennifer	Banks	F	561 Perry Cove	Moravian Falls	NC	28654	36.0788	-81.1781	3495	Psychologist, counselling	1988-03-09	0b242abb623afa9f1f131446ed193743	1325376018	NaN	NaN	0	NaN
1	2019-01-01 00:00:44	630423337322345	fraud_Heidenreich, Lesch and Ullrich	grocery_pos	107.23	Stephanie	Gill	F	43039 Riley Greens Suite 393	Sheridan	WY	82801	44.6265	-111.1029	75267	Video editor	1978-06-21	1f76529f8574b4d66fa9de6f11f7f1b9	1325376044	NaN	NaN	0	NaN
2	2019-01-01 00:01:51	38859492057661	fraud_Lind-Buckridge	entertainment	220.11	Edward	Sanchez	M	594 White Dale Suite 530	Belleville	WI	53508	42.7933	-89.4246	19116	Commercial horticulturist	1962-01-19	2d91a119a7bfd607f65f777ee8f6d356	1325376111	NaN	NaN	0	NaN
3	2019-01-01 00:03:06	3018685985130206	fraud_Kutch, Hermiston and Farrell	gas_transport	45.00	Jeremy	White	M	9443 Cynthia Court Apt. 038	Erie	PA	16509	42.1808	-79.4774	80507	Patent attorney	1967-12-11	324cc204407e8f62a8f5fb1ab00a8529	1325376186	NaN	NaN	0	NaN
4	2019-01-01 00:03:06	3575051765865946	fraud_Keeling-Crist	misc_pos	41.60	Tyler	Garcia	M	408 Bradley Rest	Buckeye	AZ	85326	33.3759	-112.5615	62799	Biomedical engineer	1993-12-30	8d7f8d6b0bb20da3d7ae8f832d8bd943	1325376186	NaN	NaN	0	NaN
5	2019-01-01 00:05:08	3534093764340240	fraud_Stokes, Winters and Keebler	gas_transport	94.63	Jennifer	Conrad	F	611 Cynthia Walk Apt. 062	Belmont	MT	59711	45.8744	-112.5264	2691	Psychologist, educational	1958-09-16	7a811d2d2d2d2d2d2d2d2d2d2d2d2d	1325376308	NaN	NaN	0	NaN

5 rows × 23 columns

```
df.columns
```

```
Index(['trans_date_trans_time', 'cc_num', 'merchant', 'category', 'amt',
       'first', 'last', 'gender', 'street', 'city', 'state', 'zip', 'lat',
       'long', 'city_pop', 'job', 'dob', 'trans_num', 'unix_time',
       'merch_lat', 'merch_long', 'is_fraud', 'merch_zipcode'],
      dtype='object')
```

dob

encode date of birth to age

new column: age

drop column: dob

```
import datetime as dt
from sklearn.preprocessing import LabelEncoder
%matplotlib inline
```

```
df['dob'] = pd.to_datetime(df['dob'])
df['age'] = dt.datetime.now().year - df['dob'].dt.year
df['age'].hist()
```

```
<ipykernel_jobs_1_0.zip_ref_ref_ref
```