In [35]:
```python
#Clean rent raw data obtianed from "https://www.huduser.gov/portal/datasets/fmr.
```

In [36]:
```python
import pandas as pd
import numpy as np
```

In [37]:
```python
rent = pd.read_excel("C:/Users/Rosario Household/Downloads/rentdatax.xlsx",0) #i
```

In [38]:
```python
rent.head(2)

#view raw data
```

Out[38]:

	zcta	geoid	name	tblid	stub	num_estimate	num_margin_error	count_40p	a
0	601	H2500US00601	ZCTA5 00601	S5413T	1 bedroom	20	24	8.0	
1	601	H2500US00601	ZCTA5 00601	S5413T	With cash rent:	20	24	NaN	

In [39]:
```python
rent2 = rent[(rent['stub'] !='1 bedroom')
    &(rent['stub'] !='2 bedrooms')
    &(rent['stub'] !='3 bedrooms')
    &(rent['stub'] !='With cash rent:')
    &(rent['stub'] !='No cash rent')
    &(rent['stub'] !='Less than $100')
    ]

#begin to clean data by subsetting data into a  new dataframe
```

In [40]:
```python
rent2.loc[:,'stub'] = rent2['stub'].str.replace(',', '')

# remove commas in numbers to extract the full integer
```

In [41]:
```python
rent3 = rent2['stub'].str.extract('(\d+)').astype(int)

# extract integers and create final dataframe
```

In [42]:
```python
rent3['num'] = pd.Series(rent2['num_estimate'])
rent3['zip'] = pd.Series(rent2['name'])
rent3['rent'] = rent2['stub'].str.extract('(\d+)').astype(int)

#import the appropiate columns
```

In [43]:
```python
rent3['bedroom'] = rent3.groupby(['zip','rent']).cumcount()+1; rent3

#calculate new column based on how the data is read
```

Out[43]:

	0	num	zip	rent	bedroom
3	100	0	ZCTA5 00601	100	1
4	150	0	ZCTA5 00601	150	1
5	200	0	ZCTA5 00601	200	1
6	250	0	ZCTA5 00601	250	1
7	300	20	ZCTA5 00601	300	1
8	350	0	ZCTA5 00601	350	1
9	400	0	ZCTA5 00601	400	1
10	450	0	ZCTA5 00601	450	1
11	500	0	ZCTA5 00601	500	1
12	550	0	ZCTA5 00601	550	1
13	600	0	ZCTA5 00601	600	1
14	650	0	ZCTA5 00601	650	1
15	700	0	ZCTA5 00601	700	1
16	750	0	ZCTA5 00601	750	1
17	800	0	ZCTA5 00601	800	1
18	900	0	ZCTA5 00601	900	1
19	1000	0	ZCTA5 00601	1000	1
20	1250	0	ZCTA5 00601	1250	1
21	1500	0	ZCTA5 00601	1500	1
22	2000	0	ZCTA5 00601	2000	1
23	2500	0	ZCTA5 00601	2500	1
24	3000	0	ZCTA5 00601	3000	1
25	3500	0	ZCTA5 00601	3500	1
30	100	0	ZCTA5 00601	100	2
31	150	0	ZCTA5 00601	150	2
32	200	0	ZCTA5 00601	200	2
33	250	35	ZCTA5 00601	250	2
34	300	35	ZCTA5 00601	300	2
35	350	25	ZCTA5 00601	350	2
36	400	35	ZCTA5 00601	400	2
...	...	...	...	...	...
359848	1000	4	ZCTA5 19979	1000	2
359849	1250	0	ZCTA5 19979	1250	2

	0	num	zip	rent	bedroom
359850	1500	0	ZCTA5 19979	1500	2
359851	2000	0	ZCTA5 19979	2000	2
359852	2500	0	ZCTA5 19979	2500	2
359853	3000	0	ZCTA5 19979	3000	2
359854	3500	0	ZCTA5 19979	3500	2
359859	100	0	ZCTA5 19979	100	3
359860	150	0	ZCTA5 19979	150	3
359861	200	0	ZCTA5 19979	200	3
359862	250	0	ZCTA5 19979	250	3
359863	300	0	ZCTA5 19979	300	3
359864	350	0	ZCTA5 19979	350	3
359865	400	0	ZCTA5 19979	400	3
359866	450	0	ZCTA5 19979	450	3
359867	500	0	ZCTA5 19979	500	3
359868	550	0	ZCTA5 19979	550	3
359869	600	0	ZCTA5 19979	600	3
359870	650	0	ZCTA5 19979	650	3
359871	700	0	ZCTA5 19979	700	3
359872	750	0	ZCTA5 19979	750	3
359873	800	0	ZCTA5 19979	800	3
359874	900	20	ZCTA5 19979	900	3
359875	1000	0	ZCTA5 19979	1000	3
359876	1250	0	ZCTA5 19979	1250	3
359877	1500	0	ZCTA5 19979	1500	3
359878	2000	0	ZCTA5 19979	2000	3
359879	2500	0	ZCTA5 19979	2500	3
359880	3000	0	ZCTA5 19979	3000	3
359881	3500	0	ZCTA5 19979	3500	3

306567 rows × 5 columns

```
In [44]: rent3['zipcode'] = rent3['zip'].str[5:]
         rent3['index'] = rent3.index

         #final columns to import
```

```
In [45]: xport_excel = rent3.to_excel (r"C:/Users/Rosario Household/Downloads/rentdata.xl

         #extract new spreadsheet for dashboard analysis and report
```